Institut National de la Statistique et des Études Économiques

Insee méthodes

Méthodologie de la statistique annuelle d'entreprises

Description du système « Ésane »

Jean-Marc Béguin
Olivier Haag

Dédicace et remerciements

A la mémoire d'Emmanuel Raulin

Le projet « Resane[1] » de modernisation de la statistique structurelle d'entreprises fut une aventure à la fois technique et humaine, toujours passionnante, souvent stressante, quelquefois même dramatique quand son maître d'ouvrage délégué et grand inspirateur ***Emmanuel Raulin*** a dû abandonner la direction opérationnelle du projet suite à des problèmes de santé, au moment crucial du basculement en production (printemps 2009), avant de décéder quelques mois plus tard. Nombre des innovations introduites dans ce projet furent les siennes. **Cet ouvrage est dédié à sa mémoire**.

J'ai eu l'honneur et la chance de porter ce projet sur les fonts baptismaux et d'en être le maître d'ouvrage, de sa naissance institutionnelle en octobre 2004 avec la nomination de ***Raoul Depoutot*** comme directeur de projet, jusqu'à la dissolution formelle de l'équipe de projet restante fin août 2012.

L'ambition du présent document est d'exposer **les grands principes** qui ont présidé au développement du projet et **la méthodologie statistique** mise en œuvre dans chacun des sous-systèmes ou des procédures qui composent le système d'information final « Ésane », méthodologie fréquemment innovante par rapport à celle mise en œuvre dans le système d'information précédent. Il est destiné aux professionnels, aux chercheurs et aux utilisateurs des résultats qui voudraient comprendre, de façon approfondie, comment sont fabriqués les agrégats publiés, à partir de données individuelles issues à la fois de données administratives exhaustives et de données d'enquêtes statistiques, sur échantillon. Leur exploitation **simultanée** constitue à la fois **l'originalité, la richesse** et en même temps **la difficulté principale** dans l'élaboration des statistiques structurelles françaises.

Son champ est défini (sauf précision contraire) par ce qui était développé au moment de la dissolution de l'équipe de projet en août 2012 mais quelques évolutions notables implantées entre 2013 et 2016 ont également été explicitées. C'est pourquoi il est principalement consacré à ce qu'on a appelé **la phase 1** du projet Resane c'est-à-dire la **réingénierie des systèmes existants** (EAE, Suse et SIE) pour former le système d'information Ésane. La **phase 2** – la prise en compte de tous les groupes et le profilage – est abordée dans le *chapitre 11* mais centrée sur le traitement des très grands groupes ; elle n'aborde pas en détail les derniers développements informatiques réalisés pour les profileurs depuis 2013 ni bien entendu un certain nombre qui restent à faire pour traiter tous les autres groupes.

Je souhaite en tout cas qu'il serve de document de référence chaque fois qu'un utilisateur des résultats ou un acteur du système voudra **comprendre précisément comment un résultat a été élaboré**.

* *
*

Je souhaite *exprimer ma reconnaissance* et rendre *un hommage* tout particulier à ***Olivier Haag***, membre puis responsable en dernier de l'équipe de projet de 2010 et 2012, qui a compulsé toute la documentation existante pour réaliser avec moi cet ouvrage.

Le nombre de statisticiens impliqués de près ou de loin dans les méthodes décrites est considérable. *Qu'ils soient ici tous remerciés* ; mais je voudrais au moins citer les noms de tous ceux dont Olivier Haag a directement réutilisé les travaux, dans l'équipe de projet statistique Resane, (*E. Raulin et R. Depoutot* déjà cités, *T. Balcone, S. Chami, M. Coudene, C. Coudrin, T. Denoyelle, S. Gilles, A. Husseini-Skalitz, H. Koumarianos, M. Martin, M. Moret,*

[1] Pour « REfonte des Statistiques ANnuelles d'Entreprises »

G. Pavlov, W. Rostam, J. Seng) ou même en dehors de l'équipe (*C. Basso, M. Beauvois, P. Brion, A. Fizzala, E. Gros, J. Nicolas, F. Piot, L. Sauvadet, et T. Trouche*).

Ma gratitude va également aux collègues qui ont relu tout ou partie de l'ouvrage pour le rendre le plus exact, pédagogique et exhaustif possible : *P. Biscourp, D. Brondel, C. Couderc, L. de Gimel, J. Khelif, Y Le Gros, B. Maillot, H. Mariotte, V. Moyne, C. Rieg, A. Rochas, C. Réveillard, A. Smyk et J. Talbot*.

Je forme le vœu pour finir, peut-être le plus ambitieux, que **cet effort de documentation et de transparence sur nos méthodes se maintienne** afin que ce document reste fidèle au fonctionnement de l'application, même dans ses évolutions futures.

Jean-Marc Béguin
Directeur des statistiques d'entreprises
(Octobre 2003-février 2013)

Insee Méthodes Ésane : sommaire

DÉDICACE ET REMERCIEMENTS.....2

INSEE MÉTHODES ÉSANE : SOMMAIRE.....4

AVANT-PROPOS : PRÉSENTATION ET OBJECTIFS DU SYSTÈME ÉSANE.....11

1.PRÉSENTATION DES SOURCES ET DE L'ARCHITECTURE GÉNÉRALE DU MODE DE CONTRÔLE DES DONNÉES ÉSANE *[1.1], [1.7], [1.10]*.....17

1.1Les différentes sources en présence.....17

1.1.1L'enquête sectorielle annuelle (Esa) *[1.6], [1.9]*.....17

1.1.2L'enquête annuelle de production (EAP).....18

1.1.3Les données fiscales *[1.8]*.....18

1.1.4Les données d'emploi.....20

1.1.4.1La source Urssaf.....20

1.1.4.2La source DADS.....21

1.1.4.3Les données de « Clap ».....21

1.1.5Des sources complémentaires pour des unités spécifiques.....21

1.1.5.1Les données intégrées à Ésane (syndicats des eaux et PMU).....22

1.1.5.2Les données non intégrées à Ésane mais transmises aux comptes nationaux pour leur usage.....22

1.2La logique de contrôle de ces données multi-sources.....23

1.2.1Le référentiel : un champ défini a priori (cf. chapitre 4 pour une explicitation complète) 26

1.2.1.1La définition du champ.....26

1.2.1.2Ocsane : le « répertoire » d'Ésane qui permet la gestion de ce champ.....27

1.2.1.3Les difficultés engendrées par la définition d'un champ a priori.....28

1.2.2Des contrôles « intra-sources » indépendants d'une source à l'autre.....28

1.2.3Une réconciliation des sources validées ou contrôles inter-sources.....30

1.2.4Une mise en cohérence finale des données de chaque source.....30

1.2.5Les avantages et inconvénients de ce principe de contrôle.....30

2. PRÉSENTATION DES GRANDS PRINCIPES MÉTHODOLOGIQUES RETENUS POUR LE CONTRÔLE DES DONNÉES : LES CQR ET LES MACRO CONTRÔLES *[1.11], [3.1], [3.2], [3.5], [3.6], [3.7], [3.9]*.....31

2.1Les différentes occurrences d'une même caractéristique.....32

2.2Les micro-contrôles.....34

2.2.1Le principe des micro-contrôles.....34

2.2.2Calcul des paramètres des micro-contrôles:.....36

2.2.2.1Paramètres des CIV.....36

2.2.2.2Paramètres des CTV.....37

2.3Qualification (Esa).....37

2.4Redressement (Esa).....38

2.4.1Types de redressement.....38

2.4.1.1Les différents types de redressements.....38

2.4.1.2Cas des questions avec filtre.....39

2.4.1.2.1Cas où le montant est qualifié.....40

2.4.1.2.2Cas où le montant n'est pas qualifié.....40

2.4.1.2.3Mode de redressement spécifique des questions-filtres :.....40

2.5Ajustement.....40

2.6Messages de contrôle des données associés aux micro-contrôles.....41

2.7Les macro-contrôles [3.3], [3.4], [3.8].....41

2.7.1Macro-contrôles de type « contribution » (ou « drop out »).....42

2.7.2Le repérage des évolutions atypiques.....43

2.7.3Calcul des agrégats et des contributions.....44

2.7.3.1Agrégats sectoriels en niveau (Année N).....44

2.7.3.1.1Assiette.....44

2.7.3.1.2Formule de calcul de l'agrégat.....45

2.7.3.1.3Contribution pour les macro-contrôles contemporains....46
2.7.3.2Agrégats sectoriels en évolution (N / N-1)....47
2.7.3.2.1Assiette....47
2.7.3.2.2Formule de calcul....47
2.7.3.2.3Contribution....47
2.7.3.3Agrégats de branche en évolution (N et N−1)....47
2.7.4Macro-contrôles par différence....48
2.7.5La définition des seuils des macro-contrôles [3.2], [3.3]....49
2.7.5.1Les principes théoriques....49
2.7.5.2Mise en œuvre pratique....50
2.7.6Repérage des entreprises fortement contributrices....50
2.8Indicateur de priorité de traitement....51
2.9Les contrôles complémentaires aux macro contrôles....53

3LE CAS PARTICULIER DES ENVELOPPES DE RESTRUCTURATION....54
3.1Qu'est-ce qu'une enveloppe de restructuration et pourquoi la prendre en compte dans les statistiques ?....54
3.1.1Rappel sur les restructurations....54
3.1.2La présentation de l'enveloppe et de ses caractéristiques propres....55
3.1.3Les enveloppes et Ésane....59
3.2Les échanges Ésane – Citrus ; l'initialisation des enveloppes dans Ésane....59
3.2.1La création des unités dans l'Esa....60
3.2.2L'estimation du flux intra enveloppe et ses liens avec le coefficient d'agrégation....61
3.2.2.1La définition du flux intra enveloppe....61
3.2.2.2Le recalcul d'un coefficient d'agrégation....62
3.2.3Le calcul des variables de la liasse et du questionnaire d'une enveloppe (à partir de celles des unités légales)....64
3.3La prise en compte des enveloppes dès la phase de contrôle....64
3.3.1Les enveloppes : une nouvelle unité de contrôle....64
3.3.2Le travail des gestionnaires sur les enveloppes....65
3.3.2.1Contrôle des contours avant les macro-contrôles....65
3.3.2.2Contrôle des enveloppes par le biais des macro-contrôles....66
3.3.3L'investissement hors apport, une variable difficile à contrôler....66

4.LE CHAMP DU SYSTÈME D'INFORMATION ÉSANE....68
4.1Les critères utilisés pour la définition du champ....68
4.2Le champ géographique....70
4.3Les quatre sous champs de production....71
4.4La définition en extension du champ....73
4.4.1Sélection sur le code APE : les sections et les groupes de la NAF représentés....73
4.4.2Sélection sur la catégorie juridique (CJ) : les CJ exclues du champ....74
4.4.3Les activités couvertes par le sous champ 1....75
4.4.3.1Champ de la maîtrise d'œuvre Insee....75
4.4.3.2Répartition des maîtrises d'ouvrage de l'Esa par « grand secteur »....75
4.4.4Les activités classées dans les sous champs 2 ou 3....76
4.5Les différents champs de la statistique d'entreprise et de diffusion d'Ésane....76
4.6Les modifications de champ entre 2008 et 2010....79
4.6.1Campagne (ou année de constat) 2009....80
4.6.2Campagne (ou année de constat) 2010....80
4.6.2.1Les modifications apportées au champ de l'Esa....80
4.6.2.2L'introduction des syndicats des eaux....80
4.6.2.3Les autres modifications de champ et la création du sous champ « 5 »....80
4.6.2.4La comparaison 2010 – 2008/2009....81
4.7Les modifications de champ et de mode de diffusion, à compter de la campagne 2011....81
4.7.1Campagne 2011 / champ 2012....82
4.7.1.1Les modifications apportées au champ (prises en compte dans la publication 2012)....82
4.7.1.2La comparaison de champ entre les campagnes 2011 - 2010....82
4.7.2Campagne 2012 / champ 2013....83
4.7.2.1Les modifications apportées au champ....83

4.7.2.2La comparaison 2012-2011-2010.....83
4.8Les modifications de la définition de l'entreprise à compter de la campagne 2013 84
4.8.1Campagne 2013 / champ 2014.....84
4.8.2Campagne 2014 / champ 2015.....84
4.8.3Campagne 2015 / champ 2016.....84
4.8.4Campagne 2016 / champ 2017.....85
4.8.5Campagnes ultérieures.....85
5.LES PRINCIPES MÉTHODOLOGIQUES PROPRES AUX ENQUÊTES (ESA ET EAP).....86
5.1Présentation de l'Esa.....86
5.1.1Champ de l'enquête et questionnaire.....86
5.1.2Stratification.....87
5.1.3Pondération et allocation de l'échantillon *[2.5]*.....87
5.1.3.1Détermination des taux de sondage initiaux.....87
5.1.3.2Détermination des seuils d'exhaustivité.....90
5.1.4Renouvellement par moitié de l'échantillon :.....91
5.2Présentation de l'EAP [1.3].....93
5.2.1Le champ de l'enquête.....93
5.2.2Le plan de sondage jusqu'en 2013.....94
5.2.3Le questionnaire.....95
5.2.4Imputation et estimation des données absentes.....95
5.3La prise en compte des données de l'EAP dans le système de production Ésane. 97
5.3.1La prise en compte de la ventilation en branche de l'EAP.....97
5.3.2La prise en compte des changements d'APE.....97
5.3.3La diffusion des résultats des entreprises industrielles.....97
5.4Les contrôles gestionnaires spécifiques de l'Esa.....98
5.4.1Le rappel des grandes non-répondantes.....98
5.4.2La codification des branches non codées.....99
5.4.2.1Une codification automatique pour les cas à faible enjeu.....99
5.4.2.2Le redressement automatique des non-réponses partielles.....100
5.4.3La validation de l'APE.....100
5.4.3.1Les différentes APE présentes dans la chaîne de production Ésane.....100
5.4.3.2Le travail de contrôle de l'APE demandé au gestionnaire.....102
5.4.4La validation du cadre de restructuration.....103
5.5Le traitement de la non-réponse totale (provisoire) pour la base de production...103
5.5.1Le principe de l'extrapolation à partir des données N−1.....104
5.5.2Le principe de l'extrapolation par hot-deck.....104
5.6Le traitement de la non-réponse totale (définitive) pour la diffusion.....104
5.6.1Le cas particulier des DOM.....104
5.6.2Correction de la non-réponse totale dans les enquêtes.....105
5.6.2.1Correction de la non-réponse totale pour les entreprises appartenant à la partie exhaustive de l'échantillon.....106
5.6.2.2Correction de la non-réponse totale par repondération pour la partie échantillonnée de l'Esa⊕ EAP.....106
5.6.2.3Gestion des unités atypiques non aberrantes par « winsorisation » [2.6].....107
5.6.2.4Calage de la partie échantillonnée de l'Esa⊕ EAP.....108
6.LES PRINCIPES MÉTHODOLOGIQUES PROPRES AUX TRAITEMENTS DES DONNÉES COMPTABLES.....110
6.1Les deux phases de contrôle des données.....112
6.1.1Élaboration des résultats provisoires.....112
6.1.2Élaboration des résultats définitifs.....113
6.2La chaîne amont.....113
6.2.1Contrôle de structure.....113
6.2.2Traitements de mise en forme.....113
6.2.3Traitement des vrais doubles.....114
6.2.4Séparation socle / hors socle.....114
6.2.5Appariement avec Ocsane.....114

6.2.6Contrôle et redressement des dates déterminant l'exercice....115
6.2.7Découpage - collage....116
6.2.7.1Définition de l'année de référence dite « principale » d'une liasse....118
6.2.7.2Année de référence dite « secondaire »....121
6.2.7.3Opérations de découpage / collage des liasses....122
6.2.7.4Conséquences du découpage collage sur les caractéristiques de la liasse....131
6.2.7.5Bilan de l'étape de découpage collage (avant dilatation-contraction ou pseudo consolidation)....135
6.2.8Micro contrôles internes....137
6.2.8.1La logique des contrôles et redressements automatiques de cette phase....137
6.2.8.2Le principe des contrôles internes....137
6.2.9Les RQSI (redressements quasi sûrs internes)....138
6.2.9.1La logique des RQSI : une succession de modules....139
6.2.9.2Essai : on teste une hypothèse....139
6.2.10Les RIA (redressements internes automatiques)....140
6.2.10.1Les deux types de cohérence rétablies par les RIA : la cohérence entre les comptes et la cohérence au sein d'un compte....140
6.2.10.2Les méthodes de redressements....140
6.3La partition noyau / hors noyau : le champ des entreprises potentiellement contrôlables manuellement....142
6.4La chaîne aval....142
6.4.1Identification des liasses....143
6.4.2Remise à 12 mois des liasses....143
6.4.3Traitement des liasses multiples....143
6.4.4Traitement des dates....144
6.4.5Traitement des restructurations....144
6.4.6Traitement des macro-contrôles....145
6.5Le traitement de la non-réponse totale [2,13]....145
6.5.1Alimentation des établissements publics (cf. § 1.1.5)....146
6.5.2Détermination des entreprises économiquement actives....146
6.5.3Imputation des entreprises économiquement actives 147

7.LES PRINCIPES DE PRISE EN COMPTE ET DE CONTRÔLE DES DONNÉES D'EMPLOI....149
7.1Les différents concepts et sources utilisées....149
7.1.1Les concepts d'emploi....149
7.1.2Les sources utilisées....150
7.2Les données structurelles d'emploi : la collaboration avec le pôle Clap (emploi en nombre de postes)....154
7.2.1Les échanges Clap Ésane....155
7.2.1.1Rappel sur les traitements faits par le pôle Clap (Connaissance Locale de l'Appareil Productif)....155
7.2.1.2Les principes de la collaboration Clap - Ésane....155
7.2.2Les traitements spécifiques faits dans Ésane....157
7.2.2.1Prendre en compte les éventuels problèmes d'identification repérés par Clap dans la source DADS-U....158
7.2.2.2Passer des concepts Clap (DADS-U ou Urssaf) à des concepts liasse fiscale....158
7.2.2.3Cas particulier des entreprises d'intérim....162
7.2.2.4Estimer l'emploi des enveloppes et des entreprises profilées....163
7.2.2.5 Lister les entreprises à envoyer au pôle Clap....163
7.2.2.6Traiter la non-réponse pour les données d'emploi en nombre de postes....163
7.3Les données structurelles d'emploi : la fourniture des DADS (emploi en ETP).....164
7.3.1La fourniture des DADS-U à Ésane ; données d'emploi dites secondaires....164
7.3.1.1Variables DADS « trimestrialisées » fournies....164
7.3.2Traitements faits dans Ésane sur les données DADS....166
7.3.2.1Problèmes de champ....166
7.3.2.2Calage sur la durée d'exercice....166
7.3.2.3Estimation de la non-réponse sur les caractéristiques DADS....166
7.4Les données conjoncturelles d'emploi (Épure2)....166

8.LES PRINCIPES MÉTHODOLOGIQUES MIS EN ŒUVRE POUR LA RÉCONCILIATION DES DONNÉES INDIVIDUELLES (RÉDI).....169
8.1La cinématique générale et la présentation du chapitre.....169
8.2La logique de la réconciliation des données individuelles ; interférence avec le calendrier.....170
8.3Les principes méthodologiques de la mise en cohérence des données du chiffre d'affaires et de sa ventilation.....171
8.3.1Construction de la ventilation agrégée en trois postes pour l'Esa.....171
8.3.2Le traitement des branches mixtes du commerce.....172
8.3.2.1Branches commerciales dans l'ESA : uniquement mixtes.....172
8.3.2.2Branches commerciales dans l'ESA : mixtes et non mixtes.....173
8.3.3Le choix des sources prioritaires pour le chiffre d'affaires et sa ventilation.....174
8.3.3.1Pour le chiffre d'affaires.....174
8.3.3.2Pour la ventilation du chiffre d'affaires.....175
8.3.4Ajustements.....178
8.3.5Le redressement des achats.....180
8.3.6Le redressement de la ventilation du chiffre d'affaires et de sa décomposition des ventes selon les modalités France et exportation.....181
8.3.7Le redressement du compte de résultat (suite à modification du chiffre d'affaires de la liasse).....185
8.3.8Rappel sur le principe de contrôle des données par les gestionnaires.....186
8.4Les principes méthodologiques de la mise en cohérence de l'emploi.....188
8.4.1Les sources prioritaires dans « Rédi-emploi ».....188
8.4.1.1Pour les rémunérations.....188
8.4.1.2Pour les effectifs (en nombre de postes et ETP).....189
8.4.2Les redressements des variables d'emploi dans Rédi.....190
8.4.2.1Redressement des effectifs lorsque l'effectif de la liasse fiscale a été jugé prioritaire.....190
8.4.2.2Redressement de la non-réponse sur l'effectif ETP.....190
8.4.2.3Redressement de l'effectif ETP.....191
8.4.3Le redressement du compte de résultat (suite à modification des rémunérations). .191
8.5Les liens entre Rédi et les sources initiales : le cas des données d'enquêtes et d'emploi.....191
8.6La nécessité d'un retraitement final (dit « Rédi balai »).....194

9.LE CALCUL ET LA DIFFUSION DES AGRÉGATS FINAUX.....195
9.1La problématique générale de l'estimation des résultats [1.4], [2.1], [2.2], [2.3], [2.4], [2.11], [2.16].....195
9.2Les principes méthodologiques retenus pour le calcul des agrégats diffusés [2.12], [2.16], [3.10], [3.11].....197
9.2.1Les estimations de niveau groupe & supra des agrégats fiscaux et d'emploi.....201
9.2.2Les estimations de niveau infra groupe (sur un domaine D) des agrégats fiscaux et d'emploi.....201
9.2.3Influence du calage sur les résultats des agrégats fiscaux.....203
9.2.4Comparaison de la précision de différents estimateurs [2.16].....203
9.2.5Cas particulier de statistiques ventilant une estimation sectorielle obtenue sur une variable administrative selon des clés de répartition issues de l'Esa – en particulier calcul des agrégats branches.....205
9.2.6Cas particulier des agrégats sectoriels des variables spécifiques à l'Esa.....208
9.3Les limites de ces estimateurs et les précautions à prendre dans leur usage [2.9] 209
9.3.1Le caractère sommable des agrégats ne peut pas être assuré systématiquement à tous les niveaux.....209
9.3.2La somme des données individuelles ne redonne pas la valeur de l'agrégat.....210
9.3.3Certains agrégats en niveau (comme le nombre d'entre-prises) sont plus variables que par le passé.....211
9.4La diffusion des résultats d'Ésane.....211
9.4.1Le calendrier de diffusion.....212
9.4.2La diffusion préliminaire pour Eurostat.....212
9.4.3La diffusion définitive sur le site de l'Insee.....213

9.4.3.1Les chiffres clés..... *213*
9.4.3.2Les chiffres détaillés..... *214*
9.4.3.2.A.Résultats sur les entreprises.....215
9.4.3.2.B.Les branches d'entreprises.....217
9.4.3.2.C.Alisse..... 218
9.4.3.2.D.Données N-1 pour comparaison.....218
9.4.3.2.E.Résultats sectoriels publiés dans Insee Référence.....219
9.4.4La création du fichier de données individuelles.....219
9.4.5La publication Eurostat.....219
9.4.6La publication de l'EAP.....220
9.5La pose du secret.....221
9.5.1Problématique.....221
9.5.2L'estimateur utilisé pour poser le secret pour les variables fiscales et pour un secteur X donné.....222
9.5.3L'estimateur utilisé pour poser le secret des variables fiscales ventilées selon une variable sectorielle de l'Esa.....222
9.5.4L'estimateur utilisé pour poser le secret pour une variable uniquement présente dans l'Esa.....223
9.6Les précautions à prendre dans l'analyse des résultats en évolution [2.7],.....223
9.6.1Présentation des différents types d'évolutions possibles.....223
9.6.1.1L'évolution apparente..... *224*
9.6.1.2L'évolution économique « en comparable »..... *224*
9.6.2La décomposition de l'évolution en comparable par population.....225
9.6.2.1Les principes généraux..... *225*
9.6.2.2La décomposition de l'évolution en comparable..... *227*
9.6.2.3La contribution des différentes sous populations à considérer..... *228*
9.6.2.4Évolution, poids et contribution..... *230*
9.6.2.5Les précautions à prendre dans l'interprétation des résultats..... *231*

10.PRÉSENTATION DE LA PHASE DE VÉRIFICATION DES DONNÉES ET DE VALIDATION DES AGRÉGATS.....233
10.1Les grands principes de cette phase.....233
10.2La mise en œuvre de la démarche qualité.....234
10.2.1Les entreprises à vérifier par le service Ésane (à Nantes).....234
10.2.2Les entreprises à vérifier par les méthodologues (parisiens).....236
10.3La nécessité d'un contrôle final des données individuelles et des agrégats.....236
10.4Le contrôle sur données individuelles.....237
10.4.1Valeurs douteuses en sortie de Rédi.....238
10.4.2Le cas des entreprises profilées issues des grands groupes.....238
10.4.3Recherche d'autres types d'anomalies.....239
10.5Le contrôle des agrégats.....239
10.5.1L'analyse des taux de variation par « output editing ».....239
10.5.1.1Un exemple de tableau global d'analyse de la croissance..... *240*
10.5.1.2Un exemple de tableau détaillé d'analyse de la croissance..... *242*
10.5.2L'analyse de la vraisemblance des agrégats.....243
10.5.2.1Analyse de l'impact de chaque sous-population sur un agrégat..... *243*
10.5.2.2Analyse de l'impact des processus sur un agrégat..... *246*

11.RESANE « PHASE 2 » : LA PRISE EN COMPTE DES GROUPES ET DE L'ENTREPRISE AU SENS ÉCONOMIQUE DANS LE SYSTÈME DE PRODUCTION ÉSANE 248
11.1Pourquoi profiler ?.....248
11.2Qu'est ce qu'une entreprise et une entreprise profilée ?.....250
11.3Qu'est ce qu'un groupe ? Le système Lifi.....251
11.4La stratégie de profilage décidée par l'Insee.....252
11.5Le traitement des entreprises profilées de la cible 1 par les profileurs.....256
11.5.1La création des périmètres des entreprises profilées.....256
11.5.2Les différents types de consolidation des données fiscales envisagés.....257
11.5.3Le cas des co-entreprises.....259
11.5.4La collecte des réponses aux enquêtes Esa ou EAP.....259

11.5.5La collecte des réponses aux autres enquêtes de la statistique publique.....260
11.6La prise en compte des entreprises de la cible 1 dans le système Ésane.....260
11.6.1La gestion des périmètres dans Ocsane.....260
11.6.2La prise en compte des questionnaires Esa ou EAP.....261
11.6.3L'intégration des liasses fiscales calculées.....261
11.6.4Le calcul des données d'emploi.....261
11.6.5Une réconciliation traditionnelle.....261
11.6.62012 : année de transition et première année de prise en compte des résultats de la cible 1.....261
11.7Le répertoire SIRUS.....262
11.7.1Les principales fonctions du répertoire *[1.2], [1.12]*.....*262*
11.7.2Les liens de Sirus avec les autres systèmes d'information.....265
11.7.3La continuité des groupes et des entreprises.....267

12.DÉROULEMENT CHRONOLOGIQUE D'UNE CAMPAGNE.....270

ANNEXES.....275

13.ANNEXE 1 : PRÉSENTATION GÉNÉRALE DU CALCUL DE L'APE.....276
13.1Principe général de détermination du code APE.....276
13.2Les spécificités liées au commerce.....278
13.2.1Le cas particulier du commerce de gros.....278
13.2.2Le cas particulier du commerce de détail.....279
13.2.3La détermination de la ventilation du chiffre d'affaires en branches.....279

14.ANNEXE 2 : LES TRAITEMENTS DADS ; POSTES ANNEXES ET NON ANNEXES ; VOLUME DE TRAVAIL ; NOTION D'ETP.....280
14.1La définition des postes et le concept de postes annexes.....280
14.2Le volume de travail et le concept d'emploi ETP.....281

15.ANNEXE 3 : BIBLIOGRAPHIE.....283

16.ANNEXE 4 : GLOSSAIRE.....285

Avant-propos : présentation et objectifs du système Ésane

Ésane (qui signifie « élaboration des statistiques annuelles d'entreprise ») est le **système d'information** qui permet d'élaborer les statistiques structurelles (c'est-à-dire annuelles) d'entreprises françaises, **à destination** à la fois des autorités politiques et administratives françaises, de la Commission européenne (Eurostat[2]), des statisticiens français et en particulier des comptables nationaux, des professionnels et partenaires sociaux du monde de l'entreprise, des média ou enfin du grand public, donc pour des **besoins nombreux, variés et différents,** relayés notamment au sein du Cnis[3]). Il sert en particulier à :

- définir et caractériser les notions et **contours des « secteurs[4] » et des « branches[5] »** ;
- mettre à jour **le code APE** des unités légales interrogées dans les enquêtes liées à Ésane ;
- calculer **la part du PIB** attribuable aux entreprises ;
- fournir des **dénombrements de sous populations** d'entreprises ;
- caractériser ces sous populations par de nombreuses **variables (i) ou ratio comptables (ii)** qu'on peut tirer des comptes sociaux des entreprises :
  i. chiffre d'affaires(CA), excédent brut d'exploitation (EBE), valeur ajoutée (VA), résultat comptable, immobilisations, marge commerciale, emploi total, emploi salarié, emploi non salarié, etc. ;
  ii. CA par salarié, VA par salarié, salaire brut annuel par tête, immobilisations par salarié, total de bilan par salarié, taux de marge brute (EBE / VA), taux d'exportation, rentabilité d'exploitation (EBE / capital engagé), investissement corporel (hors apports) / VA, etc. ;
- **décrire certaines sous populations** par des variables propres à certains secteurs seulement (surfaces de magasins pour des entreprises de commerce de détail, la répartition du chiffre d'affaires par type d'ouvrages dans le BTP, etc.).
- **répondre au règlement européen** dit « SBS » (Voir ci-dessous note 2) qui impose de fournir un certain nombre d'agrégats économiques et comptables relatifs aux entreprises françaises.

Il repose sur le concept « d'entreprise » et présuppose donc que chacun sait **ce qu'est une entreprise**. En réalité il s'agit d'une notion **délicate à définir**, différente selon que l'on adopte un point de vue économique, comptable, juridique ou statistique. Disons pour simplifier[6] que dans toutes les statistiques françaises jusqu'à l'existence d'Ésane, l'entreprise est synonyme de « l'unité légale » et recouvre donc un ensemble de **sociétés et entreprises individuelles** qui sont enregistrées dans Sirene. Ésane a été l'occasion de remplacer cette définition, essentiellement juridique, par une notion **économique** d'entreprise, faisant appel aux groupes. Les premiers résultats qui en tiennent compte ont été publiés à l'été 2015, relativement aux années 2013 et 2012. Il existe donc désormais des **entreprises, entendues comme acteurs économiques indépendants, distinctes des unités légales[7] et formées de plusieurs unités légales.**

---

[2] Dans le cadre d'un règlement européen de 1997 dit « SBS » (pour « Structural Business Statistics) n° 58/97 ; ce règlement a été plusieurs fois amendé depuis et Eurostat prépare un nouveau cadre réglementaire, dit « FRIBS », pour l'ensemble de la statistique d'entreprises européenne qui devrait être voté aux alentours de 2019. La définition des unités statistiques utilisées (entreprises, établissements ou groupes) est l'objet du règlement 696/93 et devrait rester inchangée. Il y sera fait référence au *chapitre 11*.

[3] Conseil national de l'information statistique, instance de concertation entre les producteurs et les utilisateurs de la statistique publique.

[4] Ensemble des unités légales qui ont la même activité principale (repérée par le code dit APE de l'unité légale).

[5] Ensemble des parties d'unités légales qui contribuent à une activité donnée, que ce soit l'activité principale ou une activité secondaire de l'unité légale. Le système Ésane sert, entre autres, à définir l'ensemble des activités secondaires des unités légales directement pour celles qui appartiennent à l'échantillon de l'enquête structurelle ou par estimation pour les autres. Un schéma décrivant ces notions est présenté au paragraphe *2.7.3.3*.

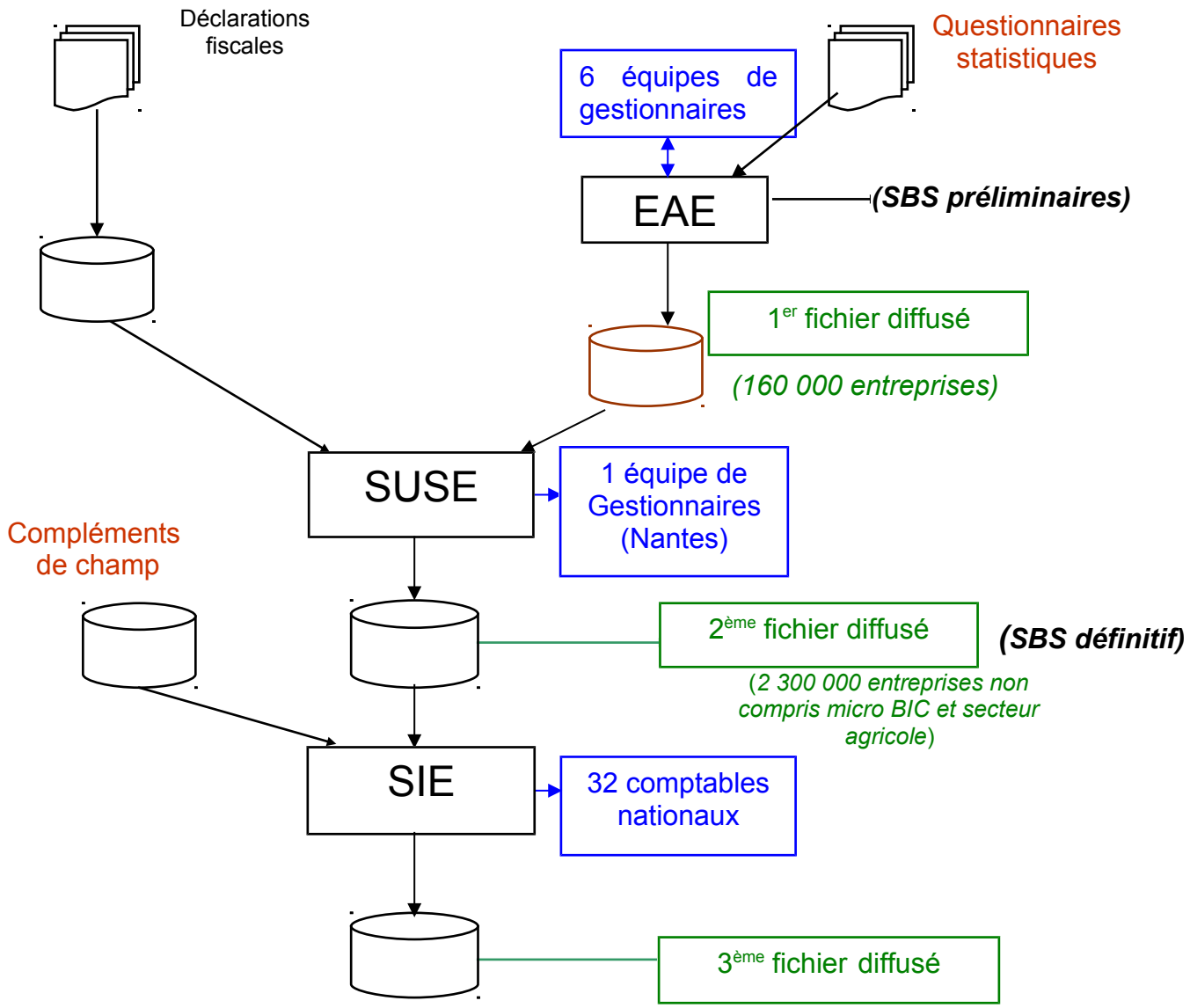
[6] Un rappel historique plus détaillé figure en introduction du chapitre 11 et plus particulièrement dans son § *11.1*. Le rapport du Cnis dont la référence est donnée dans la note 11 ci-après débat longuement de la définition de l'entreprise.

[7] Les unités légales indépendantes (non contrôlées par un groupe) continuent naturellement d'être considérées comme des entreprises.

Ésane correspond à la **quatrième ou cinquième génération de ce système d'information**. Le système précédent était en réalité constitué de **trois** parties distinctes et autonomes, donc ayant chacune une maîtrise d'ouvrage séparée, une chaîne d'exploitation informatique en propre avec des gestionnaires dédiés et une diffusion autonome de leurs résultats :

- Les **EAE** (enquêtes annuelles d'entreprises) qui en étaient à leur quatrième génération ;
- **SUSE** (système unifié de statistiques d'entreprises) qui correspondait à l'exploitation des données fiscales et dont c'était la troisième génération ;
- Le **SIE** (système intermédiaire d'entreprises) en aval de SUSE, plus récent, qui était spécifiquement destiné à la comptabilité nationale (et dont c'était la première version).

Il correspondait à l'architecture stylisée[8] suivante :



Pour un secteur donné de la nomenclature d'activités, la statistique publique publiait ainsi par exemple trois chiffres d'affaires, **distincts et généralement peu cohérents**, à l'issue de chacun de ces trois processus.

[8] Dans le schéma qui suit, les dessins qui figurent des fichiers représentent des fichiers « logiques » ; mais physiquement ceux-ci peuvent se présenter sous une forme beaucoup plus fragmentée : par exemple les fichiers fiscaux représentent des dizaines de fichiers différents qui sont livrés séparément par les centres informatiques des impôts à des moments différents et qu'il faut donc concaténer, supprimer les enregistrements en double etc. Il s'agit donc bien d'une représentation logique simplifiée.

Ésane résulte ainsi de la **refonte** simultanée de ces trois systèmes dans le cadre d'un **grand projet** de l'Insee, **le programme**[9] **Resane**[10] (Refonte de la statistique annuelle d'entreprise), qui s'est déroulé **sur sept ans**, de **septembre 2005 à août 2012**. Ce programme était découpé en **deux phases** : une *première phase de réingénierie* proprement dite des trois systèmes qu'on vient de nommer, à concepts constants, fondée sur l'unité légale comme unité statistique de base du système, identifiée à l'entreprise ; et une *deuxième phase,* fondée comme on vient de le voir sur une *évolution du concept d'entreprise* - entendue désormais comme « acteur économique » - visant à prendre en compte les groupes dans la statistique d'entreprises, au travers du « profilage » des groupes.

Cette deuxième phase a nécessité la création en 2010 d'une nouvelle division à l'Insee et d'un nouveau métier de statisticien : celui de **profileur**. Beaucoup moins centrée sur l'informatique, cette deuxième phase a donc démarré en septembre 2010 et devrait durer jusqu'à la prise en compte de ce nouveau concept d'entreprises dans la comptabilité nationale, dont le calendrier reste à fixer. Ses fondements conceptuels en sont décrits dans le **chapitre 11** de cet ouvrage[11], même si seul le traitement des très grands groupes y est décrit. Le traitement des groupes restants fait l'objet d'un **autre projet** dont la mise en production devrait se situer en 2018.

Les **objectifs de la première phase** de réingénierie étaient de :

- **diminuer la charge** statistique de réponse des entreprises (celle-ci a été divisée environ par deux) ;
- **rationaliser les moyens** consacrés par la statistique publique à cette opération (ceux-ci ont été également divisés environ par deux) ;
- **changer de stratégie de publication** en passant d'une logique d'applications séparées exploitant une source de données (les EAE, SUSE ou le SIE) à une logique de « produits », issus d'un même système d'information, produisant un résultat unique pour chaque variable du système (un seul chiffre d'affaires pour un secteur donné par exemple) ;
- **raccourcir les délais de diffusion** : seul cet objectif n'a pas été atteint à la hauteur des espérances initiales : on pensait pouvoir publier en fin d'année N+1 ; on arrive à transmettre à cette date les données de cadrage dont a besoin la comptabilité nationale ainsi que les résultats dits « préliminaires » à Eurostat à fin octobre N+1, mais la diffusion publique de l'ensemble des tableaux et données élaborées ne peut pas avoir lieu avant le mois de juin N+2.

Chacun de ces objectifs a conditionné bien entendu les caractéristiques du système. Celui-ci comprend à la fois, comme tout système d'information, **des données** issues de **sources**, **des programmes de traitement** qui s'appliquent à ces données et **des acteurs** qui les ont définis d'une part (la maîtrise d'ouvrage) et qui les mettent en œuvre d'autre part (la maîtrise d'œuvre).

Les **maîtrises d'ouvrage (dites sectorielles car responsables d'un grand secteur)** sont restées les mêmes que précédemment, des départements de l'Insee ou des services statistiques ministériels. Mais alors qu'auparavant ces maîtrises d'ouvrage assuraient également la maîtrise d'œuvre des EAE, une des grandes nouveautés du système est la ***centralisation de la maîtrise d'œuvre à l'Insee**** avec deux équipes de gestionnaires (contre huit en tout auparavant) situées à **Caen** pour l'industrie (hors IAA) et à **Nantes** pour tous les autres secteurs. Cette réorganisation territoriale a contribué grandement à la rationalisation des moyens évoquée plus haut.

La deuxième grande innovation architecturale est liée à **l'unicité du « système » d'information** : une même base de production (*la base Ésane*) est alimentée par toutes les sources du système (voir le schéma correspondant au § *1.1.5.2*). La nécessité de réduire la charge de réponse des entreprises a conduit à réduire fortement le questionnement statistique des entreprises : les anciennes EAE ont donc été remplacées par de nouvelles enquêtes, plus légères : l'enquête annuelle de production (EAP) dans l'industrie manufacturière (hors IAA) et les enquêtes sectorielles annuelles (ESA) pour les autres secteurs. Cet ensemble forme ce qu'on appelle **la « source enquête »** du système. Environ 150 000 entreprises sont ainsi interrogées chaque année. La réduction du volume d'informations collectées par enquêtes a été rendue possible par la généralisation du recours aux sources administratives dont le

[9] On appelle fréquemment "programme", de grands projets qui sont souvent constitués de plusieurs projets plus ou moins indépendants qui concourent à la réalisation du même système d'information final.

[10] Il ne faut pas confondre «Resane » – qui est le nom d'un « projet », opération limitée dans le temps et qui a été réalisée par une équipe de projet statistique et une équipe de projet informatique – avec Ésane qui est le nom du système d'information qui résulte de ce projet. Bien entendu ce système n'est pas figé et continue d'évoluer.

[11] On peut en trouver également une description très détaillée dans le rapport d'un groupe de travail du Cnis n° 107 : « Statistiques structurelles fondées sur les groupes d'entreprises et leurs sous-groupes » accessible ci-après : *http://www.cnis.fr*

cœur forme ce qu'on appelle « **la source fiscale** » (même si derrière se cachent des régimes fiscaux différents et plusieurs dizaines de fichiers) et correspond à l'ancienne application SUSE, mais on a aussi recours aux données sur les cotisations sociales et les effectifs des URSSAF et aux déclarations annuelles de données sociales (DADS) pour l'emploi, ainsi qu'aux déclarations de TVA dans l'industrie ; des sources complémentaires en provenance de la DGFiP, pour des entreprises particulières (régies des eaux, CROUS, etc..) viennent compléter l'ensemble.

Les sources administratives fournissent ainsi des informations sur près de 5 millions de sociétés ou d'entreprises individuelles[12]. L'originalité du système repose ainsi sur cette double alimentation, exhaustive pour les sources administratives mais complétées par des enquêtes sur échantillon, avec **une phase de réconciliation individuelle des données** pour les unités qui sont présentes des deux côtés. Une fois les données manquantes estimées, les données jugées fausses ou incohérentes redressées, c'est cette phase de réconciliation, **totalement nouvelle et caractéristique d'Ésane** qui permet de fabriquer les « produits de diffusion » mis à disposition du public sur le site Insee.fr (chiffres clés, tableaux agrégés ou détaillés, fiches « sectorielles », etc.).

Ces publications peuvent se résumer, pour le niveau le plus fin, à des agrégats **au niveau sous-classes** (quatre chiffres et une lettre) de la nomenclature d'activité française (Naf) et à des agrégats **au niveau groupes** (trois chiffres) **croisés avec des tranches d'effectifs** (sur quatre tranches). C'est sur ces niveaux d'agrégation de la nomenclature que l'Insee s'est engagé, en particulier vis-à-vis du Cnis[13], en termes de qualité et c'est donc **sur ces niveaux** qu'ont été optimisés les contrôles qualité opérés par les gestionnaires de la direction régionale des Pays de la Loire à Nantes.

* *

L'objet du présent ouvrage est de **décrire les données et les principes méthodologiques** qui sont derrière tous les traitements automatiques ou semi-automatiques mis en œuvre par les équipes de gestionnaires de Caen et de Nantes, ainsi que par la division Ésane située à la direction générale de l'Insee : à quoi et comment juge-t-on qu'une donnée est correcte ou non ? Dans quels cas réinterroge-t-on l'entreprise ? Comment fonctionnent les programmes d'estimation ? Comment calcule-t-on les résultats finaux alors qu'on ne dispose pas de la même information sur toutes les entreprises ? etc. :

- La présentation et la **description des sources utilisées**, **les principes « généraux » des contrôles** effectués sur ces sources ainsi que **l'architecture générale du système** font l'objet du ***chapitre 1*** ; on plante le décor en quelque sorte.
- Le ***chapitre 2*** détaille **les principes du contrôle** des sources présentées au chapitre 1 : on explique la distinction entre valeur brute / modifiée / redressée d'une caractéristique et la différence entre les micro-contrôles et les macro-contrôles.
- Les unités statistiques sur lesquelles portent ces contrôles et redressements sont les « entreprises » qui sont le plus souvent les unités légales du système[14] ; mais ces unités légales ne sont pas stables dans le temps. Elles font l'objet de « restructurations » (transferts ou cessions d'actifs entre unités légales) qui viennent gêner la comparaison temporelle de leurs caractéristiques, lesquelles fondent une grande partie des contrôles exposés au chapitre 2. On est donc amené à définir dans le ***chapitre 3*** une notion nouvelle : les **« enveloppes de restructuration »**, au sein desquelles ces comparaisons temporelles restent possibles. Ces unités statistiques sont également utilisées dans le cadre des estimations des évolutions temporelles des agrégats.
- Le ***chapitre 4*** est consacré à la description du **« champ »** du système : chacun se « représente » ce qu'est une entreprise ; pourtant la multiplicité des statuts juridiques et des unités enregistrées dans le répertoire Sirene (8,5 millions à fin 2010, plus de 10 millions en 2016) fait qu'il n'est pas simple de décider quelles unités légales doivent être prises en compte

[12] C'est le succès du statut d'auto-entrepreneur qui explique la forte croissance de ce nombre. Au démarrage d'Ésane en 2008 (avant l'existence de ce statut, on était à un peu plus de 3 millions).

[13] Extrait de la fiche de présentation de l'enquête sectorielle annuelle (Esa) pour examen d'opportunité (Cf. réunion de la Commission « statistiques d'entreprises » du Cnis du 16 octobre 2007) : « Il faut souligner que les produits standards de diffusion ne croiseront plus le niveau d'activité le plus fin (5 chiffres de la Naf rév. 2) avec les classes de taille, cette finesse ne devant plus être accessible à l'avenir que sur demande (des outils seront développés en conséquence dans le cadre du programme Resane) »
http://www.cnis.fr

[14] Comme on l'a vu plus haut, avant la prise en compte des groupes et de la notion d'entreprise au sens économique à compter de la campagne 2012, et qui est exposée au *chapitre 11*, les entreprises étaient exclusivement des unités légales. A compter de 2012 et au delà, les entreprises peuvent être soit des unités légales, soit des ensembles d'unités légales. Les contrôles dont il est question s'appliquent bien aux entreprises, quelle que soit leur composition. En revanche, les « restructurations » ne portent bien que sur les unités légales.

dans le système et lesquelles ne le doivent pas ; faut-il prendre en compte les associations par exemple ? Ce chapitre répond à ces questions et détaille également les évolutions marginales qui se produisent à chaque campagne, jusqu'à l'introduction des entreprises profilées à partir de l'année de constat 2013.

- Les chapitres suivants décrivent précisément **les traitements** effectués respectivement sur **la source enquête**, (***chapitre 5***) (plan de sondage, estimation de la non-réponse, validation du code APE, winsorisation et calage, etc.), sur **la source fiscale** (***chapitre 6***) (les dix étapes de la chaîne amont avec la définition du socle et du hors socle, le découpage collage des liasses pour se ramener à des déclarations sur l'année civile, les RQSI[15] et les RIA[16], les entreprises du noyau et du hors noyau, la chaîne aval, l'estimation des liasses manquantes) et enfin sur les **sources de données d'emploi** (***chapitre 7***) (échanges et collaboration avec la source dite Clap ; alimentation directe par les DADS ; données conjoncturelles des Urssaf).
- Après avoir traité indépendamment chacune des sources, il s'agit de **« *réconcilier* »** les quelques données communes à plusieurs sources (chiffres d'affaires et emploi essentiellement) et ainsi de disposer, pour une unité donnée, d'un ensemble cohérent de ses caractéristiques quelle qu'en soit la provenance (statistique ou administrative). Les principes retenus font l'objet du ***chapitre 8***.
- Une fois qu'on dispose de données individuelles qu'on estime fiables et complètes, il faut bien sûr **calculer des agrégats** pour les **diffuser**. Le ***chapitre 9*** présente ainsi le principe des estimateurs composites d'Ésane (dont la formule de calcul est différente selon le niveau de nomenclature auquel on s'intéresse), la pose du secret statistique et se termine par la présentation des tableaux diffusés sur le site Internet de l'Insee. Ce chapitre contient quelques formules mathématiques non triviales.
- Dans le cadre d'une **démarche qualité** appliquée à l'ensemble du processus, le ***chapitre 10*** explique comment on **vérifie** le travail de contrôle des données effectué par les gestionnaires à Nantes, vérification qui s'étend également à la **validation des résultats agrégés** présentés au chapitre 9.
- Les principes de **prise en compte des groupes** dans la statistique structurelle d'entreprises constituent, comme on l'a vu plus haut, le ***chapitre 11***. On y présente la définition du « profilage » des groupes, les trois cibles de groupes retenues par l'Insee et la stratégie de traitement adaptée à chacune d'entre elles. C'est aussi l'occasion de décrire le **répertoire statistique SIRUS** qui a été développé à l'occasion du programme Resane et qui permet désormais de piloter l'ensemble des enquêtes de la statistique publique auprès des entreprises.
- L'ouvrage se termine par la présentation du **calendrier d'une campagne** dans le ***chapitre 12*** qui permet de voir que de nombreuses tâches « de production statistique » se déroulent en parallèle ; on y fait le lien lorsque c'est possible, avec les paragraphes de l'ouvrage qui décrivent la méthodologie sous-jacente aux tâches correspondantes.
- Deux ***annexes*** viennent compléter certains chapitres pour présenter des concepts utiles à la compréhension de la chaîne mais indépendants d'Ésane ***(annexe 1*** sur le ***calcul du code APE*** et ***annexe 2*** sur la **définition des « postes annexes *» dans le calcul de l'emploi*** et les traitements effectués dans les DADS) ; l'***annexe 3*** présente en ***bibliographie*** la quarantaine d'articles auxquels renvoient les indications figurant entre crochets dans les intitulés des chapitres ou des paragraphes. Ils ont été présentés dans des colloques ou publiés à l'occasion des réflexions autour du programme Resane**.**

Ésane constitue au final un système assez complexe, dont la maîtrise est en plus quelquefois obstruée par le recours à de nombreux sigles ou abréviations ; ceux-ci sont généralement expliqués dans le texte mais on a pensé qu'un **glossaire** pouvait également permettre à un lecteur novice de mieux s'y retrouver. Celui-ci se trouve dans **l'*annexe 4*.**

Chaque chapitre commence par une présentation du contenu et du plan du chapitre qui décrit l'objet de chaque paragraphe. Même si le lecteur ne s'intéresse qu'à un point particulier du chapitre, il est conseillé de lire cette présentation pour comprendre la logique globale du chapitre.

Le système Ésane se caractérise par la mise en œuvre d'un grand nombre d'**innovations** par rapport aux pratiques antérieures dont beaucoup ont été présentées au Cnis ou dans des colloques nationaux ou internationaux. Ces innovations ont touché tous les domaines, que ce soit celui de **l'organisation** avec la mise en place d'une démarche d'assurance qualité pour les équipes de gestionnaires nantais

[15] Redressements quasi sûrs internes
[16] Redressements internes automatiques

ou celui de **l'ergonomie** avec la mise à disposition de deux écrans par gestionnaire ou l'inclusion de la gestion de tous les courriers au sein même de l'application.

Sur le plan de la seule **méthodologie statistique**, une liste indicative en est donnée dans l'encadré ci-dessous avec la mention du paragraphe qui en traite dans l'ouvrage. On a cherché à reprendre puis à généraliser les innovations qui avaient pu être mises en place précédemment dans la statistique d'entreprises, tant au Sessi pour la gestion des enquêtes industrielles [**traitement des restructurations juridiques** (6) ou **calculs des évolutions à champ constant** (18) par exemple] qu'à l'Insee dans le cadre de la quatrième génération des EAE [utilisation d'un **langage informatique de programmation des contrôles** mis à disposition des statisticiens, le LSE[17]] ou de la troisième génération de SUSE [**principes des contrôles de la liasse fiscale** (14)]. Mais on a cherché aussi à pallier les défauts connus des applications existantes [**primauté donnée à l'enquête** sur la liasse fiscale dans la ventilation du chiffre d'affaires en branches (16), introduction de nouveaux principes de redressement [par exemple **recalcul des exercices comptables sur 12 mois** (13) ou prise en compte des **exercices décalés** (12) et des **branches des petites entreprises** (17)]. On a enfin développé de nombreuses innovations totalement originales [définition d'un champ a priori (3), systématisation des macro-contrôles (5), estimateurs composites (18), indépendance des modules jusqu'à l'étape de réconciliation individuelle (2) et bien sûr changement de définition de l'unité statistique « entreprise » avec la prise en compte des groupes (21)].

Ces **innovations** sont en principe indiquées dans le texte par une mention en ***italique gras suivie d'une astérisque****. Dans certains cas, celles-ci avaient été mis en place dans certaines EAE et l'innovation réside dans l'extension de certaines EAE à l'ensemble de la statistique d'entreprises.

Phase 1 :

1. Transfert des maitrises d'œuvre des SSM vers l'Insee : Cf. *avant propos* ci-dessus
2. Indépendance des différents modules ou sous-processus de Ésane : Cf. § *1.2.2* et § *1.2.5*.
3. Définition du champ « a priori » des enquêtes ESA et EAP : Cf. §*1.2.1* et § *4.1*.
4. Primauté des corrections ou des validations gestionnaires sur les redressements automatiques (même en cas de valeur atypique) : Cf. § *2.1*
5. Systématisation des macro-contrôles : Cf. introduction du *chapitre 2* et § *2.7*.
6. Prise en compte des restructurations et donc des enveloppes dans les contrôles : Cf. § *3.3*.
7. Estimation d'un flux intra-enveloppe : Cf. § *3.2.2.1*.
8. Suppression de la stratégie « petites entreprises » des EAE qui conduisait à une évolution heurtée des agrégats : Cf. § *5.1.3.1*.
9. Rééquilibrage des échantillons entre secteurs (et homogénéisation de la précision des estimateurs entre secteurs) : Cf. § *5.1.3.1*.
10. Renouvellement de l'échantillon par moitié chaque année : Cf. § *5.1.4*.
11. Interrogation des petites entreprises industrielles : Cf. § *5.2.1*.
12. Prise en compte des exercices décalés sur l'année civile où ils comptent le plus de mois : Cf. § *6.2.7.1*).
13. Recalcul des exercices comptables sur 12 mois : Cf. § *6.2.7.3* et § *6.4.2*.
14. Principes de contrôle de la chaîne amont (6.2.1 à 6.2.10 mais plus particulièrement micro-contrôles internes) (§ *6.2.8*), redressements quasi sûrs internes (§ *6.2.9*) et redressements internes automatiques (§ *6.2.10*).
15. Utilisation de sources administratives « sociales » pour l'estimation de l'emploi : Cf. § *7.1*
16. Primauté de l'enquête sur la liasse fiscale pour la ventilation du chiffre d'affaires en branches : *Cf. § 8.3*.
17. Prise en compte des branches des petites entreprises : Cf. § *8.3.3.2*.
18. Exploitation simultanée des sources « enquêtes » et « administratives » pour le calcul des agrégats grâce à l'utilisation d'estimateurs dits « composites » : Cf. *§ 9.1* et *9.2*
19. Analyse des évolutions par sous population (démographie, modification de champ, restructurations et pérennes) : Cf. § *9.6.2*.
20. Mise en place d'une démarche d'assurance qualité : Cf. *chapitre 10*.

Phase 2 :

21. Changement de définition de l'unité statistique « entreprise » : prise en compte des groupes : Cf. *chapitre 11*.

[17] Il s'agit d'une innovation principalement informatique, donc non décrite ici, mais qui se traduit par une pratique nouvelle des statisticiens qui ne se contentent plus de "spécifier" des contrôles mais les programment réellement.

1. Présentation des sources et de l'architecture générale du mode de contrôle des données Ésane *[1.1], [1.7], [1.10]*

Comme tout système d'information, Ésane se procure les données qu'il traite ensuite auprès d'un certain nombre de « sources », externes au système. L'originalité du système est de mixer des sources administratives et des sources statistiques tout en leur appliquant des traitements communs. L'objet de ce premier chapitre est de décrire ces sources, (§ *1.1*) une par une, puis les principes de ces traitements communs (§ *1.2* sur le contrôle des données, leur champ, référentiel du système).

1.1 Les différentes sources en présence

1.1.1 L'enquête sectorielle annuelle (Esa) *[1.6], [1.9]*

L'Enquête Sectorielle Annuelle (Esa) est formellement nouvelle bien qu'en fait, elle soit une redéfinition sensible des Enquêtes Annuelles d'Entreprises (EAE) sur les mêmes champs sectoriels dans le cadre du programme Resane.
Un des objectifs principaux de ce programme étant l'allègement de la charge de réponse des entreprises via une large mobilisation des sources administratives, toutes les données jusqu'alors collectées dans les EAE et qui sont de fait déjà disponibles dans diverses sources administratives, notamment fiscales, ne font plus l'objet d'une enquête statistique. L'Esa correspond ainsi à une EAE très allégée, notamment de toutes les données comptables que l'on peut mobiliser sur les liasses fiscales et des principales données d'emploi.

Les principaux objectifs de cette enquête sont les suivants :

- comme pour les EAE précédemment, repérer les différentes activités exercées par les entreprises, via la ventilation de leur chiffre d'affaires en branches, et en déduire alors leur activité principale (APE). Cet objectif est premier tout d'abord en ce qu'il conditionne le bon classement sectoriel des entreprises et par conséquent la qualité des statistiques sectorielles. Ensuite, il conditionne un bon passage secteur/branches, sur lequel repose l'élaboration des comptes nationaux de la France ;
- compléter la liasse fiscale sur certains aspects liés à l'investissement, notamment dans sa composante immatérielle et par voie « d'apport », en cas de restructuration. Lorsque les sources administratives disponibles ne permettent pas d'obtenir les variables du compte de résultat ou celles du bilan, l'Esa sert alors à les collecter[18].
- identifier les restructurations juridiques qui affectent la vie des entreprises et dont le repérage est essentiel pour produire de bonnes statistiques en évolution ;
- décrire au travers de variables spécifiques les principales caractéristiques de chaque secteur.

Cette enquête concerne globalement les secteurs économiques du commerce, des services, des industries agro-alimentaires, des exploitations forestières et des scieries, des transports et de la construction ; toutes les tailles d'entreprise sont dans le champ de l'enquête (Cf. le *chapitre 4* pour plus de détails). Le champ a bougé une fois en 2010 (Introduction des entreprises du spectacle vivant et suppression des loueurs de fond personnes physiques Cf. § *4.6*).

Le nombre d'unités enquêtées (120 000 environ) est inférieur à celui des EAE. Il s'agissait au départ d'une enquête exclusivement postale. La possibilité de répondre par Internet a été développée à compter de 2013 sur l'exercice 2012 par remplissage d'un formulaire identique au questionnaire papier mais en format « libre office » exclusivement, ce qui a limité fortement le taux de réponse par Internet. A partir de l'année de constat 2013, on accepte le format PdF d'Adobe. Il est prévu de proposer un questionnaire « web » en 2018 sur l'exercice 2017.

Les questionnaires sont envoyés en 4 vagues successives. Sur l'année de constat N, la première est envoyée en fin février (N+1) et la dernière fin juin (N+1). C'est la date de clôture d'exercice N-1 qui détermine l'appartenance d'une entreprise à une vague.

[18] C'est le cas de certaines coopératives agricoles ou du commerce de gros ainsi - à compter de l'année de constat 2013 - que des entreprises profilées au sein des grands groupes (Voir *chapitre 11*)

- La vague 1 regroupe les entreprises qui ont une date de clôture comprise entre septembre et décembre N.
- La vague 2, la plus importante, regroupe les entreprises qui ont une date de clôture au 31/12 de l'année N.
- La vague 3 regroupe les entreprises qui ont une date de clôture comprise entre janvier et mars N+1.
- La dernière vague regroupe les autres entreprises.

Le plan de sondage est « classique » : stratifié à un degré, en utilisant les critères « code APE » et tranche de taille (en effectifs salariés) comme critères de stratification. Au sein des strates ainsi définies, le même taux de sondage est appliqué pour l'ensemble des régions françaises, à l'exception des Dom.

Jusqu'en 2012 compris (sur l'année de constat 2011), sont interrogées des **unités légales**. À compter de 2013 (sur 2012), des **entreprises « profilées »** (Cf. *chapitre 11*), correspondant à des groupes ou des parties de groupes, ont **commencé d'être interrogées et prises en compte dans l'élaboration des agrégats publiés**[19].

1.1.2 L'enquête annuelle de production (EAP)

Cette enquête est équivalente à l'Esa pour les **entreprises des secteurs industriels hors IAA**. Elle a, en plus, un **objectif supplémentaire** qui est de répondre au règlement européen **ProdCom**. Cette enquête remplace donc à la fois l'EAE dans l'industrie et l'ancienne Enquête Annuelle de Branche pour les entreprises industrielles[20], mais elle est plus proche dans son contenu des enquêtes annuelles de branche que de l'ancienne EAE (à la différence des ESA).

L'utilisation de l'EAP pour répondre au règlement Prodcom rend les rôles de ces deux enquêtes en partie **différents** en ce sens que les agrégats issus de l'EAP sont publiés alors que ceux de l'ESA ne le sont pas en tant que tels ; seuls sont disponibles des agrégats issus du système Ésane complet. La chaîne de contrôle validation des données collectées dans l'EAP est différente de celle d'Ésane et n'est d'ailleurs pas décrite dans cet ouvrage. Seule l'intégration des données EAP dans Ésane sera décrite. On notera cependant que :

- Le tirage de l'EAP est réalisé en même temps que le tirage de l'Esa. Certaines entreprises peuvent recevoir un questionnaire de l'EAP (appelée dans ce cas EAbis) et un questionnaire de l'Esa. Il s'agit d'entreprises non industrielles mais qui ont une activité secondaire industrielle suffisamment importante pour justifier d'une interrogation spécifique dans le cadre de l'EAP.
- L'EAP fournit à Ésane la ventilation en branche et l'APE des entreprises industrielles afin de pouvoir calculer les agrégats branches complets.
- Comme les liasses fiscales des entreprises industrielles sont traitées dans le cadre d'Ésane, la réconciliation des données (décrite au *chapitre 8*) est faite de la même façon qu'avec l'Esa.
- Les informations relatives aux restructurations impliquant des entreprises industrielles connues par l'EAP sont fournies au système Ésane comme celles connues par l'Esa via l'application interne dite Citrus (cf. *chapitre 3*).

1.1.3 Les données fiscales *[1.8]*

Les unités légales, sociétés ou entreprises individuelles, ont **l'obligation** de tenir une comptabilité (article L123-12 du code du commerce)

Elles doivent **acquitter l'impôt** chaque année à raison des bénéfices réalisés cette année-là (article 12 du code des impôts) :

- impôt sur les sociétés (IS) pour les sociétés ;
- impôt sur le revenu (IR) pour les entreprises individuelles.

La DGFiP et l'Insee ont une convention pour que **la DGFiP fournisse ces données, brutes, à l'Insee**, qui les utilise pour des besoins statistiques.

L'obligation comptable diffère selon le secteur, la forme juridique et la taille de l'entreprise.

[19] La première publication qui les prend en compte porte sur l'année 2013 ; mais elle comprend également une mise à jour des agrégats de l'année 2012 calculés sur les mêmes entreprises interrogées qu'en 2013 pour rendre les évolutions significatives.

[20] Elle est traitée par des gestionnaires de la Direction régionale de l'Insee de Normandie à Caen qui traitaient ces enquêtes au sein du service statistique du ministère de l'industrie (Sessi) avant son rattachement à l'Insee.

Il en résulte différents types de déclarations de bénéfice :

- les bénéfices agricoles (BA) ;
- les bénéfices industriels et commerciaux (BIC) ;
- les bénéfices non commerciaux (BNC).

Et différents régimes selon l'importance de l'entreprise :

- le régime normal ;
- le régime simplifié ;
- le régime micro (ou au forfait pour le BA).

Le champ précis du système d'information Ésane est fixé par une note interne détaillée de la responsabilité de la division « Infrastructure et répertoires statistiques » ; en fonction de paramètres qui caractérisent chaque unité légale (son code APE, sa catégorie juridique, etc.), de sorte que l'on sache, pour chaque liasse reçue ou attendue, si elle doit être prise en compte par le système ou pas et s'assurer ainsi de l'exhaustivité des données et de l'absence de double compte. Un chapitre entier (Cf. *chapitre 4*) est dévolu à la description du champ qui est présenté plus succinctement au § *1.2.1*.

A titre illustratif sur l'exercice 2008, le nombre d'entreprises dans les différentes sources fiscales était le suivant[21] :

Régime ou source	**Nb total d'entreprises**	**Nb d'entreprises du champ Ésane**	**% du CA total du champ Ésane**	**Nb variables disponibles**
Bénéfice réel normal (BIC-RN)	700 000	700 000	**90**	1000
Bénéfice réel simplifié (BIC-RSI)	1 200 000	1 200 000	5	400
Bénéfices non commerciaux (BNC)	500 000	500 000	2	100
Entreprises profilées	3	3	3	1000
Régimes micro (BIC-micro et BNC-micro)	400 000	400 000	ε	0
DGCP	500	500	ε	1000
Coopératives IAA et commerciales hors DGI	500	500	ε	200
Bénéfices agricoles (BA-RN / BA-RSI)	300 000	200	ε	400
TOTAL	3 100 000	**2 800 000**	**100**	**1200**

Ces données sont fournies en 4 vagues :

- les liasses dites « anticipées » livrées en juin N+1 pour les entreprises qui télé-déclarent ;
- les liasses dites « intermédiaires » livrées mi-septembre N+1 ;
- les liasses dites « normales » livrées mi-octobre N+1 ;
- les liasses dites « complémentaires » livrées en mars N+2 ;

et par quatre centres de saisie des impôts (Marseille, Clermont-Ferrand, Nantes et Reims). Le tout représente donc un nombre important de fichiers à traiter, d'autant qu'il y a autant de fichiers que de types de régime et de centres de saisie.

La volumétrie des livraisons observée pour l'exercice 2009 était par exemple la suivante :

[21] Le nombre d'entreprises a considérablement augmenté depuis 2008 plus particulièrement dans les régimes micros qui ne sont plus hors champ Ésane avec la création du régime de l'auto-entrepreneur devenu en 2015 micro-entrepreneur.

Remontée	Régime	Nb de liasses	Total
Anticipées 2008[22]	BIC RN	703	1 047 796
Intermédiaires 2008		3 520	
Normales 2008		14 557	
Complémentaires 2008		98 424	
Anticipées 2009		539 901	
Intermédiaires 2009		274 481	
Normales 2009		87 636	
Complémentaires 2009		28 574	
Normales 2008	BIC RSI	39 952	1 654 125
Complémentaires 2008		144 700	
Normales 2009		1 405 168	
Complémentaires 2009		64 305	

N.B. : ces proportions ont évolué depuis, allant dans le sens d'une meilleure couverture dès la fourniture des anticipées en juin.

Quelques sources complémentaires sont utilisées pour les unités non assujetties à l'impôt sur les bénéfices (régies des eaux, quelques coopératives des industries agro-alimentaires par exemple, Cf. § *1.1.5* ci-après).

Le processus de production des résultats se partage en trois grandes phases :

- l'élaboration de **résultats provisoires en octobre** qui permet de répondre aux estimations provisoires demandées par le règlement SBS d'Eurostat[23], à un niveau agrégé (groupe de la NAF) et sur un nombre d'agrégats limité (6 agrégats retenus : CA, VA, EBE, salaires, résultat comptable et investissement) ;
- l'élaboration des **agrégats nécessaires aux comptables nationaux** (centrés sur la valeur ajoutée) pour le calcul du compte semi-définitif de l'année N, pour le 31/12 de l'année N+1 ;
- l'élaboration de **résultats complets** sur plus d'une centaine d'agrégats à des niveaux détaillés (NAF700 ou NES116*classes de taille), tenant compte à la fois des informations fiscales et des données sectorielles fournies par les enquêtes Esa et EAP pour la fin juin de l'année N+2.

1.1.4 Les données d'emploi

Ésane s'appuie essentiellement sur **deux sources administratives**, les « URSSAF », les « DADS » qui résultent de déclarations administratives remplies par les entreprises (unités légales) et dont l'Insee dispose d'une **copie** des déclarations, et « Clap » qui est un **système d'information interne à l'Insee**, lui-même obtenu par rapprochement et mise en cohérence des sources Urssaf et DADS. Il existe également des informations sur l'emploi dans les liasses fiscales, mais de médiocre qualité (beaucoup d'informations manquantes en particulier) ; d'où le fait qu'on s'est tourné vers des sources administratives spécifiques. Ces sources sont progressivement remplacées, par la « déclaration sociale nominative » (DSN) ce qui nécessitera une évolution importante du système Ésane.

1.1.4.1 La source Urssaf

Il s'agit de données fournies par les établissements aux Urssaf lors du paiement des cotisations sociales et mises à disposition d'Ésane par le système d'information « Épure » de l'Insee.
Épure fournit l'effectif de fin de trimestre (le nombre de personnes physiques salariées) et les rémunérations payées aux salariés lors du trimestre. Les données sont regroupées au niveau « entreprise » (Épure dispose de l'information à l'établissement). Le trimestre T est fourni à T+80j en provisoire (ex. le 3e trimestre 2012 (fin sept.) sera disponible le 20 décembre 2012). Le trimestre précédent est fourni en version définitive en même temps (ex. le fichier du 20 décembre 2012 comprendra donc aussi le second trimestre de 2012 définitif).

Ces données permettent d'obtenir une estimation plus précoce que les autres sources de l'emploi et des rémunérations au niveau « entreprise » pour le champ Ésane et elles servent :

[22] Les liasses fournies en 2008 par les impôts mais qui ont plus de mois d'activité en 2009 (y compris pour les durées d'exercices inférieures à 12 mois), sont comptabilisées pour l'année 2009 dans Ésane. C'est pourquoi elles figurent dans ce tableau des liasses 2009.
[23] Les estimateurs définitifs requis par le règlement prévoient un délai de 18 mois et sont donc tirés des estimateurs définitifs d'Ésane, qui sont produits en décembre puis juin.

- de variables de stratification pour le tirage de l'échantillon de l'Esa de l'année suivante ;
- de variables auxiliaires pour le contrôle des données de l'Esa ;
- de variables auxiliaires pour l'imputation des données d'emploi provenant des DADS (cf. paragraphe suivant).

En revanche, ces données d'estimation d'emploi conjoncturelles (source Urssaf) ne sont pas diffusées par Ésane et **ne font pas l'objet de contacts avec l'entreprise** pour être validées.

1.1.4.2 La source DADS

Il s'agit d'**une formalité administrative obligatoire, remplie par les unités légales, destinée à plusieurs organismes :**

- **CNRACL** : retraite des agents des collectivités locales ;
- **CPAM :** ouverture des droits des salariés à l'assurance maladie ;
- **CRAM :** accidents du travail, droits à la retraite ;
- **DGFiP :** taxes sur les salaires, IRPP ;
- **IRCANTEC :** retraite complémentaire agents non titulaires État et CL ;
- **UNEDIC :** contributions assurance chômage ;
- **URSSAF :** calcul des cotisations de Sécurité sociale ;

Cette source contient des informations sur l'établissement employeur (N° Siret, activité etc.) et sur chacun des salariés qu'il emploie (rémunération, période d'emploi etc.). C'est la source la plus complète dont on peut disposer.
Les déclarations portant sur l'année N sont traitées par l'Insee d'avril N+1 à fin janvier N+2. Mais une transmission est faite pour Ésane fin novembre N+1[24]. Cette livraison couvre quasiment l'intégralité du champ Ésane car ce sont surtout les secteurs agricoles et publics qui sont disponibles plus tard.
Il s'agit de données d'emploi et de salaires trimestrialisées (on aura les valeurs des 4 trimestres de l'année civile pour toutes les variables fournies) regroupées au niveau entreprise et qui concernent la ventilation des effectifs et des rémunérations :

- selon le temps de travail (complet / partiel) ; les DADS sont en particulier **la seule source** qui fournit des effectifs en **« équivalents temps plein** » (ETP) ;
- selon les grandes catégories socioprofessionnelles (cadres / professions intermédiaires / employés / ouvriers).

1.1.4.3 Les données de « Clap »

Clap est un système d'information de l'Insee dont l'acronyme signifie « connaissance locale de l'appareil productif ». En pratique, il est essentiellement tourné vers une connaissance de la localisation fine des effectifs salariés. Clap utilise pour cela à la fois les données Urssaf et DADS regroupées par établissement qu'il localise à la commune. Clap jouera un rôle intermédiaire entre les sources administratives brutes et Ésane :

- en arbitrant entre données élémentaires Urssaf et DADS lorsqu'elles sont incohérentes ;
- en aidant Ésane pour la réconciliation avec les données fiscales, lorsque l'incohérence vient visiblement des sources d'emploi et non des sources fiscales (l'emploi est rattaché à l'ensemble des données comptables fournies par la liasse fiscale par l'intermédiaire de la masse salariale).

Les données d'emploi fournies par Clap sont des données en nombre de personnes physiques salariées **mesurées au 31/12, ainsi que les masses salariales correspondantes**. Les données des DADS sont utilisées pour avoir par ailleurs également des données **en « équivalent temps plein** ». Tout ceci est décrit en détail au *chapitre 7* et au § *8.6*.

1.1.5 Des sources complémentaires pour des unités spécifiques

Il s'agit essentiellement des entités parapubliques pour lesquelles on ne dispose pas de liasse fiscale.

Plusieurs types d'entités liées à la sphère publique mais considérées comme des Entreprises Non Financières par les comptables nationaux ne déposent pas de liasse fiscale. Le « système Ésane » continue à collecter des informations sur les entités figurant précédemment dans le SIE[25] mais **ne les intègre pas toutes** dans ses propres résultats (les autres sont directement transmises aux comptes nationaux). En effet, l'ensemble des variables (notamment comptables) nécessaires à leur intégration correcte dans le système Ésane standard n'est pas toujours disponible. Le coût de l'intégration s'est également parfois avéré très élevé.

[24] Via la division Exploitation de Fichiers Administratifs sur l'emploi et les revenus (EFA).
[25] Système Intermédiaire d'Entreprises : Cf. Avant-propos.

1.1.5.1 Les données intégrées à Ésane (syndicats des eaux et PMU)

- Les (environ) 3500 « **syndicats des eaux** » : les données comptables des établissements publics de coopération intercommunale (EPCI) sont transmises par le bureau CL2A de la DGFiP début novembre n+1 (donc trop tard pour les résultats préliminaires). Les syndicats des eaux (qui mènent des activités de traitement et distribution d'eau, ou de collecte et traitement des eaux usées, respectivement Naf 36 et 37) sont pleinement intégrés dans Ésane depuis l'exercice 2010[26]. La sélection des unités intégrées s'effectue au niveau des établissements. On consolide[27] les établissements précédemment sélectionnés sur leur Siren. Par conséquent, les entités figurant dans Ésane (avec un numéro Siren comme identifiant Cf. la liste d'exceptions au § *13.1*) peuvent avoir des contours plus étroits que celles figurant dans Sirene et des APE différentes. Ces unités sont ensuite traitées de façon automatique dans la chaîne Ésane. Toutefois, des traitements manuels complémentaires sont le plus souvent nécessaires pour aboutir à des données définitives fiables[28]. Il n'est pas toujours possible de réaliser de telles estimations par manque d'information disponible et il arrive donc d'avoir in fine des trous de collecte.

- **Le PMU** envoie à l'Insee chaque année par courrier un rapport interne confidentiel. Le PMU est intégré à Ésane depuis l'actualisation des résultats 2010 concomitante au définitif 2011. Son chiffre d'affaires y est évalué hors gains des parieurs.

1.1.5.2 Les données non intégrées à Ésane mais transmises aux comptes nationaux pour leur usage

- **Les 28 CROUS** (centres régionaux des œuvres universitaires et scolaires) : les données comptables des « opérateurs de l'État »[29] sont transmises par le bureau CE2B de la DGFiP entre octobre et novembre n+1, les CROUS en sont extraits et transmis directement aux comptables nationaux.
- **L'Institut Pasteur** : des données sont publiées sur le site Internet de l'Institut, et les principales variables sont **transmises aux comptables nationaux**.
- **Les ports et aéroports** qui restent gérés par des collectivités locales ou des CCI (hors champ car non marchandes) auraient dû être traités à part pour que l'on puisse suivre quand même la partie marchande de leur activité, un peu comme les syndicats des eaux qui répondent sur un périmètre différent de celui sous lequel ils sont enregistrés dans Sirene. Finalement, le SOeS dont ils dépendent n'a pas transmis les informations qui auraient permis de les repérer. De moins en moins nombreux, **ils ne sont donc pas** repérés spécifiquement.

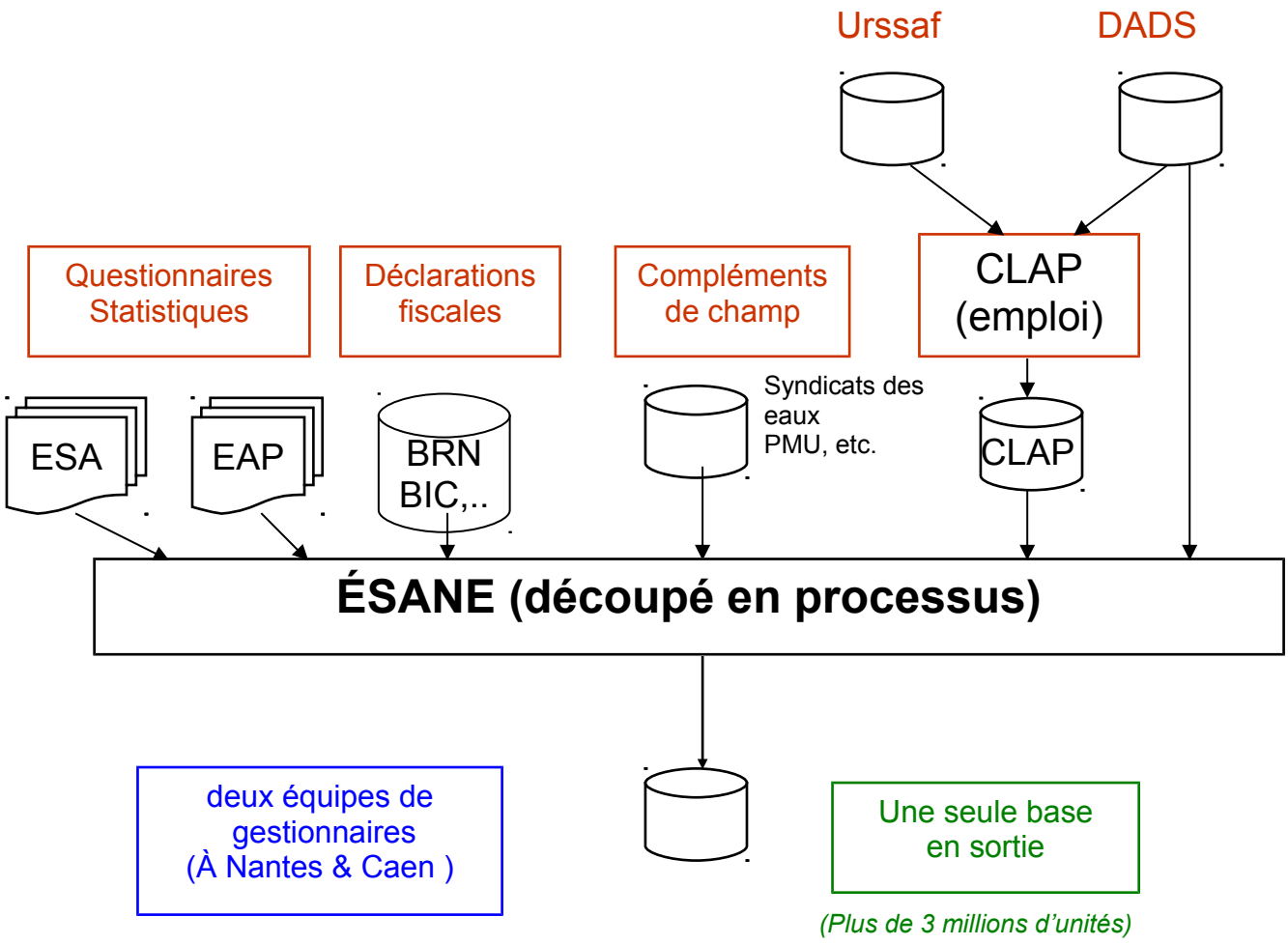
[26] En « sous-champ 2 » et repérables à leur source « 41- ex DGCP » ; les différents sous champs sont expliqués un peu plus loin au § *1.2.1.1*.

[27] Dans le cas où il y a plusieurs établissements pour un même Siren, leurs informations fournies par la DGFiP sont cumulées.

[28] Ainsi, une régie peut conclure un accord avec le fisc pour que celui-ci ne considère comme lucrative ou concurrentielle qu'une partie de son activité (après sa reprise des activités auparavant exercées par le secteur privé). Du coup, la liasse qui nous parvient ne porte que sur une portion de son activité et il reste impossible de récupérer le complément par ailleurs. Il est donc nécessaire d'essayer de reconstituer « à la main » une liasse complète. Ce travail peut se faire à partir de la réponse de l'unité à partir de l'ESA et du rapport d'activité.

[29] Les *opérateurs de l'État* sont des organismes distincts de l'État, au statut juridique public ou privé, auxquels est confiée une mission de service public de l'État et rattachés à un ou plusieurs programmes du budget de l'État, ils sont notamment régis par le décret du 29 décembre 1962.

L'architecture stylisée du nouveau système est donc la suivante :



1.2 *La logique de contrôle de ces données multi-sources*

L'architecture de contrôle des données adoptée dans Ésane ainsi que le calendrier de contrôle sont présentés dans les deux graphiques ci-après[30] ;l'objectif de ces graphiques est de visualiser l'enchaînement des tâches.

Afin de pouvoir diffuser des données provisoires, données qui permettent également de répondre au règlement SBS, il a été décidé de scinder la validation des données d'enquête et des liasses fiscales en deux étapes, une première où l'on traite des données diffusées en provisoire et une seconde où on traite l'ensemble.

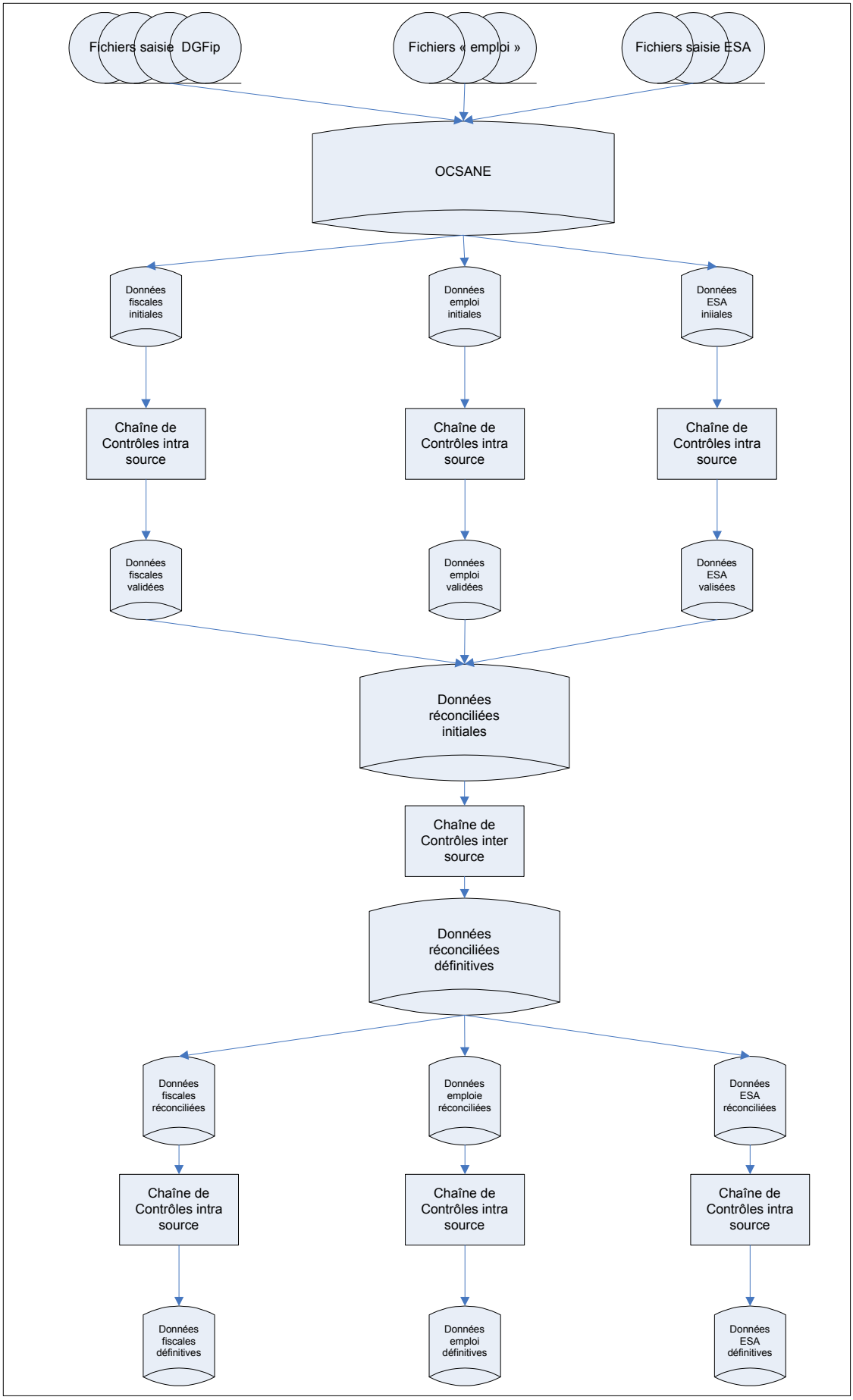
Pour l'Esa, les données provisoires concernent :

- le chiffre d'affaires,
- sa ventilation en branches et par voie de conséquence l'APE
- les données relatives aux restructurations.

Pour les données de la liasse fiscale, les données provisoires sont :

- le chiffre d'affaires
- la valeur ajoutée hors taxe
- l'excédent brut d'exploitation (EBE)
- les salaires et traitements
- le résultat comptable
- la variation d'immobilisations corporelles

[30] Un planning plus détaillé figure au *chapitre 12*.


Fichiers saisie DGFip
Fichiers « emploi »
Fichiers saisie ESA
OCSANE
Données fiscales initiales
Données emploi initiales
Données ESA iniiales
Chaîne de Contrôles intra source
Chaîne de Contrôles intra source
Chaîne de Contrôles intra source
Données fiscales validées
Données emploi validées
Données ESA valisées
Données réconciliées initiales
Chaîne de Contrôles inter source
Données réconciliées définitives
Données fiscales réconciliées
Données emploie réconciliées
Données ESA réconciliées
Chaîne de Contrôles intra source
Chaîne de Contrôles intra source
Chaîne de Contrôles intra source
Données fiscales définitives
Données emploi définitives
Données ESA définitives

ID	Nom de tâche	N-1		N												N+1	
		nov.	déc.	janv.	févr.	mars	avr.	mai	juin	juil.	août	sept.	oct.	nov.	déc.	janv.	févr.
1	Initialisation du référentiel N-1 (campagne N)																
2	**Chaine de validation des données ESA**																
3	Chargement de l'échantillon ESA																
4	Édition et routage des questionnaires [vague 1]																
5	Édition et routage des questionnaires [vague 2]																
6	Édition et routage des questionnaires [vague 3]																
7	Édition et routage des questionnaires [vague 4]																
8	Collecte (y compris gestion des relances)																
9	Intégration de la saisie																
10	Contrôle intra source des données pour les estimations des résultats provisoires																
11	Contrôle intra source : chiffre d'affaires et ventilation en branches																
12	Contrôles intra source des données sectorielles																
13	Contrôles intra source post réconciliation																
14	**Chaîne de validation des données fiscales**																
15	Livraison des liasses fiscales anticipées par la DGFip																
16	Contrôle intra source des données provisoires																
17	Livraison des liasses fiscales intermédiaires par la DGFip																
18	Contrôle intra source des données définitives																
19	Contrôles intra source post réconciliation																
20	Livraison des liasses fiscales normales par la DGFip																
21	**Chaîne de validation des données d'emploi**																
22	**Contrôles intra sources des données d'emploi**																
23	Echange de données avec CLAP																
24	**Chaine de validation des données inter sources**																
25	**Ensemble des données ESANE validées**																

1.2.1 Le référentiel : un champ défini a priori (cf. *chapitre 4* pour une explicitation complète)

La « population de référence » Ésane, c'est-à-dire l'ensemble des entreprises traitées par le système, est constituée à partir du répertoire statistique Sirus (Cf. § *11.7*) **à chaque début de campagne Ésane** et le répertoire statistique en assure la gestion pendant toute la campagne. Ceci constitue une **nouveauté** par rapport à ce qui était fait dans le passé. En effet, dans le système précédent, faisait partie du champ toute unité pour laquelle on recevait une liasse fiscale en provenance des impôts. Le champ de la statistique structurelle d'entreprise était donc défini a posteriori, en fin de campagne. Dans le cas d'Ésane, la logique est inversée et c'est bien **a priori** que sont définies les unités que l'on souhaite prendre en compte pour l'élaboration des statistiques structurelles d'entreprises. Il s'agit d'une ***innovation forte**** d'Ésane qui permet de mieux contrôler les évolutions N/N-1.

En pratique, cette définition du champ a priori s'est avérée beaucoup plus complexe que ce qui était prévu à l'origine, comme on le verra au *chapitre 4* qui présente ce champ en détail. De plus Ésane répond à des besoins différents et son champ total est donc l'union :

- du champ qui doit répondre aux besoins de la comptabilité nationale, que l'on pourrait qualifier de champ DCN ;
- du champ qui doit permettre de satisfaire au règlement de statistiques structurelles d'Eurostat, que l'on pourrait qualifier de champ SBS ;
- du champ de la statistique d'entreprise au sens large, que l'on pourrait qualifier de champ DSS[31] (qui prend notamment en compte un secteur de l'agriculture).

Cette difficulté à définir très précisément le champ et cette multiplicité des besoins ont conduit à une certaine variabilité du champ Ésane durant les premières campagnes ce qui a généré des problèmes pour les comparaisons temporelles et les validations des données diffusées. Ces difficultés sont évoquées plus en détail dans le *§ 1.2.1.3*.

1.2.1.1 La définition du champ

Il s'agit des **sociétés et entreprises individuelles, non agricoles[32] et non financières** (exceptés les auxiliaires financiers et d'assurances ainsi que les holdings qui font partie du champ depuis 2010). Il comprend les sociétés juridiques considérées dans le répertoire statistique comme marchandes[33] et exploitantes[34] ou non exploitantes participant au système productif. Le critère marchand retenu, notamment pour qualifier les associations, est celui défini par les comptables nationaux. En outre, un critère supplémentaire est appliqué : celui de la transparence fiscale. Ainsi, sont éliminées du champ toutes les sociétés définies par la DGFiP comme transparentes fiscalement. Cela concerne principalement des sociétés civiles immobilières et des SEP[35].
Certaines catégories juridiques sont exclues du champ dans leur totalité soit parce qu'elles concernent uniquement des unités non marchandes, soit parce qu'elles font double emploi avec d'autres unités productives. D'autres sont exclues du champ quand elles n'ont pas d'effectif salarié.

Le champ géographique d'Ésane est la France entière, y compris les Dom.

Le champ Ésane est **découpé en quatre sous champs**. Ces sous champs correspondent à des **considérations de production statistique** (c'est-à-dire que les procédures de contrôle ou de redressements qui sont appliquées aux unités légales de ces sous-champs ne sont pas les mêmes : certaines sont automatiques ; d'autres font l'objet d'un contrôle humain) mais n'ont pas vraiment de sens en termes de diffusion, à l'exception du sous-champ 1. La description précise de ces sous-champs est donnée au chapitre 4 mais il importe de savoir qu'une même sous-classe de la Naf est

[31] Le département des synthèses sectorielles (DSS) de l'Insee est un des plus importants utilisateurs internes du système, en particulier pour les commissions des comptes de la nation dont il est rapporteur.
[32] Les exploitations agricoles sont dans Ocsane mais ne font pas partie du champ Ésane à l'exception des exploitations forestières 02.20Z. Les liasses fiscales des BAS et BAN (bénéfices agricoles simplifiés et normaux) transmises à l'Insee par la DGFiP, avec les autres régimes fiscaux, sont traitées automatiquement donc sans intervention des gestionnaires Ésane. Aucun traitement n'est mis en œuvre pour extrapoler les données en cas d'absence de liasse fiscale parce qu'aucune information n'est diffusée pour ces secteurs. Ces données sont collectées pour la comptabilité nationale.
[33] Une unité est considérée comme *marchande* lorsque ses coûts de production sont couverts à plus de 50% par ses ventes.
[34] Une unité est *exploitante* si elle exploite elle-même des moyens de production pour produire des biens ou services sur un marché.
[35] Société en participation. Ce sont des sociétés qui ne sont pas personnes morales.

fréquemment répartie sur plusieurs sous-champs, ceux-ci se distinguant souvent par la catégorie juridique de l'unité légale ou par sa taille :

- Le sous champ Ésane 1 (ou *champ enquête*) correspond au champ de l'Esa et de l'EAP. Les secteurs couverts représentent **594 sous-classes** de la NAF (350 pour l'Esa, 244 pour l'EAP)[36]. Ils peuvent être considérés comme les secteurs les plus complètement décrits par le système Ésane : en général, ils sont aussi couverts par la source fiscale et les sources d'emploi. Les informations « sectorielles » qui ne sont disponibles que via les enquêtes ne sont, par définition, disponibles que sur ce sous-champ 1.
- Le sous champ Ésane 2 (ou *champ gestionnaire*) constitue une première extension du sous champ Ésane 1. Il contient des secteurs non couverts par l'Esa ou l'EAP mais correctement couverts, en termes de poids économique essentiellement, par la source fiscale. Ces secteurs font l'objet d'une demande de statistiques (Services statistiques des ministères, Département des Comptes Nationaux ou autres) telle qu'y **consacrer des moyens gestionnaires Ésane est justifié**.
- Le sous champ Ésane 3 (ou *champ automatique)* : Il s'agit du reste c'est-à-dire de secteurs couverts par la source fiscale mais ne faisant l'objet d'aucune demande particulière. Par rapport au sous champ Ésane 2, la différence essentielle provient de ce qu'aucun moyen gestionnaire Ésane n'y est consacré ; tous les traitements sont automatisés.
- Un sous champ Ésane 5[37] a été ajouté à partir de la campagne 2010 et constitue une dernière extension du champ. Il correspond aux unités qui ont été ajoutées au champ initial au fil des années (auto-entrepreneurs, etc.). Il traduit parfaitement le problème de variabilité évoqué ci-dessus. Comme pour le sous-champ « 3 », aucun traitement ne sera effectué par les gestionnaires Ésane. On pourra toutefois calculer des agrégats sur ce sous-champ, ce qui permettra, notamment pour le Département des Comptes Nationaux et le Département des Synthèses Sectorielles, de reconstituer leur champ d'intérêt propre.

1.2.1.2 Ocsane : le « répertoire » d'Ésane qui permet la gestion de ce champ

Ocsane est un répertoire **interne au système** qui a été développé pour gérer les fonctions d'identification des unités qui constituent sa population de référence. Le répertoire statistique SIRUS (Cf. § *11.7*) met hebdomadairement à jour Ocsane lorsqu'il modifie la population de référence (ou des caractéristiques de ses unités). Dans l'autre sens, Ocsane demande à SIRUS de mettre à jour les caractéristiques qu'il gère tout particulièrement (APE, chiffre d'affaires voire effectifs). Certaines de ces demandes de mise à jour ont d'ailleurs lieu en temps réel et peuvent le cas échéant être refusées si une source prioritaire selon Sirus donne une information contradictoire. Ainsi par exemple Ocsane ne peut pas demander la réactivation d'une unité à Sirus si cette dernière est déclarée cessée dans Sirene. En revanche pour l'APE et le chiffre d'affaires de l'année d'exercice en cours de traitement dans Ésane, Ocsane est considéré comme la source prioritaire par Sirus et ses demandes de mises à jour sont systématiquement acceptées.

Ocsane a donc des fonctions dédiées à Ésane [38] alors que SIRUS gère des caractéristiques qui sont partagées par l'ensemble des applications de la statistique d'entreprises. Ocsane :

- permet de **gérer des informations spécifiques à Ésane**, (adresses de collecte pour l'Esa, etc.) qui ne sont pas d'un intérêt collectif au-delà des statistiques structurelles annuelles, et donc n'ont pas vocation à être remontées dans le répertoire statistique ;
- comprend des **fonctions transversales sur les unités** qui sont utiles à différents sous-systèmes d'Ésane :
  - stocker les identifiants des sources administratives (liasses fiscales, données sociales) des unités légales du référentiel Ésane ;
  - prendre en compte (recevoir) les « restructurations[39] » provenant d'un autre système d'information de l'Insee, baptisé Citrus. Ces restructurations sont transversales et ont en effet un impact sur le contrôle des données et la diffusion des résultats ;
  - gérer les contacts avec les entreprises (assistance à l'appel téléphonique) ;

[36], Dont trois sous-classes des arts du spectacle vivant ajoutées en 2010 (en ESA).
[37] Un sous-champ 4 a existé transitoirement ; aussi ce champ rajouté en 2010 porte t-il le numéro 5.
[38] C'est en quelque sorte le successeur d'Océan (Outil de Coordination des Enquêtes ANnuelles), utilisé pour les anciennes EAE, sauf en ce qui concerne les aspects de tirage d'échantillon. De la même façon, FUL (Fichier Unique de Lancement) assure le même genre de fonctions pour l'EAP.
[39] On appelle ainsi les modifications de contours qui affectent les unités légales (fusions, scissions, transferts d'actifs). Elles font en particulier l'objet du chapitre 3.

- gérer les commentaires des gestionnaires, ce qui répond à un **besoin de traçabilité** tout au long de la chaîne de production, d'autant plus fort que plusieurs gestionnaires peuvent être amenés à traiter différents aspects d'une même entreprise au cours d'une même campagne.

1.2.1.3 Les difficultés engendrées par la définition d'un champ a priori

Le fait de définir un champ a priori permet d'assurer un résultat final **indépendant** des processus de collecte. Les entreprises sont au départ sur le même pied d'égalité. Toutefois, dans le cas d'Ésane, la difficulté de la gestion de ce champ a priori vient du fait que l'on ne dispose pas d'une variable provenant des impôts ou renseignée par les entreprises qui permette d'être sûr à 100% que chaque unité répond ou non à la définition du champ que l'on souhaite appréhender. Il est donc nécessaire d'estimer cette appartenance au champ à partir de variables statistiques telles que l'APE, la catégorie juridique et le code marchand, dont on dispose d'une valeur pour chaque entreprise. Toutefois ces caractéristiques peuvent être parfois entachées d'erreur. De plus, dans certains cas très particuliers, le croisement de ces variables ne permet pas d'identifier exactement la population voulue. Ainsi par exemple, l'idéal serait de ne retenir dans le champ Ésane que les Sociétés Civiles Immobilières (SCI) qui remplissent une liasse fiscale car elles détiennent dans certains cas l'ensemble des immobilisations d'un groupe. En revanche, les SCI familiales n'ont aucun intérêt en statistique d'entreprises. La distinction entre ces deux types de SCI n'est toutefois pas possible par le simple croisement des variables de définition du champ. Il est donc parfois nécessaire de gérer des listes de Siren à part avec toutes les difficultés qu'entraîne ce genre de traitement et notamment la mise à jour annuelle de ces listes.

Ce problème est en partie à l'origine du sous champ 5 décrit ci-dessus. Ce dernier permet en effet d'identifier des populations **non prises en compte en 2008** et réintroduites les années suivantes. Les résultats de ces entreprises sont bien pris en compte dans les statistiques d'entreprises publiées ultérieurement.

Enfin ce problème de définition du champ est particulièrement prégnant au moment du traitement des non-réponses.

En effet, en fin de campagne, une fois que la DGFiP a livré l'ensemble des liasses qu'elle a traitées, il reste dans le référentiel Ésane (Ocsane) des unités sans liasse fiscale pour lesquelles il faut se poser la question de l'imputation ou non d'une liasse :

- La liasse est-elle absente car mal identifiée, ou non encore traitée par la DGFiP ?
- La liasse est-elle absente parce que l'entreprise n'est pas dans l'obligation d'en remplir une ? Cette entreprise n'aurait donc pas dû être dans le champ a priori.

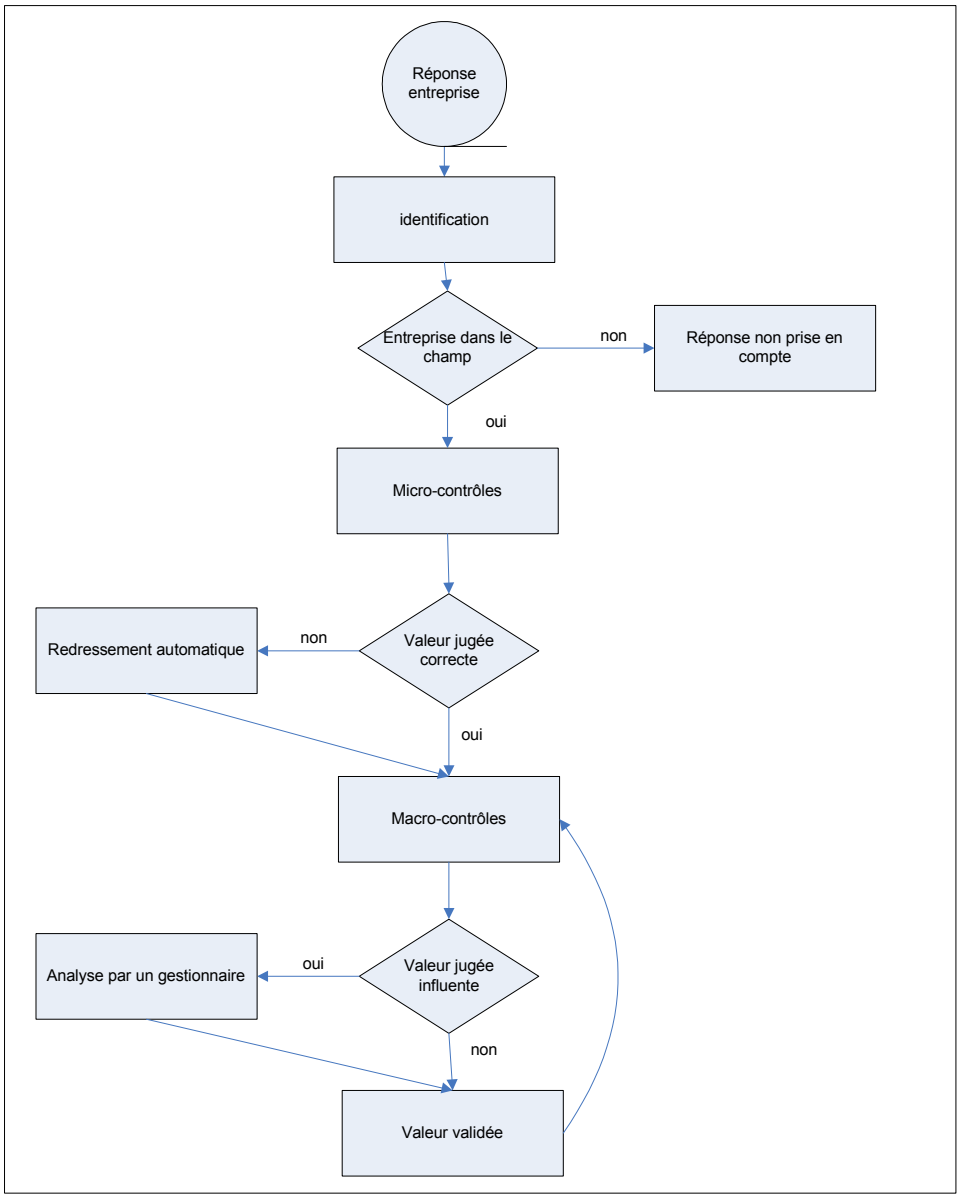
La résolution de ce problème est traitée au § *6.5.3*.

1.2.2 Des contrôles « intra-sources » indépendants d'une source à l'autre

Afin de conserver une certaine adaptabilité du système face à la diversité de sources **qui peuvent évoluer indépendamment** de l'Insee, ***le système d'information est conçu de façon modulaire****. Chaque source subit, de façon indépendante, une chaîne de contrôle. L'objectif de cette chaîne est d'optimiser le partage des tâches entre l'homme et la machine pour rechercher et corriger les erreurs qui se sont glissées dans l'information collectée. Ne sont confiés à l'action humaine que les dossiers pour lesquels l'automatisation est moins performante. L'automatisation trouve ses limites surtout en matière d'entreprises dont le poids économique relativement à l'ensemble de la population est important.

Les contrôles portent bien évidemment sur les **unités statistiques** qui sont présentes dans la source et qu'on appelle souvent indifféremment **unités légales ou entreprises**, partant de ce que, jusqu'à la campagne 2013, les deux notions étaient synonymes. Lorsqu'on a introduit les entreprises profilées issues de la cible 1 (Cf. *chapitre 11*) , les contrôles évoqués par exemple au § *11.6.5* portent sur les entreprises profilées. Lorsqu'on généralisera le profilage aux cibles 2 et 3, les contrôles de l'ensemble de la chaîne porteront bien sur les valeurs relatives aux entreprises profilées introduites ; mais ils pourront donner lieu à des corrections **au niveau des unités légales qui les composent** car les valeurs des variables de ces entreprises profilées sont obtenues de fait à partir de la consolidation des données des unités légales.

Le principe de la chaîne de contrôle pour une source donnée est décrit dans la figure ci-après.



Cette chaîne de contrôle se déroule en deux étapes une fois que l'entreprise a bien été identifiée comme faisant partie du champ de l'enquête :

- la première a pour objectif de **traiter la non-réponse partielle** et de **corriger des valeurs jugées aberrantes**. Elle est entièrement automatique et met en œuvre des **micro-contrôles** de cohérence et de vraisemblance puis, le cas échéant, des redressements automatiques se basant sur la réponse de l'année précédente ou sur les réponses des autres entreprises appartenant à la même strate de contrôle (obtenue par croisement entre l'activité principale et la tranche de taille de l'entreprise). Contrairement à ce qui était fait dans l'EAE, ces micro-contrôles sont assez « lâches ». Ils ne donnent lieu à redressement automatique d'une valeur renseignée que si cette dernière est vraiment jugée très atypique. En outre, comme on le verra par la suite, un macro-contrôle spécifique (dit macro-contrôle par différence) permettra de contrôler ces redressements automatiques. Il faut en effet éviter le biais qui consiste à supprimer toute valeur atypique en la considérant comme aberrante !
- la seconde vise à identifier les unités qui doivent être **contrôlées par un gestionnaire**. Cette sélection s'opère principalement[40] à l'aide de **macro-contrôles** qui vont permettre d'identifier les unités les plus fortement contributrices aux agrégats diffusés.

[40] Voir le chapitre suivant : quelques variables spécifiques sont aussi contrôlées de façon micro économique par les gestionnaires.

1.2.3 Une réconciliation des sources validées ou contrôles inter-sources

Une fois les caractéristiques des différentes sources validées, il convient d'assurer une **cohérence globale** de ces caractéristiques multi sources. C'est le but du processus de **réconciliation des données individuelles** (dit Rédi). Il s'assure que les caractéristiques communes à deux sources ont, in fine, la même valeur. C'est ***une des grandes innovations de Resane***.

Dans le cas où il existe une divergence initiale sur ces caractéristiques d' « accrochage », le processus affecte **une valeur unique** après un **arbitrage automatique** qui privilégie une source selon différentes règles de priorité (ces règles sont détaillées dans le *chapitre 8* de ce document). Cette réconciliation concerne le chiffre d'affaires et sa ventilation en types de ventes (biens, marchandises, services) mais également l'emploi et les rémunérations. Sur ce contrôle de données inter sources, comme pour le contrôle des données intra sources, il a été décidé d'appliquer des procédures de macro-contrôles conduisant à demander une expertise manuelle des gestionnaires uniquement pour les cas jugés les plus impactants[41].

1.2.4 Une mise en cohérence finale des données de chaque source

Pour les variables d'accrochage, une fois la valeur unique déterminée, elle est réinjectée dans chaque source dans laquelle elle figure. La chaîne de contrôle des données intra-source est alors lancée dans chacune des sources afin de contrôler et de redresser automatiquement les données **en prenant en compte le résultat de la réconciliation** des données.

1.2.5 Les avantages et inconvénients de ce principe de contrôle

La nécessité d'une telle réconciliation en aval du processus de production repose sur ***l'un des piliers méthodologiques**** de Resane à savoir **la modularité** et **l'indépendance des différents sous-systèmes**. Ainsi, par principe, on s'interdit d'utiliser, pour la validation des caractéristiques d'une source donnée (Enquête ou liasse fiscale par exemple), des caractéristiques provenant d'une autre source pour le même exercice. On aurait pu imaginer, par exemple, que la valeur du chiffre d'affaires de l'Esa (connu avant celui de la liasse fiscale), s'il était disponible au moment du contrôle des données de la liasse, soit mobilisée afin que le gestionnaire fasse de la réconciliation individuelle en cours de traitement. Mais ce principe a été rejeté d'emblée dans Resane.

Ce principe permet ainsi d'éviter tout blocage complet des chaînes de contrôles, en cas de problème sur l'une des sources. Si, une année donnée par exemple, la DGFiP ne nous livrait pas les données fiscales ou le faisait avec retard, on pourrait sans problème valider quand même les données de l'enquête et les données d'emploi. De même, si on avait un gros souci dans la gestion de l'enquête, on pourrait exploiter la source fiscale. Si l'ensemble avait été intégré, ceci n'aurait pas été possible.

[41] Par exemple si l'écart de chiffre d'affaires entre les deux sources est supérieur à x% du chiffre d'affaires du secteur.

2. Présentation des grands principes méthodologiques retenus pour le contrôle des données : les CQR et les macro contrôles *[1.11], [3.1], [3.2], [3.5], [3.6], [3.7], [3.9]*

L'objet de cette partie est de détailler les grands principes présentés au § *1.2.2*. Il s'agit donc de décrire la chaîne **dite des CQR** (Contrôles Qualification Redressement)[42] puis la mise en œuvre des **macro contrôles**.

Les principes de base des CQR, constituant ce qu'on appelle les **micro contrôles,** sont pour la plupart repris des EAE (Enquêtes Annuelles d'Entreprises). L'idée est de corriger **automatiquement** les données qui apparaissent erronées au regard d'un certain nombre de contrôles de cohérence ou de vraisemblance. On commence donc par une phase de «micro- contrôles », décrite dans le § *2.2* telle que chaque variable de l'enquête est l'objet d'au moins un micro-contrôle. Une note est attribuée selon le résultat de chaque micro-contrôle. On somme ensuite toutes les notes obtenues pour une variable donnée dans une nouvelle variable qui caractérise la « qualité » de la variable. Les variables qui ont atteint un certain « seuil » de qualité sont dites « qualifiées » et ne sont pas modifiées ; c'est la phase de « qualification » décrite dans le § *2.3*. A l'inverse, les variables qui n'ont pas atteint ce seuil de qualité (c'est-à-dire qu'aucun contrôle n'a jugé « bonnes ») sont corrigées automatiquement. C'est la phase de redressement qui fait l'objet du § *2.4*. C'est en particulier le cas des non-réponses partielles qui, par définition, ne peuvent pas satisfaire ces contrôles ; c'est aussi le cas des variables grossièrement fausses (erreurs d'unités, euros à la place de kilo euros par exemple, ou erreurs résultant de décalages dans la saisie des données). A l'issue de ces redressements automatiques, certaines égalités comptables entre variables peuvent ne plus être vérifiées (elles ne l'étaient d'ailleurs pas forcément au départ). Il faut donc « ajuster » plus ou moins automatiquement ces redressements comme l'explique le § *2.5*. Lorsque ces ajustements sont trop importants, cela donne lieu à de messages d'erreurs qui sont décrits au § *2.6*.

Ils sont suivis par des **« macro-contrôles »** qui mesurent **l'influence d'une variable** et donc d'une entreprise **sur un agrégat final qui sera publié.** La principale ***nouveauté**** **des contrôles** mis en œuvre dans Ésane est la place **prépondérante** laissée à ces macro-contrôles (Cf. § *2.7*) qui signalent aux gestionnaires **les entreprises à contrôler manuellement**[43] **en priorité**.

Dans l'EAE, les macro-contrôles n'intervenaient qu'en fin de processus. Les premiers rappels d'entreprises se faisaient à partir des résultats des seuls micro-contrôles.
Dans Ésane, **ce sont principalement les macro-contrôles qui décident de la nécessité ou non de l'examen de la réponse d'une entreprise par un gestionnaire**. Quelques **contrôles de contribution simples** (Cf. § *2.9*) des principales entreprises contributrices à certains agrégats accompagnent également les macro-contrôles.

L'objectif de la phase de contrôle des données par un gestionnaire est double :

- **corriger** des erreurs éventuelles ;
- **obtenir des explications** de la part de l'entreprise quant à son comportement jugé atypique par les contrôles.

Les macro-contrôles mettent en effet en évidence des unités dont la réponse est jugée **atypique** par rapport aux autres unités qui lui ressemblent (en termes d'activité et de tranche de taille) et qui ont un poids non négligeable dans les agrégats diffusés.

[42] Il s'agit d'une notion reprise de l'EAE dont il existe une description très précise dans l'article suivant : « Enquêtes annuelles d'entreprise : à la rencontre du 4ème type », P. Rivière, Courrier des Statistiques n°78 page 17, août 1996 :*https://www.epsilon.insee.fr/jspui/bitstream/1/14348/1/cs78.pdf*

[43] On qualifie de « manuels » les contrôles qui sont réalisés par les gestionnaires, par opposition aux contrôles qui sont réalisés automatiquement par des programmes. Lors des contrôles « manuels », des programmes fournissent des éléments de diagnostic et le gestionnaire exerce alors sa responsabilité par rapport à ce diagnostic soit en confirmant la valeur d'une variable, soit en la modifiant.

2.1 Les différentes occurrences d'une même caractéristique

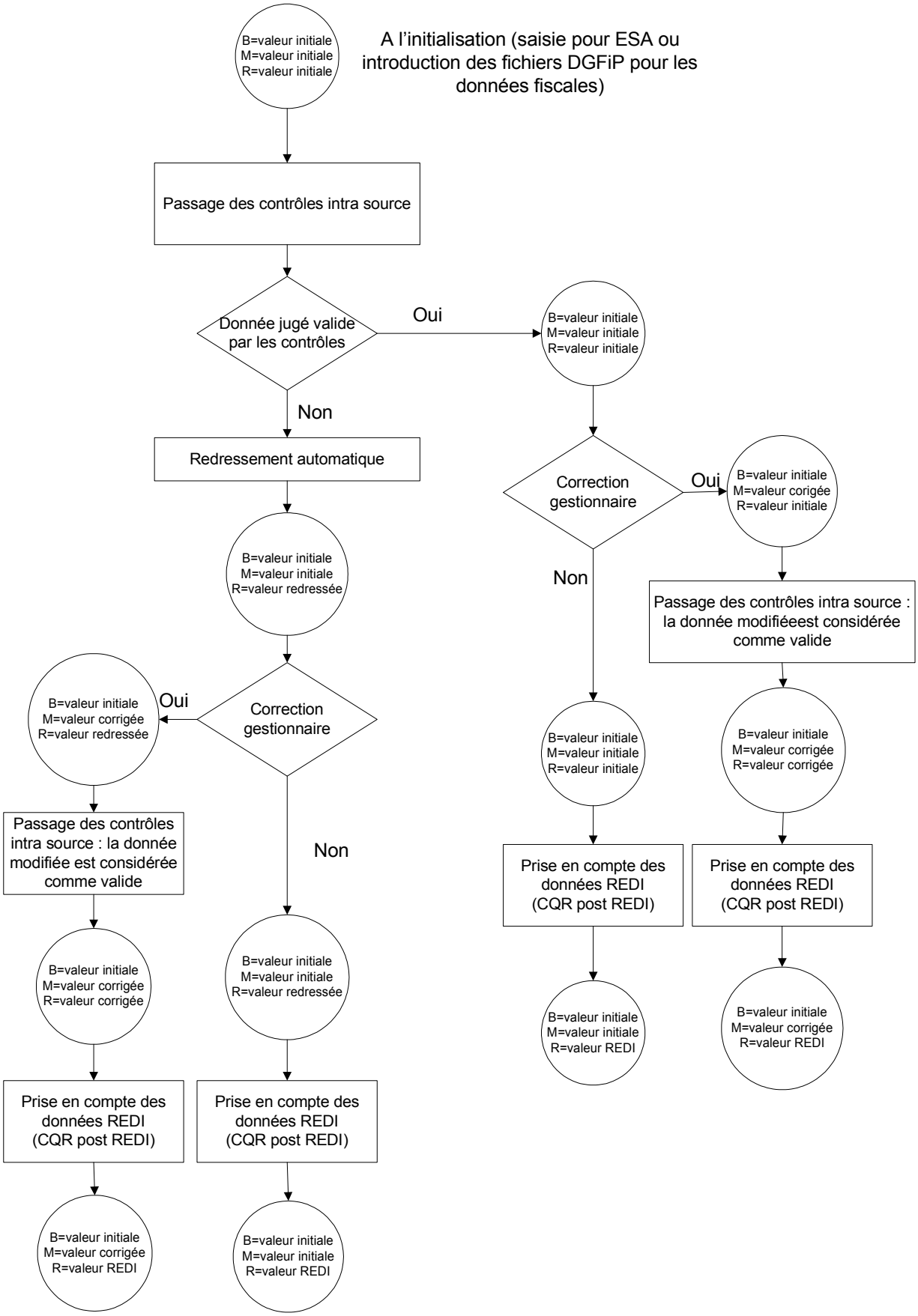
Dans Ésane, les différentes valeurs successives qu'une même caractéristique peut prendre au cours du processus de production, au sein duquel s'enchaînent des traitements automatiques et des traitements manuels, sont les suivantes :

- **Sa valeur brute**[44] : c'est celle qui résulte du processus de mesure utilisé dans une source. En pratique, la variable brute correspond à la variable saisie par le prestataire de saisie en cas de retour de questionnaire ou de prise en compte de la liasse provenant du fichier de la DGFiP. En revanche, les données qui seraient collectées par téléphone suite à une non-réponse totale ou partielle ne sont pas considérées comme données brutes, car elles sont nécessairement obtenues à l'occasion d'un échange avec un gestionnaire qui implique une validation de sa part.
- **Sa valeur redressée ;** c'est celle qui résulte du traitement automatisé de redressement dont l'objectif est de limiter au maximum les erreurs de mesure. On distingue en la matière deux grands types d'erreur de mesure : l'erreur de *non-réponse* et l'erreur *sur réponse*. C'est la valeur de cette occurrence qui est en sortie de la chaine des CQR[45].
- **Sa valeur modifiée** : c'est celle qui résulte de l'intervention du gestionnaire, pour éliminer au maximum les erreurs de mesure qui ont pu l'affecter. Dans le concept de valeur « modifiée » on range également les valeurs brutes validées, c'est-à-dire celles qui paraissent suspectes, mais pour lesquelles un contact avec l'entreprise a permis d'établir qu'elles ne sont pas entachées d'erreur de mesure. Ces données « validées » bien qu'atypiques ne subiront donc pas de redressement automatique. Cette décision qui responsabilise le gestionnaire est une ***nouveauté**** de Resane (par le passé, dans l'EAE par exemple, le redressement automatique pouvait éventuellement corriger une correction faite par un gestionnaire si cette correction restait atypique). C'est la valeur de cette occurrence qui est en entrée de la chaine des CQR, lorsqu'un contrôle est exécuté après une intervention d'un gestionnaire. C'est également la valeur de cette occurrence qui sera en entrée de la chaine de réconciliation (Rédi)
- **Sa valeur réconciliée** : c'est celle qui résulte de la réconciliation individuelle de deux sources en aval du processus de production. C'est l'occurrence des caractéristiques, lorsqu'elle existe qui sera diffusée. Ainsi par exemple si un gestionnaire modifie une valeur du chiffre d'affaires d'une entreprise dans un sous-système puis modifie une seconde fois la valeur du chiffre d'affaires au moment de la réconciliation, ce sera cette dernière valeur qui sera retenue pour la diffusion et qui figurera de fait dans l'occurrence redressée de l'ESA (cf. § *8.5*)

[44] On distingue pour chaque variable numérique la valeur « manquante » de la valeur « nulle ».

[45] Ainsi, si le gestionnaire modifie une réponse de l'entreprise, comme cette correction est alors jugée correcte, ce sera la valeur corrigée par le gestionnaire qui figurera dans la donnée redressée. En effet, toute modification gestionnaire annule de fait un éventuel redressement automatique.

Le graphique ci-après illustre ce point :



La **valeur finalement utilisée dans le calcul d'un agrégat diffusé**[46] sera dans l'ordre suivant :

- l'occurrence réconciliée pour les caractéristiques qui figurent dans plusieurs sources ;
- sinon l'occurrence modifiée, si le gestionnaire est intervenu sur la donnée (en confirmant ou corrigeant la valeur proposée par le système de production) ;
- sinon l'occurrence redressée si la valeur brute a été jugée atypique par les programmes de contrôles ;
- sinon la donnée brute.

2.2 *Les micro-contrôles*

2.2.1 Le principe des micro-contrôles

Si les formules de calcul et les principes sont les mêmes pour le contrôle des données de la liasse fiscale et de l'Esa, la finalité des micro-contrôles diffère quelque peu entre ces deux sources :

- Dans l'Esa, les micro-contrôles attribuent des notes de 0 à 10 aux variables sur lesquelles ils sont lancés, permettant ainsi, en sommant ces notes, de savoir si ces variables sont qualifiées ou non (voir *§ 2.3* sur la Qualification), et **donc si elles subiront un redressement automatique** (voir *§ 2.4* sur les Redressements). Suivant la note attribuée, un message de contrôle pourra être généré ; cependant l'affichage de ce message au gestionnaire sera conditionné par le résultat du macro-contrôle correspondant au micro-contrôle. Ce sont en effet les macro-contrôles qui vont permettre de sélectionner les unités et les variables de ces unités à contrôler manuellement. Les messages des micro-contrôles sont à considérer comme **des aides au diagnostic du gestionnaire** et non comme des éléments déclencheurs des contrôles manuels.
- Pour le contrôle des données de la liasse fiscale, ils servent uniquement d'éléments de diagnostic. Ils ne donnent pas lieu à redressement. Ceci vient du fait qu'une phase de contrôle et redressement spécifique et systématique a lieu dans la liasse avant l'introduction des données dans la base (cf. description de la chaîne amont *§ 6.2*).

Les micro-contrôles sont de quatre types :

- **Contrôles internes de cohérence** (CIC)
  - Ils permettent de détecter des incohérences au sein du questionnaire (ou de la liasse) d'une entreprise, en comparant deux valeurs, dont l'une en général est la somme des montants dans une ventilation et l'autre le total déclaré par ailleurs.
  - Ils ne nécessitent pas de paramètres ; une marge d'erreur de 5 ou 10 % sera accordée.
  - Les notes accordées à ce type de contrôle sont fortes (en général 10 si l'écart est acceptable et 0 sinon). Ainsi, les données seront jugées correctes au niveau de la phase des micro-contrôles à partir du moment où les équilibres comptables les impactant sont respectés.
  - Exemple d'erreur détectée par un CIC : incohérence entre le chiffre d'affaires déclaré par l'entreprise et la somme des montants déclarés dans la ventilation en branches de son chiffre d'affaires.
- **Contrôles internes de vraisemblance** (CIV)
  - Ils permettent de détecter des invraisemblances entre la valeur d'un ratio pour une entreprise et les valeurs de ce ratio dans la strate à laquelle cette entreprise appartient.
  - Les CIV nécessitent l'utilisation d'une table de paramètres contenant :
    - la médiane du ratio du CIV notée Q2 ;
    - les bornes Q1 et Q3, correspondant aux premier et troisième quartiles ;
    - les bornes B1 et B3, correspondant par exemple aux premier et neuvième déciles [voir plus bas : calcul des paramètres des micro-contrôles].
  - La note attribuée à la variable dépendra de la valeur du ratio par rapport aux bornes :
    - entre Q1 et Q3, la note sera élevée (en général : 5) ;

[46] Pour des raisons de coûts informatiques et de gestion il a été décidé de ne pas créer une occurrence « diffusée » en plus des 3 occurrences qui existaient déjà dans les bases de données de l'enquête ESA. Ainsi, la donnée « diffusée » est recopiée et vient donc « écraser » l'occurrence « redressée » in fine.

        - entre B1 et B3 mais hors de l'intervalle Q1 Q3, la note sera moyenne (en général : 3) ;
        - en dehors de l'intervalle B1 B3, la note sera nulle.
    - Ces bornes seront calculées en début d'enquête grâce aux données N − 1 puis recalculées en cours d'enquête.
    - On utilisera des strates du type : 3 premiers chiffres de l'APE x tranche de taille (en 4 tranches de tailles) pour le calcul des bornes[47].
    - Les messages seront susceptibles d'être activés dès que le ratio de l'unité sera hors des bornes Q1 Q3.
    - On pourra utiliser des estimations pour données incomplètes, par exemple en ramenant la durée d'exercice à 12 mois.

- **Contrôles temporels de vraisemblance** (CTV) :
    - Ils permettent de détecter des invraisemblances entre la valeur d'une variable ou d'un ratio d'une entreprise en N et sa valeur en N − 1 ;
    - Les CTV nécessitent l'utilisation d'une table de paramètres contenant :
        - la médiane du ratio (ou de son évolution) du CTV notée Q2 ;
        - les bornes Q1 et Q3, correspondant aux premier et troisième quartiles ;
        - les bornes B1 et B3, correspondant par exemple aux premier et neuvième déciles.
    - La note attribuée à la variable (ou la variable active dans le cas d'un ratio) dépendra de l'évolution de la valeur de la variable (ou du ratio) par rapport aux bornes (les notes seront paramétrables) :
        - entre Q1 et Q3, la note sera élevée (par exemple : 7) ;
        - entre B1 et B3 mais hors de l'intervalle Q1 Q3, la note sera moyenne (par exemple : 5) ;
        - en dehors de l'intervalle B1 B3, la note sera nulle ;
        - les notes attribuées aux contrôles temporels sont plus fortes que celles des contrôles internes. Ainsi, la cohérence temporelle est jugée plus fiable que la cohérence interne. Autrement dit une entreprise ayant le même comportement atypique deux années de suite sera considérée comme « normale » la 2e année.
    - Ces bornes seront calculées en début d'enquête grâce aux données N − 1 et N − 2 puis recalculées en cours d'enquête.
    - On utilisera des strates du type : 3 premiers chiffres de l'APE x tranche de taille (grandes ou petites entreprises i.e. plus ou moins de 20 salariés).
    - On pourra utiliser des estimations pour données incomplètes, par exemple en ramenant la durée d'exercice à 12 mois.

Remarque : les contrôles temporels ne sont faits que pour les entreprises présentes en N et N − 1 et n'appartenant pas à une enveloppe N (voir plus loin : si elles appartiennent à une enveloppe, les contrôles portent sur l'enveloppe).

- **Contrôles temporels de cohérence** (CTC)
    - Ces contrôles ne sont effectués que pour les liasses fiscales. Ils permettent de repérer les incohérences d' « accrochages temporels » des liasses fiscales. Ainsi par exemple, le total des immobilisations de fin d'exercice de l'exercice N-1 doit être égal au total des immobilisations de début d'exercice N.
    - Ils ne nécessitent pas de paramètres ; une marge d'erreur de 5 ou 10 % sera accordée.
    - Les notes accordées à ces contrôles sont fortes comme pour les CIC (10 si l'écart est tolérable et 0 sinon).

Une variable **peut subir plusieurs contrôles** de types différents voire plusieurs contrôles de même type. Prenons l'exemple du chiffre d'affaires dans l'Esa. Il intervient dans les micro-contrôles suivants :

[47] Si la strate contient moins de 3 unités alors le calcul des bornes se fera sur la strate élargie obtenue en croisant les deux premiers chiffres de l'APE avec la tranche de taille. S'il y a encore moins de 3 unités dans cette strate, des bornes par défaut sont utilisées.

- CIC entre le chiffre d'affaires et la somme des montants déclarés dans la ventilation en produits ou en activités ;
- CIC entre le chiffre d'affaires et la somme des montants déclarés dans la ventilation par forme de vente (seulement dans le secteur du Commerce) ;
- CIV sur le rapport : chiffre d'affaires sur effectif ;
- CTV sur le chiffre d'affaires ;
- CTV sur le rapport : chiffre d'affaires sur effectif.

2.2.2 Calcul des paramètres des micro-contrôles:

Le principe de **calcul des paramètres de contrôle** des données des différentes sources (l'Esa, la liasse fiscale et les données d'emploi) a été **repris à l'identique de l'EAE**.

2.2.2.1 Paramètres des CIV

Pour un ratio donné, et dans une strate donnée, les trois quartiles Q1, Q2 et Q3 de la distribution observée sont calculés. Pour déterminer les intervalles de validité, dans la mesure où l'intervalle interquartile est jugé trop sévère et conduirait à redresser automatiquement trop de réponses, il est « étendu » des deux côtés, en prenant pour base, d'un côté (Q2 - Q1), et de l'autre (Q3 - Q2).

Pour ce faire, deux paramètres d'écartement, β (beta) et γ (gamma) sont utilisés. On construira le petit intervalle, [q1, q3] avec le paramètre γ, et le grand intervalle, [b1, b3], avec le paramètre β. Une première idée consistait donc à poser, par exemple, b3 comme valant Q2 + β. (Q3 - Q2), et b1 comme valant Q2 - β.(Q2 - Q1). Ainsi, l'idée d'élargissement de l'intervalle interquartile initial est conservée, tout en respectant sa dissymétrie éventuelle, puisqu'on se centre sur Q2. La même idée serait mise en œuvre pour l'intervalle le plus strict entre Q1 et Q3, en utilisant γ au lieu de β.

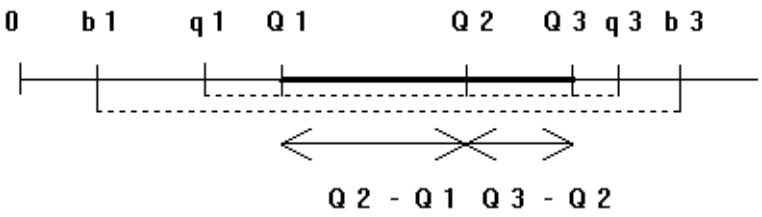
Cette technique a cependant un inconvénient lorsque les quartiles sont très resserrés : (Q3 - Q2) et (Q2 - Q1) sont très faibles, et l'élargissement de l'intervalle sera négligeable. Dans ce cas, il a été décidé de tenir compte de la valeur même de la médiane. Ainsi, dans le cas extrême d'un minuscule intervalle interquartile ce n'est plus Q3 - Q2 (respectivement Q2 - Q1) qui serait utilisé, mais une proportion de Q2, α.Q2.

Au final donc, si (Q3 - Q2) [respectivement (Q2 - Q1)] est inférieur à α.Q2, c'est l'écartement α.Q2 qui sera pris en compte ; c'est donc lui qui sera multiplié par β pour le grand intervalle, par γ pour le petit.

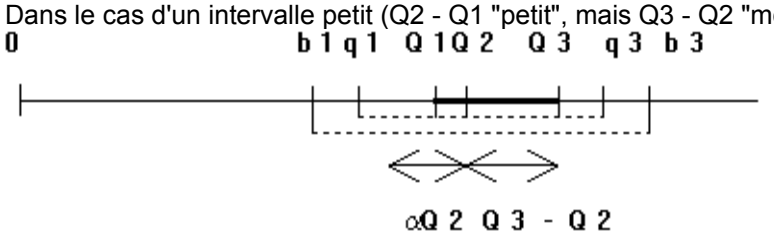
En résumé :

q1 = Q2 - γ max(Q2 - Q1,α.Q2)
b1 = Q2 - β max(Q2 - Q1,α.Q2)
q3 = Q2 + γ max(Q3 - Q2,α.Q2)
b3 = Q2 + β max(Q3 - Q2,α.Q2)

Voici ce que cela donne dans le cas où γ =1,4 et β=2 (l'intervalle étant suffisamment large pour que le terme en α.Q2 n'intervienne pas). L'intervalle interquartile y est représenté en traits forts, et les deux intervalles de validité (petit et grand) du ratio sont en pointillés.



Dans le cas d'un intervalle petit (Q2 - Q1 "petit", mais Q3 - Q2 "moyen") :



On voit qu'ici, c'est le terme en αQ2 qui l'emporte pour chaque borne inférieure, et le terme en Q3 - Q2 qui domine pour chaque borne supérieure.

Valeurs retenues pour α, β, γ:

Leur choix doit obéir à quelques principes simples. Tout d'abord, il faut que les deux intervalles soient plus grands que l'intervalle interquartile, trop sévère : γ et β doivent donc être tous deux supérieurs à 1. À contrario, il ne faut pas non plus qu'ils soient trop élevés, sinon les contrôles ne serviraient plus à rien. De plus, le grand intervalle contenant le petit, il est nécessaire que $\beta > \gamma$. En ce qui concerne α, il doit jouer un rôle dans certains cas particuliers (voir plus haut) tout en évitant d'aboutir à des intervalles énormes, d'autant plus qu'il sera ensuite multiplié, dans les calculs, par β ou γ. En outre, il est soumis à la contrainte forte $\alpha\beta < 1$, car sinon le grand intervalle contiendrait toujours la valeur 0.

De nombreux tests ont été effectués pour l'EAE4G sur une petite dizaine de ratios, sur données réelles. Ceci avait permis de fixer des valeurs raisonnables pour ces paramètres.

À l'issue de ces tests, les valeurs retenues pour ces divers paramètres sont :
alpha = 0,1 ; beta = 2,5 ; gamma = 1,8

Remarque : le grand intervalle [b1 , b3] correspond, en gros, aux *5e et 95e centiles*, alors que le petit intervalle [q1 , q3] est approximativement placé sur les *10e et 90e centiles*. Il s'agit d'intervalles "minimaux", puisque α ne joue pas : α n'interviendra (et élargira, dans ce cas, les deux intervalles) que si la distribution est très resserrée, relativement à Q2.

2.2.2.2 Paramètres des CTV

Dans le cas des CTV, dans la mesure où il s'agit de taux de croissance, la distribution est très dissymétrique. Une possibilité (qui n'a pas été retenue) aurait été de travailler sur le logarithme du taux de croissance ; à la place de cela, le choix a été fait d'introduire six coefficients, i.e. trois de chaque côté : $\alpha 1$, $\beta 1$, $\gamma 1$, $\alpha 2$, $\beta 2$, $\gamma 2$.

Les formules deviennent donc :
$q1 = Q2 - \gamma 1 \max(Q2 - Q1, \alpha 1.Q2)$
$b1 = Q2 - \beta 1 \max(Q2 - Q1, \alpha 1.Q2)$
$q3 = Q2 + \gamma 2 \max(Q3 - Q2, \alpha 2.Q2)$
$b3 = Q2 + \beta 2 \max(Q3 - Q2, \alpha 2.Q2)$

Les valeurs retenues pour les divers paramètres des CTV sont :
alpha1 = 0,1 ; beta1 = 3,4 ; gamma1 = 2,2 ;
alpha2 = 0,2 ; beta2 = 4 ; gamma2 = 2,5

Remarque : le grand intervalle [b1 , b3] correspond peu ou prou aux *4e et 95e centiles*, alors que le petit, [q1 , q3], s'ajuste à peu près sur les *10e et 90e centiles*.

2.3 Qualification (Esa)

Cette phase de qualification **n'a lieu que dans l'Esa**.

Chaque micro-contrôle attribue une note aux variables qui sont concernées par le contrôle (une seule variable dans le cas d'un contrôle temporel et deux pour un contemporain). La qualification d'une variable consiste à sommer ses notes et à **comparer ce total à un seuil**.

- Si le total des notes est supérieur ou égal au seuil alors la variable est dite qualifiée et sa valeur est conservée pour la suite des traitements.
- Si le total des notes est strictement inférieur au seuil alors la variable est **redressée automatiquement** (cf. paragraphe suivant). C'est le cas **notamment des variables en non-réponse partielle** pour lesquelles la somme des notes est nulle.

Dans le cadre d'Ésane, le seuil de notation des variables est fixé à 10.
La note des contrôles est ensuite adaptée de telle sorte qu'une variable est en général qualifiée si :

- elle obtient la valeur maximale à un CIC et ou au moins un CTV ;
- elle obtient deux notes moyennes à au moins 2 contrôles (CTV et / ou CIV).

Par rapport à l'EAE, la qualification des variables de l'Esa est **moins stricte**. L'idée est de redresser le moins possible les réponses initiales des entreprises. Les macro-contrôles permettront dans un deuxième temps de repérer des problèmes éventuels.

Cette phase de qualification **vise** essentiellement à **corriger la non-réponse partielle** d'une part et à **corriger les grosses erreurs** du type erreur de saisie ou erreur dans l'unité de réponse (réponse en € au lieu de K€ par exemple).

Une autre ***décision majeure*** a été prise dans le cadre de Resane et qui diffère par rapport à l'EAE. Dans Ésane, lorsqu'un gestionnaire intervient manuellement sur la valeur d'une caractéristique soit en la corrigeant, soit en la confirmant, cette variable est considérée comme qualifiée même si après intervention, la somme de ses notes est inférieure à 10 (conséquence des micro-contrôles décrits au § *2.2*).
En d'autres termes, une intervention gestionnaire ne sera jamais corrigée automatiquement et ce même si elle génère une incohérence.

Le gestionnaire doit donc s'assurer que les corrections qu'il fait ne créent pas d'incohérence dans le questionnaire. Des messages d'erreur sur son écran le préviendront en cas d'incohérences mais une fois sa correction enregistrée, elle sera prise en compte en l'état sans plus aucune remise en cause.

2.4 Redressement (Esa)

Dans la chaîne des CQR de l'Esa, à la suite de la qualification des variables a lieu le **redressement des variables non qualifiées** et donc notamment des variables en non-réponse partielle.

2.4.1 Types de redressement

2.4.1.1 Les différents types de redressements

Lorsqu'un redressement est requis, on tente un premier redressement ; si les conditions nécessaires à son application ne sont pas réunies, alors on tente un autre redressement, et ainsi de suite jusqu'à temps que les conditions (de moins en moins exigeantes) nécessaires à un des redressements de la liste soient réunies. On s'arrête dès qu'un redressement est effectué. Quoi qu'il arrive, en fin de course un redressement aura lieu (si un redressement est requis, évidemment).

Il existe différents types de redressement, qui sont tentés dans cet ordre :

- des **redressements déterministes** :
  - ils consistent à remplacer la valeur de la variable par une autre valeur ou une combinaison de valeurs propres à l'entreprise ;
  - par exemple : nombre total de magasins remplacé par la somme des nombres de magasins déclarés.

  Ce redressement est accepté à partir du moment où la correction proposée permet de résoudre le problème, c'est-à-dire que le redressement permet de qualifier la variable. Si tel n'est pas le cas, un autre type de redressement devra être réalisé.
  Vu que le questionnaire Esa évite au maximum toute redondance d'informations, ce genre de redressement est assez rare.

- des **redressements par estimation** :
  - **par tendance auxiliaire** :
    - ils consistent à faire évoluer la valeur N-1 de la variable à redresser selon le taux d'évolution observé sur une autre variable du questionnaire (appelée variable auxiliaire) corrélée à la variable à redresser ;
    - ces redressements supposent que la valeur en N − 1 de la variable à redresser n'avait pas subi de redressement ou avait été redressée par tendance auxiliaire ;
    - ils supposent aussi que les valeurs en N et N-1 de la variable auxiliaire sont de bonne qualité (variable auxiliaire qualifiée en N) ;
    - par exemple : remplacer le chiffre d'affaires en N par le chiffre d'affaires en N − 1 multiplié par le rapport Effectif en N / Effectif en N − 1.

- ou **par moyenne de strate** :
  - la valeur redressée est obtenue en multipliant la valeur N d'une caractéristique auxiliaire par le ratio obtenu en divisant le total de la caractéristique à redresser par le total de la caractéristique auxiliaire calculée au niveau de la strate d'appartenance de l'unité (groupe de l'APE croisé avec la tranche de taille)[48] ;
  - ce type de redressement suppose que la valeur en N de la variable auxiliaire est de bonne qualité (variable auxiliaire qualifiée en N)[49] ;
  - par exemple : redresser le chiffre d'affaires à partir du ratio chiffre d'affaires par salarié. Pour cela on va remplacer le CA en N par les effectifs en N × [Total CA(N)/Total Effectif N]. Ce ratio de chiffre d'affaires par salarié est calculé sur les entreprises répondantes de la strate. Il est initialisé en début d'enquête par le rapport des totaux observés en N-1 dans la strate.

Les redressements peuvent traiter plusieurs variables en même temps.

Prenons l'exemple de la ventilation en branches :

Cas 1. Si l'entreprise a donné son CA et rempli son cadre ventilation :
Lorsque le CA est validé, il sera prioritaire par rapport au total calculé à partir de la ventilation en branches si ce dernier diffère du chiffre d'affaires déclaré[50] ; la ventilation devra donc être calée sur le CA.

Cas 2. Si l'entreprise a donné son CA mais n'a pas rempli son cadre ventilation :
- Si l'on dispose de la ventilation de l'entreprise en N − 1 et qu'elle est de qualité suffisante, on reprend cette ventilation, en la calant sur le CA en N (redressement par tendance auxiliaire).
- Sinon, on prend la ventilation moyenne de la strate que l'on applique sur le chiffre d'affaires.

Cas 3. Si l'entreprise n'a pas donné son CA mais a rempli son cadre ventilation :
- Si l'entreprise a ventilé son CA en montants, on prend comme CA la somme des montants dans la ventilation. Il s'agit d'un redressement déterministe.
- Si l'entreprise a ventilé son CA en pourcentages : on procède aux redressements classiques pour le CA. On applique ensuite au CA redressé la ventilation donnée par l'entreprise pour obtenir le montant des CA par branche.

Cas 4. Si l'entreprise n'a pas donné son CA et n'a pas rempli son cadre ventilation :
- Estimation du CA (par tendance auxiliaire ou moyenne de strate).
- Puis estimation de la ventilation (par tendance auxiliaire puis moyenne de strate).

2.4.1.2 Cas des questions avec filtre

Le questionnaire Esa a été amélioré par rapport à celui de l'EAE afin d'essayer de **mieux distinguer** les valeurs manquantes qui sont en fait des 0, des vraies valeurs manquantes.
Pour ce faire, il a été décidé de rajouter pour chaque question quantitative une question filtre (oui/non) afin de savoir si l'unité est concernée ou non par la question. Quand une entreprise n'est pas concernée par une question, on sait ainsi mieux interpréter l'absence de réponse à la question.
Ce filtre est bien entendu utilisé lors des redressements automatiques comme ceci est expliqué ci-dessous.

[48] Les paramètres des CQR pour l'exercice N (bornes pour les micro-contrôles et moyennes pour les redressements) sont initialisés en début de campagne à partir des résultats des exercices N-1 et N-2. Ils peuvent être mis à jour en cours de campagne à partir des résultats observés sur N et N-1 à partir du moment où l'on considère disposer de suffisamment de questionnaires (ou liasses) validés pour l'exercice N.
[49] Si aucune variable auxiliaire n'est trouvée, la caractéristique est redressée en affectant sa moyenne dans la strate. Ce qui revient à prendre la variable « 1 » comme variable auxiliaire !
[50] Le questionnaire demande bien sûr au répondant de vérifier que le chiffre d'affaires total est bien égal à la somme de ses composantes. Mais il arrive fréquemment que ce ne soit pas le cas.

2.4.1.2.1 Cas où le montant est qualifié

Ce cas sous-entend que le montant est renseigné.

- Si le montant est non nul :
  - si la réponse à la question-filtre est Oui : pas de redressement
  - si la réponse à la question-filtre n'est pas Oui (i.e. soit Non, soit pas de réponse) : on redresse la réponse à la question-filtre à Oui.
- Si le montant est nul :
  - si la réponse à la question-filtre est Non : pas de redressement
  - si la réponse à la question-filtre n'est pas Non : on redresse la réponse à la question-filtre à Non.

2.4.1.2.2 Cas où le montant n'est pas qualifié

- Si le montant n'est pas renseigné :
  - s'il y a une réponse à la question-filtre :
    - si la réponse à la question-filtre est Oui : on redresse le montant de manière classique, selon la méthode indiquée au *§ 2.4.1.1* en s'assurant en plus que le montant redressé est non nul ;
    - si la réponse à la question-filtre est Non : on redresse le montant à 0.
  - s'il n'y a pas de réponse à la question-filtre : on redresse la réponse à la question-filtre en utilisant le **mode de redressement des questions-filtres** décrit ci-dessous, puis on redresse le montant de manière classique, selon la méthode indiquée décrite au *§ 2.4.1.1*.
- Si le montant est renseigné :
  - s'il y a une réponse à la question-filtre :
    - si la réponse à la question-filtre est Oui : on redresse le montant de manière classique, selon la méthode indiquée au *§ 2.4.1.1;*
    - si la réponse à la question-filtre est Non : on redresse le montant à 0.
  - s'il n'y a pas de réponse à la question-filtre :
    - si le montant est strictement positif : on redresse la réponse à la question-filtre à Oui puis on redresse le montant de manière classique, selon la méthode indiquée au *§ 2.4.1.1* en s'assurer en plus que le montant redressé est non nul
    - si le montant est nul : on redresse la réponse à la question-filtre en utilisant la **mode redressement des questions-filtres** décrit ci-dessous, puis on redresse le montant de manière classique, selon la méthode indiquée au *§ 2.4.1.1*.

2.4.1.2.3 Mode de redressement spécifique des questions-filtres :

Pour redresser des réponses aux questions-filtres, on tente successivement les deux redressements suivants :

1) **Déterministe** :
   Ce redressement consiste à prendre pour l'année N la réponse de l'année N − 1 si celle-ci existe.
2) Par estimation par **moyenne de strate** :
   - ce redressement consiste à imputer une réponse aléatoirement ;
   - au préalable, à l'aide des données N − 1, on calcule le ratio R du nombre de réponses Oui, sur le nombre de réponses total (Oui + Non) pour chaque question-filtre et cela pour chaque strate ; puis un nombre aléatoire entre 0 et 1 est généré ; on attribue la réponse Oui si le nombre tiré est inférieur à R et Non sinon.

2.5 Ajustement

Dans le questionnaire, après la phase de redressement automatique, toutes les ventilations de variables sont **« ajustées » automatiquement sur le total,** si ce dernier figure dans le questionnaire. Ceci s'applique également à tous les équilibres comptables de la liasse fiscale. Toutefois pour cette dernière, l'ensemble de ces ajustements a lieu lors du passage de la chaîne amont (cf. *§ 6.2*).

Lors d'un ajustement, le total indiqué T est considéré comme une valeur de référence sur laquelle les éléments sommés doivent être calés. Par exemple, l'ajustement de la ventilation du chiffre d'affaires

en produits (ou en activités) consiste à pondérer le montant X_i de chaque ligne i par une même constante α afin que la somme $\sum\alpha X_i$ soit égale au chiffre d'affaires T.

Dans le cas où un gestionnaire modifie des caractéristiques (par exemple la ventilation du chiffre d'affaires), il se peut que leur somme vienne à différer du total indiqué par ailleurs (dans l'exemple le chiffre d'affaires) ; la procédure d'ajustement proposée est alors la suivante :

- Si l'écart est inférieur à 5 %, les données modifiées sont enregistrées telles quelles dans la base, et un ajustement automatique a lieu afin de caler la ventilation sur le total. Les corrections affectent en priorité les montants de la ventilation non modifiés par le gestionnaire. Ceci permet de respecter la règle qui veut qu'on ne touche pas une valeur modifiée par un gestionnaire. Il s'ensuit que la correction n'est pas homothétique. On suppose de fait que le gestionnaire qui corrige une ventilation s'assure que le total est bien égal à la somme des composantes.
- Si l'écart est compris entre 5 et 10 %, un message sera affiché dans l'application pour avertir le gestionnaire de cet état de fait ; celui-ci aura le choix soit de modifier la ventilation pour rendre cohérentes les caractéristiques du questionnaire, soit de lancer un ajustement.
- Si l'écart est supérieur à 10 %, un message sera affiché par l'IHM pour avertir le gestionnaire de cet état de fait ; le gestionnaire devra modifier la ventilation pour rendre cohérentes les caractéristiques du questionnaire, sans quoi elles ne pourront pas être enregistrées dans la base.

2.6 Messages de contrôle des données associés aux micro-contrôles

Chaque micro-contrôle en erreur (c'est-à-dire les micro-contrôles pour lesquels la note obtenue n'est pas maximale) donne lieu à un message d'erreur à l'attention des gestionnaires.

Exemple de message :
`La valeur des Intérimaires pour l'entreprise est de : 90 809. La valeur des Effectifs pour l'entreprise est de : 24. On a donc ici un rapport Intérimaires sur Effectif de 3 783.708. Or pour 90 % des entreprises du secteur (groupe de la NAF), le rapport est compris entre 0 et 0.17. La réponse de l'entreprise est donc atypique.`

(Dans le cas présent, il y a de grandes chances que ce message ait été généré suite à un décalage de zone à la saisie).

Comme on l'a mentionné au § 2.2.1 ces messages sont considérés comme des « aides au diagnostic » pour les gestionnaires.
Toutefois pour une liasse ou un questionnaire à contrôler donné, ne sont **présentés aux gestionnaires** sur leur écran que les messages des micro-contrôles qui sont **associés à un macro contrôle en erreur**.

Ainsi, par exemple, si un micro-contrôle est en erreur pour la variable investissement, mais qu'en parallèle, aucun macro-contrôle impactant la variable investissement ne pose problème, alors le message associé au micro contrôle en erreur sur l'investissement ne sera pas présenté au gestionnaire. On cherche ainsi à ne pas passer du temps à corriger des erreurs qui n'auraient pas d'impact visible sur les résultats finaux pour se concentrer sur celles qui en ont.

2.7 Les macro-contrôles [3.3], [3.4], [3.8]

Dans le cadre de Ésane, **les macro-contrôles servent à déterminer quelles entreprises seront traitées par les gestionnaires et, au sein d'un dossier d'entreprise, quelles variables**. Il faut se rappeler en permanence qu'il est impossible de faire contrôler par les gestionnaires les centaines de variables des liasses fiscales des plus de trois millions d'entreprises françaises ni l'ensemble des données des quelque 150 000 entreprises enquêtées. Tout le problème est donc de sélectionner le plus efficacement possible celles qui feront l'objet d'un contrôle effectif.

Les macro-contrôles ont pour objectif de **signaler aux gestionnaires les entreprises** dont les réponses ont **une influence notable** sur les agrégats qu'il est prévu de diffuser[51]. Charge à eux de

[51] Toutefois, les formules de calcul des agrégats servant aux contrôles, qui sont données un peu plus loin, sont fondées sur des variables internes à une source. Elles sont donc différentes des formules utilisées finalement

vérifier soit directement auprès des entreprises soit dans des bases de données, la véracité des réponses ou de les corriger.
Les macro-contrôles passent après la chaîne des CQR décrite ci-dessus, sur des données qui ont donc éventuellement pu être redressées automatiquement.

Le principe (***nouveau par rapport aux contrôles des applications précédentes****) des macro-contrôles est le suivant. Il consiste principalement à calculer pour chaque variable de chaque entreprise une **contribution** ou un **score** (cf. § *2.7.3.1.3* et suivants pour les formules de calcul). Si la contribution **de l'entreprise dépasse un certain seuil, un message propre au contrôle est produit** pour le signaler au gestionnaire (Cf. § *2.7.5* pour la définition des seuils). Secondairement, des macro-contrôles par différence permettront de s'assurer que les redressements automatiques individuels n'ont pas été trop forts et si c'était le cas, attireront également l'attention des gestionnaires.

Ces messages permettront au gestionnaire de savoir quelles sont les variables à contrôler et **donnent quelques éléments de diagnostic pour contacter l'entreprise** (par mail ou téléphone).

L'idée est la suivante : dans un secteur donné, la conjoncture devrait faire varier les principales grandeurs économiques **à peu près de la même façon** : dans un secteur où les CA croissent de 10%, une entreprise qui croît de 12% est moins atypique qu'une entreprise qui croît de +3% quand les autres décroissent de 5%. Il s'agit donc de contrôler prioritairement des entreprises qui modifient fortement les variations qu'auraient eues les agrégats **sans elles**.

Exemple de message de macro-contrôle :
`Le taux d'évolution de CA Total du secteur (Niveau sous-classe) est de 1.665%. Sans cette entreprise, ce taux d'évolution serait de 1.002%. L'entreprise fait évoluer son secteur de +0.663 point(s).`

Dans le cadre de l'Esa, le calcul des contributions ne concerne que les unités répondantes (le traitement de la non-réponse totale se faisant par repondération, il est inutile de contrôler les entreprises non-répondantes).
En revanche comme le traitement de la non-réponse totale des liasses fiscales se fait par imputation, le calcul des contributions concerne toutes les unités du noyau (cf. § *6.3* pour la définition du noyau).

Pour les contrôles de données dans Ésane, **deux types de macro-contrôles** sont donc mis en œuvre :

- **Par contribution** (ou « drop out ») ; ils s'appliquent à deux sortes d'écart :
  - écart contemporain (par exemple : rapport Chiffre d'affaires sur Effectif en N)
  - écart en évolution (par exemple : Chiffre d'affaires entre N − 1 et N, ou rapport Chiffre d'affaires sur Effectif entre N − 1 et N), utilisant éventuellement les enveloppes (cf. *chapitre 3*).

- **Par différence**, pour contrôler les redressements individuels (par exemple : mesure de l'écart entre le chiffre d'affaires initial et le chiffre d'affaires obtenu après le passage du CQR).

2.7.1 Macro-contrôles de type « contribution » (ou « drop out »[52])

Comme on vient de le voir, ils s'appliquent soit à des écarts en évolution soit à des écarts contemporains. En ce qui concerne les macro-contrôles en évolution, il faut distinguer les entreprises qui participent au calcul des agrégats (N – 1) et N (l'assiette du macro-contrôle) et les entreprises qui peuvent être sélectionnées par le macro-contrôle (entreprises éligibles).

Les agrégats seront sectoriels sauf dans le cas des macro-contrôles portant sur l'évolution du Chiffre d'affaires d'une branche[53] ; dans ce cas les agrégats utiliseront le Chiffre d'affaires ventilé par activité (ou par produit).

Un macro-contrôle est rattaché à au plus deux variables actives (par exemple : Chiffre d'affaires en N, Effectif en N). Pour chaque macro-contrôle, la ou les variables actives seront identifiées. Il porte toujours sur un ratio (soit entre 2 variables contemporaines, soit sur l'évolution temporelle d'une variable.

dans la diffusion de Ésane, qui seront exposées au chapitre 9.
[52]. Car on compare ce qui se passerait si on « enlevait » l'unité qu'on contrôle avec ce qui se passe quand elle est là. On mesure donc son « influence » sur un agrégat.
[53] Sur la distinction – fondamentale en statistique d'entreprises – entre secteurs et branches, voir par exemple les notes 2 et 3 de l'Avant-propos ou le schéma du § *2.7.3.3*.

Un macro-contrôle est lié à un niveau d'agrégation.

Les contributions seront calculées à deux niveaux de nomenclature *paramétrables*, parmi différentes valeurs fixées a priori :

- au niveau sectoriel uniquement (à 3 ou 5 caractères) ;
- au niveau sectoriel (à 3 caractères) × tranche de taille (en 5 tranches de tailles).

2.7.2 Le repérage des évolutions atypiques

Le principe du macro-contrôle « drop out » tel qu'il a été décrit ci-dessus est de repérer les entreprises ayant un impact fort sur l'évolution d'un agrégat diffusé.
En d'autres termes, outre les entreprises ayant un fort poids dans l'agrégat, le macro-contrôle permet aussi d'identifier les unités qui ont une évolution très différente de celle de leur secteur et qui impacte cette dernière.

Prenons l'exemple du secteur suivant :

	Valeur N-1	Valeur N
entreprise A	1000	900
entreprise B	900	700
Entreprise C	700	1200
Entreprise D	800	710000
Entreprise E	1500	1300
total	**4900**	**714100**

A priori, il y a, une erreur de saisie pour l'entreprise D (qui a probablement renseigné son questionnaire en euros au lieu de K€[54]). On souhaite donc corriger en premier le cas de l'entreprise D qui fait varier à elle seule l'agrégat du secteur, le rendant donc très variable selon que l'entreprise D est ou n'est pas prise en compte.

Pour cela, une première phase de « **repérage » des évolutions atypiques** est mise en œuvre (décrite dans le § *2.9*) afin d'identifier les cas semblables à ceux de l'entreprise D et de ne commencer la phase des macro-contrôles proprement dite qu'une fois les agrégats jugés stabilisés.

Ce problème de variabilité des agrégats est d'ailleurs accentué dans le cadre **du contrôle des résultats de l'enquête**. En effet, contrairement aux liasses fiscales pour lesquelles le premier chargement des données permet d'avoir des agrégats suffisamment stables en termes de recouvrement (on a plus de 80% de la valeur ajoutée[55]), pour l'enquête, les questionnaires rentrent au fil de l'eau et les plus grosses entreprises ne répondent pas forcément les premières. Ainsi, il est nécessaire d'attendre d'avoir **un minimum de réponses validées** pour pouvoir avoir des agrégats suffisamment **représentatifs et robustes**. Ce seuil a été fixé à 30% de questionnaires validés mais ceci peut cacher des déséquilibres sectoriels importants. Et même si un traitement des non-réponses, basé sur les réponses N-1, permet d'obtenir un agrégat total du secteur, ce dernier reste fragile tant que les plus grosses entreprises n'ont pas répondu.

[54] De telles erreurs de saisie sont normalement redressées automatiquement car repérées par des micro-contrôles temporels aberrants. Un tel cas ne peut se produire que si l'ensemble du questionnaire est erroné et que des équilibres comptables sont respectés ce qui permet de qualifier la variable. Comme on l'a en effet vu précédemment, les micro-contrôles sont suffisamment lâches pour éviter des redressements abusifs, les macro-contrôles devant permettre de repérer les « erreurs ».
[55] Ceci vient du fait que les résultats des grosses entreprises figurent dans le fichier dit « des anticipées », livré en juin, car elles sont dans l'obligation de télé déclarer, et leur traitement par la DGFiP est donc plus rapide.

Une fois l'entreprise D corrigée, la situation se présente ainsi :

	Valeur N-1	Valeur N
entreprise A	1000	900
entreprise B	900	700
Entreprise C	700	1200
Entreprise D	800	710
Entreprise E	1500	1300
total	**4900**	**4810**

C'est désormais l'entreprise C qui apparaît comme atypique et que le macro-contrôle doit donc désigner comme « à vérifier » : c'est en effet la seule entreprise de la strate qui voit sa caractéristique augmenter fortement alors que toutes les autres unités de la strate voient leur caractéristique baisser ! Voyons maintenant comment on peut calculer les contributions qui vont mettre en évidence de tels cas.

2.7.3 Calcul des agrégats et des contributions

On distingue trois types d'agrégats :

- agrégats sectoriels en niveau (ne concernent que l'année N, quelle que soit la variable que l'on agrège)
- agrégats sectoriels en évolution (concernent l'évolution entre N−1 et N de variables sectorielles)
- agrégats de branche en évolution (idem mais pour des branches qui par définition couvrent l'ensemble des secteurs)

Le détail du calcul des agrégats et des contributions est décrit ci-dessous.
Les formules présentées concernent le cas d'un échantillon (enquête Esa). Dans le cas de données exhaustives (cas du contrôle des liasses fiscales), il suffit de remplacer l'échantillon par la population totale et considérer que le poids est égal à 1 dans les formules.

Le traitement de l'enquête est particulier dans la mesure où le traitement de la non-réponse totale se fait par repondération et calage pour la diffusion des résultats (cf. *§ 5.6*). Toutefois, pour gagner du temps dans les développements au début du programme Resane, il avait été décidé de conserver la logique de traitement de la non-réponse de l'EAE, fondée sur des imputations des valeurs manquantes, pour le contrôle des données de l'Esa. Ainsi, le traitement de la non-réponse de l'Esa dans la base de production se base sur une imputation par N-1 lorsque la réponse de l'entreprise existait et était de bonne qualité en N-1 et par hot Deck sinon (cf. *§ 5.5*).
Le calcul des agrégats pour les macro-contrôles utilise ces valeurs imputées auxquelles on applique le poids de sondage au lancement de l'enquête, appelé par la suite « poids avant calage ». En revanche comme ceci a déjà été évoqué plus haut, seules les entreprises répondantes se voient calculer des contributions.

2.7.3.1 Agrégats sectoriels en niveau (Année N)

2.7.3.1.1 Assiette

Bien qu'on travaille sur des agrégats en niveaux, on va en fait se recaler sur les évolutions de l'agrégat par rapport à sa valeur de l'année précédente qui, elle, est connue et stable. La difficulté est en effet que les agrégats spontanés qu'on peut calculer au fur et à mesure de la rentrée des questionnaires en cours de campagne fluctuent beaucoup en fonction du nombre de questionnaires rentrés et sont en particulier sensibles aux données extrêmes (et qui arrivent non corrigés).
Afin de les stabiliser, et d'obtenir ainsi des contributions qui soient robustes aux points extrêmes et donc des contributions relativement « stables » au cours de la campagne, on utilise la formule de calcul ci-dessous fondée sur la médiane des évolutions par rapport à N-1 de la variable que l'on veut contrôler ; on sait que la médiane est robuste aux valeurs extrêmes.

Sont ainsi prises en compte toutes les entreprises présentes à la fois en N et en N-1 et classées dans le secteur S en N. C'est-à-dire toutes les unités présentes en N et en N-1 pour lesquelles l'APE utilisée pour la diffusion[56] est une APE du secteur S en N (x tranche de taille éventuellement).

[56] Une description précise des différentes APE utilisées dans Ésane figure au *§ 5.4.3* . Schématiquement, comme l'enquête sert à la mise à jour des codes APE des entreprises, l'APE utilisée pour la diffusion est l'APE

2.7.3.1.2 Formule de calcul de l'agrégat

L'agrégat en N de la grandeur Y sur le secteur S est calculé selon la formule suivante :

$$\boxed{Y_S^N \approx Y_S^{N-1} * \mathrm{Med_evo}_Y(S)}$$

où $\mathrm{Med_evo}_Y(S)$ représente la médiane des évolutions individuelles de la variable Y entre N et N-1 pour les unités (hors enveloppes) appartenant au secteur S l'année N et présentes en N-1.

Pour l'agrégat N − 1, Y_S^{N-1}, on cherche à se placer sur un champ comparable à celui de N pour pouvoir lui appliquer des évolutions. Sont donc prises en compte :

- Toutes les entreprises hors enveloppe[57] en N avec leurs données N − 1, qui :
  - ne sont pas en cessation au cours de l'année N,
  - et étaient classées dans le secteur cible S l'année N − 1 et n'ont pas été reclassées dans un autre secteur en N, ou sont classées dans le secteur cible en N et étaient classées ailleurs en N − 1. Pour le calcul de l'agrégat N-1 on utilise le poids calculé après traitement de la non-réponse par repondération en N-1 (qui sera explicité au § 5.6.2.4). Ainsi en N-1 seules les unités répondantes en N-1 sont prises en compte.
- Les valeurs N-1 des enveloppes de restructuration (cf. chapitre 3) classées dans le secteur S en N.

Y_S^{N-1} est ainsi calculé selon la formule :

$$\boxed{\begin{aligned} Y_S^{N-1} = & \sum_{i \in \text{Échantillon N-1}} (\mathrm{POIDS_APRES_CALAGE}_i^{N-1} \times Y_i^{N-1} \times (I_{i \in S}^N + I_{i \in S}^{N-1} \times I_{i \notin \text{Échantillon N}}) \times I_{i \notin mort}^N \times I_{i \notin env}^N) \\ & + \sum_{e \in \text{Enveloppes}} (Y_e^{N-1} \times I_{e \in S}^N) \end{aligned}}$$

où $\mathrm{POIDS_APRES_CALAGE}_i$ sont les poids de sondage après le calage et

$$I_{x \in S}^N = \begin{cases} 1 \text{ si l'entreprise (ou l'enveloppe) x appartient au secteur S en N} \\ 0 \text{ sinon} \end{cases}$$

$$I_{i \in S}^{N-1} = \begin{cases} 1 \text{ si l'entreprise } i \text{ appartient au secteur S en N-1} \\ 0 \text{ sinon} \end{cases}$$

$$I_{i \notin \text{Échantillon } N} = \begin{cases} 1 \text{ si l'entreprise } i \text{ n'appartient pas à l'échantillon en N} \\ 0 \text{ sinon} \end{cases}$$

$$I_{i \notin nais}^N = \begin{cases} 1 \text{ si l'entreprise } i \text{ n'est pas née en N} \\ 0 \text{ sinon} \end{cases}$$

$$I_{i \notin mort}^N = \begin{cases} 1 \text{ si l'entreprise } i \text{ n'est pas morte en N} \\ 0 \text{ sinon} \end{cases}$$

$$I_{i \notin env}^N = \begin{cases} 1 \text{ si l'entreprise i n'appartient pas à une enveloppe en N} \\ 0 \text{ sinon} \end{cases}$$

En d'autres termes, les créations/cessations observées l'année N (y compris en cours d'année) ne sont pas prises en compte. Les entreprises ayant changé de secteur entre N-1 et N seront considérées comme faisant partie du secteur d'arrivée (N) pour les deux années pour travailler à

après son éventuelle mise à jour en cours de campagne. Celle-ci est faite à chaque passage d'un CQR qui peut modifier la ventilation du chiffre d'affaires qui est à la base du calcul de l'APE.

[57] Cette notion d'enveloppe est décrite en détail au chapitre 3 de ce document.

assiette constante : leurs données N − 1 seront donc incluses dans le calcul de l'agrégat N − 1 du secteur d'arrivée N.

De même, les unités de la partie renouvelée[58] de l'échantillon (N-1) (c'est-à-dire qui n'appartiennent pas à l'échantillon N) sont prises en compte ; en revanche, les cessations au cours de l'année N de la partie conservée de l'échantillon ne sont pas prises en compte pour l'agrégat N − 1.
Par principe, l'hypothèse faite est que le processus de production permet d'identifier l'ensemble des restructurations d'une année. Les enveloppes (qui sont bien prises en compte dans l'agrégat N-1 mais pas dans la médiane) ont donc un poids de 1. De ce fait, pour le calcul des valeurs des caractéristiques d'une enveloppe, le poids de chaque entreprise de l'enveloppe est de 1.

Par exemple dans le cas du chiffre d'affaires, la formule devient :

$$CA_S^N = CA_S^{N-1} * \text{Med_évo}_{CA}(s)$$

2.7.3.1.3 Contribution pour les macro-contrôles contemporains

La formule de calcul d'une **contribution pour une unité i_0** est la différence (prise en valeur absolue) entre le ratio contemporain considéré et **le même ratio calculé sans l'unité i_0**, à savoir :

$$\boxed{CTR(i_0) = \left| \frac{Y_S^N}{X_S^N} - \frac{Y_S^N - POIDS_AVT_CALAGE_{i_0}^N \times Y_{i_0}^N}{X_S^N - POIDS_AVT_CALAGE_{i_0}^N \times X_{i_0}^N} \right|}$$

où X_S^N est défini de la même façon par :

$$X_S^N \approx X_S^{N-1} * \text{Med_evo}_X(S)$$

Les évolutions ne peuvent être calculées que sur des entreprises présentes à la fois en N et en N-1 ; en revanche, les estimateurs ainsi obtenus Y_S^N et X_S^N sont des estimateurs de l'agrégat, hors créations toutefois, donc qui sous-estiment (pour des variables positives) les vrais agrégats. On les utilise en tout cas pour toutes les entreprises i_0 du secteur S[59]. En d'autres termes la population éligible est bien celle de tout le secteur S.

Par exemple, dans le cas du ratio Chiffre d'Affaires (CA) sur Effectif (Eff), la formule devient :

$$CTR(i_0) = \left| \frac{CA_S^N}{EFF_S^N} - \frac{CA_S^N - POIDS_AVT_CALAGE_{i_0}^N \times CA_{i_0}^N}{EFF_S^N - POIDS_AVT_CALAGE_{i_0}^N \times EFF_{i_0}^N} \right|$$

Vu la façon dont ils sont calculés, CA_S^N et EFF_S^N ne vont pas bouger beaucoup si un questionnaire atypique pour une entreprise I_0 rentre. En revanche le terme [poids * $CA(I_0)$] sera lui atypique et contribuera à faire ressortir sa contribution.

[58] L'échantillon de l'enquête Esa est constitué de 2 parties : une partie dite exhaustive, interrogée tous les ans et une partie dite échantillonnée qui est renouvelée par moitié tous les deux ans. Cf. § *5.1.4*.
[59] Sous l'hypothèse implicite que les créations d'une année ne devraient pas constituer le gros des agrégats économiques contrôlés. L'important pour ces contrôles est d'avoir le bon ordre de grandeur.

2.7.3.2 Agrégats sectoriels en évolution (N / N-1)

2.7.3.2.1 Assiette

On utilise les mêmes formules que précédemment ; l'assiette et les entreprises éligibles sont donc les mêmes que dans le cas précédent.

2.7.3.2.2 Formule de calcul

Les formules de calcul des agrégats en N et N − 1 de la grandeur Y sur le secteur S sont donc de même :

$$\boxed{Y_S^N \approx Y_S^{N-1} * \mathrm{Med_evo}_Y(S)}$$

et

$$\boxed{\begin{aligned} Y_S^{N-1} &= \sum_{i \in \text{Échantillon N-1}} (\mathrm{POIDS_APRES_CALAGE}_i^{N-1} \times Y_i^{N-1} \times (I_{i \in S}^N + I_{i \in S}^{N-1} \times I_{i \notin \text{Échantillon N}}) \times I_{i \notin \text{mort}}^N \times I_{i \notin \text{env}}^N) \\ &+ \sum_{e \in \text{Enveloppes}} (Y_e^{N-1} \times I_{e \in S}^N) \end{aligned}}$$

2.7.3.2.3 Contribution

La formule de calcul d'une contribution pour une unité i_0 est, de la même façon que précédemment :

$$\boxed{\mathrm{CTR}(i_0) = \left| \frac{Y_S^N}{Y_S^{N-1}} - \frac{Y_S^N - \mathrm{POIDS_AVT_CALAGE}_{i_0}^N \times Y_{i_0}^N}{Y_S^{N-1} - \mathrm{POIDS_AVT_CALAGE}_{i_0}^N \times Y_{i_0}^{N-1}} \right|}$$ soit

soit :

$$\boxed{\mathrm{CTR}(i_0) = \left| \mathrm{Med_evo}_Y(Y) - \frac{Y_S^N - \mathrm{POIDS_AVT_CALAGE}_{i_0}^N \times Y_{i_0}^N}{Y_S^{N-1} - \mathrm{POIDS_AVT_CALAGE}_{i_0}^N \times Y_{i_0}^{N-1}} \right|}$$

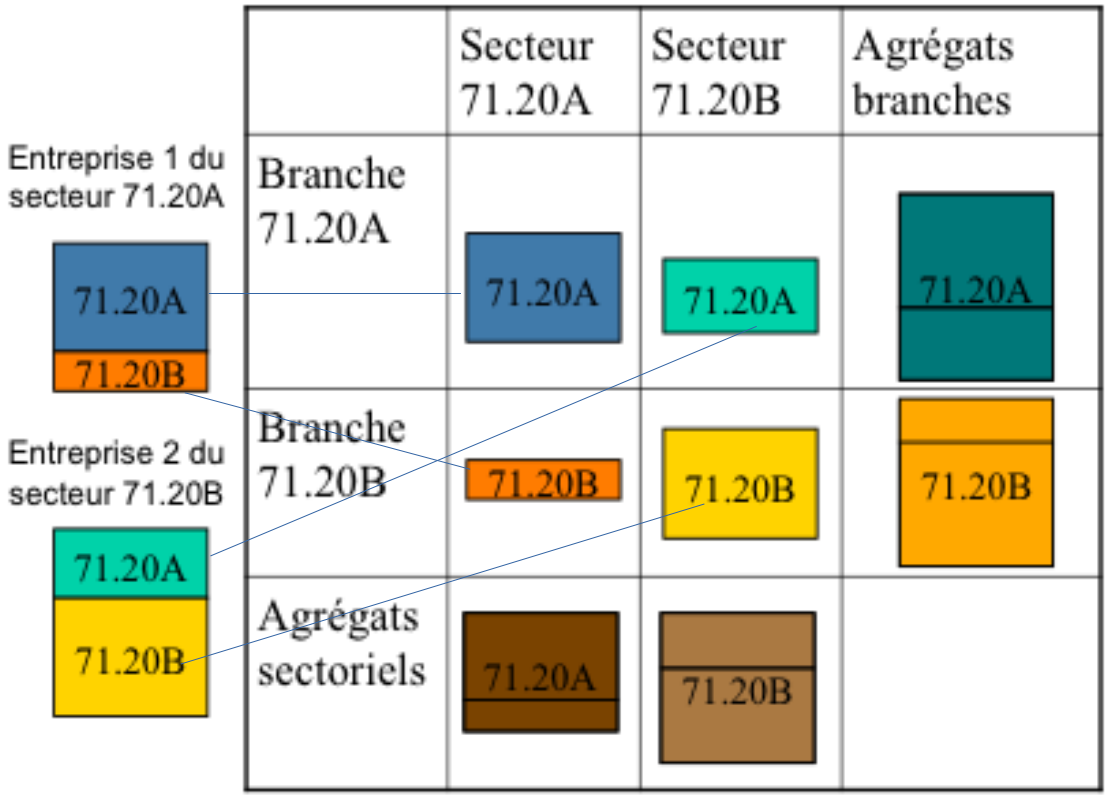
Pour mesurer l'évolution individuelle, il semble préférable de le faire sans faire intervenir les corrections de poids mais uniquement en mesurant l'impact des déclarations. C'est pour cela qu'on utilise **le même poids en N et N-1** et on prend celui de N avant calage car au moment des contrôles on ne dispose pas du poids après calage N.

Par exemple, dans le cas de l'évolution du CA, la formule devient :

$$\mathrm{CTR}(i_0) = \left| \frac{CA_S^N}{CA_S^{N-1}} - \frac{CA_S^N - \mathrm{POIDS_AVT_CALAGE}_{i_0}^N \times CA_{i_0}^N}{CA_S^{N-1} - \mathrm{POIDS_AVT_CALAGE}_{i_0}^N \times CA_{i_0}^{N-1}} \right|$$

2.7.3.3 Agrégats de branche en évolution (N et N−1)

Le schéma ci-après montre la différence d'approche entre les agrégats secteurs et branches :



Le schéma figure deux entreprises des secteurs 71.20A et 71.20B. Leur chiffre d'affaires est réparti entre ces deux secteurs mais pour la première, sa part dans le secteur 71.20A en fait son activité « principale » quand c'est l'inverse pour la seconde. Par définition, la part de chiffre d'affaires dans une activité autre que l'activité principale d'une entreprise constitue une de ses « branches ». La surface des rectangles est proportionnelle aux montants de chiffre d'affaires concernés. Le schéma montre sur un cas simple à 2 entreprises ce que sont les agrégats sectoriels et de branche.

Le principe de calcul des agrégats branche en évolution est le même que celui décrit pour le calcul des agrégats sectoriels décrit précédemment.

La seule distinction vient du fait que la sommation sur (N-1) **n'est plus contrainte par le secteur**.

L'agrégat est obtenu par sommation des entreprises qui réalisent du chiffre d'affaires dans la branche concernée et ce quel que soit leur secteur d'activité.

2.7.4 Macro-contrôles par différence

Le principe est de vérifier, pour chaque unité et pour chaque variable, que les redressements effectués n'ont pas été « trop forts », c'est-à-dire visibles au niveau de l'agrégat. Le redressement va en effet naturellement « normaliser » la réponse des entreprises, mais il ne faudrait pas arriver à l'extrême inverse, à savoir supprimer les réponses atypiques mais correctes. L'objectif de ce contrôle est de **détecter les entreprises et les variables** pour lesquelles le redressement, même s'il est fort, est juste car si ce redressement est destiné à corriger une erreur de saisie ou d'unité de réponse, il ne doit pas corriger un comportement atypique. Pour cela, pour toute unité i, une contribution est calculée à partir de l'écart entre la valeur redressée de la variable Y et la valeur brute de Y.

La formule de calcul est la suivante[60] :

$$\boxed{CTR(i_0) = POIDS_AVT_CALAGE_i \times \left| Y_i^{red} - Y_i^{brute} \right| / Y_S^N}$$

où Y_S^N est l'agrégat en niveau du secteur S pour l'année N de la variable Y. Cet agrégat peut être calculé au niveau groupe, sous-classe ou groupe croisé avec les tranches de taille.

[60] Bien que ce ne soit pas mathématiquement obligatoire, on notera que cette contribution est très généralement inférieure à 1 car on a au dénominateur un terme qui a la dimension de sommes du type « poids * Y_i ».

Le principe **est ensuite le même** que pour les autres macro-contrôles : le calcul des contributions pour toutes les entreprises i permettra de repérer les entreprises les plus fortement contributrices (Cf. § 2.7.6 ci-après). Un indicateur de priorité de traitement (Cf. § 2.8) permettra alors de sélectionner les messages d'erreur qui signaleront aux gestionnaires les contributions les plus suspectes que ceux-ci devront alors valider ou au contraire corriger.

2.7.5 La définition des seuils des macro-contrôles [3.2], [3.3]

2.7.5.1 Les principes théoriques

La difficulté est d'initialiser le système car on peut ensuite s'appuyer sur les seuils définis les années précédentes soit pour les relâcher soit pour les durcir.

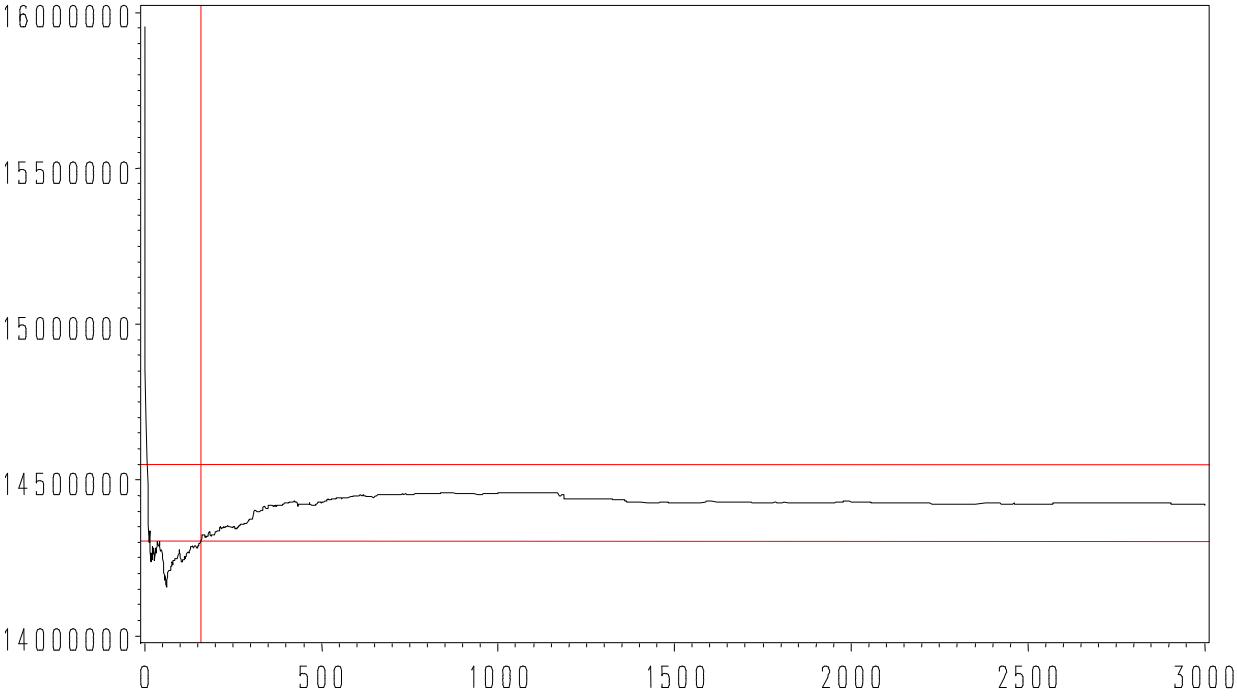
La méthode a consisté à procéder par simulation à partir des données des enquêtes EAE passées[61]. En effet, on disposait sur le passé (en pratique sur les années 2007 et 2006) à la fois des données brutes, à contrôler, et des données finales, considérées comme contrôlées manuellement et donc correctes[62], sur l'ensemble des unités.

Grâce à ce matériau, il a été possible de calculer **rétrospectivement** chaque score à partir des données brutes, puis d'observer l'évolution de chaque agrégat (dans l'exemple ci-après, le chiffre d'affaire de la branche « commerce de détail spécialisé ») en fonction du nombre d'entreprises contrôlées (i.e. pour lesquelles la donnée brute initiale a été remplacée par une donnée finale différente), celles-ci étant classées par ordre de score décroissant :

- on part de l'agrégat calculé à l'aide de toutes les valeurs brutes, avant toute correction ;
- puis, on considère la valeur de l'agrégat obtenu en corrigeant l'entreprise correspondant à la contribution la plus élevée ;
- puis la valeur de l'agrégat obtenu en corrigeant la suivante et ainsi de suite.

On obtient alors une courbe similaire à celle du graphique 1, de laquelle on déduit, par examen visuel, la valeur du seuil : ce dernier est choisi de manière à rendre négligeable l'erreur de mesure résiduelle, qui résulte de la non-vérification de certains questionnaires.

Graphique 1 : Évolution de l'estimateur du chiffre d'affaires 2007 de la branche « Commerce de détail alimentaire en magasin spécialisé » en fonction du nombre d'entreprises contrôlées



[61] Pour les macro contrôles sur les liasses fiscales, on a procédé de même, mais à partir des données de SUSE. Les données fiscales posent en pratique moins de problème car toutes les entreprises ont un poids de 1.

[62] Le contrôle des données dans les EAE reposait principalement sur des micro-contrôles et ceux-ci étaient presque exhaustifs. (Ils requéraient donc davantage de ressources en gestionnaires. L'objectif de Resane d'atteindre une qualité du même ordre en économisant des moyens humains impliquait de nouvelles pratiques).

Sur cet exemple, on constate que le contrôle des 500 unités les plus influentes pour le score local considéré – soit moins de 20% du total des unités contrôlables – suffit pour obtenir un estimateur robuste : **l'impact de contrôles supplémentaires sur des unités moins influentes est quasiment nul**. En conséquence, le seuil de ce contrôle peut être défini comme la valeur du score de la 500ème unité.

Le nombre de seuils à calculer dans le cadre du système Ésane se comptant en centaines de milliers[63], il était évidemment impossible de procéder par examen visuel pour chaque série ; la méthode artisanale de détermination des seuils décrite ci-dessus a donc été automatisée.

Cette procédure automatique s'appuie toujours sur l'analyse de l'évolution de la statistique d'intérêt en fonction du nombre d'entreprises contrôlées, mais le caractère négligeable de l'erreur de mesure résiduelle est cette fois-ci **mesuré à l'aune de l'erreur d'échantillonnage** : le seuil est automatiquement déterminé **de telle sorte que l'erreur de mesure résiduelle soit dominée par l'écart-type de l'estimation**. En pratique, la valeur du seuil correspond au score de la première entreprise (repérée par la droite verticale rouge dans l'exemple du graphique 1 ci dessus) à partir de laquelle l'estimation reste à l'intérieur d'un intervalle centré sur l'estimateur final et de magnitude égale à un pourcentage donné de l'écart-type de l'estimation (intervalle délimité par les deux droites horizontales rouges sur le graphique 1).
Dans l'exemple ci-dessus, on voit qu'à partir de la 153ème entreprise, le chiffre d'affaires évolue entre son estimation finale plus ou moins une « erreur » qui est du même ordre que l'erreur d'échantillonnage. Il ne servirait donc à rien de contrôler davantage de données en espérant améliorer la précision du chiffre d'affaires de la branche. On voit aussi que par rapport aux EAE où on contrôlait environ 3000 entreprises, on n'a plus à en contrôler que 20 fois moins pour une qualité a priori équivalente.

2.7.5.2 Mise en œuvre pratique

L'application de cet algorithme aux macro-contrôles du système Ésane a conduit dans la majeure partie des cas à des résultats satisfaisants et conformes à l'esprit de la vérification sélective des données : les seuils ainsi déterminés **amènent à contrôler un nombre d'unités en général assez restreint**.
Toutefois, cette procédure automatique se révèle évidemment sensible à la qualité des données passées. Ainsi, il arrive, pour certains secteurs, que les erreurs de mesure soient dès le départ négligeables devant l'erreur d'échantillonnage – soit parce que la variance de l'estimateur était très importante sur le passé, soit parce les entreprises avaient directement bien répondu lors de l'enquête précédente – ce qui conduit à des seuils particulièrement élevés, d'où un risque de sous-contrôle. Afin de prévenir ce risque, un « filet de sécurité » a été mis en place, sous la forme d'un niveau de seuil maximal toléré, ce qui assure qu'aucune unité importante ne passe au travers des contrôles.

Dans certains cas concrets, lors de la première campagne en 2008, il a été nécessaire de **relever empiriquement certains seuils** car ils aboutissaient à un trop grand nombre de dossiers à contrôler.

2.7.6 Repérage des entreprises fortement contributrices

Comme vu précédemment, ce sont les macro-contrôles qui structurent le plus profondément le travail des gestionnaires. Lorsqu'une entreprise est transmise à un gestionnaire pour examen, différents messages de contrôle des données peuvent apparaître pour indiquer au gestionnaire qu'il doit procéder à un examen du cas de l'entreprise.

Les différents **statuts techniques** d'un macro-contrôle pour une caractéristique donnée sont le résultat de deux conditions : **dépasser (ou non**) un seuil d'influence d'une part (présent / non présent) et concerner des caractéristiques **non validées (ou validées**) d'autre part (actif[64] /inactif).

Les macro-contrôles peuvent donc prendre les **trois statuts** suivants[65] :

---

[63] Autant qu'il y a de croisements possibles : type de score # variable à contrôler # niveau de validation

[64] On dit que le macro contrôle est actif au sens où si une des variables au moins sur lesquelles il porte n'est pas validée, le gestionnaire a quelque chose à faire (valider ou invalider la variable). Une fois que la caractéristique est validée, le contrôle n'est plus utile.

[65] La combinaison « non présent » « actif » n'est pas possible puisque si la contribution est inférieure au seuil (macro-contrôle non présent) le gestionnaire n'a rien à faire et ne peut pas valider ou invalider la ou les variables en cause.

- « non présent » « inactif » (la contribution est inférieure au seuil) ;
- « présent » « actif » (la contribution est supérieure au seuil et l'une au moins des variables du macro-contrôle n'est pas validée) ;
- « présent » « inactif » (la contribution est supérieure au seuil et les variables actives macro-contrôle sont toutes validées). Ces messages concernent donc les unités qui ont un comportement atypique. Comportement qui a été « validé » par le gestionnaire après examen des données voire rappel de l'entreprise.

Les contributions, les messages et leur statut sont recalculés lors de chaque passage du programme du CQR. Cela permet instantanément au gestionnaire de voir sur son écran l'impact de la modification d'une caractéristique dans la mesure où il a la possibilité de déclencher manuellement un CQR sur l'unité qu'il est en train de contrôler.

Les statuts (présent actif, présent inactif, non présent) permettent le calcul de l'indicateur de priorité de traitement. Pour cela, **on recalculera les priorités de traitement** indiquées plus bas (voir ci-après le § *2.8*) **en tenant compte uniquement des données non validées**.

Ainsi, **si le dossier est jugé finalement non prioritaire, il ne sera jamais à traiter par un gestionnaire** même si au moins une variable de sa réponse a dépassé le seuil de contribution d'un macro-contrôle la concernant.

2.8 Indicateur de priorité de traitement

Pour chaque ratio contrôlé par un macro-contrôle de type écart et pour chaque macro-contrôle par différence, on dispose d'un seuil de contrôle SC.

Pour les entreprises ou les enveloppes[66] (on parle alors de ***dossiers*** dans Ésane) dont la contribution dépasse ce seuil pour la variable d'intérêt X_i, on considère qu'il est important de valider la variable en question. On peut alors caractériser le dossier pour cette variable par un « état », noté $E(X_i)$ qui prend alors la modalité « I » (comme important).

Pour gérer la production, il est nécessaire de calculer **une priorité globale** pour chaque dossier. L'objectif étant de faire **traiter en priorité** par les gestionnaires les **dossiers jugés les plus douteux**. Ainsi, on saura ordonner dans le temps les dossiers des plus importants vers les moins importants, de sorte qu'on soit sûr que les plus importants l'ont été, quand on doit fournir un résultat provisoire à un demandeur extérieur (Eurostat ou comptables nationaux par exemple). Comme on l'a vu en introduction, le nombre de dossiers potentiels est tel qu'il est de toutes façons impossible de les contrôler tous.

Le calcul de l'indicateur de priorité de traitement se fait donc en trois temps :

1. **Pour une variable donnée X_i**, on regarde le niveau de gravité maximal des macro-contrôles qui impliquent un ratio dans lequel cette variable est au numérateur, pour les macro-contrôles contemporains et temporels, ainsi que le macro-contrôle par différence[67]. On retient alors le degré de gravité le plus élevé comme caractérisant l'état $E(X_i)$ de la variable X_i. On ne considère **que les variables non validées par le gestionnaire**, c'est-à-dire les macro-contrôles présents actifs.

2. On agrège ensuite **les états de toutes les variables X_i** d'un même dossier avec la formule suivante pour définir un état global quantitatif QEG (Quantification de l'État Global), fondé sur les paramètres K_i :

$$QEG = \frac{\sum_{\text{var i}} K_i . \text{score_max}(X_i) . 1\mathrm{I}_{\{E(X_i)=I\}}}{\sum_{\text{var i}} K_i}$$

où score_max(X_i) est défini comme **la valeur maximum** entre les trois contributions suivantes :

[66] Voir le *chapitre 3*.
[67] On rappelle qu'une variable peut être impliquée dans plusieurs macro contrôles.

- la contribution relative au macro-contrôle par différence portant sur la variable X_i (en excluant les macro contrôles par différence sur la non-réponse partielle) ; on a vu en note de bas de page du § 2.7.4 que cette quantité était le plus souvent <1 ;
- la contribution relative au macro-contrôle drop-out temporel portant sur la variable X_i (en normalisant par l'évolution observée X_S^N / X_S^{N-1} d'ailleurs elle-même souvent proche de 1, pour que le résultat soit à coup sûr inférieur à 1);
- la contribution relative au macro-contrôle drop-out contemporain (normalisée de même par le ratio impliqué dans ce contrôle pour que le résultat soit inférieur à 1).

et où les K_i, fixés pour chaque variable, représentent **l'importance de chaque variable**. Ces paramètres sont par nature un choix de la maîtrise d'ouvrage qui reflète les priorités de l'Insee vis-à-vis du contrôle des variables. Ils ne résultent pas d'un calcul.

La présence des indicatrices fait qu'on ne prend en compte que les macro-contrôles correspondant aux variables qui dépassent un seuil de macro contrôle. La formule fait intervenir à la fois **le nombre de « problèmes » détectés** (c'est-à-dire le nombre de fois où l'unité i est influente sur divers macro-contrôles) et en même temps **l'ampleur de ces « problèmes »** par la « valeur » des contributions dans score_max : une unité qui présente une valeur extrêmement aberrante pour une seule variable (associée à un K_i non nul) peut quand même être repérée par un QEG important car la contribution de cette valeur aberrante sera élevée.
Le dénominateur permet de normer cet indicateur afin de le ramener à un ordre de grandeur proche de 1, le maximum du numérateur étant de $\sum_i K_i$, il correspond au cas où toutes les variables d'une entreprise dépasseraient le seuil élevé (SE) (et où tous les score_max seraient eux-mêmes égaux à 1 !).

3. On définit ensuite un état global qualitatif **EG, pour chaque entreprise** (ou dossier), à l'aide d'une codification en tranches de QEG :

EG=P(rioritaire) si *QEG* > s1,
EG=I(mportant) si s1>*QEG*>s2,
EG=S(econdaire) si s2>=*QEG*>s3
EG=A(automatisable) si s3>=*QEG*

avec s1, s2, s3 compris entre 0 et 1.

Seules les entreprises **dont la priorité globale est jugée suffisante** sont examinées par les gestionnaires[68]. Lorsqu'une entreprise est sélectionnée, toutes les variables pour lesquelles un macro-contrôle dépasse le seuil sont examinées. Ainsi, lorsqu'un dossier est à contrôler, il est demandé au gestionnaire, lors de son contact avec l'entreprise, de passer en revue **toutes les réponses de l'entreprise** qui ont semblé suspectes au vu des macro-contrôles.

Les seuils s1, s2 et s3 sont des **paramètres globaux de réglage** qu'utilise l'encadrement du service Ésane à Nantes, en relation avec la maîtrise d'ouvrage parisienne, pour gérer la stratégie de travail des gestionnaires, en fonction du calendrier. Ils ont vocation à être modifiés au cours d'une campagne, alors que pour les seuils SE et SM les modifications en cours de campagne doivent rester rares. Ces seuils SE et SM vont dépendre des niveaux de nomenclature, des variables, des ratios (temporels ou contemporains) alors que s1, s2, s3 sont unidimensionnels.

Quant aux K_i, s'agissant d'une priorité stratégique, leur modification doit rester exceptionnelle (et de la seule responsabilité de la maîtrise d'ouvrage).

[68] en particulier, on ignore les contributions élevées sur les variables que l'on considère de peu d'intérêt pour les utilisateurs, et pour lesquelles on a mis les paramètres K_i à zéro.

2.9 *Les contrôles complémentaires aux macro contrôles*

Comme on l'a vu en introduction, quelques micro-contrôles, du même type que ceux qui étaient mis en œuvre dans les EAE ont été maintenus et sont mis en œuvre par les gestionnaires pour les quelques cas où les macro-contrôles ne sont pas efficaces. Pour que les macros contrôles soient efficaces, il faut en effet qu'il y ait :

- soit au moins une caractéristique corrélée parmi les autres caractéristiques collectées afin de pouvoir mettre en place un macro-contrôle contemporain efficace ;
- soit une cohérence temporelle dans la réponse à la variable pour pouvoir mettre en place un macro-contrôle temporel efficace.

Pour certaines caractéristiques **telles que les investissements** par exemple, qui ne sont corrélées à aucune autre caractéristique comptable, aucune de ces deux conditions n'est vraiment respectée ce qui conduit à un contrôle inefficace de cette variable.

Pour de telles variables un contrôle simple **sous forme de liste des plus gros contributeurs** à l'agrégat **en niveau** ou la recherche de valeurs aberrantes (investissements nets fortement négatifs par exemple) **a donc été maintenu**.

Dans le même ordre d'idées, il est très difficile de contrôler par macro-contrôles des caractéristiques pour lesquelles l'agrégat est proche de 0 ou peut changer de signe. Des soldes, tel que l'Excédent Brut d'Exploitation par exemple, en sont le parfait exemple. Les contributions sont trop volatiles et leur interprétation devient impossible.
Pour de telles variables, il est préférable de se concentrer sur le contrôle des variables qui participent au calcul du solde et non au solde lui-même.

3 Le cas particulier des enveloppes de restructuration

On a été conduit, à plusieurs reprises, au chapitre précédent à mentionner l'existence « d'enveloppes de restructurations ». Une des grandes difficultés de la statistique d'entreprises est que les objets qu'elle observe, les entreprises, changent de contour, se déforment dans le temps, se « restructurent » par transfert d'activités. Ceci est très gênant dès que l'on veut comparer une situation entre deux instants t_1 et t_2, que ce soit à des fins d'analyse de résultats publiés ou, comme on l'a vu dans le chapitre précédent, à des fins de contrôle de ce qui s'est passé à l'année N, par rapport à l'année N-1, car on ne sait pas toujours comment interpréter les évolutions.

Les enquêtes structurelles (que ce soit les anciennes EAE comme désormais l'Esa ou l'EAP) se sont toujours intéressées au suivi des restructurations. Mais ***la prise en compte des restructurations dans les contrôles*** est une nouveauté d'Ésane, dont la description est l'objet de ce chapitre, qui vient compléter le chapitre précédent sur ce point. L'utilisation des enveloppes dans l'interprétation des évolutions des agrégats sera présentée au chapitre 9 (§ *9.6.1.2*) ; s'agissant d'un sujet très délicat, elle n'a cependant été diffusée qu'en interne et est principalement utilisée pour valider les résultats obtenus. Des évolutions surprenantes sont en effet parfois le signe d'une erreur non repérée durant le processus de production des données. Il est toutefois possible à terme de publier des évolutions utilisant les enveloppes.

Le plan du chapitre est d'abord de rappeler ce que sont les « restructurations » qui affectent les unités légales (§ *3.1.1*), puis le concept « d'enveloppes » qui permet de se ramener à champ constant entre une année N et une année N-1 même en présence d'une restructuration (§ *3.1.2*). On verra ensuite comment le système Ésane s'interface avec la base de données interne à l'Insee (dite Citrus) où sont stockées toutes les informations dont nous disposons sur les enveloppes (§ *3.2*). On terminera donc par l'utilisation des enveloppes aux fins de contrôle des données (§ *3.3*).

3.1 Qu'est-ce qu'une enveloppe de restructuration et pourquoi la prendre en compte dans les statistiques ?

3.1.1 Rappel sur les restructurations

Une restructuration économique se caractérise par un ensemble d'opérations qui a un impact sur l'activité productive courante (opérations de production et de formation brute de capital fixe) des entreprises par le biais de **transferts d'activités**.
Les restructurations les plus connues sont les **fusions**, les **absorptions, les sessions partielles d'actifs** et les **scissions partielles ou totales**. On peut bien sûr avoir une combinaison de plusieurs sortes de restructurations. Elles ont un impact non négligeable sur l'évolution entre deux années consécutives des agrégats sectoriels.

> Prenons l'exemple d'un secteur dans lequel toutes les unités voient leur chiffre d'affaires croître de 10% entre deux années consécutives et où en parallèle, la plus grosse entreprise du secteur qui pesait 50% du chiffre d'affaires en N-1 disparaît car elle est absorbée par une autre unité d'un secteur différent.
> Une simple comparaison des résultats en niveau des deux années consécutives peut conduire à un diagnostic erroné. En effet, entre les deux années, le chiffre d'affaires du secteur a baissé de 45 %. On pourrait donc conclure au fait que le secteur se porte mal alors qu'en réalité les entreprises du secteur sont toutes en croissance. Deux messages contradictoires sont donc possibles.

À partir de ce constat et afin de contrebalancer les discours qui se basent sur les évolutions apparentes[69], il a été décidé de prendre en compte, ou plutôt d'essayer de neutraliser, les restructurations afin de **rétablir des évolutions économiques sectorielles cohérentes** entre 2 années successives. Pour ce faire, l'objectif est de recalculer la valeur des caractéristiques en N-1 **en se plaçant dans la même structure en termes d'entreprise qu'en N**. il s'agit en quelque sorte de faire comme si la restructuration avait déjà eu lieu en N-1. Un nouveau concept d'unité statistique a

[69] Par évolution apparente il faut comprendre l'évolution entre les agrégats en niveau calculés en N-1 et N.

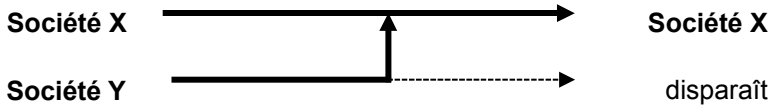
donc été créé. Il s'agit de **l'enveloppe de restructuration** qui est géré à l'Insee dans un outil baptisé Citrus[70].

3.1.2 La présentation de l'enveloppe et de ses caractéristiques propres

L'enveloppe **se compose de deux listes d'entreprises ou Siren** (entreprises de l'année N-1 avant la restructuration ; entreprises de l'année N après la restructuration). Elle peut être vue comme la somme des unités légales qui prennent part à une ou plusieurs restructurations. La règle de base étant qu'une entreprise ne peut appartenir qu'à une enveloppe et une seule, une enveloppe peut contenir des entreprises qui n'ont eu aucune relation directe entre elles.

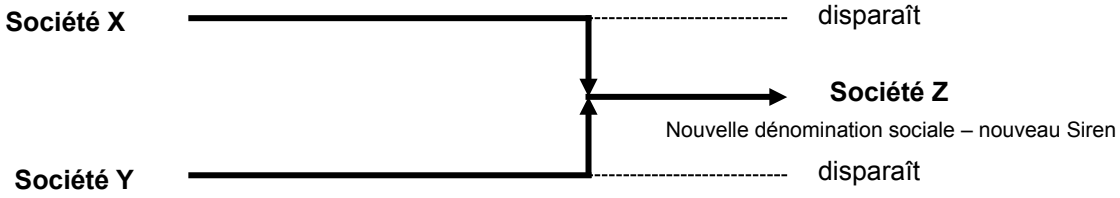
Exemples d'enveloppes :

- *Dans le cas d'une absorption* :



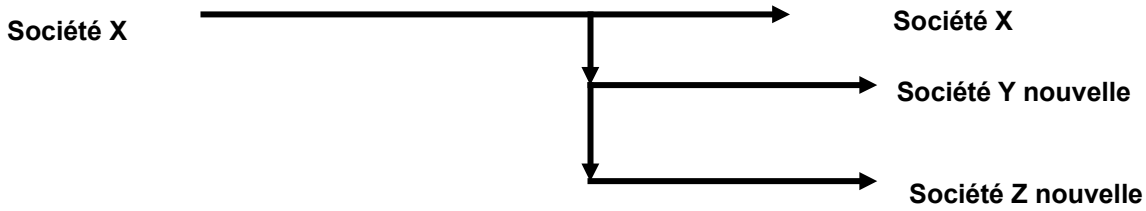
L'enveloppe contient les entreprises X et Y. En N, l'enveloppe correspond à l'entreprise X et en N-1, au cumul des entreprises X et Y.

- *Dans le cas d'une fusion* :



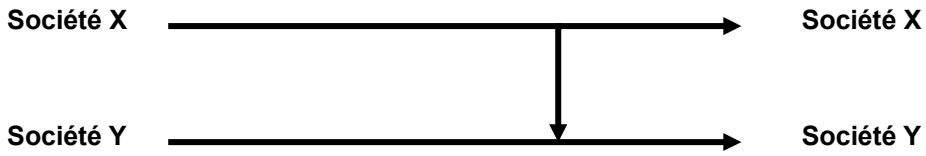
L'enveloppe contient les entreprises X, Y et Z. En N, l'enveloppe correspond à l'entreprise Z et au cumul des entreprises X et Y en (N-1).

- *Dans le cas d'une scission partielle* :



L'enveloppe contient les entreprises X, Y et Z. En N l'enveloppe correspond au cumul des entreprises X, Y et Z et en N-1 à l'entreprise X.

- *Dans le cas le plus fréquent d'un apport partiel d'actifs* :

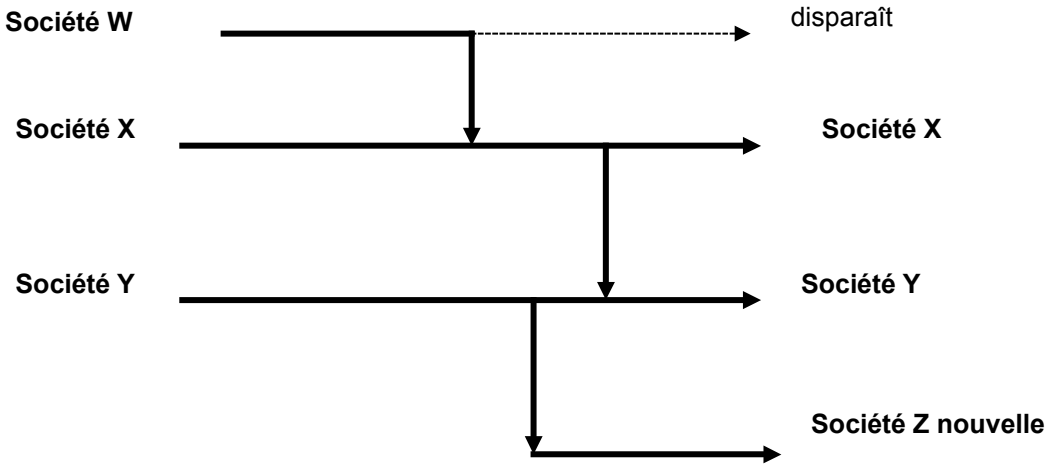


[70] Pour en savoir plus, il est possible de lire l'article ***Citrus - Système d'information sur les restructurations d'entreprises,*** M. Beauvois, Courrier des statistiques n° 95-96, 2000 :

La société X apporte une partie de ses actifs à la société Y mais il n'y a pas apport d'actif de Y en contrepartie. Le paiement est effectué en titre, en cash ou avec des financements hybrides.

L'enveloppe contient les entreprises X et Y. En N et en N-1, l'enveloppe correspond au cumul des entreprises X et Y

- *Dans un cas complexe :*



L'enveloppe contient les entreprises W, X, Y et Z. En N l'enveloppe correspond au cumul des entreprises X, Y et Z et en N-1 au cumul des entreprises W, X et Y.

À partir du moment où le contour d'une enveloppe est connu, il est ensuite possible de calculer des **caractéristiques** statistiques **au niveau de l'enveloppe**, caractéristiques qui doivent permettre de neutraliser l'effet de la restructuration. Il s'agit essentiellement de calculer :

- une APE pour l'enveloppe dont l'objectif est de neutraliser les évolutions sectorielles liées à la restructuration. L'APE est obtenue à partir des APE des unités légales qui la composent l'année N. Et afin de neutraliser des changements de secteurs liés à la restructuration, l'APE de l'enveloppe en N-1 est par principe identique à celle obtenue en N.
- un coefficient d'agrégation (appelé Kagre) dont l'objectif est de réussir à raisonner à structure constante entre les deux années, c'est-à-dire de permettre de neutraliser des flux entre unités qui sont présents (respectivement absents) avant la restructuration et qui ont disparu (respectivement sont apparus) après la restructuration. Ainsi, la comparaison des données N-1 et N de l'enveloppe traduira bien une évolution économique réelle du secteur de l'enveloppe et non une évolution fictive liée à un changement organisationnel au sein d'une entreprise. Cette neutralisation des flux se fait, dans Citrus, automatiquement, par l'estimation d'un coefficient d'agrégation. Ce dernier se base sur le principe que la valeur ajoutée est additive et donc indépendante de la structure d'une entreprise. Ainsi, toutes choses égales par ailleurs, c'est-à-dire indépendamment des évolutions temporelles de la valeur ajoutée liée à celle des variables[71], la somme des valeurs ajoutées des entreprises de l'enveloppe en N-1 est égale à la somme des valeurs ajoutées des entreprises de l'enveloppe en N. Ceci n'est pas du tout vrai au niveau du chiffre d'affaires (à cause d'éventuels échanges entre les unités de l'enveloppe en N et/ou N-1). Le coefficient d'agrégation vise à corriger le chiffre d'affaires N-1 au niveau de l'enveloppe **de telle sorte que l'évolution du chiffre d'affaires de l'enveloppe soit identique à celle de sa valeur ajoutée**. La formule exacte du coefficient d'agrégation (Kagre) est donc :

[71] Voir l'exemple ci-après pour comprendre sur un exemple ce qui est exprimé ici.

$$Kagre = \frac{CA_N^{env}}{CA_{N-1}^{env}} * \frac{VA_{N-1}^{env}}{VA_N^{env}}$$

où $CA_N^{env} = \sum_E CA_N^E * 1\mathrm{I}_{E \in env\, en\, N}$

avec CA_N^E le chiffre d'affaires d'une entreprise en N et

$$\mathbb{I}_{E \in \text{env en N}} = \begin{cases} 1 \text{ si l'entreprise E appartient à l'enveloppe "env" en N} \\ 0 \text{ sinon} \end{cases}$$

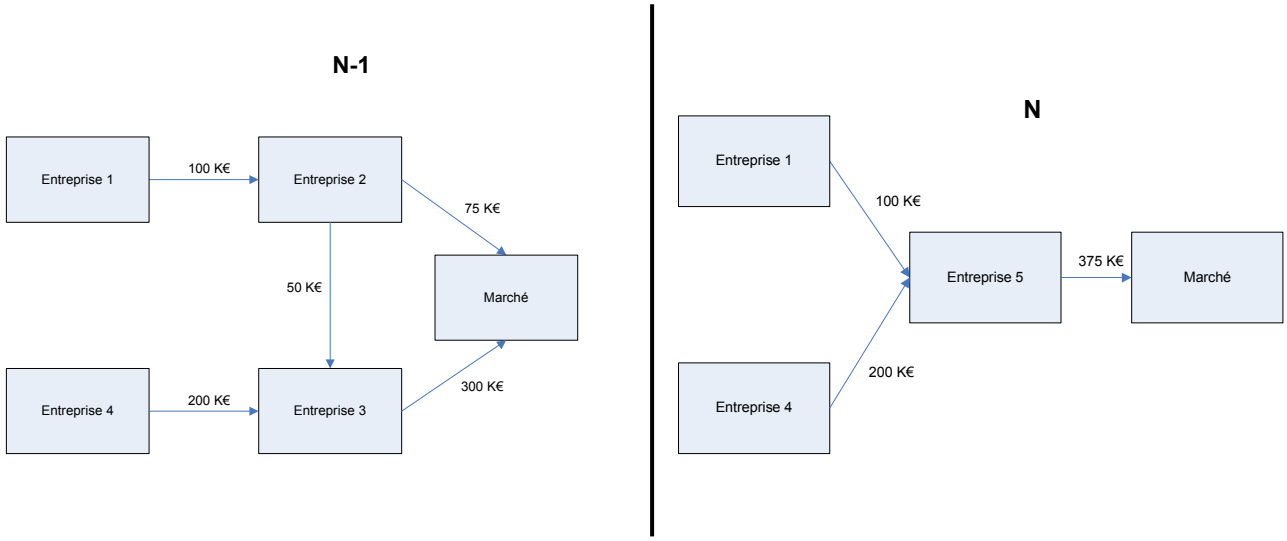
et $CA_{N-1}^{env} = \sum_E CA_{N-1}^E * 1\mathrm{I}_{E \in env\, en\, N-1}$

La valeur « corrigée » du chiffre d'affaires N-1 de l'enveloppe devient alors :

$$CAdef_{N-1}^{env} = Kagre * CA_{N-1}^{env}$$

Le coefficient d'agrégation s'applique ainsi à toutes les variables dites non additives (les différentes ventes et achats du compte de résultats par exemple). En revanche, les investissements par exemple connaissent un traitement particulier, décrit en fin de chapitre (Cf. § 3.3.3).

Exemple de calcul de coefficient d'agrégation dans le cas de la fusion des entreprises 2 et 3 dans l'entreprise 5 :



L'enveloppe associée à cette restructuration est composée des entreprises 2 et 3 en N-1 et de l'entreprise 5 en N. Les entreprises 1 et 4 sont des fournisseurs dans cet exemple.

En première approximation les achats et chiffres d'affaires de l'enveloppe en N-1 sont obtenus en sommant les achats et chiffres d'affaires des entreprises 2 et 3 et ceux de N correspondent à ceux de l'entreprise 5. « Toutes choses égales par ailleurs » signifie qu'on ne fait que comparer la structure des échanges ; **dans la réalité**, le montant des chiffres d'affaires en N et en N-1, ainsi que des valeurs ajoutées, **seront évidemment différents** et seront calculés à l'aide des valeurs observées des variables et **le calcul du Kagre en tiendra compte**.

On obtient donc le tableau suivant :

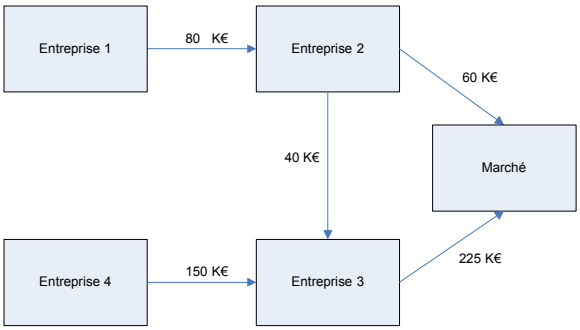
	N-1	N
Achats	100 + (200 + 50)[72]= 350	100 + 200 = 300
Chiffre d'affaires	(75 + 50)[73] + 300 = 425	375
Valeur ajoutée[74]	425-350 = 75	375 - 300 = 75

Comme prévu, on voit que, toutes choses égales par ailleurs, la valeur ajoutée est identique dans les deux configurations alors que le chiffre d'affaires baisse artificiellement à cause de la suppression du chiffre d'affaires réalisé par l'entreprise 2 avec l'entreprise 3 en N-1, chiffre d'affaires qui est devenu interne à l'entreprise 5 en N.

Dans cet exemple le coefficient d'agrégation vaut donc : (375*75)/(425*75) = 0.88
On en déduit que le chiffre d'affaires N-1 de l'enveloppe vaut : 0.88 * 425 = 375.

Dans la réalité, on n'est pas « toutes choses égales par ailleurs » et les chiffres en N-1 sont évidemment différents des chiffres en N. Imaginons que le CA « toutes choses égales par ailleurs » de l'entreprise 2 en N (soit 125k€) résulte d'une hausse de 25% (donc valait 100k€ en N-1) et celui de l'entreprise 3 (de 300k€) d'une hausse de 33%[75] (donc valait 225 k€). On fait de plus l'hypothèse que les taux de valeur ajoutée restent constants entre N-1 et N pour estimer des flux plausibles des entreprises 1 et 4 vers 2 et 3.
Les flux réellement observés deviennent :

N-1



La valeur ajoutée de l'enveloppe, observée en N-1 devient :
100-80 = 20k€ pour l'entreprise 2 et
225-190=35k€ pour l'entreprise 3, soit 55k€ pour l'enveloppe complète.

Le calcul du coefficient d'agrégation qui sera réellement fait dans Citrus sera donc :
Kagre = (375/325) x (55/75) = 0.8461

C'est un calcul qui corrige le chiffre d'affaires N-1 de l'enveloppe de façon à remplacer 325 par 325*0.8461=275k€

Cette valeur de 275k€ n'est pas la vraie valeur « consolidée » de l'enveloppe (qui serait 285k€) **mais elle en est beaucoup plus proche que la valeur observable de 325k€.**

[72] L'entreprise 3 achète 200 K€ à l'entreprise 4 et pour 50 à l'entreprise 2.
[73] L'entreprise 2 vend pour 50 K€ à l'entreprise 3 et pour 75 à d'autres acteurs économiques.
[74] Pour l'exemple on réduit le calcul de la valeur ajoutée à sa plus simple expression à savoir ventes moins achats.
[75] Chiffres certes peu réalistes mais qui permettent de vérifier les calculs aisément.

3.1.3 Les enveloppes et Ésane

Le processus de production Ésane fait intervenir les enveloppes à 3 niveaux :

i) Par le biais du questionnaire Esa qui dispose d'un cadre spécifique lié aux restructurations d'une part, et par l'examen des comptes des entreprises d'autre part, les gestionnaires Ésane vont repérer des restructurations d'entreprises. Ces dernières seront communiquées à Citrus qui sera alors mis à jour et créera, le cas échéant, de nouvelles enveloppes (Cf. *§* 3.2 sur l'initialisation des enveloppes dans Ésane). Sur ce point Ésane reprend les rôles anciennement tenus par les EAE, SUSE et le SIE.

ii) Pour la première fois en statistiques d'entreprises, les enveloppes sont prises en compte dans le processus de contrôle des données (cf. § 3.3 ci-dessous pour plus de détails). Cette prise en compte permet notamment :
   - de mieux cibler les unités à contrôler ;
   - de récupérer, par le biais des gestionnaires Ésane (ou ceux qui contrôlent l'EAP), des informations sur les flux existants entre les entreprises participant à une restructuration et ainsi d'améliorer l'estimation du coefficient d'agrégation qui a souvent, par le passé, été jugé de mauvaise qualité dans le cas des « grosses » restructurations ;
   - de valider les immobilisations corporelles échangées au moment de la restructuration (on parle ici « d'apport »). Cette information ne figure pas dans la liasse fiscale. Elle est obtenue via Citrus ou l'Esa. Elle permet ensuite de calculer l'investissement « hors apport » qui est celui qui est plus pertinent au niveau de l'analyse économique.

iii) Pour calculer des agrégats en évolution comparable et à champ constant et permettre ainsi de prendre en compte les restructurations comme facteur explicatif de l'évolution d'un secteur, même si ces agrégats ne sont utilisés qu'en interne. Ce point est détaillé au *§ 9.6.1.2*.

3.2 Les échanges Ésane – Citrus ; l'initialisation des enveloppes dans Ésane

L'ensemble des enquêtes auprès des entreprises – et entre autres les enquêtes infra-annuelles – peuvent donner l'occasion à des gestionnaires de collecter des informations sur les restructurations. Ils doivent alors **les centraliser dans la base de données Citrus**, qui a elle-même des canaux privilégiés d'information par l'exploitation des annonces du Balo[76] et du Bodacc[77]. Ésane a donc besoin de **récupérer** de l'information **de Citrus** pour vérifier si des restructurations déjà connues ne touchent pas des entreprises qu'on est en train de traiter.

Comme on vient de le mentionner au § précédent, l'Esa permet aussi de collecter systématiquement de l'information sur les restructurations. Ésane doit alors **envoyer** cette information **vers Citrus** qui centralise toutes ces informations pour l'ensemble des statisticiens d'entreprises. Il y a donc des échanges réguliers entre Citrus et Ésane au terme desquels les deux systèmes partagent les informations suivantes :

- la liste des entreprises de l'enveloppe en N ;
- la liste des entreprises de l'enveloppe en N-1 ;
- la liste des couples d'entreprises « bénéficiaires - cédantes » avec le montant des apports ;
- les caractéristiques de l'enveloppe (Ape, coefficient d'agrégation, effectifs, dates de la modification de structure). Le coefficient d'agrégation est recalculé dans Ésane à partir de la notion de flux intra enveloppe qui est décrite ci-dessous au § *3.2.2*. En effet, lorsque les gestionnaires Esa récupèrent de l'information sur les flux intra enveloppes, cette information « remplace » la notion de Kagre pour le chiffre d'affaires. En outre comme cette information est collectée auprès de l'entreprise elle est donc « vraie » et donc de meilleure qualité que le Kagre qui lui est estimé. Ainsi, afin que cette information de meilleure qualité soit profitable à l'ensemble des variables du questionnaire, le Kagre est recalculé à partir du flux (cf. § *3.2.2.2*).

[76] Le Bulletin des Annonces Légales Obligatoires (ou Balo) est édité par la Direction de l'information légale et administrative (la Dila). Il recense les informations légales des sociétés faisant publiquement appel à l'épargne. On y trouve notamment la publicité sur les opérations financières qui correspondent, entre autres, aux restructurations.

[77] Le Bulletin Officiel des Annonces Civiles et Commerciales (ou Bodacc), également édité par la Dila, publie les actes enregistrés au Registre du Commerce et des Sociétés (RCS) parmi lesquels les ventes et les cessions nous intéressent particulièrement. Pour des raisons de moyens, on ne peut pas exploiter la totalité de ces annonces qui ne sont par ailleurs pas exhaustives.

Toutes les enveloppes de Citrus ne sont pas prises en compte dans Ésane. En effet, **seules** les enveloppes contenant **au moins un couple d'entreprises du champ d'Ésane sont retenues**. Ainsi, par exemple, si une entreprise financière absorbe une entreprise du champ d'Ésane, cette restructuration ne donne pas lieu à la création d'une enveloppe dans Ésane. L'entreprise absorbée est considérée dans les calculs des agrégats comme étant cessée.

Cette décision a été prise dans la mesure où l'objectif majeur des enveloppes dans Resane est de pouvoir raisonner à champ comparable. Il devient inutile alors de créer des enveloppes contenant des entreprises pour lesquelles on ne dispose d'aucune information (ni liasse ni questionnaire).

Cette décision est plus discutable si l'on considère l'autre objectif à savoir une estimation fiable de l'investissement hors apport. Mais même dans ce cas, l'absence de liasse des unités hors champ Ésane est rédhibitoire, c'est un problème plus général d'exclusion des entreprises financières du champ d'Ésane.

3.2.1 La création des unités dans l'Esa

Contrairement aux liasses fiscales pour lesquelles on dispose peu ou prou de l'exhaustivité des informations, ceci est bien entendu faux pour l'Esa où il est possible d'être dans une des situations suivantes :

- une entreprise prenant part à l'enveloppe ne faisait pas partie de l'échantillon Esa en N-1 ;
- une entreprise prenant part à l'enveloppe ne fait pas partie de l'échantillon Esa en N ;
- une entreprise prenant part à l'enveloppe ne fait partie de l'échantillon Esa ni en N, ni en N-1.

Afin d'éviter le plus possible cette situation, le plan de sondage prend en compte les restructurations connues. En effet, au moment du tirage de l'Esa pour l'exercice N, Citrus envoie au DMS[78] la liste des unités qu'il sait être en restructuration pour l'année N au moment du tirage de l'échantillon[79]. Ces dernières sont alors isolées dans des strates de tirage spécifiques dont le taux de sondage est de 100%.
Le nombre d'unités en restructuration et absentes de l'échantillon en N est donc réduit, **mais il en existe toutefois.**
Ainsi, si on apprend en cours de campagne qu'une unité est en restructuration au cours de l'exercice N et qu'elle ne figure pas dans l'échantillon, les opérations suivantes ont lieu :

- forcement de l'unité dans l'échantillon avec un poids de 1 ;
- envoi éventuel d'un questionnaire à l'entreprise. Ceci est **à la discrétion du service Ésane de Nantes** et dépend de la date à laquelle cette introduction a lieu et de la taille de l'unité introduite. Ainsi, plus l'unité est petite et la collecte avancée et plus la probabilité d'envoi d'un questionnaire diminue ;
- extrapolation éventuelle du questionnaire. Cette extrapolation est « forcée » par N-1 à partir du chiffre d'affaires de la liasse fiscale de l'exercice N (Cf. § 5.5.1) ;
- recalcul des poids des entreprises échantillonnées de la strate de tirage de l'unité ainsi « ajoutée » selon la formule : $poids_i = \frac{\left(\sum_{i \in strate} poids_i\right) - 1}{\left(\left(\sum_{i \in strate} 1_{i \in ESA}\right)\right)}$ de façon que le poids total de la strate reste, lui, inchangé.

Lorsque l'entreprise était absente de l'échantillon en N-1, elle y est rajoutée de façon fictive et un questionnaire est extrapolé selon la méthode de l'extrapolation par N-1 (cf. § 5.5.1 ci-dessous).

Ainsi, à l'issue de ces différentes étapes, **chaque unité** appartenant à une enveloppe **dispose** d'un questionnaire Esa pour la (ou les) années pour la (ou les)quelle(s) elle faisait partie du contour de l'enveloppe de restructuration.

[78] Département des méthodes statistiques : unité de l'Insee en charge du tirage de l'échantillon de l'ESA et de l'EAP.
[79] C'est-à-dire au mois d'octobre de l'année N.

3.2.2 L'estimation du flux intra enveloppe et ses liens avec le coefficient d'agrégation

3.2.2.1 La définition du flux intra enveloppe

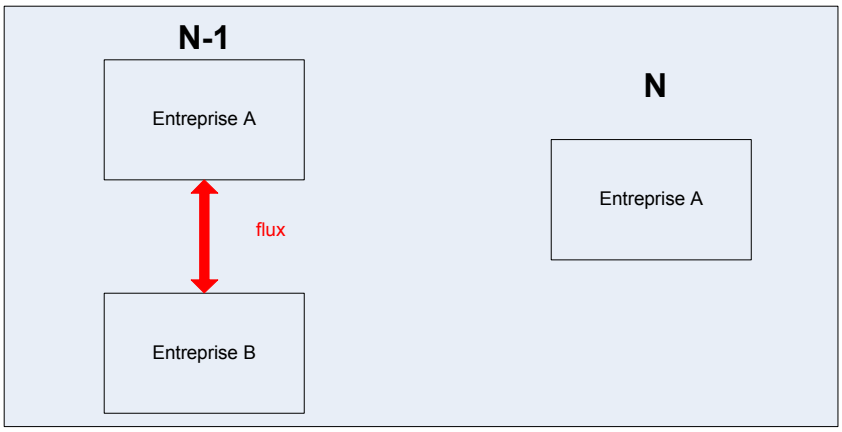
L'une des « nouveautés » d'Ésane est **l'estimation** et la validation éventuelle par les gestionnaires d'un ***flux intra enveloppe****. Le changement de structure juridique peut en effet se traduire par la disparition ou la suppression de chiffre d'affaires ; certaines activités pouvant passer du statut de marchande à celui d'intégrée dans le cas d'une fusion par exemple.
Ces flux intra enveloppe perturbent le calcul en évolution. Car si l'on veut raisonner à structure constante entre N et N-1, ce qui est l'objectif de l'enveloppe, il faut réussir à le neutraliser. Pour ce faire, il faut donc **réussir à en obtenir une estimation**, vu que l'on ne dispose pas dans les bases Ésane des variables permettant de le calculer. Le seul moyen d'obtenir la valeur réelle du flux est le **rappel** téléphonique (ou via e-mail) de la (ou des) entreprise(s) concernée(s).
Ces flux intra enveloppe ne concernent que les variables dites non additives. C'est-à-dire les variables qui sont dépendantes de la modification de structure.
Dans le cas d'Ésane ces flux concernent avant tout **le chiffre d'affaires et sa ventilation** qui sont en outre les moins complexes à obtenir.
Le flux sur le chiffre d'affaires, s'il est obtenu après rappel de l'entreprise, sera utilisé pour estimer le coefficient d'agrégation Ésane qui sera appliqué à l'ensemble des caractéristiques non additives de l'Esa et de la liasse fiscale.

Par convention, **le flux sera appliqué au chiffre d'affaires N-1 de l'enveloppe** (à l'instar du coefficient d'agrégation). Il pourra donc être positif dans le cas où il aurait disparu à cause de la restructuration (cas d'une fusion par exemple) ou négatif dans le cas où il serait apparu à l'issue de la restructuration (cas d'une scission par exemple).

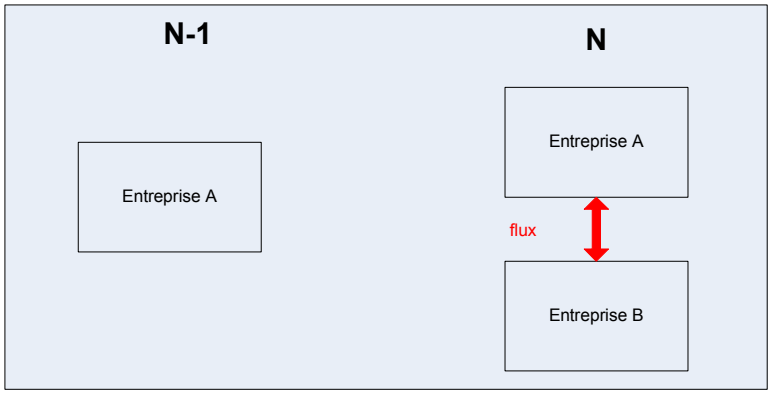
Ainsi on aura, pour la valeur corrigée du CA de l'enveloppe :

$$\boxed{CAdef_{N-1}^{env} = CA_{N-1}^{env} - flux}$$

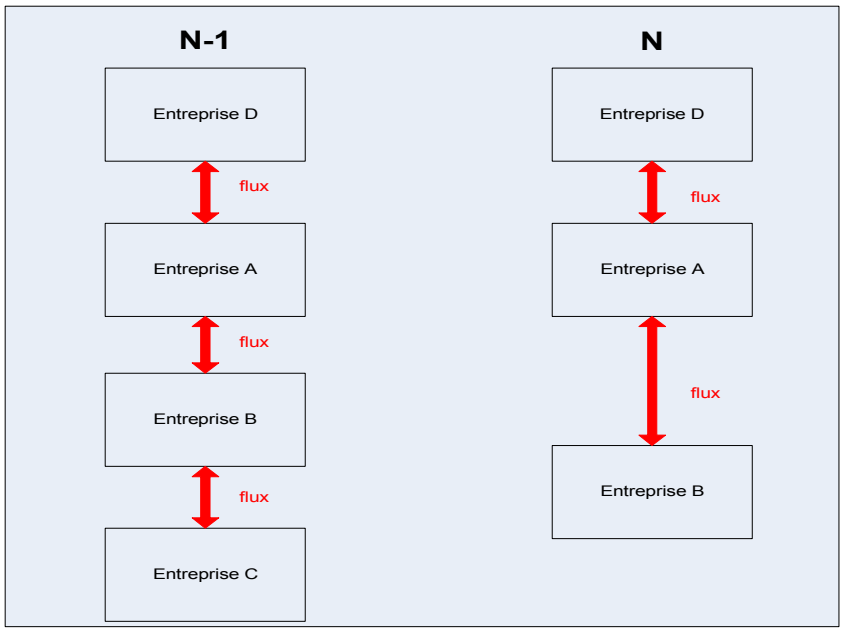
Explicitation du flux dans le cas d'une absorption : Flux = $CA_{(N-1)}$ [A->B] + $CA_{(N-1)}$ [B->A]
(les échanges peuvent avoir lieu dans les deux sens entre A et B d'où la double flèche dans le graphique ci-dessous)



Explicitation du flux <u>dans le cas d'une scission</u> : Flux= - (CA_N[A->B] + CA_N[B->A])



Exemple de flux dans un cas complexe



Dans cet exemple on suppose qu'il y a des transferts d'actifs entre les 4 sociétés (A, B, C et D) et des échanges à la fois en N et en N-1 entre ces différentes sociétés.
Dans ce cas le calcul du flux est plus complexe et peut être estimé de la façon suivante.
Flux = [CA (A->B) (N-1) - CA (A->B) (N)] + [CA(B->C) (N-1)] + [CA (A->D) (N-1) - CA (A->D) (N)]

3.2.2.2 Le recalcul d'un coefficient d'agrégation

Le **flux intra enveloppe** tel que défini ci-dessus n'est utilisé pour le moment **que dans le système de production Ésane**. Mais, il a le **même rôle** que celui dévolu au **coefficient d'agrégation** que Citrus calcule.
C'est pourquoi, **à l'initialisation**, on va faire une première estimation du flux à partir du coefficient d'agrégation que nous fournit Citrus.
En utilisant les deux formules suivantes on obtient :

$$\left.\begin{array}{l} CAdef_{N-1}^{env} = CA_{N-1}^{env} - flux \\ \text{et} \\ CAdef_{N-1}^{env} = Kagre * CA_{N-1}^{env} \end{array}\right\} flux = CA_{N-1}^{env} * (1 - Kagre).$$

Cette première estimation est **ensuite corrigée** dans les cas suivants :

- lors d'une fusion ou absorption, si le flux est supérieur au plus petit chiffre d'affaires (N-1) des deux entreprises concernées par la restructuration alors le flux est égal à ce chiffre d'affaires ;
- lors d'une scission, si la valeur absolue du flux est supérieure au plus petit chiffre d'affaires N des deux entreprises concernées par la restructuration alors le flux est égal à l'opposé de ce chiffre d'affaires.

Par la suite, **à chaque mise à jour du flux** par un gestionnaire (à la suite d'un contact avec une des entreprises en restructuration), le coefficient d'agrégation est **recalculé** à partir de la formule suivante :

$$\left.\begin{array}{l} CAdef_{N-1}^{env} = CA_{N-1}^{env} - flux \\ \text{et} \\ CAdef_{N-1}^{env} = Kagre * CA_{N-1}^{env} \end{array}\right\} \boxed{Kagre = 1 - \frac{flux}{CA_{N-1}^{env}}}.$$

Cette estimation du coefficient d'agrégation permet d'estimer toutes les caractéristiques N-1 non additives de l'enveloppe.

Reprenons l'exemple présenté dans le § *3.1.2* :

Le flux réel en N-1 entre les entreprises 2 et 3 qu'un gestionnaire Ésane pourra apprendre par contact avec l'entreprise 5 résultant de leur fusion était de 40k€ (avions nous supposé). Il en résulte un nouveau Kagre qui vaut d'après la formule ci-dessus :
Kagre = 1 – (40/325) = 0.8769.
On vérifie bien que le chiffre d'affaires (N-1) corrigé de l'enveloppe, qui vaut donc 0.8769*325 = 285k€ **est le vrai chiffre d'affaires consolidé des entreprises 2 et 3**. La connaissance exacte du flux a bien permis de reconstituer le vrai chiffre d'affaires (N-1) de l'enveloppe, qui est « comparable » au chiffre d'affaires N de cette même enveloppe.

Les premiers macro-contrôles seront lancés avec ces estimations :

- Si l'évolution du chiffre d'affaires de l'enveloppe n'est pas à contrôler (car pas suffisamment contributif à l'évolution de l'agrégat selon les macro-contrôles), c'est cette estimation du flux qui sera conservée pour la diffusion des résultats en évolution.
- En revanche si le chiffre d'affaires de l'enveloppe est à contrôler (l'enveloppe est jugée fortement contributrice pour le macro-contrôle temporel du chiffre d'affaires), le gestionnaire devra rappeler au moins une entreprise de l'enveloppe pour obtenir une évaluation de ce flux (à moins que l'évolution aberrante soit due à une réponse aberrante en N d'une des entreprises de l'enveloppe).

En théorie, un flux existe pour toute variable non-additive. Toutefois, pour des raisons de simplification, dans le cadre d'Ésane, les flux intra sur lesquels il est demandé aux gestionnaires d'intervenir concernent **uniquement** :

- le chiffre d'affaires ;
- la ventilation du chiffre d'affaires en activités.

Les autres variables non additives (sous-traitance etc.) sont estimées à partir du coefficient d'agrégation (éventuellement mis à jour à partir de la correction du flux du chiffre d'affaires obtenu par les gestionnaires) qui est appliqué à la somme des valeurs N-1 des entreprises de l'enveloppe.

3.2.3 Le calcul des variables de la liasse et du questionnaire d'une enveloppe (à partir de celles des unités légales)

À partir du contour des enveloppes en unités légales d'une part et du caractère additif ou non d'une caractéristique d'autre part, la valeur d'une caractéristique X d'une enveloppe est obtenue de la façon suivante :

$$X_N^{env} = \sum_i X_N^i * 1\mathrm{I}_{i \in env}$$

$$X_{N-1}^{env} = \sum_i X_{N-1}^i * 1\mathrm{I}_{i \in env}$$ si X est une caractéristique additive

$$X_{N-1}^{env} = Kagre * (\sum_i X_{N-1}^i * 1\mathrm{I}_{i \in env})$$ si X est une caractéristique non-additive

et $CAdef_{N-1}^{env} = CA_{N-1}^{env} - flux$ pour les variables de chiffre d'affaires associées à un flux. Cette formule est équivalente à celle-ci-dessus puisque le Kagre est recalculé à partir des flux de chiffre d'affaires

C'est sur ces valeurs que les macro et micro contrôles temporels vont être mis en œuvre.

Pour le calcul des variables des **liasses,** il existe quelques **subtilités supplémentaires,** non détaillées ici et assez complexes, qui ont été traitées en examinant tous les cas possibles de consolidation. En effet, certaines unités en restructuration peuvent avoir des exercices décalés et/ou des exercices différents de 12 mois[80]. Enfin, il peut aussi y avoir une prise en compte « rétroactive » de la restructuration dans les comptes sociaux. Ainsi, par exemple, si une entreprise A absorbe une entreprise B en mars, elle peut ne faire qu'une déclaration fiscale pour toute l'année en incluant l'activité de l'entreprise B sur les 3 premiers mois de l'année dans son activité totale de l'année. Dans ce cas de figure il ne faudra donc pas « extrapoler » une liasse pour l'entreprise B pour les 3 premiers mois de l'année sous peine de faire un double compte. La question de la rétroactivité de la restructuration est posée dans le questionnaire Ésane. Dans les faits, il arrive que l'on passe à côté. Mais si le gestionnaire s'en aperçoit au moment de ses contrôles, il a la possibilité de considérer l'entreprise B comme étant cessée statistiquement à partir du 1er janvier.

3.3 La prise en compte des enveloppes dès la phase de contrôle

3.3.1 Les enveloppes : une nouvelle unité de contrôle

La prise en compte des enveloppes dès la phase de contrôle des données est ***chose nouvelle à l'Insee****. Elle vise principalement trois choses :

1. permettre d'avoir des agrégats de contrôles en évolution entre N et N-1 plus proches des agrégats en évolution diffusés et ainsi d'améliorer la pertinence des contrôles temporels ;
2. demander aux gestionnaires d'obtenir des informations sur les flux intra-enveloppes à gommer (ou ajouter) pour les enveloppes les plus influentes afin d'obtenir la meilleure estimation de l'évolution possible ;
3. gommer des évolutions sectorielles « atypiques » au niveau des unités légales en prenant en compte les enveloppes de restructuration et ainsi améliorer le processus d'identification des unités à contrôler. Ne pas utiliser l'enveloppe conduirait à contrôler manuellement à tort des unités légales ayant des évolutions surprenantes uniquement dues à la restructuration. Ainsi, une entreprise absorbante peut voir son chiffre d'affaires fortement augmenter d'une année sur l'autre alors que si on compare ce chiffre d'affaires à la somme en N-1 des chiffres d'affaires

[80] Ces cas ont été réglés en répertoriant tous les cas de restructurations possibles selon qu'il existe 0, 1 ou 2 entreprises pérennes dans la restructuration et selon qu'il s'agit d'une scission, d'une fusion, d'une absorption ou d'une cession partielle. Le principe général est de donner une date d'effet à la restructuration qui soit le 31 décembre de l'année N-1 ou de l'année N ; on part des liasses réelles et on les contracte ou on les dilate selon des principes qui seront exposés au *chapitre 6* de façon à se ramener à des données sur 12 mois coïncidant avec une année civile permettant à nouveau d'utiliser les formules ci-dessus.

des entreprises absorbées et de son propre chiffre d'affaires, l'évolution paraitra beaucoup moins atypique.

Le principe des contrôles est donc le suivant pour les enveloppes et les unités légales qui les composent[81] :

- les contrôles (macro et micro) **temporels** sont lancés :
  - sur les unités légales hors enveloppes pour lesquelles on dispose de données pour les exercices N et N-1[82] ;
  - sur les enveloppes ;
- les **autres contrôles** sont lancés sur les **unités légales uniquement**.

L'enveloppe devient donc une unité de contrôle à part entière. En outre, comme une restructuration forme « un tout » qui nécessite un examen d'ensemble pour comprendre son impact économique, il a été décidé :

- de confier l'examen des enveloppes à des **gestionnaires spécialisés** ;
- de concentrer sur le dossier enveloppe tous les problèmes qui concernent soit directement l'enveloppe (problème d'évolution temporelle) soit l'une des entreprises du contour de l'enveloppe.

Enfin, il avait également été décidé que dans le cas d'un problème temporel sur les enveloppes, les corrections suivantes seraient possibles :

- s 'il s'agit d'une erreur au niveau de la valeur N de la caractéristique de l'enveloppe, la correction de cette dernière doit être faite au niveau de la (ou des) unité(s) légale(s) en erreur
- s'il s'agit d'un problème d'estimation du flux, ce dernier doit être corrigé au niveau de l'enveloppe ;
- s'il s'agit d'une erreur au niveau de la valeur N-1 de la caractéristique de l'enveloppe, le gestionnaire doit demander une correction au niveau de la (ou des) unité(s) légale(s) en erreur dans la base de production N-1[83].

3.3.2 Le travail des gestionnaires sur les enveloppes

Les enveloppes à traiter figurent dans la liste des dossiers à traiter ou dans l'en-cours du gestionnaire comme un dossier classique d'unité légale. Le gestionnaire peut néanmoins facilement repérer les enveloppes car leur identifiant commence par un « E ».

Après sélection de l'enveloppe, le gestionnaire sait s'il doit :

- contrôler des unités légales de l'enveloppe uniquement
- contrôler l'enveloppe uniquement (pour un problème d'évolution temporelle atypique) ;
- contrôler à la fois l'enveloppe et une ou plusieurs unités légales de l'enveloppe.

Il doit alors commencer par traiter les unités légales de l'enveloppe. Le traitement préliminaire des unités légales de l'enveloppe peut en effet conduire à la résolution du problème au niveau de l'enveloppe[84]. Les traitements des gestionnaires spécifiquement liées aux restructurations sont les suivants.

3.3.2.1 Contrôle des contours avant les macro-contrôles

Avant la phase de contrôle par macro contrôles, il existe des contrôles dans l'Esa et d'autres au niveau de la liasse fiscale qui permettent **d'identifier des entreprises appartenant à des enveloppes**. L'objectif de ces contrôles est de « découvrir » ou de compléter des restructurations et d'enrichir Citrus. À l'issue de ces phases de contrôle, le gestionnaire doit valider la composition de l'enveloppe en unités légales.

[81] Les unités légales hors enveloppes subissent bien sûr tous les types de contrôles (contemporains et temporels)

[82] Ainsi une unité légale appartenant à une enveloppe, même si elle dispose de données N et N-1, ne subira aucun contrôle temporel.

[83] Il y a ici une subtilité technique de nature informatique : l'année N-1 n'est pas accessible en écriture de façon standard pour les gestionnaires dans la campagne N car les résultats N-1 sont censés être définitifs. Elle ne peut du coup être faite que de façon spécifique par une équipe spécialisée. Cette correction n'impactera les bases de diffusion que si la correction est vraiment très importante ; sinon le gestionnaire se contentera de valider la valeur N en ayant compris pourquoi il a une évolution N/N-1 anormale.

[84] Une correction d'un chiffre d'affaires d'une unité légale de l'enveloppe conduit en effet au recalcul du chiffre d'affaires N de l'enveloppe. Cette correction peut corriger une évolution temporelle erronée. Le gestionnaire doit toutefois attendre le lendemain pour connaître l'impact de sa correction au niveau de l'enveloppe car le calcul des caractéristiques des enveloppes se fait la nuit.

1 Dans l'Esa, ces contrôles visent à **assurer la cohérence** entre les différentes informations sur les restructurations présentes sur le questionnaire Esa. Ce dernier comprend en effet plusieurs sous cadres sur le sujet :

- un cadre de description générale concernant la restructuration qui contient des informations telles que le type de restructuration et sa date ;
- un cadre relatif à la description des partenaires de la restructuration (n° Siren du partenaire de la restructuration, montant des apports échangés etc.) ;
- un cadre relatif aux montants totaux des investissements et cessions par voie d'apport.

2 Dans la liasse fiscale, ces contrôles de contour visent à assurer un accrochage temporel des liasses pour une unité légale donnée. Pour tout décrochage de plus de 1500K€ des immobilisations et amortissements[85], l'unité légale doit être contrôlée par le gestionnaire. Ce dernier doit alors **vérifier que cette unité ne fait pas partie d'une restructuration**. Pour ce travail il consulte les sources (privées) à sa disposition et rappelle l'entreprise en dernier recours.

3.3.2.2 Contrôle des enveloppes par le biais des macro-contrôles

On se place dans le cas où l'enveloppe doit être contrôlée parce qu'elle a été repérée par au moins un macro contrôle temporel comme ayant une évolution atypique et ayant un poids important dans un des agrégats contrôlé.
Dans ces conditions, le gestionnaire doit commencer par vérifier la complétude de l'enveloppe. En règle générale, le gestionnaire peut avoir des doutes à ce niveau si l'enveloppe a été repérée comme atypique pour plusieurs macro-contrôles. Ceci peut en effet **cacher l'oubli d'une unité légale** dans le contour de l'enveloppe en N et/ou N-1. Pour valider le contour, le gestionnaire **peut consulter des sources externes** commerciales dans lesquelles figurent des faits marquants sur les unités légales. Si ces sources ne permettent pas de conclure, l'appel d'une des entreprises présentes en N dans l'enveloppe reste l'unique solution pour régler le problème.

Une fois le contour validé, s'il reste des incohérences, trois pistes sont encore possibles :

- des **données N erronées sur une des unités légales** du contour. Dans ce cas le gestionnaire les corrige directement depuis son poste de travail au niveau de l'unité légale, après un rappel téléphonique éventuel de l'entreprise ;
- des **données N-1 erronées sur une des unités légales** du contour. Dans ce cas, comme on l'a vu, le gestionnaire doit demander une correction des données N-1 ;
- une évolution erronée d'une variable non-additive associée au flux à cause **d'une estimation erronée du flux** à partir du coefficient d'agrégation. Dans ce cas, le gestionnaire doit rappeler l'entreprise qu'il juge « la plus importante » de l'enveloppe afin d'obtenir une estimation du flux intra à partir des montants de chiffres d'affaires qui sont apparus ou ont disparu au sein de l'enveloppe entre N-1 et N. Cette correction du flux se traduira également par une mise à jour du coefficient d'agrégation. Ceci se traduira donc par une correction de toutes les valeurs N-1 des caractéristiques non additives de l'enveloppe.

3.3.3 L'investissement hors apport, une variable difficile à contrôler

Lorsqu'on s'intéresse au comportement d'investissement des entreprises dans l'analyse économique, la variable la plus pertinente est le montant des investissements en dehors des apports obtenus par fusion ou absorption partielle. Il faut donc retirer le montant des investissements obtenus par apport du montant total des investissements ; et pour cela il faut bien les connaître. C'est l'un des objectifs visés par le contrôle des enveloppes, comme on l'a vu au point ii) du § *3.1.3*. Or, le contrôle des investissements n'est pas chose aisée dans la mesure où il s'agit d'une variable sans cohérence temporelle (une entreprise peut investir fortement une année donnée et pas les années suivantes ou précédentes, sans que cette discontinuité révèle un problème sur les données).

[85] Immobilisations et amortissements de fin de l'exercice N-1 différents des immobilisations et amortissements de début de l'exercice N

Pour cette raison, les macro-contrôles temporels sur les investissements des enveloppes ne permettent donc pas bien d'identifier les cas d'unités légales ayant des apports de mauvaise qualité et il a été nécessaire de mettre en place une alternative aux macro-contrôles temporels.

L'idée est de demander aux gestionnaires de **contrôler les plus gros investissements en valeur absolue**.

Les unités correspondantes sont sélectionnées **par l'intermédiaire de listes obtenues à partir de l'un des cinq critères suivants :**

1. Pour les unités légales appartenant à une enveloppe :

    a. Chez la bénéficiaire, si le montant des apports est supérieur ou égal au montant brut des augmentations des immobilisations et que le montant des apports est supérieur à *seuil_1* (seuil_1 est pour l'instant égal à 1 500 K€) ;
    b. Chez la cédante, si le montant des apports est supérieur ou égal au montant brut des cessions des immobilisations et que le montant des apports est supérieur à *seuil_2* (seuil_2 est pour l'instant égal à 1 500 K€).
    c. Chez la bénéficiaire et chez la cédante, si montant des apports > *seuil_3* : il s'agit de vérifier les dossiers avec un montant des apports très important (seuil_3 est pour l'instant égal à 100 millions d'euros)

2. Pour toutes les entreprises (qu'elles appartiennent à une enveloppe ou pas)

    a. montant d'investissement corporel hors apport supérieur à *seuil_4* (*seuil_4* est pour l'instant égal à 50 millions d'euros)
    b. montant d'investissement corporel hors apport inférieur à *seuil_5* (avec *seuil_5* négatif, - 20 millions d'euros pour l'instant, car il s'agit de contrôler les investissements corporels fortement négatifs).

Au final, les unités repérées sont des unités légales et non des enveloppes. Il n'y a aucun traitement au niveau enveloppe, **les corrections sont faites au niveau des unités légales,** même si les problèmes rencontrés sont liés à une restructuration. À noter qu'un dossier peut être à traiter pour un seul ou plusieurs des cinq contrôles cités. Après ces corrections, les caractéristiques des enveloppes se retrouvent, de fait, cohérentes.

4. Le champ du système d'information Ésane

Le champ du système a déjà été présenté succinctement au § *1.2.1.1*. On va en donner ici en plus les principes détaillés en abordant successivement :

- Les critères utilisés pour la définition du champ (§ *4.1*).
- Le champ géographique (§ *4.2*).
- Les sous champs (distingués dans les traitements) (§ *4.3*).
- La définition du champ « en extension » (§ *4.4*).
- Les champs distingués dans la diffusion (§ *4.5*).
- Les modifications de champ entre 2008 et 2010 (§ *4.6*).
- Les modifications de champ et de mode de diffusion à compter de la campagne 2011 (§ *4.7*).
- Les modifications de la définition de l'entreprise à compter de la campagne 2013 (§ *4.8*).

4.1 Les critères utilisés pour la définition du champ

Le champ d'Ésane est constitué « globalement[86] » **des sociétés et entreprises individuelles, marchandes et productives ou participant au système productif[87], non agricoles[88], et non financières** (exceptés les holdings (classe 6420Z de la NAF rev. 2) et les auxiliaires financiers et d'assurances – division 66 - qui font partie du champ)**. *Ce champ est défini a priori pour assurer les comparaisons N/N-1*.***

- *Les statistiques afférentes aux exploitations agricoles* sont élaborées et publiées directement par le service statistique du ministère de l'agriculture.
- *La partie « financière » des statistiques structurelles d'entreprises* est élaborée par la Banque de France, et simplement « ajoutée » quand Eurostat publie les résultats français.

Comme on l'a vu au *chapitre 1*, **le champ du système Ésane** est en pratique **la réunion de tous les champs** correspondants à des agrégats calculés par le système pour un **utilisateur ou un public particulier**. C'est pourquoi il n'est pas particulièrement signifiant en lui-même. Il est parfaitement concevable que tel agrégat livré aux comptables nationaux ou à un ministère soit légèrement différent du même agrégat calculé pour Eurostat, car ils diffèrent sur un certain champ. Mais pour pouvoir faire des calculs, il faut bien que les données élémentaires aient été intégrées au système. Comme on va le voir ci-dessous, un sous champ particulier a même été créé en 2010 pour pouvoir calculer des agrégats pour des publics spécifiques. La différenciation du champ Ésane complet **en ses différents champs de publication** sera débattue au § *4.5*.

Dans le détail, le champ Ésane, c'est-à-dire le fait pour une unité d'être marchande et productive ou participant au système productif, est **défini de façon assez complexe à partir de 5 critères** qui s'appliquent **aux unités légales** du répertoire Sirus (Cf. *11.7*) lors de la constitution du référentiel Ocsane.

Les cinq critères, définissant le **champ a priori** pour une unité légale donnée, sont les suivants :

- **Le code d'activité principale exercée[89] (APE)** : celui-ci peut prendre a priori 732 valeurs définies par la nomenclature officielle française (la NAF) ; le code APE d'une unité légale est défini lors de sa création ; il peut ensuite être mis à jour :
  - soit **à l'initiative de l'unité légale** elle-même (il existe une formalité administrative simple disponible sur le site de l'Insee pour cela) ;

[86] Au sens où cette phrase ne définit pas de façon suffisamment précise le champ d'Ésane comme les paragraphes suivants le montreront. Il y a des exceptions qui font justement l'objet du chapitre.
[87] Tous ces termes seront expliqués précisément dans la suite du paragraphe.
[88] Les exploitations agricoles sont dans Ocsane mais ne font pas partie du champ Ésane sauf les exploitations forestières. Les liasses fiscales des BAS et BAN (bénéfices agricoles simplifiés et normaux) transmises à l'Insee par la DGFiP, avec les autres régimes fiscaux, sont traitées automatiquement donc sans intervention des gestionnaires Ésane. Aucun traitement n'est mis en œuvre pour extrapoler les données en cas d'absence de liasse fiscale.
[89] Voir l'*Annexe 1* pour une description de l'algorithme utilisé pour la détermination de ce code lorsque l'unité légale considérée exerce plusieurs activités.

- soit à la **suite des enquêtes d'amélioration du répertoire Sirene** (qui sont des enquêtes administratives sur l'initiative du service Sirene de l'Insee, menées dans le cadre d'opérations de maintien de la qualité du répertoire) ;
- soit à la **suite des réponses de l'unité légale aux enquêtes Esa ou EAP**.
- ***Seuls 666 secteurs sont représentés dans Ésane** et certains ne le sont que partiellement pour certaines informations qui ne sont collectées que via les enquêtes ESA ou EAP ; des indications plus précises sont données au § 4.4.1.*

➢ **La catégorie juridique (CJ)** : celle-ci est définie parmi une liste de plus de 270 valeurs possibles[90], à la création de l'unité légale ou à la suite d'un changement de catégorie que l'unité signale dans le cadre du système Sirene. La loi crée de nouvelles catégories juridiques presque chaque année. *Les catégories juridiques exclues sont explicitées au § 4.4.2.*

➢ **Le nombre de salariés (NB-SAL)** : ce critère intervient uniquement pour distinguer des parties de secteurs sur lesquelles on va collecter des informations complémentaires par voie d'enquêtes (ESA ou EAP) donc pour distinguer les sous champs de production (Cf. § *1.2.1.1* et § *4.3*). L'idée est que, au-dessus d'une certaine taille dans un secteur donné, il est justifié d'avoir recours aux enquêtes pour avoir un certain détail d'informations ; mais en dessous de certains seuils, l'information fiscale suffit et **il coûterait trop cher de vouloir enquêter des unités trop petites** ; mais du coup, on ne dispose pas des informations collectées exclusivement dans les enquêtes pour ces petites entreprises. Cela vaut exclusivement pour certaines catégories juridiques (personnes morales de droit étranger, associations). La variable d'emploi utilisée est l'emploi **mesuré en équivalent temps plein** (Cf. *chapitre 7*). *Les trois seuils principaux pour différencier des sous populations d'unités sont les seuils de 0[91], 5 et 30 salariés*.

➢ **Le caractère marchand ou non marchand (MA)** : une unité marchande est une unité qui retire au moins 50% de ses ressources de ventes de biens et services sur un marché.
Le code marchand est défini à partir du « secteur institutionnel » de la comptabilité nationale[92] auquel appartient toute unité légale, La table de correspondance entre secteur institutionnel et code marchand est la suivante :

Secteur institutionnel	Intitulé en clair	Code marchand
S11	Sociétés Non Financières	MA
S12	Institutions Financières et Monétaires Interprétation : toutes les unités des croisements S12 sont des auxiliaires financiers – donc à inclure au champ Ésane – lorsqu'elles ne sont pas sur les listes officielles (unités agréées) des établissements de crédit (S12A) et des assurances (S125)	MA
S13	Administrations Publiques	PU
S14AA	Entrepreneurs Individuels Non Financiers	MA
S14AF	Entrepreneurs Individuels Financiers	MA
S14B	Ménages	Hors champ non codé
S15	Institutions Sans But Lucratif au Service des Ménages	PR
S2	Reste du monde	Hors champ non codé
HC		Hors champ non codé

En d'autres termes, sont considérés comme **marchandes** les unités classées en S11, S12, et S14A.

[90] La nomenclature des catégories juridiques est publiée sur le site de l'Insee : *https://www.insee.fr/fr/information/2028129*
[91] Cf. en particulier le § *5.2.1* qui définit le champ de l'EAP lequel exclut les unités légales sans salarié (jusqu'en 2015).
[92] La *comptabilité nationale* française se conforme au cadre comptable du système européen de comptabilité (SEC) qui distingue, entre autres, les secteurs institutionnels figurant dans le tableau ci-dessus.

Jusqu'à la campagne 2015, l'appartenance de chaque unité légale à un secteur institutionnel était elle-même fonction d'un croisement CJ[93] x APE[94] que les comptables nationaux déterminent à partir de leur expertise. A l'issue de cette expertise, on sait donc quelles sont les unités marchandes, non marchandes privées (PR) ou non marchandes publiques (PU). Outre le fait qu'**elles sont susceptibles de varier chaque année** (à cause de l'appréciation du seuil de 50%), elles comprennent un certain nombre de cas particuliers et d'exceptions de sorte que ce caractère marchand ou non marchand finit par être une caractéristique propre des unités légales que l'on gère au sein du répertoire Sirus[95] (Cf. *11.7*) et qui est susceptible d'évoluer chaque année à la marge. C'est ainsi qu'à compter de la campagne 2015, le **S13** (liste des administrations publiques) est défini directement chaque année en extension, à partir de la liste des numéros Siren des unités correspondantes (fournie par le DCN[96] en accord avec la direction générale des finances publiques). A compter de la campagne 2017, ce sera également le cas du S12 qui sera défini par la Banque de France et transmis par le biais du DCN.
On ne retient en tout cas dans le champ d'Ésane que les unités marchandes.

- **Le caractère « productif » ou non d'une unité légale** **(PRODUCT)** : même en ne retenant que les unités marchandes, l'idée est que les répertoires Sirene et Sirus contiennent encore **trop d'unités** qui n'intéressent pas les statisticiens pour mesurer l'activité économique (unités qui sont marchandes mais ne participent pas au système productif) et qu'on souhaite donc écarter (il s'agit des SCI « familiales », des associations sans salarié et des sociétés étrangères non inscrites au RCS n'ayant pas de liasse fiscale). Deux notions complémentaires nous intéressent donc pour écarter ces cas :
  - le **fait** de participer ou non au système productif pour produire des biens ou services sur un marché; o*n ne garde donc que les unités productives ;*
  - ou le fait de posséder soi-même ses propres moyens de production (tels que locaux, machines, personnel) ou d'utiliser au contraire des moyens de production, mis à disposition par d'autres unités (qui ne les exploitent pas mais contribuent ainsi indirectement à la production : exemples de sociétés que l'on souhaite conserver: loueurs de fonds ; SCI appartenant à des groupes et/ou déposant des liasses fiscales et/ou employeuses ; holdings financiers). On va aussi garder dans le champ d'Ésane *les unités non productives elles-mêmes mais participant au système productif indirectement.*

On note au passage que cette définition sur la base de ces cinq critères permet de sélectionner par exemple **un certain nombre « d'associations** » (de l'ordre de 12 à 13000 chaque année). Les associations sont caractérisées par une catégorie juridique de la forme 92XX qui n'est pas exclue du champ. Lorsqu'elles appartiennent à des secteurs retenus par le système, qu'elles sont à la fois marchandes et productives (par exemple elles sont assujetties à la TVA), il est tout à fait logique qu'elles fassent partie du champ d'Ésane.

4.2 Le champ géographique

Le champ géographique d'Ésane est la **France entière, y compris les Dom**.

Les entreprises de **Saint Barthélemy** et **Saint Martin**[97] sont **maintenues** dans le champ Ésane, ce qui représentait en 2010 respectivement 3 782 et 6 914 entreprises.
Les entreprises localisées **à Mayotte** (6 199 entreprises en 2010) sont intégrées dans Ocsane mais **ne font pas partie du champ Ésane** en 2016 (elles devront être intégrées un jour, mais ne le sont toujours pas fin 2016, des améliorations du répertoire et des fichiers fiscaux étant nécessaires au préalable).

[93] Catégorie juridique (codée dans Sirene) ; on utilise la CJ à 4 caractères.
[94] Activité principale exercée (codée dans Sirene) : la table repose sur les deux premiers caractères de la NAF (les divisions) à l'exception de quelques divisions qui sont détaillées au niveau sous-classe.
[95] Il s'agit d'une caractéristique statistique et pas juridique, gérée par conséquent par les comptables nationaux dans le cadre des règlements internationaux de comptabilité nationale.
[96] Département des Comptes Nationaux de l'Insee.
[97] Depuis le 15 juillet 2007, Saint Barth et Saint Martin sont des Collectivités d'Outre Mer, qui ne font plus partie du Département d'Outre Mer de la Guadeloupe.

4.3 Les quatre sous champs de production

Le champ Ésane est découpé en quatre sous champs comme on l'a vu au § 1.2.1.1 que l'on reprend donc ici. Ces sous champs correspondent à des considérations de **production statistique** mais n'ont pas vraiment de sens en termes de diffusion **à l'exception du sous champ 1** ; en revanche, la combinaison de critères qui sert à définir le champ complet d'Ésane s'applique de façon différenciée aux quatre sous champs ; c'est ce que l'on précisera ensuite ; du fait des cinq critères à regarder pour savoir une unité légale appartient à un sous-champ ou pas, il faut garder entête qu'il est rare qu'un secteur complet de la NAF appartienne à un sous-champ :

- Le sous champ Ésane 1 (ou *champ enquête*) correspond au champ de l'Esa et de l'EAP. Les secteurs couverts représentaient en 2012 595 sous-classes sur 732 de la NAF (351 pour l'Esa, 244 pour l'EAP) dont trois sous classes des arts du spectacle vivant qui ont été ajoutées à partir de 2010. Ils sont listés dans le fichier Excel que l'on peut consulter ou télécharger à l'URL ci-dessous. Ils peuvent être considérés comme **les secteurs les plus complètement décrits** par le système Ésane puisqu'ils sont couverts à la fois par les enquêtes sectorielles, la source fiscale et les sources d'emploi. Ce sont évidemment les seuls secteurs (ou fractions de secteurs) pour lesquels on dispose des informations correspondant aux questions « sectorielles » des enquêtes. Le champ de l'EAP est également décrit en détail au § 5.2.1.

  « Champ des enquêtes » (fichier xls)

- Le sous champ Ésane 2 (ou *champ gestionnaire*) constitue une première extension du sous champ Ésane 1. Il contient 33 secteurs (ou fractions de secteurs) non couverts par l'Esa (ou l'EAP) mais correctement couverts, en termes de poids économique essentiellement, par la source fiscale. Ces secteurs **font l'objet d'une demande de statistiques** (services statistiques des ministères, Département des Comptes Nationaux de l'Insee ou autres) telle qu'y consacrer des moyens gestionnaires Ésane est justifié. Ce champ est donc décrit a priori uniquement par des sources administratives. Sa moindre "qualité" par rapport au sous champ Ésane 1 provient de ce qu'il n'y a ni contrôle, ni mise à jour systématique de l'APE, ni évidemment recensement des branches présentes (pas d'enquête sur ces secteurs). En revanche, ce champ bénéficiera d'un bon suivi des unités mortes économiquement (à la fois via des EAR de Sirene, via le FRP[98] de la DGFiP, voire pour ses plus grandes unités via un travail des gestionnaires Ésane). De même, ce champ bénéficiera des **traitements gestionnaires via les macro-contrôles** avec retour possible à l'entreprise.

- Le sous champ Ésane 3 (ou *champ automatique)* : Il correspond à 209 secteurs (ou fractions de secteurs) restants, couverts par la source fiscale mais ne faisant l'objet d'aucune demande particulière. Par rapport au sous champ Ésane 2, la différence essentielle provient de ce **qu'aucun moyen gestionnaire Ésane** n'y est consacré car **l'absence de demande de statistique** sur ce sous champ spécifique ne justifierait pas d'y consacrer des moyens humains ; tous les traitements sont automatisés.

- Le sous champ Ésane 5[99] a été ajouté à partir de la campagne 2010 et constitue une extension du champ. Il correspond aux unités qui ont été ajoutées au champ initial des années 2008 et 2009 (auto entrepreneurs, etc.). Comme pour le champ « 3 », aucun traitement ne sera effectué par les gestionnaires Ésane. **Ce sous-champ est pris en compte pour le calcul des agrégats publiés sur le site Insee et envoyés à Eurostat.** Il n'est toutefois pas pris en compte à ce stade par les comptables nationaux du Département des Comptes Nationaux ou du Département des Synthèses Sectorielles qui raisonnent sur un champ constant depuis 2008 pour leurs analyses en évolution[100].

In fine, l'explicitation du champ Ésane, défini par les modalités des cinq critères (APE x CJ x NB-SAL x MA x PRODUCT) que l'on retient, parmi les quatre sous champs ci-dessus est fournie par une **combinaison**, très complexe dans le détail, puisqu'il faut préciser, pour chaque entreprise pour laquelle on connaît la valeur des cinq critères, à quel sous champ elle appartient. Même s'il y a beaucoup de combinaisons impossibles, on voit que le nombre de combinaisons théoriques possibles, 732x250x2x2x2 **est tout à fait considérable**. La table détaillée qui décrit toutes les combinaisons réelles utilisées **contient plus de 2200 combinaisons.**

[98] Fichier des Redevables Permanents.

[99] Ce sous-champ supplémentaire porte le numéro 5 ! Comme on l'a vu dans le chapitre 1, il était prévu à l'origine un sous-champ 4 qui n'a jamais été finalement utilisé ; mais du coup on a conservé le numéro 5 pour ce sous champ supplémentaire.

[100] Cela ne signifie bien évidemment pas que les auto-entrepreneurs ne sont pas pris en compte par la comptabilité nationale mais qu'ils le sont par une autre source qu'Ésane.

Le tableau suivant fournit la répartition du champ Ésane selon les sous champs et ce, en termes d'entreprises et de chiffre d'affaires **en 2010** (ces chiffres sont assez structurels et ne varient qu'à la marge d'année en année, sauf pour le nombre des auto-entrepreneurs qui augmente de deux à trois cent mille chaque année mais beaucoup restent inactifs).

Sous champs	Unités légales		Chiffre d'affaires 2010	
	Nombre	%	Montant (en millions d'€)	%
1	2 726 306	66,3	3 456 636	94,1
2	99 651	2,4	84 974	2,3
3	614 631	14,9	83 872	2,3
5	673 333	16,4	46 798	1,3
Total	**4 113 921**	**100,0**	**3 672 281**	**100,0**

Comme on peut le voir sur le tableau, les deux tiers des entreprises du champ Ésane font partie du champ de l'Esa ou de l'EAP, ce qui représente **94% du chiffre d'affaires total**. Le sous champ « 5 » représente une part non négligeable du champ Ésane en termes d'entreprises (16%) mais son poids en termes de chiffre d'affaires ou de valeur ajoutée est marginal (1%). Ce sous champ est constitué principalement **d'auto entrepreneurs** (devenus micro-entrepreneurs[101] fin 2014) (89% des entreprises du sous champ « 5 », qui par définition n'existaient pas au démarrage en 2008).

Si on rajoute la répartition des sous-champs par grand secteur de l'économie, on obtient le tableau ci-dessous :

- Dans les IAA, la Construction, le Commerce, les transports et l'Industrie, le champ de l'Esa constitue l'intégralité ou la quasi-intégralité du champ Ésane en termes de chiffre d'affaires. Dans ces secteurs, la part des entreprises du sous champ « 1 » dans le champ Ésane varie de 86% à 94%. Les autres entreprises de ces secteurs sont essentiellement des auto-entrepreneurs et sont donc classées dans le sous champ « 5 ».
- Dans les services, seule la moitié des entreprises font partie du champ de l'Esa, ce qui représente les trois quarts du chiffre d'affaires du secteur. Les entreprises des sous champs « 3 » et « 5 » représentent respectivement 26,4% et 18,8% des entreprises des services, soient respectivement 10% et 5,3% du chiffre d'affaires du secteur.

Secteurs d'activité	Sous champs	Unités légales		Chiffre d'affaires 2010	
		Nombre	%	(en millions d'euros)	%
IAA	1	27 887	94,1	162 129	100,0
	5	1 745	5,9	43	0,0
	Total	**29 632**	**100,0**	**162 172**	**100,0**
Construction	1	471 667	85,2	257 229	99,3
	3	2	0,0	1	0,0
	5	81 645	14,8	1 723	0,7
	Total	**553 314**	**100,0**	**258 953**	**100,0**
Commerce	1	777 641	86,4	1 408 654	100,0
	5	122 810	13,6	666	0,0
	Total	**900 451**	**100,0**	**1 409 320**	**100,0**
Industrie	1	176 472	87,7	814 391	100,0
	2	259	0,1	132	0,0
	3	46	0,0	41	0,0
	5	24 450	12,2	108	0,0
	Total	**201 227**	**100,0**	**814 672**	**100,0**
Services	1	1 178 123	50,5	640 499	76,3
	2	99 363	4,3	70 955	8,5
	3	614 575	26,4	83829	10,0
	5	438 564	18,8	44 239	5,3
	Total	**2 330 625**	**100,0**	**839 522**	**100,0**
Transports	1	94 516	95,8	173 735	92,6
	2	29	ε	13 886	7,4
	3	8	ε	1	ε
	5	4 119	4,2	18	ε
	Total	**98 672**	**100,0**	**187 641**	**100,0**
Total	**1 à 5**	**4 113 921**	**100,0**	**3 672 281**	**100,0**

[101] A ne pas confondre avec l'ensemble des micro entreprises au sens fiscal (Cf. par exemple le § 6.5.3).

(Où ε < 0.05 %)

En 2010, les entreprises sont des unités légales, à l'exception permanente de cinq groupes dits « historiques » qui ont été profilés de façon expérimentale avant les autres (Cf. § *11.1*).

4.4 La définition en extension du champ

Comme on l'a vu, l'explicitation complète, en extension, du champ en fonction des cinq critères et des quatre sous champs contient 2200 combinaisons. Ce qui compte essentiellement est la distinction entre le sous champ 1 et les autres sous champs. En effet, pour le sous champ 1, on dispose directement ou indirectement (par pondération des unités répondantes de l'échantillon) des résultats des enquêtes Esa ou EAP ; **les résultats dits sectoriels ne sont donc connus que pour ces secteurs**. Les sous champs 2 et 3 inversement, ne se distinguent que par l'intervention des gestionnaires.
On se contente donc ici d'une « vue » du champ par critère ou sous champ, indépendamment les uns des autres, visant à expliciter pour chacun d'entre eux, les modalités retenues ou inversement exclues. Mais il ne faut pas perdre de vue que **c'est bien la combinaison complète** qui permet de savoir comment un secteur est représenté ou comment une entreprise est prise en compte.

Exemples : on va voir au *§ suivant 4.4.1* que les sections P « enseignement » ou Q « santé humaine et action sociale » sont « représentées » dans Ésane mais il ne faut pas oublier que ne seront suivies que les entreprises « marchandes » et « participant au système productif sur un marché » ; on n'aura donc aucune statistique complète dans Ésane sur l'enseignement public et privé ou sur la santé puisqu'on ne suit pas du tout la partie publique.

4.4.1 Sélection sur le code APE : les sections et les groupes de la NAF représentés

La sélection du champ sur le code APE correspond, en Naf rév. 2, aux activités :

- des sections :
  - B - Industries extractives
  - C - Industrie manufacturière
  - D - Production et distribution d'électricité, de gaz, de vapeur et d'air conditionné
  - E - Production et distribution d'eau ; assainissement, gestion des déchets et dépollution
  - F - Construction
  - G - Commerce ; réparation d'automobiles et de motocycles
  - H - Transports et entreposage
  - I - Hébergement et restauration
  - J - Information et communication
  - L - Activités immobilières
  - M - Activités spécialisées, scientifiques et techniques
  - N - Activités de services administratifs et de soutien
  - P - Enseignement
  - Q - Santé humaine et action sociale
  - R - Arts, spectacles et activités récréatives
  - S - Autres activités de services
- des groupes à trois chiffres de la NAF :
  - 02.2 - Exploitation forestière (section A)
  - 66.1 - Activités auxiliaires de services financiers, hors assurance et caisses de retraite (section K)
  - 66.2 - Activités auxiliaires d'assurance et caisses de retraite (section K)
- Et du 64.20Z (Holdings) (section K)

- **La section A** n'est donc suivie qu'au travers des exploitations forestières (à la demande du ministère de l'agriculture et parce que les exploitations forestières sont souvent couplées avec des scieries et s'apparentent donc davantage à des entreprises industrielles qu'à des entreprises agricoles). Elles sont donc suivies par l'Insee dans le cadre d'Ésane. Le reste du champ agricole est suivi par le ministère de l'agriculture, en dehors d'Ésane.

- La section K (activités financières) est suivie au travers de la seule division 66 (hors le 66.3 gestion de fonds). Les entreprises du secteur « des services financiers, hors assurance et caisses de retraite » (division « 64 ») et des « assurances » (division « 65 ») ont été ajoutées dans Ocsane mais ne font pas partie du champ de diffusion Ésane. Les liasses fiscales de ces entreprises sont chargées

dans Ésane mais ne font l'objet d'aucun traitement. **Les holdings (Ape « 64.20Z ») font partie du champ Ésane à compter de 2010** comme on le verra au *§ 4.6*. En 2010, 11 332 entreprises appartiennent au secteur des services financiers et 1 238 au secteur des assurances.

Les divisions **64** (hors le 64.20Z) et **65** sont suivies par la **Banque de France**.

- **Les sections O** (administration publique) **T** (Activités des ménages en tant qu'employeurs ; activités indifférenciées des ménages en tant que producteurs de biens et services pour usage propre) et **U** (Activités extra territoriales[102]) **ne sont pas suivies du tout**.

Au total, finalement, **il y a 666 sous-classes qui sont suivies dans Ésane** (mais dans différents sous champs). Leur liste est donnée dans le fichier Excel que l'on peut consulter ou télécharger à l'URL suivante :

« Sous-classes NAF Ésane » (fichier xls)

4.4.2 Sélection sur la catégorie juridique (CJ) : les CJ exclues du champ

Catégories juridiques exclues du champ	Observations et exceptions
1800- Associé - gérant de société	
2210 - Société de fait entre personnes physiques	Réintégrée à partir de la campagne 2012
2220 - Société créée de fait avec une personne morale	hormis les unités profilées
2310 - Société en participation entre personnes physiques	Si non employeur hormis une liste d'unités définie par la division Ésane
2320 - Société en participation avec une personne morale	
6316 - Coopérative d'utilisation de matériel agricole (CUMA)	
6589 - Société civile de moyens (SCM)	
71XX - Administration de l'État	
72XX - Collectivité territoriale	
73XX - Établissement public administratif sauf exceptions (Cf. ci-contre)	La catégorie juridique « 7371 »[103] est dans le champ Ésane. Les catégories juridiques « 7353 »[104], « 7354 »[105], « 7355 »[106] (correspondant aux syndicats des eaux) sont dans le champ Ésane pour les divisions « 36 », « 37 » et « 96 » de la Naf rév. 2. Les 28 CROUS classés dans la catégorie juridique « 7385 »[107] sont inclus dans le champ Ésane.
74XX - Autre personne morale de droit public administratif	
81XX - Organisme de protection sociale	
82XX - Organisme mutualiste	
83XX - Comité d'entreprise	
91XX - Syndicat de propriétaires	
93XX - Fondation	
99XX - Autre personne morale de droit privé	

[102] Il s'agit des activités des ambassades par exemple qui sont bien des unités françaises mais situées à l'étranger.
[103] Office public d'habitation à loyer modéré (OPHLM).
[104] Syndicat intercommunal à vocation unique (SIVU).
[105] Syndicat mixte communal.
[106] Autre syndicat mixte.
[107] Autre établissement public national administratif à compétence territoriale limitée.

Les catégories juridiques **exclues du champ** sont définies dans le tableau ci-dessus : comme l'indique la deuxième colonne, il existe toutefois des exceptions aux exclusions.

La mention « XX » signifie simplement que l'exclusion vaut pour toutes les catégories juridiques de la classe définie par les deux premiers caractères.

4.4.3 Les activités couvertes par le sous champ 1

4.4.3.1 Champ de la maîtrise d'œuvre Insee

Le sous champ 1 est donc par définition le champ des enquêtes (Esa ou EAP). **L'Esa** couvre précisément en 2012 **351 sous-classes** NAF et **l'EAP 244** (soit 595 au total). Leur liste figure dans le fichier Excel joint ci-dessus au § 4.3.

Mais attention : dire qu'une sous-classe est couverte par une enquête ne signifie pas que la totalité du secteur est couvert par l'enquête : d'une part, **les auto-entrepreneurs** (devenus micro-entrepreneurs) **ne sont jamais interrogés dans les enquêtes Esa ou EAP**. D'autre part, il est fréquent, comme on l'a vu, qu'on ne suive dans l'enquête pour certaines sous-classes **que les entreprises de certaines catégories juridiques de plus de 5 salariés ou de 30 et plus** et donc que les résultats portant sur les questions purement sectorielles ne concernent qu'une partie des entreprises de ces sous-classes.

4.4.3.2 Répartition des maîtrises d'ouvrage de l'Esa par « grand secteur[108] »

Une des caractéristiques d'Ésane et de l'Esa en particulier est la **séparation** entre **la maîtrise d'œuvre** de la responsabilité du département de l'Insee « Répertoires, infrastructures et statistiques structurelles », des **différentes maîtrises d'ouvrage concernées (MOA)**, au nombre de six. On en donne ci-dessous la répartition par rapport à la nomenclature officielle :

IAA [MOA : le Service de la statistique et de la prospective (SSP) au ministère de l'agriculture et de la pêche]
Divisions : 10 (sauf 10.13B, 10.71B, 10.71C, 10.71D), 11 et 12 ;

Sous classes : Exploitations forestières et Scieries : 02.20Z,
Sciage et rabotage du bois hors imprégnation 16.10A
Imprégnation du bois : 16.10B ;

BTP [MOA : la sous-direction des statistiques du logement et de la construction (SDSLC) au SOeS]
BTP (41.2, 42, 43) et Promotion immobilière (41.1)

Services [MOA : le département des synthèses sectorielles (DSS) de l'Insee]
Divisions : 55 (sauf 559), 56, 58, 59, 60, 61, 62, 63, 68 (sauf 68.32B), 69, 70, 71, 73, 74, 77, 78, 79, 80, 81 (sauf 81.30Z), 82, 95, 96
À partir de 2010, il faut rajouter les sous-classes : *90.01Z, 90.02Z, 90.04Z* (activités du spectacle vivant, soutien au spectacle vivant et gestion de salles de spectacles).

Transport [MOA : la sous-direction des statistiques de transport (SDST) au SOeS)
Divisions : 49, 50, 51 (y ***compris APE 51.22Z*** *- transports spatiaux), 52,*
Sous-classe : 53.20Z (autres activités de poste et de courrier)

Commerce [MOA : le département des synthèses sectorielles (DSS) de l'Insee]
Divisions : 45, 46 (y compris les coopératives agricoles des groupes 46.2, 46.3 et du 46.75Z), 47 + 10.13B, 10.71B, 10.71C, 10.71D

Industrie (pour les DOM) [MOA : les directions régionales de l'Insee correspondantes]
Divisions : 05, 06, 07, 08, 13 à 33 (sauf 16.10A et 16.10B), 35, 37, 38, 39
Sous-classes : 09.10Z, 09.90Z, 36.00Z

[108] Rappelons que l'industrie (hors DOM) est suivie au travers de l'EAP et non l'Esa.

4.4.4 Les activités classées dans les sous champs 2 ou 3

- Les **33 activités** avec les valeurs correspondantes de la catégorie juridique ou des effectifs classées (en 2012) dans **le sous champ 2** figurent dans le fichier Excel que l'on peut consulter ou télécharger à l'URL suivante :

« Sous-champ 2 complet » (fichier xls)

- Les **209 activités** avec les valeurs correspondantes de la catégorie juridique ou des effectifs sont classées (en 2012) dans **le sous champ 3** figurent dans le fichier Excel que l'on peut consulter ou télécharger à l'URL suivante :

« Sous-champ 3 complet » (fichier xls)

4.5 Les différents champs de la statistique d'entreprise et de diffusion d'Ésane

Comme on l'a souvent expliqué, Ésane est un système d'information sur les entreprises qui permet de **calculer et de diffuser des agrégats**, restreints à des champs particuliers, pour un public que l'on suppose intéressé par ces champs particuliers. Comme on le verra un peu plus loin, il y en a principalement **trois**, emboités les uns dans les autres. En tant que système d'information, le champ d'Ésane complet est donc la réunion de tous ces champs ; il est lui-même contenu dans le champ global de la statistique d'entreprises qui comprend le secteur agricole (alors que, comme on l'a vu, Ésane ne comprend que le secteur agricole des exploitations forestières 02.20Z). Passons en revue les champs les plus couramment utilisés[109] :

Le champ du répertoire Sirene est le plus large de tous les champs : la réglementation conduit à immatriculer des unités que la statistique n'interroge pas car elles n'exercent pas véritablement d'activité économique ou celle-ci est observée dans d'autres structures juridiques (sociétés civiles immobilières crées par les ménages, sociétés civiles de moyens, etc.). Rappelons que Sirene est un répertoire administratif dont l'utilisation dépasse de loin la statistique.

Le champ du répertoire SIRUS (qui est bien, lui, un répertoire statistique Cf. § *11.7*) est inclus dans le champ de Sirene mais plus large que le champ de la statistique d'entreprises. En effet, il doit également répondre aux besoins des statistiques sur **l'emploi** (utilisées par le DERA[110] et la Dares[111]). Il est donc formé par les unités constituant la statistique d'entreprises ainsi que par les unités **employeuses, même hors du champ de la statistique d'entreprises** (organismes non marchands, secteur public, ISBLSM[112], syndicats mixtes, syndicats de copropriété et autres unités employeuses).

Le champ de la statistique d'entreprises est formé des unités **participant au système productif sur un marché[113].**

Le champ SBS complet est le champ sur lequel la France doit envoyer des résultats à Eurostat au titre du règlement « Structural Business Statistics » (règlement 295/2008[114] dit SBS recast). Comme on l'a vu en introduction du chapitre, il **comprend le secteur financier** mais les données correspondantes sont élaborées et envoyées à Eurostat par la Banque de France (voir plus précisément le § *4.4.1* pour le partage entre la Banque de France et l'Insee) ; il ne comprend pas le secteur agricole.

---

[109] Le graphique ci-après donne une estimation des ordres de grandeur du volume de ces champs, mesurée en 2010 ; mais leur taille augmente assez vite puisque de 2010 à 2016 le nombre d'unités de Sirene par exemple est passé de 8,5 millions à plus de 10 millions.

[110] DERA : Département de l'emploi et des revenus d'activités (Insee DSDS)

[111] Dares : Direction de l'animation de la recherche, des études et des statistiques au ministère du travail, de l'emploi et de la formation professionnelle.

[112] Institutions sans but lucratif au service des ménages.

[113] Caractérisées par le code « PRODUCT = Oui » dans le répertoire SIRUS.

[114] *http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:097:0013:0059:FR:PDF*.
Ce règlement est constitué de différentes « annexes » qui précisent les agrégats que doivent fournir les pays et le champ de ces agrégats. L'annexe I correspond aux statistiques communes à tous les secteurs ; les annexes II à VIII à des secteurs particuliers et l'annexe IX à la démographie d'entreprises. Il devrait être remplacé à compter de 2019 environ par un nouveau règlement cadre dit « FRIBS » pour « Framework regulation integrating business statistics ».

- **Pour « le module commun », correspondant à l'annexe I,** relatif à l'ensemble des statistiques structurelles annuelles publiées par Eurostat, le champ **ne comprend pas les sections P, et Q** de la Nace correspondant aux secteurs de l'enseignement, de la santé humaine et d'action sociale, et de services aux ménages. Il faut donc **faire attention** quand on compare les chiffres France publiés par l'Insee avec ceux publiés par Eurostat. Leur champ est notablement différent (Cf. aussi ci-dessous les champs de diffusion et le § *9.4.5*).
- **Pour la démographie d'entreprises, correspondant à l'annexe IX, le champ comprend, lui, les sections P et Q** ce qui fait que le nombre total d'entreprises diffère selon les annexes que l'on consulte.

Le champ Ésane complet[115] est défini comme l'ensemble des unités participant au système productif sur un marché et marchandes[116], à l'exclusion dans un premier temps du secteur agricole, mais en conservant toutefois les unités « agricoles » codées en Apen[117] 02.20Z (exploitations forestières, incluant notamment les scieries). C'est le champ SBS complété par ce secteur particulier.

Les trois principaux champs de diffusion d'Ésane sont les suivants (avec leur intitulé tel qu'il figure dans les tableaux de diffusion le plus souvent) :

1. **Secteurs marchands non agricoles** : il s'agit du champ de diffusion le plus large ; il correspond à la déclinaison des secteurs listés en tête du § 4.4.1 sans le secteur de l'exploitation forestière. Il est plus petit que le champ SBS complet car il ne comprend que la division 66 et les holdings 64.20Z pour représenter la section K des activités financières [il manque le 65 (assurance) et le reste du 64 (activités des services financiers)] ;
2. **Secteurs principalement marchands non agricoles** : par rapport au champ précédent, on enlève les sections P et Q (activités d'enseignement, de santé humaine et d'action sociale) comme pour l'annexe I du règlement SBS ; le plus gros de la valeur ajoutée de ces secteurs est en effet réalisée par le secteur public et sa partie privée n'est donc pas représentative des secteurs en question ;
3. **Secteurs principalement marchands non agricoles et non financiers** : par rapport au champ précédent, on enlève la partie couverte de la section K (le 66 et le 64.20Z) car les données comptables des entreprises financières ne sont pas homogènes avec celles des autres secteurs.

Le champ d'intérêt des maîtrises d'ouvrage sectorielles sera défini comme la restriction du champ Ésane complet aux secteurs d'activités de la compétence de celles-ci.

Le tableau ci-après synthétise la situation et le graphique qui suit donne une idée des ordres de grandeur en termes de nombre d'entreprises :

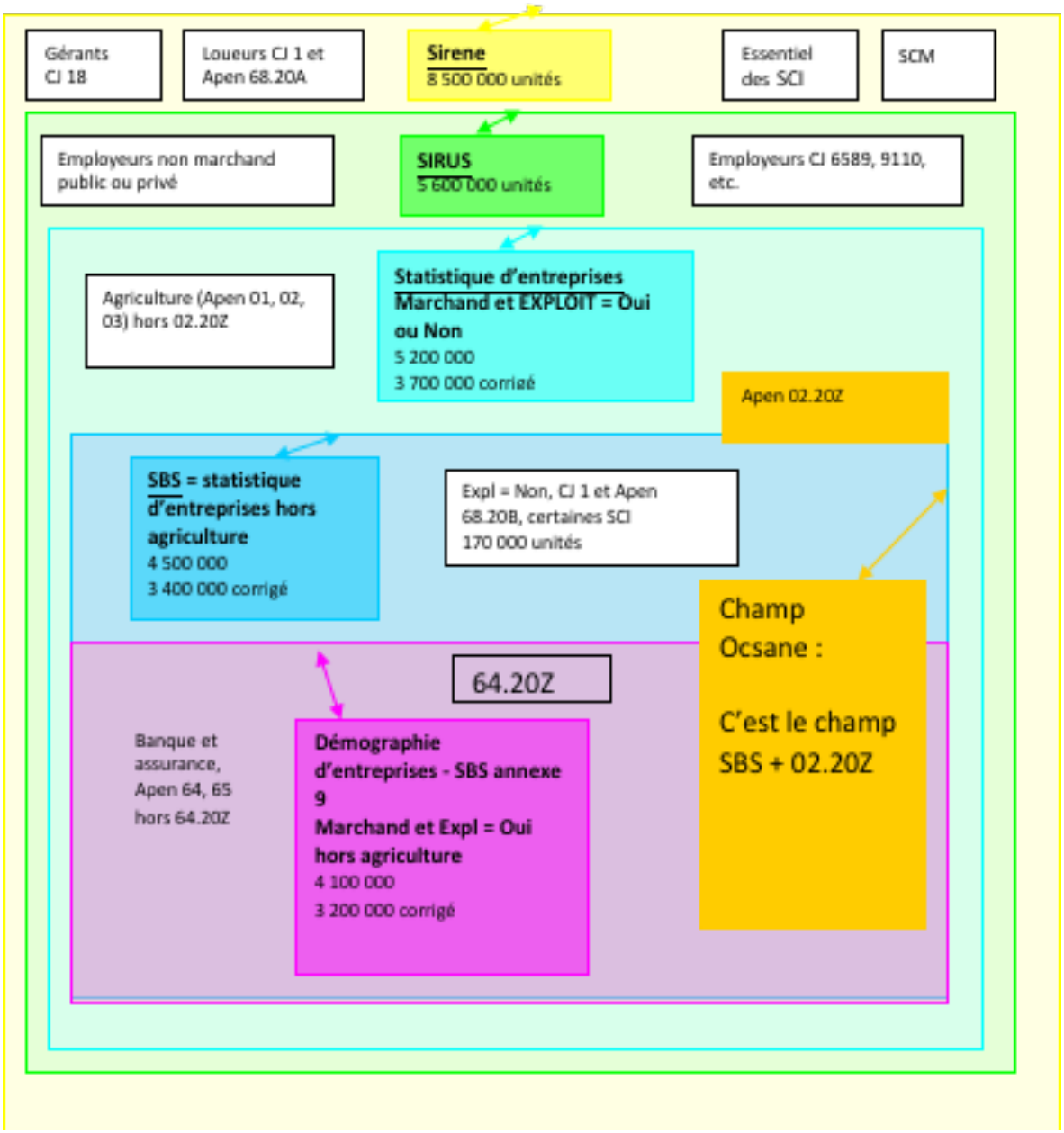
[115] Comme on va le voir au § *4.6*, les champs Ocsane (sur lequel on calcule les agrégats d'Ésane) et donc SBS (dont la définition est moins précise en termes de CJ en particulier) peuvent évoluer à la marge selon les années.
[116] Caractérisées par le code « Marchand = MA » dans le répertoire SIRUS.
[117] L'Apen désigne le code d'activité principale de niveau « entreprise » par opposition au code Apet qui désigne le code d'activité d'une unité locale (établissement). On ne raisonne dans Ésane que sur des « entreprises » (unités légales ou unités profilées) qui possèdent donc un code Apen.

Tableau résumant la définition des différents champs <u>pour le décompte du nombre d'entreprises :</u>

Champs	Définition	Principales exclusions
Champ **Sirene**	Toutes unités	Aucune
Champ **SIRUS**	Marchand et participe au système productif + unités employeuses	Gérants, essentiel des SCI, SCM, partie des groupements sans personnalité morale, loueurs personnes physiques
Champ global **statistique d'entreprises**	Marchand et participe au système productif	Non marchand, ne participe pas au système productif
Champ **SBS complet** - hors agriculture	Marchand et participe au système productif hors Apen 01, 02 et 03	Apen 01, 02, 03
Champ **démographie d'entreprises** (Annexe IX du règlement SBS)	Marchand, (produit des biens et services sur un marché) hors Apen 01, 02 et 03	(n'exploite pas des moyens de production)
Champ **Ésane complet**	C'est le champ SBS complet, mais en y rajoutant le secteur 02.20Z	
Secteurs marchands non agricoles	C'est le champ Ésane complet sans le 02.20Z, le secteur 64 (à l'exception du 64.20Z) ni le secteur 65.	Secteurs 02.20Z, 64 (sauf 64.20Z), 65
Secteurs principalement marchands non agricoles	Secteurs marchands non agricoles moins sections P, Q	Secteurs 02.20Z, 64 (sauf 64.20Z), 65, P, Q
Secteurs principalement marchands non agricoles et non financiers	Secteurs principalement marchands non agricoles moins le 64.20Z et le 66	Secteurs 02.20Z, K, P, Q

Graphique illustrant ce tableau : imbrication des différents champs définis pour le seul décompte du nombre d'entreprises :

Les décomptes donnés ci-dessous sont des **estimations fin 2010**. A noter que le dynamisme de la création d'entreprises fait croître tous ces nombres chaque année. En 2016, Sirene compte plus de 10 millions d'unités.



Les chiffres dits « corrigés » dans le schéma ci-dessus sont obtenus en éliminant les unités sans déclaration fiscale sur les 3 dernières années, considérées comme des faux actifs. (Cf. § 11.7.1 alinéa 8 pour la définition des fausses actives)

4.6 Les modifications de champ entre 2008 et 2010

Les résultats des campagnes successives mettent en évidence certains manques dans le champ défini au départ pour la campagne 2008 qu'on a donc ensuite voulu "corriger" (on appelle "**campagne N" les résultats afférant à l'année de constat N** ; mais bien entendu cette campagne a lieu pendant l'année N+1 et même jusqu'à l'année N+2). Les évolutions réglementaires (cas des auto-entrepreneurs), ou les demandes des maîtrises d'ouvrage amènent aussi à faire évoluer le champ **plus ou moins à la marge chaque année**, comme on va le voir. Dans l'optique de la statistique structurelle qui décrit comme son nom l'indique des **effets de structure à base de ratios comptables peu évolutifs, cela n'est pas très ennuyeux.** La façon de faire a cependant changé à

partir de la campagne 2011, raison pour laquelle on a séparé les trois premières campagnes des campagnes suivantes qui sont désormais gérées différemment (Cf. § *4.7* et *4.8*).

4.6.1 Campagne (ou année de constat) 2009

C'est le cas le plus simple. Le champ 2009 est **strictement identique à celui de l'année 2008**. On peut donc sans problème comparer les résultats 2009 aux résultats 2008.

Mais cela signifie aussi qu'on n'a pas introduit les auto-entrepreneurs qui étaient un régime nouveau créé en 2009. Ils ont été ensuite rajoutés globalement par les comptables nationaux dans les comptes nationaux français mais ne figurent pas dans les résultats Ésane de l'année 2009.

4.6.2 Campagne (ou année de constat) 2010

4.6.2.1 Les modifications apportées au champ de l'Esa

- les **entreprises du spectacle vivant** (APE « 9001Z - Arts du spectacle vivant », « 9002Z - Activités de soutien au spectacle vivant » et « 9004Z - Gestion de salles de spectacle ») ont été intégrées dans le champ de l'Esa (sous champ 1) à la demande du ministère de la culture. Lors des campagnes précédentes, ces entreprises faisaient partie du sous-champ « 2 ». En 2010, 17 233 entreprises étaient classées dans ces secteurs d'activités.

- les **loueurs de fonds personnes physiques** (APE « 6820B - Location de terrains et d'autres biens immobiliers » et catégorie juridique commençant par « 1 ») sont passés dans le sous-champ « 3 ». Jusqu'alors, ils faisaient partie du champ de l'Esa. 60% des entreprises classées dans l'activité « Location de terrains et d'autres biens immobiliers » sont en réalité des personnes physiques (soit 64 861 entreprises sur 107 209) qui ne s'y retrouvaient pas du tout dans la réponse à l'enquête et répondaient donc très mal ou pas du tout.

4.6.2.2 L'introduction des syndicats des eaux

Ce qu'on appelle couramment les syndicats des eaux sont des établissements publics qui gèrent le service public de l'eau potable généralement pour un ensemble de communes, ou quelquefois d'autres structures de collectivités locales (intercommunalités, communauté d'agglomération, etc.). Leur catégorie juridique est donc généralement 7353 (SIVU) mais peut être aussi 7354 (syndicat mixte communal) ou 7355 (autre syndicat mixte). Ils ont été intégrés à partir de 2010 dans le champ d'Ésane (sous champ 2) (cela représente environ 3000 unités légales).

4.6.2.3 Les autres modifications de champ et la création du sous champ « 5 »

Ce sous champ a été créé pour la campagne 2010, essentiellement pour pouvoir comparer les résultats de l'année 2010 avec ceux de l'année 2009 à la fois sur un champ identique ou avec la totalité du champ de 2010. On a donc mis dans ce sous champ 5 toutes les modifications du champ 2010 (il n'y avait que des ajouts) et on a calculé à part les agrégats sur ce sous champ de façon à pouvoir calculer les agrégats finaux avec et sans ce sous-champ.

Hormis les deux modifications du sous champ de l'Esa ci-dessus, il comprend tous les autres ajouts du champ en 2010, à savoir :

- 1) Les auto-entrepreneurs (catégories juridiques « 1290 », « 1390 » et « 1990 »). Ces entreprises, dont le statut a été créé en 2009, ne faisaient pas partie du champ Ésane en 2009. On ne prend en compte **que les auto-entrepreneurs qui exercent réellement une activité**, c'est-à-dire qui ont déclaré un chiffre d'affaires positif à l'Acoss et dont l'Acoss nous fournit la liste exhaustive[118].

- 2) Les Sociétés civiles immobilières (SCI) (catégories juridiques « 6540 - Société civile immobilière », « 6541 - Société civile immobilière de construction - vente » et « 6599 - Autre société civile ») ayant un chiffre d'affaires positif en 2008[119] : ces sociétés (plus d'un million) avaient été écartées au départ car on pensait qu'il s'agissait essentiellement de sociétés transparentes fiscalement et servant presque exclusivement les ménages. Puis on s'est rendu compte que de

[118] Cette restriction ne change évidemment pas les variables comptables mais modifie assez notablement le nombre d'unités légales du champ car seule environ la moitié des 300 000 auto-entrepreneurs qui se créent tous les ans exerce une activité.
[119] Le dernier qui était disponible et validé au moment de la constitution du référentiel 2010 (en novembre 2010).

nombreux groupes utilisaient également cette forme juridique pour gérer leur patrimoine immobilier ; ils font bien sûr payer cette mise à disposition et réalisent donc du chiffre d'affaires ; d'où le critère utilisé pour leur réintégration. Ainsi, a priori, seules les sociétés civiles immobilières **déclarant un chiffre d'affaires positif** font partie du champ Ésane.

Le chiffre d'affaires devient donc, de fait, un nouveau critère d'appartenance au champ (en plus des cinq déjà évoqués plus haut) tant pour les auto-entrepreneurs que pour les SCI. Toutefois, **l'utilisation explicite de ce critère dans Ocsane n'est pas possible** avant le lancement de la campagne puisqu'il nécessite de connaître les liasses fiscales et oblige à des **traitements spécifiques ex post des sous populations correspondantes** :

- S'agissant des **auto-entrepreneurs**, l'information est de plus exogène au système et fournie par l'Acoss. Elle conduit à un traitement spécifique tout à fait en fin de chaîne.
- S'agissant des **SCI**, l'information est disponible dans le système (présence ou absence de liasse dans les livraisons DGFiP). Le traitement consiste à modifier le fonctionnement normal des imputations de façon à ne pas imputer à tort comme liasse manquante, la totalité des SCI qu'on aurait dans Sirene.

- 3) Les holdings financières (Ape « 64.20Z ») : à la suite du changement de nomenclature de 2008, certaines unités classées auparavant en « Administration d'entreprises » (741J) en Naf rév. 1.1, ont été reclassées dans le secteur financier (64.20Z) en Naf rév. 2, alors qu'elles n'apparaissent pas dans la liste des sociétés financières repérées par la Banque de France. En conséquence, la distinction entre holdings financières et sièges sociaux paraissant floue, il a paru préférable d'intégrer les holdings financières au champ d'ÉSANE pour éviter des trous de collecte.

Les résultats d'Ésane 2010 ont été publiés à l'aide de l'ensemble des sous champs (donc y.c. le sous champ 5). Toutefois, la distinction de ce qui a été spécifiquement rajouté a permis de publier quelques tableaux de cadrage explicitant les évolutions qui s'expliquent par l'évolution du champ. **A partir de 2011, on a modifié la stratégie de publication** lorsque le champ est modifié une année donnée, en publiant systématiquement l'année N sur le champ de l'année N-1 pour faciliter les comparaisons. En revanche, **il ne faut pas comparer sans précaution N et N-2**.

4.6.2.4 La comparaison 2010 – 2008/2009

La comparaison des champs (sur les bases de sondage[120]) 2010 et 2008/2009[121] ne fait pas apparaître de grosses divergences au niveau global. En 2008, un peu moins de 3,5 millions d'entreprises (3 454 187 exactement) faisaient partie du champ Ésane contre 4,1 millions en 2010. Cette hausse du nombre d'entreprises est essentiellement liée à l'intégration des auto-entrepreneurs en 2010.

La répartition des sous champs notamment en termes de nombre d'unités légales se trouve modifiée suite à la création du sous-champ « 5 ». En effet, 66% des entreprises du champ Ésane font partie du champ de l'Esa en 2010 contre 81% en 2008. Par contre, la structure en termes de chiffres d'affaires a peu varié. En effet, le chiffre d'affaires des entreprises du champ de l'Esa représente 94,1% du champ Ésane en 2010 contre 95,5% en 2008.

4.7 Les modifications de champ et de mode de diffusion, à compter de la campagne 2011

Comme il ne paraissait pas optimal de rajouter un sous champ différent, chaque année, **à partir de la campagne 2011**, on a procédé autrement en intégrant les nouveautés à ce même sous-champ 5 et en publiant en fait **les résultats de chaque année deux fois** : une fois sur le champ de l'année précédente pour avoir des résultats en comparable avec N-1, puis un an plus tard en même temps que l'année N+1, sur le « champ nouveau » **qu'on baptise du coup champ « N+1 »** puisque les résultats sont publiés à l'occasion de la publication de l'année N+1. En d'autres termes, un changement de champ observé en N ne sera pris en compte qu'en N+1 car on a besoin de deux années d'observation pour pouvoir calculer des statistiques en évolution qui aient du sens.

Ainsi, **2011** a d'abord été publié (à l'été 2013 Cf. § *9.4.1*) sur un champ 2011 **identique à celui de 2010** (ne prenant donc pas encore en compte les évolutions propres à 2011) ; puis une deuxième version a été diffusée lors de la publication des résultats de 2012 (à l'été 2014) qui eux prenaient en

[120] On compare là les comptages au moment de la constitution de la base de sondage d'Ésane et du tirage des échantillons de l'Esa. Les résultats finaux d'Ésane sont évidemment un peu différents car en un an on prend connaissance (avec retard) de la disparition de nombreuses entreprises.

[121] Les chiffres qui suivent sont issus d'une note de l'époque reprenant donc les données les plus rapidement mobilisables. Rappelons que les champs 2008 et 2009 étaient les mêmes.

compte toutes les évolutions de champ qu'on avait programmé d'introduire lors de la « campagne 2011 » Mais comme ces modifications sont introduites **lors de la première publication de 2012**, on l'appelle « 2011 champ 2012 ». Au moment de la publication de 2012, on peut donc comparer les agrégats de 2011 et de 2012 qui sont calculés sur le même champ.

Cette façon de faire a **été ensuite généralisée les années suivantes :** chaque année on diffuse en même temps que les résultats de l'année N sur le champ N les résultats de l'année N-1 recalculés sur le champ N. On constitue ainsi des maillons annuels qui sont à **champ et méthodologie constants** ; en revanche, on ne peut pas comparer N et N-2 sans précaution.

4.7.1 Campagne 2011 / champ 2012

4.7.1.1 Les modifications apportées au champ (prises en compte dans la publication 2012)

La campagne de 2011 a, comme celle de 2010, connu quelques évolutions de champ notamment suite à la redéfinition des codes sous-jacents à la définition du champ de la statistique d'entreprises. Les codes **Marchand** (unité tirant au moins 50 % de ses ressources de ventes sur un marché) et **Product** (unité appartenant au système productif directement ou indirectement) **ont été définis dans Sirus** pour tenir compte d'une expertise plus approfondie du champ des entreprises. Ceci a provoqué l'entrée dans le champ :

- de nouvelles SCI **qui n'avaient pas été intégrées en 2010 ;**
- et d'un nombre important de personnes physiques.

Les principales modifications (entrées) de champ sont détaillées ci-dessous.

Il y avait déjà quelques SCI dans Ésane à partir de 2008 (entre 4000 et 5000, en principe celles qui sont employeuses, et c'est à l'occasion de la campagne (l'exercice) 2011 (en plus de celles de 2010) que les autres SCI au service des entreprises (appartenant à un groupe ou émettant une déclaration fiscale) ont vraiment été réintégrées dans le champ Ésane en sous champ 5. Ceci représente environ 23 000 unités en plus des 4 ou 5 000 initiales.
S'agissant des personnes physiques, notamment celles issues du "CFE[122] Impôts", elles étaient très systématiquement mises hors champ de la statistique d'entreprise auparavant et donc hors champ d'Ésane à l'origine. Leur entrée dans le champ 2012 est **massive en nombre** mais **pas en poids** dans l'économie car ce sont toutes de très petites unités mais qu'il est légitime de prendre en compte puisqu'elles participent au système productif.
Ces personnes physiques se trouvent très largement dans 4 sous-classes NAF : 90.01Z, 90.03A, 90.03B (professions artistiques telles que spectacle vivant, création artistique, etc.) et 68.20B (loueurs d'autres biens immobiliers) (hors logement, par exemple fonds). Les nombres d'unités en sont respectivement, 33 000, 37 000, 26 000 et 62 000. Les personnes physiques des 4 secteurs cités (90.01Z, 90.003A, 90.03B et 68.20B) sont mises en sous-champ 5 depuis la campagne 2011.

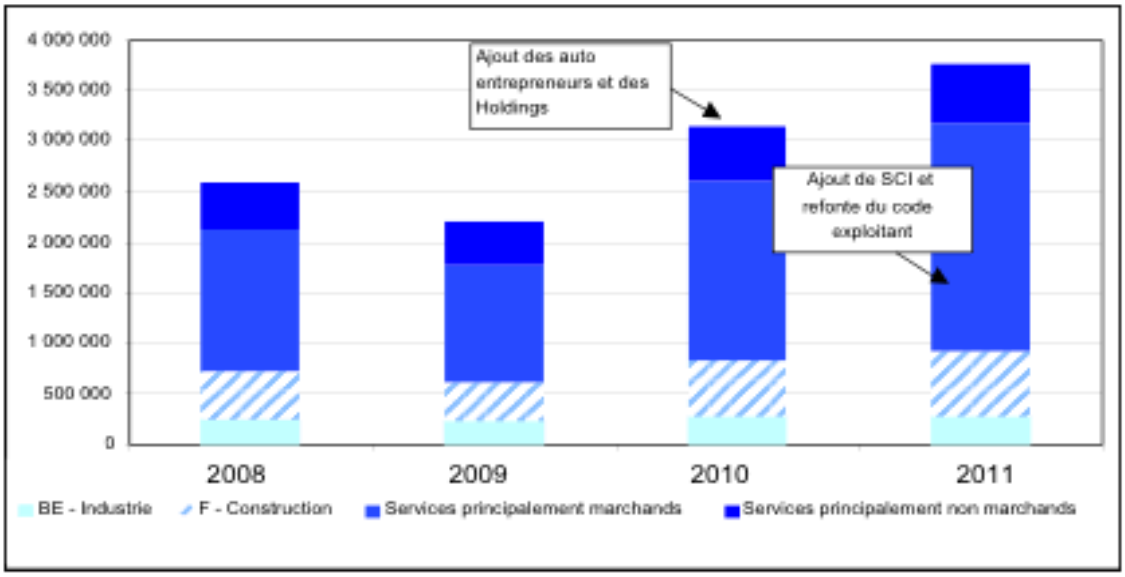
4.7.1.2 La comparaison de champ entre les campagnes 2011 - 2010

Afin de faciliter les comparaisons annuelles, les données définitives d'Ésane 2011 publiées à l'été 2013 ont été publiées « à champ constant » par rapport à 2010, **elles ne tiennent donc pas compte des entrées de champ décrites ci-dessus**. Techniquement, il a donc fallu repérer le champ 2010 dans Ésane 2011 et « retirer » en quelque sorte ces entrées de champ, pour pouvoir faire des calculs d'agrégats à champ constant.
Ce n'est qu'à l'été 2014 avec les résultats 2012, que les agrégats tenant compte de ces entrées de champ ont été publiés pour l'année 2011. En d'autres termes, c'est en 2014 lors de la publication des données 2012, qu'a été sensible la modification de champ introduite en 2011 ; en revanche, les modifications de champ propres à 2012 n'ont été prises en compte que l'année suivante lors de la publication de 2013 .

Le graphique suivant montre l'évolution quantitative du champ Ésane en cumulant les effets économiques (entre 2008 et 2009 la baisse des effectifs est exclusivement liée à la crise puisque les champs sont identiques) et les effets de champ (en 2010 et en 2011).

[122] Centre de formalités des entreprises : c'est par leur intermédiaire que les unités légales sont créées dans Sirene. L'administration fiscale peut également dans certains cas en créer et constitue donc un CFE particulier.

Tableau : évolution du champ Ésane entre 2008 et 2010 (en nombre d'unités légales)



4.7.2 Campagne 2012 / champ 2013

4.7.2.1 Les modifications apportées au champ

1) Il n'y a eu que des changements très mineurs de champ Ésane lors de la campagne 2012.

Ont été introduites :

- Les sociétés de fait entre personnes physiques (CJ 2210) ; cela concerne environ 13 200 unités légales.
- Les indivisions entre personnes physiques (CJ 2110) : cela concerne 29 unités légales.
- Les sociétés en participation entre personnes physiques (CJ 2310) : cela concerne 24 unités.

Ont été au contraire exclus :

- Les auto-entrepreneurs loueurs de logement (Ape 68.20A) qui ont été exclus du champ.

Ces corrections sont marginales : **elles ne sont pas décelables au niveau des agrégats**.

2) Quelques catégories d'unités ont changé de sous-champ : toutes les SCI ont été basculées en sous-champ 5 ; toutes les activités de location (Ape débutant par 77…) exercées par des personnes physiques ont été basculées en sous-champ 3.

3) La liste des entreprises, immatriculées en France mais ayant leurs activités hors du territoire français et qui sont donc exclues du champ Ésane, a été modifiée, comme chaque année. Il s'agit d'unités codées en exploitation pétrolière (Ape 06.10Z) le plus souvent. Une partie de ces unités était déjà hors champ sur les exercices 2008 à 2011. D'autres sont exclues pour la première fois sur la l'exercice 2012.

4) Enfin les entreprises de Mayotte ont été introduites en 2012 dans Ocsane (mais pas dans les résultats diffusés d'Ésane Cf. § *4.2*) puisque cette entité est devenue un département d'Outre-mer en mars 2011.

4.7.2.2 La comparaison 2012-2011-2010

1) Pour permettre aux comptables nationaux de travailler à champ constant, la division Ésane a produit des fichiers définitifs 2011 et semi-définitifs 2012 dont le contour est proche de celui de la campagne 2010.
Pour obtenir des fichiers 2011 et 2012 proches du champ 2010, les corrections suivantes ont été apportées aux données d'origine :

- les unités entrées dans le champ d'Ésane en 2011 du seul fait du changement de code exploitant ont été retirées des fichiers des données individuelles de 2011 et de 2012 à l'exception de celles qui ont été interrogées par l'Esa ou l'EAP[123] ;
- les sociétés entrées dans le champ d'ÉSANE en 2012 suite à l'introduction de nouvelles catégories juridiques ont également été retirées du fichier 2012.

2) Une nouvelle méthode d'imputation des liasses fiscales a été introduite. Celle-ci n'est pas véritablement un problème de « champ ». Elle modifie toutefois les comparaisons que l'on peut faire entre deux années consécutives. Elle est décrite à la fin du § *6.5.3.*

4.8 Les modifications de la définition de l'entreprise à compter de la campagne 2013

4.8.1 Campagne 2013 / champ 2014

Cette campagne est marquée par la **première prise en compte des groupes dans Ésane,** décrite dans le *chapitre 11*, et la première publication, à l'été 2015, dans laquelle les **entreprises ne sont pas toutes assimilées à des unités légales**.

Les comptes consolidés de 109 entreprises profilées, issues du profilage de 44 groupes, y remplacent dans Ésane ceux (non consolidés par définition) des 4224 filiales qui composent ces 109 entreprises profilées (Cf. *§ 11.6.6*). En termes d'entreprises, on peut considérer que ces nouvelles entreprises respectent les critères qui ont été énoncés pour former le champ de Ésane et ne constituent donc pas une « modification de la définition du champ » ; en revanche, il s'agit d'une modification plus essentielle puisque c'est **la notion même d'unité statistique de base et d'entreprise** qui est modifiée.

Contrairement aux cas précédents en effet, où une différence de champ peut se concentrer sur un sous-champ que l'on ajoute ou retire pour faire une comparaison, le changement « d'unité statistique » est beaucoup plus profond et oblige à tout recalculer, en prenant en compte l'ensemble des entreprises. L'impact de ce changement nécessite une véritable étude de comparaison des agrégats finaux[124].

En termes d'unités légales, cependant, les contours du champ peuvent être légèrement différents : par exemple, telle filiale agricole d'un groupe des IAA peut entrer dans le champ car elle est filiale du groupe alors qu'elle en est exclue en tant qu'unité légale. A l'inverse telle autre peut sortir du champ car elle est exerçait une activité de service dans une partie « bancaire » ou d'assurance d'un groupe ; en tant qu'unité légale elle était donc dans le champ d'Ésane ; mais l'entreprise profilée bancaire va elle sortir du champ d'Ésane, ce qui va la sortir du champ final. Ces différences de contour sont cependant **extrêmement marginales**.

Il n'y a eu **aucune autre introduction ou exclusion d'unités légales** du champ Ésane pour cette campagne 2013.

Conformément à ce qu'on a expliqué au début du § *4.7* et au § *11.6.6*, **les résultats de l'année 2012 ont été recalculés** en prenant en compte ces mêmes entreprises profilées en 2012, de façon à évaluer des évolutions 2013/2012 économiquement significatives, calculées sur un même champ **d'entreprises.**

4.8.2 Campagne 2014 / champ 2015

En termes d'entreprises profilées comme en termes d'unités légales indépendantes, **le champ 2015 est resté identique à celui de 2014.**

4.8.3 Campagne 2015 / champ 2016

Le champ de la cible 1 des très grands groupes (Cf. § *11.4*) continue (et continuera probablement) à évoluer à la marge, au fur et à mesure que des groupes ou des parties de groupes, qui ont été contactés mais n'ont pas pu franchir toutes les étapes des traitements décrits au § *11.6*, rentreront dans le système.

[123] Les unités enquêtées ont été conservées afin de ne pas fausser le plan de sondage.
[124] Voir par exemple « les entreprises en France Édition 2015 » article de J. Deroyon : « De nouvelles données issues du profilage des groupes [...] » : *https://insee.fr/fr/statistiques/1906539*

La campagne 2015 a ainsi vu **le nombre de groupes** de la cible 1 passer de **44 à 51** et celui des entreprises profilées de **109 à 119** (dont on a les comptes en 2014 pour publier à champ constant sur 2015 et 2014).

Pas de modification du champ **en termes d'unités légales**.

4.8.4 Campagne 2016 / champ 2017

Il n'y a **pas d'évolution de champ en termes d'unités légales ni de groupes supplémentaires qui vont rentrer dans la cible 1**
Il y a cependant des évolutions économiques qui se traduisent par l'apparition de trois entreprises profilées nouvelles par rapport à la campagne 2015 et suppression de 6 déjà existantes. Comme pour les unités légales, le tissu des entreprises profilées est soumis aux mêmes types d'évolutions démographiques avec des entreprises **pérennes**, des **nouvelles** et des entreprises qui **disparaissent**. Les contours et l'organisation des groupes évoluent en effet en permanence ; la seule différence est que les restructurations de groupes portent éventuellement sur des transferts d'actifs de groupe à groupe et s'accompagnent aussi de réorganisations internes : tel groupe peut céder une entreprise à un autre groupe, telle partie peut devenir plus autonome ou tel groupe peut achever la consolidation d'une fraction qu'il vient d'acquérir ce qui lui permet de rentrer véritablement dans Ésane etc. Pour la campagne 2016, les 51 groupes correspondent à 116 entreprises profilées dont les comptes consolidés remplacent ceux non consolidés des 4149 unités légales qui correspondent à leurs filiales.

4.8.5 Campagnes ultérieures

La prochaine grosse modification de champ, qui devrait être la dernière d'importance, sera la généralisation du profilage à tous les groupes, ce qu'on a appelé dans le chapitre 11 le traitement des cibles 2 et 3. L'ensemble des groupes hormis ceux de la cible 1, seront alors assimilés à des entreprises. Ce basculement pourrait avoir lieu à l'occasion de la campagne 2017.

* *
*

Il faut retenir en tout cas de ce chapitre que **le champ d'Ésane est susceptible d'évolutions d'une année sur l'autre** pour de multiples raisons, même s'il n'y en pas eu depuis la campagne 2013. Il y a eu en revanche depuis cette date une modification des **unités statistiques** avec la prise en compte des groupes, sous forme d'entreprises profilées et cette modification joue beaucoup sur la vision qu'on peut avoir du système productif[125].
S'agissant de fabriquer des statistiques structurelles, les modifications marginales de champ ne sont en principe pas très gênantes car elles jouent peu sur les principaux ratios sectoriels mais cela peut gêner des utilisateurs particuliers (internes comme externes) qui s'intéressent en fait principalement aux évolutions d'une année sur l'autre. On diffuse alors systématiquement à leur intention des fichiers et des résultats **calculés à champ et méthodologie constants, par maillon de deux années consécutives.** Il reste toutefois délicat de mener des études sur moyenne période.
A l'inverse, le changement d'unités statistiques opéré depuis 2013 a, lui, **modifié l'image du tissu productif** et la répartition sectorielle de nombreuses variables. L'introduction des cibles 2 et 3 pour les campagnes 2 et 3 jouera un rôle semblable même si l'effet devrait être inférieur.

[125] Voir par exemple Insee Première n°1399 de mars 2012 « Un tissu productif plus concentré qu'il ne semblait » : *https://www.insee.fr/fr/statistiques/1281373*

5. Les principes méthodologiques propres aux enquêtes (Esa et EAP)

Cette partie présente les différents choix méthodologiques spécifiques aux enquêtes qui ont été mis en œuvre pour l'Esa d'une part et pour la prise en compte des données de l'EAP de l'autre. Les objectifs de ces enquêtes ont été présentés au chapitre 1 de présentation des sources (Cf. § *1.1.1* et *1.1.2*). Rappelons que l'EAP est également exploitée en elle-même dans le cadre du règlement Prodcom et possède donc son exploitation et ses résultats en propre, publiés par ailleurs dans le cadre du règlement Prodcom. On fera néanmoins une description du champ et du tirage de l'EAP également puisqu'elle fait aussi partie du système Ésane.
On abordera successivement 6 points :

- Présentation de l'Esa : § *5.1* ;
- Présentation de l'EAP : § *5.2* ;
- Prise en compte des données de l'EAP dans le système Ésane : § *5.3* ;
- Les contrôles spécifiques à l'Esa (avec un accent mis sur ceux en lien avec le calcul de l'APE) : § *5.4* ;
- Le traitement de la non-réponse totale (provisoire) dans la base de production : § *5.5* ;
- Le traitement de la non-réponse totale (définitive) pour la diffusion (qui est différent) : § *5.6*.

Le mode de tirage des unités légales et le traitement de la non-réponse **ont été modifiés** assez notablement pour la **campagne 2016** afin de préparer, entre autres, la généralisation de la mise en œuvre de l'unité statistique « entreprise », à l'occasion de la campagne 2017 (phase 2 de Resane ; traitement des cibles dites 2 et 3 des groupes ; voir *chapitre 11* et § *11.4*). En effet, lorsqu'on publiera les résultats de cette campagne 2017, on s'est engagé à pouvoir en comparer les résultats avec ceux de 2016 à champ et méthodologie constants (Cf. § *4.7*) ; il faut donc faire un tirage des unités statistiques « groupes » dès 2016. En termes d'unités légales c'est donc un tirage **à deux degrés** : lorsqu'un groupe (assimilé à une entreprise) est tiré, il faut alors interroger toutes ses filiales pour pouvoir reconstituer la réponse du groupe, ce qui modifie assez notablement le plan de sondage ; au passage, on s'aperçoit qu'on est alors obligé de tirer en même temps les échantillons de l'Esa et de l'EAP puisqu'une entreprise profilée peut avoir à la fois des filiales qui relèvent des deux enquêtes. Ces modifications **ne sont pas décrites ici** mais dans on peut les trouver dans [2,15].

5.1 Présentation de l'Esa

5.1.1 Champ de l'enquête et questionnaire

Il est décrit au chapitre précédent (au § *4.4.3*). On notera que sa composition doit exclusivement à l'histoire de l'EAE et pas du tout à une réflexion autonome sur la pertinence d'enquêter telle ou telle sous-classe de la NAF en totalité ou en partie.

Il y a six grands **« types »** de questionnaires correspondant aux six grands secteurs économiques suivants : Industries Agro-Alimentaires (IAA), Industrie[126], Construction, Commerce, Services et Transports (on retrouve les anciennes EAE). Ces six types se divisent ensuite en « **sous-types** » qui dépendent éventuellement de trois critères (secteur économique précis au sein d'un grand secteur, taille et catégorie juridique de l'entreprise).
Exemple : au sein des questionnaires de *type* Commerce, on rencontrera le *sous-type* « Hypermarché », destiné uniquement à ces entreprises.

Au sein d'un questionnaire, les questions sont regroupées dans des **rubriques** (qui sont donc les briques élémentaires d'un questionnaire) qui composent eux-mêmes des **cadres**. Il existe 104 rubriques possibles et la combinatoire de toutes ces rubriques et cadres forme un ensemble d'environ **108 questionnaires différents** (en 2013).

[126] En fait ce type ne recouvre que les « exploitations forestières et scieries » (NAF 02.20Z), qui, bien qu'appartenant à l'agriculture, s'apparentent en fait à des entreprises industrielles et sont interrogées en Esa. L'industrie (hors IAA) proprement dite est couverte par l'EAP qui sera décrite au paragraphe *5.1* suivant.

On identifie deux grandes sortes de rubriques (et donc de questions) : les rubriques de « **tronc commun** » communes à plusieurs types (généralement les six), et les rubriques « **sectorielles** » qui ne sont utilisées que dans un seul type (voire un seul sous-type).
Principales rubriques du tronc commun : **la ventilation du chiffre d'affaires en branches ou en produits détaillés** ; le cadre « restructurations » ; la mesure de l'emploi. Exemples de rubrique sectorielle : la surface de magasin pour le commerce de détail ; les dépenses particulières (assurance dans le transport, frais de publicité) ; les types de clientèle, etc.

Certaines rubriques sont de plus **personnalisées** c'est-à-dire que le nombre de cases à remplir et leur signification varie selon l'entreprise. L'ensemble des questionnaires est donc personnalisé à la fois en termes d'identification (identifiant de l'entreprise), d'adressage (sur la première page), mais aussi de pré-remplissage de certaines rubriques du questionnaire. C'est le cas de la ventilation du chiffre d'affaires en branche pour les entreprises qui sont interrogées deux années de suite (Cf. § *5.1.4*).

5.1.2 Stratification

La strate de tirage était définie jusqu'à la campagne 2015 incluse[127] par le croisement suivant **: APE x tranche d'effectif x région**. Dans chaque secteur, il existe une strate exhaustive (toutes les entreprises de cette strate ont par définition un poids de 1) et des strates non exhaustives pour lesquelles les entreprises ont donc un poids différent de 1 qui peut aller jusqu'à 3 ou 400 (avant les phases de winsorisation et de calage qu'on verra respectivement aux § *5.6.2.3* et *5.6.2.4*).

Pour la métropole, les taux de sondage ont été définis au niveau APE x tranche d'effectif en distinguant les coopératives agricoles des autres types d'unités légales[128] (à la demande du SSP qui souhaite pouvoir publier des résultats en propre sur la sous population des coopératives agricoles). Pour obtenir le nombre d'entreprises à tirer par région, on applique, pour une APE x tranche d'effectif donné, le même taux de sondage dans toutes les régions.
Pour les DOM, les taux de sondage de la Guyane, Guadeloupe et Martinique sont les mêmes, contrairement à ceux de la Réunion.

5.1.3 Pondération et allocation de l'échantillon *[2.5]*

Les taux de sondage de l'Esa 2008 ont été définis en se basant sur les objectifs de taille d'échantillon de la maîtrise d'ouvrage de Resane. Leur méthodologie de calcul est évoquée en détail ci-dessous. Ces taux sont ensuite reconduits chaque année et corrigés à la marge afin d'obtenir une taille d'échantillon acceptable. Les secteurs du spectacle vivant sont entrés dans le champ de l'Esa 2010. Pour ces trois sous-classes, les taux de sondage sont basés sur une allocation de Neyman portant sur la variable chiffre d'affaires. Ils sont également reconduits chaque année.

5.1.3.1 Détermination des taux de sondage initiaux

Les **taux de sondage initiaux** (ceux de l'Esa 2008) ont été déterminés par simulation selon la procédure suivante.

Les simulations ont été effectuées à partir des dernières données définitives issues de l'EAE et des fichiers administratifs, connues au moment où ont été faites ces simulations (2006 voire 2005 pour certains secteurs qui n'étaient pas interrogés tous les ans). Elles portaient sur tous les secteurs de l'Esa, à l'exception des Industries agricoles et alimentaires et de l'Industrie, secteurs non couverts par les EAE pour les moins de 20 salariés. **Trois variables ont été étudiées** : le chiffre d'affaires sectoriel, l'investissement et le chiffre d'affaires branche de la branche principale.

La démarche retenue était la suivante :

- pour chaque variable d'intérêt, la précision de l'estimateur utilisé dans l'EAE a été calculée ;
- les données d'enquête ont ensuite été « calées » sur les données administratives selon une méthode similaire à celle mise en œuvre dans Ésane (Cf. § *5.6.2.4*) ;

[127] Comme on l'a signalé au début du chapitre, le plan de sondage a été complètement refondu à compter de la campagne 2016 et la région (dont les limites ont été redéfinies suite au découpage de 2014) n'est plus une variable de stratification.
[128] Le 10ème caractère de la strate de tirage vaut « C » pour les coopératives agricoles et « A » pour les autres types d'entreprises.

- enfin, deux types d'estimateurs après calage ont été calculés pour chaque variable, d'une part l'estimateur pondéré classique et d'autre part **l'estimateur « composite** » approprié[129], et leur précision évaluée.

La comparaison des précisions des estimations pré et post calage a ainsi permis, au prix d'une hypothèse simplificatrice détaillée ci-après, de déterminer l'impact potentiel d'une réduction de la taille de l'échantillon sur la précision des estimations.

Les hypothèses et options méthodologiques retenues dans le cadre de ces simulations étaient les suivantes :

- Les simulations ont **été menées en NAF Rév.1, alors que le plan de sondage de l'Esa est défini en NAF Rév.2**. Toutefois, il est vraisemblable que les résultats obtenus en termes de gains de précision associés au calage demeurent valables en nouvelle nomenclature.

- En outre, **les simulations ont été effectuées en utilisant la stratification et les seuils d'exhaustivité alors en vigueur dans l'EAE 2006**. La question de l'optimisation des strates de tirage et de la modification des seuils d'exhaustivité qui en découle n'a pas été abordée dans cette étude.

- **Le calage**, qui concerne la partie échantillonnée de l'EAE (donc hors partie exhaustive), **a été réalisé sur le seul chiffre d'affaires sectoriel tel que connu via la source administrative, avec application au niveau groupe**[130] (trois premiers caractères de la NAF). Ainsi, les poids $\widetilde{w}_i$ obtenus près calage respectent la condition suivante pour tous les groupes X (*APErep* désignant le code APE connu dans le répertoire) :

$$\sum_s \widetilde{w}_i \, CA_{adm}(i) \, 1_{APErep=X}(i) = \sum_U CA_{adm}(i) \, 1_{APErep=X}(i)$$

  Le calage a été effectué à l'aide de modèles logit bornés, permettant de contenir la dispersion des rapports de poids (poids après calage / poids avant calage) entre les bornes présentées dans le tableau suivant :

Secteur	Borne inférieure	Borne supérieure
Construction	0,85	1,50
Commerce de gros	0,75	2,50
Commerce de détail	0,65	6,00
Services aux entreprises	0,40	5,00
Services aux particuliers	0,85	1,50
Transport	0,85	2,00

  Remarque : depuis, les procédures de calage ont été améliorées, et le calage s'effectue dorénavant à la fois **sur le chiffre d'affaire et le nombre d'unités légales par groupe**[131] (modulo quelques regroupements) en limitant, dans la très grande majorité des groupes, la dispersion maximale des rapports de poids à un facteur deux (Cf. § *5.6.2.4*).

- Enfin, afin de déterminer l'impact potentiel d'une réduction de la taille de l'échantillon sur la précision des estimations, nous avons posé **l'hypothèse simplificatrice d'une variance des estimations inversement proportionnelle à la taille de l'échantillon,** une réduction de moitié de l'échantillon (hors partie exhaustive) conduisant ainsi à un doublement de la variance des estimations.

[129] Cet estimateur composite, caractéristique d'Ésane, sera présenté en détail au § *9.2*. Il s'agit ici d'un estimateur « par différence » pour les variables sectorielles CA et investissement, et d'un estimateur « par ventilation » pour le CA de la branche.
[130] Un calage au niveau NAF4 a également été testé, mais conduisait à des rapports de poids beaucoup trop dispersés. En conséquence, l'idée de caler à un niveau aussi fin a été abandonnée.
[131] Le classement sectoriel étant toujours celui issu du répertoire.

Au final, les résultats de ces estimations ont montré que **le calage sur données fiscales et l'utilisation des estimateurs composites** (Cf. § 9.2) mis en œuvre dans Ésane **permettraient de réduire de moitié la taille de l'échantillon, hors exhaustif, sans perte significative de précision pour la majorité des variables, aux niveaux groupe**[132] (de la nomenclature) **et supérieurs**.

Appliquée à l'ensemble des secteurs de l'Esa – hors IAA –, la stratégie de réduction de moitié de l'échantillon hors exhaustif conduit aux volumes d'échantillon suivants :

Secteur	Situation passée (EAE 2008/2007)				Echantillon ESA proposé - stratégie « demi échantillon h exhaustif »		
	Echantillon total (1)+(2)	Exhaustif (1)	Hors exhaustif (2)	Taux de sondage	1/2 échantillon hors (3)	Nouvel échantillon total (3)+(1)	Taux de sondage
Construction	19819	10639	9180	5,00%	4590	15229	3,84%
Commerce de	24160	7173	16987	14,23%	8494	15667	9,23%
Commerce de	38281	9209	29072	7,13%	14536	23745	4,42%
Services aux	33523	10428	23095	7,02%	11548	21976	4,60%
Services aux	34410	6115	28295	7,09%	14148	20263	4,17%
Transport	13833	6697	7136	17,29%	3568	10265	12,83%
Total hors DOM, Industrie et IAA	164026	50261	113765	7,64%	56884	107145	4,99%

On aboutissait donc à un échantillon total, hors DOM, Industrie et IAA, d'environ 107 000 entreprises. En prenant également en compte l'exhaustif des IAA – 4022 entreprises – ainsi que la part de l'échantillon de l'EPEI (Enquête sur les Petites Entreprises Industrielles, réalisée du temps de l'EAE sur les moins de 20 salariés) relevant des IAA[133] et non de l'Industrie – estimée à 12% en se basant sur la répartition en NES36 de l'échantillon EPEI 2001 (le dernier enquêté), soit 2193 entreprises –, on obtenait un échantillon total, hors DOM et Industrie, d'environ **113 000 entreprises**, soit une diminution de l'ordre de 35% pour l'échantillon total, par rapport à l'EAE 2008/2007. **Cette réduction importante de la charge statistique pesant sur les entreprises a permis de *supprimer la stratégie dite « petites entreprises »** alors en vigueur dans l'EAE**[134] et qui conduisait à des évolutions heurtées et peu réalistes dans les séries.

Enfin, les simulations menées dans cette étude ont également permis de mettre en évidence un déséquilibre au niveau de la précision des estimations entre les différents grands secteurs de l'EAE : en particulier, les estimations du secteur des « Transports » se révélaient systématiquement très précises, à l'inverse de celles relatives au secteur des « Services aux particuliers ». La mise en place du système Ésane a été l'occasion de ***revoir la répartition des allocations**** entre les différents grands secteurs, afin d'opérer un **rééquilibrage** au niveau de la précision des estimations. Cette réallocation a été effectuée sous les hypothèses suivantes :

- on a supposé ici encore que la précision des estimations était inversement proportionnelle à la taille de l'échantillon ;
- la réallocation a été effectuée au niveau des échantillons hors exhaustif de l'EAE, de façon à uniformiser au maximum la précision des estimations du chiffre d'affaires sectoriel entre les différents grands secteurs ;
- quel que soit le secteur considéré, on s'est interdit de réduire de plus de 50% la taille de l'échantillon hors exhaustif EAE initial.

Et l'on a ensuite procédé à la réduction de moitié des échantillons EAE hors exhaustif ainsi « équilibrés », ce qui nous a conduit aux échantillons Esa « équilibrés » suivants

[132] Des simulations ont également été menées au niveau classe NAF (4 caractères de la NAF) pour les variables sectorielles chiffre d'affaires et investissement. Elles conduisaient à des résultats similaires, les estimations composites post-calage avec échantillon réduit de moitié se révélant plus précises dans 99% des cas.

[133] Rappelons qu'à l'inverse du reste de l'industrie, interrogée en EAP, les entreprises des IAA sont désormais interrogées en ESA. Mais comme les petites entreprises des IAA n'étaient pas interrogées auparavant en EAE, on ne pouvait pas faire un calcul identique aux autres secteurs pour les IAA.

[134] Elle consistait à n'enquêter les petites entreprises (celles n'appartenant pas aux strates exhaustives) qu'une année sur deux alternativement, dans les secteurs du commerce de détail puis de gros ainsi que dans les services aux entreprises puis aux ménages. L'année de « non interrogation », l'estimation des réponses des entreprises était faite à partir de leurs liasses fiscales.

			Chiffre d'affaires sectoriel		Investissement		CA branche principale	
Secteur	**Echantillon ESA initial**	**Echantillon ESA « équilibré »**	**CV_{ESA} initial**	**CV_{ESA} « équilibré »**	**CV_{ESA} initial**	**CV_{ESA} « équilibré »**	**CV_{ESA} initial**	**CV_{ESA} « équilibré »**
Construction	15229	12934	0,003 %	0,003 %	0,014 %	0,016 %	0,018 %	0,022 %
Commerce de gros	15667	15667	0,011 %	0,011 %	0,122 %	0,122 %	0,020 %	0,020 %
Commerce de détail	23745	18900	0,012 %	0,015 %	0,014 %	0,017 %	0,021 %	0,026 %
Services aux entreprises	21976	26821	0,011 %	0,009 %	0,203 %	0,167 %	0,051 %	0,042 %
Services aux particuliers	20263	24342	0,009 %	0,008 %	0,066 %	0,055 %	0,046%	0,038 %
Transport	10265	8481	0,001 %	0,002 %	0,000 %	0,000 %	0,005 %	0,006 %

Rappel : CV, coefficient de variation, est égal au rapport de l'écart type de l'estimateur de la variable sur sa moyenne. Il caractérise en quelque sorte la précision d'un estimateur.

L'opération de rééquilibrage de la précision des estimations du chiffre d'affaires sectoriel opérée sur les échantillons EAE s'est donc traduite sur les échantillons Esa par un léger rééquilibrage des précisions entre les différents grands secteurs. On constate que seule la taille de l'échantillon du commerce de gros est restée inchangée ; 3 secteurs ont diminué : construction, commerce de détail et transports, au profit des services aux entreprises et aux particuliers.

5.1.3.2 Détermination des seuils d'exhaustivité

On a repris les seuils d'exhaustivité tels qu'ils existaient dans les EAE. Pour une sous-classe NAF donnée, ceux-ci sont fonction de deux variables : le chiffre d'affaires et l'effectif salarié. Font partie de la strate exhaustive toutes les entreprises qui sont au-dessus d'au moins un des deux seuils.

Les seuils d'exhaustivité sont définis comme ci-après (en règle générale ils se situent à 20 ou à 30 salariés, mais dans le détail c'est beaucoup plus varié) :

Service enquêteur	APEN	CA minimum (en k€)	Effectif salarié minimum
Commerce	Coopératives agricoles	38 000	10
	Classes 10.13B, 10.71C, 10.71D, 45.20A, 45.20B, 46.31Z, 46.34Z, 46.61Z, 46.69B, 46.73A, 47.11D	38 000	50
	Autres secteurs	38 000	20
Construction	Tous les secteurs	15 000	30
IAA	Coopératives agricoles	5 000	10
	Tous les secteurs sauf coopératives agricoles	5 000	20
Services	Classe 55.20Z	1 000	6
	Classe 55.30Z	1 000	10
	Groupes 411, 681, 683, classe 70.10Z, divisions 77, 79	8 000	30
	Division 61	5 000	6
	Groupe 682	15 000	50
	Classe 78.20Z	5 000	200
	Groupe 812	5 000	100
	Classes 90.01Z, 90.02Z, 90.04Z	1 000	30
	Autres secteurs	5 000	30
Transports	Tous les secteurs	3 000	20
DOM	*Tous les secteurs (dont industrie)*	*800*	*10*

De plus, afin de mieux suivre les restructurations, les entreprises appartenant au champ de l'Esa et **détectées en restructuration** au cours de l'année de référence (date d'effet comptable) dans Citrus sont mises de force **dans l'exhaustif**, même si leurs caractéristiques (APE, effectif) ne les y mettent point. Ces entreprises sont repérables à partir de la modalité de leur strate de tirage : « FORC EXH ».

Étant donné le nombre important de strates, **les arrondis** effectués pour déterminer les allocations dans chaque strate sont **contrôlés**, en particulier sur les marges APE et tranche d'effectif. Vu le grand nombre de strates, on aboutit pour certaines d'entre elles à une allocation nulle. Lorsque cela est le cas, des **regroupements de strates** sont effectués afin d'obtenir des allocations non nulles :

- Pour la métropole (y compris coopératives agricoles) on regroupe les strates en commençant par les régions d'une même Zeat, puis si nécessaire les Zeat et des tranches d'effectifs.
- Pour chaque DOM, on regroupe en commençant par la tranche d'effectif, puis la sous-classe, la classe et le groupe d'activité de l'entreprise.

Les strates regroupées sont repérables par la modalité « G » à la place de la variable de regroupement dans le nom de la strate. Malgré cela, des unités (de l'ordre de 1 000 chaque année, situées majoritairement dans les DOM) conservent une probabilité de tirage nulle. **Le nombre final de strates est de l'ordre de 8 000 (de l'ordre de 25 000 avant les regroupements).**

5.1.4 Renouvellement par moitié de l'échantillon :

Le renouvellement par moitié des unités de l'échantillon est ***une des innovations d'Ésane****. Il ne concerne par définition que la partie non exhaustive de l'échantillon : une entreprise dont les caractéristiques la placent dans la partie exhaustive y reste si ses caractéristiques la maintiennent dans la partie exhaustive. Il a pour objectif de permettre de **calculer des résultats en évolution N/N-1** tout en **limitant la charge de réponse des petites entreprises**. Le renouvellement consiste :

- d'une part à **conserver** l'une des deux parties **de la base de sondage** (dite partie conservée) de l'année N-1, privée des unités qui ne sont plus dans le champ de l'enquête en N, ainsi que la partie correspondante de l'échantillon (dite partie conservée de l'échantillon) ;
- et d'autre part à **tirer un nouvel échantillon dans une base de tirage** constituée des unités légales de l'autre partie de la base de sondage (dite partie renouvelée) de l'année N-1, également privée des unités qui ne sont plus dans le champ de l'enquête en N, auxquelles sont ajoutées les unités nouvelles, c'est-à-dire entrées dans le champ de l'enquête en N. On attribue à ces unités nouvelles un numéro hexal permanent qui peut être pair ou impair.

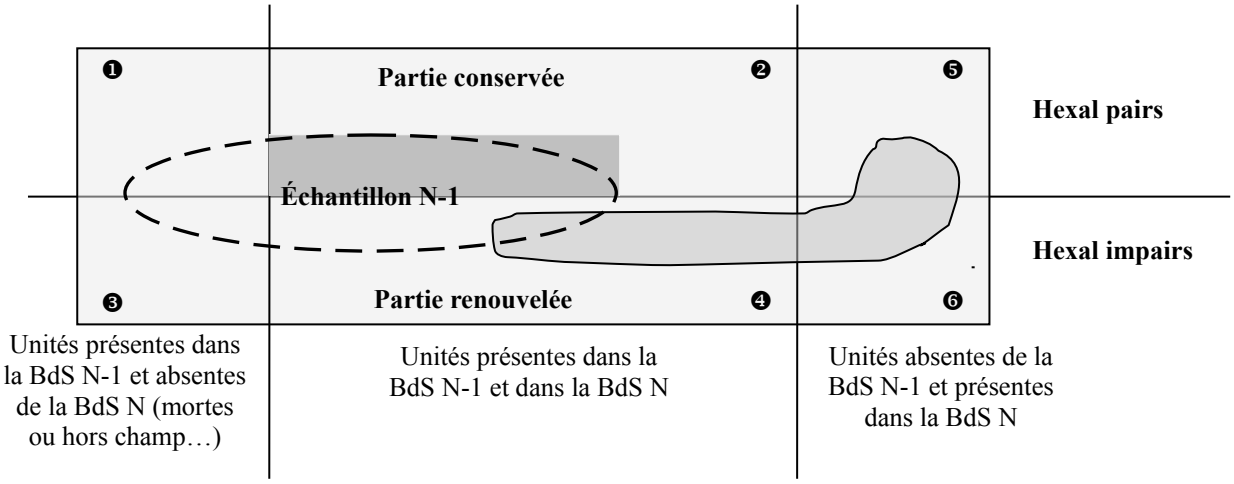
Le renouvellement se fait par le biais du **numéro hexal** (numéro entier compris entre 1 et 60, affecté aléatoirement à chaque unité de la base de sondage dans Ocsane lors de sa première apparition). L'unité conserve donc ce numéro aléatoire à vie[135]. La base de sondage est alors scindée en 2 (hexal pair, hexal impair). Pour l'Esa, on conserve les entreprises dont le numéro hexal est de même parité que l'année de campagne.

Par exemple, si l'année de campagne est impaire (campagne 2011 traitée en 2012 par exemple), on va conserver (pour l'ESA 2012) les entreprises de numéro hexal impair et le tirage des entreprises entrantes ne sera effectué que sur les entreprises de n° hexal pair et les entrantes de n° hexal pair ou impair de la base de sondage. Notons par ailleurs que les entreprises qui sont conservées, conservent aussi leur poids de tirage d'origine (même si leurs nouvelles caractéristiques les placeraient dans une nouvelle strate).

Aucune coordination négative n'est effectuée sur la partie renouvelée de l'échantillon car la base de sondage est mise à jour avec les données de l'enquête (il y a en effet dans ce cas un risque de biais). En d'autres termes, on effectue **un tirage avec remise** ce qui fait qu'une entreprise qui vient d'être interrogée deux années de suite peut très bien être interrogée pour deux années supplémentaires. Chaque année, environ 10% (soit de l'ordre de **3000 unités légales**)[136] des unités de la partie renouvelée non exhaustive de l'échantillon étaient déjà présentes dans l'échantillon l'année précédente.

[135] On n'utilise ici que la parité du numéro hexal. Le choix 1 ou 2 aurait donc pu suffire. Mais ce numéro, multiple également de 3, 4 et 5 a été choisi pour pouvoir tiré au besoin des échantillons coordonnés et renouvelés par tiers, quarts, cinquièmes ou sixièmes, sans changer de numéro.
[136] Ce nombre élevé vient du fait que pour certaines strates, les taux de sondage dépassent les 50%.

Le schéma suivant récapitule la situation, dans le cas d'un renouvellement concernant les numéros hexal impairs : le grand rectangle gris clair figure la base Ocsane ; la partie grise foncée l'échantillon N, l'ellipse, l'échantillon N-1 ; l'exhaustif n'est pas représenté ; dans la partie renouvelée, l'intersection de l'ellipse N-1 avec l'échantillon N représente les 10% de la partie renouvelée qui sont à nouveau tirés l'année suivante :



⇒ Le tirage effectivement réalisé l'année N n'a concerné que les unités légales de numéro hexal impair présentes à la fois dans la base de sondage de l'année N-1 et de l'année N – partie ❹ –, ainsi que les unités légales nouvelles, c'est-à-dire absentes de la base de sondage en N-1, que leur numéro hexal soit pair ou impair – parties ❺ & ❻ –. Les unités effectivement tirées en N correspondent donc à la « banane grise ». L'intersection de cette banane avec l'ellipse figurant l'échantillon N-1 correspond aux unités qui sont interrogées quatre ans de suite en fait ; en revanche leurs poids sont ceux de ce nouveau tirage.

Poids à utiliser

En termes de poids, la situation est la suivante :

- Les unités appartenant à la partie renouvelée de l'échantillon en N (la « banane grise » sur le schéma) ont été effectivement tirées en N, avec une stratification utilisant des informations relatives à l'année N et selon des poids w_k^N correspondant aux poids de tirage en N → la partie renouvelée de l'échantillon en N, pondérée par les poids w_k^N est extrapolable à l'univers ❹ + ❺ + ❻, i.e. pour une variable Y quelconque :

$$E\left[\sum_{i \in \text{échantillon renouvelé}} w_i^N Y_i\right] = \sum_{i \in 4 \oplus 5 \oplus 6} Y_i$$

- Les unités appartenant à la partie conservée de l'échantillon en N (la partie grisée de l'échantillon N-1 sur le schéma) ont été effectivement tirées en N-1, avec une stratification utilisant des informations relatives à l'années N-1 et selon des poids w_k^{N-1} correspondant aux poids de tirage en N-1 → la partie conservée de l'échantillon en N, pondérée par les poids w_k^{N-1} est extrapolable à l'univers ❷, i.e. pour une variable Y quelconque :

$$E\left[\sum_{i \in \text{échantillon conservé}} w_i^{N-1} Y_i\right] = \sum_{i \in 2} Y_i$$

Par conséquent, l'échantillon complet de l'année N, pondéré par les poids

$$w_k = \left| \begin{array}{l} w_k^N = \text{poids de tirage en N si l'entreprise } i \in \text{échantillon renouvelé} \\ w_k^{N-1} = \text{poids de tirage en N-1 si l'entreprise } i \in \text{échantillon conservé} \end{array} \right.$$

est extrapolable à l'univers ❷ + ❹ + ❺ + ❻ qui correspond bien au champ de l'enquête en N (plus précisément au champ de l'enquête en N au lancement ≡ base de sondage en N).

⇒ Pour résumer, les poids à utiliser sont les poids « historiques », i.e. relatifs à la date de tirage effectif des unités. En particulier, les unités légales qui sont conservées gardent leurs caractéristiques et en particulier leur poids de tirage de l'année N-1 (et ce même si leurs caractéristiques ont évoluées depuis et les placent en N dans une « strate actualisée » différente de leur strate de tirage N-1 initiale)

Contrairement aux échantillons stratifiés habituels (c'est-à-dire renouvelés entièrement chaque année), la somme des poids de tirage de l'ensemble de l'échantillon d'une année N ne coïncide pas avec la taille de la base de sondage N.
Dans la partie renouvelée de l'échantillon, la somme des poids de tirage correspond bien à la taille N_R de la partie renouvelée de la base de sondage. En revanche, dans la partie conservée de l'échantillon, la somme des poids de tirage n'est qu'une estimation de la taille N_C de la partie conservée de la base de sondage. En effet, les sorties de champ ont pu toucher les unités de l'échantillon différemment des unités qui ne sont pas dans l'échantillon. Par conséquent, la somme des poids de tirage de l'ensemble de l'échantillon d'une année N ne coïncide pas avec la taille de la base de sondage N mais c'est bien un estimateur sans biais de N. L'écart entre cet estimateur et N pourra être corrigé au moment du calage.

Plus formellement, pour la variable indicatrice C valant 1 si i est dans la base de sondage l'année N et 0 sinon, on a les propriétés suivantes :

$$\sum_{i \in \text{échantillon renouvelé}} w_i^N C_i = \sum_{i \in 4 \oplus 5 \oplus 6} C_i = N_R \text{ et } E\left[\sum_{i \in \text{échantillon conservé}} w_i^{N-1} C_i\right] = \sum_{i \in 2} C_i = N_C .$$

5.2 *Présentation de l'EAP* [1.3]

L'EAP est gérée indépendamment du processus de production Ésane par le service statistique national des entreprises (SSNE) de la direction régionale de Caen, qui a pris la suite du Sessi[137] lorsque celui-ci a été rattaché à l'Insee.
Le tirage de l'échantillon et les principes de contrôle et redressements des réponses sont restés différents de ceux de l'Esa de 2008 à 2013. Ils sont décrits succinctement ci-dessous.
Les données de l'EAP sont donc validées en dehors de la chaîne Ésane, pour la publication de Prodcom[138]. Toutefois, afin de pouvoir diffuser des résultats sur l'ensemble du champ des activités des entreprises françaises (y compris l'industrie), il est nécessaire de **récupérer la ventilation en branches des entreprises industrielles** collectée via l'EAP. La prise en compte des données de l'EAP dans le système Ésane est décrite au § *5.3*. Elle bénéficie d'une **publication à part** sur le site de l'Insee, du fait de l'existence du règlement Prodcom, pour la publication de la production industrielle française à un niveau très détaillé (Cf. § *9.4.6*).

5.2.1 Le champ de l'enquête

Le champ « **théorique** » de l'enquête EAP devrait être celui des entreprises marchandes et exploitantes (Cf. la note de bas de page du *§ 1.2.1.1* pour la définition de ce terme) ayant au moins un jour d'activité dans l'année d'interrogation, avec leur siège social en France métropolitaine et ayant pour APE dans le référentiel Ésane une activité industrielle classée dans les sections suivantes :

B - Industries extractives
C - Industrie manufacturière (sauf les secteurs des IAA qui sont dans le champ de l'ESA : division 10 : industries alimentaires, division 11 : fabrication de boissons, division 12 : fabrication de produits à base de tabac, sauf les sous-classes 1610A et 1610B sciage, rabotage et imprégnation du bois)
D - Production et distribution d'électricité, de gaz, de vapeur et d'air conditionné
E - Production et distribution d'eau ; assainissement, gestion des déchets et dépollution

[137] Nom du service statistique du ministère de l'industrie, puis de l'économie, des finances et de l'industrie, qui était responsable de l'EAE dans l'industrie puis de l'EAP à sa création, ce qui explique la différence de méthodologie initiale avec l'Esa.
[138] Voir § *5.2.3*.

En 2010[139], ce champ EAP « **théorique** » comportait **162 859 unités**. La ***nouveauté*** par rapport à l'ancienne EAE industrielle consiste en ***l'interrogation des petites entreprises industrielles****[140].

Toutefois, en pratique, compte tenu de leur très faible poids, par mesure de simplification, le Sessi qui avait alors la responsabilité de l'enquête, **a écarté** du champ de l'enquête, en amont du tirage de l'échantillon, les entreprises **sans salarié** avec un chiffre d'affaires inférieur à 5 M€ ainsi que les entreprises de moins de 20 salariés avec un chiffre d'affaires inférieur à 5 M€ appartenant à des NAF comportant 0 ou un seul « produit industriel » (défini par la nomenclature fine de l'enquête et appelé génériquement ProdEap). Cet ensemble d'unités, **dénommé « EAP0[141] »,** représentait **88 315 unités en 2010**. Au final, le champ **réellement couvert** par l'enquête EAP comportait donc uniquement en 2010, **74 544 unités** (162 859 moins 88 315).

Comme on le voit, même si cet ensemble pesait très peu globalement en termes de CA ou de VA, il était loin d'être négligeable en nombre d'unités légales et ne l'était pas du tout dans certaines sous-classes particulières. Cette exception aux principes de tirage compliquait le calcul final des agrégats (car ces unités font évidemment partie des sous champs 2, 3 ou 4 d'Ésane définis au chapitre précédent et cela scindait donc toutes les sous-classes industrielles en deux). C'est pourquoi on a décidé **d'aligner le plan de sondage de l'EAP sur celui de l'Esa à compter de la campagne 2014.**

Par ailleurs, l'enquête est complétée, pour répondre au règlement **ProdCom**, par l'interrogation exhaustive de quelques centaines **d'entreprises non industrielles** mais ayant une production industrielle significative comme activité « secondaire » (cette population est appelée **EABIS** par la suite). Dans le cadre de l'EABIS 2010, **1 061 entreprises** ont ainsi été interrogées. En outre, ces entreprises sont interrogées en parallèle dans l'Esa. Il a donc été décidé **de ne pas prendre en compte**, dans le système de production Ésane, la réponse à l'EAP[142] de ces unités mais uniquement leur réponse à l'ESA.

5.2.2 Le plan de sondage jusqu'en 2013

Jusqu'à l'année d'observation 2013, le plan de sondage s'apparente, en théorie du moins, à un plan de sondage **stratifié à un seul degré avec sondage à probabilités égales au sein de chaque strate**. Le critère de stratification retenu est le secteur d'activité de l'entreprise (code APE) au niveau **5 caractères de la NAF**[143].

1. **La strate exhaustive** qui comptait **26 809 unités en 2010** (hors EABIS), appelée « EAP1 », est constituée comme suit :
   - ✓ les entreprises industrielles de 20 salariés et plus **ou** ayant un chiffre d'affaires supérieur à 5M€ sont enquêtées exhaustivement (comme c'était le cas auparavant dans l'EAE) ;
   - ✓ en complément, dans le cas où il existe plus d'un produit (ProdEap) dans l'activité considérée, il est procédé à une vérification du taux de couverture en chiffre d'affaires par secteur. Pour les secteurs dont le taux de couverture est inférieur à 85%, la sélection est complétée[144], secteur par secteur, par les entreprises les plus significatives afin d'atteindre cette couverture de 85% par la strate exhaustive.
2. À cette strate exhaustive de l'EAP, **s'ajoute donc l'EABIS**, ensemble exhaustif d'entreprises non industrielles réalisant dans au moins une branche industrielle un montant de chiffre d'affaires significatif. La détermination de ces entreprises non industrielles fortement contributrices aux résultats par branches se fait **sur la base des seuils[145] de chiffre d'affaires définis pour l'EAP1 par secteur**. Ainsi, pour une NAF S industrielle donnée,

[139] On donne ici des évaluations relatives à 2010, mais les ordres de grandeur sont évidemment les mêmes tous les ans.

[140] Dans l'ancien système, leur suivi était assuré par une enquête ad-hoc, effectuée par l'Insee, intitulée EPEI (Enquête sur les Petites Entreprises Industrielles), réalisée en moyenne tous les trois ans.

[141] Dans le paragraphe 9.2 de ce document, on verra qu'un traitement particulier est réalisé au moment de la diffusion des résultats pour prendre en compte cette non couverture du champ complet de l'industrie par l'EAP.

[142] Pour ces entreprises, le SSNE ne dispose que du chiffre d'affaires des branches industrielles. Seule l'ESA donne la ventilation du chiffre d'affaires complet. Les deux services se coordonnent toutefois en fin d'exploitation pour vérifier que les réponses fournies sont cohérentes (et se retournent vers l'entreprise dans les cas où elles ne le seraient pas).

[143] Les critères « chiffre d'affaires » et « effectifs salariés » interviennent également pour la définition de la strate exhaustive.

[144] I.e. le seuil d'exhaustivité en termes de chiffre d'affaires, initialement fixé à 5 M€, est abaissé de façon à assurer une couverture, en termes de chiffre d'affaires total du secteur, supérieure ou égale à 85%.

[145] Pour mémoire, ce seuil X_S vaut, pour un secteur S donné :

- 5 M€ si l'ensemble des entreprises du secteur ayant un CA ≥ 5 M€ représente plus de 85% du CA total du secteur, ou si le secteur comporte 0 ou 1 ProdEap ;

toutes les entreprises non industrielles, présentant un chiffre d'affaire dans la branche S supérieur au seuil sectoriel X_S, sont retenues.

3. **La partie échantillonnée** de l'EAP, **appelée EAP2**, constitue ***une nouveauté par rapport à l'ancienne EAE***[146] qui n'interrogeait que les entreprises appartenant au champ EAP1 et résulte de la sélection, au sein de chaque secteur d'activité (défini par le code APE au niveau 5 caractères de la NAF), d'un sixième[147] des unités du « champ EAP2 »[148], selon un schéma de tirage rotatif visant à assurer l'interrogation de l'ensemble du « champ EAP2 » en six campagnes.

Ce schéma, complexe en pratique et surtout différent de celui de l'ESA, a été abandonné, comme on l'a vu, au bout d'un cycle et le tirage de l'EAP a été aligné sur celui de l'Esa à compter de la campagne 2014. Il **subsiste** bien entendu **une strate exhaustive EAP1** et **une strate échantillonnée EAP2** mais le tirage de cette strate échantillonnée se fait sur le champ complet de l'EAP cette fois-ci (**donc en abandonnant le principe de l'EAP0**) selon les principes exposés plus haut concernant l'ESA et non plus selon un principe de tirage rotatif[149].
En d'autres termes, à compter de la campagne 2014, le **champ de l'EAP correspond bien à son champ théorique**.

5.2.3 Le questionnaire

Le questionnaire de l'EAP contient, jusqu'à la campagne 2011, quasi exclusivement des informations sur la répartition du chiffre d'affaires de l'entreprise en « produits » avec donc une nomenclature d'environ 4000 produits (les ProdEap), propre à l'enquête et élaborée à partir de la nomenclature du règlement ProdCom[150], ainsi pour beaucoup d'entre eux qu'une évaluation de la production en quantités physiques. Il n'y a aucune autre question « dite sectorielle » comme il en existe dans l'Esa. À compter de la campagne 2012, des questions sur la mesure des investissements ont été rajoutées et il y en a d'autres sur la mesure de l'emploi (distinction salarié et non salarié à partir de la campagne 2013)
Les principaux cadres du questionnaire concernent :

- la vente de produits industriels y compris les services industriels ;
- la production totale (en quantité) ;
- les ventes au détail ;
- l'installation et la pose des produits industriels ;
- la réparation et la maintenance ;
- les ventes de produits et services non industriels ne relevant pas des industries manufacturières ou extractives ;
- des cadres spécifiques selon l'APE.

Comme pour l'Esa, **chaque entreprise a un questionnaire « sur mesure »** avec uniquement les produits qui la concernent.

5.2.4 Imputation et estimation des données absentes

La procédure d'estimation retenue dans le dispositif EAP s'appuie sur une méthode d'imputation de masse. La base de données de calcul du dispositif EAP, dénommée BEC (base élémentaire de calcul), contient ainsi **l'ensemble des entreprises** industrielles actives **du champ théorique** de l'EAP (y.c. l'EAP0 donc), ainsi que les unités de l'EABIS. Pour chaque unité, sont stockées dans la BEC à la fois les informations **relatives aux produits et celles relatives aux branches** :

- X M€, avec X déterminé de façon à ce que l'ensemble des entreprises du secteur ayant un CA ≥ X M€ représente plus de 85% du CA total du secteur.

[146] Avant Ésane, les petites entreprises industrielles n'étaient interrogées que tous les deux ou trois ans au travers d'une enquête spécifique appelée EPEI (enquête sur les petites entreprises industrielles). Cette enquête a été supprimée à cette occasion puisqu'elles font désormais partie du champ de l'EAP.
[147] Depuis l'EAP 2010. Pour les EAP 2008 et 2009, le taux de sondage de l'EAP2 était de 1/5.
[148] Ce champ EAP2 correspond donc au champ de l'EAP hors EAP0, privé de l'exhaustif EAP1 et de l'EABIS.
[149] Qui ne permettait pas en pratique de respecter une équi-probabilité de tirage de 1/6 pour les entreprises du champ EAP2, à cause de la nécessité d'interrogation des entreprises « créées ».
[150] Le règlement européen « ProdCom » définit **chaque année** la liste détaillée des produits dont les États membres doivent envoyer la production évaluée en valeur et en quantités physiques. Le nombre de produits suivis varie donc un peu chaque année, généralement à la baisse.

- ✓ pour une entreprise donnée, sont ainsi présents dans la BEC, si on en dispose, **l'ensemble des produits** ayant été recensés pour cette entreprise au cours des différentes campagnes de collecte de l'EAP ou lors de la dernière enquête annuelle de branche de 2007.

- ✓ **en ce qui concerne les branches**, chaque entreprise se voit associer la liste connue des branches la concernant. Cette liste, qui peut être réduite à une seule branche, est soit issue de la collecte liée à une campagne précédente, quel que soit le millésime, soit constituée d'une unique branche, celle correspondante à son APE, pour les entreprises n'ayant jamais été enquêtées ou répondantes à une première campagne EAP. Toutes les unités possèdent ainsi au moins une branche associée (**au prix d'une hypothèse sur le caractère monobranche** des unités pour lesquelles on ne dispose d'aucune information[151]).

Dès l'initialisation de la campagne, les données individuelles de l'année N sont **imputées pour l'ensemble des unités actives** ou présumées actives de la BEC, selon une procédure d'imputation s'appuyant sur les données N-1 qui sera décrite ci-dessous. En cours d'enquête, les données collectées et validées viennent se substituer aux imputations pour les unités répondantes à l'EAP. Ainsi, en fin de campagne, les données dans la BEC sont, soit issues de l'enquête, soit estimées pour les entreprises non enquêtées et les non répondantes. Le calcul des agrégats s'effectue alors par simple sommation (puisque la BEC est exhaustive). À noter que les agrégats branches calculés à partir des données branches de la BEC ne sont pas publiés, et servent juste en interne à des fins d'aide à l'analyse pour les gestionnaires.

Les données relatives aux unités non enquêtées ou non répondantes sont imputées selon la procédure suivante :

❶ les unités *présumées cessées* selon la source TVA[152] voient leurs données (produits comme branches) mises à blanc, et ne participeront donc pas aux estimations ;

❷ les unités *présumées actives* et pour lesquelles des données, déclarées ou imputées, *sont disponibles*[153] dans la BEC N-1, voient leurs données estimées à partir des données N-1, soit actualisées par le coefficient d'évolution temporelle $CA_{TVA}^{N}(i)/CA_{TVA}^{N-1}(i)$ pour les unités pour lesquelles les sources TVA et EAP sont jugées comparables, soit reconduites à l'identique pour les autres unités ;

❸ les unités *présumées actives* de la BEC N pour lesquelles les données *ne sont pas présentes* dans la BEC N-1[154], pour les estimations branches uniquement, et pour lesquelles on dispose de données fiscales N-1 sont automatiquement considérées comme monobranches, et se voient attribuer comme chiffre d'affaires N le montant de production de biens disponible dans les liasses fiscales N-1, éventuellement actualisé par le coefficient d'évolution temporelle $CA_{TVA}^{N}(i)/CA_{TVA}^{N-1}(i)$ pour les unités pour lesquelles la source TVA est disponible.

Ainsi, à l'issue de cette phase, les unités non répondantes à l'EAP, sans données disponibles dans la BEC en N-1 et pour lesquelles on ne dispose pas de données fiscales en N-1[155] **se retrouvent sans aucune information imputée**. Par ailleurs, même présentes dans la BEC, les entreprises de l'EAP0[156] n'ont pas d'information sur les produits.

En pratique, **ce défaut d'imputation n'est guère préjudiciable** à la qualité et à la diffusion des résultats dans le cadre du dispositif actuel de l'EAP. En effet, le dispositif de l'EAP répond à un double objectif :

[151] On a vu plus haut qu'il ne pouvait s'agir que d'entreprises qui viennent d'être créées ou qui appartiennent à l'EAP2 donc qui sont petites.
[152] Les fichiers issus des déclarations mensuelles de TVA nous sont fournis par la DGFiP. Ils ne constituent pas véritablement une « source » d'Ésane mais sont utilisés dans cette phase d'estimation des données de l'EAP. Lorsqu'une entreprise est absente des fichiers de TVA, elle est présumée cessée.
[153] Il s'agit donc, pour les données branches, des unités présumées actives en N et en N-1 hors créations et entrées de champ, et pour les données produits, des mêmes unités restreintes à celles possédant au moins un produit associé.
[154] Les entreprises en entrée de champ soit à cause d'un changement d'APE soit à cause d'un accroissement du nombre de salariés.
[155] Cela concerne donc essentiellement les unités créées en N.
[156] Rappelons qu'il s'agit des unités légales du champ de l'EAP en termes d'APE mais sans effectif salarié.

- d'une part produire des données fines sur la production industrielle, principalement pour répondre aux exigences du règlement ProdCom. Or, le règlement ProdCom impose un taux de couverture minimal de 90 % de la production nationale en termes de chiffre d'affaires par sous-classe d'activité en NAF Rév.2, et ne vise donc pas l'exhaustivité. Ainsi, l'absence de données déclarées ou imputées pour certaines unités dans le dispositif EAP, qui concerne très essentiellement de petites unités, n'a que peu d'impact sur la qualité des résultats fournis au titre de la réponse au règlement ProdCom ;
- d'autre part repérer les différentes activités exercées par les entreprises, via la ventilation de leur chiffre d'affaires en branches, et en déduire leur activité principale exercée (APE).

5.3 La prise en compte des données de l'EAP dans le système de production Ésane

5.3.1 La prise en compte de la ventilation en branche de l'EAP

Les branches issues de l'EAP sont envoyées au système de production Ésane. Ces données sont considérées comme validées et ne rentrent donc **pas dans la chaîne de contrôle** et de redressement automatique de l'Esa. En revanche, ces données sont prises en compte pour le calcul des agrégats et **subissent des macro-contrôles sur les branches non-industrielles**.
Si une entreprise est à contrôler selon les règles des macro-contrôles et qu'un rappel de l'entreprise semble nécessaire, **ce rappel sera réalisé à Caen** par un gestionnaire en charge de l'EAP (hors Ésane), afin d'éviter que l'entreprise ne soit interpelée par des gestionnaires auxquels elle n'a généralement pas affaire.

5.3.2 La prise en compte des changements d'APE

Un calcul de l'APE[157] a lieu à partir de la réponse en branche de l'EAP, dans la base de production de l'EAP (les principes du calcul de l'APE, communs à tous les secteurs, sont rappelés dans l'*annexe 1* de cet ouvrage). Une fois l'APE validée, elle est envoyée à Ocsane. Ce principe d'échange est le même que celui qui existe entre l'Esa et Ocsane.

5.3.3 La diffusion des résultats des entreprises industrielles

Les liasses fiscales des entreprises industrielles sont introduites dans le système de production Ésane comme toute autre liasse fiscale. Elles sont validées par la division « industrie et IAA » du service Ésane de Nantes.

La **réconciliation** des données de l'EAP et des liasses fiscales se déroule selon la **même procédure** que la réconciliation entre les données de l'Esa et de la liasse (cf. *chapitre 8* ci-dessous). Toutefois à ce stade si l'entreprise doit être rappelée, elle le sera par le gestionnaire en charge de l'EAP.

Enfin, l'absence (que l'on vient de voir) de données imputées pour certaines unités non-répondantes de l'échantillon EAP ne pose quasiment pas problème[158]. En effet, pour la partie échantillonnée de l'échantillon Esa⊕EAP, le système Ésane traite la non-réponse par repondération, et **n'utilise donc pas les données imputées par le dispositif de l'EAP**. Quant à la partie exhaustive, pour laquelle la non-réponse est gérée par imputation dans l'Esa et où l'on veut donc avoir des données imputées pour l'ensemble des non répondantes de l'exhaustif Esa⊕EAP, en particulier sur la ventilation en branches, l'absence de données imputées pour certaines non répondantes de l'exhaustif EAP est traité par le processus Rédi puisqu'il impute automatiquement des « branches Rédi » à ces unités (cf. chapitre *8* ci-après pour plus de détails).

Au final, la **diffusion** des données des entreprises industrielles se fait donc **de la même façon** que celles des entreprises des autres secteurs.

[157] Activité Principale Exercée. On a indiqué au tout début de l'ouvrage (Cf. § *1.1.1*) que l'un des principaux objectifs de ces enquêtes est la mise à jour (éventuelle) du code APE des entreprises interrogées.
[158] Pour les résultats définitifs en tout cas. Pour les résultats provisoires envoyés à Eurostat fin octobre, c'est un peu plus problématique car le traitement de la non-réponse se base alors sur les données imputées et non sur la repondération. Mais vu le poids des EAP0 ce n'est pas dramatique au niveau des résultats provisoires.

5.4 Les contrôles gestionnaires spécifiques de l'Esa

En plus des micro et macro contrôles qui ont déjà été décrits dans le chapitre 2, les contrôles supplémentaires faits par les gestionnaires sur l'Esa visent principalement à :

- obtenir une réponse de l'entreprise (Cf. § 5.4.1) ;
- compléter la réponse initiale faite par l'entreprise (Cf. § 5.4.2) ;
- valider l'APE de l'entreprise Cf. § 5.4.3) ;
- repérer un maximum de restructurations Cf. § 5.4.4).

5.4.1 Le rappel des grandes non-répondantes

Le service Ésane a défini une **liste de Très Grandes Entreprises** (TGE) qui sont contrôlées par des gestionnaires **spécialisés**. Ainsi, par exemple, c'est le nom du gestionnaire qui figure sur le questionnaire Esa de ces entreprises afin de faciliter les relations avec l'entreprise. De plus, tous les traitements susceptibles d'être réalisés sur cette entreprise tout au long de la campagne, seront automatiquement confiés au même gestionnaire spécialisé, chargé des TGE.

La liste des TGE est **établie chaque année** par la division Ésane en fonction des taux de couverture minimum par groupes de la NAF (3 positions).

- Est considérée comme TGE une entreprise dont le chiffre d'affaires représente **au moins 1%** du total de son groupe Naf et **au moins 100 millions d'euros**. Une « très très grande entreprise » (**TTGE**) est une TGE dont le chiffre d'affaires dépasse **1 milliard d'euros**.

  ⇒ Les quelque 1500 TGE représentent à elles seules environ 40% du total du champ Ésane en chiffre d'affaires.

- Parmi les TGE, environ 70 TTGE, soit environ 10 par division sectorielle feront l'objet d'un suivi **par le chef de division sectorielle** lui-même. Un dossier personnalisé pour chaque TTGE permettra de répondre à toute demande d'information de la part de la maîtrise d'ouvrage ou du DCN, dans un délai très court (1 journée).

  ⇒ Sur **2008**, les 220 TTGE représentent à elles seules 800 milliards de CA soit un peu plus du quart du total du champ Ésane.

Nombre	Nb TGE	CA TGE k€	Nb TTGE	CA TGE k€
Total Champ Ésane	**1 430**	**1 202 487 531**	**221**	**804 679 444**
Commerce	160	308 452 948	88	277 409 968
Construction	63	18 453 205	1	1 187 893
Industrie IAA	832	594 708 443	92	358 282 529
Services aux entreprises	222	154 617 872	20	83 654 918
Services aux particuliers	97	45 462 638	7	18 809 705
Transports	56	80 792 425	13	65 334 431

Ces entreprises ont donc un poids économique tel dans les agrégats diffusés qu'il n'est pas possible, pour ces unités, d'avoir recours à un traitement de la non-réponse totale traditionnel (décrit ci-dessous). Ces unités ont en effet un comportement atypique qu'il est quasiment impossible de modéliser. C'est pourquoi, en cas d'absence de questionnaire, il est demandé pour ces unités aux gestionnaires spécialisés de les **rappeler** afin d'obtenir impérativement une réponse.
Dans le cas où l'entreprise refuse de répondre, le gestionnaire doit faire son possible pour essayer d'obtenir des ordres de grandeur et ce notamment pour la ventilation du chiffre d'affaires en branches. De ces estimations dépend en effet la qualité globale des agrégats des branches concernées.

5.4.2 La codification des branches non codées

Dans le questionnaire de l'enquête figure un cadre dans lequel il est demandé aux entreprises de répartir leur chiffre d'affaires par activité (plus ou moins fines).

Cette partie du questionnaire est **personnalisée selon l'activité principale exercée (APE)** par l'entreprise, connue au lancement de l'enquête. Dans les secteurs du commerce et des services, cette personnalisation est identique pour l'ensemble des unités d'une même APE. En revanche, pour les secteurs du BTP, du transport et des IAA, cette personnalisation contient des activités par défaut mais également la liste des activités pour lesquelles l'entreprise avait, le cas échéant, renseigné un montant l'année précédente.
Enfin, cette partie du questionnaire contient également des **zones de « texte libre** » dans lesquelles l'entreprise **peut décrire une activité qu'elle ne trouve pas** dans sa liste pré-imprimée et y associer un montant.
Dans ce cas, au retour du questionnaire, il est nécessaire de procéder à une **codification** du libellé en clair saisi afin d'obtenir une ventilation complète du chiffre d'affaires de l'entreprise.

Ce cadre du questionnaire est le plus important de ce dernier. Il permet en effet de calculer :

- l'APE sur laquelle reposent les statistiques sectorielles ;
- les agrégats en branches, informations nécessaires au DCN pour le calcul du PIB.

Ainsi, lorsqu'un questionnaire comporte une activité non codée, ayant un montant jugé significatif, il est demandé au gestionnaire de coder le libellé saisi par l'entreprise. Pour ce faire, il est parfois nécessaire de **rappeler l'entreprise pour avoir des précisions** et choisir le bon code de la NAF correspondant. Sinon, on opère une codification automatique fruste (Cf. § 5.4.2.1).

Enfin, il arrive également que l'entreprise ne réponde pas du tout à cette question. Dans ce cas, il est également demandé au gestionnaire de recontacter l'unité afin d'obtenir sa ventilation du chiffre d'affaires en branche si on considère que l'unité a un poids important[159]. Sinon, on procède à un redressement automatique de cette non-réponse partielle (Cf. *5.4.2.2*).

5.4.2.1 Une codification automatique pour les cas à faible enjeu

Afin de réduire le nombre de dossiers à traiter par les gestionnaires, il a été décidé de réaliser une codification automatique des libellés **dans le cas où les montants en jeu étaient jugés faibles ou négligeables**. Cette codification est pilotée par le service Ésane de Nantes en fonction des ressources disponibles pour chaque division sectorielle.
Le principe de cette codification est le suivant :

Sont considérées comme **pouvant être codées automatiquement** toutes activités telles que :

- le montant pondéré associé reste inférieur à un seuil strict et paramétrable[160] ;
- le montant pondéré associé reste inférieur un seuil moins strict[161] mais le poids de la branche dans le chiffre d'affaires de l'entreprise dépasse les 30%[162].

Le codage se fait ensuite de la façon suivante

- la ventilation du chiffre d'affaires de l'Esa est répartie en plusieurs cadres sur le questionnaire[163] ;
- pour **chaque cadre du questionnaire et chaque APE**, le service MOA concerné[164] fournit **une activité par défaut**. C'est cette activité qui viendra prendre la place du libellé saisi et ce indépendamment du contenu du libellé. Dans le cas où cette activité existe déjà dans la réponse de l'entreprise, le montant associé au libellé est cumulé au montant déjà existant pour l'activité. Il s'agit donc d'une codification **assez systématique**. Mais elle est supposée être du second ordre vu les montants en jeu.

[159] Les critères de définition de cette notion d'importance sont décrits ci-dessous.
[160] Pour la campagne 2010, par exemple, il était fixé à 15 000 K€ pour tous les secteurs sauf pour le commerce de détail où il était de 27 000 K€ et le commerce de gros où il valait 50 000 K€.
[161] Pour la campagne 2010 il était fixé à 10 000 K€ pour tous les secteurs sauf pour le commerce de gros où il valait 30 000 K€.
[162] Ce seuil est également paramétrable mais est pour le moment constant pour chaque division.
[163] En général il y a 3 cadres l'un relatif aux activités de services, un second dédié aux activités de commerce et un dernier aux activités productives de biens.
[164] Département des Synthèses Sectorielles (pour commerce et services), SSP (pour les IAA), SOeS (pour transport et BTP).

5.4.2.2 Le redressement automatique des non-réponses partielles

Il est demandé aux gestionnaires de **rappeler toutes** les unités pour lesquelles le **chiffre d'affaires pondéré est supérieur à un seuil**. Ce seuil, paramétrable, est différent selon les divisions sectorielles du service Ésane de Nantes. En 2010 ce seuil était de 15 000 K€ pour tous les secteurs sauf pour le commerce de détail où il était de 27 000 K€ et le commerce de gros où il valait 50 000 K€.

Pour les unités pour lesquelles le chiffre d'affaires pondéré est inférieur au seuil, la ventilation manquante **est estimée** par tendance auxiliaire, c'est-à-dire maintien de sa structure antérieure ou par moyenne de strate si l'entreprise n'a pas répondu à l'Esa en N-1 (c'est-à-dire qu'on applique la répartition moyenne de la strate à laquelle elle appartient à son chiffre d'affaires total).

5.4.3 La validation de l'APE

La mise à jour de l'Activité Principale Exercée (APE) est **l'un des produits phare de l'Esa** dans la continuité de ce qui était fait auparavant dans les EAE. Il s'agit **en théorie** de l'activité **la plus génératrice de valeur ajoutée**. En pratique, comme la valeur ajoutée est très difficile à mesurer par activité, c'est **le chiffre d'affaires par activité** qui est utilisé comme **proxy**. L'APE est donc calculée à partir de la ventilation du chiffre d'affaires par activité (rappelons que les principes du calcul de l'APE sont exposés dans l'*annexe 1* de ce document).

Ce code permet de classer correctement les entreprises dans leur activité majoritaire, activité qui est à l'origine du calcul des statistiques sectorielles. Ce classement sectoriel, issu de la réponse de l'entreprise à la ventilation du chiffre d'affaires par activité, est une **originalité française**. Dans les autres pays européens, ce classement est le plus souvent purement **déclaratif**.

La chaîne de production Ésane **gère plusieurs types d'APE** dont les descriptions sont faites ci-dessous. Et il est demandé au gestionnaire **un travail spécifique de contrôle de cette variable** qui est cruciale. L'APE « statistique » calculée par le biais de l'Esa a d'ailleurs pris encore plus de poids dans le cadre d'Ésane que par le passé, à cause de l'estimateur composite qui est utilisé pour le calcul des agrégats (cf. chapitre *9*). En effet, les changements d'APE repérés sur l'échantillon Esa **sont ensuite « extrapolés** », comme on le verra dans ce chapitre 9, à l'ensemble de la population.

5.4.3.1 Les différentes APE présentes dans la chaîne de production Ésane

La chaîne de calcul traite de sept variables se référant à l'APE selon les moments du traitement et les résultats à la fois des traitements et du dialogue avec l'entreprise. Cela ne signifie évidemment pas que ces sept APE sont différentes. On verra en conclusion toutefois **qu'il peut y en avoir 3 différentes pour une même unité répondante au cours d'une campagne**. A contrario bien entendu, l'APE individuelle d'une des quelques millions d'entreprises figurant dans la base Ésane ne peut être modifiée que si l'entreprise est répondante à l'Esa ou à l'EAP.

1. **APE en sortie d'Ésane baptisée simplement « APE » pour l'exercice N** :

C'est celle qui figure dans Ocsane et elle est déterminée dans l'ordre suivant par :

- L'APE forcée par un gestionnaire si elle existe.
- Sinon, l'APE calculée stabilisée de l'Esa (et confirmée par le gestionnaire).
- Sinon, l'APE mise à jour dans Ocsane (suite à un sous-processus PSCD, PSCP ou Rédi emploi).
- Sinon, l'APE au lancement.

C'est cette variable qui est **transmise au répertoire statistique** Sirus (Cf. § *11.7*) à chaque mise à jour et qui est **ensuite envoyée à SIRENE**. A compter de la campagne 2015, dans quelques cas exceptionnels, cette APE « statistique » du répertoire Sirus peut être différente de celle de Sirene (Voir 4. Ci-dessous).

2. **APE calculée (APE_C)** :

Elle est calculée **à partir des ventes branches** (et quelques variables auxiliaires) selon un algorithme spécifique assez complexe en réalité dont les grandes lignes sont indiquées en *annexe* 1.

- Elle est recalculée à chaque fois que l'on met en œuvre un CQR qui risque de modifier les variables intervenant dans le calcul de l'APE (telles que chiffre d'affaires ou sa répartition en branches).

- Cette variable est stockée dans la base de gestion de l'Esa et n'existe que pour les entreprises répondantes à l'Esa. Elle est très rarement utilisée pour remplacer l'APE calculée stabilisée si cette dernière ne peut être calculée.

3. **APE calculée stabilisée (APE_C_S)**

Le but de cette APE est d'éviter de faire le yoyo entre deux activités de poids équivalents d'une année sur l'autre.

- Pour ce faire, on ajoute fictivement 20% de chiffre d'affaires à la branche correspondant à l'APE de départ dans l'algorithme de calcul. Pour que l'APE bascule dans une autre « branche » il faut donc que la nouvelle soit véritablement plus importante que l'ancienne.
- Il s'agit de l'APE « officielle » de l'Esa. C'est cette dernière **qui va mettre à jour** la variable APE décrite ci-dessus. Elle est stockée dans la base de gestion de l'Esa et n'existe que pour les entreprises répondantes à l'Esa et reflète la ventilation en branches de l'entreprise.

4. **APE forcée :**

Lorsque le calcul de l'APE calculée stabilisée amène à un résultat différent de l'APE de lancement, l'entreprise devrait donc changer d'APE et les gestionnaires ont pour consigne de **vérifier qu'il n'y a pas eu d'erreur de déclaration** de l'entreprise à l'enquête (voir paragraphe *5.4.3.2* suivant). Il s'ensuit un dialogue entre le gestionnaire et l'entreprise. Quelles qu'en soient les raisons, jusqu'à la campagne 2014 incluse, s'ils se mettent d'accord sur le fait que l'APE calculée stabilisée n'est pas la bonne et qu'il faut donc lui en attribuer une autre (généralement celle qu'il avait au départ[165]), le gestionnaire a **la possibilité de « forcer » une APE différente** de celle fournie par le calcul.

A compter de la campagne 2015, **on a supprimé cette notion d'APE forcée**. On a bien sûr maintenu la phase de vérification que l'entreprise n'avait pas commis d'erreur de déclaration. Mais s'il s'avère, après cette vérification, que l'APE calculée stabilisée est différente de l'APE au lancement, l'APE calculée stabilisée devient automatiquement l'APE « statistique » de l'entreprise, utilisée pour calculer les agrégats d'Esane. C'est également elle qui sera utilisée l'année suivante comme APE au lancement de la campagne. Elle est stockée dans Sirus qui la transmet au service Sirene lequel le notifie à l'entreprise[166].

5. **APE au lancement (LANC_APE) :**

Elle est connue dans Ocsane au moment du lancement de la campagne N et sert à stratifier l'échantillon de l'Esa. C'est l'APE du répertoire statistique **au moment de la constitution du référentiel** Ésane en novembre N.

6. **APE de diffusion (APE_DIFF)** :

C'est l'APE qui sert **pour les calculs des agrégats sectoriels définitifs**. Elle est calculée et stockée dans la base de diffusion uniquement. **C'est l'APE du FARE**[167]. Elle diffère de l'APE, car, on utilise pour la diffusion l'APE calculée stabilisée plutôt que l'APE forcée afin d'assurer la cohérence entre ventilation en branches du chiffre d'affaires et APE. Elle est définie de la façon suivante :

- C'est l'APE calculée stabilisée pour les entreprises répondantes à l'enquête ;
- Sinon c'est l'APE au lancement pour les entreprises du sous-champ 1 au lancement.

7. **APE de référence (APE_REP[168])** :

C'est une APE « intermédiaire de calcul » qui sert au calcul des agrégats avec la méthode de l'estimateur composite qui sera explicitée au § *9.2*, en liaison avec le fait que des entreprises peuvent être tirées dans l'échantillon deux années consécutives, comme on l'a vu au § *5.1.4*, et comporter par conséquent deux APE au lancement qu'il faut savoir distinguer.

---

[165] Changer de code APE peut être lourd pour une entreprise du simple fait que le code APE figure sur tous les papiers à en-têtes de l'entreprise (courriers, factures) dont elle peut avoir des stocks, ainsi que sur des annuaires professionnels, sur son site Internet éventuel, etc. Changer de code implique de mettre à jour tout cela.

[166] Dans certains cas très exceptionnels, si l'entreprise réagit très négativement à la notification de sa nouvelle APE et justifie le maintien de l'ancienne, il se peut alors que des entreprises aient des APE dans Sirene et dans Sirus différentes.

[167] Fichier approché des résultats d'Ésane : il s'agit du fichier contenant les valeurs individuelles des variables du système, diffusé en interne de l'Insee ainsi qu'aux chercheurs qui en obtiennent le droit via le comité du secret et le Centre d'accès sécurisé. Il s'appelle « approché » car il ne permet pas de calculer les agrégats diffusés par simple addition des valeurs des variables. Il faut pour cela disposer d'un programme spécifique (appelé Kit de diffusion Ésane, présenté au chapitre *9.7*) qui reconstitue les formules des estimateurs composites.

[168] « REP » car c'est celle du répertoire OCSANE lorsque l'entreprise est échantillonnée.

L'APE de référence va permettre de bien mesurer et extrapoler à la population totale les changements d'APE, mesurés donc par rapport à cette APE_REP. Elle concerne les entreprises du sous champ 1 enquêté et est obtenue de la manière suivante :

- Pour les entreprises de la partie nouvelle de l'échantillon l'APE_REP est égale à l'APE au lancement en N
- Pour les entreprises de la partie conservée de l'échantillon, elle est égale à l'APE au lancement en N-1. Ceci est dû au fait qu'il ne faut pas utiliser dans la formule de l'estimateur composite une information qui a été mise à jour par l'enquête elle-même, **sans quoi on introduit un biais**.

Cette APE ne sert que dans la formule de calcul des agrégats composites. **Elle n'intervient pas dans les caractéristiques** d'une unité légale.

Au final, jusqu'à la campagne 2014 incluse, une unité légale est décrite par :

- **son APE au lancement** (soit A) : elle détermine en particulier le questionnaire que reçoit l'entreprise ;
- le calcul de **son APE calculée stabilisée** qui peut bien entendu être différente (soit B) : c'est son APE « statistique », sur laquelle sont fondés les agrégats et les fichiers diffusés ;
- **son APE forcée** s'il y a eu dialogue entre un gestionnaire Ésane et l'entreprise (soit C)[169]. C'est l'APE « administrative » de l'entreprise, qui est remontée dans Sirene si elle diffère de A.

Selon les usages dont on a besoin (participation à un agrégat, à un contrôle, à un fichier de diffusion, envoi à Sirene) on utilisera tantôt A, tantôt B, tantôt C.

A compter de la campagne 2015, une unité légale est décrite par :

- **Son APE au lancement** (soit A) : il s'agit de son APE statistique à l'issue de la campagne précédente ; elle détermine en particulier le questionnaire que reçoit l'entreprise ; elle figure dans le répertoire statistique **Sirus** ;
- **Son APE calculée stabilisée,** calculée à l'issue de l'enquête, qui peut bien entendu être différente de A (soit B) et qui devient son **APE statistique** à l'issue de l'enquête. Celle-ci **met à jour Sirus** et sert au calcul des agrégats ; elle est transmise aux gestionnaires de Sirene.
- Son APE figurant dans le répertoire Sirene ; elle est désormais sans influence sur le déroulement d'Ésane.

5.4.3.2 Le travail de contrôle de l'APE demandé au gestionnaire

L'APE est au cœur du système de diffusion des statistiques structurelles d'entreprises puisqu'elle définit les secteurs ; il est nécessaire de contrôler les « changements » d'APE afin de s'assurer de leur crédibilité. C'est pourquoi il est nécessaire que les gestionnaires **valident les changements d'APE** en allant jusqu'à **rappeler l'entreprise** pour qu'elle confirme sa déclaration.

À l'origine donc, tout changement d'APE se doit d'être contrôlé et ce quelle que soit la taille de l'entreprise. Si les gestionnaires doivent toujours « contrôler » l'ensemble des changements d'APE[170], quelle que soit la taille de l'entreprise, il leur est aujourd'hui seulement demandé de **s'assurer de la validité de la ventilation et de sa cohérence avec la nouvelle APE**, sans forcément avoir recours au rappel systématique des entreprises. Cette décision a été prise :

- d'une part pour limiter la charge de traitement des changements d'APE pour le service Ésane de Nantes ;
- d'autre part parce que les entreprises ont toujours la possibilité de contacter Sirene pour contester ce changement d'APE le cas échéant. L'expérience montre que les contestations sont souvent dues à des problèmes administratifs (changement de convention collective notamment), ce qui ne remet pas en cause la qualité de la réponse à l'Esa (et ne devrait d'ailleurs pas justifier le non-changement d'APE).

Il est en pratique impossible de contacter toutes les entreprises en validation d'APE mais par principe on ne change pas l'APE d'une très grande entreprise sans la contacter, c'est également souvent le

[169] Un autre cas, possible mais improbable, serait d'avoir une nouvelle APE « D » mise à jour dans SIRENE après contact avec l'unité (après envoi de l'avis de situation passant l'entreprise du secteur A à C). D serait alors bien l'APE de lancement Ésane pour N+1.

[170] L'ordre de grandeur c'est 10 000 par an.

cas pour les grandes entreprises. Par ailleurs, l'activité principale déclarative de l'entreprise est collectée dans un cadre spécifique du questionnaire (saisie d'un libellé en clair), mais elle n'est pas prise en compte directement dans le calcul de l'APE. Elle est juste consultée par le gestionnaire, ce qui facilite la compréhension de l'activité de l'entreprise et peut permettre de valider ou non un changement d'APE.

5.4.4 La validation du cadre de restructuration

Comme vu précédemment, le questionnaire Esa comporte trois cadres spécifiques sur les restructurations (cf. § 3.3.2.1). Des contrôles spécifiques de ces cadres permettent d'en assurer la cohérence globale.
En cas d'incohérence, il est demandé au gestionnaire de rappeler l'entreprise afin de récupérer le plus d'information possible sur les restructurations.

Ces contrôles ont toutefois été simplifiés au cours du temps car les premiers mois de collecte ont montré que certaines entreprises avaient tendance (en lisant sans doute trop rapidement le questionnaire) à confondre « investissement par apport » avec « investissement total » ou encore à mettre leur date de création comme date de restructuration, même si leur création n'était pas issue d'une restructuration, etc. Au final, plus que la recherche d'une cohérence complète entre les différents cadres, c'est la description des partenaires de la restructuration qui est privilégiée (et donc indirectement la création des enveloppes de restructuration qu'on a vues au chapitre 3).

Le gestionnaire **doit rappeler** l'entreprise uniquement **dans les cas d'incohérence suivants** :

- l'entreprise déclare un « type » de restructuration mais ne décrit pas de partenaires[171] ;
- l'entreprise ne donne pas toutes les informations sur le partenaire (oubli du Siren ou de la date de la restructuration ou du montant des apports) ;
- l'entreprise déclare des investissements par apport supérieurs à 1 500 K€ mais ne décrit aucun partenaire ;
- l'entreprise déclare des cessions par apport supérieurs à 1 500 K€ mais ne décrit aucun partenaire.

5.5 *Le traitement de la non-réponse totale (provisoire) pour la base de production*

Afin d'obtenir des agrégats de production proches des agrégats définitifs pour le calcul des contributions des macro-contrôles, on met en œuvre une extrapolation des non-réponses totales **hebdomadairement** en base de production, **en attendant que toutes les réponses arrivent** (celles ci s'échelonnent en tout sur plus de 6 mois).

La méthode utilisée est un **traitement de la non-réponse totale par imputation** des réponses manquantes (et non par repondération). La méthode d'estimation par repondération, utilisée pour redresser les non-réponses définitives (et qui sera présentée au § 5.6 suivant) est en effet plus efficace mais plus complexe à mettre en œuvre et nécessite le recours à des experts méthodologues, ce qui n'aurait pas de sens dans le cadre d'une procédure à utiliser hebdomadairement.

Dans le cadre de l'Esa, seules les entreprises de l'échantillon en non-réponse totale (ou quasi-totale[172]) sont extrapolées :

- Si l'entreprise a répondu en N − 1, on privilégiera une extrapolation par les données N − 1 (Cf. § 5.5.1) (en supposant toutefois que ces données N − 1 sont de bonne qualité, *i.e.* ne proviennent pas déjà d'une extrapolation). En outre, les grandes entreprises seront forcément extrapolées par N-1 quel que soit le statut de leur réponse en N-1.
- Sinon on procèdera à une extrapolation par hot-deck (Cf. § 5.5.2). Les strates d'extrapolation sont obtenues par le croisement de l'APE (de lancement) et des tranches de taille (effectif de lancement si on ne dispose pas de l'effectif pour l'année N).

Enfin, dès que la réponse d'une entreprise arrive, elle vient bien entendu « écraser », le cas échéant, les données qui lui avaient été attribuées par l'imputation.

[171] A l'inverse une entreprise qui décrit des partenaires mais pas le type de restructuration n'est pas rappelée.
[172] Les entreprises ne déclarant ni chiffre d'affaires, ni ventilation, ni effectif, bien qu'ayant renvoyé un questionnaire non-vierge, sont considérées en non-réponse quasi-totale.

5.5.1 Le principe de l'extrapolation à partir des données N−1

Dans un premier temps, le chiffre d'affaires de l'entreprise est estimé pour l'année N à partir de sa valeur au lancement (qui correspond peu ou prou à sa valeur en N-1) et d'un taux de croissance « médian » calculé sur les entreprises répondantes de la même strate au lancement (APE au niveau groupe, croisée avec une tranche de taille). Ensuite, la chaîne des CQR est lancée sur le questionnaire dont toutes les variables sont manquantes, sauf le chiffre d'affaires. Les valeurs des caractéristiques sont donc naturellement redressées automatiquement une par une, en tant que non-réponses partielles, par les techniques habituelles présentées au § *2.4* (qui s'appuient sur les données N-1 si elles existent). Le CQR est donc identique à celui décrit plus haut au chapitre 2, à ceci près qu'aucun message de contrôle des données n'est produit.

Une telle méthode a plusieurs avantages. Elle permet d'assurer une cohérence entre redressements de non-réponses partielles et redressements de non-réponses totales. Et elle n'est pas si éloignée que cela d'une méthode consistant à appliquer partout un taux de croissance médian : le redressement par tendance auxiliaire est largement employé, sauf si la donnée N − 1 n'est pas bonne.

Pour les très grandes entreprises (celle de l'exhaustif) une extrapolation par N-1 est forcée. Le chiffre d'affaires est estimé comme vu ci-dessus et si l'unité était non-répondante en N-1, les autres caractéristiques du questionnaire seront estimées par moyenne de strate à partir du chiffre d'affaires estimé.
Enfin, une fois que le processus Rédi est mis en œuvre, c'est le chiffre d'affaires issu de Rédi qui sert de référence.

5.5.2 Le principe de l'extrapolation par hot-deck

Pour les unités qui ne satisfont pas les contraintes de l'extrapolation par N-1, une extrapolation par hot-deck stratifié[173] est mise en œuvre. Pour chaque strate d'extrapolation, on constitue un ensemble d'entreprises, dites « extrapolantes », qui sont non atypiques et jugées de qualité suffisante en N. Pour chaque entreprise en non-réponse **on sélectionne, au hasard**, une entreprise extrapolante de sa strate. Une entreprise pourra servir d'extrapolante au maximum 3 fois.
Dans le cas où il n'y a plus d'extrapolante disponible dans sa strate, on en cherche une dans une strate plus « large » en cherchant une entreprise de même tranche de taille et du même groupe d'APE au lancement, etc.

À noter que pour le calcul des agrégats (de production toujours et pour le calcul des macro-contrôles), l'APE utilisée dans ce cas est bien l'APE de l'entreprise extrapolante et non celle de l'entreprise extrapolée.

5.6 *Le traitement de la non-réponse totale (définitive) pour la diffusion*

Pour la **diffusion des résultats**, il a été décidé de mettre en œuvre une méthode de traitement de la non-réponse (qui est cette fois-ci une non-réponse définitive) différente de celle mise en œuvre en production, qui ne visait que la réalisation des macro contrôles. Elle est présentée ci-dessous.

5.6.1 Le cas particulier des DOM

Initialement, la correction de la non-réponse totale par repondération et le calage devaient concerner l'ensemble de l'enquête Esa⊕EAP, DOM inclus : l'Esa DOM devait donc bénéficier des mêmes traitements que ceux effectivement mis en œuvre pour la métropole. Cependant, les délais de traitements des données de l'Esa DOM, pour les premières campagnes, se sont révélés en pratique incompatibles avec le calendrier de publication des résultats définitifs « France entière » d'une campagne Ésane standard. Aussi a-t-il été décidé dans ce contexte de n'effectuer ces opérations statistiques que sur l'enquête Esa⊕EAP métropole. Les statistiques sectorielles relatives à la partie « DOM » du champ Ésane sont calculées, pour les publications « France entière » selon une procédure **d'estimation simplifiée**, basée exclusivement sur les données fiscales et le découpage sectoriel du répertoire, et similaire à celle mise en œuvre sur les sous champs Ésane 2 à 4.

[173] Les strates sont définies en croisant l'APE au lancement de l'entreprise et sa tranche d'effectif

Parallèlement à ce traitement particulier des DOM dans le cadre des statistiques Ésane « France entière », l'Esa DOM fait également l'objet d'une exploitation spécifique centrée sur les seuls DOM. À cette occasion, des opérations de correction de la non-réponse totale par repondération et de calage **sur le champ des seuls DOM** sont réalisées.

A terme, le calendrier de traitement des données des Esa DOM pourra être aligné sur celui de la métropole et l'objectif initial sera alors atteint.

Dans la suite du § 5.6, on raisonne donc uniquement sur le sous-champ Ésane 1 hors DOM, et sur l'échantillon Esa⊕EAP hors DOM.

5.6.2 Correction de la non-réponse totale dans les enquêtes

On s'intéresse ici à la correction de la non-réponse totale dans **l'ensemble des enquêtes, i.e. Esa⊕EAP**, notée S dans la suite de ce document.

Le statut des unités, déterminé lors de l'étape de caractérisation des liasses fiscales, constitue la référence pour distinguer les unités actives dans le champ Ésane des unités hors-champ Ésane cf. §. Ce statut, combiné aux données de gestion de l'enquête relatives aux questionnaires retournés et à leur qualité, permet de répartir les entreprises de l'échantillon S en trois catégories :

Partition de l'échantillon Esa⊕EAP en « répondants » / « non-répondants » / « hors-champ »

❶ les unités répondantes dans le champ de l'enquête ⇒ catégorie R : il s'agit des entreprises appartenant au champ de l'enquête et ayant retourné un questionnaire exploitable ;

❷ les unités non-répondantes dans le champ de l'enquête ⇒ catégorie NR : il s'agit soit des entreprises ayant retourné un questionnaire inexploitable[174], soit des entreprises actives ou présumées actives[175] n'ayant pas retourné de questionnaire ;

❸ les unités hors sous-champ Ésane 1 ⇒ catégorie HC : il s'agit essentiellement des entreprises qui ont répondu (elles sont au départ dans l'échantillon) mais dont la réponse révèle qu'en réalité elles sont hors-champ Ésane ou présumées hors-champ Ésane, ainsi que plus marginalement des entreprises actives dans le champ Ésane mais n'appartenant plus au sous champ Ésane 1, en général du fait d'un changement d'APE que révèle leur réponse à l'enquête.

À noter que la confrontation du statut des unités (actif donc à extrapoler ou statistiquement cessé donc à ne pas extrapoler) selon une procédure s'appuyant quasi-exclusivement sur des sources d'informations externes à l'enquête – et des données d'enquête révèle un problème concernant certaines unités ayant répondu à l'enquête[176] : en effet, certaines entreprises ayant retourné un questionnaire et *appartenant au sous-champ Ésane 1* d'après leurs réponses sont par ailleurs présumées *inactives* ou *hors-champ Ésane*. Dès 2008, l'examen de la liste de ces entreprises au comportement atypique avait permis de mettre en évidence la présence dans l'échantillon d'unités manifestement hors du champ de l'enquête : personnes physiques appartenant à une société civile professionnelle (avocats ou notaires appartenant à un cabinet par exemple) enquêtées par ailleurs à titre individuel[177], entreprises ayant renvoyé un questionnaire avec un chiffre d'affaires nul et un commentaire indiquant une cessation d'activité, etc. Afin de pallier ce problème, il a été décidé de **considérer ces unités comme hors-champ**, le statut « présumées cessées ou hors-champ Ésane » primant donc sur la réponse à l'enquête. Cette approche a été reconduite les années suivantes[178].

[174] Cette catégorie contient en particulier les unités ayant renvoyé un questionnaire sans renseigner ni leur chiffre d'affaires ni sa ventilation par activité.
[175] On néglige donc ici le cas des entreprises actives ou présumées actives dans le champ Ésane mais qui seraient hors du champ de l'enquête (sous champ Ésane) pour des raisons de changement d'APE, de CJ ou de statut marchand. Au vu du champ particulièrement large de l'enquête ESA⊕EAP en termes de secteurs, cette hypothèse ne paraît ni déraisonnable ni excessivement forte.
[176] Ce problème concernait par exemple environ 1200 entreprises en 2009, principalement hors exhaustif.
[177] Mais répondant quand même pour l'ensemble de l'activité de leur SCP (société civile professionnelle) en général.
[178] Ce qui est d'autant plus logique que cela est en parfaite cohérence avec la décision de conférer au statut déterminé lors de l'étape de caractérisation des liasses fiscales un caractère de référence.

5.6.2.1 Correction de la non-réponse totale pour les entreprises appartenant à la partie exhaustive de l'échantillon

La correction de la non-réponse totale sur la partie **exhaustive** de l'Esa⊕EAP se fait **par imputation**, (car pour les grandes entreprises de cette partie exhaustive, on a besoin de pouvoir travailler[179] sur des fichiers exhaustifs de données individuelles) à partir des données N-1 de façon quasi-systématique pour éviter les changements d'APE intempestifs qui se produiraient si on procédait par hotdeck de strate. On utilise évidemment le même traitement que celui opéré pour la base de production (Cf. § *5.5.1*).

5.6.2.2 Correction de la non-réponse totale par repondération pour la partie échantillonnée de l'Esa⊕EAP

En ce qui concerne la partie **échantillonnée** de l'Esa⊕EAP, la correction de la non-réponse totale est effectuée **par repondération** selon la méthode des **groupes de réponse homogènes** (GRH).

Étudions d'abord comment sont formés ces GRH. Une première régression logistique, effectuée **chaque année,** permet de déterminer les variables les plus **explicatives de la non-réponse** parmi les variables auxiliaires disponibles[180]. Cinq variables reviennent le plus souvent : la présence ou l'absence de liasse fiscale, le secteur au lancement, l'effectif au 31/12 issu d'Épure[181], la zone géographique d'implantation de l'entreprise[182] et la durée d'existence de l'entreprise. Les modalités des variables explicatives ont été choisies en fonction de la distribution des taux de réponse observés et de manière à garantir un nombre minimal de répondants par modalité en vue de la constitution des groupes de réponse homogènes :

- Pour la **liasse fiscale**, on a deux modalités : présence ou absence ;
- pour **le secteur au lancement**, la situation est très différente entre l'Esa et l'EAP. Pour l'industrie, du fait de la faible taille de la partie échantillonnée dans l'EAP, on raisonne ainsi systématiquement au niveau division ou groupement de divisions de la nomenclature. Hors industrie, le niveau de nomenclature retenu dépend fondamentalement des allocations initiales de l'échantillon : niveau « groupe » dans la majeure partie des cas, et regroupement de groupes, voire de divisions sinon ;
- cinq tranches ont été constituées pour **l'effectif issu d'Épure** : « 0 salarié », « 1 salarié », « 2 à 5 salariés », « 6 à 9 salariés » et « 10 salariés et plus » ;
- on distingue deux **zones géographiques** : « Île de France » *versus* « Reste de la France » ;
- enfin, pour **la durée d'existence**, deux catégories d'entreprises sont considérées : les entreprises ayant plus de 10 ans d'existence et les autres.

Les groupes de réponse homogènes (GRH) sont ensuite constitués à partir des **croisements des modalités de ces différentes variables explicatives**. Afin d'obtenir des estimations de probabilité de réponse un minimum robustes, des regroupements de modalités sont effectués au besoin en vue d'assurer un nombre minimal de **50 répondants par GRH**.

Une fois ces groupes constitués, les probabilités de réponse sont supposées indépendantes d'un groupe à l'autre et **constantes** au sein de chaque groupe. La probabilité de réponse $\hat{\pi}_i^{rep}$ d'une unité i appartenant à un groupe h donné est alors estimée, en notant d_k les poids de sondage initiaux, par :

$$\hat{\pi}_i^{rep} = \hat{\pi}_h^{rep} = \frac{\sum_{k \in R} d_k \; \mathbb{1}_{k \in \text{groupe } h}}{\sum_{k \in R \oplus NR} d_k \; \mathbb{1}_{k \in \text{groupe } h}}$$

[179] C'est vrai des comptables nationaux qui effectuent des retraitements sur données individuelles et des chargés d'étude qui veulent pouvoir suivre des populations.
[180] Il s'agit de diverses variables issues de la base de sondage ou de sources externes comme Épure, et donc disponibles pour l'ensemble des unités de l'échantillon : présence ou absence de liasse fiscale, APE au lancement, effectifs au lancement, effectifs issus d'Épure, valeur du CA fiscal, catégorie juridique, nombre d'établissements, région, type de l'unité légale, date de création de l'entreprise, etc. Les cinq qui sont détaillées sont celles qui sont le plus souvent significativement explicatives du comportement de non-réponse.
[181] Cette source exploitant les déclarations aux Urssaf a été présentée au § *1.1.4.1*.
[182] On se fonde sur l'adresse de son siège social. Rappelons qu'on est sur la partie échantillonnée donc sur de petites entreprises peu souvent multi-régionales.

À l'issue de cette étape, les poids des entreprises répondantes ont été ajustés de façon à prendre en compte le phénomène de non-réponse. La pondération d'une unité i donnée suite à cette correction de la non-réponse totale (CNRT) vaut donc :

$$w_i^{CNRT} = \begin{cases} \dfrac{d_i}{\hat{\pi}_i^{rep}} \text{ si l'entreprise i est répondante} \\ d_i \text{ si l'entreprise i est hors champ Esane 1} \end{cases},$$

et l'ensemble « répondants » ⊕ « hors sous champ Ésane 1» de la partie échantillonnée, pondéré par les poids w_i^{CNRT}, est extrapolable à l'univers de tirage – i.e. la base de sondage – hors DOM et hors exhaustif à l'aide de ces poids.

5.6.2.3 Gestion des unités atypiques non aberrantes par « winsorisation » *[2.6]*

On rencontre toujours dans les enquêtes des entreprises « atypiques », mais non aberrantes, marquées par exemple par des ratios de chiffre d'affaires par salarié anormalement élevés par rapport au reste des entreprises de même secteur et de même taille[183]. Lorsqu'elles appartiennent à la partie exhaustive de l'échantillon, elles ne posent pas de problème puisqu'elles existent et sont prises en compte pour ce qu'elles sont. En revanche, lorsqu'elles appartiennent à la partie échantillonnée, le caractère atypique est multiplié par leur poids, ce qui accroît de façon excessive la variance des agrégats qu'on peut calculer. On cherche alors à réduire leur influence, ici par une technique dite de « winsorisation[184] », qui va consister à diminuer leur poids[185].

On notera donc désormais dans la suite R l'ensemble des entreprises répondantes *de la partie échantillonnée*[186] de l'Esa⊕EAP hors DOM, et HC l'ensemble des entreprises hors sous champ Ésane *de la partie échantillonnée* de l'Esa⊕EAP hors DOM.

La gestion de ces unités atypiques non aberrantes est effectuée après l'étape de correction de la non-réponse totale (CNRT) que l'on vient de voir, par winsorisation portant sur la variable « chiffre d'affaires », issue de liasses fiscales. Plus précisément, les poids issus de la correction de la non-réponse totale ont été modifiés comme suit :

$$w_i^{Winsor} = \begin{cases} 1 + \dfrac{K_h\left(w_i^{CNRT} - 1\right)}{CA_i^{IEG}} & \text{si } CA_i^{IEG} \geq K_h \text{, pour une unité } i \in \text{strate h} \\ w_i^{CNRT} \text{ sinon} \end{cases}$$

où CA_i^{IEG} désigne le chiffre d'affaires de la liasse fiscale[187].

La détermination des seuils K_h est effectuée chaque année par des experts du département des méthodes statistiques de l'Insee, en suivant une procédure proposée par Kokic et Bell (1994). Il ne s'agit pas d'une procédure totalement automatique qui serait intégrée à la chaîne Ésane. Celle-ci consiste à choisir les valeurs des seuils qui minimisent l'erreur quadratique moyenne – somme de la variance et du carré du biais – des statistiques que l'on souhaite produire. Kokic et Bell parviennent ainsi à déterminer le jeu de seuils optimaux en se plaçant dans un cadre asymptotique, bien adapté au contexte des ESA.

---

[183] Il peut y avoir de multiples raisons, entre autres le fait qu'une sous-classe de la nomenclature peut recouvrir des activités assez hétérogènes… « Atypique » correspond à quelques ordres de grandeur en plus ; « aberrant » correspond par exemple à un facteur 1000. Les données aberrantes doivent bien entendu être corrigées.

[184] Du nom de l'ingénieur et biostatisticien britannique Charles P. Winsor (1895–1951). Le terme de « winsorisation » a été créé en 1960 par Wilfrid J. Dixon dans sa publication *Simplified estimation from censored normal samples*, Annals of Mathematical Statistics, vol. 31, pp. 385–391.

[185] On peut montrer que cette technique est plus efficace que par exemple une troncature des valeurs atypiques au delà de certains seuils.

[186] Attention, il y a là un léger changement des notations par rapport au début du § *5.6.2* où elles ont été introduites. Les notations R, NR et HC, ne comprennent plus la partie exhaustive de l'échantillon. Ce paragraphe et le suivant visent en effet à corriger des poids, ce qui n'a évidemment pas de sens pour la partie exhaustive pour laquelle ceux-ci sont, par définition, égaux à 1. Les entreprises de la partie exhaustive de l'échantillon ne représentent qu'elles-mêmes.

[187] Le traitement des liasses fiscales sera exposé dans le chapitre 6. Il s'effectue dans un sous-système d'Ésane appelé Informations Économiques Générales d'où vient le nom « IEG ».

5.6.2.4 Calage de la partie échantillonnée de l'Esa⊕EAP

Cette opération ne concerne toujours que la partie échantillonnée de l'Esa⊕EAP. Le calage va consister à modifier les poids associés aux unités de l'échantillon (eux mêmes déjà modifiés par les opérations d'estimation de la non-réponse puis de winsorisation, par rapport aux poids du tirage initial) pour tenir compte des unités inutilisables et faire en sorte que le redressement des unités exploitables permette de retrouver l'univers de tirage initial (lui même corrigé des unités qui se sont révélées hors champ, ex post, par une source externe à l'enquête).

Les unités présumées cessées, ainsi que les unités détectées comme hors-champ de façon certaine par une source externe à l'enquête, ne participent donc pas au calage, contrairement aux unités hors-champ détectées par l'enquête – présentes à la fois dans l'échantillon utilisé pour le calage et dans l'univers de calage –.

En conséquence, le calage porte sur les unités de l'ensemble R⊕[HC source Esa], et consiste en une modification des poids w_i^{Winsor} de façon à **retrouver, à partir de l'échantillon** R⊕[HC source Esa] extrapolé, **des quantités connues sur l'univers** U_1 hors DOM, hors exhaustif et privé des unités hors-champ ou présumées hors-champ par une source externe à l'enquête.

Rappelons toutefois l'existence, jusqu'à la campagne 2013 comprise, de l'EAP0, qui constitue un défaut de couverture de l'échantillon sur le sous champ Ésane 1. L'EAP0 est constitué des entreprises industrielles réalisant moins de 5 M€ de CA, et sans salariés ou de moins de 20 salariés appartenant à une activité comportant moins de 2 produits ProdEap. Ces entreprises sont exclues de la base de sondage lors du tirage de l'EAP, et donc non couvertes par l'enquête. Il convient donc d'enlever ces entreprises du champ retenu pour le calage[188].

Au final, l'univers de calage correspond donc au sous champ 1 Ésane au lancement (noté U_1), hors DOM, hors exhaustif et hors EAP0, et privé des unités hors-champ ou présumées hors-champ par une source externe à l'enquête.

Le calage est ensuite réalisé sur les variables « chiffre d'affaires par groupe » et « nombre d'entreprises par division »[189]. Le chiffre d'affaires est une variable de base du dispositif qui est corrélée avec pratiquement toutes les variables comptables ; le calage permet alors d'améliorer la précision des estimateurs en réduisant la variance des estimateurs des totaux de toutes les variables **corrélées au chiffre d'affaires**[190]. Inversement, le nombre d'entreprises par secteur n'est corrélé avec rien alors que c'est aussi une des variables d'intérêt du système et on risquerait d'avoir des estimateurs très imprécis pour cette variable si on ne réalisait pas ce calage. Plus précisément, les poids w_i^{Winsor} **sont ajustés de telle façon que** :

- le **chiffre d'affaires**[191] d'un secteur – <u>au niveau groupe</u> – donné, tel que défini par le répertoire, soit retrouvé par l'échantillon R⊕[HC source Esa] extrapolé ;
- le **nombre d'entreprises** d'un secteur – au niveau division – donné, tel que défini par le répertoire, soit retrouvé par l'échantillon R⊕[HC source Esa] extrapolé.

Les poids $w_i^{calé}$ issus du calage – réalisé à l'aide de la macro Calmar2[192] par le département des méthodes statistiques de l'Insee à l'issue de l'étape de winsorisation – respectent donc les contraintes de calage suivantes, pour tout groupe G / division D appartenant au sous champ Ésane :

[188] À noter que ce défaut de couverture a également des conséquences sur les estimations (cf. chapitre 9).
[189] Modulo quelques regroupements pour gérer des groupes/divisions à effectifs réduits.
[190] Le calage permet en quelque sorte de « purger » la variance de l'estimateur du total d'une variable comptable de sa partie expliquée par le chiffre d'affaires.
[191] Le chiffre d'affaires utilisé ici est le chiffre d'affaires issu d'IEG, intégrant les traitements opérés lors des phases de réception et de contrôles (CQR, macro-contrôles PSCP) des données administratives.
[192] Il s'agit d'un programme écrit en langage SAS ©, développée à l'Insee pour l'exploitation des enquêtes.

$$\begin{cases} \sum_{i \in R \oplus HC\ source\ ESA} w_i CA^{IEG}(i)\ \mathbb{1}_{groupe_rep=G}(i) = \sum_{i \in U_1 \setminus [exhaustif \oplus DOM \oplus EAP0 \oplus HC\ externes]} CA^{IEG}(i)\ \mathbb{1}_{groupe_rep=G}(i) \\ \sum_{i \in R \oplus HC\ source\ ESA} w_i\ \mathbb{1}_{div_rep=D}(i) = \sum_{i \in U_1 \setminus [exhaustif \oplus DOM \oplus EAP0 \oplus HC\ externes]} \mathbb{1}_{div_rep=D}(i) \end{cases}$$

[La première équation correspond au calage des chiffres d'affaires, la seconde au nombre d'entreprises[193]. Le membre de gauche correspond au redressement de l'échantillon avec les poids $w_i^{calé}$ que l'on cherche et le membre de droite est la valeur connue dans l'univers sur laquelle on cale. Pour ces deux variables, calculer un agrégat à partir de l'échantillon ou à partir des liasses fiscales donne, par construction des poids, le même résultat].

Il y a davantage d'inconnues (les $w_i^{calé}$) que d'équations ce qui laisse des **degrés de liberté pour essayer de déformer le moins possible les poids**. Notons que les divisions formant une partition de l'univers, ces opérations de calage peuvent être réalisées de manière indépendante division par division. En procédant de la sorte, le calage peut être effectué en limitant la dispersion maximale des rapports de poids (poids après calage / poids avant calage) à un facteur deux[194] pour 80 à 90% des 63 divisions. Quant aux divisions récalcitrantes, elles font l'objet de calages moins stricts mais qui demeurent parfaitement acceptables, puisque dans le cas le moins favorable, la dispersion des rapports de poids **reste contenue entre les bornes 0.4 et 4.**

[193] Dans les termes de sommation, le symbole \ signifie « moins »
[194] À la hausse comme à la baisse.

6. Les principes méthodologiques propres aux traitements des données comptables

Comme vu précédemment, les **liasses fiscales** correspondant aux données comptables figurant dans les déclarations fiscales des « unités légales » sont **fournies à l'Insee par la DGFiP**. Cette dernière fournit au total 80 fichiers[195]. Ces fichiers sont extrêmement lourds car ils contiennent non seulement beaucoup de lignes mais également beaucoup de colonnes.

C'est au final **plus de 4 milliards de données** que l'Insee récolte de la sorte chaque année.

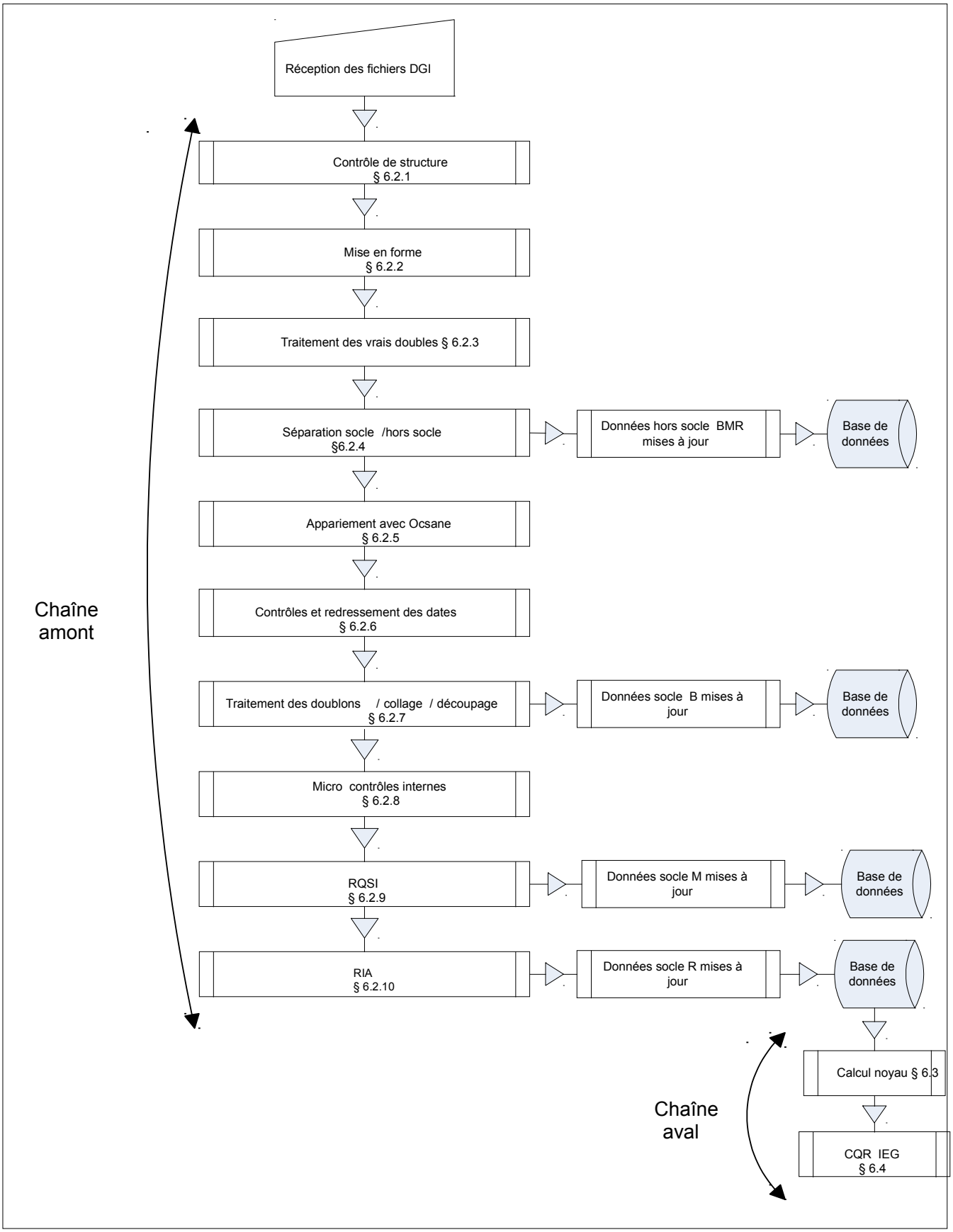
De premiers traitements qui sont une reprise **à l'identique de ce qui était fait dans « Suse[196] »** (Système Unifié des Statistiques d'Entreprises) ont lieu uniquement sur des fichiers plats hors base de données. Cette partie de la chaîne de production est communément appelée **la chaîne amont** et ne génère **aucune intervention manuelle[197] des gestionnaires**. Cette décision de travailler « en amont » de la base de données qui abrite le système d'information Ésane a été prise uniquement pour des raisons de temps de traitements informatiques.

Une deuxième salve de traitements intervient après chargement des données dans la base de données. Il s'agit de **la chaîne aval** qui met en œuvre les « **Contrôles, Qualification, Redressements** » (dits CQR[198]) tel que décrits ci-dessous au § *6.4* et qui sont à l'origine des traitements manuels effectués par les gestionnaires du service Ésane de Nantes.

L'objet de ce chapitre est de **décrire les deux chaînes de contrôle** qui portent donc sur les données comptables des entreprises.

Le graphique suivant présente la succession de ces deux phases et présente du même coup la plupart des paragraphes dans lesquels elles seront décrites :

[195] Il y a quatre centres de traitement des impôts différents qui fournissent un fichier par type de régime (5) pour chaque livraison (4). La première livraison ne concernait qu'un seul type de régime (BIC-RN) jusqu'à la campagne 2010 ; depuis la campagne 2011, tous les régimes fiscaux sont reçus dès la première livraison.
[196] Voir *l'avant propos*.
[197] En termes de vocabulaire, on dit que les gestionnaires assurent un travail ou un traitement « manuel » par opposition aux traitements « automatiques » réalisés par des algorithmes programmés dans des chaînes de traitements automatisées.
[198] Le principe général des CQR a été présenté en détail au *chapitre 2*.On parle ici de la mise en œuvre spécifique à la source fiscale.



RQSI : redressements quasi sûrs internes ; RIA : redressements internes automatiques
Données B : brutes ; M : modifiées ; R : redressées (cf. § 2.1)

Ce chapitre du document va :

- présenter les **deux** grandes **phases du contrôle** des données fiscales (chaîne amont, chaîne aval) (§ *6.1*) ;
- présenter succinctement les **dix étapes** différentes de la chaîne amont (§ *6.2.1* à § *6.2.10*) ;
- expliquer les différentes populations étudiées (définition du **noyau**) (§ *6.3*) ;
- présenter les principaux traitements automatiques et manuels de la chaîne **aval** qui viennent compléter les micro et macro contrôles qui ont été décrits plus haut (traitement des doublons, traitement des dates, traitement des suspicions de restructuration, « contrôle global » versus « contrôle var ») (§ *6.4*) ;
- présenter la méthodologie du traitement de la **non-réponse totale** mise en œuvre (§ *6.5*).

6.1 Les deux phases de contrôle des données

Le processus de production des résultats se partage en deux grandes phases :

- l'élaboration de **résultats provisoires** programmée pour avant la fin octobre de l'année N+1 pour la campagne N, au niveau des groupes de la Naf sur un nombre d'agrégats limité (6 agrégats retenus : CA, VA, EBE, salaires, résultat comptable et investissement). Cette diffusion provisoire permet de fournir, avant le 31 octobre de l'année N+1, les estimateurs provisoires[199] requis par le règlement SBS[200] d'Eurostat, pour les données de l'année N.
- l'élaboration de **résultats complets** entre octobre N+1 et fin juin N+2 à des niveaux détaillés (NAF700 ou NES[201]116*classes de taille).

6.1.1 Élaboration des résultats provisoires

Les résultats provisoires sont **diffusés à l'automne**. Ces résultats sont élaborés à partir des liasses comptables du régime réel normal des Bénéfices Industriels et Commerciaux dites "anticipées"[202] jusqu'à la campagne 2010, qui ont été envoyées par procédure télématique à la DGFiP[203] et que celle-ci nous fournit mi-juin. À partir de la campagne 2011, ce sont tous les régimes fiscaux qui sont désormais reçus dès les anticipées de juin.
On utilise également des données issues du « profilage » réalisé par l'Insee sur cinq grands groupes qui ont accepté historiquement de fournir des données consolidées avant que l'Insee ne généralise cette pratique à l'ensemble des grands groupes (cf. *chapitre 11* pour plus de détails) puis ultérieurement à l'ensemble des groupes ; ils sont traités dans ces campagnes comme s'il s'agissait d'unités légales (c'est-à-dire que les profileurs introduisent directement leurs comptes consolidés dans le système d'information Ésane et on ne traite pas les comptes des unités légales de leurs contours, reçus de la DGFiP). À compter de la campagne 2013, une quarantaine de groupes correspondants à une grosse centaine d'entreprises « profilées » est traitée de la sorte (Cf. à nouveau le chapitre 11, § *11.6.6*).

La production de ces résultats se fait conformément aux étapes décrites ci-dessous (chaîne amont puis chaîne aval).

À l'issue de ces traitements provisoires, les données sont transmises au sous-système **réconciliation des données individuelles** (Rédi décrit dans le *chapitre 8* ci-dessous) qui comparera les caractéristiques des comptes qui sont également mesurées par d'autres sources (le chiffre d'affaires par les enquêtes ESA et EAP ou l'emploi et les salaires par les sources d'emploi) et les mettra en cohérence.

[199] Les estimateurs définitifs requis par le règlement prévoient un délai de 18 mois et sont donc tirés des estimateurs définitifs d'Ésane, qui sont produits en décembre.
[200] Structural Business Statistics : il s'agit du règlement n° 58/97 du Conseil du 20 décembre 1996 relatif aux statistiques structurelles sur les entreprises. Un nouveau règlement dit « FRIBS » pour *Framework Regulation Integrating Business Statistics* est en préparation et devrait être voté *en 2019 environ*. Il devrait augmenter le volume des données provisoires et surtout mettre l'accent sur une nouvelle définition de l'entreprise prenant en compte les groupes (Cf. *chapitre 11*)
[201] La NES (nomenclature économique de synthèse) est un regroupement exclusivement français des sous-classes de la Naf qui comprend 116 postes, particulièrement utilisé par les comptables nationaux.
[202] Des fichiers anticipés qui couvrent plus des trois quarts de la population attendue en poids économique (même si c'est beaucoup moins en nombre d'entreprises) et dont la couverture augmente d'année en année.
[203] Cela concerne en particulier toutes les grandes entreprises (relevant de la direction des grandes entreprises DGE), ce qui permet à ce fichier de peser une part économique importante de l'ensemble France Entière.

6.1.2 Élaboration des résultats définitifs

Les résultats définitifs N sont arrêtés en fin du deuxième trimestre N+2. Il s'agit d'environ 130 caractéristiques comptables validées à plusieurs niveaux assez fins des secteurs auxquels appartiennent les entreprises : NAF700 d'une part et croisement trois premiers chiffres du code APE *classes de tailles d'autre part.

Le **principe de flexibilité** est valable pour cette phase d'élaboration :

- flexibilité vis-à-vis de la phase provisoire : l'exécution ou non des traitements provisoires n'est pas bloquante pour celle des traitements définitifs qui peuvent s'appliquer à des fichiers anticipés qui n'auraient pas pu être pris en compte durant l'été.
- flexibilité vis à vis des fichiers traités : ils sont nombreux lors de la phase définitive : deux vagues d'envoi de la DGFiP, en septembre et en octobre, ceci multiplié par 5 régimes fiscaux et 4 centres de saisie des impôts (CSI), plus des fichiers de complétions de champ (coopératives IAA par exemple). Au total, plus de 40 fichiers vont être traités pour la phase définitive, le processus ne doit donc pas être bloqué par le retard ou l'absence de l'un d'entre eux.

À ce principe de flexibilité s'ajoute un **principe de capitalisation entre les phases provisoires et définitives**. On doit pouvoir exécuter les traitements définitifs sans que les provisoires l'aient été, mais à l'inverse, si les traitements provisoires ont été réalisés, les traitements définitifs doivent en tenir compte pour ne pas faire le travail deux fois.
Les types de traitements entre les deux phases sont similaires, les adaptations venant essentiellement de la multiplicité des sources et du nombre d'agrégats et strates à valider plus important.

À l'issue de ces traitements définitifs, les données des liasses fiscales non traitées pour les résultats provisoires sont transmises à Rédi.

6.2 La chaîne amont

Cette partie décrit en détail les différents éléments de la chaîne amont dont l'enchaînement a été présenté dans le graphique ci-dessus.

6.2.1 Contrôle de structure[204]

Chaque année, aux environs du mois d'avril, la DGFiP prévient l'Insee des modifications qui sont intervenues sur les liasses fiscales (ajout ou suppression de variables). Le dessin de fichier est donc connu avant la première réception qui intervient mi-juin en général.
Ce premier contrôle « de structure » vérifie que les fichiers reçus par la DGFiP correspondent bien au dessin annoncé. Dans le cas contraire, la chaîne s'arrête jusqu'à sa correction ou au renvoi d'un nouveau fichier conforme de la part de la DGFiP.

6.2.2 Traitements de mise en forme

Il s'agit de faire le lien entre les variables des liasses fiscales, et les variables de notre base de données. Pour chaque régime, ces traitements récupèrent les variables correspondant aux différents bigrammes des fichiers de la DGFiP[205], et les affectent aux variables comptables Ésane correspondantes.
En outre, la phase de réception - mise en forme comprend un contrôle préliminaire, basé essentiellement sur des comptages, vérifiant que les données transmises sont d'une qualité suffisante pour être chargées dans la base de données.

[204] Depuis la campagne Ésane 2010, un fichier test est reçu courant avril pour contrôler et vérifier au plus tôt nos programmes d'intégration de liasses.
[205] Dans les liasses fiscales, les différentes variables sont caractérisées par deux lettres. Par exemple, FS pour les achats de marchandises ou DA pour le capital social.

6.2.3 Traitement des vrais doubles

Dans sa livraison successive de fichiers, **la DGFiP ne raisonne pas en flux mais en stock**. Ainsi les liasses présentes dans les fichiers de la livraison des anticipés se retrouvent dans les fichiers de la livraison intermédiaire etc. Il est donc inutile de conserver et traiter les liasses que l'on a déjà reçues dans une livraison antérieure.
Ainsi, en sortie de mise en forme, le fichier traité est comparé au fichier de cumul du même régime et du même centre de traitement des impôts complété à la vague précédente (ce fichier s'implémente au fur et à mesure des vagues de réception). Les liasses qui sont en doublon entre ces deux fichiers **sans avoir été modifiées** sont supprimées. Une liasse est considérée comme modifiée à partir du moment où au moins un montant des caractéristiques du socle[206] diffère. Il s'agit d'une comparaison chaîne de caractères par chaîne de caractères.

6.2.4 Séparation socle / hors socle

La partition socle / hors socle est une partition dans les variables des liasses[207]. Le but de ces traitements est de passer des données propres à chaque régime, à un ensemble de données, appelé "**socle**", **commun pour tous les régimes**, dont on a besoin pour le calcul des agrégats finaux et sur lequel la suite des traitements s'effectuera. Ainsi, si pour un régime donné, une variable du socle ne correspond à aucune variable de la liasse, mais peut être obtenue par sommation d'autres variables présentes, cette opération sera faite au sein de ces traitements.
Les variables du socle comprennent les variables utiles pour les analyses économiques et financières et notamment :

- celles du compte de résultat (cadre 3 de la liasse fiscale) ;
- les principales du bilan ;
- les principales d'immobilisations ;
- les principales des amortissements.

Les variables hors socle concernent plus des informations purement « fiscales » :

- les variables permettant la détermination du résultat fiscal (différent du résultat comptable) ;
- les résultats du groupe ;
- les variables d'impôt ;
- les variables sur la détermination de la valeur ajoutée (cadre 16 de la liasse fiscale).

En sortie de ces traitements, il y a un fichier contenant toutes les variables du socle (environ **800 variables**), directement récupérées des liasses fiscales ou calculées, et un fichier contenant les variables "hors socle" (environ **800 variables également**).

Les caractéristiques du fichier « hors socle » sont jugées secondaires et ne présentent pas d'intérêt statistique majeur. Ces données sont donc chargées en l'état dans la base de données Ésane. Elles ne subiront **ni contrôles et redressements automatiques ni interventions manuelles**. Elles sont disponibles dans la base de production mais pas dans le fichier des résultats individuels dit FARE[208] et aucun agrégat correspondant n'est publié.

6.2.5 Appariement avec Ocsane

Rappelons qu'Ocsane est le répertoire interne du système (présenté dans le § *1.2.1.2*). Ce traitement avec Ocsane consiste en un appariement des données fiscales ou profilées avec le référentiel et en un traitement des rejets par un envoi au « système d'identification automatique de masse » (Siam) du répertoire Sirene [les liasses contiennent un identifiant Sirene (le Siren) de l'entreprise déclarante ; si cet identifiant n'est pas présent dans Ocsane, c'est soit que la liasse correspond à une unité hors champ ; il ne faut donc pas la retenir ; soit que le Siren indiqué par la DGFiP n'est pas le bon identifiant Sirene de l'entreprise[209] ; on a donc besoin « d'identifier » à qui ce Siren correspond ; c'est ce que permet de faire ce système d'identification comme on va le préciser ci-après].

[206] La définition du socle est détaillée ci-dessous.
[207] Il lui correspond une partition Noyau / hors noyau au niveau des unités légales (Cf. § *6.3*)
[208] Voir sa définition au § *9.6.4.2*
[209] De nombreux événements peuvent conduire une unité légale à changer de Siren ; la déclaration peut alors avoir été faite avec un ancien Siren de l'entreprise. Environ 1% des Siren sont ainsi erronés.

L'objectif de cet appariement est de trois ordres :

- associer un identifiant Siren de la meilleure qualité possible à chaque déclaration fiscale ;
- retenir les déclarations fiscales qui font partie du champ d'étude d'Ésane[210] ;
- récupérer les caractéristiques d'Ocsane utiles pour les futurs traitements des comptes en particulier l'APE statistique, mais aussi le nombre de liasses déjà appariées.

L'appariement se fait d'abord sur le numéro Siren[211]. Si le Siren fiscal n'existe pas dans Ocsane, l'appariement se fait alors sur l'IFRP[212] et l'on affecte le Siren figurant dans Ocsane et correspondant à l'IFRP de la liasse.

Les unités qui n'ont pu être appariées ni sur Siren ni sur IFRP sont envoyées au système d'identification automatique de masse (Siam) de Sirene. À partir de caractéristiques de l'unité légale figurant dans la liasse (raison sociale, enseigne, adresse), **le Siam permet d'associer un ou plusieurs Siren** du répertoire Sirene (qu'on appelle des échos) au rejet d'identification envoyé.

Ces résultats renvoyés par le Siam peuvent être de trois types :

- il y a un écho de bonne qualité qui permet une identification fiable du rejet[213] ;
- il y a plusieurs échos mais de qualité insuffisante pour que l'identification puisse être considérée fiable[214] ;
- aucun écho n'est renvoyé pour l'entreprise en rejet.

Dans les deux derniers cas, l'identification automatique est en échec. Les échos multiples ou douteux sont néanmoins conservés pour aider d'éventuels traitements interactifs.
Dans le premier cas, la liasse prend alors le Siren correspondant à cet écho. Elle prend un statut d'appariement ad-hoc. Les doublons éventuels produits par cette procédure sont traités plus loin, en même temps que les autres doublons du fichier d'origine.
L'ensemble des rejets ainsi traités est à nouveau confronté à Ocsane afin de ne **retenir in fine que des liasses dans le champ Ésane**. Ces unités nouvellement identifiées et faisant partie du champ Ésane sont alors **ajoutées aux liasses déjà identifiées** de manière à reconstituer le fichier complet avant identification.

6.2.6 Contrôle et redressement des dates déterminant l'exercice

Ces traitements vont s'efforcer de redresser les dates d'ouverture et de clôture de l'exercice ainsi que la durée de celui-ci, s'il s'avère que des contrôles lancés au sein de ce programme révèlent des anomalies.

Pour ce faire, on tente d'abord de trouver une cohérence entre les trois variables liées au début de l'exercice (date déclarée de début de l'exercice, date déclarée de clôture de l'exercice précédent et date de clôture de la liasse active enregistrée l'année précédente) :

- Si ces trois dates sont bien cohérentes[215], alors on considère que **la date de début d'exercice est bonne** et en cas d'incohérences entre cette date de début, la date de clôture et la durée de l'exercice (en mois), ces deux dernières variables seront redressées à partir de la date de début, en ajoutant la durée d'exercice à la date de début pour obtenir la date de clôture.
- Si au contraire, il n'a pas été possible de fixer une date de début à cause d'incohérences entre les trois variables liées indiquées au paragraphe précédent, c'est que les trois dates sont différentes. Le programme **va alors tenter de se baser sur la date de clôture** et la

[210] Typiquement, les unités légales des contours des 5 entreprises profilées suivies depuis le début des années 2000 et plus généralement des entreprises profilées décrites au chapitre 11 ne font pas partie de ce champ pour éviter un double compte avec les comptes de ces entreprises qui sont rajoutés à la fin.
[211] Il existe en fait deux Siren dans une déclaration : le Siren fiche (Siren du référentiel de la DGFiP) qui est pré-rempli sur la déclaration et le Siren liasse (Siren rempli manuellement par l'entreprise). Le Siren retenu est le Siren fiche, considéré comme de meilleure qualité ; le Siren liasse n'est utilisé que si le Siren fiche est vide ou égal à '000000000'.
[212] L'IFRP est l'identifiant interne des entreprises dans les systèmes informatiques de la DGFiP. Depuis le décret Madelin qui oblige toutes les administrations qui gèrent des entreprises à utiliser le Siren, la DGFiP utilise le Siren mais elle continue d'utiliser aussi l'identifiant interne qui lui est propre : l'IFRP.
[213] Par exemple on trouve une unité ayant la même raison sociale à la même adresse.
[214] Par exemple on trouve plusieurs sociétés à la même adresse mais aucune avec la raison sociale indiquée.
[215] Les deux dernières doivent être identiques et précéder d'un jour la date déclarée de début du « nouvel » exercice. On considère en fait que si deux dates parmi les trois sont cohérentes, c'est qu'elles sont bonnes et on aligne alors la troisième sur les deux premières ; cela suffit alors à déterminer une date de début « valide ».

durée de l'exercice pour chercher une cohérence avec l'une des trois dates de début possibles :

- Si l'une des trois dates de début possibles est cohérente avec la date de clôture et la durée de l'exercice, on considère que cette date est la bonne et on valide le triplet date de début, durée, date de fin.
- Si on n'y arrive avec aucune des trois dates et la durée d'exercice déclarée, on essaie avec une durée d'exercice de 12 mois et à nouveau on validera un triplet si l'un des dates de début est cohérente avec la date de clôture et une durée d'exercice de 12 mois.
- En cas de nouvelle impossibilité, le programme va calculer une "durée d'exercice prévue" en la rapprochant de celle de l'année précédente et en comparant avec l'évolution des salaires et traitements ; logiquement la masse salariale (charges comprises) est proportionnelle au premier ordre à la durée de l'exercice : on calcule donc la durée d'exercice « prévue » de l'année N par rapport à l'évolution du montant comptable des salaires et traitements soit :

  EXER_DUR_PREVUE = EXER_DUREE(N-1) * SalEtTrait(N) / SalEtTrait(N-1)

  Il va ensuite se baser sur la date de clôture et les trois dates de début existantes pour recalculer les différentes durées d'exercices possibles pour cette liasse. Il choisira ensuite celle qui est la plus proche de la durée prévue.

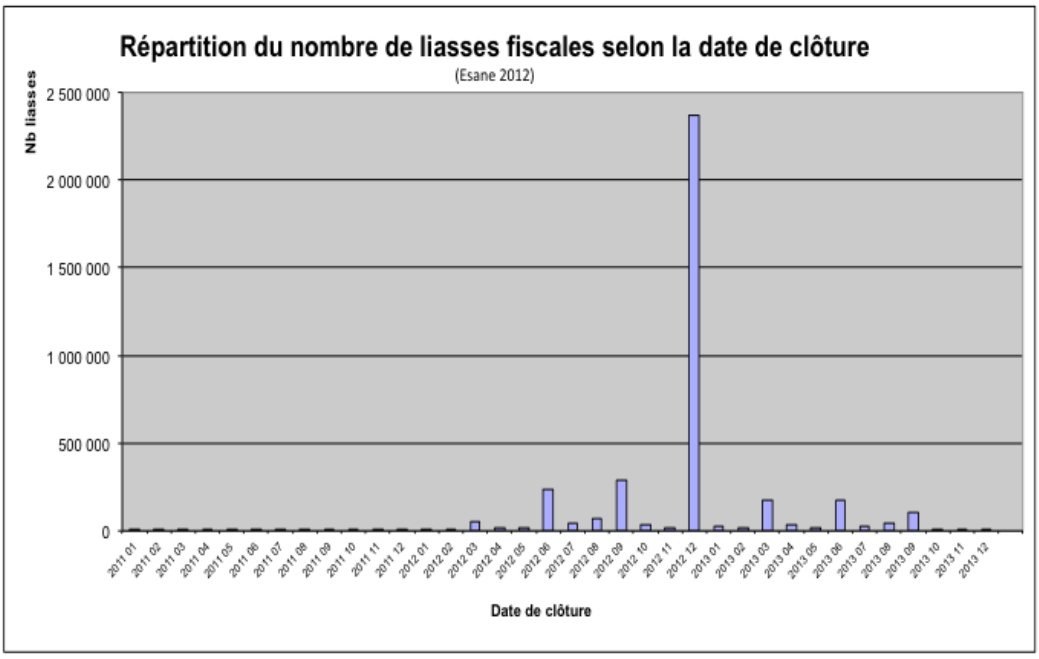
- **Si on n'a pas trouvé de date de début valide et s'il n'y a pas de date de clôture à la liasse fiscale**, alors le programme va remplir ces valeurs par défaut, en fonction des données existantes, en prenant au pire pour date d'ouverture, le premier janvier de l'année d'exercice, en date de clôture, le 31 décembre, et en durée, 12 mois.

Un cas particulier est fait pour les entreprises qui viennent d'être créées (pour lesquelles le raisonnement précédent ne marche pas) : la date de début d'exercice est fixée à la date de création de l'entreprise, telle qu'elle est définie dans le référentiel (Ocsane, et donc Sirene au départ). Les valeurs des autres variables sont calculées en fonction de leur cohérence avec cette date.

6.2.7 Découpage - collage

Cette partie présente les raisons et les décisions prises concernant le «**découpage** » et le « **collage** » des liasses fiscales partant de ce que dans la situation initiale de réception des liasses en provenance de la DGFiP, **il n'y a pas de bijection** pour une unité légale, entre liasse fiscale et année civile comme le montrent le tableau et le graphique suivant qui donnent le nombre de liasses fiscales reçues pour l'année 2012 en fonction de la date de clôture indiquée sur la liasse :

date de clôture de la liasse	Nombre de liasses reçues
2011 01	44
2011 02	6
2011 03	60
2011 04	65
2011 05	13
2011 06	13
2011 07	4
2011 08	4
2011 09	16
2011 10	3
2011 11	1
2011 12	324
2012 01	8 622
2012 02	6 873
2012 03	53 589
2012 04	19 049
2012 05	14 584
2012 06	237 976
2012 07	47 540
2012 08	66 587
2012 09	285 162
2012 10	39 284
2012 11	18 438
2012 12	2 365 454
2013 01	23 002
2013 02	18 173
2013 03	172 199
2013 04	32 569
2013 05	20 289
2013 06	176 739
2013 07	28 411
2013 08	41 986
2013 09	102 487
2013 10	5 121
2013 11	1 442
2013 12	2 901
Total	**3 789 030**



On constate que relativement à une campagne N donnée (ici 2012) on reçoit des liasses dont la date de clôture peut aller de janvier N-1 (ici janvier 2011) à décembre N+1 (ici décembre 2013), avec des pointes en fin de trimestre ; les liasses relatives à N-1 sont évidemment très minoritaires. Pour une année civile donnée, on peut d'ailleurs avoir plusieurs liasses fiscales de la même unité légale, ou au contraire aucune clôture d'exercice dans l'année[216]. Par rapport à l'ensemble des liasses reçues, le nombre de celles qui clôturent en décembre 2012 et qui couvrent donc majoritairement l'année civile 2012[217], est très majoritaire mais ne représente malgré tout que 62,4% des liasses reçues. Ce ratio d'environ 3/5 des liasses reçues, clôturant au 31/12 de l'année de campagne considérée, est assez stable dans le temps.

Or, la statistique d'entreprise vise à calculer des agrégats **se référant à des années civiles**. Par des opérations de « découpage » lorsqu'une liasse couvre des périodes trop longues ou (dans certains cas) ne coïncidant pas avec l'année civile[218], ou de « collage » lorsqu'il faut concaténer plusieurs liasses pour atteindre les 12 mois, les traitements décrits dans ce paragraphe, ***pour l'essentiel nouveaux par rapport à ce qui existait auparavant dans Suse***, visent donc à se ramener à une situation où on peut **faire se correspondre au mieux année civile et déclaration fiscale**.

Si l'habitude la plus répandue est de faire des exercices comptables sur des durées de 12 mois, et souvent sur l'année civile, **les règles de clôture d'exercice** figurant dans le code du commerce (CC) ou le code général des impôts (CGI) **sont assez peu restrictives dans l'absolu** (Article L123-12 du code du commerce). On peut, en théorie, avoir des exercices comptables de durée différente tous les ans, plusieurs déclarations par an ou une année civile sans aucune clôture d'exercice. Pour ce dernier cas, s'il ne s'agit pas d'une création d'entreprise, il est cependant requis de faire une déclaration aux impôts (article 37 du code général des impôts).

Seules deux contraintes doivent être respectées :

- La **continuité de la comptabilité** (CC L123-19) : c'est-à-dire qu'il ne peut y avoir au cours de l'existence de l'entreprise de période non couverte par un exercice comptable (une clôture un 31 mars et l'ouverture le 1er septembre suivant est impossible par exemple). Réciproquement, deux liasses ne peuvent recouvrer une même période (hors liasse rectificative, mais celle-ci annule la liasse d'origine). Enfin, il y a un accrochage comptable entre les liasses consécutives : le bilan d'ouverture d'une liasse est égal au bilan de clôture de la liasse qui la précède.
- **La périodicité des déclarations** auprès de la DGFiP : il y a toujours **au moins une déclaration déposée aux impôts par année civile** (CGI 12, 36 et 37), **à l'exception des créations** qui en sont dispensées leur année de création (CGI 209-I). En conséquence il ne peut y avoir de durée d'exercice dépassant 23 mois.

À partir de ces deux contraintes**,** les **différents cas possibles des liasses fiscales clôturant durant l'année N**, vont être étudiés.
L'objectif est de passer d'une série de liasses comptables fiscales de l'entreprise à une série de liasses comptables Ésane de cette entreprise qui recouvrent au mieux les années civiles. L'objectif que l'on vise au final est donc **d'avoir une liasse « Ésane », et une seule, par année de référence et, dans la mesure du possible, que cette liasse Ésane couvre 6 mois au moins de l'année civile**.

Le principe « général » est d'affecter une liasse **inférieure ou égale à 12 mois à une année civile** et de **couper si nécessaire**[219] **les liasses de durée supérieure à 12 mois** pour assurer en retour que **toutes les années civiles se sont bien vu affecter au moins une liasse** de façon à assurer la bijection mentionnée au début du paragraphe.

[216] A l'expérience, le problème a même été compliqué par le fait qu'on retrouve quelquefois des liasses se référant au même Siren dans des régimes fiscaux différents, et qui a nécessité des adaptations de la chaîne. Ce problème est traité au § *6.4.3*. De même, il faut détecter et éliminer les liasses « vides » (c'est-à-dire sans aucun montant rempli) car elles pourraient sinon venir écraser des liasses valides.
[217] La majorité des liasses correspond à une durée d'exercice de 12 mois.
[218] Les liasses à cheval sur deux années ne sont pas systématiquement découpées. On verra justement dans quels cas on est amené à le faire (voir note suivante).
[219] On ne coupe pas systématiquement les liasses de durée supérieure à 12 mois. On ne le fait que si on n'a aucune autre liasse à affecter aux années civiles concernées par le débordement. Par contre, si on ne les a pas coupées, on les contracte comme on le verra.

Après avoir donné la définition de l'année de référence « principale » d'une liasse fiscale (§ 6.2.7.1), puis « secondaire » (§ 6.2.7.2) les opérations (complexes[220]) de découpage - collage seront détaillées en fonction de tous les cas possibles de dates de début et de fin d'une liasse par rapport aux années civiles (§ 6.2.7.3). Enfin, on étudiera les conséquences de ces opérations de découpage et de collage sur les variables elles-mêmes de la liasse en fonction du statut de chaque caractéristique (caractéristiques additives, stocks, etc.) (§ 6.2.7.4).

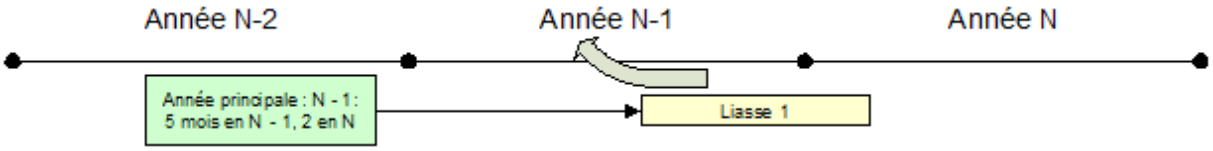
6.2.7.1 Définition de l'année de référence dite « principale » d'une liasse

1) ***Lorsqu'une liasse fiscale est à cheval sur deux années***, on dit qu'elle est affectée « **en principal** » à l'année civile avec laquelle elle a le plus de mois en commun.

Ce principe peut paraître de bon sens. ***Il correspond cependant à une innovation d'Ésane**** par rapport à Suse, **complexe** à mettre en œuvre : en effet dans Suse, une liasse était systématiquement affectée à l'année correspondant au mois de clôture. Ainsi une liasse clôturant fin janvier N+1, donc couvrant 11 mois en N était en réalité prise en compte dans la campagne N+1. Ceci pouvait introduire des biais, en cas de retournement conjoncturel, qu'on a cherché à corriger ici.

En revanche, comme on le verra plus loin, une même liasse peut couvrir plus de 6 mois de 2 années consécutives et comme on l'affecte en principal à celle où elle en a le plus, on peut alors se retrouver sans liasse correspondant à cette entreprise pour l'autre année. C'est ce qui conduira à définir une année de référence secondaire (Cf. § 6.2.7.2).

Par définition, l'année à laquelle elle est affectée **en principal** correspond à **la campagne Ésane où elle est prise en compte**.



Si la liasse allait à l'inverse de novembre N-1 à avril N, elle serait affectée en principal à N.

2) ***Les liasses qui comptent autant de mois sur une année civile que sur la suivante*** sont affectées **par convention** comme suit :

- durée d'exercice de 2, 4 ou 6 mois : liasse affectée en principal à l'année N-1 ;
- durée d'exercice de 8, 10, 12 mois ou plus : liasse affectée en principal à l'année N.

Si on appliquait ces principes aux liasses reçues de la DGFiP relativement à la campagne 2012 qu'on a vues au début du § 6.2.7 et en supposant, pour simplifier (ce qui n'est pas le cas dans la réalité), que toutes ces liasses aient des durées d'exercice de 12 mois, on aboutirait à la situation suivante (rappelons que ces liasses nous sont livrées en quatre fois, entre juin 2013 et mars 2014, comme on l'a vu au § 1.1.3) :

[220] La compréhension de ce paragraphe est indépendante de celle du reste des traitements. On peut donc sans inconvénient admettre la validité des traitements et passer au paragraphe 6.2.8.

date de clôture de la liasse	Nombre de liasses reçues	Totaux partiels
2011 01	44	
2011 02	6	
2011 03	60	
2011 04	65	
2011 05	13	
2011 06	13	
2011 07	4	
2011 08	4	
2011 09	16	
2011 10	3	
2011 11	1	
2011 12	324	553
2012 01	8 622	
2012 02	6 873	
2012 03	53 589	
2012 04	19 049	
2012 05	14 584	102 717
2012 06	237 976	
2012 07	47 540	
2012 08	66 587	
2012 09	285 162	
2012 10	39 284	
2012 11	18 438	
2012 12	2 365 454	
2013 01	23 002	
2013 02	18 173	
2013 03	172 199	
2013 04	32 569	
2013 05	20 289	3 326 673
2013 06	176 739	
2013 07	28 411	
2013 08	41 986	
2013 09	102 487	
2013 10	5 121	
2013 11	1 442	
2013 12	2 901	
		359 087
Total	**3 789 030**	3 789 030

- **Les 553 liasses qui clôturent en 2011** n'ont par définition aucun rapport avec l'année 2012 (il s'agit probablement de liasses rectificatives de liasses déjà envoyées). **On ne les prend pas compte**. On reçoit ces liasses 2012 à compter de juin 2013, moment où les résultats de 2011 sont en principe publiés ; il est donc trop tard pour pouvoir les prendre en compte et il est fort probable que cela n'aurait, de toutes façons, que très peu d'influence sur les résultats.
- **Les 102 717 liasses qui clôturent avant juin 2012,** sous l'hypothèse qu'elles ne seront pas recollées avec d'autres liasses de la même unité légale (voir plus loin), seront dans leur immense majorité affectées en principal à 2011 puisque si ce sont des liasses de 12 mois d'exercice, elles comptent davantage de mois en 2011 qu'en 2012. **On ne les prend donc pas non plus en compte**. Ceci marque une différence majeure avec Suse qui prenait lui en compte les liasses clôturant en 2012, quelle que soit leur durée.
- **Les 3 326 673 liasses qui clôturent entre juin 2012 et Mai 2013** ont – toujours sous l'hypothèse simplificatrice qu'il s'agirait de liasses de 12 mois – plus de mois en 2012 qu'en 2011 ou 2013 ; (voir le § 2 ci-dessus : les liasses de juillet 2011 à juin 2012 sont conventionnellement affectées à 2012). Ce sont donc **ces liasses qui seront prises en compte dans la « campagne » 2012**.

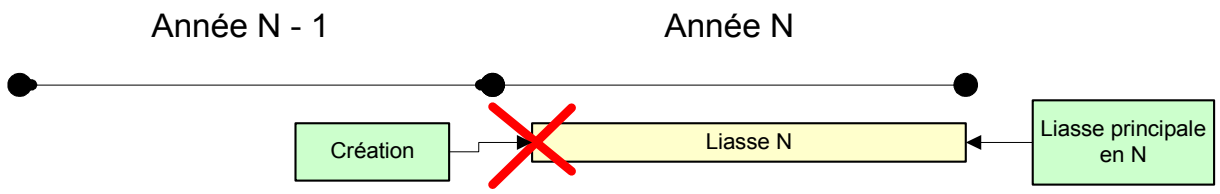
Là encore il s'agit d'un changement par rapport à Suse qui n'aurait pas pris en compte les liasses clôturant entre janvier et mai 2013 (soit 272 816 liasses !).

- **Enfin, les 359 087 liasses qui clôturent entre juin et décembre 2013,** ont toutes chances d'être affectées en principal à 2013 ; la DGFiP nous retransmet les liasses 2012 jusqu'en mars 2014 (livraison dite complémentaire) ce qui explique qu'on puisse avoir des liasses clôturant jusqu'à fin 2013. **On garde ces liasses qui seront réutilisées lors de la campagne suivante** (les liasses de 2012 reçues en 2011 ont été comptabilisées dans les 3 789 030 ci-dessus).

Dans la pratique, **ces décomptes sont un peu modifiés** par le fait que les liasses, comme on l'a vu, n'ont pas toutes une durée de 12 mois ce qui entraîne les découpages et les collages qui font l'objet de ce chapitre. Mais ce décompte **fait comprendre quelles sont les liasses qui sont utilisées dans les calculs**.

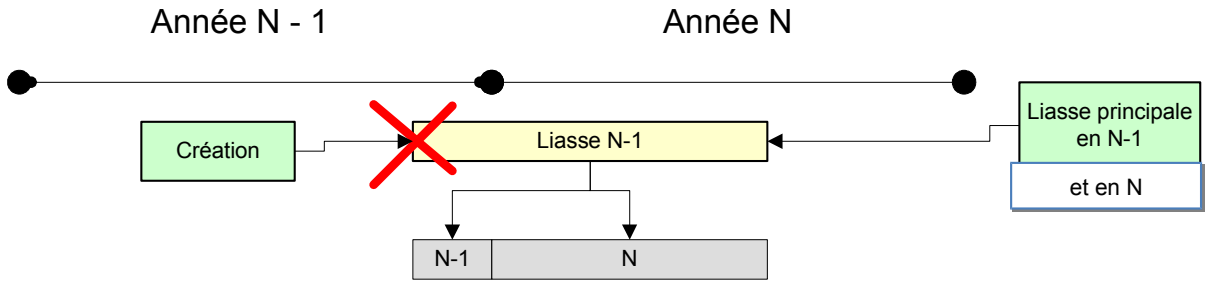
3) *Le cas particulier des créations ex nihilo[221] :*

- si sa **création est en N**, elle est affectée en **principal à l'année N** (principe général)**:**



(La croix rouge désigne la date de création ou de cessation que l'on commente)

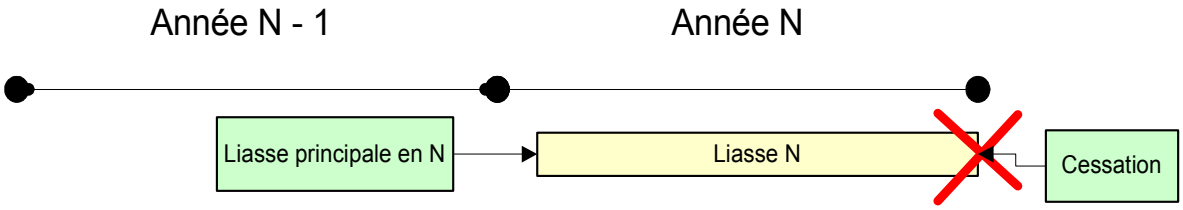
- mais si sa **création est en N-1** la liasse est coupée au 31/12/N-1 (Cf. § *6.2.7.3* pour la définition des coupures de liasse), même si liasse fait moins de 12 mois ; la première partie est affectée en principal à l'année N-1 et ce, quelle que soit la durée en commun avec l'année N-1 ; la deuxième partie est aussi affectée en principal mais à l'année N. En d'autres termes, cette liasse est prise en compte en N-1 et en N.



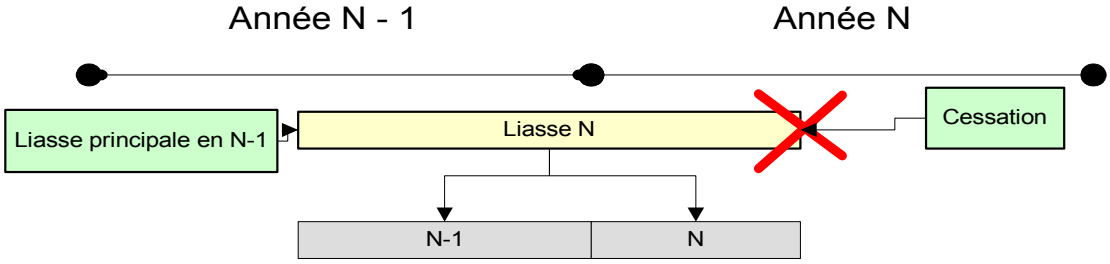
4) *Le cas particulier des cessations sans reprise :*

Réciproquement pour **les cessations en N sans reprise** :

- si sa dernière ouverture est en N, elle est affectée en principal à l'année N (principe général):



- mais si sa dernière **ouverture est en N-1**, elle est découpée même si sa durée n'est pas supérieure à 12 mois, la première partie est affectée en **principal à l'année N-1,** et la deuxième est aussi affectée en principal **à l'année N:**



[221] Créations qui ne résultent pas d'une restructuration.

6.2.7.2 Année de référence dite « secondaire »

Les règles précédentes n'assurent pas que toute année civile a une liasse qui lui soit affectée en principal (voir exemple 2 ci-dessous) ou qui couvre au moins 6 mois de l'année en question (voir exemple 1). On va donc affecter une année de référence dite secondaire aux liasses égalant ou dépassant 12 mois.

On dit qu'une liasse est **affectée en secondaire** à une année (autre que l'année principale) si **elle couvre au moins 6 mois de cette année** et que **les liasses en principal qui la précèdent** (s'il en existe) **couvrent moins de 6 mois de cette année**.

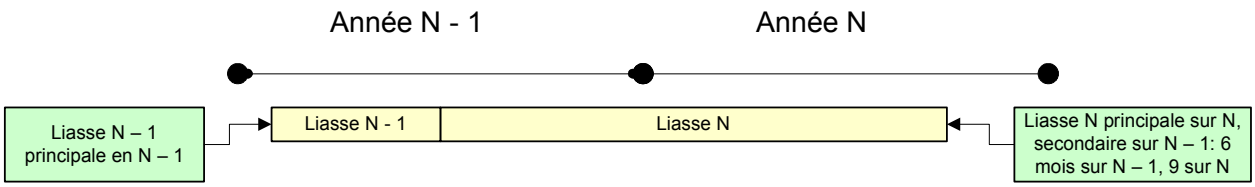
Quand une liasse est **affectée en secondaire**, **elle sera prise en compte dans les calculs pour deux exercices (principal et secondaire)** (cf. § *6.2.7.3* sur les découpages / collages de liasses).

Par convention, les créations ex-nihilo sont affectées en secondaire à leur année de création si celle-ci n'est pas principale. Réciproquement, les cessations sans reprise sont affectées en secondaire à leur année de cessation si celle-ci n'est pas leur année principale.

Exemples :

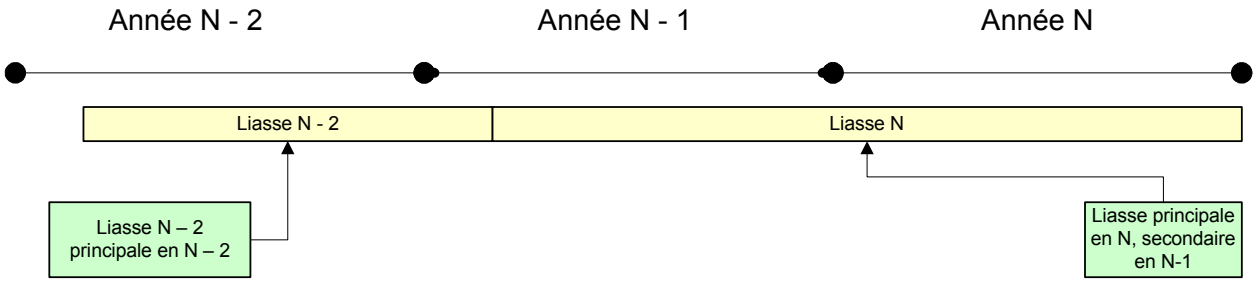
1. ***Liasse N-1 de début février N-1 à fin juin N-1 et liasse N de début juillet N-1 à fin septembre N :***

   La liasse N qui couvre 9 mois en N et 6 mois en N-1 est affectée en principal à N. La liasse N-1 ne couvre que 5 mois sur N-1. Du coup, sur N-1, on n'a aucune liasse couvrant au moins 6 mois.
   On dit qu'on fait une **affectation « secondaire » de la liasse N à N-1**



2. ***Liasse N-1 de début mars N-2 à fin février N-1 et liasse N de début mars N-1 à fin décembre N[222] :***

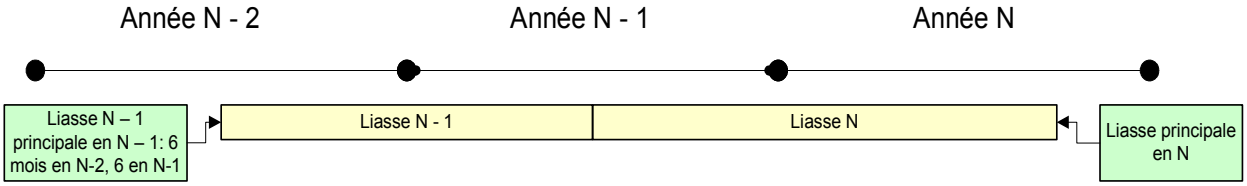
   La liasse N couvre 12 mois en N et 9 mois en N-1. Elle est donc affectée en principal à N. La liasse N-1 des impôts est affectée pour nous en principal à N-2 et on n'a donc aucune liasse qui soit affectée à N-1 !
   On dit qu'on fait une **affectation secondaire de la liasse N à N-1**



3. ***Liasse N de début juillet N-1 à fin septembre N et liasse N-1 de début juillet N-2 à fin juin N-1 :***

[222] Contrairement aux apparences, ce cas ne viole pas les règles du code des impôts : en effet, les impôts affectent une liasse à l'année de clôture de la liasse (comme à l'époque de Suse auparavant). C'est pourquoi ce qui est pour nous une liasse « N-2 » est pour les impôts une liasse « N-1 ». Dans l'exemple, on a donc bien une liasse qui se termine en N une en N-1 et une en N-2. On a donc bien une déclaration déposée par année civile.

Pas d'affectation secondaire de la liasse N à N-1 car la liasse N-1 qui est à cheval sur les deux années N-1 et N-2 est affectée en principal sur N-1 (voir plus haut). Du coup, on a déjà une liasse affectée en principal en N-1 et on n'a pas besoin d'affecter la liasse N en secondaire à N-1.



6.2.7.3 Opérations de découpage / collage des liasses

En termes d'opérations sur les liasses, on définit les traitements suivants :

- **Le découpage** : on coupe une liasse en deux liasses consécutives en prenant les montants d'ouverture pour la 1ère liasse, en calculant au prorata de la durée les montants en flux et en calculant les montants de clôture par la relation clôture=ouverture + flux. Les montants de clôture calculés de la première liasse deviennent ceux d'ouverture de la 2e liasse, les flux sont calculés au prorata de la durée et les montants de clôture sont ceux de la liasse d'origine avant scission
- **le collage** : on colle deux liasses consécutives pour n'en former qu'une en prenant en compte les montants d'ouverture de la première, les montants de clôture de la 2e et en additionnant les montants de flux.

On traite toutes les liasses **qui ont une année principale et une année secondaire**[223] en en faisant **une scission au 31 décembre de l'année la plus ancienne**. La première partie de la liasse est affectée en principal à l'année la plus ancienne, la deuxième partie de la liasse est affectée en principal à l'année la plus récente.

On est ainsi assuré d'avoir **au moins une liasse en principal par année civile**. **On colle ensuite** les liasses en principal sur la même année **s'il y en a plusieurs**. On a alors **une liasse Ésane unique par année civile**.

Remarque : une autre façon de faire – qui pourrait sembler plus simple – eût consisté à couper systématiquement les liasses au 31/12 et à recoller tous les morceaux d'une même année civile. On n'a pas procédé ainsi car on ne dispose pas de toutes les informations au bon moment et cela eût entraîné beaucoup plus d'opérations de découpage et de collage, opérations artificielles et délicates qui obligent à faire une hypothèse de pro rata temporis. Dans le cas où on a par exemple tous les ans une liasse de mars (N-1) à mars N, on l'affecte désormais en entier à N-1 (année qui a 9 mois en commun avec elle) alors que du temps de Suse on l'aurait affecté à N (qui n'avait que 3 mois). Si on découpait par année civile (ce qui serait théoriquement idéal), on affecterait 9/12ème de cette liasse à N-1, mais on ne pourrait rien faire des 3 mois sur N car il faudrait attendre la campagne N+1 (soit juin N+2) pour compléter les trois premiers mois de N. On préfère donc affecter toute la liasse (sans la découper) à N-1.

Une fois que l'on a obtenu ces liasses Ésane **uniques par année civile de campagne**, ***on les ramène ensuite à des périodes de 12 mois****, pour **les seules entreprises pérennes**[224], par une **opération de dilatation ou contraction** (décrites au § *6.2.7.4*) pour faire des comparaisons temporelles pertinentes. Les gestionnaires ont accès aux deux (liasse initiale sur période brute de longueur quelconque et liasse ramenée à 12 mois) pour faciliter leur analyse et leur dialogue avec l'entreprise.

Pour la diffusion, ce sont les liasses ramenées à 12 mois qui sont utilisées. L'idée est la suivante : si une liasse ne couvre pas l'année entière, c'est qu'**une autre liasse la couvre** (puisqu'il doit y avoir continuité de la comptabilité) que nous n'avons pas et donc qu'il s'agit d'une donnée manquante

[223] Cela représente que **très peu de liasses in fine**. Les exercices décalés par rapport à l'année civile mais de 12 mois n'ont en effet pas d'année secondaire. Ils ne sont affectés qu'à l'année sur laquelle ils ont le plus de mois d'activité. De même une liasse de 14 mois sera le plus souvent affectée à une seule année et ne sera pas découpée mais uniquement contractée sur l'année principale.

[224] Les créations et cessations continueront, elles, de compter pour le nombre réel de mois de l'année civile où elles existaient.

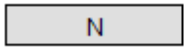
puisqu'on considère des entreprises pérennes dont nous savons qu'elles ne sont pas cessées[225]; la meilleure estimation que nous pouvons faire de ces données **manquantes** est donc de supposer que l'activité s'est poursuivie sur les mois inconnus au prorata des mois connus.

On va maintenant **passer en revue tous les cas de découpage collage (**15 au total) qu'on peut être amené à faire ; on se souviendra (Cf le graphique en début du § 6.2.7) que pour une campagne (N-1) donnée, on reçoit des liasses dont la date de clôture va de janvier (N-2) à décembre N :

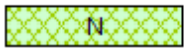
Remarque préliminaire : une entreprise peut faire de multiples déclarations clôturant la même année. Dans la pratique, ce nombre dépasse très rarement 2 et dans les cas exceptionnels où l'on a plus de deux liasses, on se ramènera à deux liasses en collant la 2e liasse et les suivantes (qui sont toutes incluses totalement dans l'année N). L'étude qui suit ne considérera donc que les cas avec une ou deux liasses clôturant en N.

Légende : on utilisera la symbolique de couleurs suivante :

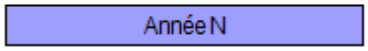
Liasse DGFiP



Liasse Esane sur une durée brute affectée à l'année N issue de la liasse DGFiP ci-dessus



Autre Liasse Esane sur une durée brute affectée à l'année N, qui n'est pas issue de la liasse DGFiP ci-dessus

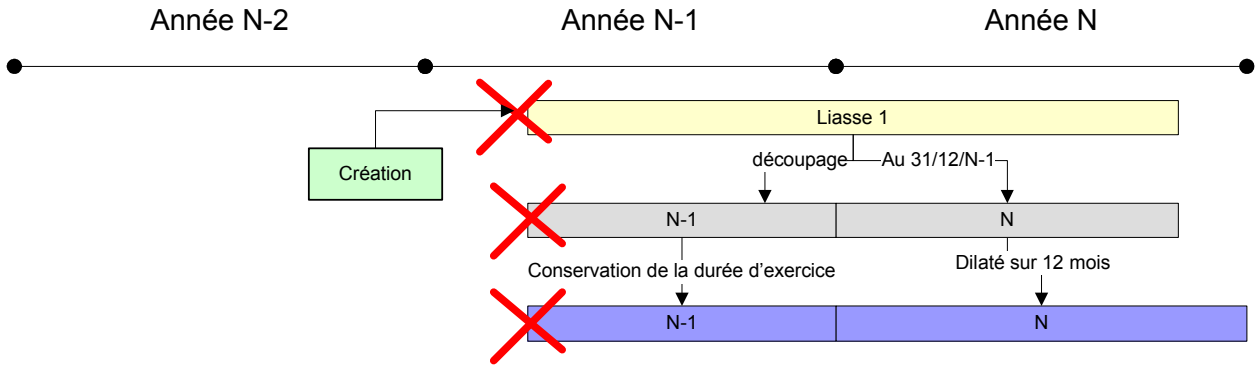


Liasse Esane sur une durée de 12 mois pour l'année N, issue de liasses Esane sur des durées brutes

Correspond à la date de création ou de cessation de l'entreprise suivante le cas (indiqué dans un cartouche vert).

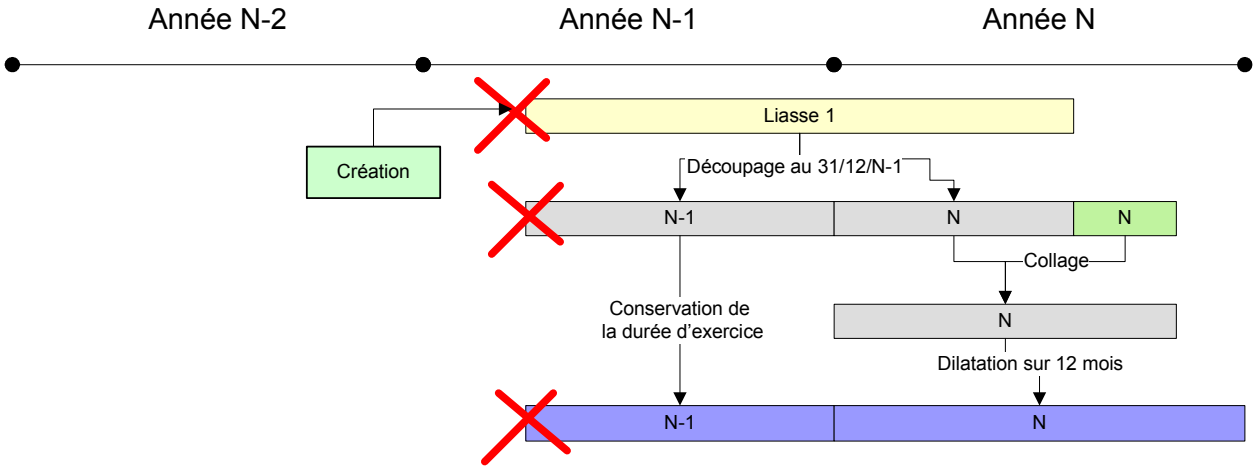
[225] À la date finale de bouclage des calculs, donc mars n+2, on connaît la quasi-totalité des entreprises qui ont cessé (de l'ordre de 97 %). Il pourrait exister marginalement quelques entreprises qui sont cessées alors que nous ne le savons pas. On ne dilate pas en revanche les entreprises cessées.

1. ***Création d'entreprise ex nihilo : l'entreprise est créée en N-1 et n'a pas d'autre liasse principale sur N***



La liasse est découpée en deux parties. La coupure se fait au 31 décembre N-1 La première partie est affectée en principal à l'année N-1, avec la durée correspondant au découpage. La deuxième partie est affectée en principal à l'année N. En l'absence d'autre liasse sur N, elle est dilatée sur 12 mois.

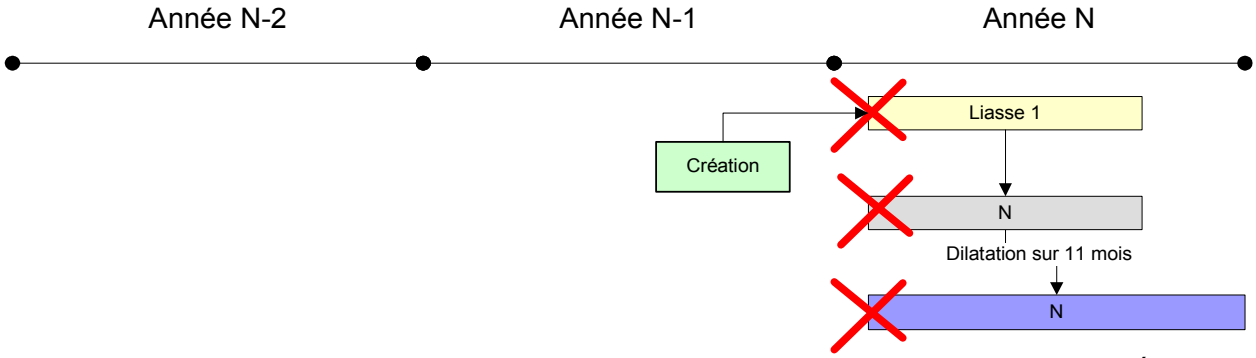
2. ***Création* d'entreprise ex nihilo : l'entreprise est créée en N-1 et a une autre liasse principale *sur N***



La liasse est de la même façon découpée en deux parties, au 31/12/N-1. La première partie est affectée à l'année N-1, sans modification de durée. La 2e partie de la 1ère liasse et la deuxième liasse Ésane sont ensuite collées puis la liasse résultante est dilatée sur 12 mois.

3. *Création d'entreprise ex nihilo : l'entreprise est créée en N et n'a pas d'autre liasse principale* sur N

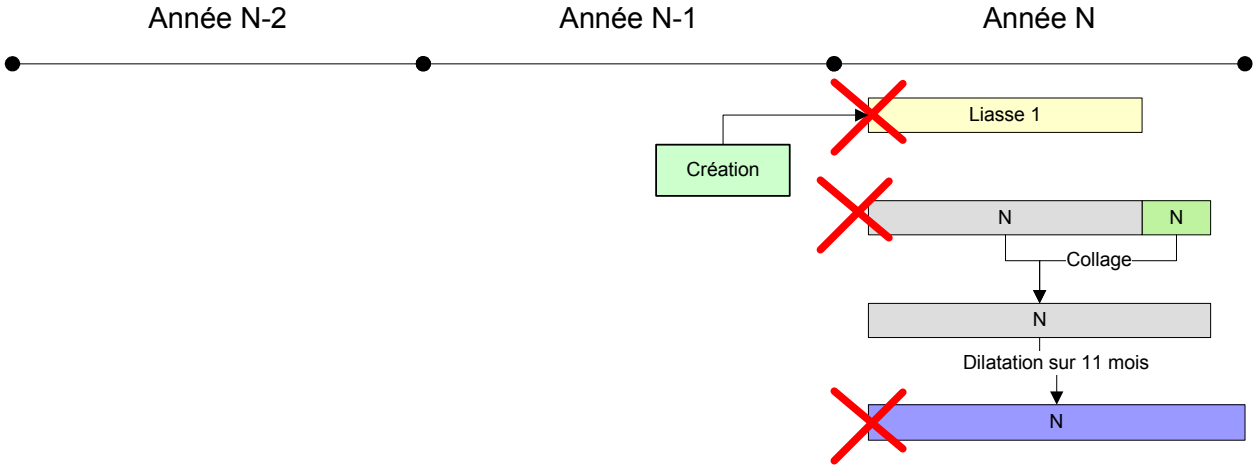
On suppose sur la figure que la création a lieu au 1er février.



La liasse est affectée à l'année N, et dilatée de façon à ce que la durée de la liasse Ésane corresponde à la durée qu'il y a entre la date de création et le 31/12/N.

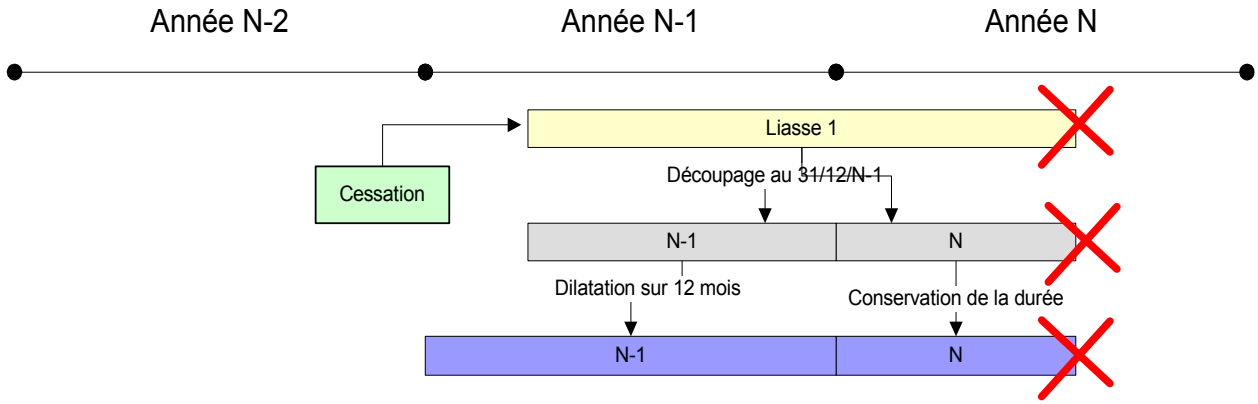
4. *Création d'entreprise ex nihilo : l'entreprise est créée en N et a d'autres liasses principales sur N*

On suppose de même que la création a lieu au 1er février.



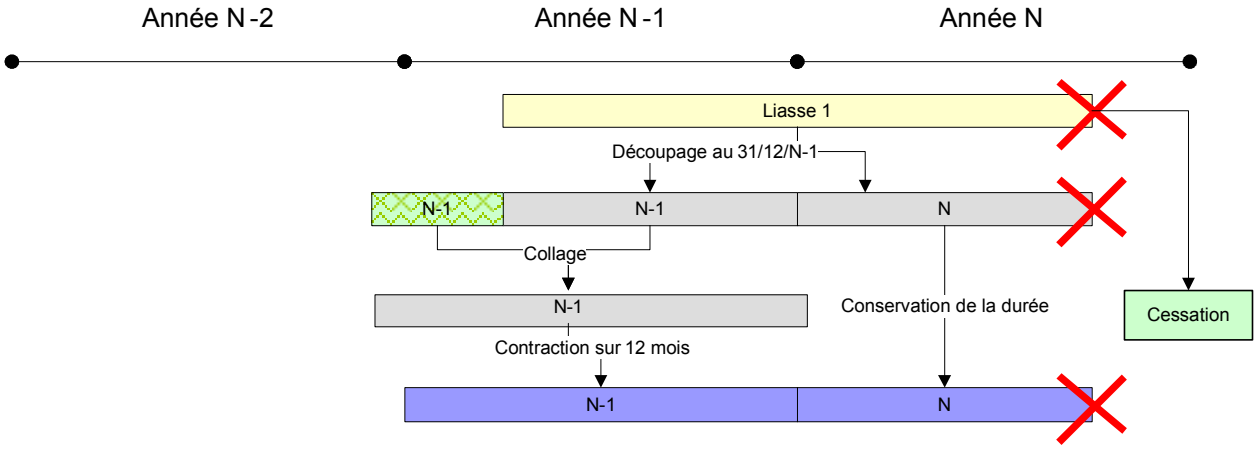
Dans le cas où il y a une autre liasse principale sur N, les deux liasses sont collées avant d'être affectées à l'année N. Le collage est ensuite contracté ou dilaté de façon à présenter une durée égale à la durée qu'il y a entre la date de création et le 31/12/N.

5. ***Cessation d'entreprise sans reprise. La liasse débute en N-1 et il n'y a pas d'autre liasse principale en N-1***



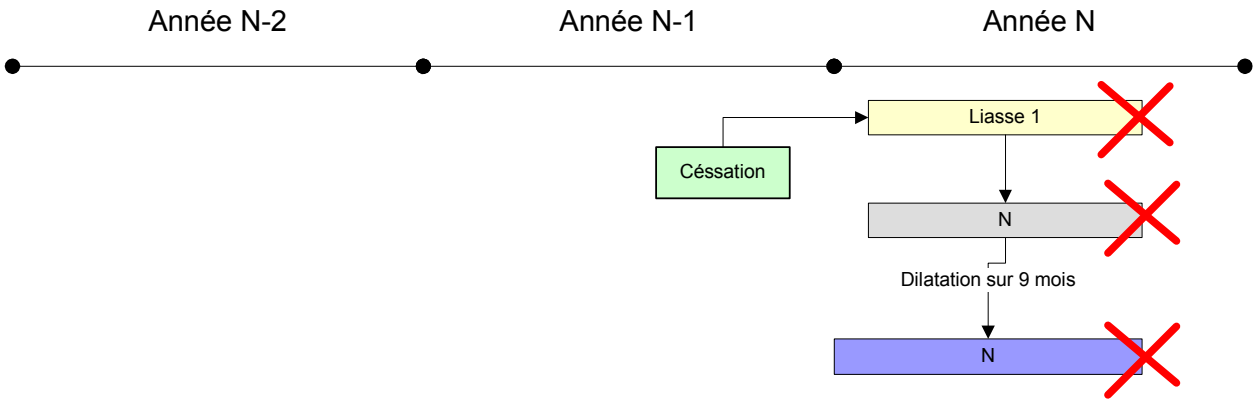
La liasse est découpée en deux parties, au 31/12/N-1. La première partie est affectée en principal à l'année N-1 et dilatée sur 12 mois, la deuxième partie est également affectée en principal à l'année N (cas 4 du § *6.2.7.1*) avec la durée correspondant au découpage (là il ne s'agit pas d'une entreprise pérenne puisqu'elle est cessée en N). La règle de la dilatation ne joue donc pas. En revanche au regard de l'année N-1, l'entreprise est pérenne puisque censée exister en N-2. Le premier morceau doit donc être dilaté).

6. ***Cessation d'entreprise sans reprise. La liasse débute en N-1 et il y a d'autres liasses principales en N-1***



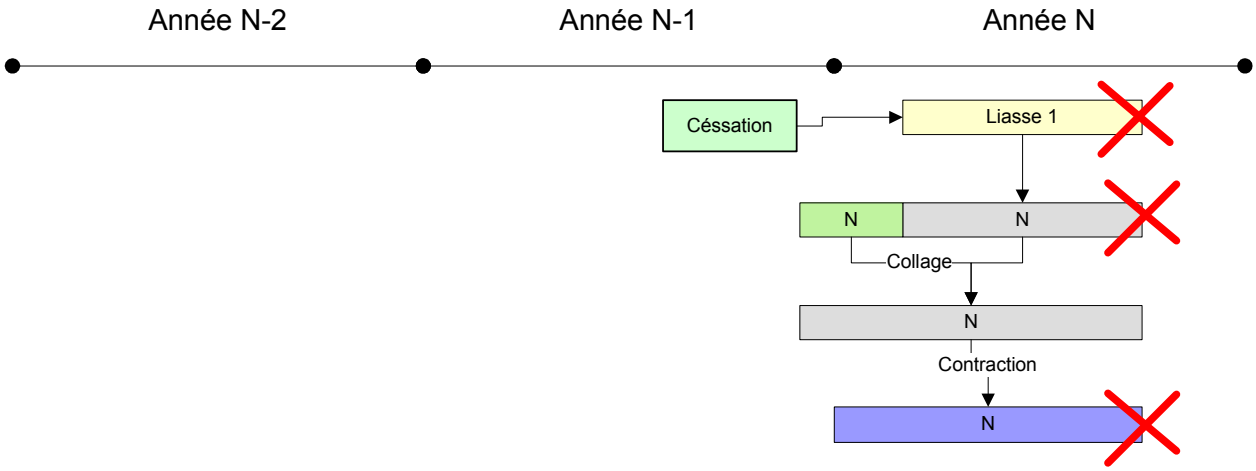
En cas de présence d'une autre liasse principale sur N-1, la première liasse Ésane, issue du découpage de la liasse fiscale au 31/12/N-1, est collée à l'autre liasse N-1. Le collage est affecté en principal à l'année N-1 puis contracté ou dilaté sur 12 mois. La deuxième liasse Ésane issue de la liasse fiscale est affectée à l'année N sans modification de durée.

7. ***Cessation d'entreprise sans reprise. La liasse débute en N et il n'y a pas d'autres liasses principales en N.***



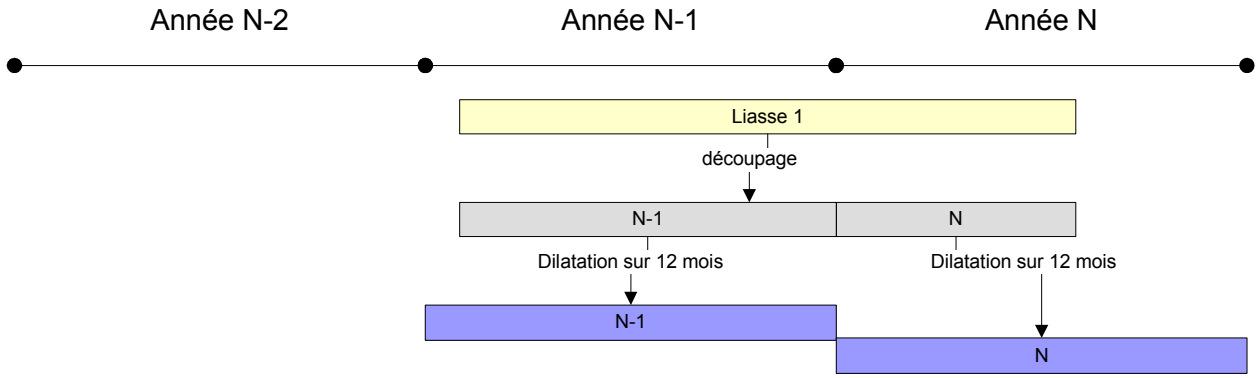
La liasse est affectée à l'année N. Elle est ensuite dilatée de façon à ce que sa durée corresponde à celle qui existe entre le 01/01/N et la date de cessation.

8. ***Cessation d'entreprise sans reprise. La liasse débute en N et il y a d'autres liasses principales en N.***



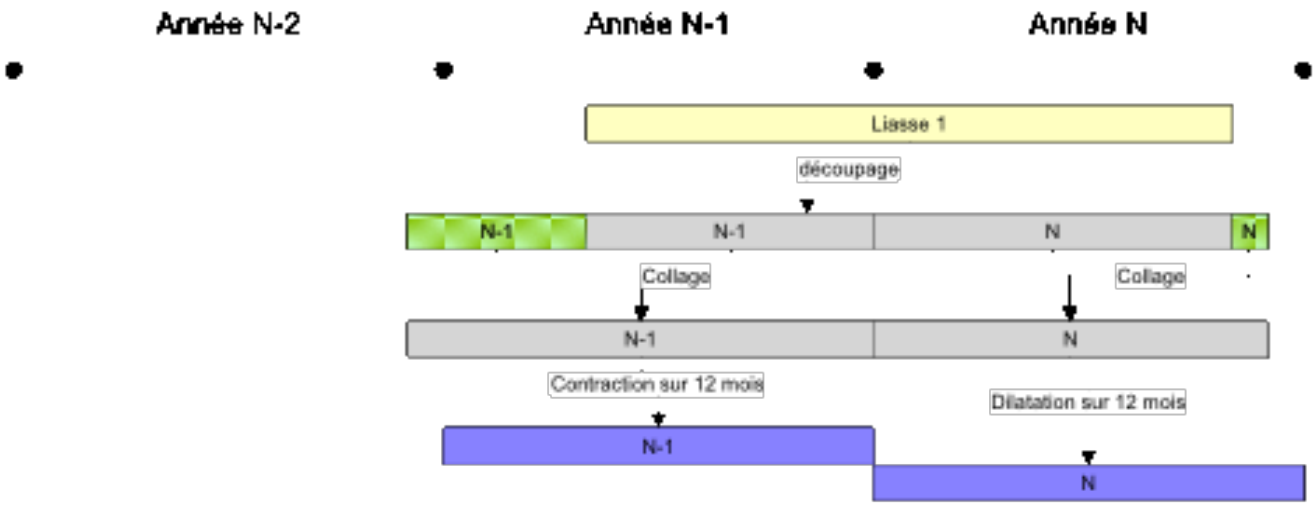
La liasse est affectée à l'année N et collée à l'autre liasse principale sur N. Ce collage est ensuite contracté de façon à présenter une durée égale à la durée qu'il y a entre la date de cessation et le 01/01/N.

9. *Liasse principale sur une année et secondaire sur une autre (hors création et cessation) et absence d'autre liasse principale*



La liasse fiscale est d'abord affectée en principal à l'année N-1 (règle générale) ; puis on s'aperçoit qu'on n'a pas de liasse principale en N ; elle est donc affectée en secondaire à N (Cf. § *6.2.7.2*). Pour cela, elle est coupée en deux au 31 décembre N-1, la première partie est affectée en liasse principale à l'année N-1 et la 2e partie affectée en liasse secondaire à l'année N.
En absence d'autres liasses principales, ces deux liasses Ésane de durée brute inférieure à 12 mois sont dilatées sur 12 mois.

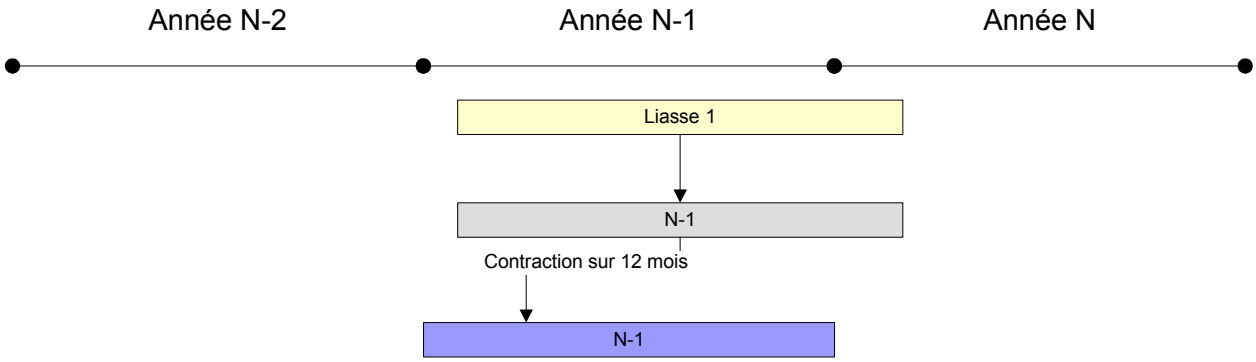
10. *Liasse principale sur une année et secondaire sur une autre (hors création et cessation) et présence d'autres liasses principales sur N-1 et/ou sur N*



Comme dans le cas précédent, la liasse fiscale est scindée en deux au 31 décembre N-1, la première partie est affectée en liasse principale à l'année N-1 et la 2e partie affectée en liasse secondaire à l'année N.
En présence d'une liasse Ésane brute principale sur N-1 (venant d'une liasse DGFiP précédente, ayant éventuellement subi une scission ; en vert sur le schéma), les liasses Ésane de durées brutes[226] affectées à N-1 sont collées et ce collage est ensuite contracté ou dilaté sur 12 mois (sur le schéma contracté, parce que la liasse N-1 commence en N-2).
En présence d'une liasse Ésane de durée brute, principale sur N (venant d'une autre liasse DGFiP), ces liasses de durées brutes sont collées, et le collage est ensuite éventuellement contracté ou dilaté sur 12 mois (sur le schéma dilaté car la deuxième liasse se termine avant la fin de l'année N).

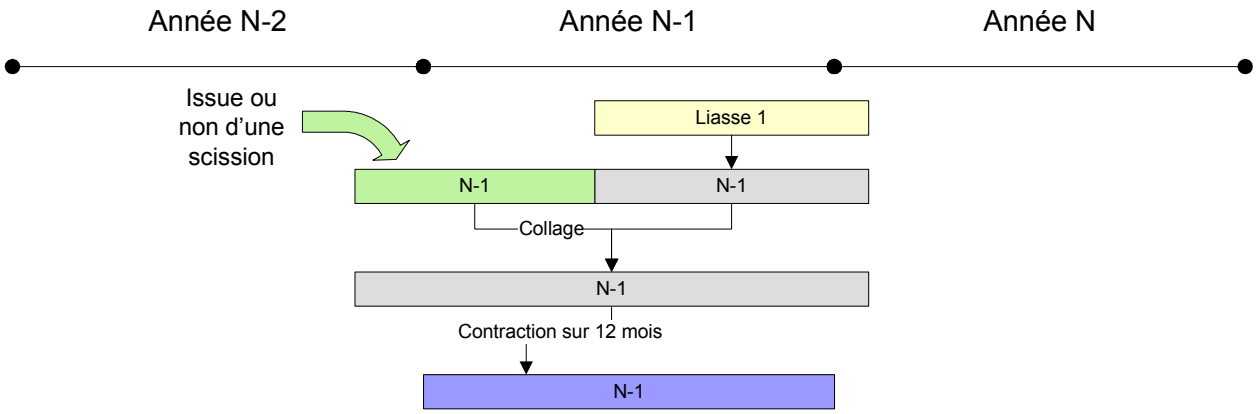
[226] On appelle ainsi les liasses reçues de la DGFiP. A l'issue du processus de découpage, puis de collage et de contraction/dilatation, les liasses n'ont plus du tout les mêmes durées qu'originellement.

11. *Liasse uniquement principale sur une année[227] (hors création et cessation) : Liasse principale sur N-1 et absence d'autre liasse principale sur N-1*



La liasse principale en N-1 est l'unique représentante de l'année N-1, elle est donc contractée ou dilatée sur 12 mois s'il y a lieu.

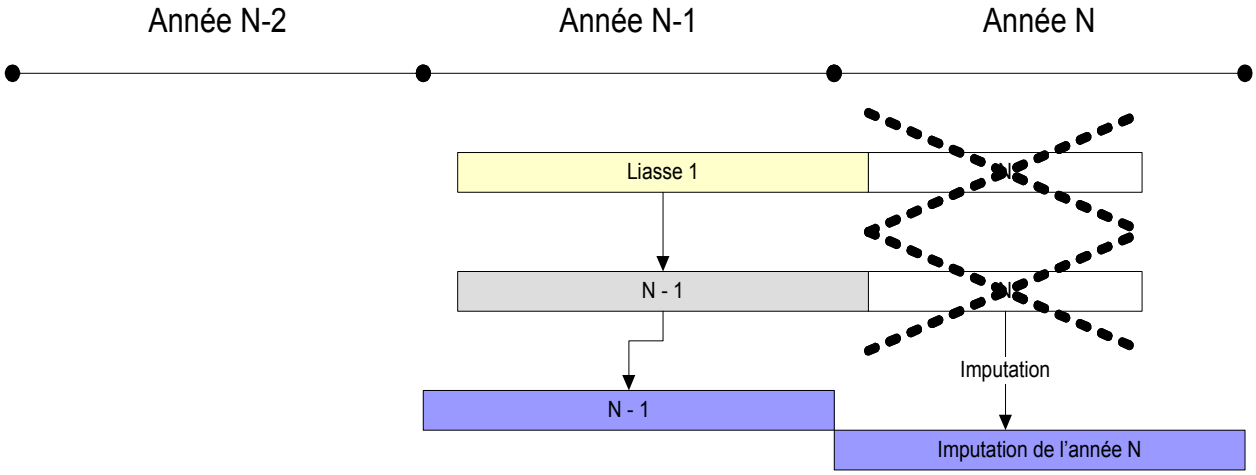
12. *Liasse uniquement principale sur une année (hors création et cessation) : Liasse principale sur N-1 et présence d'autre liasse principale sur N-1*



En présence d'autres liasses principales sur N-1, toutes ces liasses sont collées, puis contractées ou dilatées sur 12 mois s'il y a lieu.

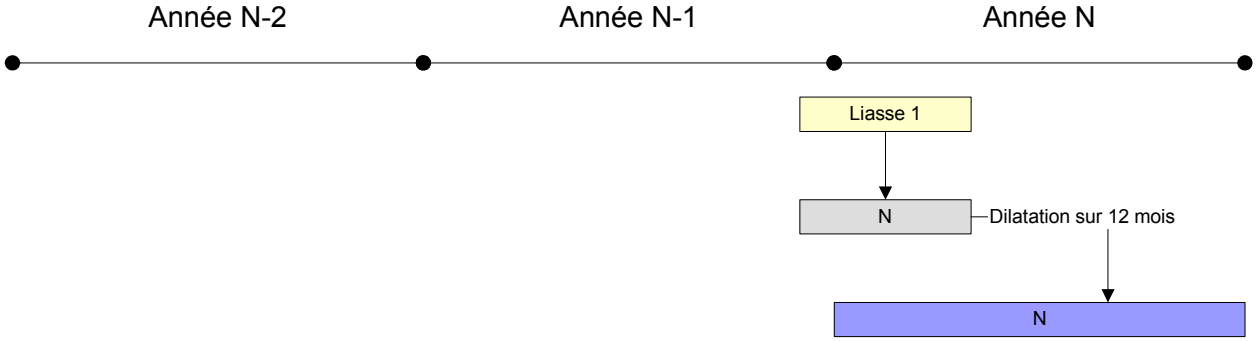
[227] C'est-à-dire qu'elle n'est pas découpée pour être affectée en principal ou en principal et secondaire à deux années.

13. *Liasse uniquement principale sur une année (hors création et cessation) : liasses principales sur N-1 et absence de liasse principale sur N (hors cessation)*



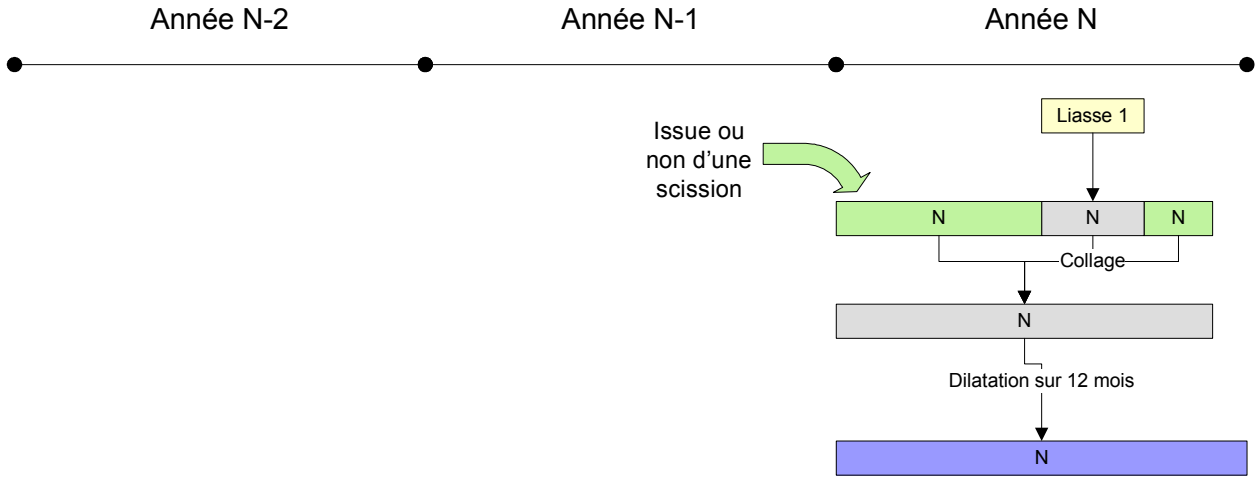
Ce cas est le cas général des liasses des unités légales pratiquant un « exercice décalé » (de mars à mars ou de mai à mai par exemple). On a vu que du temps de Suse, on prenait en compte ces liasses sur l'année N alors qu'aujourd'hui on les reporte sur l'année civile N-1 avec laquelle elles ont le plus de mois en commun.
En cas **d'absence de liasse principale sur l'année N, les données N seront alors imputées**. [En pratique, ce cas est assez rare : si on se réfère en effet aux liasses reçues en 2012 (Cf. début du § 6.2.7), on peut voir qu'on a beaucoup plus de liasses clôturant en mars 2013 (176 896) qu'en mars 2012 (56 298) ; ces dernières sont logiquement affectées en principal à 2011 ; ce n'est que si on ne trouvait pas la liasse correspondante clôturant en mars 2013 qu'on serait amené à l'imputer].

14. *Liasse uniquement principale sur une année (hors création et cessation) : Liasse principale unique sur N*



Quelle que soit la taille de la liasse principale sur N, en l'absence d'autre liasse principale sur N, elle est dilatée ou contractée sur 12 mois s'il y a lieu.

15. ***Liasse uniquement principale sur une année (hors création et cessation) : Liasse principale sur N et présence d'autres liasses principales sur N.***



En cas de multiples liasses principales sur N, ces liasses sont collées, puis dilatées ou contractées sur 12 mois.

6.2.7.4 Conséquences du découpage collage sur les caractéristiques de la liasse

Cette partie décrit les principes généraux des opérations nécessaires aux traitements d'Ésane pour prendre en compte les durées d'exercice différentes de 12 mois. Il existe **quatre types d'opérations** distincts au niveau du calcul des caractéristiques des liasses :

- les collages ;
- les découpages ;
- les dilatations / contractions ;
- les pseudo-consolidations dans le cas du collage de deux liasses principales de la même année (cas des réponses d'une même unité dans deux régimes fiscaux différents).

Typologie des caractéristiques comptables :

Grâce à une table de correspondance, les caractéristiques d'Ésane sont classées en **quatre grandes catégories** :

- A : les flux, qui se décomposent eux-mêmes en 2 sous-catégories selon le positionnement de 2 unités en cas de restructurations :
  1. flux additifs (ex : les rémunérations)
  2. flux non additifs de type produit (ex : le chiffre d'affaires sur la période)
  3. flux non additifs de type charge (ex : les achats de marchandises sur la période)

  Pour chaque flux non additif de type produit, il existe un flux non additif de type charge qui lui correspond et la différence de ces deux flux non additifs est un flux additif.
- B : les niveaux de stocks de l'exercice, qui se décomposent en :
  1. stocks de début (ex : immobilisations en début d'exercice)
  2. stocks de fin (toutes les variables du bilan après affectation du résultat)

  Pour chaque stock en ouverture et il existe un stock en clôture avec la relation :
  stock ouverture (N)=stock clôture (N-1) (principe de continuité comptable).
- C : les niveaux de stock à une date donnée (l'effectif au 31/12 de l'année de référence)
- D : les moyennes sur la période d'exercice (nombre moyen de salariés)

On va caractériser chacune des quatre opérations par le traitement qui est fait dans chacune d'elle des quatre catégories de variables
En termes de vocabulaire, on utilisera le terme de **liasse(s) "avant"** pour désigner la(les) liasse(s) avant l'opération de collage, découpage, dilatation etc. et de **liasse(s) "après"** pour désigner les liasses obtenues par l'opération.

Pour les collages on appellera **liasse prédécesseur** et **liasse successeur** les deux liasses "avant" que l'on veut coller en une liasse unique (prédécesseur et successeur se rapportent aux périodes sur lesquelles portent ces liasses et non avant et après l'opération de collage elle-même). Les mêmes adjectifs sont utilisés pour les deux liasses "après" obtenues par un découpage.

1. Les collages

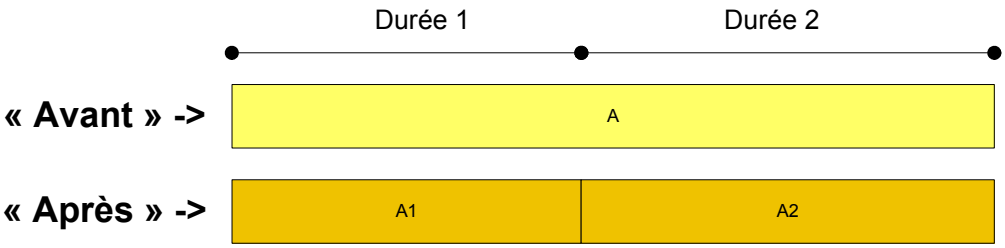
« Avant » ->

« Après » ->

Type de la caractéristique	Opération à réaliser pour le collage, pour chaque caractéristique
A1 et A2 et A3	addition des données du prédécesseur et du successeur
B1	on conserve le stock d'ouverture du prédécesseur et supprime celui du successeur. Un test préalable entre le stock d'ouverture du successeur et le stock de fermeture du prédécesseur qui lui correspond est réalisé et on affecte les différences non nulles à des apports ou des réévaluations pour certaines variables (immobilisations)
B2	on conserve le stock de clôture du successeur et on supprime celui du prédécesseur
C	Inchangé
D	on fait une somme pondérée par les durées d'exercice

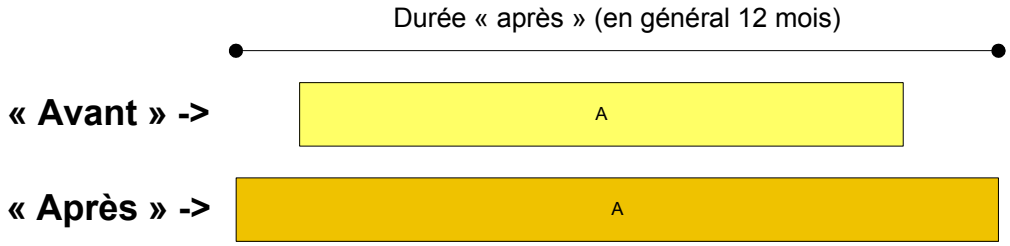
2. *Les découpages*

On doit indiquer en préalable les durées respectives (en mois) des deux liasses "après" (leur somme faisant naturellement la durée de la liasse "avant") et l'on fait une hypothèse que l'activité est la même sur tous les mois de la période et les « prorata » calculés par la suite sont donc obtenus par rapport des durées d'exercices calculées (en mois).



Type de la caractéristique	Opération à réaliser pour le collage, pour chaque caractéristique
A1 et A2 et A3	Les variables flux sont coupées en deux au prorata de la durée d'exercice des deux liasses "après"
B1	On va « proratiser » les variations de stock comme si elles étaient linéaires en fonction du temps. C'est certainement faux mais on ne peut guère faire d'autres hypothèses. Pour la liasse prédécesseur, on conserve les informations de la liasse "avant". Pour la liasse successeur on procède comme suit : 1. on calcule la différence entre le stock de fermeture et le stock d'ouverture de la liasse "avant" 2. on effectue la même règle de trois que pour le « A1 à A3 » ci-dessus à cette différence 3. on ajoute au stock de début d'exercice de la liasse "avant" la variation de stock dilatée précédemment pour obtenir le stock d'ouverture du successeur.
B2	Pour la liasse prédécesseur on procède comme suit : 1. on calcule la différence entre le stock de fermeture et le stock d'ouverture de la liasse "avant" 2. on effectue la même règle de trois que pour le « A1 à A3 » ci-dessus à cette différence 3. on ajoute au stock de début d'exercice de la liasse "avant" la variation de stock dilatée précédemment pour obtenir le stock de clôture du prédécesseur. Pour la liasse successeur, on conserve les informations de la liasse "avant".
C	on conserve le même chiffre "avant" pour les deux liasses "après"
D	on conserve le même chiffre "avant" pour les deux liasses "après"

3. *Les dilatations/contractions*



Type de la caractéristique	Opération à réaliser pour la dilatation, pour chaque caractéristique
A1 et A2 et A3	multiplier par durée "après" et diviser par durée "avant"
B1	On procède comme suit : 1. on calcule la différence entre le stock de fermeture et le stock d'ouverture de l'exercice "avant" 2. on la divise par la durée d'exercice de la liasse "avant" et on la multiplie par la durée séparant l'ouverture "avant" moins l'ouverture "après" 3. on ajoute au stock d'ouverture d'exercice "avant" la variation de stock dilatée précédemment pour obtenir le stock d'ouverture "après" 4. les variables de stock à la date d'ouverture sont inchangées
B2	On procède de façon symétrique à la procédure ci-dessus en partant des stocks et dates de clôture.
C	on ne conserve que la valeur "avant" qui correspond à la date considérée (31/12/n) si n est l'année de référence
D	on conserve la valeur "avant"

Il est à noter que ce traitement de contraction/dilatation ***qui est nouveau par rapport aux pratiques de Suse antérieures***, est à faire deux fois sur les liasses N différentes de 12 mois :

- la 1ère année, pour la campagne N, en conservant ses stocks ouverture, pour faire le lien avec la liasse N-1
- et la campagne suivante N+1, s'il ne s'agit pas d'une entreprise cessée, pour faire le lien entre ses stocks clôture et ceux d'ouverture de la liasse N+1.

Concrètement cette opération de dilatation contraction est effectuée dans la « chaîne aval » qui enchaîne toute une série de traitements avant les macro-contrôles (remise à 12 mois, traitement des liasses multiples, des dates, des restructurations : Cf. § *6.4.2* et suivants).

Pour les macro et micro-contrôles
Ce sont systématiquement les données ramenées à 12 mois qui sont utilisées (mais l'on affiche les données sur durée brute pour le gestionnaire au cas où il souhaite dialoguer avec l'entreprise sur les données qu'elle a fournies, avec possibilité d'afficher aussi à 12 mois s'il le souhaite).

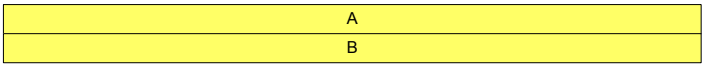
Pour la diffusion : le problème en diffusion est de savoir si l'on doit utiliser la donnée sur la durée brute ou sur la donnée ramenée à 12 mois.
La règle est la suivante (selon le raisonnement explicité au début du § *6.2.7.3*) :

- S'il s'agit d'une création ou d'une cessation, il faut prendre en compte les données correspondant **à la durée brute** dans les agrégats de diffusion.
- S'il s'agit d'une entreprise pérenne, il faut prendre en compte les données correspondant **à la durée ramenée à 12 mois** dans les agrégats de diffusion.

4. *Les pseudo-consolidations*

Dans le cas d'une déclaration dans deux régimes différents, ces dernières sont, dans un premier temps, éventuellement ramenées sur 12 mois (cf. dilatation ou contraction décrite ci-dessus) car la pseudo consolidation ne peut se faire que sur des liasses de même durée donc après calage sur 12 mois.

« Avant » ->



« Après » ->

Type de la caractéristique	Opération à réaliser pour la pseudo-consolidation, pour chaque caractéristique
A1	somme simple des liasses "avant"
A2	somme des liasses "avant" et multiplication par le coefficient d'agrégation. La différence entre la somme simple et la somme pondérée par le coefficient d'agrégation est conservée
A3	somme des liasses "avant" et on déduit la différence calculée ci-dessus pour la caractéristique de type produit qui est en relation avec la caractéristique de type charge traitée ici
B1	somme simple des liasses "avant"
B2	somme simple des liasses "avant"
C	somme simple des liasses "avant"
D	somme simple des liasses "avant"

6.2.7.5 Bilan de l'étape de découpage collage (avant dilatation-contraction ou pseudo consolidation)

On peut observer sur les données de la campagne 2012 quel est le bilan de l'ensemble des opérations de découpage collage. Les ordres de grandeur sont inchangés d'année en année. Le tableau qui suit montre à quel point cette étape est fondamentale dans le traitement des liasses.

L'opération de découpage d'une liasse crée deux liasses à partir d'une liasse initiale. L'opération de collage crée une liasse à partir de deux liasses initiales. Le tableau montre que **le solde net** de cette opération de création et de suppression est **positif d'environ 80 000 liasses** : on crée nettement plus de liasses qu'on n'en supprime.

Mais surtout on voit que sur près de 4 millions de liasses reçues au total, on n'en conserve finalement qu'un peu moins de 3 : on **laisse donc tomber de l'ordre d'un million de liasses** ! Celles-ci se composent :

- Des liasses qui sont affectées en principal en 2011 : il s'agit de toutes les liasses qui clôturent en 2011, et de la plupart des liasses qui clôturent avant juin 2012 qui, si elles ont une durée d'exercice de 12 mois, ont plus de mois sur 2011. Toutefois, il peut bien évidemment rester des liasses qui clôturent dès janvier 2012 (correspondant aux cessations de ce mois), ou des liasses de durée inférieure à 12 mois d'exercice qui clôturent début 2012 (par exemple une liasse de 6 mois de novembre 2011 à avril 2012).
- De liasses qui clôturent entre juin 2012 et mai 2013 mais qui sont abandonnées quand même parce qu'elles sont ensuite recollées avec une autre (rappelons qu'il ne peut y avoir finalement qu'une seule liasse par numéro Siren et par campagne), ou qui sont contenues dans une liasse plus longue ou qui sont en double etc. Mais c'est évidemment sur cette

période qu'on a le plus de liasses actives relativement à la campagne 2012. La fraction de liasses actives finalement retenues qui clôturent au 31/12 est de 70,1%.

- Des liasses qui sont affectées en principal à 2013 : on constate logiquement que sur plus de 180 000 qui clôturent en juin 2013, on n'en garde que très peu (6410) : les autres sont soit collées avec d'autres liasses, soit le plus souvent conservées pour être utilisées dans la campagne 2013 qui suivra. Les quelques liasses actives qui clôturent à partir de juin 2013 sont a priori des liasses qui ont des durées d'exercice supérieures à 12 mois et davantage de mois en 2012 qu'en 2013 (par exemple « la » liasse qui clôture en novembre 2013 est une liasse de 23 mois qui a donc effectivement plus de mois en 2012 qu'en 2013).

Rappelons qu'après toutes ces opérations de collage, de découpage et de réaffectation sur la bonne année de campagne, il y a encore une étape de contraction ou de dilatation qui n'est pas prise en compte dans les comptages ci-dessous et qui sera effectuée dans la chaîne aval (Cf. § *6.4.2*) :

date de clôture de la liasse	Nombre de liasses brutes DGFIP	Nombre de liasses créées par le DC (*)	Nombre de créations nettes	Nb liasses actives après DC
2011 01	44	50	6	0
2011 02	6	14	8	0
2011 03	60	132	72	0
2011 04	65	151	86	0
2011 05	13	32	19	0
2011 06	13	23	10	0
2011 07	4	7	3	0
2011 08	4	7	3	0
2011 09	16	17	1	0
2011 10	3	3	0	0
2011 11	1	2	1	0
2011 12	324	377	53	0
2012 01	8,622	8,912	290	2728
2012 02	6,873	7,215	342	1951
2012 03	53,589	56,298	2,709	7947
2012 04	19,049	20,582	1,533	3729
2012 05	14,584	15,508	924	4030
2012 06	237,976	251,951	13,975	196570
2012 07	47,540	49,711	2,171	38886
2012 08	66,587	69,719	3,132	55779
2012 09	285,162	297,807	12,645	245428
2012 10	39,284	41,007	1,723	36672
2012 11	18,438	19,377	939	17146
2012 12	2,365,454	2,393,595	28,141	2003487
2013 01	23,002	23,386	384	20292
2013 02	18,173	18,572	399	15148
2013 03	172,199	176,896	4,697	156091
2013 04	32,569	33,201	632	26638
2013 05	20,289	20,761	472	15193
2013 06	176,739	183,072	6,333	6467
2013 07	28,411	28,992	581	584
2013 08	41,986	42,568	582	582
2013 09	102,487	102,953	466	466
2013 10	5,121	5,136	15	14
2013 11	1,442	1,443	1	1
2013 12	2,901	2,903	2	0
Total	**3,789,030**	**3,872,380**	**83,350**	**2,855,829**

(*) : Découpage - Collage

6.2.8 Micro contrôles internes

Cette phase a lieu après les traitements de découpage collage qu'on vient de voir (et donc sur les données qu'on a finalement calculées à l'issue de ces traitements). Il s'agit quasiment des mêmes traitements qui étaient effectués dans Suse[228] mais qui ont été ***intégralement réécrits****.

6.2.8.1 La logique des contrôles et redressements automatiques de cette phase

1) **Les micro-contrôles**[229] **internes** permettent de vérifier que les **équilibres comptables sont bien respectés**. Un indicateur, auquel est attribuée une note qui va de 1 à 9, est associé à chacun des contrôles. La note attribuée à l'indicateur est d'autant plus **élevée** que **l'erreur constatée est importante**. Il s'agit en fait de contrôles internes de cohérence.

2) Le programme **RQSI** (Redressements Quasi Sûrs Internes) permet de redresser de manière quasi-sûre des erreurs détectées lors du passage des micro-contrôles. À la fin de ce programme, les micro-contrôles internes sont relancés. Si des erreurs persistent, **le programme RIA sera lancé**.

3) Le programme **RIA** (Redressements Internes Automatiques) permet de redresser de manière brutale les erreurs qui persistent. À la fin de ce programme, les micro-contrôles internes sont relancés pour vérifier qu'il ne reste aucune incohérence interne. **S'il en reste**, elles seront **traitées par les gestionnaires** lors de la **chaîne aval**.

6.2.8.2 Le principe des contrôles internes

- Plusieurs niveaux de contrôle

On distingue :

a- Les contraintes d'intégrité référentielle
b- La cohérence des variables ou la mesure des écarts dans le cas :
- des équilibres comptables ;
- du contrôle de la vraisemblance.

a. Les contraintes d'intégrité "référentielle"

Ces contraintes d'intégrité ne sont pas à proprement parler des contrôles. Il n'y a pas de possibilité d'erreur. Au niveau de certaines entités, le contenu peut varier suivant la structure. Exemple type, dans le compte de résultat le concept "achat d'animaux" n'a de sens que pour les régimes agricoles, il ne peut y avoir de données associées à ce concept dans les autres régimes.
Ce contrôle permet donc d'**identifier**, pour un régime donné, **les variables à contrôler**.

b. La cohérence des variables

Il y a deux grandes familles de micro contrôles internes, les équilibres comptables et les contrôles dits de vraisemblance qui s'accrochent sur les relations économico-comptables pouvant lier des données entre elles (inégalité entre variables). Le premier problème qui se pose est celui de la mesure des écarts. Les réflexions faites dans le cadre de Suse avaient conduit à privilégier des mesures d'écart absolu plutôt que des écarts relatifs. Ceci est conforme à la démarche de Suse qui tend à corriger prioritairement les grandes entreprises, là où les écarts absolus sont les plus importants et peuvent perturber le plus les comptes des secteurs. Ce principe a donc été conservé pour cette phase.

[228] Les premières exploitations de Suse remontent à une période où les données des liasses fiscales étaient saisies au départ par des dactylo-codeuses. Il s'ensuivait un certain type d'erreurs (décalages, erreurs de signes, permutations de chiffres) que ces programmes essaient de corriger de façon automatique.
[229] Micro : au sens de « portant sur des données individuelles » et par opposition aux macro-contrôles qui portent sur des agrégats. Il ne s'agit en aucun cas de « petits » contrôles !

- La mesure des écarts

Les principales relations économico-comptables sont de la forme :

1. A=B
2. A≤B (ou A<B)
3. A≥B (ou A>B)

Il s'agit au niveau des contrôles de mesurer les écarts entre A et B. Chacun de ces termes A ou B représente une donnée comptable ou une expression arithmétique de plusieurs données comptables.

1. A=B peut s'écrire A-B=e1
En théorie e1 est nul, en pratique une faible valeur pour e1 sera jugée négligeable, une très forte valeur pour e1 conduira impérativement à corriger A et/ou B. Dans ce cas e1 peut être positif ou négatif, la mesure portera donc sur la valeur absolue de e1.

2. A≤B peut s'écrire A-B=e2 avec e2≤0
Dans ce cas la mesure de l'écart n'a d'intérêt que s'il y a une anomalie, c'est-à-dire si e2>0. On mesurera donc l'anomalie au travers de l'expression Sup(e2, 0).

3. A≥B peut s'écrire A-B=e3 avec e3≥0
Cas inverse du précédent, la mesure n'a d'intérêt que si e3<0. On mesurera donc l'anomalie au travers de l'expression -Inf(e3, 0).

On rapproche ensuite cet écart d'un **plafond maximum** (MAXI) **pour obtenir une mesure en %** par rapport à ce plafond. Puis suivant le % obtenu, une table donne une valeur pour l'indicateur d'erreur associé à ce contrôle.

La table la plus utilisée est la suivante :
MAXI=1 500 000 euros

Écart en euros	Écart en %	Indicateur d'erreur
[0, 1 500 [	[0 ; 0,1 [	1
[1 500, 15 000 [	[0,1 ; 1 [	2
[15 000, 30 000 [	[1 ; 2 [	3
[30 000, 75 000 [	[2 ; 5 [	4
[75 000, 150 000 [	[5 ; 10 [	5
[150 000, 300 000 [	[10 ; 20 [	6
[300 000, 750 000 [	[20 ; 50 [	7
[750 000, 1 500 000 [	[50 ; 100 [	8
[1 500 000, +∞ [	[100 ; +∞ [	9

En outre dans le programme, quatre notes jouent un rôle très important :
Dans le programme des RQSI :

- La note 3 est associée au seuil de redressement.
  Un contrôle interne ne peut déclencher un redressement (donc un RQSI) que si sa note est supérieure ou égale à 3.
- La note 2 est associée au seuil d'égalité. Un essai au sein d'un RQSI (Cf. ci-dessous § *6.2.9.2*) est « validé » si tous les indicateurs associés aux contrôles internes définis au préalable sont inférieurs à 2.

Dans le programme des RIA :

- La note 1 est associée au seuil de cohérence. Un contrôle interne ne peut déclencher des corrections automatiques que si l'indicateur, qui lui est associé, est strictement supérieur à 1.
- La note 9 est associée au seuil d'incohérence. Un contrôle interne ne peut déclencher des corrections automatiques que si l'indicateur, qui lui est associé, est strictement inférieur à 9.

6.2.9 Les RQSI (redressements quasi sûrs internes)

Pour les différents comptes de la liasse fiscale (compte de charge, compte d'immobilisation, compte de résultats etc..), le sous-programme des RQSI est composé de **modules** qui s'intéressent aux

égalités comptables qui caractérisent ces comptes. Chaque module permet de **redresser une égalité comptable qui n'est pas respectée** en « testant » des hypothèses de causes d'erreurs qui pourraient expliquer le non-respect de l'égalité (du type changement de signe, décalage de zones etc.). Il va donc **essayer** successivement un certain nombre de corrections. Si l'une d'elles « marche », compte tenu de la très faible probabilité que ceci soit le fruit du hasard, on va considérer qu'on a trouvé **la** cause de l'erreur et modifier le compte en conséquence ; c'est pourquoi on parle de « redressements « quasi sûrs ». Si aucune ne marche, on va laisser le compte inchangé et on passera au programme de redressement interne automatique.

6.2.9.1 La logique des RQSI : une succession de modules

Chaque module est rattaché à un **contrôle**, qui teste le respect ou non de cette égalité comptable. Si le contrôle indique que l'égalité comptable n'est pas respectée, le module, **qui regroupe l'ensemble des essais qui tentent de rétablir cette égalité comptable**, sera déclenché.

L'exemple suivant sera suivi tel un fil rouge tout au long de ce paragraphe.

Le module RQSIB006 permet de redresser l'égalité comptable suivante, si elle n'est pas respectée.
Total de l'actif brut (B100) - Total des amortissements et des provisions (B200) = Total du passif (B300)

Soit B100 – B200 = B300

Le contrôle rattaché au module RQSIB006 est le contrôle interne (CI) CIB006. Il teste donc si la relation suivante est respectée : B100 - B200 = B300. Si ce n'est pas le cas et que l'écart constaté entre les deux membres est supérieur au montant 15 000 euros, la note attribuée au contrôle devient supérieure à 2 (note associée au seuil d'égalité) ce qui déclenche le module RQSIB006.

Un module est composé d'un ou de plusieurs « essais ». L'utilisation du terme essai n'est pas anodine. Elle indique bien que, lors des redressements, on teste **différentes hypothèses quant à l'origine de l'erreur**, mais que jamais on n'impose une solution et ce, contrairement aux redressements automatiques internes qu'on verra ensuite. Si les hypothèses supposées quant à l'origine de l'erreur s'avèrent non vérifiées, on laisse le compte en l'état.
L'erreur sera alors supprimée par les redressements internes automatiques.

Cette méthode de fonctionnement est très importante puisqu'elle explique la raison d'être des RQSI : les RQSI ne concernent que les données de la liasse fiscale pour lesquelles la cause de l'erreur est parfaitement identifiée.

6.2.9.2 Essai : on teste une hypothèse

La logique de l'essai est la suivante : **si normalement A = B, que je suis certain de A et que A n'est pas égal à B, je suis certain que l'erreur provient de B**.
Avant de lancer un essai, on s'assure donc que l'un des membres de la relation est correct (c'est-à-dire qu'il y a au moins un autre contrôle relatif à cette variable qui a une note égale à 1). Nous appellerons ***appui*** l'ensemble des variables qui constituent ce membre. Cet appui nous permet d'identifier le membre sur lequel porte l'erreur mais non l'erreur en tant que telle. Si ce membre est composé de plusieurs variables, nous ne sommes pas en mesure de dire quelle est la variable qui est en erreur. Cela ne suffit donc pas.
A ce stade, on ne peut que faire des suppositions sur la situation dans laquelle on se trouve :

- Une variable qui est théoriquement une somme de sous-montants n'est pas égale à la somme de ces sous-montants. On peut essayer de la remplacer par la somme des sous-montants.
- La véritable valeur d'une variable est celle qui est inscrite dans un autre compte. (Les montants de certaines variables sont demandés plusieurs fois dans la liasse fiscale). On peut essayer de la remplacer par cette autre valeur.
- Une variable est renseignée alors qu'elle devrait être nulle. On peut tester la valeur zéro.
- Il y a eu un décalage de ligne. On peut essayer de supprimer le décalage.
- Une variable est négative (respectivement positive) à tort. On peut tester de modifier son signe.

Presque tous les essais testent une de ces éventuelles situations ou une combinaison de ces dernières.
Un **essai est validé** si le **nouveau montant du membre B est égal au membre A** ou que l'écart entre ces deux membres est inférieur à un certain montant.

Exemple :
Illustration avec le premier essai du module RQSIB006.

Le point d'appui est le membre de l'équation B300 (on suppose donc le total du passif juste).

- Le premier essai proposé pour corriger l'erreur repose sur l'hypothèse que la somme des sous-totaux de l'une des deux variables B100 ou B200, n'est pas bonne. La solution consiste donc à recalculer ces sommes.
  ➔ B100=stB100 et B200=stB200[230]
  Une fois les modifications faites, tous les contrôles sont relancés. Elles seront validées si jamais le membre B100-B200 est égal au membre B300.
- Si le contrôle indique que l'égalité n'est pas faite, l'essai suivant sera lancé. L'essai suivant correspond au fait que le sous-total B104 est renseigné à tort. Un recalcul de B100 avec B104=0 a donc lieu et les contrôles sont relancés.
- On itère avec tous les essais que contient le module RQSIB006.

Fait très important : si une variable sert de point d'appui lors d'un redressement, elle est considérée comme correcte dans la suite des RQSI. Parallèlement une variable modifiée ne peut plus l'être et est considérée comme correcte dans la suite des RQSI.

6.2.10 Les RIA (redressements internes automatiques)

Comme leur nom l'indique, l'objectif des RQSI était de faire des redressements dont on peut se dire "quasi certains". À l'issue de ces derniers, des incohérences sur la liasse fiscale peuvent bien entendu persister puisqu'on n'est pas du tout assurés d'avoir trouvé la nature de l'erreur. L'objectif des RIA est de les faire disparaître en proposant des redressements "moins sophistiqués" et donc plus systématiques.

6.2.10.1 Les deux types de cohérence rétablies par les RIA : la cohérence entre les comptes et la cohérence au sein d'un compte

- La cohérence entre les comptes

Certains comptes ont des variables en commun. On pense notamment au résultat comptable présent dans le compte de résultat, le bilan et le compte de détermination du résultat fiscal. L'un des premiers objectifs des redressements internes automatiques est de faire en sorte que la cohérence entre les comptes sur ces variables soit respectée.

- La cohérence au sein d'un compte

Le second objectif des redressements internes automatiques va être de rétablir la cohérence au sein de chacun des comptes. Les incohérences peuvent être de trois types :

- un total non égal à la somme des variables qui le composent (exemple : total non égal à la somme de ses sous-totaux ⇔ B200 ≠ B202+B203)
- une inégalité non respectée (Ex. les amortissements des immobilisations inscrites au bilan doivent être inférieurs aux immobilisations brutes ⇔ B214 ≤ B114)
- si une variable est différente de zéro, l'autre est automatiquement nulle (cf. le résultat comptable est soit déficitaire soit bénéficiaire ⇔ Si F100[231]>0 alors F200=0 et si F200>0 alors F100=0)

6.2.10.2 Les méthodes de redressements

Les redressements sont de type **déterministes** ou **équilibrage.** Ils utilisent au mieux les variables « **validées** » (c'est-à-dire les variables qui ont des **contrôles associés corrects**). À l'issue de chaque redressement, le contrôle en erreur à l'origine du redressement est relancé. S'il est corrigé, on passe au redressement suivant, dans le cas contraire on procède à un autre redressement.

Au sein des RIA, les variables ont été divisées en deux groupes : les variables dites **principales** et les variables dites **secondaires**.

[230] stB100 et stB200 sont des conventions d'écriture : stB100 regroupe la somme de tous les sous-totaux de B100 dans la liasse fiscale. Il est donc équivalent d'écrire stB100 ou B101+B102+B103+B104+B105+B106. Il en va de même pour stB200 et de tous les totaux précédés de *st*.
[231] F100 est le bénéfice comptable et F200 la perte comptable.

Ce sont d'abord les variables « inter-comptes » qui sont redressées puis dans un second temps les comptes sont redressés les uns après les autres.
En outre, si les variables principales d'un compte ne sont pas cohérentes, alors les variables secondaires de ce même compte ne peuvent pas être redressées.

Un exemple de redressement : le chiffre d'affaires et sa ventilation (fiscale) :

Dans un premier temps, si le chiffre d'affaires est égal 0, on le recalcule à l'aide des postes détails non nuls du tableau de Chiffre d'Affaires. Les contrôles sont relancés.
Dans un deuxième temps, s'il persiste une erreur sur le tableau de Chiffre d'Affaires, on équilibre le tableau en se basant sur le chiffre d'affaires redressé.
Enfin, s'il y a une incohérence au niveau du total des produits, on reporte la différence sur les totaux des produits ainsi que sur les charges pour maintenir l'équilibre du compte.

Détaillons ces calculs :

Premier temps : le chiffre d'affaires (R310) est nul
Si R310=0, on tente de le recalculer de la manière suivante :

- Si la colonne « Total des composants du CA » n'est pas cohérente avec le CA nul, alors on privilégie cette ventilation en vente de biens (R402), de marchandises (R401), de services (R403) pour recalculer le total R310 = R401 + R402 + R403
- Sinon, si la colonne « Total » est nulle, mais que la ventilation France/Export, elle, est non nulle, on utilise cette dernière pour calculer le chiffre d'affaires :
  R310=R410+R420
- Si ces deux ventilations sont nulles, on s'appuie sur d'autres variables non nulles du tableau[232] pour calculer les sous totaux et le chiffre d'affaires total.

On relance alors les contrôles qui sont impactés par les modifications faites ci-dessus.

Deuxième temps : équilibrage du tableau à partir de R310
S'il y a une incohérence en colonne « Total » c'est-à-dire si R310 ≠ R401 + R402 + R403 et/ou si R310≠R410+R420, alors :

S'il y a une incohérence en ligne France/Export alors :

- on reporte le montant du chiffre d'affaires sur un des postes biens, marchandises, services en fonction du code APE[233]. La modification est effectuée pour les deux colonnes « Total » et « France », la colonne « Exportations » étant mise à 0.

Sinon:

- si les lignes France export sont cohérentes ; on se contente de recalculer les postes :
  R401 = R411+R421
  R402 = R412+R422
  R403 = R413+R423
- Sinon on reporte le chiffre d'affaires sur un montant de la colonne « Total » selon l'APE, et on équilibre en conséquence sur les variables du détail.

Sinon (la colonne « Total » est cohérente) :

S'il y a incohérence en ligne France/Export alors :

- Si l'incohérence est totale alors : on reporte les montants de la colonne « Total » sur la colonne « France », et les Exportations sont mises à 0.
- Si la ventilation « française » est correcte alors toutes les variables de la colonne « Exportations » sont recalculées en ligne : CA Exportation = CA total - CA France
- Si la ventilation « Export » est correcte, alors toutes les variables de la colonne « France » sont recalculées en ligne : CA France = CA total - CA Exportation

Sinon (cohérence du total de la ventilation France / export) :

- Si la ventilation « française » est correcte alors il faut recalculer les détails de la colonne « Exportations » comme «Total » - « France»
- Si la ventilation « Export » est correcte alors il faut recalculer les détails de la colonne « France » comme «Total » - « Exportations »

[232] On a vu plus haut que les programmes de RIA ne sont lancés que si un micro contrôle interne a détecté une incohérence quelque part. Il y a donc forcément une variable non nulle quelque part sinon le chiffre d'affaires n'aurait pas été « à redresser » car tout le tableau aurait été cohérent (avec des variables nulles partout).
[233] Si l'APE est commerciale on met le chiffre d'affaires en ventes de marchandises etc.

- Si les deux ventilations sont incohérentes alors on lisse les montants détail de manière proportionnelle aux distributions « total ligne » et « total colonne » :
  R411=R401*R410/R310
  R412=R402*R410/R310
  R413=R403*R410/R310
  R421=R401*R420/R310
  R422=R402*R420/R310
  R423=R403*R420/R310

À la fin de cet algorithme, **le tableau est équilibré**, car on est forcément passé dans un des cas.

6.3 La partition noyau / hors noyau : le champ des entreprises potentiellement contrôlables manuellement

Comme ceci a déjà été évoqué au *chapitre 4* sur la définition des différents sous champs mis en œuvre dans le processus Ésane, certaines unités légales n'ont pas vocation à être contrôlées individuellement par des gestionnaires.

En outre, pour des raisons de **temps de calcul**, la population des entreprises disposant d'une liasse fiscale dépassant largement les 3 millions d'unités au final, **il n'est pas possible de mettre à jour** les valeurs des agrégats, des contributions des macro-contrôles, des micro-contrôles, des messages, etc., **sur l'ensemble** de la population car cela nécessiterait beaucoup trop de gestionnaires.

Les liasses ont donc été réparties en deux groupes :

- les liasses dites « **du noyau** » sur lesquelles on exécutera la chaîne aval décrite ci-dessous et qui seront susceptibles d'être « corrigées » **par un gestionnaire**. Le calcul des agrégats utiles pour les macro-contrôles se fera chaque nuit sur cette population. Le noyau contient exclusivement des unités potentiellement contrôlables, donc des sous-champs 1 et 2.
- les liasses, dites **hors noyau,** qui ne subiront plus d'autres traitements que ceux décrits dans la chaîne amont. L'agrégation de ces liasses sera faite une fois pour toutes et sera ajoutée aux agrégats des liasses « du noyau » **chaque soir** afin d'obtenir des agrégats « totaux ». Le hors noyau contient des unités **de l'ensemble des sous champs** dont **l'intégralité des sous champs 3 à 5**.

Seuls les calculs initiaux des bornes de micro-contrôles se feront sur l'ensemble de la population.

L'objectif du noyau est donc de définir simplement toutes les populations de référence pour les différents calculs relatifs aux contrôles.
Ce noyau est calculé une première fois à l'issue de la chaîne amont puis est recalculé par la chaîne aval à chaque modification de la liasse.

Est considérée dans le noyau toute unité ayant :

- *été interrogée dans une enquête (ESA ou EAP) ou*
- *un effectif supérieur à 300* (pour le sous champ 2) ou
- *au moins un montant de sa liasse supérieur à 15 000 K€.*

En 2010, il y avait **165 000 liasses actives dans le noyau** contre 2 400 000 liasses actives hors noyau (sans compter les liasses imputées).

6.4 La chaîne aval

Contrairement à l'Esa, la chaîne de CQR qui passe sur les liasses (chaîne aval) ne fait que des **contrôles** (micro et macro) donnant lieu à des messages aux gestionnaires mais **ne réalise aucun redressement automatique**. Toutefois, comme pour l'Esa, en plus des dossiers à traiter parce qu'ils sont jugés influents par les macro-contrôles, il existe d'autres causes d'examens d'une liasse pour les gestionnaires.
Les principaux traitements spécifiques aux liasses sont les suivants.

6.4.1 Identification des liasses

Lors de la phase d'appariement automatique (Cf. § *6.2.5*), il peut rester des liasses fiscales non reliées à un Siren du répertoire statistique Ocsane[234]. Le but de cette procédure est de traiter les liasses non identifiées par la procédure automatique en leur affectant un Siren unique actif et en s'assurant de la présence de l'entreprise dans le champ Ésane. Dans le cas contraire, la liasse sera rejetée et donc non prise en compte dans le processus de production.

Cette activité ne concerne que les liasses non identifiées **qui appartiennent au noyau**.

6.4.2 Remise à 12 mois des liasses

Les liasses de durée supérieure strictement à 12 mois, ainsi que celles de durée strictement inférieures à 12 mois **qui ne sont ni des créations ni des cessations**, et qui n'auraient pas été traitées par les processus de découpage - collage qu'on a vu au *§ 6.2.7.3* sont **contractées ou dilatées sur 12 mois**. En effet, on sait que si ce ne sont ni des créations ni des cessations, s'il nous manque une période, celle-ci sera couverte par une liasse arrivant ultérieurement et si la période couverte est trop longue, on imputerait sur l'année en cours une activité indue. Ce ***redressement**** (une innovation d'Ésane) est donc effectué dans le but de **garantir une certaine cohérence** dans l'évolution temporelle des variables comptables, en particulier pour les macro-contrôles.

6.4.3 Traitement des liasses multiples

Le traitement des liasses multiples a pour but d'associer à chaque Siren **une liasse fiscale et une seule**.
Il existe deux types de doublons :

- **les vrais doubles** : les liasses fiscales présentent le même Siren et plusieurs autres informations identiques. **C'est la même unité légale ou entreprise**, même si les périodes peuvent différer. Il faudra donc soit en supprimer une, soit les regrouper. Quelques exemples :
  - les liasses fiscales sont parvenues sous des identifiants fiscaux différents (ce peut être le cas quand l'établissement siège a été transféré en cours d'année) ;
  - les mêmes comptes de l'entreprise ont été chargés 2 fois dans Ésane car ils ont été transmis 2 fois à la DGFiP par l'entreprise (ex : une déclaration sous forme papier et l'autre par télé déclaration) ;
  - Les entreprises peuvent envoyer plusieurs liasses sur une même année : du 01/01/N au 30/06/N puis du 01/07/N au 31/12/N par exemple ;
  - Une des deux liasses est sur N+1 ;
  - Une liasse vient corriger l'autre : il faut en supprimer une ;
  - Une des deux liasses est sur N-1.

  On peut aussi trouver des déclarantes à plusieurs régimes.

- **les faux doubles** : les liasses fiscales présentent le même Siren et une majorité d'informations différentes. A priori, **ce n'est pas la même unité légale ou entreprise**. Quelques exemples :
  - le même Siren a été affecté à la bonne entreprise et (à tort) à une autre entreprise ayant la même adresse ou la même raison sociale ;
  - le même Siren a été affecté à la société d'exploitation et à l'entreprise holding ou tête de groupe ou loueur de fonds ou société en participation selon le cas ;
  - l'entreprise a changé de Siren en cours d'année suite à une restructuration et la liasse fiscale a conservé l'ancien Siren.

Le traitement dans la chaîne amont des liasses en double (Cf. *6.2.3*) ne concernait que des vrais doublons intra régime et intra centre (CSI). Après le chargement en base de données de la chaîne amont, il reste donc un nombre important de liasses multiples.

[234] Il s'agit de liasses non sirenisées par la DGFiP et pour lesquelles aucun appariement automatique avec l'identifiant FRP n'a été possible.

i) Pour les liasses hors noyau, les doublons sont gérés **de façon automatique** en privilégiant la source (les régimes normaux puis simplifiés etc.) puis la remontée (la livraison la plus **récente est prioritaire**).

ii) Les cas de liasses du noyau sont soumis à l'analyse des gestionnaires via l'IHM.

Ce dernier a plusieurs actions possibles mais à l'issue de son traitement il ne doit plus y avoir **qu'une seule liasse active par Siren.**

Le gestionnaire peut :

- affecter une liasse à l'exercice N+1. Cette liasse alimentera ultérieurement la base N+1 ;
- rejeter une liasse (si elle correspond à N-1 par exemple[235]) ;
- coller plusieurs liasses après avoir corrigé les dates le cas échéant (Il s'agit des cas où une entreprise a fait plusieurs déclarations au cours d'une année que les procédures automatiques de découpage et collage n'ont pu résoudre) ;
- consolider plusieurs liasses. Ce cas correspond aux unités qui ont fait des déclarations dans plusieurs régimes différents, une déclaration au BIC et une au BNC par exemple ;
- attribuer une liasse à un autre Siren.

6.4.4 Traitement des dates

Le but de l'activité sur les dates et durée est d'assurer autant que possible la **continuité temporelle des périodes comptables** auxquelles correspondent des liasses. L'objectif de ce traitement est de corriger les erreurs de date pour rétablir la cohérence ''durée = (date de clôture – date d'ouverture)'', ou pour permettre le rattachement de liasses consécutives.

Les corrections de dates et durée ne sont possibles que sur l'année N de la campagne en cours. Si l'incohérence porte sur les date et durée N-1, aucune correction n'est possible. Dans ce cas, on conserve la discontinuité des exercices comptables.

Ces cas, **à traiter par un gestionnaire**, correspondent aux cas qui n'auront pu être traités de façon automatique dans la chaîne amont et qui sont dans le noyau. Pour réaliser ce traitement, le gestionnaire peut *consulter une source de données privée* à laquelle l'Insee s'abonne[236], pour définir les bonnes dates et durée d'exercice ou *rappeler l'entreprise (téléphone voire mail désormais)*.

6.4.5 Traitement des restructurations

Ce traitement concerne les unités pour lesquelles **l'accrochage des immobilisations n'est pas cohérent** (immobilisations de fin d'exercice N-1 différentes des immobilisations de début d'exercice N de plus de 1500K€). Ce point a déjà été évoqué au § *3.3.2.1*.
On part du constat selon lequel une incohérence dans l'accrochage des immobilisations est fréquemment l'indice d'une restructuration qui a touché l'entreprise. L'objectif de ce traitement est donc de mettre en évidence des enveloppes de restructuration non encore découvertes par ailleurs.

Les principes suivants sont mis en œuvre (pour limiter les interventions des gestionnaires aux dossiers les plus importants) :

- Si la liasse est imputée en N-1 ou si la liasse N-1 ne contient que des valeurs nulles sur la partie « immobilisations » ou si la liasse N ne contient que des valeurs nulles sur la partie « immobilisations », alors le dossier **ne doit pas être contrôlé**. Les liasses imputées ou vides n'ont aucun intérêt dans cette activité qui a pour but de détecter des couples de restructuration.
- Si le ratio « Montant des acquisitions / Montant des immobilisations brutes en début d'année N » est supérieur à *0.2* et que le montant des acquisitions est supérieur à 10 millions d'euros, **alors le dossier est à traiter**.
- Si le ratio « Montant des cessions / Montant des immobilisations brutes en début d'année N » est supérieur à *0.2* et que le montant des cessions est supérieur à 10 millions, **alors le dossier est à traiter**.

[235] Comme on l'a d'ailleurs déjà vu, la réception des liasses de la campagne N pourrait conduire à modifier à la marge les résultats de la campagne N-1. Ce serait toutefois tellement marginal qu'on ne le fait pas, sauf découverte d'erreurs importantes (cela a été le cas de la campagne 2010 qui a été rectifiée à l'occasion de la publication de la campagne 2011).

[236] Ces bases de données d'entreprises, privées et donc payantes, exploitent les informations issues de la publicité légale des sociétés (RCS, journaux officiels, presse nationale etc.).

Comme ceci a déjà été évoqué dans le paragraphe spécifique aux enveloppes, un contrôle spécifique de l'investissement hors apport a également été mis en place pour corriger ou mettre en évidence des restructurations (Cf. § 3.3.3).

6.4.6 Traitement des macro-contrôles

Comme ceci a déjà été évoqué plus haut, les difficultés rencontrées pour le contrôle des liasses fiscales via le processus des macro-contrôles concernent le nombre important de variables à contrôler. Ce problème se pose de fait uniquement dans la 2e phase de contrôle pour l'élaboration des résultats définitifs (les résultats provisoires en octobre ne concernent que 6 variables).

Cette contrainte a amené à distinguer **deux types de dossiers** à contrôler pour lesquelles l'approche du gestionnaire est différente. Ces deux types de dossiers sont les suivants :

- **ceux qui contiennent des macro-contrôles et des micro-contrôles actifs** (on n'affiche aux gestionnaires que les micro-contrôles actifs dont la variable est déjà dans un macro-contrôle actif cf. § 2.6). Pour ces dossiers le gestionnaire devra contrôler en détail chaque variable en micro-contrôle. Il pourra utiliser des sources externes pour essayer de trouver une explication avant de contacter l'entreprise si nécessaire.
- **ceux qui ne contiennent que des macro-contrôles sans aucun micro-contrôle.** Ces dossiers correspondent donc à des liasses fiscales sans incohérence interne a priori mais qui ont un poids important dans certains ratios sectoriels ou dans leur évolution. En général pour ces dossiers il y a donc un grand nombre de macro-contrôles actifs. Dans ce cas, le gestionnaire ne cherchera pas à contrôler chaque variable mais plus à essayer de trouver un fait marquant qui peut expliquer cette situation. S'il trouve une explication dans une source externe ou après rappel de l'entreprise, il consignera ce fait marquant dans les commentaires associés au dossier et il validera la liasse en l'état. Ces commentaires pourront être utilisés au moment de la validation des dossiers ou par les utilisateurs internes des données au moment de leur publication.

6.5 Le traitement de la non-réponse totale [2,13]

À l'issue du chargement des liasses fiscales provenant de la DGFiP, il reste en pratique de très nombreuses unités du référentiel actives, pour lesquelles on attend des liasses fiscales a priori et **pour lesquelles on ne dispose d'aucune information** ; pour fixer un ordre de grandeur, on parle ici de **800 000 liasses en 2008 ou 2009**, **davantage à partir de 2010** avec le problème des auto-entrepreneurs qui sont typiquement dans ce cas-là.
La question est de savoir si ces unités sont réellement actives – auquel cas il est légitime d'estimer leur liasse qui apparaît manquante – ou si ce sont de « fausses actives » auquel cas il ne faut pas estimer de liasse manquante. Ces cas peuvent être assimilés à de la non-réponse totale. Le traitement de ces unités se fera en 3 étapes :

- La première étape consiste à **essayer de récupérer** des informations fiscales en complément de champ pour des unités **qui n'ont pas d'obligations fiscales** (régie des eaux, cf. § 1.1.5). Ce sera l'objet du § 6.5.1 ;

- La deuxième étape consiste à distinguer parmi les non-répondants **les unités ayant une consistance économique** (autrement dit les unités dites *économiquement actives* par la suite) des autres que l'on va considérer comme hors champ des statistiques structurelles. Le problème est en effet que lorsqu'on attend une liasse parce que le numéro Siren correspondant est sélectionné comme faisant partie du champ d'Ésane, et qu'on ne reçoit pas cette liasse de la DGFiP, on ne sait pas a priori si c'est l'indication que l'entreprise n'est **en réalité plus active** ou simplement que la DGFiP ne l'a pas encore traitée (ou que la déclaration a été faite sur un Siren qui n'est pas dans le champ par exemple) **mais que l'entreprise est bien active**. Il s'agit le plus souvent de très petites unités mais très nombreuses. L'impact individuel d'une liasse mal affectée est donc très faible mais leur nombre rend le problème pas négligeable du tout au niveau macro-économique. On détaillera ces cas au § 6.5.2.

- La dernière étape consiste à **extrapoler une liasse fiscale** aux unités que l'on a considérées comme ayant une consistance économique pour l'exercice considéré (§ 6.5.3).

6.5.1 Alimentation des établissements publics (cf. § 1.1.5)

Il s'agit des syndicats des eaux qui sont des Établissements Publics de Coopération Intercommunale (EPCI) menant des activités de traitement et distribution d'eau (division 36 de la Naf) ou de collecte et traitement des eaux usées (division 37 de la Naf). Ils sont de longue date considérés comme des Entreprises Non Financières du secteur marchand (S11 en comptabilité nationale) et figurent dans le sous champ 2 d'Ésane. Les comptes des EPCI sont **collectés au niveau des établissements** par le bureau CL2A de la DGFiP qui, depuis l'exercice 2008, les transmet à l'Insee à ce même niveau de détail (en normes comptables publiques). Seuls les établissements ayant un chiffre d'affaires non nul[237] et une activité principale correspondant aux divisions 36 ou 37 de la Naf sont intégrés à Ésane. Par conséquent, ces unités légales dans Ésane peuvent avoir des contours plus étroits que ceux des unités légales figurant dans Sirene, avec parfois des APE différentes (car seuls certains établissements de ces unités légales sont pris en compte dans Ésane (cf. § 1.1.5.1).
Seules les variables du socle sont récupérées. Ces données subissent ensuite la chaîne aval.

Il y avait environ **3500 syndicats des eaux en 2010** et ils représentaient en gros les trois quarts du nombre d'unités et des immobilisations dans les divisions 36 et 37, pour seulement un septième du chiffre d'affaires (913 millions d'euros). Certains syndicats semblent confier au moins partiellement l'exploitation de leurs installations aux géants du secteur (Veolia, Suez, etc.).

6.5.2 Détermination des entreprises économiquement actives

Les entreprises pour lesquelles on souhaite réaliser une **imputation totale** de leur liasse fiscale **sont définies** comme étant **à la fois** :

- les entreprises **appartenant au champ** Ésane pour l'année considérée ;
- les entreprises pour lesquelles **aucune liasse n'a été reçue**, (les liasses déjà présentes dans la base, saisies par un gestionnaire qui aurait joint une entreprise par exemple, ne seront pas imputées) ;
- les entreprises « **économiquement actives** ».

On a vu au *chapitre 4* comment était déterminé le champ d'Ésane. Au fur et à mesure qu'on reçoit des liasses de la DGFiP, on les repère dans Ocsane ; on sait donc à tout moment quelles sont celles pour lesquelles on a reçu au moins une liasse et celles pour lesquelles on n'a rien reçu mais toutes les entreprises du référentiel Ocsane ne sont pas économiquement actives. Le problème restant à résoudre est donc la détermination des entreprises économiquement actives.

Sont considérées comme économiquement actives toutes les entreprises :

- ayant eu une liasse active en N-1 ou N-2 ;
- ayant fait une déclaration de TVA pour l'année N ou N-1 ;
- naissantes en N ;
- qui ne sont pas enregistrées dans le fichier des micro-entreprises de la DGFiP.

Ces critères sont en fait assez larges puisqu'on voit qu'il suffit d'avoir fait une déclaration en N-2 pour être considéré comme actif encore 2 ans plus tard. **Ils étaient déjà utilisés du temps de Suse et on les a conservés par souci de continuité**.
NB. : Le fait qu'une entreprise ait répondu à l'enquête (Esa ou EAP) n'est pas considéré comme un signe d'activité économique. Il peut en effet arriver qu'une unité réponde à l'enquête sur un mauvais Siren.

Toutes les entreprises qui ne vérifient aucune de ces conditions sont donc considérées comme économiquement et donc statistiquement inactives. Elles sont donc considérées comme hors du champ d'intérêt d'Ésane, et ne seront **pas extrapolées**.

Présents dans le système à partir de l'exercice 2010, **les auto-entrepreneurs** sont traités **à part**. C'est le **fichier que fournit l'Acoss** et qui liste les chiffres d'affaires des auto-entrepreneurs qui est utilisé pour **déterminer les auto- entrepreneurs économiquement actifs**. Ainsi, pour l'exercice étudié, tout auto entrepreneur qui a réalisé du chiffre d'affaires selon l'Acoss sera considéré comme actif et les autres seront cessés statistiquement et donc non extrapolés. À noter que par la suite ces auto-entrepreneurs seront imputés **selon le régime des micro-entreprises** tel que décrit ci-dessous.

[237] Ceci est une simplification du critère exactement employé.

Le chiffre d'affaires fourni par l'Acoss n'est *pour le moment*[238] pas utilisé au moment de l'extrapolation des auto-entrepreneurs.

6.5.3 Imputation des entreprises économiquement actives

Le programme d'imputation est lancé **une fois, après réception de l'intégralité des régimes** (à l'exception des complémentaires, qui sont livrés en mars N+2). Ce choix est justifié pour les raisons suivantes :

- les ratios d'évolution servant au calcul des macro-contrôles (donc avant qu'on n'ait lancé le programme d'imputation) sont calculés sur les présentes les deux années; les « vrais » ratios devraient tenir compte des imputées mais ils les prendraient alors en compte les deux années, donc l'absence d'imputation dans les ratios utilisés ne joue que très peu par rapport aux ratios théoriques complets ;
- il n'y a pas besoin d'estimation en niveau avant la fin du chargement de tous les régimes ;
- l'absence d'une entreprise dans un fichier DGFiP donné ne signifie pas qu'elle doit nécessairement être imputée : elle peut arriver avec la livraison suivante du même régime, avoir changé de régime fiscal ; on a donc intérêt à attendre la dernière livraison pour avoir le maximum de chances de trouver les liasses qu'on attend ;
- **pour les complémentaires**, il a été décidé de **ne pas rouvrir les dossiers déjà traités**. Les données brutes ne seront chargées qu'à titre d'archive pour l'année suivante. Elles ne servent donc pas pour le calcul des agrégats de la campagne N (et sont donc estimées) mais par contre les évolutions N+1/N correspondantes seront bien calculées avec les bonnes valeurs de N.

On doit distinguer d'une part :

- l'imputation en présence de données N-1 valides ;
- l'imputation en présence de données N-1 imputées ;
- l'imputation en l'absence de données en N-1 ;

D'autre part :

- l'imputation en présence de données d'emploi (rémunérations) ;
- l'imputation en l'absence de données d'emploi ;

Et enfin :

- l'imputation des données comptables de flux ;
- l'imputation des données comptables de stock ;
- l'imputation des variables spéciales ;
- l'imputation des données caractéristiques.

Il existe **trois types d'imputation**[239] ; celles-ci s'appuient sur la strate [groupe (3 chiffres de la Nace) * tranche de taille] à laquelle appartient l'entreprise. Mais comme, par définition, on ne dispose pas de sa liasse fiscale ni fréquemment de variables d'emploi au moment où sont constituées les strates d'imputation, on utilise l'information **auxiliaire** qu'on trouve dans Ocsane pour préciser sa strate (on utilise en particulier le chiffre d'affaires N-1 ou N-2 pour imputer une tranche d'effectifs plausible si on n'a pas d'effectif pour l'unité considérée puisqu'on a besoin de sa tranche de taille) :

1. **L'imputation par rapport à la liasse N-1**. Cette imputation est appliquée dès que l'on dispose d'une **liasse en N-1** pour l'unité, que **celle-ci soit issue d'une déclaration ou déjà d'une imputation**[240]. Le principe est de faire évoluer les données de la liasse N-1 selon **la médiane d'évolution du chiffre d'affaires de la strate de l'entreprise**. L'estimation d'une caractéristique X de flux vaut donc : $\hat{X}_N = med(\frac{CA_N}{CA_{N-1}})_{strate} * X_{N-1}$

   et l'estimation d'une caractéristique X de stock vaut : $X_N = X_{N-1} + \Delta_X$ où Δ_X est la variation du stock X qui est estimée comme une variable de flux.

[238] D'une part pour des raisons de délais de livraison des données par l'Acoss et d'autre part en raison des coûts de maintenance associés.

[239] Ces imputations fondées sur les médianes a été jugée de meilleure qualité que les imputations par la moyenne, très sensibles aux valeurs extrêmes, qui étaient appliquées avant 2011.

[240] Si on n'a ni liasse en N, ni liasse en N-1 (alors que l'entreprise existe en N-1), ni liasse en N-2, c'est que l'entreprise est inactive.

2. **L'imputation par la médiane de strate pour les créations et les entrées de champ de l'année N** : ce sont les seules unités susceptibles d'avoir des liasses fiscales manquantes et pour lesquelles on ne peut pas appliquer la règle précédente.
   - les créations ne sont imputées qu'à partir des créations ayant renvoyé une liasse fiscale en N ; de même pour les entrées de champ ;
   - pour chaque création (respectivement entrée de champ) une classe d'imputation est déterminée. Celle-ci doit contenir au moins 10 créations (respectivement entrées de champ) dont au moins la moitié doivent avoir renvoyé une liasse fiscale. La classe d'imputation choisie est la plus fine possible ;
   - les liasses fiscales des créations (respectivement entrées de champ) sont **imputées par application de la médiane de la strate**. Comme la médiane n'est pas linéaire, les liasses fiscales obtenues **ne respectent pas les contraintes comptables**. Aussi **les soldes comptables sont-ils recalculés** à partir des variables élémentaires dans chaque liasse fiscale sur la base des formules comptables utilisées dans le calcul des agrégats composites.

3. **L'imputation spécifique des entreprises appartenant au régime des micro-entreprises**[241] pour lesquelles la DGFiP ne fournit aucune déclaration individuelle. Le principe de cette extrapolation est le suivant :

   - Les micro-entreprises à extrapoler sont identifiées à l'aide du fichier que la DGFiP constitue à la demande de l'Insee et qu'elle tire de son propre référentiel. Ce fichier a été utilisé depuis de nombreuses années à la fois dans la division SSE et à la DGCIS[242] pour l'étude des TPE il est donc jugé de bonne qualité.
   - L'imputation se fait ensuite de la façon suivante. On applique aux micro-entreprises la structure moyenne observée sur les entreprises du **RSI**[243] **sans salarié** ayant un **chiffre d'affaires inférieur à 80 K€** tous secteurs confondus. On fait ici **une hypothèse forte** selon laquelle **les micro-entreprises se comporteraient comme les petites entreprises du RSI**. On distingue les deux régimes BIC et BNC. Par la suite le chiffre d'affaires moyen des micro-entreprises est calculé à partir de l'agrégat du chiffre d'affaires fourni par la DGFiP pour les micro-entreprises. L'estimation d'une caractéristique X vaut donc :
   - $$\hat{X}_N = \left(\frac{\sum_{i \in U} CA * 1\mathrm{I}_{i \in micro}}{\sum_{i \in U} 1\mathrm{I}_{i \in micro}} \right) * \left(\frac{\sum_{i \in U} X_N * 1\mathrm{I}_{i \in RSI\,0\,sal\,et\,CA<80}}{\sum_{i \in U} CA_N * 1\mathrm{I}_{i \in RSI\,0\,sal\,et\,CA<80}} \right)$$

   Ainsi **toutes les micro-entreprises d'un même régime** (BIC ou BNC) ont **la même imputation** quelle que soit leur APE.

   Ce point est fondamental pour les chargés d'étude qui travaillent sur données individuelles : on génère bien des estimations individuelles dont l'agrégation redonne bien l'agrégat que l'on cherche mais **il ne faut pas rapprocher ces données individuelles d'autres caractéristiques** car ce ne sont pas les vraies caractéristiques de ces unités.

---

[241] Le régime des micros-entreprises est réservé aux petites entreprises dont le chiffre d'affaires annuel hors taxes de l'année civile précédente est inférieur ou égal à :
- 32 900 € pour les activités de prestations de services ;
- 82 200 € pour les activités d'achat-revente, de vente à consommer sur place et de fourniture de logement (hors location meublée autre que meublé de tourisme, gîte rural ou chambre d'hôte).

Seuls peuvent en bénéficier les entrepreneurs individuels imposés à l'impôt sur le revenu (IR) dans la catégorie des bénéfices industriels et commerciaux (BIC) et bénéficiant de la franchise en base de TVA. Comme ils ne déclarent leur chiffre d'affaires qu'à l'IR, on ne dispose d'aucune liasse fiscale et donc d'aucune donnée comptable réelle.

[242] Direction générale de la compétitivité, de l'innovation et des services. Il s'agit de la direction générale du ministère de l'économie et des finances qui a en charge les PME et les TPE.

[243] Il s'agit du régime réel simplifié d'imposition. Ce régime s'applique aux entreprises qui ont opté pour ce régime et dont le chiffre d'affaires hors taxes est compris entre :
- 32 900 € et 236 000 € pour les activités de prestation de services ;
- 82 200 € et 783 000 € pour les activités d'achat-revente, de vente à consommer sur place et de fourniture de logement.

7. Les principes de prise en compte et de contrôle des données d'emploi

7.1 Les différents concepts et sources utilisées

7.1.1 Les concepts d'emploi

Comme on l'a rappelé dans *l'avant-propos*, Ésane sert non seulement à calculer des variables comptables issues directement des comptes sociaux des entreprises[244], et parmi celles-ci l'emploi et la masse salariale sont évidemment des variables importantes, mais aussi de nombreux ratios économiques faisant intervenir l'emploi, publiés en tant que tels, comme **caractéristiques** de la statistique structurelle d'entreprise[245]. Que ce soit en niveau ou pour des ratios, ces informations ne sont calculées dans ***aucun autre système d'information de l'Insee[246] qu'Ésane*****, au niveau micro-économique de l'entreprise** qui permet ensuite des réagrégations pour n'importe quelle sous-population.

Il est évidemment fondamental que l'emploi considéré puisse se comparer aux variables comptables de l'entreprise (pour calculer des ratios) et en particulier à la masse salariale. On a donc **besoin** de pouvoir mesurer un « **volume d'emploi** », tenant compte de toutes les formes d'emploi salarial dans le courant de l'exercice, c'est-à-dire tenant compte à la fois de la *quotité de travail* (temps plein ou temps partiel[247]) des salariés et de leur *période de travail* (sur toute l'année ou sur certaines périodes seulement[248]). Cela correspond à ce que l'Insee appelle « **l'emploi équivalent temps plein annuel** », en abrégé **ETP**[249]. On notera au passage que la masse salariale est bien influencée elle aussi par ces deux composantes (quotité et période) mais pas de façon nécessairement proportionnelle comme l'est le volume d'emploi[250].

Mais si cette mesure de l'emploi est adaptée à la mise en regard des variables comptables, elle ne suffit pas à décrire l'emploi « utilisé » par l'entreprise. Quand on dit que telle unité « emploie N salariés », on fait référence aux personnes physiques employées, indépendamment de leur quotité de travail ou de la date à laquelle ils ont été embauchés. C'est implicitement la notion d'emploi utilisée dans la presse ou dans les descriptions d'un secteur économique par un chef d'entreprise. Il faut donc compléter la notion précédente par un emploi « **en nombre de personnes physiques** » ou plus précisément en « **postes de travail physiques** ». Pour les statisticiens, cette notion correspond à l'emploi en **nombre de postes[251] à une date donnée** que l'Insee a l'habitude de mesurer au **31/12**.

---

[244] Ou plus tard des comptes consolidés des entreprises profilées.

[245] Par exemple : CA par salarié, VA par salarié, salaire brut annuel par tête, immobilisations par salarié, total de bilan par salarié, etc.

[246] En effet, les autres systèmes d'information de l'Insee qui gèrent des informations sur l'emploi, ne disposent pas simultanément des données fiscales et raisonnent à partir des données d'emploi local, en « secteurs d'établissements ». Ces derniers sont formés de tous les établissements (unités locales) qui ont la même activité principale (code dit APET). Les secteurs d'établissement ne recouvrent pas les secteurs d'entreprises ; les résultats sont donc par nature différents de ceux élaborés par Ésane.

[247] Un salarié à 80% comptera en principe pour 0,8 ETP. Pour plus de détails sur le calcul de l'emploi ETP, voir *l'annexe 2*.

[248] Un salarié qui a été embauché à temps plein le 1/10 comptera pour 0,25 ETP.

[249] A l'extérieur de l'Insee, on appelle quelquefois cette notion « l'emploi équivalent temps plein annuel **travaillé** » (ETPT) pour bien souligner qu'on prend en compte les périodes d'emploi où le salarié travaille effectivement en sus de la quotité de travail. Lorsqu'on fait cette distinction, l'emploi ETP ne décrit qu'une situation où on mesure la quotité de travail à un instant donné ; il faut alors préciser à quel moment se fait l'observation. Mais avec la mise en œuvre du « temps partiel annualisé » c'est-à-dire d'un temps partiel où le salarié choisit les périodes où il est présent au travail, les deux situations deviennent difficiles à distinguer.

[250] Un salarié qui ne travaille que m mois sur 12 recevra bien m/12 du salaire annuel mais, pour les agents à temps partiel, la fraction de rémunération à temps complet qu'ils perçoivent est généralement supérieure à la fraction de temps complet qu'ils effectuent.

[251] La définition précise des postes est donnée en *annexe 2*. Il est bien clair ici que si une personne travaille à temps partiel pour plusieurs employeurs, elle est comptée plusieurs fois en nombre de postes alors qu'il s'agit de la même personne. Seul le DERA peut faire les traitements (assez lourds et complexes) pour éliminer ces doubles comptes et faire un véritable dénombrement en personnes physiques, comparable à la mesure de la population.

Remarque : la liasse fiscale contient elle-même une variable d'emploi qu'il était évidemment très tentant d'utiliser, au moins comme l'une des deux variables d'emploi recherchées. L'emploi demandé dans la liasse fiscale est ***la moyenne arithmétique des effectifs (personnes titulaires d'un contrat de travail directement rémunérées par l'entreprise) en fin des trimestres couverts par l'exercice fiscal***. L'expérience du système SUSE antérieur avait malheureusement montré que **cette variable** – non contrôlée car généralement non utilisée par la DGFiP – est de **qualité** moyenne (elle n'est pas toujours remplie). Vu sa définition, il s'agit d'un « **emploi en nombre de postes, moyen sur l'exercice** », mesure « intermédiaire » entre les deux notions a priori recherchées :

- ce n'est pas un emploi en équivalent temps plein puisqu'il ne prend pas en compte la quotité de travail ni véritablement les périodes d'emploi puisqu'une embauche temporaire en juillet août par exemple ne sera pas prise en compte (alors qu'elle influera bien évidemment sur la masse salariale).
- Ce n'est pas non plus un emploi à une date donnée puisque c'est un emploi « moyen », sur l'exercice, calculé comme une moyenne arithmétique d'emplois en fin de trimestre.

Pour ces deux raisons (qualité moyenne et concept intermédiaire), on a décidé d'abandonner l'emploi tel qu'il était suivi auparavant dans la source fiscale à travers SUSE et ***une autre innovation d'Ésane**** a donc consisté à se tourner vers **d'autres sources administratives** dont l'exhaustivité et la disponibilité puissent se comparer à celles des liasses fiscales. Mais on verra qu'in fine, **pour l'emploi « en nombre de postes »,** on a cherché à reconstituer un emploi **qui se rapproche conceptuellement de celui des liasses fiscales**.

7.1.2 Les sources utilisées

Il existe plusieurs sources administratives sur l'emploi, constituées à partir des déclarations sociales des entreprises (Urssaf, DADS) ; elles sont utilisées dans plusieurs systèmes d'information de l'Insee (**Épure**[252], **Clap**[253]). **Au moment du démarrage du projet Resane**, le système Épure d'exploitation de la source Urssaf[254] était en refonte et le système Clap avait d'ailleurs justement comme ambition de résoudre les incohérences éventuelles entre les données des DADS et celles des Urssaf. Par ailleurs la DSE avait déjà l'habitude de mises à jour de l'emploi dans Sirene à partir des DADS. **La source administrative la plus naturelle** (car à la fois la plus exhaustive et la plus fiable**) pour estimer les données d'emploi dans Ésane était donc la source DADS-U**[255].

L'avantage majeur de la source DADS est que, partant d'une information individuelle portant sur chaque salarié et ses périodes d'emploi dans l'année, elle permettait a priori de reconstituer à la fois un emploi en ETP et un emploi au 31/12. Toutefois si l'on utilisait la source DADS initiale (c'est-à-dire avant son traitement dans Clap), on risquait de **ne pas pouvoir profiter de l'amélioration apportée par les gestionnaires Clap** et d'avoir des données d'emploi national i**ncohérentes**, conceptuellement, avec les données d'emploi publiées par les directions régionales de l'Insee. Or la réconciliation des données publiées par des sources différentes et la recherche d'une meilleure cohérence entre publications était un des axes fondateurs du projet Resane. On s'est donc orienté très tôt **vers l'utilisation de la source Clap** pour minimiser cette source potentielle d'incohérence.

[252] Système d'information de l'Insee disposant de gestionnaires en propre et reposant sur l'exploitation des Urssaf [ainsi que de la MSA pour les entreprises cotisant auprès de la MSA et de SIASP pour les agents de la fonction publique] pour fournir de l'information sur l'emploi et les revenus. Son acronyme signifie « extension du projet Urssaf sur les revenus et l'emploi » (l'extension porte justement sur la MSA et la fonction publique). Par souci de simplification, on parlera le plus souvent des Urssaf qui sont plus connues que le sigle Epure propre à l'Insee. Mais à chaque fois, c'est bien auprès du système Epure complet qu'on s'alimente.

[253] Autre système d'information disposant également de gestionnaires en propre et visant à la meilleure réconciliation de données **sur l'emploi local**. Son acronyme signifie « Connaissance locale de l'appareil productif ». Jusqu'à la campagne 2015 (exclue) les gestionnaires étaient concentrés dans un pôle situé à Toulouse à la direction régionale de Midi-Pyrénées. A compter de la campagne 2015, les mêmes opérations sont effectuées par des gestionnaires situés au Centre Statistique de **Metz**, plus précisément au sein du Centre de Statistiques Sociales et locales (CSSL). Quel que soit le nom du service, les contrôles et le travail effectué sont en principe les mêmes (jusqu'à l'arrivée de la déclaration sociale nominative (DSN) – voir encadré à la fin du § - qui changera les contrôles et la logique même du système puisque Clap disparaitra).

[254] En réalité, le système Épure s'alimente aussi auprès de la MSA et de la fonction publique.

[255] Pour Déclaration Annuelle de Données Sociales Unifiée, obligatoire depuis le 1/1/2006. Elle résulte de la fusion des DADS -TDS et des DADS-CRC (institutions de retraite complémentaire). Pour simplifier, on parle souvent de la source DADS sans préciser le U.

L'inconvénient majeur de Clap était cependant qu'il détermine la cohérence entre les DADS et les Urssaf sur la base de l'emploi **au 31/12,** dans une approche de type « **nombre de postes** » qui n'est pas le concept le plus cohérent avec le reste de la liasse fiscale.

Au final, **il a donc été décidé dès le départ de s'alimenter à trois sources à la fois,** ce qui constitue un système d'information assez complexe[256], avec **trois chaînes** qui fonctionnent en parallèle mais qui ont des interactions à certains moments clés[257] :

a. Il y a une collaboration **avec Clap** (données annuelles de la campagne en cours sur l'année N) qui vise la meilleure cohérence entre emploi local et emploi national[258] et profite de la réconciliation Urssaf – DADS pour fournir un **emploi mesuré au 31/12 de l'année N**. C'est dans le cadre de cette collaboration qu'on fixe le champ des unités pour lesquelles on attend de l'emploi salarié.
b. En parallèle, il y a une alimentation d'Ésane **directement à partir des DADS** qui permet de récupérer **un emploi en ETP** (**sur l'année N),** et qui fournit également **les données d'emploi (dites) secondaires[259].**
c. On récupère aussi **les données trimestrielles** des années N et N+1 **venant de la source Urssaf** qui servent à :
   - des corrections de durée d'exercice sur l'emploi au 31/12 et sur la masse salariale ;
   - des contrôles en cours d'année des variables comptables de l'année N (avant qu'on ne dispose des données d'emploi annuelles de la campagne N) ;
   - au tirage d'échantillon de l'ESA (N+1).

Comme on récupère des données qui viennent directement d'autres systèmes d'information où elles sont censées avoir été contrôlées, **il n'y a donc aucun contrôle manuel de l'emploi, prévu au sein de la chaîne de production Ésane** (au sens réalisé par des gestionnaires Ésane avec retour éventuel à l'unité déclarante).

En revanche, l'emploi que l'on récupère ainsi de ces diverses sources n'est pas juste « recopié ». Comme on le verra (cf. § *7.2.2.2* et § *7.3.2.2*), il est **recalé sur la date de fin d'exercice** et **transformé**, en ce qui concerne l'emploi en nombre de postes, en un **concept identique à celui des liasses fiscales**. Tout se passe comme si on n'utilisait pas l'emploi directement fourni dans la liasse fiscale en raison de sa qualité, mais qu'on cherchait à **le reconstituer**[260] à partir d'autres sources plus fiables.

Ainsi, ce qu'on baptise « **emploi au 31/12** » est en réalité **une estimation de l'emploi moyen sur l'exercice** (au sens du concept de la liasse).

Du coup, avec ce **recalage de période** (exercice au lieu de année civile) **et ce changement de concept**, on n'aura **jamais une identité parfaite** entre l'emploi publié par l'Insee **au 31/12** par le DERA ou les régions et l'emploi publié par Ésane ; mais **on peut viser une cohérence globale** se manifestant par le fait qu'on sait **expliquer les divergences**.

On savait aussi dès le départ que le choix de la DADS-U **empêchait conceptuellement** de calculer certaines caractéristiques, absentes de la DADS-U :

- l'emploi non salarié;
- la mise à disposition du personnel;
- l'emploi intérimaire dans les sociétés utilisatrices[261] ;
- un ensemble de caractéristiques très spécifiques à certains secteurs.

Ces caractéristiques manquantes dans la source administrative **devaient donc être collectées par voie d'enquête**, (dans l'ESA et l'EAP par conséquent), pour compléter les données d'emploi salarié.

[256] D'autant que Clap est elle-même une application qui confronte les Urssaf et les DADS !
[257] Voir en particulier les § *7.2.2.2*, *7.3.2.2* et *8.4.2* et le graphique ci-après.
[258] Théoriquement seulement parce qu'il y a un certain nombre de cas, comme on le verra, où on ne retient pas in fine la donnée Clap ; on explique aussi au § 7.2.2.2 comment on passe des concepts « Clap » au 31/12 aux concepts « liasse » en fin d'exercice, ce qui modifie les données. De plus les publications sur l'emploi régional comprennent aussi fréquemment l'emploi public non marchand.
[259] Cf. § *7.3.1.1* : il s'agit de précisions sur le partage temps plein / temps partiel, ou sur le pourcentage de cadres d'ouvriers, etc.
[260] Principalement par souci de continuité avec l'emploi qui était publié avec les données Suse précédemment.
[261] Les salariés sont payés par les agences d'intérim (classées dans le secteur 78.20Z) ; ce sont elles qui les déclarent dans leurs DADS mais cela n'indique pas dans quelle entreprise ils travaillent véritablement.

Deux dysfonctionnements du système ont toutefois rendu inopérante cette collecte pour la publication :

- jusqu'à l'année de constat **2012 comprise**, **l'EAP ne posait aucune question sur l'emploi**. On n'a donc aucune de ces informations (qu'on avait avec l'EAE) sur le champ de l'industrie jusqu'en 2012. On est dans un cas où le souci de diminuer la charge de réponse des entreprises a conduit à trop restreindre le questionnement ;
- jusqu'à l'année de constat **2012 comprise,** une imprécision dans la formulation des questions correspondantes de l'Esa a rendu **inexploitable les réponses de l'ESA[262].**

La conclusion de ces remarques est que sur la période 2008-2013 en pratique, **seul l'emploi salarié a pu être calculé et publié**. Ce n'est qu'à compter de **la campagne 2014** (qui a été publiée à l'été 2016)**, qu'on a pu publier[263] des données sur l'emploi total** (y.c. non-salariés et intérimaires).

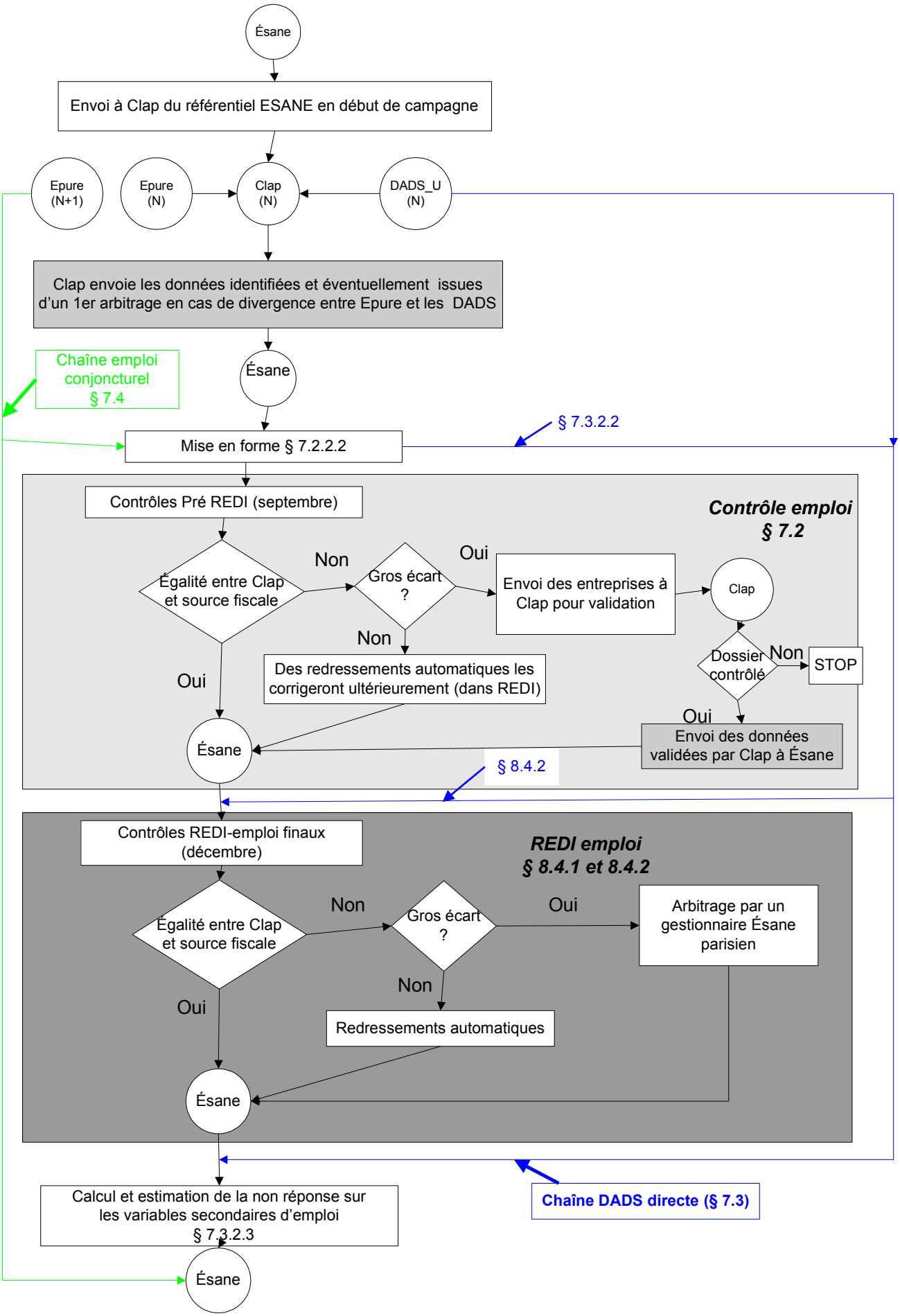
S'agissant des problèmes d'identification et de champ, il ne faut pas perdre de vue que la majorité des unités légales n'ont pas de salariés. On a donc d'un côté entre 3 et 4 millions de liasses attendues et de l'autre entre 1 million et 1,5 million d'unités légales qui déclarent de l'emploi salarié.

La norme **N4DS[264]** qui a remplacé la DADS-U le 1er janvier 2012 et surtout la **déclaration sociale nominative (DSN), mensuelle[265]**, instaurée par la loi Warsmann (qui va entraîner la disparition des sources DADS et Urssaf séparées) **vont bouleverser les systèmes d'information existants** (dans les entreprises et à l'Insee). Par définition en particulier la source Clap, qui visait à la meilleure réconciliation des DADS et des Urssaf, va disparaître. Le principe d'indépendance du traitement des sources (Cf. § *1.2.2*) devrait alors s'avérer fort utile pour garantir le bon fonctionnement d'Ésane même si certaines des sources qui l'alimentent sont modifiées voire indisponibles.
Lorsqu'on sera parvenu à un certain régime de croisière, il sera **nécessaire de revoir l'articulation** du futur système d'information sur l'emploi avec Ésane, car l'information disponible sera très différente.

La **suite du chapitre** décrit les différentes étapes du processus de production relativement aux variables d'emploi. Il commence par décrire les différents échanges qui ont lieu entre Clap et Ésane (§ *7.2*) avec les traitements spécifiques de la chaîne Ésane, pour l'estimation de l'emploi en nombre de postes. On passe ensuite à la description de la fourniture des DADS par le DERA et des traitements correspondants pour l'estimation de l'emploi en ETP (§ *7.3*).
Le chapitre se termine par la description (§ *7.4*) de l'alimentation directe de la chaîne l'emploi conjoncturel, mesuré sur l'année N+1. Comme on l'a vu, cette alimentation au fil de l'eau en données trimestrielles n'a pas pour but d'estimer directement des variables d'emploi de la campagne N mais de compléter les estimations Clap ou DADS pour les aligner sur la durée de l'exercice fiscal qui ne coïncide pas toujours avec l'année civile, ainsi que d'aider au contrôle des variables comptables de la campagne N (quand on n'a pas encore les données de Clap ou des DADS) et de préparer le tirage des échantillons d'enquête de la campagne N+1.

Le graphique ci-après schématise toute la chaîne : il présente l'ensemble des étapes du contrôle des caractéristiques de l'emploi dans les trois chaînes du système de production Ésane. Au centre, en noir sur fond gris clair, on a d'abord les étapes des échanges avec Clap ; elles débouchent sur la phase Rédi emploi (sur fond gris foncé) qui est décrite en détail dans le chapitre suivant (*cf. § 8.4.1* et § *8.4.2*). La chaîne d'alimentation directe des DADS figure en bleu à droite (Cf. § *7.3*) et celle de l'emploi conjoncturel en vert à gauche (Cf. § *7.4*). [Cependant les traitements spécifiques de ces deux chaînes ne sont pas détaillés (sauf leur lien avec la chaîne principale de l'emploi en nombre de postes)].

[262] Pour de nombreuses entreprises dont seul le dirigeant est non salarié, la question sur les non salariés ne précisait pas suffisamment clairement s'il devait se compter ou pas. Les résultats 2008 à 2012 se sont du coup avérés incohérents avec ce qu'on savait par ailleurs.
[263] Les données des années 2011, 2012 et 2013 sur les non salariés ont été transmises cependant à Eurostat sur la base d'une estimation à partir du nombre d'entreprises individuelles (EI), en faisant l'hypothèse que chaque EI correspond à un non salarié.
[264]Norme pour les déclarations dématérialisées des données sociales : c'est une évolution de la DADS-U.
[265] Elle a démarré dès 2013 sur la base du volontariat pour une trentaine d'entreprises.


Ésane
Envoi à Clap du référentiel ESANE en début de campagne
Epure (N+1)
Epure (N)
Clap (N)
DADS_U (N)
Clap envoie les données identifiées et éventuellement issues d'un 1er arbitrage en cas de divergence entre Epure et les DADS
Chaîne emploi conjoncturel § 7.4
Ésane
§ 7.3.2.2
Mise en forme § 7.2.2.2
Contrôles Pré REDI (septembre)
Contrôle emploi § 7.2
Égalité entre Clap et source fiscale
Non
Gros écart ?
Oui
Envoi des entreprises à Clap pour validation
Clap
Non
Des redressements automatiques les corrigeront ultérieurement (dans REDI)
Dossier contrôlé
Non
STOP
Oui
Oui
Envoi des données validées par Clap à Ésane
Ésane
§ 8.4.2
Contrôles REDI-emploi finaux (décembre)
REDI emploi § 8.4.1 et 8.4.2
Égalité entre Clap et source fiscale
Non
Gros écart ?
Oui
Arbitrage par un gestionnaire Ésane parisien
Non
Oui
Redressements automatiques
Ésane
Calcul et estimation de la non réponse sur les variables secondaires d'emploi § 7.3.2.3
Chaîne DADS directe (§ 7.3)
Ésane

7.2 Les données structurelles d'emploi : la collaboration avec le pôle Clap (emploi en nombre de postes)

Contrairement aux autres caractéristiques du système d'information, les données concernant l'emploi sont contrôlées **en collaboration avec le pôle Clap**[266], mais principalement par les gestionnaires du pôle CLAP puisqu'on a vu que les gestionnaires Ésane ne faisaient pas de contrôle spécifique sur l'emploi.

Avant de décrire les différentes étapes de la collaboration avec le pôle, on rappelle ici les caractéristiques des déclarations des entreprises :

Celles-ci doivent établir - soit trimestriellement soit mensuellement selon que l'entreprise occupe moins de 10 salariés (tenant compte de leur quotité de travail) ou pas au 31/12 - un **bordereau récapitulatif de cotisation (BRC)** adressé aux Urssaf en même temps qu'elles doivent s'acquitter du paiement des cotisations et contributions sociales correspondantes. En fin d'année, elles remplissent la « **déclaration annuelle de données sociales**[267] qui comprend les durées d'emploi et les rémunérations nominatives brutes et plafonnées de chaque salarié sur lesquelles sont calculées les cotisations patronales et salariales. Elle permet de reconstituer :

1) **"L'effectif au 31/12" de l'unité légale : c'est un effectif en nombre de postes.** Dans cet effectif au 31/12, un salarié à 80% compte donc pour 1 s'il est présent au 31/12. Cet effectif ne tient pas compte des périodes de l'année où le salarié n'a éventuellement pas travaillé ni par définition des salariés qui ne sont plus présents au 31/12. **Il ne tient pas compte** non plus de **l'emploi de courte durée**, caractérisé dans les DADS par la **notion de postes « annexes** ». L'idée est de ne pas considérer comme emploi permanent caractéristique de l'entreprise un salarié qui n'aurait été employé que quelques jours dans l'année. (Cf. § *7.2.1.1* et surtout *l'annexe 2* où cette notion développée par l'Insee est explicitée).

2) **"L'effectif ETP de l'unité légale" calculé sur l'année civile** (on dit aussi quelquefois **EQTP** pour équivalent temps plein) : il tient compte à la fois de la **quotité** de travail de chaque salarié et de la **période** pendant laquelle il a travaillé. Il est calculé à la fois en tenant compte et **sans tenir compte des postes annexes** ; mais les valeurs qui sont transmises à Ésane, cohérentes avec l'effectif au 31/12, sont celles **calculées sans tenir compte des postes annexes** (Cf. § *7.3.1.1*). Il est calculé selon la formule suivante (sauf pour le secteur de l'Intérim) :

EFF (ETP) = Sigma (sur i) [*Min (heures annuelles du salarié i / norme annuelle ; 1)*]

[L'usage du Min est là pour éviter que le recours aux heures supplémentaires fasse qu'un salarié puisse être compté comme plus qu'un ETP sur l'année ; quel que soit son nombre d'heures un salarié ne peut pas compter pour plus de 1. Le sigma ne porte que sur les postes de salariés occupant des postes non annexes].

Il faut retenir de ces deux paragraphes que le calcul de l'emploi fait à partir des DADS **ne prend pas en compte les postes annexes.** Pour le calcul des ratios « par ETP » on peut donc considérer qu'on a **une légère sous-estimation du volume d'emploi total tel qu'il est pris en compte dans la masse salariale**. D'après les chiffres fournis par le DERA, le poids total des postes annexes (tous secteurs confondus) serait de **1,2% de l'emploi total** (annexe + non annexe). Évidemment ce poids peut être plus fort dans certains secteurs spécifiques où cette forme d'emploi est plus répandue.

[266] À compter de la campagne 2015, au CSSL de Metz. Voir la note « Clap » du § *7.1.2*.

[267] Ainsi qu'un « **tableau récapitulatif » (TR)** qui mentionne les rémunérations brutes et plafonnées de l'ensemble des salariés de l'établissement, - le montant des cotisations et contributions dues, - le montant des cotisations et contributions versées au cours de l'année aux Urssaf et, le cas échéant, le montant de la régularisation, ainsi que deux notions d'emploi, **l'effectif global de l'entreprise au 31/12, ainsi que l'effectif moyen au 31/12 qui sont différents et qu'il ne faut pas confondre avec ceux sur lesquels nous travaillons.**

7.2.1 Les échanges Clap Ésane

7.2.1.1 Rappel sur les traitements faits par le pôle Clap[268] (Connaissance Locale de l'Appareil Productif)

Les deux objectifs centraux de Clap sont de faire la mise en cohérence des sources DADS-U et Urssaf[269], puis de déterminer un emploi « local » par établissement, cohérent avec l'effectif total de l'unité légale. Le deuxième objectif ne concerne pas directement Ésane. Cette mise en cohérence DADS - Urssaf est faite sur une caractéristique "emploi" et une caractéristique "rémunération".

La mise en cohérence Clap sur les effectifs se fait entre les deux variables suivantes :

- L'effectif au 31/12 issu de la source DADS-U. Le fichier « périodes », rempli par les unités légales contient une déclaration pour chaque période d'emploi de chaque salarié dans une unité légale donnée. À partir de ce fichier détaillé, le DERA calcule un fichier « postes » (sur le croisement salarié x établissement[270]) et caractérise ensuite **les postes en « Annexes » et « Non Annexes »**. Les postes annexes correspondent comme on l'a vu à de l'emploi de durée courte dont il n'est **pas tenu compte** pour la détermination de l'emploi au 31/12 (en « personnes physiques ») .
- L'effectif au 31/12 issu de la source Urssaf. Dans la source Urssaf, une personne salariée compte pour 1 et ce, quelle que soit sa quotité de temps de travail également. La **source Urssaf ne tient pas compte de la distinction annexe / non annexe**.

La mise en cohérence Clap sur les rémunérations se fait entre les deux caractéristiques suivantes :

- Masse salariale issue de la source DADS-U ; celle-ci ne tient pas compte non plus des postes annexes.
- Masse salariale issue de la source Épure2.

S'il n'y a **pas « d'incohérence »** entre DADS-U et Urssaf, c'est **la source DADS-U qui est systématiquement retenue**. Du fait de la prise en compte des postes annexes dans l'une et pas dans l'autre, elles peuvent être assez fréquemment différentes, tout en étant le plus souvent « cohérentes ». Dans tous les cas, **Clap propose un montant de rémunération et d'effectif issu de l'une des sources Urssaf ou DADS-U[271]**.

À l'issue de cette mise en cohérence, les caractéristiques retenues par Clap se nomment "rémunération Clap" et "effectif Clap" (au 31/12)**.**

Les résultats de la mise en cohérence des sources DADS-U et Urssaf par Clap seront pris en compte par Ésane.

7.2.1.2 Les principes de la collaboration Clap - Ésane

La collaboration se fait en 4 étapes :

1. Ésane envoie la liste des unités de l'ensemble du champ Ésane au pôle Clap
2. Ésane identifie les entreprises à vérifier par le pôle Clap (étape dite Pré Rédi)
3. Le pôle Clap traite ces unités
4. Ésane intègre les corrections de Clap et met en cohérence ces données avec les autres sources à sa disposition.

(On peut aussi suivre ces étapes sur le graphique ci-dessus).

[268] À compter de la campagne 2015, au CSSL de Metz. Voir la note « Clap » du § *7.1.2*.

[269] Dit très rapidement, on peut penser que la source Urssaf est plutôt plus exhaustive et la source DADS plutôt plus précise (mais elle a quelques trous de déclaration). La mise en cohérence bouche donc des trous des DADS et apporte aux Urssaf la précision des DADS.

[270] En d'autres termes un même salarié (caractérisé par son NIR) employé successivement dans 2 établissements différents (Siret) aura deux enregistrements dans le fichier poste. En revanche, plusieurs périodes d'emploi successives d'un même salarié dans le même établissement seront regroupées dans le même « poste », caractérisé par une durée totale d'emploi et une rémunération totale.

[271] Les gestionnaires Clap ne peuvent pas proposer une valeur différente de celle de l'une des deux sources.

1 - Identification du champ Ésane dans Clap

Le champ Ésane (défini au sein du répertoire Ocsane) est, rappelons-le, un champ défini a priori, inclus dans celui de Clap[272]. Le pôle Clap doit donc **s'assurer que toutes les unités du champ Ésane fourni sont présentes dans les données provenant des DADS ou des Urssaf**. Si Clap reçoit une DADS qui appartient au champ conceptuel d'Ésane (i.e. marchande), et qui ne figure pas dans le référentiel fourni, cela correspond certainement à une déclaration sur un autre SIREN que celui sur lequel nous l'attendons. Il faut donc que Clap essaie d'identifier ce SIREN et nous le signale.

Ceci permet également au pôle Clap **d'identifier les unités** sur lesquelles il devra nous fournir des données. C'est le champ sur lequel Ésane se met d'accord avec Clap.

2 - L'identification des dossiers à vérifier par le pôle Clap (Pré Rédi)

Une fois que Clap a traité les unités légales, il nous les renvoie, avec donc un effectif et une rémunération « Clap » au 31/12. Après un traitement automatique pour aligner ces variables sur la durée d'exercice de la liasse fiscale (Cf. § *7.2.2.2*), Ésane évalue l'écart éventuel entre les effectifs (et rémunérations) Clap ainsi corrigés et les effectifs (et la masse salariale) fiscaux[273]. Ces tests sont faits dans la logique de « réconciliation » de source qui sera exposée au chapitre 8 suivant et s'appuient sur des macro-contrôles donc sur l'influence qu'a le changement éventuel de valeur sur un agrégat dans une strate (Cf. en particulier § *8.3.8*). Ésane détermine, donc quelles sont les entreprises pour lesquelles :

- Il y a égalité entre Clap et source fiscale ; il n'y a aucun traitement à faire ;
- il existe une « petite » divergence qui sera traitée automatiquement (« petites » entreprises) ;
- il existe une « grosse » divergence **qui justifie un réexamen par un gestionnaire Clap.** On considère que c'est le cas si la contribution du macro-contrôle est supérieure à 1%[274]. La liste de ces entreprises est renvoyée au pôle Clap pour examen (Cf. § *7.2.2.5*).

Ésane transmet donc à Clap la liste des entreprises à contrôler, ainsi que les évaluations des salaires et de l'emploi figurant dans les liasses fiscales. Ces informations sont donc relatives à des exercices fiscaux. Ésane transmet aussi les données sociales (Rémunérations totales sur l'exercice et emploi) recalées sur l'exercice fiscal issues du traitement du § *7.2.2.2*. Un traitement particulier est conduit pour les entreprises en restructuration et les entreprises impliquées seront signalées à Clap.

3 - Le travail d'analyse réalisé par les équipes Clap

Comme les **gestionnaires Clap ne peuvent pas analyser les incohérences liées à la source fiscale,** ils n'analyseront que les entreprises de cette liste pour lesquelles une variation anormale, des effectifs Clap ou des rémunérations Clap, est constatée entre N et N-1, ou pour lesquelles données Épure et données DADS sont incohérentes, donc pour lesquels ils peuvent justifier d'appeler l'entreprise ; les autres entreprises seront traitées par **Ésane** au moment de la réconciliation des données.

Ce partage tient compte du fait que, au sein des entreprises, les interlocuteurs Clap appartiennent plutôt au service du personnel qui ne connaissent pas la déclaration fiscale de l'entreprise, tandis que les gestionnaires Ésane contactent plutôt des personnes du service comptabilité.

Pour cette sélection d'entreprises, Clap réalise donc une première confrontation individuelle des données d'emploi et de salaires entre les sources « Clap » et « fiscale » : il examine les unités

[272] Du moins pour toutes les unités ayant de l'emploi puisque toutes les unités employeuses (marchandes et non marchandes y compris toutes les associations) sont obligées de remplir une DADS et des bordereaux Urssaf. En revanche on n'aura pas de DADS pour les unités légales d'Ésane sans salarié (qui sont les plus nombreuses).

[273] Égaux comme on l'a vu à la moyenne arithmétique des effectifs de fin de trimestre. On note que conceptuellement les effectifs de la liasse et de Clap sont différents. Ce ne peut donc être qu'un test de « cohérence » puisqu'il n'y a pas de raison que ces effectifs soient identiques s'il y a eu des variations d'effectifs au cours de l'année.

[274] Sur le schéma précédent, cela correspond au losange « gros écart ? ».

divergentes et éventuellement change la source retenue (DADS ou Urssaf)[275] pour obtenir une masse salariale cohérente avec celle transmise par Ésane. Il est probable que les données Clap seront rarement modifiées, sauf en cas d'erreur de choix de source ou de restructuration non détectée. Pour faciliter la gestion de la cohérence, le **Pôle Clap priorise** les anomalies à traiter, ainsi :

- **Priorité 2 (la catégorie la plus prioritaire) :** + ou - 10% d'écart entre les rémunérations Clap et fiscale
- **Priorité 1 :** + ou - 5% d'écart entre les rémunérations Clap et fiscale, mais moins de 10%
- **Priorité 0 :** écart entre les rémunérations Clap et Ésane compris entre -5 et 5%

Si les informations mises à disposition des gestionnaires Clap ne permettent pas d'expliquer les divergences, ils recontacteront, s'ils le jugent nécessaire, les entreprises. Les entreprises contactées dans cette phase seront celles pour lesquelles il y a à la fois incohérence Clap et entre Clap et Ésane. En cas de rappel des entreprises, les gestionnaires Clap ne mentionneront pas les données de la source fiscale.

En cas de restructuration, un traitement particulier tient compte de la différence de dates de référence pour la prise en compte des restructurations entre la source fiscale et les sources sociales. Ésane envoie en outre à Clap les contours des restructurations qu'il a utilisés dans un fichier spécifique. Clap signale dans un fichier de même type si des enveloppes transmises par Ésane sont incomplètes ou manquantes.

Clap signale à Ésane, dans la zone commentaires, les entreprises rappelées, à quelle date, par quelle personne, pour assurer un suivi des contacts.

4 - Travail de réconciliation par Ésane (Rédi emploi)[276]

Après l'expertise Clap : Ésane (plus précisément le sous-système réconciliation des données dit « Rédi emploi ») procède alors à une seconde mise en cohérence entre les données d'emploi et de salaires venant de Clap et celles venant de la liasse

- Pour les entreprises pour lesquelles il y a égalité, il n'y a évidemment pas de sujet.
- Pour les entreprises *en dessous* des seuils d'examen par les gestionnaires Ésane, ce sous-système aligne automatiquement les effectifs Ésane en nombre de postes sur les effectifs Clap et laisse la masse salariale inchangée (elle correspond donc à la masse salariale figurant sur la liasse fiscale). C'est le cas le plus usuel. Voir le paragraphe *8.4.1*.
  Ensuite, il calcule les caractéristiques secondaires (ventilation des salariés par quotité de temps de travail et par CSP Cf. § *7.3.2.3*) sur les niveaux de l'emploi et des salaires mis en cohérence ainsi que des caractéristiques secondaires en ETP tirées des DADS[277].

Pour les entreprises *au-dessus* des seuils d'examen (branche « arbitrage par un gestionnaire ÉSANE parisien » du schéma général ci-dessus), on a une divergence forte entre la source fiscale et les sources sociales et ces divergences n'ont pas pu être expliquées par les échanges précédents entre les équipes Clap et Ésane (faisant en principe référence aux décalages de période ou aux restructurations). Ces cas doivent donc être complexes. Ils sont centralisés à Paris et confiés aux membres de la division Ésane. C'est le cas où on a une intervention manuelle pour contrôler des variables d'emploi visant à assurer une **cohérence globale** avec la liasse fiscale. Mais il n'est pas véritablement fait appel aux entreprises.

7.2.2 Les traitements spécifiques faits dans Ésane

Au cours du second semestre, il faut réceptionner les caractéristiques en provenance de Clap, prendre en compte les éventuels problèmes d'identification repérés par Clap dans la source DADS-U,

[275] On se souvient que les gestionnaires Clap ne peuvent que choisir une des deux sources ; ils ne peuvent pas retenir un troisième chiffre.

[276] Ce travail est décrit en détail dans les *§ 8.6.*1 et *§ 8.6.2*. Il ne s'agit ici que d'un résumé qui présente la logique.

[277] Les ETP et les caractéristiques secondaires sont transmises comme on l'a vu directement par le DERA à Ésane.

mettre à niveau les caractéristiques DADS-U secondaires et passer des concepts DADS-U/Urssaf à des concepts liasse fiscale.
Enfin, il faut également fournir à Clap la liste des unités à contrôler.

7.2.2.1 Prendre en compte les éventuels problèmes d'identification repérés par Clap dans la source DADS-U

Lors de la mise en cohérence des sources DADS-U et Épure2, il est fort probable que Clap repère des erreurs d'identification dans l'une de ces deux sources (c'est-à-dire des déclarations qui ne sont pas faites sur le Siren sur lequel nous l'attendons (problème des SEP[278] ou des GIE[279] par exemple). Il est alors nécessaire de prendre en compte ces erreurs d'identification lors de l'appariement des liasses en provenance de Clap.

7.2.2.2 Passer des concepts Clap (DADS-U ou Urssaf) à des concepts liasse fiscale

Dans le cadre de Resane, des études internes à l'Insee ont été faites sur la **cohérence des sources Clap et les données fiscales**. Il est apparu qu'une majorité des divergences entre les sources s'explique par :

- des différences de concepts sur les rémunérations (i) mais pas sur les effectifs pris en compte (ii) ;
- et surtout des différences de période de référence (iii).

Afin d'éviter de contrôler des entreprises à cause de ce **second** type de divergence sur les périodes de référence, **il a été décidé que le système Ésane modifierait les caractéristiques Clap** afin de corriger, dans la mesure du possible, ces différences ; ceci sera explicité au point iii). En outre ces mises à jour ont lieu sur la réponse initiale de Clap ainsi que sur les éventuelles mises à jour renvoyées par Clap à l'issue de l'examen des entreprises à contrôler par les gestionnaires Clap.

Comme les différences de concept ne sont pas faciles à identifier, on explicite ci-dessous ces différences conceptuelles entre ces sources[280] et leur traitement en comptabilité nationale, même si in fine elles ne donnent lieu à aucune correction.

- i) Les contenus des caractéristiques **des rémunérations**, issues de ces deux sources, **ne sont pas strictement identiques**. Les principaux écarts sont les suivants (surligné en turquoise) :

[278] Sociétés en participations [même si a priori ce n'est pas une structure pour avoir des salariés puisque son principal intérêt est de ne pas être déclaré (au RCS par exemple)].
[279] Groupement d'intérêt économique.
[280] Telles qu'elles existent à fin 2014. Ce tableau n'engage évidemment pas les traitements qui pourraient être faits ultérieurement par l'Acoss ou l'Insee. Tout ceci sera également à revoir avec l'arrivée de la DSN.

Récapitulatif des écarts conceptuels entre les différentes sources

Sources	Acoss, Clap/DADS (au 31/12)	Acoss	Ésane Liasse fiscale	Comptes nationaux
Variable	**Assiette CSG-CRDS Salaires bruts**	**Assiette déplafonnée**	**Variable Salaires et traitements [FY]**	**D11**
Salaires et indemnités				
• Salaires, appointements, commissions de base, y compris heures supplémentaires et complémentaires	OUI	OUI	OUI	OUI
• Indemnités de congés payés et indemnités compensatrices de congés payés non pris	OUI	OUI	OUI	OUI
Primes et gratifications				
• 13ème mois, primes de risque, de poste, de vacances, d'ancienneté, d'assiduité	OUI	OUI	OUI	OUI
• Primes liées à un événement familial	OUI	OUI	OUI	OUI
• Primes accordées lors d'un événement exceptionnel	OUI	OUI	OUI	OUI
• Primes individuelles ou collectives liées aux résultats	OUI	OUI	OUI	OUI
Sommes liées à l'épargne salariale				
• Sommes attribuées au titre de la participation	OUI	NON	NON	OUI (2)
• Sommes versées au titre de l'intéressement	OUI	NON	OUI	OUI
Indemnités et avantages divers				
• Sommes versées à l'occasion de la modification d'un contrat de travail	OUI	OUI (3)	OUI	OUI
• Indemnités de salaires versées en cas de maladie	OUI (4)	NON (4)	NON (4)	NON (4)
• Indemnités de chômage partiel	OUI	NON	OUI	OUI
• Indemnités de licenciement, de préavis non effectué, de fin de contrat (dont rupture conventionnelle)	OUI pour Montant >Y (5)	OUI pour Montant > X (5)	OUI (5)	OUI
Avantages en nature				
• Avantages en nature (nourriture, logement véhicule...)	OUI	OUI	OUI/NON	OUI/NON
• Pourboires	OUI	OUI	OUI	OUI (6)
• Contribution patronale aux titres-restaurant	OUI	NON (12)	NON	OUI (1)
• Autres avantages en nature (voyages offerts, prix…)	OUI	OUI	OUI	OUI/NON (11)
• Supplément familial	OUI	OUI	OUI	OUI
• Frais professionnels, remboursements de frais de déplacements, missions etc.	OUI	NON	NON (en général) (13)	NON (en général) (13)
• Remboursement des dépenses de transport domicile-travail	OUI	NON pour Montant <X (7)	NON	OUI (1)
• Contributions des employeurs à des régimes complémentaires de prévoyance ou de retraite supplémentaires (à l'exception des retraites « chapeau »)	OUI	NON	NON	OUI (1)
• Avantages attribués par le comité d'entreprise	OUI	OUI	NON	OUI

Sources	Acoss, Clap/DADS (au 31/12)	Acoss	Ésane Liasse fiscale	Comptes nationaux
Variable	**Assiette CSG-CRDS Salaires bruts**	**Assiette déplafonnée**	**Variable Salaires et traitements [FY]**	**D11**
• Participation de l'employeur aux chèques vacances pour les entreprises de plus de 50 salariés	OUI	OUI	NON	OUI (1)
• Abondement de l'employeur aux PEE, PEI, PERCO ou pour la souscription d'actions de société (un abattement de 3% est applicable)	OUI	NON	NON	OUI (1)
Autres				
• Allocations de préretraite	OUI	OUI	OUI	OUI
• Les stocks options	NON	OUI(8)	OUI/NON (8)	OUI/NON (8)
• Retraites chapeau	NON	OUI	NON	NON
• Rémunération des stagiaires	OUI pour Montant > X (9)	OUI pour montant > X (9)	OUI (9)	OUI
• Rémunérations allouées aux gérants majoritaires et aux administrateurs de société, revenus d'activité que les gérants de SARL ou d'autres formes particulières de sociétés déclarent dans leur déclaration fiscale dans le poste "frais de personnel"	NON	NON	OUI pour ceux imposés, à l'IS	OUI pour ceux imposés à l'IS
• Crèches financées par l'employeur via CESU (chèque emploi services)	OUI	NON pour montant < X (10)	NON	OUI (1)

(1) Via l'écart sur charges sociales
(2) Via la variable R205 de la liasse fiscale (participation des salariés aux résultats de l'entreprise)
(3) En dehors d'un plan social
(4) Concernant la prise en charge du délai de carence, cela dépend de la convention collective et de l'ancienneté : si le salarié a l'ancienneté requise et que son salaire est maintenu (convention collective), les montants versés sont comptabilisés en « salaires et traitements » dans la liasse fiscale. Pour les autres jours de congé pris par le salarié : traitement à éclaircir à nouveau avec l'agent comptable : **paragraphe 909.4 du memento Lefebvre**
(5) Dans la comptabilité d'entreprise, les indemnités de licenciement sont comptabilisées en « salaires et traitements » dans la liasse fiscale (ou peuvent être provisionnées s'il existe une obligation à la clôture). Si le licenciement est collectif, la charge peut être classée en charge exceptionnelle et non pas en « salaires et traitements », au choix de l'entreprise. Dans le cadre d'un plan de sauvegarde, les indemnités sont totalement exonérées de cotisations sociales et donc non prises en compte dans l'assiette déplafonnée de l'Acoss. Hors de ce contexte, les indemnités sont dans l'assiette déplafonnée de l'Acoss et dans la masse salariale des DADS si elles dépassent la limite de plafonds fixés par les conventions collectives de branche ou à défaut par la loi.
(6) Un complément est effectué dans le passage au compte pour les pourboires non déclarés
(7) Si inférieur à 200 € par an et par salarié
(8) Avant 2008, les stock-options n'étaient jamais en frais de personnel. Désormais ces stock-options y sont inscrits quand il s'agit d'une charge (achat d'une action existant déjà, valorisation de la charge comme la différence entre le prix d'achat et le prix payé par le salarié) et que le bénéficiaire est salarié de l'unité légale émettrice (ce qui n'est pas toujours le cas dans un groupe). Les stocks-options figurent dans les frais de personnel de la filiale employant le salarié si l'unité légale émettrice refacture à la filiale. Les enregistrements comptables sont complexes (salaires avec provisions en contrepartie) et étalés sur plusieurs années. Des charges sociales sont systématiquement dues lors de la mise en place du plan même si finalement le cours de bourse s'effondre et que personne n'utilise l'option.
(9) Les stagiaires effectuant un stage en lien avec les études peuvent ne pas être assujettis à la sécurité sociale, CSG et CRDS. La gratification mensuelle versée aux stagiaires étudiants est exonérée de cotisations (et donc absente de l'assiette déplafonnée de l'Acoss) pour la fraction n'excédant pas le produit de 12,5 % du plafond horaire de sécurité sociale et du nombre d'heures de stage effectuées au cours du mois considéré.
(10) Les aides versées par l'employeur ne sont pas soumises aux cotisations sociales (et donc absente de l'assiette déplafonnée de l'Acoss), dans la limite d'un plafond annuel de 1830 € par salarié.
(11) Les charges qui ont le caractère d'avantages en nature peuvent être enregistrées :
- Soit en consommations intermédiaires
- Soit en salaires (et consommations intermédiaires via un transfert de charges) dans les secteurs où les avantages en nature sont importants

(12) Pour être exonérée de cotisations de sécurité sociale, la contribution patronale au financement de l'acquisition des titres restaurants doit être comprise entre 50 et 60% de la valeur nominale du titre et ne pas excéder 5,29 euros (en 2013). Au-delà, cette contribution sera prise en compte dans la masse salariale déplafonnée de l'Acoss.
(13) Les remboursements de frais au personnel (déplacements, missions…) sont en principe à classer selon leur nature, sauf les éléments ou quote-part qui pourraient être assimilés à des compléments de rémunérations et donc supporter des cotisations sociales (le classement serait alors en indemnités et avantages divers donc en salaires).

- ii) **Les salariés** pris en compte dans les effectifs des deux sources semblent en revanche identiques :

Types de salariés :	**Clap (DADS ou BRC*) Effectif au 31/12**	**Liasse fiscale 2058-C Effectif moyen du personnel**[281]
- Intérimaires	NON	NON
- Apprentis	OUI	OUI
- Élèves et étudiants qui effectuent un stage obligatoire en entreprise dans le cadre de leur scolarité, et ce, quel que soit le montant de la gratification allouée	NON	NON
- Stagiaires de la formation professionnelle	NON	NON
- Salariés percevant des sommes après la rupture de leur contrat de travail	NON	?
- VRP multicartes	NON[282]	NON
- Salariés régulièrement détachés en France en vertu de conventions ou de règlements internationaux	NON	NON (si durée < 12 mois)
- Salariés handicapés		OUI

*BRC : Bordereau récapitulatif de cotisation ; il s'agit du nom du bordereau Urssaf rempli par les entreprises

Nous n'avons pas les moyens avec les éléments à notre disposition dans Ésane de corriger les écarts conceptuels de la masse salariale mentionnés en i). Ils sont toutefois considérés comme étant du 2e ordre. **Aucune correction n'a donc lieu à ce niveau** pour harmoniser les concepts des deux sources. **Ceci peut donc expliquer donc quelques écarts résiduels entre les deux sources**. Les comptes nationaux en revanche peuvent en revanche chiffrer macro-économiquement certains d'entre eux.

- iii) Les **périodes de référence** de ces deux sources, comme on l'a vu dans l'introduction, ne sont pas identiques.

La période de référence retenue par Ésane est l'exercice comptable (qui plus est éventuellement modifié par les opérations de découpage collage). Cette période de référence est **différente** de celle des sources qui alimentent Clap et qui sont calendaires.

- Les données Urssaf constituent une source trimestrielle.
- Les DADS sont une source annuelle mais qui dispose de données individuelles au niveau salarié, précisant pour chaque salarié la ou les périodes d'emploi au cours de l'année et les rémunérations correspondantes.

Du fait d'un impact non négligeable de la période de référence sur l'effectif moyen et la rémunération au niveau des sous-classes de la NAF700, **il a été décidé de ramener, le cas échéant, les effectifs et rémunérations issus de Clap sur l'exercice comptable**.

Les principes de correction sont les suivants ; on cherche en fait **à se rapprocher des concepts de la liasse**[283] :

- Si les effectifs et rémunérations fournis par Clap sont égaux à ceux de la liasse fiscale aucun traitement n'est fait. Comme on l'a vu plus haut (étape pré Rédi dans le § *7.2.1.2*), conceptuellement ces effectifs sont différents. Toutefois, dans les cas (nombreux) où l'effectif salarié d'une unité légale reste constant dans le temps, l'effectif Clap au 31/12 sera bien égal à la moyenne des effectifs de fin de trimestre figurant dans la liasse.

- En revanche s'ils ne sont pas égaux, alors les effectifs et rémunérations pris en compte sont estimés de la façon suivante[284] :

[281] On rappelle la définition de l'emploi demandé dans les liasses fiscales : c'est la moyenne arithmétique des effectifs (personnes titulaires d'un contrat de travail directement rémunérées par l'entreprise) en fin des trimestres couverts par l'exercice fiscal.
[282] A noter qu'ils sont exclus de l'effectif au 31/12 mais inclus dans l'effectif ETP des DADS.
[283] C'est ce qui fait que la valeur de cette variable est in fine beaucoup plus proche d'un effectif moyen sur la durée de l'exercice que de l'effectif à la fin de l'exercice.

Lorsque la date de clôture d'exercice est :

- égale ou postérieure au 31/12/N, on corrigera les données Clap au 31/12 **en prenant en compte les données Urssaf si elles sont disponibles** (car la source trimestrielle Urssaf est livrée en avance par rapport aux données DADS : Cf. § 7.4). Les effectifs et rémunérations Ésane sont donc les suivants :

$$X_{ESANE} = X_{CLAP} * \frac{X_{URSSAF\ EXERCICE\ COMPTABLE}}{X_{URSSAF\ ANNEE\ CIVILE}}$$

  où $X_{URSSAF\ EXERCICE'\ COMPTABLE}$ correspond à la somme (des effectifs ou des rémunérations) des trimestres correspondants à l'exercice comptable[285] et $X_{URSSAF'\ ANNEE'\ CIVILE}$ correspond à la somme (des effectifs ou des rémunérations) des 4 trimestres de l'exercice N[286].

- antérieure au 31/12/N, **on prendra en compte les données DADS**[287] trimestrialisées pour cette correction. En effet, le DERA calcule, pour chaque unité légale, les caractéristiques DADS demandées par Ésane sur chacun des trimestres de l'année civile (Cf. § 7.3.1.1 : on a les effectifs présents à la fin de chaque trimestre). Ésane se chargera ensuite d'estimer ces caractéristiques sur les exercices comptables. La formule est donc la même que celle figurant ci-dessus en remplaçant Urssaf par DADS.

Dans les deux cas, on se sert des Urssaf ou des DADS pour calculer un **coefficient multiplicateur qu'on applique à la donnée Clap** pour tenir compte du décalage d'exercice. On ne reconstitue pas la donnée annuelle à partir de composantes trimestrielles prises dans une autre source, ce qui aurait été le cas si on avait écrit directement : $X_{ESANE}=X_{URSSAF\ EXERCICE\ COMPTABLE}$.

On note aussi que pour effectuer cette correction, on a besoin **des trois sources emploi en même temps** (Cf. schéma avant le § 7.2)

7.2.2.3 Cas particulier des entreprises d'intérim

Pour les entreprises d'intérim (division 78 de la NAF rev. 2), la liasse fiscale prend en compte les intérimaires dans les effectifs et les rémunérations.

En revanche pour Clap, les intérimaires sont bien pris en compte dans les rémunérations, mais ils ne sont pas comptabilisés dans les effectifs[288] car on sait qu'ils ne travaillent pas dans les établissements des entreprises d'intérim et c'est l'emploi local que l'on cherche à appréhender.

Afin de corriger cette incohérence, pour les secteurs dont l'APE commence par 78 (les agences d'intérim), l'effectif Ésane en nombre de postes est estimé par :

$$EFF_{ESANE} = EFF_{CLAP} + E160$$

où E160 est effectif **moyen des intérimaires** sur l'exercice comptable **livré par les DADS**.

C'est ce qui explique que **l'effectif en nombre de postes d'Ésane comprend les intérimaires alors que l'effectif ETP ne les comprend pas** (Cf. fin du § 7.3.1.1).

---

[284] On notera qu'à la suite de ces calculs, l'effectif intitulé « au 31/12 », devient fréquemment **non entier** alors que sa dimension d'emploi en personnes physiques aurait pu laisser penser qu'il devait être entier.

[285] Cette approximation est efficiente, car les dates de clôture d'exercice correspondent dans la majeure partie des cas à des fins de trimestre, comme on l'a vu au chapitre 6.

[286] Ainsi si une entreprise clôture en mars N+1 :

$$EFF_{URSSAF\ EXERCICE\ COMPTABLE} = \frac{EFF^{N}_{2e\,trim} + EFF^{N}_{3e\,trim} + EFF^{N}_{4e\,trim} + EFF^{N+1}_{1er\,trim}}{4}$$ et :

$$REM_{URSSAF\ EXERCICE\ COMPTABLE} = REM^{N}_{2e\,trim} + REM^{N}_{3e\,trim} + REM^{N}_{4e\,trim} + REM^{N+1}_{1er\,trim}$$

[287] On aurait pu évidemment utiliser exclusivement les données Épure trimestrielles, sur le passé également. Mais chaque fois qu'on le peut, on utilise les DADS, en principe plus précises et qui assurent aussi une cohérence avec les données secondaires.

[288] Seuls sont comptabilisés les effectifs permanents de ces entreprises.

7.2.2.4 Estimer l'emploi des enveloppes et des entreprises profilées

Les caractéristiques de l'emploi des enveloppes et des entreprises profilées se calculent par un cumul simple des caractéristiques de l'emploi des unités légales qui les composent.

Les enveloppes et entreprises profilées subissent également les macro-contrôles décrits ci-dessous et sont envoyés au pôle Clap pour contrôle.

Un écart important au niveau d'une enveloppe peut traduire le fait que l'enveloppe est incomplète. Il est en effet possible qu'il manque une unité légale dans laquelle est enregistré l'emploi pour les DADS alors que cet emploi a été comptabilisé dans la liasse fiscale d'une autre unité légale, liasse dans laquelle figure les rémunérations de ces effectifs.

En outre, le système gère la complétude des enveloppes (respectivement des entreprises profilées). Seules les enveloppes (respectivement les entreprises profilées) complètes subissent les macro-contrôles.
Une enveloppe (respectivement une entreprise profilée) est considérée comme complète si on dispose de toutes les liasses fiscales et de tous les effectifs en provenance de Clap des unités légales qui composent leur contour en N.

Il est en effet assez facile d'imaginer un écart important au niveau de l'enveloppe qui viendrait du fait qu'à un instant « t » on dispose d'un côté de l'ensemble des liasses fiscales des unités légales et de seulement 50% des unités légales au niveau des effectifs fournis par Clap. Ces cas ne seront bien évidemment pas envoyés à Clap à l'instant « t », on attendra d'obtenir toutes les réponses de Clap avant de leur faire subir les macro-contrôles.

7.2.2.5 Lister les entreprises à envoyer au pôle Clap

La définition des unités légales à envoyer à Clap se fait par le biais de deux macro-contrôles de type « par différence » (cf. § 2.7.4), l'un sur les effectifs et l'autre sur les rémunérations.
La formule de la contribution individuelle de ces macro-contrôles est la suivante :

$$ctr_i = \frac{\left| X_i^{Clap} - X_i^{Liasse} \right|}{\sum_i (X_i^{Liasse} * 1\mathrm{I}_{i \in groupe\,APE})}$$

C'est la même formule que celle qui est présentée au § 8.3.8 pour sélectionner les dossiers à faire contrôler par les gestionnaires Ésane
Une unité légale est envoyée à Clap si sa contribution est **supérieure à 1 %** (c'est-à-dire l'écart entre la valeur Clap et la valeur de la liasse est supérieur à 1 % de l'effectif total du groupe[289] d'APE auquel cette unité appartient).

7.2.2.6 Traiter la non-réponse pour les données d'emploi en nombre de postes

Le traitement de la non-réponse des caractéristiques « effectif Ésane » et Rémunération Ésane », [c'est-à-dire de la situation où la source Clap pour les effectifs au 31/12 ou la source liasse pour les rémunérations ne contiennent pas d'information pour une unité légale donnée], se fait **au moment de la réconciliation des données** (cf. § 8.4.1 dans le chapitre suivant). Il n'y a **pas véritablement de traitement de la non-réponse** au sein de la source « Clap » elle-même.

[289] Poste de la nomenclature ayant les 3 premiers caractères du code APE de l'unité légale considérée.

7.3 *Les données structurelles d'emploi : la fourniture des DADS (emploi en ETP)*

7.3.1 La fourniture des DADS-U à Ésane ; données d'emploi dites secondaires

Les maîtrises d'ouvrage sectorielles étaient demandeuses d'une information sur **le partage de l'emploi à temps plein et à temps partiel** ainsi que sur **des qualifications de l'emploi** (part des cadres, des ouvriers etc.). Il n'est pas possible d'obtenir cette information à partir des informations agrégées au 31/12 qui sont traitées par Clap à partir du moment où il ne s'agit pas de variables agrégées préexistantes dans les sources administratives. Par contre cette information est reconstituable à partir de la description des périodes d'emploi de chaque salarié au sein des DADS. C'est donc l'objet entre autres de cette collecte directe d'information dans les DADS.

Cette alimentation directe du système Ésane à partir des DADS, fournies par le DERA, vise donc à récupérer les **données d'emploi dites secondaires avec la variable emploi salarié en ETP,** sous forme **de données trimestrialisées** (en ETP et en fin de trimestre) pour pouvoir s'aligner sur les durées d'exercice de la liasse (Cf. § *7.2.2.2*). Comme pour les données Clap, **les données ETP trimestrialisées fournies ne tiennent pas compte des postes annexes**.

Par contre à la différence des travaux avec le pôle CLAP, il s'agit d'une fourniture directe sans aller et retour avec rétroaction du système Ésane sur les DADS. S'il y a des données manquantes, on les redressera simplement en fin de traitement, lors de l'étape de réconciliation (Cf. § *7.3.2.3*).

7.3.1.1 Variables DADS « trimestrialisées » fournies

Les livraisons faites par le DERA à Ésane sont liées à celles faites à Clap. Elles se font par lots, au fur et à mesure de l'avancement des travaux manuels des gestionnaires DADS au cours de l'année ; pour chaque lot, elles portent sur **les mêmes unités légales que celles livrées à Clap**. Dans le cas d'une nouvelle déclaration ou d'un traitement manuel, une unité peut être re-livrée dans un lot suivant.

Pour recaler les données Clap et DADS sur les exercices comptables (Cf. § *7.2.2.2* et *7.3.2.2* ci-après), Ésane souhaite disposer de données « **trimestrialisées** ». Les montants de la DADS étant annuels, cette trimestrialisation est donc conventionnelle et faite exclusivement **à la demande de la division Ésane**. Elles ne sont pas analysées et encore moins publiées par le DERA.

Le calcul s'appuie sur **le volume de travail annuel** *qui sert au calcul de l'emploi en ETP*. Il est possible d'estimer un nombre d'heures **rémunérées dans le trimestre** (au prorata de la durée passée dans le trimestre à partir des dates de début et de fin de poste, et pour les postes non annexes seulement) et **de le rapporter à une référence annuelle** divisée par 4.

La variable « ETP » annuelle qui caractérise le volume de travail est elle-même une variable « **statistique** », calculée par le DERA (et donc différente de ce qui apparaît sur le contrat de travail du salarié qui est le taux d'activité « déclaré »). Son calcul est explicité dans l'*annexe 2* (pour la statistique d'entreprise, c'est une donnée exogène).

- Exemple 1 :

Un salarié a travaillé à temps complet du 1er janvier au 31 octobre.

- il comptera donc pour 1 ETP pour les trois premiers trimestres et pour 1/3 ETP au quatrième trimestre (un tiers du trimestre).
  Son équivalent temps plein annuel sera de (1 +1+1+1/3)/4 = 10/12.

- Exemple 2 :

Un salarié a travaillé à mi-temps du 1er au 31 mai.

- il comptera donc pour 1/6 ETP au cours du second trimestre (mi-temps sur un tiers du trimestre), et 0 ETP pour les autres trimestres.
  Son équivalent temps plein annuel sera de (0+1/6+0+0)/4 = 1/24.

Cela signifie pour les DADS, que le calcul va s'appuyer sur la variable annuelle ETP, qui sera proratisée en fonction de la durée passée dans le trimestre **telle que calculée dans les exemples ci-dessus.**

Les variables suivantes sont fournies **de façon trimestrialisée** par le DERA pour chaque unité légale. Dire que ces variables sont trimestrialisées signifie que le DERA fournit un fichier où figurent ces variables **pour chacun des trimestres** ; tous les volumes d'emploi qui suivent constituent ce qu'on appelle **les données d'emploi secondaires** :

1. **Volume d'ETP estimé au cours du trimestre** (pour chaque trimestre, sur la base du calcul ci-dessus, mais **en ne retenant que les postes non annexes**) ;
   Ventilé par type de quotité de travail :
   - Volume d'ETP des postes à temps plein
   - Volume d'ETP des postes à temps partiel

   Ventilé selon le type d'emploi et la catégorie sociale des emplois ordinaires selon la nomenclature suivante :
   - Volume d'ETP des postes des apprentis
   - Volume d'ETP des postes des stagiaires
   - Volume d'ETP des postes des emplois aidés
   - Volume d'ETP des postes des agriculteurs, exploitants
   - Volume d'ETP des postes ordinaires des chefs d'entreprises
   - Volume d'ETP des postes ordinaires des cadres et professions intellectuelles supérieures
   - Volume d'ETP des postes ordinaires des professions intermédiaires
   - Volume d'ETP des postes ordinaires des employés
   - Volume d'ETP des postes ordinaires des ouvriers
2. Nombre de postes non annexes[290] (NA) présents la fin du trimestre. (qui donne un **effectif présent en fin de trimestre** calculé de la même façon que l'effectif DADS au 31/12) ;
3. Montant de la masse salariale brute estimée au cours du trimestre (selon un calcul semblable au calcul ci-dessus) ;
4. Montant de la masse salariale nette estimée au cours du trimestre ;
5. Montant de la masse salariale Brut fiscal (redressé) (estimée au cours du trimestre) ;
6. Montant de la masse salariale Brut sécurité sociale (déclaré) ;
7. Nombre d'heures rémunérées estimé par trimestre ;
8. Nombre de jours totaux rémunérés au cours du trimestre ;
9. Indicatrice de décalage de paie (% de postes non annexes en décalage de paie[291]).

Pour les variables **2, 3, 7**, on a **la même ventilation par quotité de travail et par CSP** que pour la variable 1.

Pour le secteur de l'intérim, **le volume d'ETP n'est pas calculé** (il faudrait pouvoir distinguer les permanents et les intérimaires du secteur). En effet, comme on l'a mentionné plus haut et explicité dans le § *13.2* de l'annexe 2, pour « calculer » un effectif ETP dans le secteur de l'intérim, il faudrait disposer de l'observation d'une durée moyenne du travail à temps plein dans ce secteur. Or, cette donnée n'a évidemment pas de sens dans le secteur de l'intérim puisque sa valeur dépend en réalité de secteurs réels dans lesquels sont employés les intérimaires, que l'on ignore.

Au final, les données ETP ne figurent ni dans les effectifs des entreprises qui les utilisent (elles ne les rémunèrent pas donc ne les déclarent pas), ni dans le secteur de l'Intérim (secteur des entreprises qui les payent).

On n'a donc pas les intérimaires dans les données ETP fournies par le DERA, donc in fine dans les résultats d'Ésane en ETP.

[290] La distinction des postes annexes et non annexes a été introduite par les statisticiens dans les DADS pour distinguer les postes qui contribuent véritablement au système productif, de postes beaucoup moins structurels. Ils sont définis par une quantité minimale d'heures et de jours travaillés ou une rémunération minimale. La définition précise en est donnée en *annexe 2*.

[291] Certaines entreprises sont autorisées dans les DADS à pratiquer un « décalage de paie » pour certains ou tous leurs salariés. Cela se traduit par une déclaration de rémunération versée sur l'année N, mais concernant une période de travail du 1er décembre N-1 au 30 novembre N. Environ 13% des postes sont chaque année dans cette situation. Cette pratique pose des difficultés de mesure de données pour le quatrième trimestre qui conduit à un calcul particulier non détaillé ici sur le quatrième trimestre

7.3.2 Traitements faits dans Ésane sur les données DADS

7.3.2.1 Problèmes de champ

Lors de la première étape de la collaboration avec Clap, on a identifié et on s'est mis d'accord avec le pôle Clap sur le champ sur lequel Clap doit nous envoyer des données (Cf. § 7.2.1.2 étape 1). Ce champ est fixé une fois pour toutes. Si le DERA nous envoie des unités en plus, on ne les prend pas en compte. Inversement, s'il en manque, ce sont des données manquantes dans les DADS qui seront traitées comme telles (Cf. § 7.3.2.3).

7.3.2.2 Calage sur la durée d'exercice

C'est le pendant du § 7.2.2.2 pour corriger les données lorsque l'exercice comptable ne correspond pas à l'année civile, **mais pour les données en ETP** et **les données secondaires**. A la réception des données DADS trimestrialisées fournies par le DERA, on recalcule donc à la fois les données en ETP et les données secondaires **sur la durée de l'exercice fiscal** :

- Si la date de fin d'exercice est inférieure ou égale au 31/12, on calcule directement l'emploi ETP sur la durée d'exercice à partir des données trimestrialisées (y compris en reprenant au besoin celles de l'année N-1).
- Si elle est postérieure, on n'a pas de donnés ETP sur l'année (N+1). On garde donc les ETP sur l'année N. On aura donc un léger asynchronisme entre l'ETP et la masse salariale.

7.3.2.3 Estimation de la non-réponse sur les caractéristiques DADS

Le traitement de la non-réponse des caractéristiques DADS se fait au niveau de la gestion de l'emploi Ésane, toujours à l'issue de la phase de réconciliation Rédi-emploi qui sera décrite dans le chapitre suivant (qui porte sur la réconciliation individuelle, en abrégé Rédi).

Il s'agit de traiter des unités pour lesquelles on dispose d'un effectif en nombre de postes, issu de Rédi, non nul[292] mais pour lesquelles on ne dispose d'aucune information en provenance des DADS[293].

Dans ce cas, l'imputation des données DADS secondaires se passe de la façon suivante :

- Soit on dispose des données DADS secondaires pour l'exercice N-1 et dans ce cas on calcule la donnée X en N de la façon suivante : $X_N = X_{N-1} * \frac{EFF_N^{ESANE}}{EFF_{N-1}^{ESANE}}$ si X est une caractéristique d'effectifs et on remplace EFF par REM si X est une caractéristique de rémunération. EFF^{ESANE} est bien entendu l'effectif Clap de l'unité.
- Soit on ne dispose pas de données N-1 et dans ce cas aucune imputation n'est faite et **les données sont manquantes**.

7.4 Les données conjoncturelles d'emploi (Épure2)

En dehors de l'alimentation du système par des données Clap ou DADS relatives à l'année N, on dispose aussi en fin d'année N+1 des données conjoncturelles d'emploi de l'année N+1 (et au long de l'année des versements trimestriels) pendant que l'on traite la campagne N. Ces données en provenance du système dit « Epure » sont uniquement utilisées pour réaliser des **contrôles internes** du système de production Ésane et **pour rectifier les données provenant de Clap en cas d'exercice fiscal décalé** (Cf. § 7.2.2.2). **Elles ne donnent lieu à aucune diffusion directe dans Ésane** (contrairement aux traitements décrits aux § 7.2 et 7.3) puisqu'elles ne portent pas sur la campagne en cours ; elles sont cependant utilisées pour avoir une première estimation d'un emploi « annuel ».

[292] Qui peut être un effectif Clap ou un effectif liasse, justement quand l'effectif Clap est absent ; cf. § 8.6.1.2.

[293] En s'alimentant directement auprès du DERA, on ne bénéficie pas du travail de mise en cohérence fait par les gestionnaires Clap – en particulier en cas de données manquantes dans les DADS. Or il arrive que certaines entreprises « oublient » de faire la déclaration DADS en fin d'année (surtout si elles sont déjà à jour de leur cotisation). Ce cas est donc plus fréquent que l'absence de déclaration aux Urssaf (condamnable au pénal)

L'estimation de l'emploi annuel (N+1) faite à partir des données trimestrielles (N+1) des Urssaf, au fur et à mesure que l'année (N+1) se déroule, sert donc uniquement à :

- **initialiser Ocsane en novembre** puis servir d'estimation précoce de l'emploi sur N+1 pour améliorer le **tirage l'échantillon de l'Esa N+1**[294] ;
- **contrôler des variables liées pour la campagne N** (comme le chiffre d'affaires) de l'enquête et des liasses fiscales[295] ;
- **Mettre à jour l'effectif Clap dit au 31/21 lorsque l'exercice fiscal est décalé** par rapport à l'année calendaire.

Cette estimation de l'emploi et des rémunérations se décompose en 2 étapes.

1 – Réception - chargement des données trimestrielles transmises par le système Épure (les entreprises peuvent déclarer à une ou plusieurs sources parmi Acoss, MSA[296] ou Siasp[297]) ; c'est cette étape qui sert aussi pour la rectification des exercices décalés ;

2 - Calcul des estimations annuelles de l'emploi et des rémunérations.

Deux effectifs sont présents dans Épure ; dans les deux cas, il s'agit d'effectifs en nombre de postes où chaque salarié compte pour un :

- les effectifs à la fin de chacun des trimestres ;
- le nombre de salariés rémunérés par l'entreprise au cours de chaque trimestre (indépendamment donc du fait que le salarié est encore présent à la fin du trimestre).

Deux masses salariales sont renseignées dans Épure (mais seule la première est diffusée) :

- la masse salariale déplafonnée ;
- la masse salariale CSG ; elle correspond à 98,25 % des revenus entrant dans le champ de l'abattement. Cet abattement est applicable à la fraction de la rémunération qui ne dépasse pas 4 fois le plafond de la sécurité sociale (soit 145 488 euros pour l'année 2012).

Quatre variables seront estimées :

- **L'effectif courant** : il correspond à la moyenne des effectifs à la fin des quatre derniers trimestres reçus. Le tableau précédent précise exactement de quels trimestres il s'agit. Il viendra alimenter la base Ocsane N+1 à sa création (en novembre N+1), et *servira au tirage de l'échantillon Esa relative à la campagne N+1*, car il est l'effectif le plus à jour.
- **L'effectif sur l'année** civile N: pendant les deux premiers trimestres, il est égal à l'effectif courant, puis, lorsque les 4 trimestres de l'année N sont reçus, il n'est plus recalculé. *Il servira comme variable auxiliaire* pour **la validation** du CA puis des variables sectorielles de l'Esa de la campagne N en cours.
- **L'effectif moyen au cours de l'exercice comptable**: il sera calculé en faisant la **moyenne** des effectifs à la fin de chacun des trimestres composant l'exercice comptable

  Remarque** : cet effectif **correspond exactement** à l'effectif demandé dans **les liasses fiscales.
- **La masse salariale déplafonnée versée aux salariés de l'entreprise au cours *de l'exercice comptable**. C'est cette masse salariale qui est retenue puisque c'est celle qui* se rapproche le plus de la masse salariale demandée dans les liasses fiscales. **Elle sert à mettre à jour la masse salariale issue de Clap, en cas d'exercice décalé**.

Les estimations sont faites en fonction des trimestres de l'exercice comptable disponibles. Si tous les trimestres sont disponibles[298], il suffira de faire la moyenne des effectifs et la somme des rémunérations des quatre trimestres disponibles. Si certains trimestres ne sont pas encore

[294] Les effectifs sont en effet un critère de stratification de la base de sondage (cf. § 5.1.2).
[295] En effet au moment des contrôles intra-source on ne dispose pas encore des données validées par Clap. Ceci permet aussi d'assurer le principe d'indépendance entre sources.
[296] Mutualité sociale agricole.
[297] Système d'information sur les agents des services publics.
[298] Exemple : nous faisons l'estimation V1 et l'entreprise clôture ses comptes en septembre N ; on a donc bien tous les trimestres. Réciproquement, si l'entreprise clôture ses comptes en décembre N, on n'aura tous ses trimestres que lors de l'estimation V2. En V1, on prendra le 4e trimestre N-1.

disponibles, ils seront estimés à la volée, en fonction des données N-1, et ce, pour prendre en comptes d'éventuels effets saisonniers.

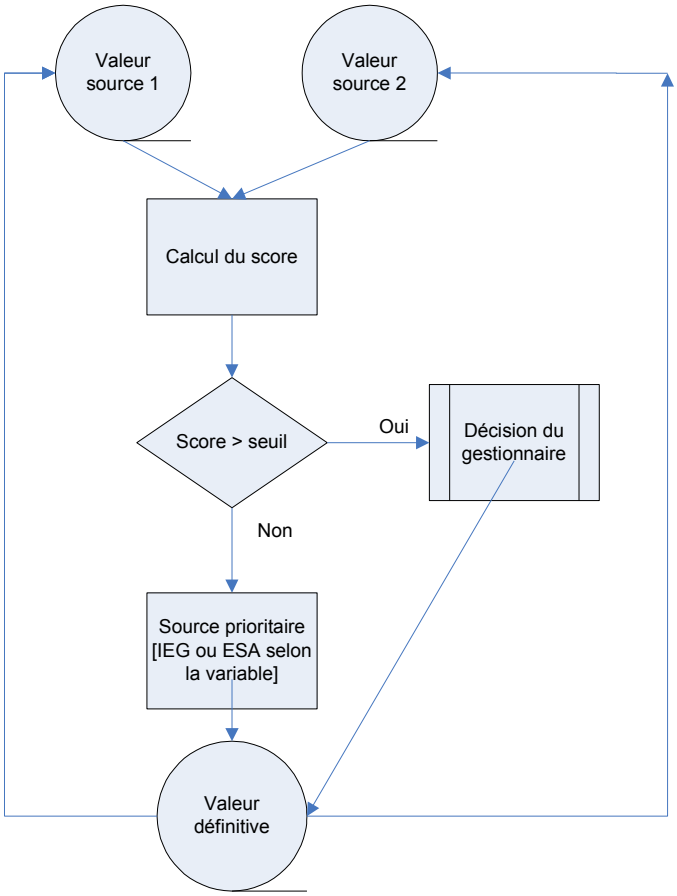
Une entreprise est soupçonnée cessée et il n'y aura donc pas d'estimation de ses trimestres absents si les deux derniers trimestres reçus (pour l'ACOSS ou pour SIASP) ou les trois derniers trimestres (pour la MSA) ont des effectifs nuls ou absents, et que cela n'est pas dû à un effet de saisonnalité.

Il est à noter que ces estimations concernent également **les enveloppes de restructuration** et **les entreprises profilées** (Cf. § *11.2*) pour lesquelles les estimations d'emploi sont faites en agrégeant les estimations d'emplois faites au niveau des unités légales qui les composent.

8. Les principes méthodologiques mis en œuvre pour la réconciliation des données individuelles (Rédi)

8.1 La cinématique générale et la présentation du chapitre

L'objectif de la réconciliation des données individuelles (Rédi) est de définir **une valeur unique** pour les variables dites d'accrochage (ou « *rédifiables* ») entre deux sources. Cette valeur est alors appelée « **valeur Rédi** » de la variable. La logique générale de cette réconciliation, schématisée ci-dessous, sera présentée au § 8.2.



En théorie les valeurs des variables qu'on cherche à réconcilier devraient être identiques dès le départ dans les deux sources mais il existe moult raisons pour lesquelles ce n'est pas toujours le cas.

Ces variables sont :

- le **chiffre d'affaires et sa ventilation** en « ventes de biens » « ventes de services » « vente de marchandises[299] » qui sont présentes dans la liasse fiscale et dans l'enquête (ESA ou EAP) ; les principes méthodologiques correspondants à la réconciliation de ces variables seront présentés au § 8.3.
- les **effectifs salariés et leurs rémunérations** qui figurent dans les données fiscales et sociales (avec leurs définitions qu'on a vues au chapitre 7). Ce sera l'objet du § 8.4.

Pour chacune de ces caractéristiques l'écart entre les valeurs des deux sources est calculé (score du graphique ci-dessus) :

- Si cet écart est jugé suffisamment important, il est demandé au gestionnaire **d'arbitrer** entre les deux valeurs, voire d'en retenir une troisième (ceci est décrit au § 8.3.8).
- Si cet écart est nul ou jugé acceptable, c'est la valeur de la **source** dite « **prioritaire** » qui est retenue (cf. § 8.3.3 pour le chiffre d'affaires et § 8.4.1 pour l'emploi)).

[299] Rappelons qu'en langage comptable, une « marchandise » est un bien revendu en l'état, donc caractéristique d'une activité commerciale.

Comme on le voit sur le schéma, la valeur définitive a vocation à être mise à jour en retour dans les sources 1 et 2. En effet, comme on va le verra dans la description du calendrier de diffusion (Cf. *§ 9.4.1*), on a besoin de continuer à travailler sur les sources après la diffusion de premières données provisoires, qui tiennent déjà compte de l'étape Rédi. Si celle-ci a modifié le chiffre d'affaires retenu, il faut en tenir compte pour le traitement du reste de la source (car le chiffre d'affaires sert à de nombreux ratios de contrôle des autres variables) et donc mettre à jour les sources initiales. Ceci sera décrit au § *8.5*.
Ce travail de réconciliation ne peut, par définition, être mené que sur les entreprises présentes dans les deux sources : il est donc, pour sa partie relative au chiffre d'affaires et à sa ventilation en activités, nécessairement limité aux entreprises faisant partie de l'échantillon des enquêtes Esa et EAP et dont les données ont été validées. Mais la réconciliation de l'emploi a vocation à se faire sur tout le champ des entreprises employeuses. Il y a donc un dernier passage de ce traitement de réconciliation, intitulé Rédi balai qui fera l'objet du § *8.6*.

8.2 La logique de la réconciliation des données individuelles ; interférence avec le calendrier

L'utilisation conjointe de données d'enquête d'une part et de données administratives d'autre part est un des points forts du dispositif Ésane, en même temps ***qu'une des innovations les plus fortes du système*** par rapport à ses précédentes versions, mais elle constitue en même temps un élément de complexité. À partir du moment où, pour une même entreprise, on peut opérer « l'accrochage » des sources, le dispositif acquiert de la robustesse : la détection d'une incohérence entre les sources conduit souvent à détecter un problème au niveau des données fournies par l'entreprise, pouvant être lié par exemple à un événement comme une restructuration.

Sur ce contrôle de données inter-sources, comme pour le contrôle des données intra-sources, on **applique des procédures de macro-contrôles** conduisant à demander une expertise manuelle des gestionnaires uniquement **pour les cas jugés les plus impactants** (cf. § *8.3.8* ci-dessous pour plus de détails). **Pour les autres cas**, s'il y a une divergence entre les deux sources, la valeur de la caractéristique réconciliée retenue au final est **celle de la source qui est jugée prioritaire** (cf. § *8.3*).

Cette réconciliation est placée **en aval du processus de production.** Ceci est la conséquence de l'un des piliers méthodologiques de Resane, à savoir **la modularité et l'indépendance des différents sous-systèmes** (Cf. § *1.2.2*). Ainsi, par principe, **on s'interdit d'utiliser**, pour la validation des caractéristiques d'une source donnée (Enquête ou liasse fiscale par exemple), des caractéristiques provenant d'une autre source pour le même exercice. On aurait pu imaginer, par exemple, que la valeur du chiffre d'affaires de l'Esa (connu avant celui de la liasse fiscale), si elle était disponible au moment du contrôle des données de la liasse fiscale, soit mobilisée afin que le gestionnaire fasse de la réconciliation individuelle en cours de traitement. Mais ce principe a été rejeté d'emblée dans Resane car contradictoire avec le principe précédent.

La réconciliation individuelle inter-source **aurait donc dû intervenir en principe en toute fin de traitement**, une fois les caractéristiques d'une unité validées dans toutes les sources concernées par les réconciliations.

En pratique, compte tenu de la nécessité de **pouvoir diffuser les résultats provisoires fin décembre** aux comptables nationaux et donc d'avoir pu faire travailler les gestionnaires sur cette phase Rédi pendant environ un mois, il est nécessaire de **prévoir une ouverture de Rédi en début novembre.** Or à cette date, il est impossible, en termes de délais, de disposer de l'ensemble des caractéristiques de l'enquête et de la liasse fiscale complètement validées, ne serait ce que parce qu'on ne reçoit la troisième livraison de la DGFiP que dans la deuxième quinzaine d'octobre. Il est donc inévitable de **prévoir un chevauchement partiel entre les traitements Rédi d'une part et les traitements** des liasses et de l'enquête d'autre part et de remonter la valeur arbitrée dans les sources que l'on continue de traiter en parallèle. Ceci sera décrit au § *8.5* comme dit plus haut.

Dans tous les cas, les valeurs diffusées (c'est-à-dire celles intervenant dans le calcul des agrégats) seront les valeurs des caractéristiques issues de la réconciliation.

8.3 Les principes méthodologiques de la mise en cohérence des données du chiffre d'affaires et de sa ventilation

La réconciliation automatique consiste à **choisir une source préférentielle** entre l'Esa et la liasse fiscale pour le chiffre d'affaires et pour sa ventilation, puis à **effectuer des redressements** de manière à obtenir des données cohérentes au sein de chaque source.

L'idée générale, qui constitue une ***innovation par rapport à l'ancien SIE*,*** est que l'on retienne **le total du CA en provenance de la liasse** (car la déclaration fiscale fait en principe foi) mais qu'on lui applique ***la ventilation de l'ESA*,*** car la définition des activités secondaires est le cœur de l'enquête et la distinction entre biens, services et marchandises est – en moyenne – mieux remplie dans l'enquête que dans la liasse fiscale lorsque l'enquête est correctement renseignée. **Dans le cas** où la liasse apparaît cependant « **plus ventilée** » que l'Esa[300] (§ *8.3.3.2*), on retient celle de la liasse (**par exception à l'idée générale**) ; dans le cas où on retient la liasse, on peut être amené à modifier cependant le classement spontané de la liasse entre biens, marchandises et services (exceptions à l'exception ; Cf. la fin du § *8.3.3.2*, cas 1 et 2 avec les exemples 5 et 6 correspondants).

Avant de passer à la réconciliation proprement dite, on s'assure que les données sont **comparables** entre les deux sources.

Concernant **le chiffre d'affaires**, on considère que celui de l'enquête et celui de la liasse fiscale sont **directement comparables** [le chiffre d'affaires est préalablement redressé sur 12 mois comme on l'a vu dans les § *6.2.7* et *6.4.2* (hors création et cessation) dans l'Esa et la liasse fiscale].

La comparaison **directe** de la **ventilation du chiffre d'affaires** est quant à elle **impossible à réaliser directement** car on a d'un côté une ventilation **détaillée en branches d'activités** dans l'Esa et de l'autre une ventilation **agrégée en trois postes** (biens, marchandises et services) dans la liasse.
En outre, certaines branches commerciales de l'Esa sont mixtes et doivent être scindées entre marchandises et services pour participer au calcul de la ventilation agrégée. C'est pourquoi, avant la comparaison des ventilations agrégées du chiffre d'affaires de l'Esa et de la liasse fiscale, il est nécessaire de réaliser certains traitements qui sont décrits ci-dessous.

8.3.1 Construction de la ventilation agrégée en trois postes pour l'Esa

Le plus grand dénominateur commun étant une ventilation en trois postes, on va devoir agréger la ventilation détaillée de l'Esa en biens, marchandises et services. Une règle a été établie pour agréger les branches en fonction de leur code d'activité de manière à se rapprocher des concepts de la liasse fiscale. Ainsi en complément des définitions des trois natures d'activité, il a été décidé, pour les données de l'Esa, de **reclasser en services** l'installation et la réparation, la promotion immobilière et les travaux d'installation ; et de **reclasser en marchandises** le commerce et la distribution d'électricité. Les règles détaillées figurent dans le tableau ci-dessous :

Code APE	Code de la branche	Nature
Tous	parmi (33, 411, 432)	Services
	parmi (3513Z, 3514Z, 3522Z, 3523Z)	Marchandise
	< 45 hors (33, 411, 432, 3513Z, 3514Z, 3522Z, 3523Z)	Bien
	= 45, 46 ou 47	Marchandise
	> 47	Services

[300] C'est fréquent lorsqu'on n'a pas la réponse à l'enquête. Mais, même en cas de réponse, il arrive que le répondant mette tout son CA dans la même branche pour aller plus vite alors que la liasse apporte une ventilation biens / marchandises / services qui est donc plus fine que celle de la liasse.

Exemple 1 : construction de la ventilation agrégée en trois postes.

Soit une entreprise A dont la ventilation de l'Esa est la suivante :

Ventilation détaillée	Chiffre d'affaires	Classement
4730Z	100	Marchandise
4632B	100	Marchandise
1071B	400	Bien
7739Z	100	Service
7712Z	200	Service
3513Z	100	Marchandise
3313Z	200	Service
3521Z	300	Bien

On aura pour cette unité : Ventes de biens = 700
Ventes de marchandises = 300
Ventes de services = 500

8.3.2 Le traitement des branches mixtes du commerce

Les branches mixtes sont des activités commerciales (au sens de la NAF) qui **mélangent** à la fois une notion **de commerce et de service**. L'objectif est de séparer au sein des branches mixtes de l'Esa le montant de chiffre d'affaires qui correspond à du commerce pur (à classer en ventes de marchandises) et le montant qui correspond à des services commerciaux (à classer en ventes de services).

Les activités d'entretien et réparation de véhicules automobiles (452), de commerce et réparation de motocycles (454) et d'intermédiaires du commerce de gros (461) **constituent les branches mixtes**. La partie commerce issue d'une branche mixte conserve le code branche d'origine et la partie service issue de cette même branche est **affectée à un code branche fictif** associé aux services.

L'algorithme s'appuie sur la comparaison avec la ventilation agrégée de la liasse fiscale pour décider du classement des branches mixtes entre ventes de marchandises et services. On ne modifie jamais le chiffre d'affaires total généré par les branches de commerce mixte, on ne fait que le répartir en deux activités le cas échéant. Le principe du traitement est décrit ci-dessous.

La ventilation agrégée de la liasse fiscale est préalablement calée sur le chiffre d'affaires de l'Esa.

8.3.2.1 Branches commerciales dans l'ESA : uniquement mixtes

Dans le cas où l'on observe, dans la décomposition en branches du chiffre d'affaires de l'entreprise dans l'Esa, **uniquement des branches mixtes,** sans branche purement commerciale alors :

i. Si le montant des ventes de marchandises de la liasse est inférieur au montant des branches mixtes de l'enquête, alors on restreint le montant des branches mixtes de l'enquête au montant des ventes de marchandises de la liasse. Le montant restant est transféré dans une nouvelle branche de services créée dans la ventilation de l'Esa. La liste des branches mixtes et des activités de services correspondantes est donnée dans le tableau Excel qui peut être consulté ou téléchargé à l'URL suivante :

   « Branches mixtes du commerce » (fichier xls)

ii. Sinon, on garde toutes les branches mixtes dans le commerce, la ventilation de l'Esa est inchangée.

Exemple 2 : cas de branches commerciales toutes mixtes.

Soient deux entreprises A (cas ii) et B (cas i) dont les chiffres d'affaires dans la liasse fiscale valent respectivement 690 et 710 et dont les ventes de marchandises de la liasse fiscale valent 100.
Leur ventilation du chiffre d'affaires dans l'Esa est la suivante :

Activité	Chiffre d'affaires de l'entreprise A	Chiffre d'affaires de l'entreprise B
1071B (biens)	400	400
7712Z (services)	200	200
4520A (branche mixte)	50	50
4612A (branche mixte)	40	60

Leur ventilation en ventes de marchandises, biens et services calculée à partir de l'Esa **en entrée de Rédi** est la suivante :

Activité	Chiffre d'affaires de l'entreprise A	Chiffre d'affaires de l'entreprise B
Ventes de biens	400	400
Ventes de services	200	210 (on transfère 10 sur les services)
Ventes de marchandises	90 (on garde les branches mixtes Esa inchangées car 90<100)	100 (on garde le montant de la liasse car 100<110)

Au niveau des branches **en entrée de Rédi** ceci se traduit par la situation suivante :

Activité	Chiffre d'affaires de l'entreprise A	Chiffre d'affaires de l'entreprise B
1071B (biens)	400	400
7712Z (services)	200	200
4520A (branche mixte)	50	45
4612A (branche mixte)	40	55
9999Q[301] (branche service associée à la branche mixte 4520A)	0	5
9999E (branche service associée à la branche mixte 4512A)	0	5

8.3.2.2 Branches commerciales dans l'ESA : mixtes et non mixtes

Dans le cas où l'on a **à la fois des branches mixtes et des branches de commerce pur** dans la décomposition en branche du chiffre d'affaires de l'entreprise dans l'Esa :

i. Si les ventes de marchandises de la liasse sont inférieures à la somme des branches de commerce pur, alors toutes les branches mixtes sont considérées comme des services et une branche service est créée dans la ventilation de l'Esa avec comme chiffre d'affaires le montant total correspondant aux chiffres d'affaires généré par les branches mixtes.
ii. Sinon :
    a. Si les ventes de marchandises de la liasse sont inférieures à la somme des branches de commerce pur et mixte, alors on restreint les branches mixtes aux ventes de marchandises de la liasse moins le total des branches pures. Le montant restant est transféré dans une nouvelle branche service créée dans la ventilation de l'Esa avec comme chiffre d'affaires le montant restant.
    b. Sinon on garde toutes les branches mixtes dans le commerce, la ventilation Esa est inchangée.

Le tableau des branches de service créées éventuellement créées associées aux branches mixtes est toujours donné dans le tableau Excel ci-dessus.

[301] cf. tableau Excel ci-dessus pour avoir la nomenclature complète.

Exemple 3 : cas d'un mélange de branches mixtes et de branches commerciales pures.

Supposons que l'on ait trois entreprises A (cas i), B (cas ii-a) et C (cas ii-b) dont les chiffres d'affaires dans la liasse fiscale valent respectivement 800, 710 et 695 et dont les ventes de marchandises de la liasse fiscale valent 100.
Leur ventilation du chiffre d'affaires dans l'Esa est la suivante :

Activité	Chiffre d'affaires de l'entreprise A	Chiffre d'affaires de l'entreprise B	Chiffre d'affaires de l'entreprise C
1071B (biens)	400	400	400
7712Z (services)	200	200	200
4520A (branche mixte)	50	50	50
4612A (branche mixte)	40	40	40
4730Z (commerce pur)	30	10	5
4632B (commerce pur)	80	10	0

Leur ventilation en ventes de marchandises, biens et services calculée à partir de l'Esa **en entrée de Rédi** est la suivante :

Activité	Chiffre d'affaires de l'entreprise A	Chiffre d'affaires de l'entreprise B	Chiffre d'affaires de l'entreprise C
Ventes de biens	400	400	400
Ventes de services	290	210	200
Ventes de marchandises	110	100	95

Au niveau des branches **en entrée de Rédi** ceci se traduit par la situation suivante :

Activité	Chiffre d'affaires de l'entreprise A	Chiffre d'affaires de l'entreprise B	Chiffre d'affaires de l'entreprise C
1071B (biens)	400	400	400
7712Z (services)	200	200	200
4520A (branche mixte)	0	45	50
4612A (branche mixte)	0	35	40
4730Z (commerce pur)	30	10	5
4632B (commerce pur)	80	10	0
9999Q (branche service associée à la branche mixte 4520A)	50	5	0
9999E (branche service associée à la branche mixte 4512A)	40	5	0

8.3.3 Le choix des sources prioritaires pour le chiffre d'affaires et sa ventilation

8.3.3.1 Pour le chiffre d'affaires

Pour le chiffre d'affaires, **la liasse fiscale est la source par défaut**. C'est-à-dire qu'à partir du moment où le chiffre d'affaires de la liasse est renseigné ou a été imputé par N-1 et qu'il n'a pas été redressé automatiquement à 0, alors **le chiffre d'affaires issu de Rédi sera systématiquement celui de la liasse**.

Le tableau ci-dessous présente les résultats observés à ce niveau lors du passage de Rédi sur l'exercice 2008. On voit que la liasse est la source utilisée dans plus de 97% des cas.

CA		choix	fréquence	% CA
source prioritaire par défaut		Liasse	97,3%	99,2%
si Esa n'est ni absente ni extrapolée et	absence de données de la liasse	Esa	0,0%	0,0%
	ou liasse extrapolée (1)	Esa	2,7%	0,8%

(1) liasse imputée par moyenne de strate mais pas par N-1 (qui reste la source par défaut dans ce cas) ou liasse redressée automatiquement avec un chiffre d'affaires à 0.

8.3.3.2 Pour la ventilation du chiffre d'affaires

Pour **la ventilation agrégée** du chiffre d'affaires, **l'Esa est la source prioritaire** (*principe général*), **quelle que soit la taille de l'entreprise**, **sauf** dans les cas particuliers suivants pour lesquels on dispose d'une ventilation dans la liasse (*exceptions au principe général*):

1. Questionnaire Esa extrapolé par hot-deck (dans le cas d'une non-réponse totale) ou redressé automatiquement par moyenne de strate (dans le cas d'une non-réponse partielle ou d'une réponse jugée aberrante) dans l'Esa. En revanche, une ventilation Esa jugée aberrante ou absente et qui a été redressée par rapport à la ventilation en N-1 de l'Esa est jugée prioritaire par rapport à la ventilation de la liasse fiscale.
2. La réponse à l'Esa correspond à l'exercice précédent (la date de clôture de l'Esa correspond à l'exercice fiscal de l'année précédente).
3. Si la ventilation détaillée de l'Esa est mono produit (ou mono branche dans le cas d'un questionnaire ventilant en branches, et non en produits) et si le chiffre d'affaires de la liasse fiscale est ventilée sur plusieurs montants (on compte uniquement les ventilations de la liasse fiscale qui ont un montant supérieur à 10% du CA, alors la ventilation retenue est celle de la liasse fiscale. Cette correction vient du fait que, dans certains cas, afin de ne pas passer trop de temps au remplissage du questionnaire Esa, certaines entreprises mettent (à tort) la totalité de leur chiffre d'affaires dans l'activité correspondant à leur APE alors que leur liasse fiscale est finalement plus conforme à la réalité.

Le fait de prendre en compte les branches de l'Esa quelle que soit la taille de l'entreprise signifie que ***les branches des petites entreprises sont prises en compte****, ce qui constitue une ***innovation très forte d'Ésane par rapport à Suse**** en termes d'impact sur les chiffres publiés. Jusque-là en effet, seules les « grandes entreprises » (c'est-à-dire environ celles de la partie exhaustive de l'échantillon) étaient prises en compte ; les petites étaient considérées comme monobranches[302].

ventilation du CA		choix	fréquence	% CA
source prioritaire par défaut		Esa	81,4%	92,5%
si liasse n'est ni absente ni extrapolée et	absence de données Esa	liasse	3,6%	0,9%
	ou unité extrapolée dans l'Esa (cas 1 décrit ci-dessus)	liasse	10,8%	4,1%
	ou date de clôture de l'Esa = année N-1	liasse	0,3%	0,1%
	ou IEG est plus ventilée que l'Esa (cas 3 décrit ci-dessus)	liasse	4,0%	2,4%

Source : Ésane 2008

Exemple 4 : cas par défaut. Priorité de la liasse fiscale pour le chiffre d'affaires et priorité de l'Esa pour la ventilation

Au niveau du Chiffre d'Affaires la priorité est donnée à la liasse fiscale

Chiffre d'Affaires de la liasse fiscale	Chiffre d'Affaires Esa	Chiffre d'Affaires Réconcilié
1000	900	1000

Pour la ventilation, la priorité est donnée à l'Esa mais elle est calée sur le chiffre d'affaires réconcilié

Ventilation agrégée	Liasse fiscale	Esa	Réconciliée
Marchandises	250	200	222[303]
Biens	550	400	445
Services	200	300	333

Ventilation détaillée	Classement (M/B/S)	Esa	Réconciliée
4730Z	M	100	111
4617A	M	100	111
1071B	B	400	445
7739Z	S	100	111
7712Z	S	200	222

[302] Concrètement, cela signifie par exemple que 100% du CA d'un coiffeur était attribué à la NAF 96.02A (coiffure) alors même que la plupart ont une activité commerciale secondaire de vente de produits de beauté, qui n'était pas prise en compte.

[303] 222 = 200 *1000 / 900 etc.

Lorsque c'est la ventilation de la liasse fiscale qui est choisie, dans le cas où il n'y a pas de branche Esa correspondant à une vente non nulle de la ventilation agrégée de la liasse fiscale (biens, marchandises ou services), la liasse fiscale ne fournit qu'un montant agrégé de ventes de biens, marchandises ou services. On a donc besoin de branches détaillées au niveau sous-classes ; **on impute donc pour cette vente une branche par défaut**, en fonction de l'APE et de la vente concernée. **Ces branches par défaut ont été fournies par les maîtrises d'ouvrage sectorielles**[304]. Pour les secteurs IAA, construction, transport, on n'a pas besoin de descendre plus fin dans la classification que la sous-classe. En revanche pour les secteurs du commerce et des services, la ventilation est codée à 7 caractères au niveau « produit ». **Un tableau Excel**[305], dont la maintenance est assurée par le Driss, donne la liste des branches et des produits (selon les cas) par défaut utilisés par le système pour cette imputation – en fonction du code APE de l'unité légale concernée.

En outre, **quand la liasse fiscale a été choisie pour la ventilation**, on modifie la ventilation agrégée de la liasse fiscale en amont dans le but **de ne pas imputer des branches jugées indésirables** (même traitement que celui qui était fait par le SIE qui était le fruit de l'expérience) dans les deux cas suivants (*exceptions à l'exception du principe général*):

1. Pour certains secteurs spécifiques (*Énergie* : APE commençant par 35 à 39, *Construction* : APE commençant par 41 à 43 et *Services immobiliers* : APE commençant par 68.1, 68.2, 68.3) on reclasse tous les montants de la ventilation agrégée dans l'item qui correspond à l'APE, comme le montre l'exemple ci-dessous

Exemple 5 : exception à l'exception dans le cas d'une entreprise d'APE spécifique.

Pour une entreprise d'APE 3522Z (distribution de combustibles gazeux par conduite), on dispose des données suivantes :

Chiffre d'Affaires de la liasse fiscale	Chiffre d'Affaires Esa	Chiffre d'Affaires Réconcilié
2000	2000	2000

Les deux chiffres d'affaires sont identiques, on a pris formellement celui de la liasse fiscale dans Rédi.

Ventilation agrégée	liasse fiscale	Esa
M	200	0
B	1500	2000
S	300	0

Ventilation détaillée	Classement M/B/S)	Esa
3522Z	B	2000

On est dans le cas où la ventilation de l'Esa est monobranche, on devrait donc choisir la ventilation de la liasse fiscale. Mais comme il s'agit d'une entreprise d'APE 3522Z, la totalité de son chiffre d'affaires est finalement mis dans cette activité. **On retrouve donc la ventilation de l'Esa**.

On obtient donc la ventilation détaillée réconciliée suivante :

Ventilation étaillée	Classement M/B/S)	Esa
3522Z	B	2000

[304] Les maîtrises d'ouvrage sectorielles sont décrites dans l'*avant-propos* et au § *4.4.3.2*.
[305] Non public.

2. Pour les autres secteurs non commerciaux (non spécifiques), on garde les ventes de marchandises de la liasse, mais on regroupe la production vendue de biens et la production vendue de services dans l'item qui correspond à l'APE. En effet, comme la distinction ventes de biens et ventes de service n'est pas considérée comme d'excellente qualité[306] dans la liasse fiscale, il a été décidé de les regrouper pour éviter de créer des branches à tort. Ainsi une entreprise des services ne pourra avoir que des ventes de services et de marchandises et les entreprises des autres secteurs n'auront que des ventes de biens et des ventes de marchandises. La branche marchandise qui est affectée au total de ventes de marchandises de la liasse fiscale est toujours donnée par le tableau Excel mentionné après l'exemple 4 ci-dessus. Ce choix permet ainsi de pouvoir continuer à calculer des marges commerciales.

Ce traitement est explicité dans l'exemple ci-dessous.

Exemple 6 : **exception à l'exception dans le cas d'une entreprise d'APE non spécifique.**

Pour une entreprise des services d'APE 5510Z (hôtels et hébergements similaires), on dispose des données suivantes :

Chiffre d'Affaires de la liasse fiscale	Chiffre d'Affaires Esa	Chiffre d'Affaires Réconcilié
5900	2800	5900

Ventilation agrégée	liasse fiscale	Esa
M	300	0
B	600	0
S	5000	2800

Ventilation détaillée	Classement (M/B/S)	Esa
5510Z11	S	2800

Le chiffre d'affaires prioritaire est celui de la liasse fiscale.

On est dans le cas où la totalité du chiffre d'affaires du questionnaire de l'Esa a été déclarée dans l'APE de l'entreprise, on choisit donc la ventilation de la liasse fiscale, plus précise, comme ventilation réconciliée. Comme il s'agit d'une entreprise du secteur des services, on va regrouper les montants des ventes de biens et services dans la vente de services. Une branche « marchandises » est imputée dans la ventilation détaillée. D'après le tableau des produits à imputer par APE (mentionné après l'exemple 4 ci-dessus), le produit principal « marchandises » pour l'APE 5510Z est le 4771Z00.

On obtient donc les ventilations réconciliées suivantes :

Ventilation agrégée	Rédi
M	300
B	0
S	5600

Ventilation détaillée	Classement (MBS)	Rédi
4771Z00	M	300
5510Z11	S	5600

[306] Parce qu'elle est faite par des comptables qui n'appliquent pas les règles des nomenclatures officielles pour distinguer les ventes de biens d'une part et de services d'autre part.

3. Pour les secteurs commerciaux enfin, on ne fait pas de modification de la ventilation agrégée de la liasse fiscale. On autorise l'imputation de branches dans les trois postes, biens, marchandises et services. Ce traitement est explicité dans l'exemple ci-dessous.

Exemple 7 : exception au principe général pour les secteurs commerciaux.

Pour une entreprise commerciale d'APE 4778B (commerce de détail de charbons et combustibles) on dispose des données suivantes :

Chiffre d'Affaires de la liasse fiscale	Chiffre d'Affaires Esa	Chiffre d'Affaires Réconcilié
300	300	300

Les deux chiffres d'affaires sont identiques, on a pris formellement celui de la liasse fiscale dans Rédi.

Ventilation agrégée	liasse fiscale	Esa
M	200	300
B	60	0
S	40	0

Ventilation détaillée	Classement (M/B/S)	Esa
4778BA1	M	300

On est dans le cas où la ventilation de l'Esa est monobranche, on choisit donc la ventilation de la liasse fiscale qui est plus détaillée. Comme il s'agit d'une entreprise commerciale et qu'il n'y a pas de branche Biens présente dans l'Esa, on va imputer une branche de biens et une de services dans la ventilation détaillée. D'après le tableau Excel des produits à imputer par APE fourni par les MOA sectorielles (mentionné après l'exemple 4), le produit principal Biens pour l'APE 4778B est le 1071CA0 et le produit principal Services est le 4941BC0.

On obtient donc la ventilation détaillée réconciliée suivante :

Ventilation détaillée	Classement (M/B/S)	Rédi
1071CA0	B	60
4778BA1	M	200
4941BC0	S	40

Nota : lorsque la prise en compte de la ventilation de la liasse fiscale conduit à un changement d'APE, **l'entreprise sera à contrôler par un gestionnaire** et ce quelle que soit sa taille.

8.3.4 Ajustements

Lors de chaque passage de la chaîne de traitement automatique Rédi et au moment de l'enregistrement de chaque modification d'un gestionnaire sur un dossier d'entreprise dans Rédi :

a. On cale la ventilation agrégée (en M/B/S) sur le chiffre d'affaires. Comme dans la majorité des cas le chiffre d'affaires retenu dans Rédi provient de la liasse fiscale et sa ventilation de l'Esa, il est fréquent que la somme de la ventilation du chiffre d'affaires Rédi ne redonne pas dans un premier temps le chiffre d'affaires Rédi. Pour remédier à ce problème, on ajuste donc la ventilation sur le chiffre d'affaires total (par une simple règle de 3).
b. On cale la ventilation détaillée sur la ventilation agrégée en ajustant séparément les branches de chaque secteur M, B et S sur son total agrégé.

Deux exemples de calage sont présentés ci-dessous.

Exemple 8 : Ajustement des ventilations de CA dans le cas général par défaut : priorité de la liasse fiscale pour le chiffre d'affaires et priorité de l'Esa pour la ventilation

Au niveau du Chiffre d'Affaires la priorité est donnée à la liasse fiscale.

Chiffre d'Affaires de la liasse fiscale	Chiffre d'Affaires Esa	Chiffre d'Affaires Réconcilié
1000	900	1000

Pour la ventilation, la priorité est donnée à l'Esa mais elle est calée sur le chiffre d'affaires réconcilié.

Ventilation agrégée	liasse fiscale	Esa	Réconciliée
M	250	200	222[307]
B	550	400	445
S	200	300	333

Ventilation détaillée	Classement (M/B/S)	Esa	Réconciliée
4730Z	M	100	111
4617A	M	100	111
1071B	B	400	445
7739Z	S	100	111
7712Z	S	200	222

Exemple 9 : Ajustement des ventilations de CA lorsque la liasse fiscale est prioritaire pour le chiffre d'affaires et pour la ventilation

L'unité a pour APE 6202A (Conseil en systèmes et logiciels informatiques). La priorité est donnée à la liasse fiscale au niveau du chiffre d'affaires.

Chiffre d'Affaires de la liasse fiscale	Chiffre d'Affaires Esa	Chiffre d'Affaires Réconcilié
1000	900	1000

Pour la ventilation, l'Esa ne donne pas de ventes de marchandises alors que la liasse si. On est donc dans le cas où il n'y a pas de branche Esa correspondant à une branche non nulle de la liasse, et dans ce cas, la priorité est également donnée à la liasse fiscale pour la ventilation. On cale la ventilation détaillée de l'Esa sur les montants de la ventilation agrégée (MBS) de la liasse fiscale. On est amené à imputer une branche marchandise dans la ventilation détaillée de cette entreprise qui appartient à un secteur non commercial.

Ventilation agrégée	Liasse fiscale	Esa	Réconciliée
M	200	0	200
B	300	200	300
S	500	700	500

Ventilation détaillée	Classement (M/B/S)	Esa	Réconciliée
4651Z01[308] (produit imputé)	M	pas de produit	200
2620Z01	B	200	300
5829A03	S	210	150[309]
6202A12	S	350	250
6203Z22	S	140	100

[307] 222 = 200 *1000 / 900

[308] Le produit principal Marchandise pour le 6202A est le 4651Z01. (Ce qu'indiquerait le tableau Excel mentionné entre les exemples 4 et 5 du § *8.3.3.2*).

[309] On applique ce qui est indiqué dans le b) : 150 = 210 * 500 / (210 +350+140) = 210*500/700 : on applique la structure des seules ventes de services au nouveau total (réconcilié) des ventes de services (500)

8.3.5 Le redressement des achats

Les achats présents dans les charges d'exploitation du compte de résultat ont une occurrence réconciliée. Ils sont constitués des achats et variations de stocks de marchandises, des achats et variations de stocks de matières premières et des autres achats et charges externes.

Ils subissent un **traitement spécifique** pour conserver équilibre avec le compte de résultat et un **ne pas perturber la marge commerciale de l'entreprise**.

On initialise les achats réconciliés par les achats de la liasse fiscale, dilatés sur le chiffre d'affaires Rédi (si celui-ci est différent du chiffre d'affaires de la liasse).

Ensuite, lorsque les ventes de marchandises de Rédi diffèrent des ventes de marchandises de la liasse **un module de redressement des achats intervient**. L'objectif de ce redressement est de **conserver** autant que faire se peut **le taux de marge**[310] **commerciale** de l'entreprise. En outre, ces redressements effectués sur les achats de marchandises ne modifient pas le total des achats car on va solder sur les deux autres postes (achats de matière première et autres achats) de façon à laisser le total inchangé.

Le détail des traitements est le suivant :

1. On détermine le taux de marge commerciale à utiliser pour le redressement. Si le taux de marge commerciale de l'entreprise (calculé sur les données de la liasse) est compris entre le premier et le troisième quartile de la distribution du taux de marge du secteur de l'entreprise, alors on garde ce taux de marge commerciale. S'il dépasse un des deux quartiles, on lui affecte la valeur du quartile le plus proche. Si le taux de marge commerciale est indéterminé, on utilise le taux de marge commerciale médian du secteur de l'entreprise.

2. On simule le redressement des achats de marchandises que l'on obtiendrait à partir des ventes de marchandises de Rédi en conservant le taux de marge commerciale défini ci-dessus. La somme des achats et des stocks de marchandises est alors égale aux ventes de marchandises divisées par 1 plus le taux de marge commerciale (Voir note de bas de page). Comme on ne souhaite pas modifier le total des achats, on solde la différence d'achats générée sur les achats de matières premières et les autres achats. Si aucun achat n'est négatif à l'issue de cette simulation, alors on retient ce redressement.

3. Si on a des achats négatifs, on réessaye l'étape 2 en utilisant le taux de marge commerciale calculé sur l'entreprise (même si celui-ci est en dehors de l'intervalle inter quartile de son secteur). Si aucun achat n'est négatif à l'issue de cette simulation, alors on retient ce redressement.

4. Sinon, on essaie d'autres méthodes :
   i. Dans le cas où seules les ventes de marchandises de Rédi sont non nulles dans la ventilation agrégée du chiffre d'affaires, on redresse les achats de marchandises en s'assurant que les achats et les stocks de matières premières sont nuls tout en gardant un total des achats inchangé (les autres achats et charges externes servent de variable d'ajustement).
   ii. Sinon, on tente de redresser à partir du taux de marge non commerciale[311]. On essaye de modifier d'abord les autres achats (non plus ceux de marchandises, mais ceux de biens et de services) puis on solde la différence sur les achats de marchandises si ceux-ci restent positifs.

5. Si malgré tout cela on obtient encore des achats négatifs, on abandonne toute réconciliation sur les achats et on garde les achats initiaux de la liasse.

[310] Le taux de marge commerciale TM est défini par la différence entre les ventes de marchandises VM et « la somme des achats et des stocks de marchandises » (A + Stock), le tout divisé par la « somme des achats et des stocks de marchandises » soit [VM-(A+Stock)]/ (A+Stock) = TM ⇔ VM = (A+Stock) (1+TM).

[311] Le taux de marge non commerciale est défini par la différence entre la production vendue de biens et services et la « somme des achats et des stocks de matières premières et autres charges externes », le tout divisé par la « somme des achats et des stocks de matières premières et autres charges externes ».

8.3.6 Le redressement de la ventilation du chiffre d'affaires et de sa décomposition des ventes selon les modalités France et exportation

Dans la liasse fiscale, le chiffre d'affaires est ventilé suivant deux critères croisés : la nature de l'activité (biens, marchandises ou services) et la destination de la production (France ou export). Cette ventilation croisée se résume de la façon suivante :

	France	Export	Total	
Ventes de marchandises (VM)	VMF	VMX	VM	(5)
Production vendue de biens (VB)	VBF	VBX	VB	6)
Production vendue de services (VS)	VSF	VSX	VS	(7)
Chiffres d'affaires nets	CAF	CAX	CA	(2)
	(3)	(4)	(1)	

On doit vérifier 7 égalités comptables reliant ces différentes variables :
VM + VB + VS = CA (1)
CAF + CAX = CA (2)
VMF + VBF + VSF = CAF (3)
VMX + VBX + VSX = CAX (4)
VMF + VMX = VM (5)
VBF + VBX = VB (6)
VSF + VSX = VS (7)

Seules les égalités (2), (3) et (4) sont assurées par les contrôles de la liasse fiscale avant Rédi.
Lors de la phase de réconciliation, on est amené à modifier (et parfois de façon considérable) la ventilation VM VB VS, et plus ponctuellement le chiffre d'affaires total CA.

L'algorithme proposé ci-dessous a pour objectif de redresser les ventilations croisées VMF, VBF, VSF, VMX, VBX et VSX tout en conservant les marges réconciliées Redi_VM, Redi_VB, Redi_VS, Redi_CAF et Redi_CAX.

On part donc d'un tableau de la forme suivante :

	France	Export	Total
Ventes de marchandises	VMF	VMX	Redi_VM
Production vendue de biens	VBF	VBX	Redi_VB
Production vendue de services	VSF	VSX	Redi_VS
Chiffres d'affaires nets	Redi_CAF	Redi_CAX	Redi_CA

On procède en 5 étapes successives. **Tant que les égalités comptables (3) à (7) ne sont pas vérifiées, on passe à l'étape suivante** (l'exemple numérique 10 qui suit illustrera les traitements proposés).

Étape 1. Si la ventilation France/Export est absente.

Ce cas intervient lorsque le chiffre d'affaires de la liasse est nul et que son occurrence réconciliée n'est pas nulle. On affecte alors toute la ventilation ventes de marchandises, biens et services dans la partie France.

Étape 2. Si la réconciliation supprime une (ou deux) branche(s) agrégée(s) (marchandises, biens ou services), l'idée est de répartir les montants de la branche supprimée vers les autres branches existantes après Rédi.

On transfère alors la ventilation France/Export de la branche supprimée par la réconciliation vers les ventilations France/Export des deux autres branches selon les marges ventes de biens, marchandises et services de Rédi.
Les ajustements se feront ensuite dans l'étape 4.

Explicitation dans le cas où la réconciliation supprime la branche « Services » :

	France	Export	Total
Ventes de marchandises	VMF	VMX	Redi_VM
Production vendue de biens	VBF	VBX	Redi_VB
Production vendue de services	VSF	VSX	**0**
Chiffres d'affaires nets	Redi_CAF	Redi_CAX	Redi_CA

On répartit le montant de la production vendue de services France *VSF* dans les deux autres branches France suivant les marges des branches Rédi :
VMF = VMF + *VSF* * Redi_VM / (Redi_VM + Redi_VB)
VBF = VBF + *VSF* * Redi_VB / (Redi_VM + Redi_VB)

Idem au niveau de la production vendue de services Export *VSX*
VMX = VMX + *VSX* * Redi_VM / (Redi_VM + Redi_VB)
VBX = VBX + *VSX* * Redi_VB / (Redi_VM + Redi_VB)

Puis, on annule la ventilation France/Export dans la branche services
VSF = 0
VSX = 0

Si les 7 égalités comptables sont vérifiées alors le traitement s'arrête là, sinon on poursuit avec l'étape suivante.

Étape 3. Si la réconciliation génère une (ou 2) branche(s) agrégée(s) (marchandises, biens ou services), on ne connaît pas la ventilation France/Export de cette branche créée lors de la réconciliation. On lui affecte par défaut la ventilation France/Export du chiffre d'affaires pour initialiser les montants sur cette branche.
L'ajustement sera assuré par l'étape 4

Exemple : La réconciliation génère une branche marchandises (Redi_VM)

	France	Export	Total
Ventes de marchandises	**0**	**0**	Redi_VM
Production vendue de biens	VBF	VBX	Redi_VB
Production vendue de services	VSF	VSX	Redi_VS
Chiffres d'affaires nets	Redi_CAF	Redi_CAX	Redi_CA

On initialise la ventilation France/Export de marchandises avec la ventilation totale France/Export :
VMF = Redi_VM * Redi_CAF / Redi_CA
VMX = Redi_VM * Redi_CAX / Redi_CA

Si les 7 égalités comptables sont vérifiées alors le traitement s'arrête là, sinon on poursuit avec l'étape suivante.

Étape 4. On effectue un calage sur marges en ajustant successivement les ventilations croisées verticalement selon les totaux France/Export puis horizontalement suivant les totaux des 3 types de ventes.
Cette technique est abandonnée si les relations comptables ne sont toujours pas vérifiées au bout de 25 itérations (ceci pour des raisons de temps de traitement d'une part et d'autre part parce qu'on a observé qu'en « médiane » l'algorithme converge en 1 itération et qu'il converge en moins de 25 itérations dans 99% des cas ; mesures faites sur 365 cas).

Étape 5. Si les étapes précédentes n'ont pas suffit à assurer les 7 égalités comptables, on n'utilise plus les ventilations issues de la liasse fiscale et on applique la ventilation marginale :

Redi_VMF = Redi_VM * Redi_CAF / Redi_CA
Redi_VBF = Redi_VB * Redi_CAF / Redi_CA
Redi_VSF = Redi_VS * Redi_CAF / Redi_CA
Redi_VMX = Redi_VM * Redi_CAX / Redi_CA
Redi_VBX = Redi_VB * Redi_CAX / Redi_CA
Redi_VSX = Redi_VS * Redi_CAX / Redi_CA

Cette étape assure les égalités comptables, mais on fait l'hypothèse forte que les ventilations observées sur les marges sont identiques au sein de la « matrice ».

Exemple 10 : Cas standard ; pas de création ni de suppression de branche au cours de la réconciliation ; illustration des étapes 4 et 5.

L'entreprise suivante ayant le CA issu de la liasse et la ventilation issue de l'enquête se retrouve avec une ventilation déséquilibrée :

	France	Export	**marges Redi**	marges de la liasse fiscale
Marchandises	1 703 473	0	**1 736 689**	1 703 473
Biens	0	0	**0**	0
Services	6 168	44 591	**17 542**	50 759
marges Redi	**1 709 641**	**44 591**	**1 754 232**	1 754 232
marges de la liasse fiscale	1 709 641	44 591		

Étape 1 :
On passe à l'étape suivante.
Étape 2 : la réconciliation ne supprime pas de branche.
On passe à l'étape suivante.
Étape 3 : la réconciliation ne crée pas de branche.
On passe à l'étape suivante.

Étape 4 :
i=1
Calage sur les marges horizontales (MBS)
On multiplie les lignes par le rapport de la marge Redi sur la marge de la liasse fiscale :

	France	Export	**marges Redi**	marges de la liasse fiscale
Marchandises	1 736 689	0	**1 736 689**	1 736 689
Biens	0	0	**0**	0
Services	2 132	15 411	**17 542**	17 542
marges Redi	**1 709 641**	**44 591**	**1 754 232**	1 754 232
marges de la liasse fiscale	1 738 821	15 411		

Calage sur les marges verticales (France/Export)
On multiplie les colonnes par le rapport de la marge Redi sur la marge de la liasse fiscale

	France	Export	**marges Redi**	marges de la liasse fiscale
Marchandises	1 707 545	0	**1 736 689**	1 707 545
Biens	0	0	**0**	0
Services	2 096	44 591	**17 542**	46 686
marges Redi	**1 709 641**	**44 591**	**1 754 232**	1 754 232
marges de la liasse fiscale	1 709 641	44 591		

Les égalités ne sont pas vérifiées et i<30 donc on recommence l'opération de calage

(Suite de l'exemple 10)
Les égalités comptables ne sont pas vérifiées, on doit redresser la ventilation croisée.

i=2
[…]

On s'aperçoit que l'algorithme ne pourra jamais converger car on souhaite avoir simultanément
VMF=1 736 689 (contrainte horizontale)
VMF<1 709 641 (contrainte verticale)

[…]

i=25
Calage sur les marges horizontales (MBS)
On multiplie les lignes par le rapport de la marge Redi sur la marge de la liasse fiscale

	France	Export	**marges Redi**	marges de la liasse fiscale
Marchandises	1 736 689	0	**1 736 689**	1 736 689
Biens	0	0	**0**	0
Services	0	17 542	**17 542**	17 542
marges Redi	**1 709 641**	**44 591**	**1 754 232**	1 754 232
marges de la liasse fiscale	1 736 689	17 542		

Calage sur les marges verticales (France/Export)
On multiplie les colonnes par le rapport de la marge Redi sur la marge de la liasse fiscale

	France	Export	**marges Redi**	marges de la liasse fiscale
Marchandises	1 709 641	0	**1 736 689**	1 709 641
Biens	0	0	**0**	0
Services	0	44 591	**17 542**	44 591
marges Redi	**1 709 641**	**44 591**	**1 754 232**	1 754 232
marges de la liasse fiscale	1 709 641	44 591		

Les égalités ne sont pas vérifiées et i=30
On abandonne le calage sur marge et on passe à l'étape suivante

Étape 5 :

On construit la ventilation marginale

	France	Export	**marges Redi**	marges de la liasse fiscale
Marchandises	1 692 545	44 145	**1 736 689**	1 736 689
Biens	0	0	**0**	0
Services	17 096	446	**17 542**	17 542
marges Redi	**1 709 641**	**44 591**	**1 754 232**	1 754 232
marges de la liasse fiscale	1 709 641	44 591		

Les égalités sont toutes vérifiées, **on arrête le redressement et on enregistre ces données avec le statut réconcilié**

8.3.7 Le redressement du compte de résultat (suite à modification du chiffre d'affaires de la liasse)

Lorsque le chiffre d'affaires retenu n'est pas celui de la liasse fiscale, mais celui de l'enquête, alors le compte de résultat est corrigé a minima afin qu'il reste cohérent. Comme on l'a vu ci-dessus, si le chiffre d'affaires ou sa ventilation change, les achats et les ventilations des ventes en France et exportées sont mis en cohérence. Mais les variables soldes contenant le chiffre d'affaires doivent également être recalculées et notamment le total des produits.
À partir de là, **le principe de base mis en œuvre est de ne pas toucher au résultat net** et de solder selon les cas sur les dotations aux provisions ou sur les reprises sur amortissements et provisions.

Ainsi on recalcule le total des charges comme étant le résultat de la différence entre le résultat net (inchangé) et le total des produits (modifié suite à la modification du chiffre d'affaires).

La différence entre ce nouveau total des charges et celui obtenu en sommant toutes les charges individuelles (y compris les achats recalculés) est soldé dans la variable « dotations aux provisions ».

Deux solutions sont alors possibles :

- Soit les dotations aux provisions ainsi calculées restent positives. Dans ce cas le traitement s'arrête
- Soit les dotations aux provisions deviennent négatives. Dans ce cas :
  - ces dernières sont mises à 0 ;
  - les soldes des charges sont recalculés ;
  - on recalcule le total des produits comme étant la somme entre le résultat net et le nouveau total des achats ainsi calculé ;
  - l'écart entre l'ancien et le nouveau total des produits (qui est forcément positif) est solé soldé dans la variable « reprise sur amortissements et provisions »

Rien de tel qu'un exemple numérique pour mieux comprendre :

Exemple 11 : redressement du compte de résultats.
Imaginons le compte initial suivant (avec deux hypothèses pour les dotations aux provisions):

Produits (hypothèse 1)	Charges (hypothèse 1)	Produits (hypothèse 2)	Charges (hypothèse 2)
Chiffre d'affaires = 100	Achats = 80	Chiffre d'affaires = 100	Achats = 80
Autres produits = 100 *- Dont* *Reprises sur amortissements et provisions = 10*	Autres charges = 90 *- Dont* *Dotations aux provisions = 20*	Autres produits = 100 *- Dont* *Reprises sur amortissements et provisions = 10*	Autres charges = 70 *- Dont* *Dotations aux provisions = 0*
Total des produits = 200	Total des charges = 170	Total des produits = 200	Total des charges = 150
Résultat net = 30		**Résultat net = 50**	

Supposons que le chiffre d'affaires finalement retenu dans Redi **soit celui de l'Esa égal à 80**.
Le total des achats recalculé est de 64 (le total des achats est recalculé proportionnellement au chiffre d'affaires).

Le compte définitif est le suivant (en bleu les corrections) :

- Dans l'hypothèse 1, pour maintenir le résultat net à 30, on fait passer le total des charges à 150. le recalcul des achats fait alors passer les autres charges de 90 à 86. On doit donc imputer ces 4 aux dotations aux provisions qui passant de 20 à 16 restent positives.
- Dans l'hypothèse 2, si on faisait le même calcul après la modification des achats à 64, les dotations aux provisions qui étaient nulles deviendraient négatives à -4. C'est donc le total des charges (134) qui va servir de pivot pour conserver un résultat net de 50. Le montant des produits devient donc égal à 184. Et c'est la variable reprise sur amortissement qu'on augmente de 4 pour assurer la cohérence de la somme des produits.

Produits (hypothèse 1)	Charges (hypothèse 1)	Produits (hypothèse 2)	Charges (hypothèse 2)
Chiffre d'affaires = **80**	Achats = **64**	Chiffre d'affaires = **80**	Achats = **64**
Autres produits = 100 *- Dont* *Reprises sur amortissements et provisions = 10*	Autres charges=86=(150-64) *- Dont* *Dotations aux provisions =**16**=[20-(90-86)]*	Autres produits = **104** *- Dont* *Reprises sur amortissements et provisions = **14** = 10+4*	Autres charges = 70 *- Dont* *Dotations aux provisions = 0*
Total des produits = **180**	Total des charges = **150** = (180-30)	Total des produits = **184** =134+50	Total des charges = **134** =(64+70)
Résultat net = 30		**Résultat net = 50**	

8.3.8 Rappel sur le principe de contrôle des données par les gestionnaires

Dans la logique des traitements Ésane, des macro-contrôles sont mis en place pour sélectionner les entreprises pour lesquelles la différence entre les données de la liasse et de l'Esa (respectivement des données sociales) est non négligeable vis-à-vis de l'agrégat du secteur. On contrôle de cette façon les écarts au niveau du chiffre d'affaires et au niveau de la ventilation agrégée en 3 postes. Pour une variable Y donnée, on calcule la contribution d'une unité de la façon suivante :

$$ctr = \frac{abs(Y_ESA - Y_IEG) \times poidsESA}{\sum_{strate} Y_IEG}$$

ou

$$ctr = \frac{abs(Y_emploi - Y_IEG) \times 1}{\sum_{Strate} Y_IEG}$$

Y_IEG désigne la valeur de la caractéristique Y *dans la liasse fiscale* (soit une composante du chiffre d'affaires, soit l'emploi de la liasse fiscale soit la masse salariale). Du côté de l'emploi, comme on est sur des sources exhaustives, il n'y a pas de poids (poids =1).
La strate est déterminée en fonction du niveau d'exigence souhaité pour les macro contrôles. Les macro-contrôles au niveau groupe de la nomenclature puis au niveau sous-classe sont effectués.

Une entreprise est **à contrôler manuellement** quand la contribution de l'entreprise pour un macro-contrôle **dépasse un seuil lâche de 1%.** Un deuxième seuil **dit strict et égal à 2%** permet également de repérer les entreprises fortement contributrices. Ce deuxième seuil est utilisé pour le calcul d'un indicateur de priorité de traitement qui permet de classer les questionnaires à contrôler en fonction de la gravité globale du dossier, à l'instar de ce qui est fait pour les contrôles intra-source. Ainsi, le nombre de macro-contrôles actifs et le dépassement d'un seuil strict influent positivement sur l'indicateur de priorité synthétique. La formule explicite de l'indicateur de priorité synthétique est la suivante :

On définit d'abord un état global quantitatif QEG pour mesurer le degré d'influence globale des données de l'entreprise par rapport à son secteur :

$$QEG=\frac{3*\sum_{strict}\left(2*1_{ventil}+1_{CA}+1_{eff}+1_{rem}\right)+\sum_{lâche}\left(2*1_{ventil}+1_{CA}+1_{eff}+1_{rem}\right)}{27} \quad , 0<QEG<1$$

Avec :

- 1_{ventil} qui vaut 1 pour un macro-contrôle actif sur une des trois variables de la ventilation agrégée et 0 sinon ;
- 1_{CA} qui vaut 1 pour un macro-contrôle actif sur le chiffre d'affaires (et 0 sinon) ;
- 1_{eff} qui vaut 1 pour un macro-contrôle actif sur l'effectif en nombre de postes (et 0 sinon) ;
- 1_{rem} qui vaut 1 pour un macro-contrôle actif sur les rémunérations (et 0 sinon) ;
- $\sum_{strict}$ la somme sur tous les macro-contrôles qui sortent avec un seuil strict de 2%.
- $\sum_{lâche}$ la somme sur tous les macro-contrôles qui sortent avec un seuil lâche de 1%.

(Les deux sommes sont en réalité exclusives puisqu'on considère, assez logiquement, qu'un macro contrôle ne sort avec un seuil « lâche » que s'il ne sort pas avec un seuil strict).

Le « 27 » au dénominateur permet de « normer » cet indicateur entre 0 et 1. En effet le maximum que peut atteindre le numérateur est 27[312] (cas où les écarts des 3 ventilations du chiffre d'affaires, des effectifs et des rémunération dépassent le seuil strict).

La formule montre qu'un macro-contrôle activé à la suite du dépassement d'un seuil strict (2%) est 3 fois plus important à traiter qu'un macro-contrôle activé par un seuil lâche (1%). De même, un macro-contrôle actif sur une variable de la ventilation agrégée est 2 fois plus important à traiter qu'un macro-contrôle actif sur les variables d'emploi ou sur le chiffre d'affaires total.

On définit ensuite un état global qualitatif EG à l'aide d'une codification en tranches de QEG, par exemple :

EG=P(rioritaire) si	*QEG* > 0.4,	(au moins 2 macro-contrôles stricts sont actifs)
EG=I(mportant) si	0.40> *QEG*>0.25,	(au moins 1 macro-contrôle strict est actif)
EG=S(econdaire) si	0.25>=*QEG*>0.1	(au moins 1 MC lâche sur le CA ou 2 sur la ventilation)
EG=A(utomatisable) si	0.10>=*QEG*	

Les valeurs de la codification en tranche du QEG affichées ci-dessus sont données à titre d'exemple puisque celles-ci peuvent varier d'une campagne à l'autre et d'une division Ésane à l'autre. En effet, **la définition des tranches du QEG permet de calibrer le nombre de dossiers à traiter** par les gestionnaires selon les priorités pour une division et une campagne donnée.

Les entreprises sont traitées par les gestionnaires par ordre de priorité. Les entreprises dont l'état global est **automatisable** ne sont **pas traitées manuellement**.

Enfin, pour éviter de traiter manuellement des petites entreprises de l'échantillon (dont la forte contribution serait essentiellement due à leur fort poids de sondage), on s'assure que l'écart entre les chiffres d'affaires de l'Esa et celui de la liasse est supérieur à 10% du chiffre d'affaires réconcilié. **Les unités ne vérifiant pas cette condition ne sont pas traitées manuellement**. Il est en effet difficile

[312] Il y a en effet 3 postes de la ventilation agrégée (biens, marchandises et services). Si une unité a des écarts forts (qui dépassent le seuil strict) pour toutes ses variables Redi, c'est-à-dire les 3 composantes du chiffre d'affaires, le chiffre d'affaires total, l'effectif au 31/12 et les rémunérations, son numérateur sera au maximum égal à 27 [3*(2+2+2+1+1+1)].

de contacter une entreprise pour lui demander une explication sur la différence de chiffre d'affaires entre sa déclaration fiscale et sa déclaration à l'enquête si l'écart entre ces deux valeurs est faible.

Le Service Ésane traite **manuellement** les cas de divergence au vu de ceux déjà effectués sur l'entreprise **lors des campagnes précédentes** ou en **recontactant directement** l'entreprise concernée. Lors d'une réconciliation manuelle, les gestionnaires disposent des données de l'Esa et de la liasse fiscale[313] pour mener à bien leur expertise. Ils ont la possibilité de valider le chiffre d'affaires issu de la réconciliation automatique ou **bien de saisir un nouveau chiffre d'affaires**. La ventilation du chiffre d'affaires est traitée au niveau détaillé. Les variables liées au chiffre d'affaires et à sa ventilation (sous-traitance incorporée et charges d'exploitation) peuvent aussi être modifiées manuellement. Après validation du dossier par un gestionnaire, **aucun traitement automatique ne peut modifier a posteriori** les données saisies manuellement.

8.4 *Les principes méthodologiques de la mise en cohérence de l'emploi*

L'objectif de cette réconciliation n'est pas d'assurer une égalité parfaite avec les autres publications faites par l'Insee sur l'emploi. On a vu au § *7.1.2* que c'est impossible. **Les agrégats de référence sur l'emploi restent ceux calculés par le DERA**.

L'idée est de bénéficier au mieux des traitements réalisés par le pôle Clap au niveau des données individuelles et d'obtenir ainsi des caractéristiques **fiscales et d'emploi cohérentes sur le champ Ésane, les deux étant liées par l'intermédiaire de la masse salariale**. On a d'ailleurs vu dans le chapitre 7 (cf. *§ 7.1*) les problèmes posés par les différents concepts d'emploi et l'impossibilité de les rendre cohérents avec les concepts Ésane du fait des décalages temporels.

8.4.1 Les sources prioritaires dans « Rédi-emploi »

Pour la confrontation des effectifs et des rémunérations des sources administratives sociales et fiscales, la **priorité sera donnée à l'effectif en nombre de postes de Clap, à l'effectif ETP des DADS et à la rémunération des liasses fiscales**.

Ce choix résulte des traitements statistiques faits sur chacune des trois sources mais il impose de vérifier ex post quelques règles de cohérence qui pourront donner lieu à des redressements qui seront l'objet du § *8.4.2*. Dans Clap, la priorité est donnée aux effectifs et non aux rémunérations et inversement dans la liasse fiscale où en dehors des rémunérations, aucune caractéristique n'est organiquement liée aux effectifs ; les DADS sont la seule source permettant de calculer un effectif ETP. **Aucun contrôle ne peut être véritablement fait dans la liasse fiscale sur les effectifs**[314]. En effet, les effectifs n'intéressant pas directement les impôts, la DGFiP n'attache guère d'importance à la qualité de cette variable, ce qui n'incite pas les entreprises à y répondre de la meilleure façon possible contrairement aux variables sur lesquelles se calcule l'impôt par exemple. De fait cette variable est même quelquefois absente dans les liasses.

Les règles de priorité dans le détail sont exposées ci-après. Comme on l'a mentionné au § *7.1.2*, il n'y a pas de contrôle manuel de l'emploi exercé par les gestionnaires Ésane et tous ces traitements sont donc automatiques.

8.4.1.1 Pour les rémunérations

Pour les rémunérations, la liasse fiscale est donc la source par défaut.

Les rémunérations provenant de ***la source emploi ne sont privilégiées*** que dans les deux cas de figure suivants :

i) s'il y a des rémunérations non nulles dans la source emploi et des rémunérations nulles dans la liasse fiscale ;
ii) si la liasse fiscale était absente et a été imputée « active » (Cf. § *6.5.3*) et que l'on dispose par ailleurs de données d'emploi provenant de Clap pour cette unité.

[313] Les entreprises sont informées par courrier (et sur le site Insee.fr) de ce que les gestionnaires Esa de l'Insee ont connaissance de leur déclaration fiscale. Historiquement, aucune entreprise ne s'en est d'ailleurs offusquée.
[314] Rappelons que la notion d'effectifs que les entreprises doivent déclarer sur leur liasse fiscale est un effectif moyen égal à la moyenne arithmétique des effectifs à la fin de chacun des trimestres de l'exercice comptable. L'effectif comprend l'ensemble des personnes titulaires d'un contrat de travail, rémunérées directement par l'entreprise.

Pour le traitement de la non-réponse de la liasse fiscale (cas ii)**,** c'est-à-dire pour l'imputation en fin de campagne de toutes les liasses manquantes par rapport au champ « Ésane » initial, il s'agit de savoir si la liasse est manquante mais l'entreprise active (auquel cas, il faut l'imputer) ou au contraire si elle n'a plus d'activité (auquel cas il ne faut pas : on appelle cela les « imputées cessées »). À compter de 2012, on a introduit dans Sirus (Cf. § *11.7.1*) les deux notions d'unités actives sans activité économique et d'unité cessée statistiquement (mais pas encore juridiquement). Le fait qu'une unité ait déclaré des effectifs montre bien qu'elle est active mais ne signifie pas qu'elle a une activité économique ; une unité peut continuer à employer des personnes pour gérer la liquidation économique par exemple. Cet effectif ne nous intéresse pas d'un point de vue économique. Lorsqu'une entreprise a été jugée sans activité économique (donc pas imputée active), c'est-à-dire en pratique **si elle n'a pas rempli de liasse fiscale p**our les **deux** derniers exercices (et qu'elle n'a fait **aucune déclaration de TVA durant les 2 dernières années**), on considère qu'elle est sans activité économique ; on ne lui impute donc pas de chiffre d'affaires[315], ni de valeur ajoutée et **ses effectifs sont mis à 0, même s'il y en a dans Clap**, afin d'assurer une cohérence globale de l'entreprise. **Ces unités légales sont de fait hors du champ de la diffusion Ésane**.

Nota : au paragraphe ci-dessus, on a écarté de l'imputation de rémunérations et donc d'emploi les unités légales inactives pour lesquelles on n'avait pas de liasse. Mais on reçoit également un assez grand nombre de **liasses avec un CA nul voire négatif**. Dans ce cas, l'unité légale reste bien dans le champ de la publication « en unités légales » faite par l'Insee. Entre 50 et 100 000 emplois correspondent chaque année à des unités légales de ce type[316].

8.4.1.2 Pour les effectifs (en nombre de postes et ETP)

Pour les effectifs dits en nombre de postes[317], la source prioritaire est Clap.

Pour les effectifs ETP, **il n'y a que la source DADS**. On n'a donc pas vraiment de problème de réconciliation ; mais les effectifs ETP sont liés aux effectifs en nombre de postes et Rédi va jouer le rôle de traitement de la non-réponse globale pour toutes les variables d'emploi.

La source CLAP n'est pas retenue pour les effectifs en nombre de postes et ***on ne retient à la place l'emploi de la liasse fiscale*** que si simultanément :

- Clap n'a fourni aucune information sur l'entreprise en question ;
- l'emploi de la liasse fiscale est strictement positif ;
- et la liasse n'a pas été imputée par la médiane de strate (pour les créations et les entrées de champ (Cf. § *6.5.3*).

Sinon, il est à « valeur manquante ».

Comme on l'a vu précédemment au chapitre 7, c'est donc Rédi qui joue le rôle de traitement de la non-réponse totale pour l'emploi (Cf. *§ 7.2.2.5* ou § *7.3.2.3*). En effet, l'effectif de la liasse qui est retenu, si on ne dispose pas d'information de la part de Clap, sera ensuite utilisé pour permettre de redresser les autres variables d'emploi (caractéristiques DADS secondaire) qui existent dans les bases Ésane.

On notera toutefois que ces règles amènent à considérer que l'emploi des unités sans données Clap et dont la liasse est redressée par la médiane de strate **est à valeur manquante**. L'emploi estimé par une liasse médiane a été jugé de trop mauvaise qualité pour être retenu.

[315] L'introduction du concept d'unité cessée statistiquement est récente dans Sirus, mais ces conditions d'imputation des liasses manquantes sont bien mises en œuvre depuis la première campagne 2008.

[316] Cette définition du hors champ peut être par la suite « complétée » et adaptée par l'utilisateur : à compter des résultats 2010, dans la diffusion qui est faite pour les résultats en « entreprises » au sens du décret LME, les unités légales ayant un CA négatif ou nul, qui faussent la compréhension des ratios économiques que l'on peut calculer, sont enlevées du champ, indépendamment de leur nombre de salariés. On « perd » ainsi ces 50 à 100 000 salariés.

[317] Rappelons (Cf. § *7.2.2.2*) que les effectifs en question sont en réalité une estimation de l'emploi de la liasse fiscale c'est-à-dire une estimation, sur la durée de l'exercice fiscal, de la moyenne des emplois en personnes physiques en fin de trimestre. L'exercice fiscal ne coïncide pas lui-même toujours avec l'année civile (Cf. § *6.2.7.1*).

8.4.2 Les redressements des variables d'emploi dans Rédi

Seules **trois variables** relatives à l'emploi sont **présentes dans Rédi** (les rémunérations, les effectifs Clap en nombre de postes et les effectifs ETP) ; comme pour l'Esa, la mise en cohérence des autres données d'emploi (caractéristiques DADS secondaires) se fait après rétrocession des données Rédi à la source emploi (voir le schéma de la fin du § 7.1 et le § 8.5).

8.4.2.1 Redressement des effectifs lorsque l'effectif de la liasse fiscale a été jugé prioritaire

Dans le cas où la source prioritaire pour la mise à jour de l'effectif est la liasse fiscale (c'est-à-dire que les conditions du § 8.4.1.2 sont réunies), cet effectif de la liasse fiscale[318] peut être redressé automatiquement s'il est jugé aberrant par rapport aux rémunérations. Comme ceci a déjà été évoqué précédemment, ce redressement se justifie par le fait que les rémunérations sont jugées de bien meilleure qualité que les effectifs sur la liasse fiscale. Les redressements sont les suivants :

- si les rémunérations sont à 0, alors on redresse l'effectif Rédi (en nombre de postes) à 0 ; il s'ensuit que l'effectif ETP sera également mis à 0 (Cf. § 8.4.2.3).
- si le ratio rémunération annuelle /effectif ETP n'est pas compris dans un intervalle entre 7 k€ et 62 k€ [319]alors on redresse l'effectif selon la formule suivante :

  $Effectif = \frac{rémunération}{salaire\ moyen}$. Le salaire moyen a été estimé à 23 K€ pour les premières campagnes (il évoluera dans le temps avec l'évolution de l'environnement économique).

8.4.2.2 Redressement de la non-réponse sur l'effectif ETP

Lorsque l'effectif ETP est manquant[320], il y a deux cas à considérer selon que l'emploi en nombre de postes de Rédi est rempli[321] ou non :

i) S'il est rempli, il joue le rôle de variable pivot dans le redressement de la non-réponse. Selon que l'effectif Clap est présent ou manquant, l'effectif Rédi contiendra l'effectif Clap ou l'effectif de la liasse ; mais dans tous les cas, **l'effectif ETP sera imputé** à partir de l'effectif 31/12 Rédi et sa valeur est obtenue en multipliant l'effectif Rédi par un coefficient (calculé évidemment sur les unités pour lesquelles on dispose des deux données) :

$$EFF_{ETP} = EFF_{REDI} * \frac{\sum_i EFF_{ETP}^{DADS} * 1\!\!1_{i \in secteur}}{\sum_i EFF_{31/12}^{CLAP} * 1\!\!1_{i \in secteur}}$$

où le secteur correspond au secteur des divisions sectorielles du service Ésane de Nantes d'appartenance de l'unité dont l'effectif est à imputer à savoir :

- BTP ;
- Transport ;
- IAA ;
- Services aux particuliers ;
- Services aux entreprises ;
- Commerce de détail ;
- Commerce de gros.

Le calcul du ratio moyen a été fait à ce niveau de la nomenclature principalement pour des raisons de facilités « informatiques » et non pour des raisons statistiques. **Il est recalculé à chaque nouvelle campagne sur les données disponibles en N-1**.

[318] Cf. la description figurant dans le 2e tableau du § 7.2.2.2 pour une description détaillée de son contenu.

[319] Il s'agit des 5e et 95e percentiles de la distribution du ratio. [Attention : ce redressement n'est effectué que lorsque l'effectif de la liasse fiscale est prioritaire !].

[320] Comme mentionné en note de bas de page au début du § 7.2.1.1, il existe quelques trous de déclaration dans les DADS. Lors de la fourniture directe des DADS par le DERA, on ne bénéficie pas de la mise en cohérence faite par le pôle Clap ; ces trous subsistent donc jusqu'à ce stade.

[321] C'est-à-dire que soit l'emploi Clap est présent soit les conditions du § 8.4.1.2 sont remplies pour qu'on ait retenu l'emploi de la liasse.

ii) Si l'emploi en nombre de postes de Rédi n'est pas rempli (c'est-à-dire si les conditions du § *8.4.1.2* ne sont pas réunies), l'emploi en nombre de postes et l'emploi ETP sont à « valeur manquante »[322].

8.4.2.3 Redressement de l'effectif ETP

On corrige éventuellement automatiquement l'effectif ETP (redressement de mise en cohérence) dans deux cas :

i) des unités pour lesquelles l'effectif en nombre de postes a été mis à 0 dans Rédi et pour lesquelles on disposait d'un effectif ETP non nul. Dans ce cas, **l'effectif ETP est mis à 0**.
ii) des unités pour lesquelles l'effectif ETP est à 0 alors que l'effectif salarié en nombre de postes correspondant est supérieur à 1, on **redresse l'effectif ETP** avec la même formule que ci-dessus au § *8.4.2.2* (cas de l'effectif Clap présent).

8.4.3 Le redressement du compte de résultat (suite à modification des rémunérations)

Dans le cas où les rémunérations viennent de la source emploi, le compte de résultat est corrigé pour prendre en compte cette modification.

Le principe de correction est le même que celui évoqué au § *8.3.7* ci-dessus sur le redressement du compte de résultat.
Le résultat net est inchangé et les soldes sont mis en priorité sur les dotations aux provisions et si ce n'est pas possible, c'est le poste « reprise sur amortissements et provisions » qui est corrigé.

8.5 Les liens entre Rédi et les sources initiales : le cas des données d'enquêtes et d'emploi

Sachant qu'au moment où les gestionnaires commencent les traitements Rédi (novembre N+1 pour la campagne sur une année N), les traitements de l'enquête d'une part et des liasses fiscales de l'autre ne sont pas terminés, la question se pose de remonter les éventuelles modifications apportées par Rédi dans les sources, afin d'en tenir compte pour traiter le reste des variables. Mais elle se pose du coup différemment pour l'enquête et la source fiscale car dans le traitement de l'enquête, les gestionnaires peuvent être amenés à recontacter l'entreprise, ce qui peut faire revenir sur l'arbitrage déjà fait dans Rédi.

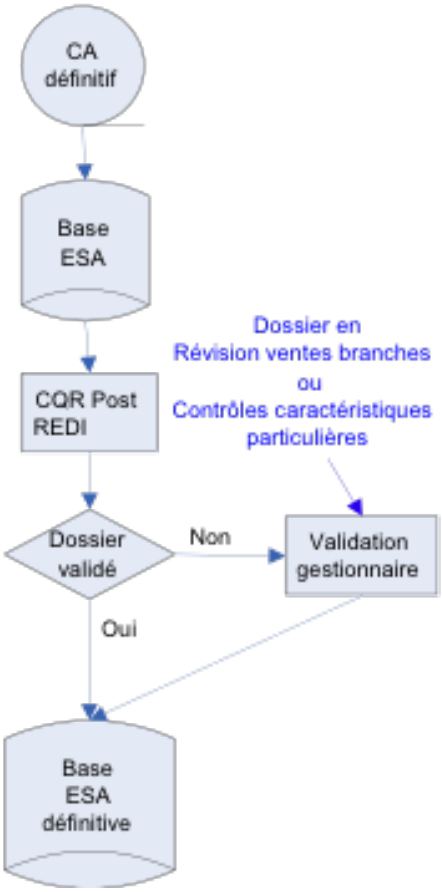
Pour l'Esa et l'emploi, la logique adoptée pour résoudre ce problème est la suivante :

- Les données qui sont prises en compte en entrée dans Rédi sont les données **redressées**[323] des sous-systèmes de départ (ESA ou emploi) ;
- Les gestionnaires **peuvent corriger dans Rédi** (i.e. dans les données en sortie de Rédi) **les caractéristiques communes aux différents sous-systèmes** et quelques **variables liées** (le chiffre d'affaires, sa ventilation en activités et produits, la sous-traitance pour l'Esa et l'effectif en nombre de postes, l'effectif en ETP et les rémunérations pour les sources d'emploi) ;
- Un CQR[324] post-Rédi (**mis en œuvre automatiquement par le système** dès l'instant qu'on passe dans le module Rédi pour une entreprise donnée) prend ensuite en compte les valeurs Rédi de ces caractéristiques dans la source de départ (Esa ou emploi) pour **redresser et ajuster automatiquement les autres caractéristiques de la source**, liées aux caractéristiques « redifiées ». Cette prise en compte de la valeur de Rédi peut avoir pour conséquence de faire « ressortir » un dossier dans la source initiale pour une caractéristique

[322] Dans les calculs ultérieurs d'agrégats, ces valeurs manquantes sont traitées comme des « zéros » car on ne sait pas faire autrement. Pour les ratios « agrégés » (calculés comme ratios d'agrégats), il en est de même. Ces cas sont peu nombreux mais peuvent être à l'origine d'une légère sous-estimation de l'emploi. Pour les ratios « individuels » en revanche (du type VA / tête pour une unité légale), si le dénominateur est manquant, le ratio sera lui aussi manquant. Dans les fichiers individuels de sortie (type FARE), ces données manquantes sont codées comme tel (c'est-à-dire avec un « . » en SAS par exemple).
[323] Rappel : pour la signification précise de ces termes, Cf. le § *2.1*. Les valeurs les plus prioritaires sont classées de haut en bas.
[324] Voir le *chapitre 2* pour une présentation complète des CQR (programmes de contrôle, qualification redressement).

qui a été modifiée, car elle était liée à une caractéristique « rédifiée[325] ». Dans ce cas le gestionnaire peut toujours corriger cette valeur le cas échéant. En revanche, **une caractéristique « rédifiée » ne peut plus être modifiée par un gestionnaire dans la source de départ**[326]. Si une correction doit être faite, elle doit avoir lieu dans Rédi qui la rétrocédera ensuite à la source de départ et un CQR sera à nouveau lancé sur la source de départ avec cette nouvelle valeur, etc.

La cinématique de ce traitement est décrite dans le schéma ci-dessous et l'exemple 12 qui suit :



L'intérêt de modifier aussi la valeur dans l'enquête est que c'est désormais cette valeur arbitrée qui servira pour contrôler ou redresser les variables de l'enquête qui ne l'ont pas encore été au moment du passage de Rédi.

[325] Par exemple une surface de magasin par rapport au chiffre d'affaires.
[326] Plus précisément, elle peut toujours l'être mais le résultat ne sera plus pris en compte ultérieurement dans les agrégats qui seront calculés. C'est donc une mise à jour qui ne servirait à rien.

Exemple 12: traitement Rédi de données ESA.

Les données avant Rédi se présentaient de la façon suivante dans l'Esa (cf. § 2.1 pour les définitions des données brutes modifiées redressées et les règles d'alimentation de ces différentes occurrences). :

Type de donnée	Chiffre d'affaires Esa
Esa Brute	100
Esa redressée	110
Esa Modifiée	110

À l'issue de Rédi, c'est le chiffre d'affaires de la liasse fiscale qui a été retenu, à savoir (supposons) 130 K€.

Type de donnée	Chiffre d'affaires Esa
Esa Brute	100
Esa Redressée	110
Esa Modifiée	110
Réconciliée	130

En réalité, le passage de Rédi est indissoluble d'un CQR post Rédi qui va remettre la valeur 130[327] comme valeur redressée[328]. Les choses se présentent alors ainsi dans la base :

Type de donnée	Chiffre d'affaires Esa
Esa Brute	100
Esa Redressée	**130**
Esa Modifiée	110
Réconciliée	**130**

Au cours du rappel téléphonique de l'entreprise pour une variable sectorielle, imaginons que le gestionnaire apprenne que le « bon » chiffre d'affaires de l'entreprise est 120.
Si le gestionnaire corrige le chiffre d'affaires dans l'Esa, la situation deviendra la suivante :

Type de donnée	Chiffre d'affaires Esa
Esa Brute	100
Esa Redressée	**130**
Esa Modifiée	**120**
Réconciliée	**130**

La valeur 120 ne serait pas prise en compte, car **la donnée Rédi est par définition la donnée la plus prioritaire**. Pour obtenir un chiffre d'affaires de 120 (une fois qu'on est passé dans Rédi), **la correction devra intervenir sur la valeur de Rédi**.
Donc, dans ce cas, le gestionnaire devra corriger le chiffre d'affaires dans Rédi et il n'aura rien à faire dans les données de l'Esa car le CQR post Rédi va remettre lui-même la valeur 120 comme valeur redressée de l'enquête.
Une fois cette correction effectuée, on se retrouvera avec les données suivantes :

Type de donnée	Chiffre d'affaires Esa
Brute	100
Esa Redressée	**120**
Esa Modifiée	110
Réconciliée	**120**

[327] Dans les faits, dans le CQR post Redi, la donnée en entrée n'est pas la valeur de l'occurrence modifiée comme dans le CQR « normal » mais la valeur de l'occurrence réconciliée qui est de plus considérée comme validée et ne subira donc aucun redressement automatique. C'est pourquoi au final la valeur de l'occurrence redressée sera égale à la valeur réconciliée.
[328] Rappelons que par définition le résultat d'un CQR (procédure automatique) alimente la valeur « Redressée ».

8.6 La nécessité d'un retraitement final (dit « Rédi balai »)

Comme on l'a vu en début de chapitre, ne sont susceptibles d'être *réconciliées* que les unités ayant **un chiffre d'affaires validé dans la liasse *et* l'enquête** et des données d'emploi validées après une éventuelle intervention de Clap. On distingue en réalité les deux procédures sur le CA et l'emploi.

A. Rédi-balai sur le CA

En fin de campagne Ésane, il reste trois types de dossiers pour lesquelles la réconciliation n'a pu se faire :

1. les dossiers pris en compte par les programmes de réconciliation **mais devant être validés par un gestionnaire** (ils ne l'ont pas été faute de temps ; mais le macro contrôle indiquait par exemple un QEG « important ») ; on dispose cependant d'une Esa et d'une liasse ;
2. les dossiers **non validés dans les 2 sources** sur les variables à réconcilier ; cette catégorie contient des unités (de l'échantillon) pour lesquelles on n'a au mieux qu'une réponse dans une source et **des données extrapolées dans l'autre** ; comme les données ne sont pas validées dans la liasse et l'enquête **le dossier n'a pas pu passer dans Rédi ;**
3. **les dossiers non présents dans l'enquête** (Esa ou EAP), qui sont naturellement les plus nombreux puisque le volume de l'enquête est bien inférieur à celui des liasses.

Les principes de **retraitement final** (plus connu en interne sous le nom de **Rédi balai**[329] car on applique des principes grandement repris du Rédi normal lorsque la liasse est prioritaire) sont largement inspirés de ceux décrits ci-dessus. La plus grosse différence vient du fait qu'il n'y a pas l'étape des-macro contrôles par les gestionnaires et que **la source jugée prioritaire est retenue sans autre forme de procès**. Mais à la différence du vrai Rédi, on n'y supprime pas de branches qui figureraient dans la ventilation initiale.

Le Rédi balai concerne donc **tout le reste du champ Ésane** (non encore pris en compte ou validé par Rédi). Ce programme est très lourd car il traite **près de 3 millions d'unités hors noyau** (cf. § *6.3* .pour la définition exacte du noyau).

- **Quand il n'y a pas d'enquête pour ces unités,** ce qui est systématiquement le cas dans le cas 3 évidemment, (ou qu'elle a été redressée par imputation dans le cas 2 ci-dessus), **c'est la liasse fiscale qui est la source prioritaire pour le niveau et la ventilation du chiffre d'affaires**.
  - Ainsi des **branches d'activités « fictives » sont créées**[330] pour chaque unité en fonction de l'APE de l'unité et de la répartition du chiffre d'affaires sur la liasse fiscale en ventes de biens, marchandises et services, comme c'est expliqué au § *8.3.3.2*. Par exemple on rajoutera une branche commerciale à une UL d'APE productrice de biens qui déclare des ventes de marchandises. (Voir exemples 6 et 7) ;
  - En revanche, quelle que soit l'APE (spécifique ou non spécifique), **on ne supprime aucune des branches** de la ventilation de la liasse, contrairement à ce que l'on faisait dans les exemples 5 et 6.
- **Quand, en revanche, on a une réponse à l'Esa** (ce qui se produit dans le cas 1 et peut se produire dans le 2 si c'est la liasse qui est imputée), on applique les principes des § *8.3.3.1* pour le niveau du CA et du § *8.3.3.2* pour sa ventilation .

B. Rédi-balai sur l'emploi

Rappelons qu'il n'y a pas de traitement manuel sur l'emploi exercé par les gestionnaires. Il y a juste une validation finale exercée par la division Ésane (Cf. chapitre 10). En l'absence de données d'emploi pour une unité jugée statistiquement active, ce sont les données d'emploi de la liasse fiscale qui sont retenues. **C'est en cela que l'on dit que la réconciliation joue le rôle de traitement de la non-réponse totale pour l'emploi**.

[329] En réalité ce terme est relativement **impropre** puisque, pour toutes les entreprises dont on n'a pas de réponse à l'enquête (cas 3), on ne réalise pas à proprement parler de « réconciliation » entre deux sources, puisqu'on n'a pas l'enquête. En revanche, on fait bien un traitement dans lequel on crée des branches d'activités fictives pour correspondre à la ventilation de la liasse.

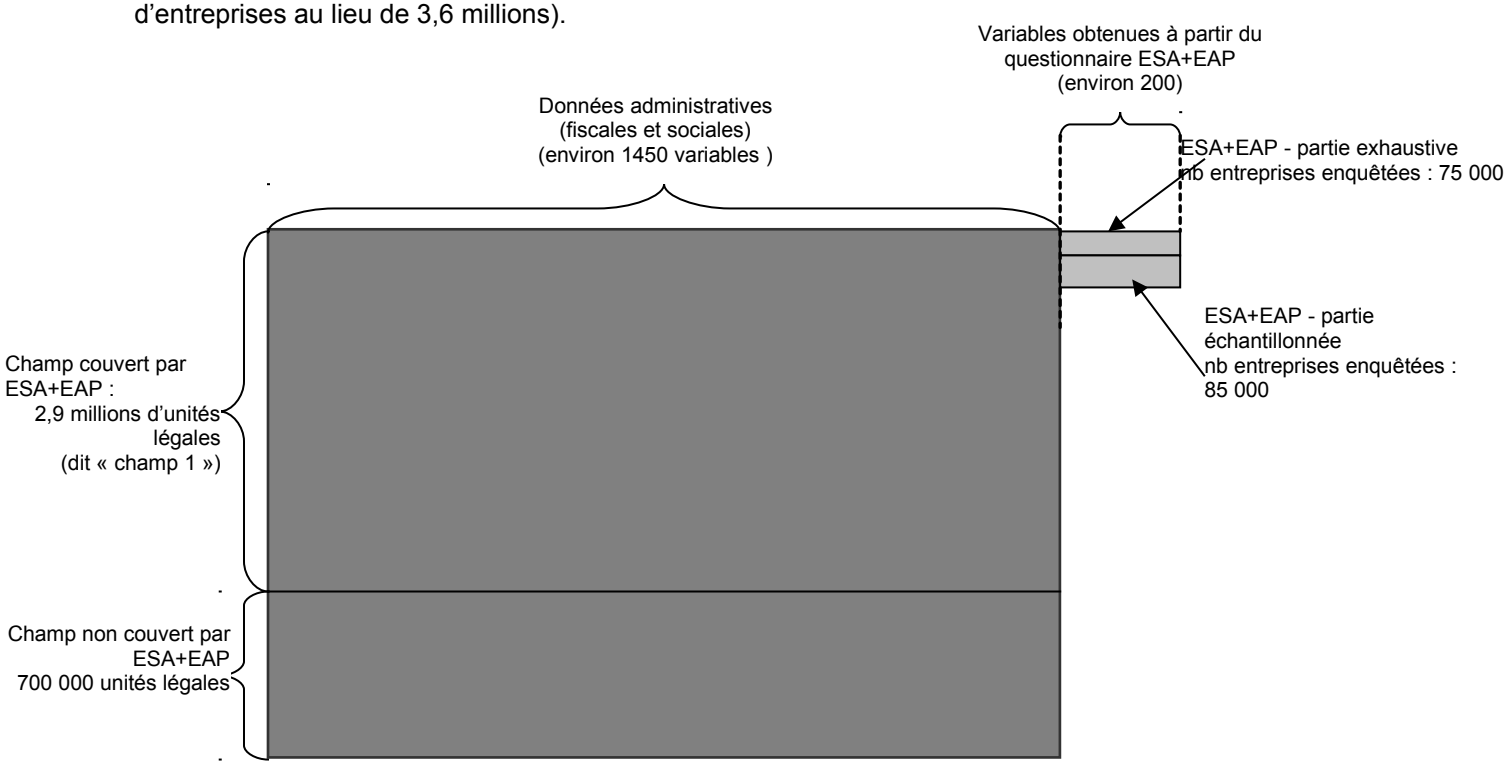
[330] Cela permet d'avoir en interne des fichiers de données individuelles homogènes mais ces données ne sont pas utilisées dans la confection de la matrice secteur branche comme on le verra au § *9.2.5* du moins sur le sous-champ 1 : seules sont utilisées les branches des unités répondantes à l'enquête. En revanche elles le sont sur les champs 2 à 5 sur les DOM et l'EAP0 (jusqu'en 2013)

9. Le calcul et la diffusion des agrégats finaux

9.1 *La problématique générale de l'estimation des résultats* [1.4], [2.1], [2.2], [2.3], [2.4], [2.11], [2.16]

La production d'estimations statistiques à partir d'un dispositif multi-sources n'est pas immédiate.

D'une part, on dispose de données plus ou moins riches selon que l'entreprise appartient à l'échantillon de l'Esa ou de l'EAP ou pas (auquel cas on ne dispose que de ses données administratives) comme le montre le schéma ci-dessous (chiffres représentatifs des années 2008 ou 2009 ; à partir de 2010, le nombre d'auto-entrepreneurs a fait nettement augmenter le nombre d'unités légales du champ non couvert par ESA+EAP ; en 2015, le champ Ésane couvre environ 4,7 millions d'entreprises au lieu de 3,6 millions).



Note de lecture : ce schéma représente à une échelle **reflétant à peu près le volume des données**, la manière dont les différentes sources sont utilisées pour produire les statistiques dans le dispositif Ésane. Une entreprise (unité légale) est représentée par une ligne de tableau ; une variable par une colonne. Ainsi, les variables obtenues à partir du questionnaire Esa+EAP concernent les 2,9 millions d'entreprises du champ total des deux enquêtes et leur calcul se base sur les réponses de l'ensemble des 75 000 entreprises enquêtées chaque année (généralement ce sont celles de 20 ou 30 salariés ou plus Cf. § *5.1.3.2*) et de la partie échantillonnée des entreprises (celles en dessous du seuil d'exhaustivité), soit environ 85 000 entreprises environ. C'est le champ non couvert par les enquêtes qui a beaucoup augmenté en nombre d'unités, ceci étant dû essentiellement aux auto-entrepreneurs.

D'autre part, comme il a été indiqué précédemment, pour certaines variables, **on dispose de valeurs dans les deux sources** : les cas du chiffre d'affaires et de sa ventilation, ou de l'emploi, qui nécessitent un arbitrage, objet du processus Rédi, ont été présentés dans le *chapitre 8*, mais la situation est du même type pour **« le classement sectoriel »** (c'est-à-dire pour la variable « APE »). En effet, pour celui-ci on dispose du code APE du répertoire et de sa valeur réévaluée à l'occasion de l'enquête ; la similitude avec les variables d'accrochage vues au chapitre 8 s'arrête toutefois là, car la valeur du code APE dans le répertoire est a priori ancienne quand celle mise à jour par l'enquête est par définition la plus récente et donc la bonne. C'est donc cette dernière qui doit être utilisée pour les entreprises interrogées dans l'enquête. Mais pour les 3,5 millions (environ) qui ne le sont pas, on ne

dispose que de la valeur du code APE qui figure dans le répertoire ; elle remonte au mieux à la dernière fois que l'entreprise a signalé au répertoire un changement d'activité ou qu'elle a été interrogée par l'Esa ou l'EAP, au pire à la date de création de l'entreprise. On a donc finalement a priori **trois façons**[331] de calculer des agrégats :

1. *On s'appuie exclusivement sur les liasses fiscales* (source exhaustive) dans lesquelles on a simplement injecté les valeurs issues de Rédi des variables pour les entreprises interrogées en enquête. C'est ce que l'on publiait auparavant quand on publiait les « résultats » de « Suse »[332]. Mais **on pourrait améliorer ces résultats** si on prenait en compte les changements sectoriels observés sur l'échantillon Esa+EAP pour les « extrapoler » à l'ensemble des entreprises.

   Prenons un exemple : imaginons que l'on observe dans l'enquête de l'année N que des débits de boisson (NAF 5630Z) ont changé d'activité et sont devenus des cybercafés (NAF 6190Z). On cherche à estimer le total Z du chiffre d'affaires des cybercafés. La source fiscale « pure » peut nous indiquer que le chiffre d'affaires total du secteur des cybercafés, connus en N-1, vaut X l'année N. Mais il faut logiquement y rajouter le chiffre d'affaires des débits de boisson en N-1, qui ont changé de secteur et sont devenus cybercafés l'année N. On ne sait pas lesquels c'est, sauf pour ceux d'entre eux qui appartiennent à l'échantillon de l'enquête. Mais connaissant le chiffre d'affaires « représenté » par ces débits de boisson devenus cybercafés, on va pouvoir corriger ce chiffre X en lui rajoutant une estimation du chiffre d'affaires dû à ce changement de secteur. L'information apportée par l'enquête est la différence (de la forme $\hat{Y} - \hat{X}$) entre le chiffre d'affaires (pondéré grâce aux poids de sondage) de l'échantillon à l'issue du processus Ésane de toutes les entreprises que l'enquête a classées en cybercafés avec le chiffre d'affaires de celles qui étaient déjà classées en cybercafés avant l'enquête.

   Au final on obtiendrait donc une formule du type[333] :

   $\bar{Z} = X + (\bar{Y} - \bar{X})$ (a) [334]

2. *On s'appuie exclusivement sur l'enquête* : on dispose a priori d'un échantillon représentatif de la population des entreprises exerçant une activité sur le sol français. On sait estimer ces résultats grâce aux poids de sondage de l'enquête redressés (correction de la non-réponse, calage) et à l'estimateur habituel $\hat{Y}$ calé (le même $\hat{Y}$ qu'au paragraphe précédent). C'est ce que l'on faisait auparavant lorsqu'on publiait les « résultats » des « EAE ». Mais **on pourrait améliorer ces résultats** si l'on prenait en compte de façon additionnelle les données fiscales exhaustives. Pour cela, on peut ajouter à l'estimateur $\hat{Y}$ du chiffre d'affaires d'un secteur, un estimateur d'espérance nulle, de la forme $(X - \hat{X})$, écart entre la somme simple des chiffres d'affaires des entreprises classées dans le secteur avant l'enquête et la somme pondérée des chiffres d'affaires des entreprises de l'échantillon qui les représentent (qui représente en quelque sorte l'information apportée par la source exhaustive à l'enquête). On peut démontrer que, sous certaines conditions qui sont ici fréquemment vérifiées, cet estimateur améliore la précision de l'estimateur calé initial. Au final on obtiendrait une formule de la forme :

   $\bar{Z} = \bar{Y} + (X - \bar{X})$ (b)

3. O*n cherche à exploiter simultanément les deux sources* :
   - l'une partielle en terme d'information (pas de ventilation en branches, pas de classement sectoriel à jour) mais disponible sur l'ensemble de la population ;
   - l'autre apportant des informations plus fraîches (en terme de classement sectoriel), mais sur un échantillon uniquement.

Il se trouve comme on vient de le voir que les améliorations envisagées dans les deux alinéas précédents conduisent **aux mêmes estimateurs dits « par différence »**, notés ici $\tilde{Z}$ que l'on peut donc interpréter d'une façon ou de l'autre, sous la forme (a) ou (b).

[331] Il y en a en réalité bien davantage, car on aurait pu penser également à imputer toutes les variables du questionnaire pour toutes les entreprises qui ne sont pas interrogées dans l'enquête et se ramener ainsi à des données complètes sous forme d'une matrice rectangulaire. P. Brion a montré dans [2.1] qu'un telle imputation de masse conduisait à des estimateurs biaisés et dont l'erreur quadratique moyenne est nettement supérieure en moyenne à celle des estimateurs retenus.

[332] Voir le schéma figurant au début de *l'Avant propos*.

[333] X, $\hat{X}$, $\hat{Y}$ et $\tilde{Z}$ sont tous des estimateurs de la quantité inconnue, Z, chiffre d'affaires des cybercafés de l'année N. X et $\tilde{Z}$ désignent des quantités calculées sur la source administrative alors que les lettres avec des chapeaux font référence à des sommes pondérées.

[334] La formule exacte, que l'on verra au paragraphe suivant, doit évidemment tenir compte non seulement des entreprises qui sont devenues « cybercafés » (en provenance de l'ensemble des secteurs possibles), mais aussi des cybercafés en N-1 qui sont, eux, sortis du secteur en N ainsi que de la « déformation » du chiffre d'affaires apportée par Redi lorsque la valeur finale (donc Redi) retenue par Ésane est différente de celle de la source fiscale seule.

Il s'agit là ***d'une des innovations les plus importantes d'Ésane et les plus impactantes sur les résultats publiés*.***

Le reste du chapitre adoptera alors le plan suivant :

- Le principe et les formules de calcul des agrégats seront présentés en détail dans le § *9.2* ci-dessous. On verra en particulier que les estimateurs retenus sont linéaires, sans biais, et maximisent la précision mesurée par les coefficients de variation.
- Mais malgré ses nombreux avantages, l'estimateur « par différence » qu'on vient de voir a quelques limites : dans certains cas, il ne permet pas de conserver les inégalités et en particulier il peut donner des estimations négatives pour des variables positives à des niveaux fins de la nomenclature, ce qui a obligé à complexifier l'estimateur final retenu, pour les niveaux fins de la nomenclature ou certains agrégats spécifiques.Le § *9.3* décrit ainsi les limites des estimateurs obtenus et les précautions à prendre dans leur interprétation.
- Le système de production Ésane donne lieu à différentes publications d'agrégats, provisoires puis définitives, sur le site Insee.fr (ainsi que sur celui d'Eurostat), qui seront décrites au § *9.4*. Ce paragraphe décrit en particulier les quatorze tableaux que l'utilisateur peut trouver sur le site de l'Insee.
- Dans ces publications, un certain nombre de données sont « sous secret statistique ». Le § *9.5* explicite les algorithmes qui sont utilisés pour déterminer les données qui doivent être masquées, compte tenu des estimateurs présentés au § 9.2.
- Les utilisateurs s'intéressent fréquemment à l'évolution entre N-1 et N des grandeurs que mesure le système Ésane même si celui-ci n'a pas été élaboré prioritairement dans cet objectif. Le § *9.6* présente la décomposition que l'on peut faire d'une population d'entreprises entre les pérennes, les entrantes et les sortantes, au regard d'un secteur donné et l'interprétation qu'elle permet de faire des évolutions sectorielles. Cette partie présente une **méthodologie non encore mise en œuvre de façon pratique à l'Insee**. Elle a donc principalement un rôle pédagogique afin d'illustrer la difficulté du calcul et de l'interprétation d'évolutions économiques à la fois pertinentes et fiables.

9.2 Les principes méthodologiques retenus pour le calcul des agrégats diffusés [2.12], [2.16], [3.10], [3.11]

Dans les formules ci-après, on utilise les notations suivantes :

- Y_i^{fiscal} est la valeur de la caractéristique Y dans la source fiscale ;
- Y_i^{Redi} est la valeur de la caractéristique Y à l'issue de la phase de réconciliation (Rédi) ;
- U_1 (comme univers du sous-champ 1) désigne l'ensemble des entreprises du sous-champ 1[335], hors DOM et hors EAP0[336] ;
- HC (comme hors champ) désigne les unités légales pour lesquelles on apprend une sortie de champ (cessation, changement d'APE ou d'autres caractéristiques entraînant une **sortie du champ Ésane**). Si cette sortie de champ ne concerne que les unités échantillonnées et que l'information a été obtenue par le biais de l'enquête on parle de **HC source Esa**[337]. En revanche si cette sortie de champ a été obtenue par une information externe indépendante de l'échantillon et accessible pour toutes les unités de l'univers, on parle de **HC externe**.
- $U_1 \setminus \text{HC externes}$ signifie U1 *privé*[338] des unités « hors champ » repérées par des sources externes (telles que SIRENE ou Sirus). À l'inverse, les unités de l'échantillon repérées hors champ par l'ESA ou l'EAP sont prises en compte dans cette population, pour éviter tout biais ;
- R (comme répondant) désigne l'ensemble des unités qui appartiennent à l'échantillon (y compris sa partie exhaustive), **répondantes aux enquêtes**, ainsi que **les non-répondantes de la partie exhaustive** qui sont **extrapolées individuellement** (cf. infra) appartenant au champ (tout se passe comme si elles avaient répondu) ;
- W_i désigne **le poids** de l'unité i de l'échantillon après traitement de la non-réponse, winsorisation et calage (cf. § *5.6.2*) (W_i vaut 1 pour les entreprises de l'exhaustif) ;

[335] Voir § *1.2.1.1* ou *4.3*. Il s'agit bien de l'ensemble des entreprises qui sont représentées par l'échantillon, pas de l'échantillon lui-même.

[336] Ainsi qu'on l'a précisé au § *5.2.2*, le principe de l'EAP0 a abandonné à compter de la campagne 2014 et le mode de tirage de l'échantillon EAP aligné sur celui de l'ESA. Dans toutes les formules qui suivent, l'EAP0 joue un rôle spécifique qui disparaitra à ce moment là. Les formules restent valables sans la restriction spécifique au rôle de l'EAP0.

[337] Il faut comprendre ici source Esa **ou EAP**, mais on n'a pas précisé pour ne pas alourdir l'écriture.

[338] C'est le sens du symbole antislash « \ ».

- Les indicatrices « rep » (pour répertoire) ou « enq» (pour enquête) sont relatives au classement sectoriel : « rep » est l'APE dans le répertoire lors du tirage de l'unité dans la base de sondage (cf. définition de l'APE_REP au § *5.4.3.1*), et « enq » le classement connu à l'issue de l'enquête et correspondant donc au classement de l'unité lors de la diffusion (cf. définition de l'APE_DIF dans ce même § *5.4.3.1*) ;
- On désigne enfin par « ech » l'échantillon des enquêtes y compris donc la partie exhaustive.

Suivant ce qu'on vient de voir au § *9.1*, pour un groupe G de la NAF (niveau à 3 caractères)[339], l'estimateur de toute statistique sectorielle ayant trait à une variable quelconque Y_i^{fiscal} de la source administrative, sera calculé de la façon suivante :

- Pour les estimations relatives au sous champ 1 (SC1) hors DOM et hors EAP0, **calcul de l'estimateur par différence** (ou estimateur **composite**[340]) suivant, de la forme X + ($\hat{Y}$ - $\hat{X}$) :

[Formule (1)] :

$$\hat{Y}^{SC1}_{compo}(G) = \sum_{i \in U_1 \setminus HC\ externes} Y_i^{fiscal} \mathbb{I}_{groupe_rep=G}(i) + \sum_{i \in R} w_i Y_i^{Rédi} \mathbb{I}_{groupe_diff=G}(i) - \sum_{i \in R \oplus HC\ source\ Esa} w_i Y_i^{fiscal} \mathbb{I}_{groupe_rep=G}(i)$$

On a pris ici l'approche par la source fiscale (cf. point 1 du paragraphe précédent) dans laquelle X est la somme simple de la variable sur les entreprises qui appartenaient au groupe G au lancement, $\hat{Y}$ est l'estimateur calé calculé sur les entreprises qui appartiennent au groupe G, suite à l'enquête, et $\hat{X}$ est la somme pondérée de la variable calculée sur les valeurs de la source fiscale connues avant l'étape de réconciliation Rédi.

- Pour les agrégats relatifs à l'EAP0[341], aux DOM ainsi qu'aux sous champs Ésane 2 à 5[342], **estimation par sommes simples non pondérées**. En effet pour ces populations, il n'y a pas lieu d'effectuer une quelconque inférence à partir des résultats de l'enquête, puisqu'aucune de ces unités n'appartient à l'échantillon de l'enquête.
- Calcul de l'estimation finale portant sur l'ensemble du champ Ésane par **sommation des deux éléments précédents**.

L'intérêt de l'approche privilégiant la source administrative est de **pouvoir la décomposer facilement en trois effets**, dus aux entrées de champ, aux sorties de champ et à la déformation apportée par Rédi, selon la formulation suivante (équivalente à la précédente) :

[339] Et par additivité bien entendu tous les niveaux plus agrégés (divisions, sections ou ensemble de l'économie).
[340] On a préféré l'indication « composite » car Diff pour « différence » prête à confusion avec « diffusion ».
[341] Voir § *5.2.1* et *5.2.2.*
[342] Voir § *1.2.1.1* ou *4.3*

[Formule (2)] :

$$+\sum_{i\in R} w_i Y_i^{Redi} * \left[\mathbb{I}_{groupe_diff=G}(i) - \mathbb{I}_{groupe\text{-}reg=G}(i)\right]$$ ← Effet changements d'APE

$$+\sum_{i\in R} w_i \mathbb{I}_{groupe_reg=G}(i) * \left[Y_i^{Redi} - Y_i^{fiscal}\right]$$ ← Effet Redi

$$-\sum_{i\in HC\ source\ Esa} w_i Y_i^{fiscal} \mathbb{I}_{groupe_rep=G}(i)$$ ← Effet sorties de champ

- **Toutes les entreprises sont dans le premier terme**. Si une entreprise n'est pas dans l'échantillon, elle apparaît uniquement dans le premier terme (elle compte pour 1). On part du total mesuré sur la source fiscale auquel on ajoute un certain nombre d'effets.
- Si une entreprise appartient à l'échantillon et ne change pas d'APE, et si Rédi n'a pas modifié le chiffre d'affaires, elle apparaît dans les 3 premiers termes ci-dessus, mais les termes « effet de changement d'APE » et « effet Rédi » s'annulent pour cette entreprise. Elle compte également pour 1.
- Si une entreprise est sortie du groupe G tout en restant dans le sous-champ 1, elle comptera pour $-w_i$ dans le second terme, et pour 1 dans le premier. Elle compte donc pour $1 - w_i$ (multiplié par la variable d'intérêt Y).
- Inversement, une entreprise qui n'était pas dans le groupe G dans le répertoire mais qui appartient au groupe G à l'issue de l'enquête comptera pour 0 dans le premier terme et pour w_i dans le second.
- Si une entreprise n'a pas changé de secteur, mais que Rédi a modifié son chiffre d'affaires, elle comptera pour 1 dans le premier terme, pour 0 dans le second, et pour $w_i * \left(Y_i^{Redi} - Y_i^{fiscal}\right)$ dans le troisième.
- Le dernier terme contient enfin les entreprises qui étaient dans le champ de l'Esa et dans le groupe G au lancement mais que la réponse à l'enquête sort du champ de l'enquête.

L'estimateur de la formule (1) ou (2) présente de nombreux avantages : il s'agit en effet d'un estimateur **sans biais** et **linéaire**[343] – propriété particulièrement intéressante dans le contexte du dispositif Ésane, dont les variables présentent la particularité d'être très fréquemment reliées entre elles par des équations comptables, qu'il convient de respecter lors des estimations – et globalement **plus efficace** (c'est-à-dire de variance inférieure) – du fait de l'utilisation d'un estimateur par différence sur la partie échantillonnée – que les estimateurs usuels de type estimateur calé (Cf. § 9.2.4).

Il ne garantit cependant pas la positivité des estimations, à cause du terme « -$\hat{X}$ », et peut donc conduire dans certains cas à des estimations négatives, alors même que toutes les données individuelles sont positives ou nulles. Ceci se révèle problématique lorsque cela concerne des variables pour lesquelles des agrégats négatifs n'ont aucun sens économique, comme les ventes de marchandises par exemple[344].

En pratique, ce type d'estimations problématiques relève majoritairement de deux cas de figures bien distincts, que met en évidence la formule (2) :

[343] A la fois par rapport aux variables Y mais aussi par rapport aux secteurs (l'indicatrice de l'union de deux secteurs est la somme des indicatrices de ces secteurs).

[344] Symétriquement, on peut observer un problème similaire dans le cas d'estimations positives de quantités censément négatives, telle la variable « Perte comptable de l'exercice ».

- d'une part l'estimation de variables fortement affectées par le processus Rédi, et pour lesquelles la valeur finale Y_i^{Redi} diffère fortement de la valeur de la source fiscale Y_i^{fiscal}. Ceci concerne les variables de ventilation agrégée du chiffre d'affaires « ventes de marchandises », « production de biens » et « production de services », ainsi que les variables d'achats et de variations de stocks correspondantes. En effet, pour ces variables, l'enquête constitue la source de référence du processus Rédi, et il est donc très fréquent que la ventilation agrégée du chiffre d'affaires post-Rédi, ainsi que les achats correspondants, s'éloignent sensiblement des déclarations fiscales ;
- d'autre part, l'estimation de statistiques portant sur des petits domaines où l'échantillon n'est pas représentatif des entrées et des sorties du domaine. Ce cas de figure se rencontre soit lorsqu'on s'intéresse à l'estimation de quantités à un niveau fin – classement sectoriel au niveau sous-classe, estimation par [groupe ⊗ tranche de taille] ou encore par [groupe ⊗ région], etc. –, soit lorsqu'on travaille sur des variables comptables à occurrences rares, du type « Autres charges et dépenses somptuaires ». Dans ces conditions, les estimations par différence sont peu robustes[345] – taille de la population U d'intérêt de l'ordre de quelques centaines, voire quelques dizaines d'unités seulement concernées dans l'échantillon – et il suffit qu'une unité de l'échantillon avec un poids élevé change de secteur pour que l'on obtienne parfois une estimation négative. C'est la raison pour laquelle nous avons **limité** l'usage de la formule (1) aux **groupes de la NAF**.

De fait, tant qu'on raisonne à un niveau suffisamment agrégé – en pratique, niveau groupe (trois premiers caractères de la NAF) ou supérieur –, **ces estimations problématiques s'avèrent rares**[346], et une procédure de corrections individuelles (par mise à zéro ponctuelle des agrégats incriminés par exemple) peut être envisagée. En revanche, dès qu'on s'intéresse à des agrégats à un niveau plus fin, le nombre d'estimations négatives – relevant cette fois-ci majoritairement du second cas de figure – s'accroît sensiblement et un traitement au cas par cas garantissant la cohérence des estimations entre les différents niveaux n'est plus envisageable. Et donc finalement :

- pour les agrégats de niveau groupe, les estimations problématiques au niveau des agrégats seront gérées selon des procédures qui seront détaillées au § *9.2.1* ;
- les estimations de niveaux supérieurs au groupe se déduiront immédiatement des estimations de niveau groupe par simple agrégation ;
- enfin, pour les estimations de niveau infra groupe, la méthode retenue, détaillée au § *9.2.2*, propose des estimateurs s'appuyant essentiellement sur les données fiscales, permettant d'assurer le signe des estimations obtenues et de garantir la cohérence avec les estimateurs composites de niveau groupe.

Cette mécanique globale d'estimation assure par construction une **cohérence parfaite des estimations entre les différents niveaux de nomenclature**, et garantit un **signe valide** aux agrégats obtenus. Elle **implique** en revanche le recours à des **méthodes d'estimations non linéaires**, que ce soit pour la gestion des estimations problématiques au niveau groupe ou pour le calcul des agrégats de niveau infra groupe. Or comme évoqué précédemment, les variables du système Ésane présentent la particularité d'être très fréquemment reliées entre elles par des équations comptables, qu'il convient de respecter lors des estimations. Aussi, dès lors que l'on envisage d'utiliser des méthodes d'estimations non linéaires, il n'est plus possible de traiter les variables indépendamment les unes des autres lors des estimations.

Par conséquent**, on procède à des estimations « directes »[347] uniquement pour les variables « élémentaires »[348], et on en déduit les estimations pour les variables non élémentaires via les égalités comptables.**

[345] Tout comme les estimateurs calés d'ailleurs, mais sur ces derniers ce problème de robustesse ne se manifeste pas de manière aussi flagrante, liée au signe de l'estimateur.
[346] L'ensemble des estimations problématiques niveau groupe – qui ne concerne que les variables de ventilation agrégée du chiffre d'affaires et les variables d'achats et de variations de stocks liées (1er cas de figure), ainsi que quelques variables à occurrences rares (2nd cas de figure) – représente moins de 0,1 % du nombre total d'estimations niveau groupe.
[347] I.e. par calcul d'un estimateur, via la procédure décrite précédemment, à partir des données individuelles.
[348] I.e. des variables n'intervenant qu'en tant que composantes et jamais comme solde dans les équations comptables.

9.2.1 Les estimations de niveau groupe & supra des agrégats fiscaux et d'emploi

L'estimation de toute statistique sectorielle s'appuyant sur une variable administrative Y donnée **au niveau groupe ou supra** repose sur la procédure suivante (en sept étapes) :

❶ Pour les estimations relatives au sous champ 1 (SC1) hors DOM et hors EAP0, **calcul de l'estimateur par différence** suivant pour les variables élémentaires :

(Formule 1 précédente) :

$$\hat{Y}^{SC1}_{compo}(G) = \sum_{i \in U_1 \setminus HC\ externes} Y_i^{fiscal} \mathbb{I}_{groupe_rep=G}(i) + \sum_{i \in R} w_i Y_i^{Rédi} \mathbb{I}_{groupe_diff=G}(i) - \sum_{i \in R \oplus HC\ source\ Esa} w_i Y_i^{fiscal} \mathbb{I}_{groupe_rep=G}(i)$$

Remarque : de même que pour le calage, les unités présumées cessées, ainsi que les unités détectées comme hors-champ de façon certaine par une source externe à l'enquête, ne participent pas au calcul[349], contrairement aux unités hors-champ détectées par l'enquête[350].

❷ Pour les agrégats de variables élémentaires relatifs aux sous champs Ésane 2, 3 et 5 ainsi qu'à l'EAP0[351] et aux DOM , **estimation par sommes simples non pondérées** de la variable Y retraitée par Rédi-balai, sur les entreprises appartenant au champ Ésane. Soit avec des notations évidentes :

(Formule finale 1bis) pour l'ensemble du champ et toutes les variables élémentaires :

$$\hat{Y}_{compo}(G) = \hat{Y}^{SC1}_{compo}(G) + \hat{Y}^{Sous-Champ\ 2_5}_{Rédi}(G) + \hat{Y}^{Dom}_{Rédi}(G) + \hat{Y}^{EAP0}_{Rédi}(G)$$

où $\hat{Y}^{champ}_{Rédi}(G)$ désigne la somme simple de la variable Y après Rédi-balai sur le champ indiqué.

On voit que la composante « composite » de l'estimateur n'est due qu'au sous-champ 1.

❸ Spécifiquement pour les variables « ventes de marchandises », « production de biens » et « production de services », gestion des estimations négatives au niveau de l'agrégat relatif au sous champ Ésane 1 (y compris exhaustif) selon la **procédure suivante** :

- mise à zéro de la variable négative ;
- report du montant négatif ainsi traité sur la variable correspondant à la branche principale de l'entreprise ;
- et ajustement en conséquence des variables d'achat et de variation de stock correspondantes.

❹ Calcul des agrégats globaux (i.e. sur l'ensemble du champ) portant sur les variables élémentaires par **sommation des agrégats** relatifs à l'échantillon du sous champ 1, à l'EAP0, aux DOM et aux sous champs Ésane 2 à 5.

❺ Une fois l'ensemble des estimations portant sur des variables élémentaires réalisées, calcul de tous les agrégats résultant d'une équation comptable par l'équation elle-même.

❻ Pour les niveaux supra-groupe (divisions, sections, etc.), les estimations s'en déduisent par **agrégation des estimations niveau groupe**.

❼ Gestion des estimations négatives (respectivement positives) « à tort » par **mise en « non significatif »** des agrégats concernés.

9.2.2 Les estimations de niveau infra groupe (sur un domaine D) des agrégats fiscaux et d'emploi

Au niveau **d'un domaine D infra groupe** (sous-classe ou groupe⊗taille ou APE⊗SNF/EI), la procédure d'estimation retenue pour les statistiques sectorielles s'appuyant sur une variable administrative Y donnée (agrégats fiscaux ou emploi) est la suivante (en quatre étapes) :

[349] Il s'agit des unités notées HC externes dans la formule ci-dessus.
[350] Il s'agit des unités notées HC source ESA dans la formule ci-dessus.
[351] Jusqu'en 2013 comme on l'a vu au § 5.2.2

❶ Pour les variables élémentaires, et pour la partie de l'agrégat relative **au sous champ Ésane 1** hors DOM et hors EAP0, on va s'inspirer de la formule a) du § 9.1 ci-dessus : pour un groupe G et un domaine D ⊂ G donnés, au lieu de prendre l'estimateur par différence brut, on va se caler sur le total de Y sur le domaine D (Y_D^{Redi}), qui est un des estimateurs "naturels" possibles, (D apprécié en APE de diffusion, c'est-à-dire en tenant compte des changements d'APE apportés par l'enquête, entrées comme sorties de champ, par rapport à D) ; on va ensuite corriger cet estimateur comme si on voulait retrouver l'estimateur par différence en rajoutant ($\hat{Y}_G^{différence}$-Y_G^{Redi}) mais mesuré sur le groupe auquel appartient D et pondéré par le "poids" de D dans G relativement à la variable Y ; comme certaines variables peuvent prendre des valeurs négatives, on prend comme poids la valeur absolue du total de la variable Y dans D, rapportée à la somme des mêmes valeurs absolues sur les domaines D composant G ; cela donne la formule suivante :

[Formule (3)[352]] :

$$\hat{Y}_D^{infra} = Y_D^{Redi} + \left(\hat{Y}_G^{différence} - Y_G^{Redi}\right) \frac{\left|Y_D^{Redi}\right|}{\sum_{D\subset G}\left|Y_D^{Redi}\right|}$$

avec :

$$Y_G^{Redi} = \sum_{i\subset U_1} Y_i^{Redi} 1I_{groupe_diff=G}(i)$$

$$Y_D^{Redi} = \sum_{i\subset U_1} Y_i^{Redi} 1I_{APE_diff=D}(i)$$

Et $\hat{Y}_G^{différence}$ l'estimateur composite de niveau groupe qui sert de référence.

Il est facile de voir que si D=G, on retrouve évidemment l'estimateur par différence. On peut voir facilement aussi que si l'estimateur par différence de niveau groupe est >0 pour une variable >0, alors l'estimateur infra est également et se ramène à $\hat{Y}_G^{différence} * Y_D^{Redi}/Y_G^{Redi}$.

Remarque : il peut arriver que le dénominateur de la formule (3) ci-dessus soit nul[353] alors que l'estimateur par différence $\hat{Y}_G^{différence}$ retenu au niveau groupe est non nul. Dans une telle situation, on a une quantité à ventiler mais il est impossible de calculer une clef de répartition (puisqu'on a zéro au numérateur et au dénominateur). Ceci peut se produire :

- lorsque l'on travaille sur une variable Y à occurrence rare[354] pour laquelle Y est non nul pour une unité appartenant à G comme APE de lancement au répertoire mais plus en diffusion (elle a changé de groupe à cause de sa réponse à l'enquête) ; dans ce cas, Y_D^{Redi} et Y_G^{Redi} sont nuls mais $\hat{Y}_G^{différence}$ ne l'est pas[355].
- lorsqu'une variable est très fortement affectée par Rédi : par exemple dans certains secteurs, Rédi refait le partage biens / services ; c'est le cas de l'exemple 5 du § *8.3.3.2* : la liasse fiscale faisait apparaître des ventes de service que Rédi reclasse en ventes de biens. La variable « ventes de services » sera donc nulle dans le groupe 352 auquel appartient l'entreprise de l'exemple alors que l'estimateur par différence des ventes de services dans ce même groupe 352 ne le sera pas parce que dans les liasses fiscales des entreprises non répondantes à l'enquête, il y a des ventes de services non nulles.

Afin de pallier ce problème, on utilise dans ce cas la formule suivante :

[352] Une autre formule, qui présentait des inconvénients dans certains cas particuliers, a été utilisée lors des premières campagnes.

[353] Comme il s'agit d'une somme de valeurs absolues, cela implique que chaque terme est nul et donc que Y_G^{Redi} aussi est nul.

[354] Du type « autres charges et dépenses somptuaires » comme l'a vu plus haut par exemple.

[355] Rappelons que l'estimateur par différence peut s'interpréter comme l'ajout de l'information apportée par l'échantillon à celle contenue dans la source administrative exhaustive. Si l'occurrence est suffisamment rare pour ne pas se rencontrer dans l'échantillon, celui-ci n'apporte aucune information, mais on a une mesure de la variable dans la source exhaustive et donc un estimateur par différence non nul.

[formule (4)] :

$$\hat{Y}_D^{infra-IEG} = Y_D^{Rep-IEG} + \left(\hat{Y}_G^{différence} - Y_G^{Rep-IEG}\right) \frac{\left|Y_D^{Rep-IEG}\right|}{\sum_{D \subset G} \left|Y_D^{Rep-IEG}\right|}$$

avec :

$$Y_G^{Rep_IEG} = \sum_{i \subset U_1} Y_i^{IEG} 1I_{groupe_rep\ =G}\ (i)$$

$$Y_D^{Rep_IEG} = \sum_{i \subset U_1} Y_i^{IEG} 1I_{APE_rep\ =D}\ (i)$$

Nota : On est dans le cas où $\hat{Y}_G^{différence}$ est non nul alors que Y_G^{Redi} l'est. C'est donc que les données individuelles non nulles intervenant dans le calcul de $\hat{Y}_G^{différence}$ sont des données IEG sur des unités ayant un groupe_rep=G, donc que les $Y_D^{Rep-IEG}$ ne sont pas tous nuls, et partant le dénominateur de la formule (4) non plus. La clef de répartition fonctionne donc bien.

❷ Pour les agrégats de variables élémentaires relatifs à l'EAP0, aux DOM ainsi qu'aux sous champs Ésane 2 à 5, **estimation par sommes simples non pondérées** de ladite variable sur les entreprises appartenant au champ Ésane.

❸ Calcul des agrégats globaux (i.e. sur l'ensemble du champ) portant sur les variables élémentaires par **sommation des agrégats** relatifs à l'échantillon du sous champ 1, à l'EAP0, aux DOM et aux sous champs Ésane 2 à 5.

❹ Une fois l'ensemble des estimations portant sur des variables élémentaires réalisées, calcul de tous les agrégats résultant d'une équation comptable **par l'équation elle-même**.

9.2.3 Influence du calage sur les résultats des agrégats fiscaux

On a vu au § *5.6.2.4* que le traitement de la non-réponse des enquêtes se terminait par une étape de « calage » dont l'objet était le recalcul des poids de l'enquête. Si l'on regarde comment est effectué ce recalcul par rapport à la formule (1) écrite sous la forme (b) du § *9.1* Y= $\hat{Y}$ + (X - $\hat{X}$), il est aisé de se rendre compte que ce sont des équations qui assurent l'égalité X= $\hat{X}$ lorsque la variable fiscale considérée est le chiffre d'affaires total ou la constante « 1 », qui permet d'estimer le nombre d'entreprises répondantes.

Il en résulte, puisque la formule (1) est utilisée au niveau groupe de la nomenclature (trois caractères) qui est aussi le niveau auquel est effectué le recalage, que **les estimateurs composites (Y) sont égaux aux estimateurs calés** ($\hat{Y}$) pour l'estimation du **chiffre d'affaires** ainsi que pour l'estimation **du nombre d'entreprises**, **au niveau groupe de la nomenclature**.

9.2.4 Comparaison de la précision de différents estimateurs [2.16]

On a vu au § *5.6.2* les transformations successives qu'on effectuait sur les poids associés aux unités sondées de l'échantillon. Partant des poids d_i, égaux à l'inverse de la probabilité de tirage de l'entreprise i dans la strate à laquelle elle appartient, on est passé :

i. à des poids w_i^{CNRT} qui tiennent compte de la correction de la non-réponse totale ;

ii. puis à des poids w_i^{winsor} qui corrigent certaines entreprises jugées atypiques par le procédé de winsorisation ;

iii. puis enfin aux poids finaux $w_i^{calé}$ qui permettent de renormer l'échantillon sur le chiffre d'affaires et le nombre d'entreprises, au niveau groupe de la nomenclature.

Il est évidemment intéressant de regarder l'impact de ces différents changements sur la précision des estimateurs du total d'une variable Y obtenus si on se contentait de redresser les résultats de

l'enquête d'une part (en utilisant les estimateurs du paragraphe 9.1) et en comparant avec l'estimateur composite final retenu dans Ésane d'autre part.

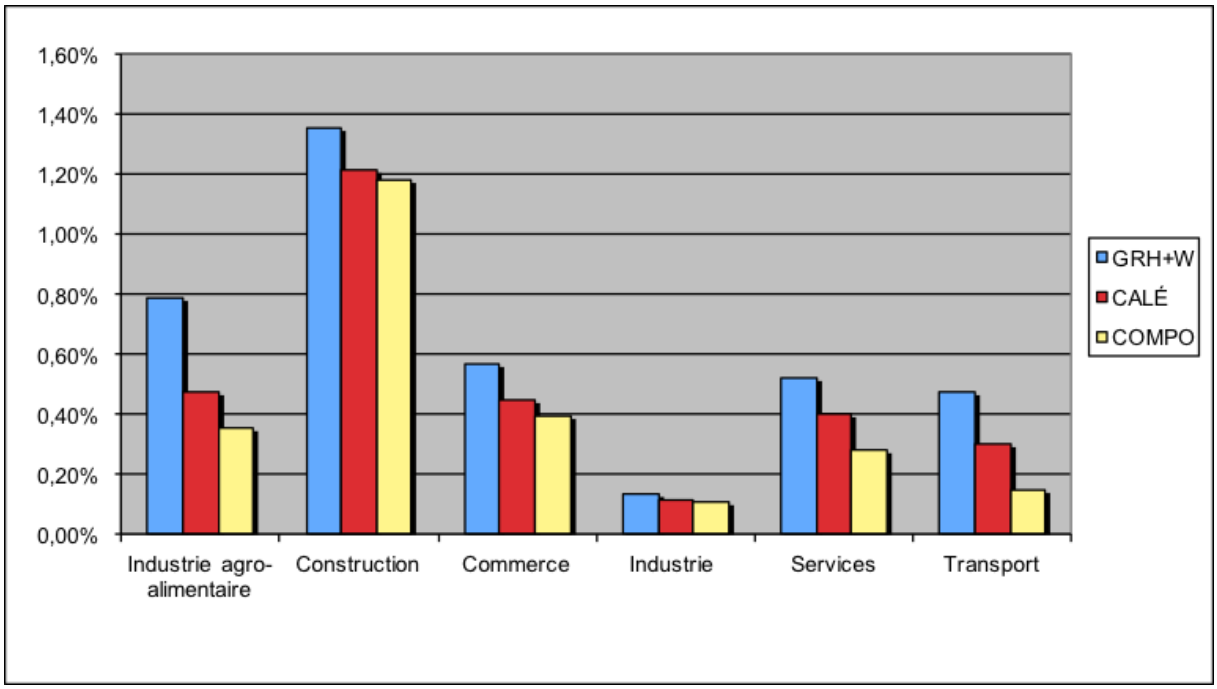
Plus précisément, on a retenu de comparer pour des niveaux sectoriels S :

1) les estimateurs pondérés classiques à l'issue des phases i) et ii) [donc munis des poids w_i^{winsor}] qu'on va appeler estimateurs "**GRH+W**" puisque les poids résultent à la fois de la correction de la non-réponse par la méthode des groupes de réponse homogène (GRH) et de la winsorisation qu'on fait subir aux poids initiaux d_i[356] ;
2) les estimateurs pondérés classiques à l'issue de la phase iii) [donc munis des poids $w_i^{calé}$] qu'on va qualifier d'estimateurs "**CALÉ**" ; (donc a priori les meilleurs estimateurs pondérés) ;
3) enfin, les estimateurs composites retenus dans Ésane qu'on a vus au § 9.2.1 ci-dessus et qu'on va appeler estimateurs "**COMPO**".

P. Brion et E. Gros ont effectué cette comparaison dans [2.16] pour un petit groupe de huit variables : nombre d'entreprises, chiffre d'affaires, frais de personnel, valeur ajoutée HT, excédent brut d'exploitation, total de l'actif et total du passif, et investissements corporels bruts. On vient de voir au § précédent (Cf. 9.2.3) que les deux estimateurs (CALÉ et COMPO) étaient identiques pour les variables chiffre d'affaires et nombre d'entreprises au niveau des "groupes" de la nomenclature (et donc aussi aux niveaux supérieurs). Il n'en est évidemment rien pour les autres variables.

La comparaison de la précision de ces estimateurs va se faire sur **le coefficient de variation**[357] associé à l'estimation de ces huit variables, pour différents niveaux de la nomenclature[358]. On se contente de donner ici les résultats concernant les deux variables Valeur ajoutée (graphique 1) et Investissement corporel (graphique 2) dont le total est mesuré sur les six grands secteurs du système productif.

Graphique 1 : *Coefficients de variation (exprimés en %) des trois types d'estimateurs (GRH+W, CALÉ, COMPO) relatifs à l'estimation de* ***la VA totale*** *pour les six grands secteurs suivis dans Ésane :*



(chiffres Ésane **2010**)

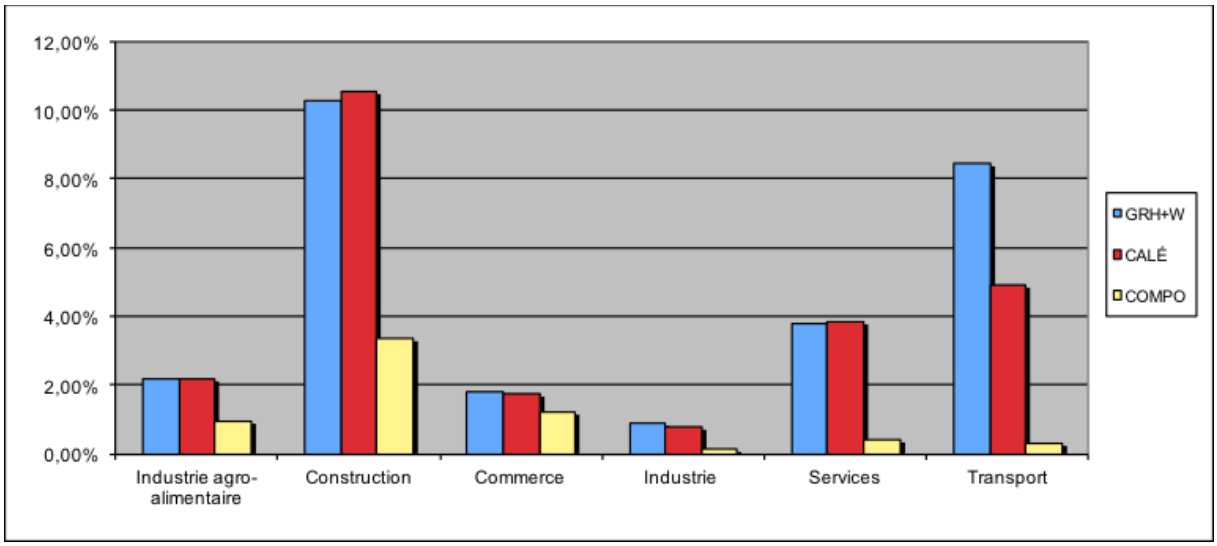
[356] On aurait pu mesurer à part l'influence de la winsorisation mais on sait qu'elle introduit un peu de biais pour gagner en variance. Les seuils de winsorisation utilisés ont été déterminés de façon à minimiser l'erreur quadratique moyenne de l'estimateur GRH+W (réaliser le meilleur arbitrage possible entre le gain en variance et l'introduction d'un biais).

[357] Rapport de l'écart type de l'estimation à l'estimation elle-même. Il permet de comparer des distributions de valeurs dont les échelles ne sont pas comparables, ce qui est le cas avec les variables et secteurs concernés ici.

[358] Le calcul de l'écart type de chaque estimateur est assez complexe et repose sur l'utilisation d'une macro ad-hoc © SAS qui tient compte du plan d'échantillonnage, du modèle de groupes de réponses homogène retenu et de l'impact de la winsorisation.

À ce niveau agrégé de la nomenclature, les estimateurs composites sont plus précis que les deux autres et CALÉ l'est davantage que GRH+W : la procédure de calage a bien amélioré la précision des estimateurs tenant compte de la non-réponse car la valeur ajoutée est bien corrélée avec le chiffre d'affaires (Cf. § 5.6.2.4) ; les estimateurs composites améliorent encore ce résultat d'un ordre de grandeur comparable.

Graphique 2 : *Coefficients de variation (exprimés en %) des trois types d'estimateurs (GRH+W, CALÉ, COMPO) relatifs à l'estimation de* ***l'investissement*** *pour les six grands secteurs suivis dans Ésane :*



(chiffres Ésane **2010**)

Cette fois-ci la procédure de calage n'améliore pas systématiquement les estimateurs initiaux car l'investissement n'est pas fortement corrélé au chiffre d'affaires (ni au nombre d'entreprises par secteur[359] !). En revanche, l'amélioration de la précision des estimateurs composites est spectaculaire ; ceci est principalement dû à la composante exhaustive de l'estimateur composite qui s'appuie sur l'ensemble des chiffres de la source fiscale. (les CV des estimateurs composites sont du même ordre de grandeur dans les deux cas ; ce sont les autres qui sont beaucoup plus élevés dans le graphique 2 : attention les échelles ne sont pas les mêmes).

A des niveaux plus fins de la nomenclature (groupes), Brion et Gros ont pu montrer que les résultats étaient qualitativement les mêmes : dans plus de 80% des cas, les estimateurs composites sont plus précis que les estimateurs GRH+W ; mais cela n'est plus vrai dans les 20% de cas restants.

9.2.5 Cas particulier de statistiques ventilant une estimation sectorielle obtenue sur une variable administrative selon des clés de répartition issues de l'Esa – en particulier calcul des agrégats branches

La clé de répartition provenant de l'Esa ou de l'EAP, elle ne porte que sur le sous champ Ésane 1, hors DOM[360] et hors EAP0[361]. En conséquence, les statistiques sectorielles ainsi ventilées sont celles **relatives au seul sous champ Ésane 1** hors DOM et hors EAP0.

Remarque préalable sur la ventilation du chiffre d'affaires en branches[362] : afin de garantir une cohérence parfaite entre la ventilation du chiffre d'affaires en branches au niveau fin, qui découle de l'enquête, et la ventilation du chiffre d'affaires en branches au niveau agrégé (ventes de marchandises / production vendue de biens / production vendue de services) disponible dans les données fiscales, on ventile en branches **séparément** les ventes de marchandises, la production vendue de biens et la production vendue de services. La ventilation du chiffre d'affaires total en branches **s'en déduit immédiatement**.

[359] Ce serait le cas si l'investissement était une constante ne dépendant que du secteur.
[360] Pour lesquels les résultats de l'enquête ne sont pas disponibles à temps.
[361] La note relative à l'EAP0 au début du § 9.2 s'applique toujours ici.
[362] Pour la définition des branches, on peut se reporter à la note 5 de l'*avant-propos* et au schéma du § 2.7.3.3.

On donne ici l'exemple de la ventilation de la production de services en branches (et on pourrait facilement étendre l'exemple au cas d'une branche de production de biens ou de ventes de marchandises). On s'intéresse aux ventes d'une branche de production de service b donnée, pour les entreprises appartenant à un secteur S donné (c'est-à-dire un élément de la fameuse matrice de passage secteur – branche).

Exemple : on veut calculer les ventes de films pour le cinéma par les entreprises qui produisent des films pour la télévision. Ces dernières constituent le secteur S=5711A. La production de films pour le cinéma correspond à la branche b= 5711C.

On cherche donc à estimer la quantité (**) :

$$\sum_{i \in U_1 \setminus \text{DOM \& EAP0}} \text{VS}(\text{b,i})\, 1\text{I}_{i \in \text{champ}}\, 1\text{I}_{\text{secteur}=S}(i) \quad (**) \ ,$$

où VS(b,i) désigne les ventes de services réalisées par l'entreprise i dans la branche b.

$1\text{I}_{i \in \text{champ}}$ permet de ne prendre en compte que les unités dans le champ au final. Ainsi par exemple les unités non répondantes considérées comme statistiquement inactives étaient présentes dans le champ initial mais ne font plus partie du champ de diffusion et cette indicatrice vaut 0 pour de telles unités. Par définition, cette indicatrice est inutile lorsque l'on ne somme que des entreprises répondantes qui sont par définition dans le champ et pour lesquelles cette indicatrice vaut toujours 1.

Compte tenu de la remarque préalable précédente, cette quantité (**) sera estimée par l'estimateur composite suivant :

[formule (5)] :

$$\hat{V}S_{\text{composite}}(\text{S,b}) = \left(\hat{V}S^{\text{ech}}_{\text{diff}}(S)\right) \frac{\sum_{i \in R} w_i\, \text{VS}^{\text{Redi}}(\text{b,i})\, 1\text{I}_{\text{Secteur_enq}=S}(i)}{\sum_{i \in R} w_i\, \text{VS}^{\text{Redi}}(i)\, 1\text{I}_{\text{Secteur_enq}=S}(i)}$$

Sur la partie relative au sous champ Ésane 1 hors DOM et hors EAP0, on applique donc à l'estimateur composite des ventes de services des entreprises appartenant au secteur S une clef de répartition en branche, mesurée au sein des branches de services, estimée à partir des estimateurs calés issus de l'enquête.
À noter que ce faisant lors de ce « calage », on est le plus souvent amené à **modifier les niveaux des branches fines** tels qu'ils ressortaient des estimateurs calés de l'enquête, puisque l'on applique une structure à un estimateur différent d'une ventilation agrégée.

Exemple :

Prenons l'exemple suivant où l'on a d'un côté la ventilation du chiffre d'affaires avec son estimateur composite, et de l'autre l'estimateur calé calculé sur l'échantillon.

Les données des unités appartenant à l'échantillon ont été réconciliées, et les égalités individuelles sont donc respectées entre CA dans l'échantillon et CA fiscal mais cela ne nous assure pas la cohérence des estimateurs qui n'existe qu'au niveau « groupe » (Cf. § 9.2.3).

Estimateur composite sectoriel des variables fiscales : [*obtenu par la formule (3)*]			**Estimateur calé des branches (issues de l'ESA ou l'EAP)**	
Chiffre d'affaires	**800**	≠	**Chiffre d'affaires**	**880**
Vente de marchandises	300	≠	Branche 1 – M	330
Production de biens vendue	100	≠	Branche 2 – B	220
Production de services vendue	400	≠	Branche 3 – S	110
			Branche 4 – S	220

On est confronté à 2 cas :

i) soit la ventilation fine est possible (il existe des branches fines pour ventiler tous les postes de la ventilation agrégée). Dans ce cas, la somme des ventilations fines (somme des branches au sein de chaque poste de la ventilation agrégée) n'est généralement pas égale à

l'estimateur de la ventilation agrégée du chiffre d'affaires. On va donc dilater (ou contracter) les montants des branches, de façon à ce que les totaux soient égaux.

donc le suivant :

Estimateur composite final des branches [*Obtenu par la formule (5)*]	
Chiffre d'affaires	**800**
Branche 1 – M (contractée) (300 * 330/330)	300
Branche 2 – B (contractée) (100 * 220/220)	100
Branche 3 – S (dilatée) (400 * 110/330)	133
Branche 4 – S (dilatée) (400 * 220/330)	266

ii) Soit la ventilation fine n'est pas possible (par exemple il n'existe pas de branches de marchandises dans les réponses des entreprises enquêtées, alors que l'estimateur composite des ventes de marchandises est non nul). Dans ce cas, on va générer une branche agrégée par défaut (on utilise les branches par défaut de Rédi, listées dans le tableau Excel mentionné au § *8.3.3.2*) pour que les totaux soient égaux (cf. exemple du tableau ci-dessous). Cette branche n'existe pas dans les ventilations individuelles des entreprises de ce secteur.

entreprise du secteur S de l'échantillon n'a de branche commerciale alors que dans la source fiscale on trouve au moins une entreprise du secteur S qui vend des marchandises. L'estimateur composite de la ventilation agrégée du chiffre d'affaires est alors calculé :

- par la formule (3) si aucun des postes de la ventilation agrégée du chiffre d'affaires n'est nul au niveau du groupe de la nomenclature (poste à 3 chiffres) auquel appartient le secteur S c'est-à-dire si aucun de ces postes n'est à occurrence rare pour ce groupe. On notera que dans ce cas, l'estimateur des ventes de marchandises du secteur S est calculable (le dénominateur de la formule (3) est non nul) **mais il est nul** puisqu'il n'y a pas de ventes de marchandises. La ventilation fine est donc tout à fait possible comme dans l'exemple précédent ;
- par la formule (4) si l'un des postes est à occurrence rare c'est-à-dire si on ne trouve pas de branche commerciale dans l'échantillon dans aucune des autres sous-classes du groupe G auquel appartient le secteur S. Dans ce cas, on ne peut pas utiliser la formule (3) et l'estimateur des ventes de marchandises du secteur **S sera non nul** puisque le numérateur qui est calculé sur les ventes de marchandises des liasses du secteur est par hypothèse non nul. Imaginons qu'on ait la répartition en branches fines suivante

Estimateur composite sectoriel des variables fiscales [*obtenu par la formule (4)*]			**Estimateur calé des branches (issues de l'ESA ou l'EAP)**	
Chiffre d'affaires	800	≠	Chiffre d'affaires	880
Vente de marchandises	300	≠		0
Production de biens vendue	100	≠	Branche 1 – B	220
Production de services vendue	400	≠	Branche 2 – S	110
			Branche 3 – S	330
			Branche 4 – S	220

L'estimateur composite final des branches sera le suivant ; on conserve les postes de la ventilation agrégée comme dans le cas précédent, ce qui oblige à « inventer » une branche fine de ventes de marchandises et on ventile le total des ventes de services au pro rata des branches observées dans l'échantillon :

Estimateur composite final des branches [*Obtenu par la formule (5)*]	
Chiffre d'affaires	**800**
Branche 1 – M (**la branche fine va devoir être créée**)	300
Branche 2 – B (contractée)	100
Branche 3 – S (contractée) (400 * 110/660)	67
Branche 4 – S (dilatée) (400 * 330/660)	200
Branche 5– S (contractée) (400 * 220/660)	133

Le calage des poids (cf. § *9.2.3*) assure une bonne correspondance des estimateurs de chiffre d'affaires au niveau groupe de la NAF.
Par conséquent, les estimations entre branches fines et ventilations agrégées ne sont pas trop mauvaises sur le « cœur de métier » où la corrélation avec le chiffre d'affaires du secteur est forte (par exemple correspondance entre branches commerciales et vente de marchandises pour les secteurs du commerce). En revanche, les estimations divergent plus en dehors du cœur de métier (branches de services et production de services vendue pour les secteurs du commerce).

i) Dans le premier cas (dilatation/contraction des branches) on conserve les niveaux de la ventilation agrégée issus de l'estimateur composite, et la structure des branches fines issue de l'enquête.
Pour l'année 2009 par exemple, 68 % des coefficients de dilatation/contraction utilisés étaient compris entre 0,9 et 1,1.

On peut donc supposer que dans les secteurs où :

- on a peu d'entreprises dans l'échantillon et/ou
- l'activité est plus concentrée,

on peut avoir quelques répondantes utilisées une année, avec éventuellement une dilatation forte des branches, et l'année suivante, pas de branches fines permettant de ventiler les ventes de marchandises par exemple. Dans ce cas on utiliserait alors la branche Rédi par défaut (pas nécessairement égale à la branche dilatée l'année précédente).

Cette technique est généralisable à toute variable résultant de la ventilation d'une estimation sectorielle, obtenue sur une variable comptable, selon une clé de répartition issue de l'Esa.

En ce qui concerne spécifiquement **la ventilation du chiffre d'affaires en branches**, (un des agrégats importants calculé et publié par Ésane) des branches sont imputées par le mécanisme Rédi pour l'ensemble des unités appartenant à des champs non couverts par l'enquête : EAP0, DOM et sous champs Ésane 2 à 5. En conséquence, il est possible de calculer une ventilation du chiffre d'affaires en branche pour l'EAP0, les DOM et les sous champs Ésane 2 à 5, et **donc une ventilation pour l'ensemble du champ Ésane** par sommation de ces quantités avec la ventilation relative au sous champ Ésane 1 hors DOM et hors EAP0 calculée précédemment.

9.2.6 Cas particulier des agrégats sectoriels des variables spécifiques à l'Esa

L'enquête sectorielle annuelle contient un certain nombre de variables dites « sectorielles » correspondant à des questions propres aux six grands secteurs de l'enquête (Cf. § *5.1.1*). Le calcul des agrégats relatifs à ces variables n'a donc de sens que pour les secteurs du sous-champ 1 et pour autant que ces secteurs soient bien couverts par l'Esa (Cf. § *4.4.3*). L'estimateur naturel pour ces agrégats est un estimateur calé classique.

Cependant, comme on l'a vu au § *9.1*, on peut améliorer l'estimateur calé du chiffre d'affaires par un terme correctif mobilisant deux termes calculés sur des données administratives seulement. De la même façon, on peut souhaiter améliorer les estimateurs calés des variables de l'ESA, grâce à l'information obtenue dans la source fiscale. De nombreuses variables de l'enquête sont d'ailleurs très fortement corrélées au chiffre d'affaires, voire sont directement une ventilation de ce chiffre d'affaires. Il serait alors peu cohérent de publier des informations liées au chiffre d'affaires non cohérentes avec le chiffre d'affaires estimé et publié par ailleurs.

C'est pourquoi les estimateurs calés des variables de l'Esa sont « dilatés » selon une variable auxiliaire dont on dispose sur le sous champ 1 et corrélée avec la variable concernée, « Y », selon le processus suivant :

- calcul de l'estimateur calé du total de Y, sur le domaine de diffusion voulu D, sur l'échantillon de l'Esa ;
- pondération de cet estimateur par le ratio entre l'estimateur par différence sur le même domaine et l'estimateur calé sur le même domaine **d'une variable auxiliaire A adaptée au cas de chaque variable spécifique**.

La formule est donc la suivante :

[Formule (6)] :

$$\boxed{\hat{Y}_{ratio}^{ESA} = \hat{Y}_{HT}^{ech}(D)\frac{\hat{A}_{Diff}^{ESA}(D)}{\hat{A}_{HT}^{ech}(D)}}$$

Remarque : si D est un domaine inclus dans un groupe G de la nomenclature alors il faut remplacer : $\hat{A}_{Diff}^{ESA}(D)$ par $\hat{A}^{ESA}_{infra}(D)$. En effet, l'estimateur composite de A est alors l'estimateur infra-groupe [(formule (3) ou (4)] et non la formule (1).

Il existe 4 variables auxiliaires utilisées selon les cas :

- Le chiffre d'affaires. Il est utilisé pour toutes les variables de l'enquête qui sont des ventilations du chiffre d'affaires. *Exemple :* la ventilation du chiffre d'affaires par type d'ouvrage dans le bâtiment.
- Le nombre d'entreprises. Il est utilisé pour les agrégats relatifs à un nombre d'entreprises. *Exemple :* le nombre d'entreprises ayant fait des dépenses de publicité.
- Le total des achats. Il est utilisé pour toutes les variables de dépenses. *Exemple :* le montant des frais de publicité.
- L'effectif ETP. Il est utilisé pour toutes les variables sur l'emploi. *Exemple :* effectif non salarié, effectif des intérimaires etc.

Cet estimateur permet bien :

- de prendre en compte les effets de l'estimateur par différence qui mobilise plus d'informations que l'estimateur calé simple ;
- d'être calé sur des agrégats fiscaux ou d'emploi connus par ailleurs lorsque la variable d'intérêt est une ventilation d'une variable fiscale ou d'emploi. Ceci permet d'assurer par exemple que la somme d'une ventilation du chiffre d'affaires redonne bien le chiffre d'affaires total diffusé pour le sous champ 1 du domaine étudié.

9.3 Les limites de ces estimateurs et les précautions à prendre dans leur usage [2.9]

9.3.1 Le caractère sommable des agrégats ne peut pas être assuré systématiquement à tous les niveaux

Ce point a déjà été évoqué ci-dessus mais c'est le problème majeur de l'estimateur final retenu : il est non linéaire sauf au niveau groupe. Ainsi il est impossible d'assurer, systématiquement, **à la fois une sommabilité verticale et horizontale des variables étudiées en dessous du niveau groupe**.

L'estimateur choisi permet d'assurer une additivité verticale ; ainsi, pour une variable donnée, la somme des agrégats au niveau infra groupe redonne l'agrégat calculé au niveau du groupe.

En revanche, il ne permet pas d'assurer systématiquement une additivité horizontale. Ainsi, s'il existe deux ventilations d'une même variable, ou, si une variable peut être calculée comme solde de deux équations comptables différentes, le respect des relations liant les différentes variables est assuré au niveau du groupe **mais ne le sera pas nécessairement au niveau infra-groupe**.

Prenons l'exemple du bilan où le total de l'actif se doit d'être égal au total du passif. Dans l'exemple ci-dessous, le groupe est bien la somme de ses sous-classes et l'actif est égal au passif au niveau groupe mais pas au niveau sous-classe.

Pour arriver à une égalité, il faut accepter de « solder » l'écart sur une variable. L'agrégat de cette variable ne respecte alors plus la méthodologie de calcul décrite ci-dessus pour le niveau infra groupe. Dans le cas du bilan, on solderait l'écart sur la variable « report à nouveau » des capitaux propres. Cette variable a le bon goût de pouvoir être positive ou négative. En revanche la modifier, simplement pour équilibrer des tableaux, aurait un impact non négligeable sur les capitaux propres eux-mêmes, qui sont une autre variable d'intérêt de l'enquête. **C'est pourquoi cette correction n'est pas faite** ; mais, si on est plus intéressé par l'équilibre actif-passif que par le contenu des capitaux propres, elle pourrait l'être de la façon suivante :

		valeur (groupe)	sous classe 1 (en %)	sous-classe 2 (en %)	sous classe 1 (en montant)	sous-classe 2 (en montant)	total sous classes
Actif	**Var1**	300	80	20	240	60	300
Actif	**Var2**	200	60	40	120	80	200
Actif	**Var3**	100	30	70	30	70	100
	total	**600**			**390**	**210**	**600**
Passif	**Var4**	150	90	10	135	15	150
Passif	**Var5**	250	70	30	175	75	250
Passif	**Var6**	200	50	50	100	100	200
	total	**600**			**410**	**190**	**600**

On suppose qu'il s'agit de la variable 6 dans notre exemple et on obtiendrait alors la ventilation définitive suivante : elle assure bien l'équilibre actif – passif au niveau sous-classe, mais les agrégats correspondant à la variable 6 ne sont plus alors calculables à partir des variables 6 des entreprises des deux sous-classes.

		valeur (groupe)	sous classe 1 corrigé (en montant)	sous-classe 2 (en montant)	total sous classes corrigées
actif	**Var1**	300	240	60	300
actif	**Var2**	200	120	80	200
actif	**Var3**	100	30	70	100
	total	**600**	**390**	**210**	**600**
Passif	**Var4**	150	135	15	150
Passif	**Var5**	250	175	75	250
Passif	**Var6**	200	80	120	200
	total	**600**	**390**	**210**	**600**

9.3.2 La somme des données individuelles ne redonne pas la valeur de l'agrégat

La formule de l'estimateur par différence est plus complexe que la formule des estimateurs qu'on utilisait auparavant dans le cadre de Suse ou des EAE (Cf. *Avant-propos*). Dans Suse, par exemple, l'estimateur du chiffre d'affaires d'un domaine était simplement la somme des chiffres d'affaires des entreprises du domaine. Désormais on tient compte des entreprises qui sont rentrées et sorties du domaine au cours de l'année considérée mais, comme on ne les connaît pas exhaustivement, on raisonne en inférant de ce que l'on peut observer du comportement des entreprises du domaine D dans l'échantillon, par rapport à ces entrées sorties ainsi que par rapport à la déformation des données entre l'enquête et la source fiscale.
La formule (1) demande d'utiliser deux valeurs différentes pour la même caractéristique (valeur issue de l'enquête et valeur au moment du tirage de l'échantillon pour l'APE, valeur de la liasse et valeur après réconciliation pour les données fiscales, etc.). Ceci n'est pas nécessairement intuitif pour un utilisateur qui n'est pas rompu à toute la méthodologie mise en œuvre pour l'élaboration de ces statistiques.
C'est pourquoi le nom du fichier de diffusion de données individuelles est **FARE** pour « Fichier **Approché** des Résultats d'Ésane » car **il ne permet pas de retrouver** par simple sommation les estimateurs que l'Insee publie par ailleurs. Ce fichier contient un certain nombre de caractéristiques pour l'ensemble des unités du champ. Il ne comprend par essence que les valeurs **issues de la**

réconciliation et **l'APE de diffusion**. Ce fichier est parfait pour réaliser des études économétriques sur les entreprises par exemple, que ce soit en interne de l'Insee ou pour les chercheurs qui peuvent accéder à ces données après autorisation du comité du secret (au travers du Centre d'accès sécurisé distant aux données).

Le chapitre 10 dans lequel on explicite les méthodes utilisées pour vérifier la vraisemblance des agrégats calculés donne une idée de l'impact de Rédi et des changements d'APE (en général quelques pour cents) (Voir par exemple le tableau 4 du § *10.5.2.2*).

9.3.3 Certains agrégats en niveau (comme le nombre d'entreprises) sont plus variables que par le passé

Par le passé, le calcul des agrégats fiscaux sectoriels se faisait à partir du fichier Suse par une simple somme des caractéristiques des unités d'un secteur donné.
L'APE des entreprises du fichier Suse était la suivante :

- celle issue de l'EAE pour les unités répondantes de l'échantillon ;
- celle de Sirene pour les autres unités.

L'hypothèse implicite sous-jacente faite était donc **qu'aucune unité hors échantillon ne changeait de secteur au cours de l'année d'observation**, hypothèse évidemment fausse mais qui assurait une grande stabilité apparente aux résultats. En effet, hormis les changements d'APE en provenance de l'EAE et certaines EAR[363] spécifiques, les APE de Sirene sont rarement mises à jour à l'initiative des entreprises et sont donc le plus souvent fixes.

L'un des principes forts de l'estimateur composite est au contraire **d'extrapoler les changements d'APE observés sur l'échantillon Esa** à l'ensemble de la population et donc de considérer qu'ils sont représentatifs des comportements généraux de mobilité sectorielle. Sa faiblesse corrélative est que, sur de petits domaines, certains changements d'APE atypiques peuvent, en raison de leur poids, exagérer certains mouvements.

La principale incertitude concerne l'estimation **du nombre d'entreprises** : en effet, pour les agrégats fiscaux les résultats des entreprises sont en quelque sorte « pondérés » par la valeur de leur caractéristique. De ce fait, **la stabilité des agrégats reste de mise**, car les grandes entreprises d'un secteur changent rarement d'activité et sont donc stables ; et lorsqu'elles en changent, on n'a pas de mal à expliquer l'impact sur les agrégats sectoriels. Le changement de secteur des entreprises de plus petite taille est du deuxième ordre dans l'agrégat du secteur {là encore le chapitre 10 permet de quantifier les poids respectifs de l'exhaustif (les grandes) par rapport à la partie échantillonnée (les petites) ; voir le tableau 3 du § *10.5.2.1*) colonnes [1] et [2]}.

En revanche pour le nombre d'unités, les entreprises, qu'elles soient grandes ou petites, comptent pour « un ». Dans ce contexte, l'aléa de sondage joue un rôle plus important, et les estimations de nombre d'entreprises par secteur sont moins précises et par conséquent peuvent être **relativement variables d'une année sur l'autre**.

Ce phénomène est très nettement atténué au niveau de l'agrégat du chiffre d'affaires (ou des grandes autres variables fiscales) dans la mesure où ces petites entreprises ne représentent généralement qu'un faible pourcentage du chiffre d'affaires d'un secteur ; l'impact du départ ou de l'arrivée de telles unités est donc bien moindre sur ce type de variable.

9.4 La diffusion des résultats d'Ésane

L'objectif d'un système d'information comme Ésane est bien entendu de diffuser les résultats agrégés et les estimations que les formules précédentes ont permis de calculer. L*'avant-propos* a indiqué les nombreux publics et les objectifs variés que ce système vise à satisfaire. L'un d'entre eux est d'avoir le plus tôt possible (Cf. § *9.4.1*, le calendrier de diffusion) une « première idée » de la performance globale du système productif français, une année donnée.

C'est la raison pour laquelle une première estimation de quelques variables clefs (emploi, production, valeur ajoutée) alimente entre autres ce qu'on appelle les comptes semi-définitifs de la comptabilité

[363] Ces « enquêtes d'amélioration du répertoire » (EAR) sont des opérations administratives (et non statistiques) destinées à améliorer la qualité du répertoire. Mais elles ne touchent annuellement que quelques dizaines de milliers d'unités (sur un stock d'environ 5 millions).

nationale[364] (Cf. § 9.4.2). Ces résultats préliminaires sont également demandés et publiés par la Commission européenne dans le cadre du règlement SBS.

On a rappelé dans le schéma du début du chapitre le volume considérable de données élémentaires collectées chaque année. Même si on se limite aux variables du socle (Cf. § 6.2.4), on a en tout près de 1000 variables, se référant à 666 sous-classes (Cf. § 4.1). Le volume d'agrégats que l'on peut calculer est donc lui aussi considérable surtout si l'on veut « croiser » des critères (sous-classe x taille ou catégorie juridique). Il a donc fallu concevoir des « produits de diffusion » particuliers, adaptés pour certains à la nécessité de pouvoir synthétiser l'information et pour d'autres à celle d'avoir en parallèle une information à des niveaux de détail fins. Leur description sera l'objet du § 9.4.3.

Mentionnons également l'existence dans la collection « **Insee références** » de l'ouvrage « ***Les entreprises en France*** », (à compter de l'édition 2013 pour les résultats 2011) qui a pour vocation d'offrir une vue complète de notre système productif en s'appuyant sur des données individuelles d'entreprises. Une partie importante des statistiques diffusées dans cette publication s'appuie sur le dispositif Ésane.

N'oublions pas enfin l'EAP, qui joue un rôle parallèle à l'Esa, dans l'industrie, mais qui bénéficie d'une publication autonome concernant la production annuelle de produits industriels à un niveau très fin (Cf. § 9.4.6).

9.4.1 Le calendrier de diffusion

Le calendrier **cible** type correspondant à l'année de constat N est indiqué dans le tableau ci-dessous.

Liste des différentes diffusions	Calendrier pour une campagne N
Diffusion Eurostat « préliminaire »	Fin octobre N+1
Diffusion DCN 31/12 (dite semi-définitive)	31/12/N+1
Diffusion Eurostat définitive	Fin juin N+2
Diffusion définitive (Insee.fr)	Été N+2[365]

On pourra également se référer au déroulement chronologique détaillé d'une campagne présenté au chapitre 12.

9.4.2 La diffusion préliminaire pour Eurostat

Cette diffusion concerne un nombre limité de variables et une méthodologie de calcul des agrégats beaucoup plus simple dans la mesure où :

- on ne dispose que de la livraison (en juin) des liasses anticipées qui ont été contrôlées par les gestionnaires ; avec les progrès de l'informatisation des déclarations, cette livraison couvre cependant une part de plus en plus grande de l'économie ;
- le traitement de la non-réponse totale de l'enquête Esa n'a pas encore eu lieu ;
- les traitements de réconciliation des données n'ont pas encore commencé ;
- on ne dispose pas de données d'emploi validées par Clap.

Les variables diffusées sont : le chiffre d'affaires, la VA, l'EBE, les salaires, le résultat comptable et les investissements de la liasse fiscale ;

Un traitement de la non-réponse totale particulier est fait pour les liasses fiscales absentes. Il reprend la même méthodologie que celle exposée dans le § 6.5.3 : il privilégie une extrapolation par N-1 à partir du moment où il existe des données en N-1 même si ces dernières ont été obtenues par imputation. Ainsi seules les unités « créées » sans liasse fiscale sont estimées par la médiane.

Les agrégats obtenus sont calculés par une simple somme pour les données fiscales.

La diffusion à ce niveau est faite **au niveau groupe de la NAF**.
Ces estimations permettent de **répondre à la demande de résultats préliminaires du règlement SBS**.

[364] Les données qui leur sont fournies précisément pour le 31/12 ne sont pas détaillées ici.
[365] Parce que la chaîne aboutissant à la mise à disposition sur le site est assez complexe, il arrive en fait assez souvent que les chiffres clefs soient publiés mais que les chiffres détaillés ne le soient que plus tard.

9.4.3 La diffusion définitive sur le site de l'Insee

Cette diffusion annuelle intervient :

- une fois que l'ensemble des traitements à réaliser par les gestionnaires l'ont été ;
- une fois que l'ensemble des traitements automatiques de fin de campagne l'ont été (Rédi balai, CQR Esa et emploi post Rédi) ;
- après validation des résultats (Cf. *chapitre 10*) ;
- après la « pose du secret » (Cf. § *9.5*) ;
- et enfin une fois que la direction de la diffusion a mis à jour le site de l'ensemble des agrégats correspondants.

Elle va permettre :

- d'alimenter le site Insee.fr ;
- d'alimenter la base (interne) des groupes (LIFIUS) pour calculer des résultats en entreprises (cf. *chapitre 11* pour plus de détails) ;
- de mettre à disposition des chercheurs qui auront reçu l'agrément du comité du secret, les données individuelles des entreprises.

Avant de rentrer dans le détail de ce qui est publié, il faut d'abord garder en tête la coupure introduite par les années 2012 et 2013, à la fois dernières années de publication des résultats où l'unité statistique était l'unité légale (lors de la première publication des résultats de 2012) et première année de publication des résultats en entreprise avec la prise en compte de la cible 1 des groupes (Cf. § *11.6.6* et § *4.8.1*), lors de la publication des résultats de 2013 à l'été 2015 (où on a également fait des calculs d'agrégats en unités légales). **A compter de l'été 2015, les résultats d'Ésane sont donc publiés « en entreprises »**, les entreprises étant soit des unités légales indépendantes, soit des parties autonomes des très grands groupes de la cible 1. En attendant un traitement similaire de l'ensemble des groupes, en principe à l'été 2019 lors des résultats de la campagne 2017, les unités légales, filiales des autres groupes, sont donc encore considérées comme des entreprises également. Au delà il n'y aura plus que des entreprises au sens économique d'acteurs autonomes sur un marché.

Les résultats portant sur la période 2009-2013, **en unités légales donc**, sont toutefois toujours accessibles sous la rubrique « **caractéristiques comptables, financières et d'emploi des unités légales de 2009 à 2013** »[366] qu'on ne décrit pas en détail, mais on y retrouve in fine les mêmes tableaux que ceux décrits ci-après. En revanche, sur cette période, les données ne sont **pas strictement comparables** d'une année sur l'autre puisqu'il n'y a qu'une série de tableaux par année ; il faut faire attention aux variations de champ (Cf. § *4.6*). Les évolutions en champ comparable ne sont disponibles qu'à compter de l'année de constat 2012.

On décrit ci-après les données mises en ligne dans la **nouvelle version du site Insee.fr** datant de novembre 2016. Elles se décomposent en deux parties relativement disjointes : des **chiffres clés** d'un côté, des **chiffres détaillés** de l'autre. Les premiers sont évidemment contenus dans les seconds. On se contentera donc de décrire comment on accède aux premiers, sans rentrer dans le détail des tableaux (qui sont de plus *susceptibles d'évolutions*).

9.4.3.1 Les chiffres clés

Depuis l'onglet « STATISTIQUES » de la page d'accueil du site Insee.fr, on accède aux données que l'on souhaite en spécifiant un certain nombre de critères dans le bandeau de gauche. Mais la spécificité de cette recherche est que l'on n'a pas besoin de connaître ni d'indiquer la « source » des données (ici Esane) qui n'est précisée que sous les tableaux eux-mêmes des chiffres clés sélectionnés.

- Dans la rubrique « THÈMES » il faut choisir [« **entreprises** » puis **« caractéristiques des entreprises »** (comme sous-rubrique)] ou [« **secteurs** » puis « **un des six grands secteurs de l'Esa/EAP**[367] »] selon que l'on souhaite rester au niveau global ou spécifier un secteur ou non (les tableaux obtenus ne sont pas les mêmes) ;
- Dans la rubrique « NIVEAU GÉOGRAPHIQUE », on spécifie « **France** » ;
- Dans la rubrique « CATÉGORIES » on spécifie « **données** » (et non pas publications), puis « **chiffres clés** » ;

[366] *https://www.insee.fr/fr/statistiques/2410446?sommaire=2410634* . Les données de 2008 ne sont accessibles que dans la base Alisse. Voir § *9.4.3.2.C* .
[367] Commerce, Industrie (dont IAA), Construction, Services, Transports.

On a alors accès à une liste d'un certain nombre de petits tableaux ou graphiques (s'agissant de chiffres clés) visibles à l'écran dès que l'on clique dessus mais que l'on peut également télécharger en format Excel ; ils sont **repérés par un titre** qui décrit le contenu mais pas la ou les sources dont ils sont issus.

La première séquence (*Entreprises/Caractéristiques des entreprises/France/Données/Chiffres clés*)[368] donne accès – parmi éventuellement d'autres propositions ; l'Insee peut rajouter des tableaux à tout moment) – aux **neuf tableaux** ci-dessous ; les **sept premiers** sont **issus d'Ésane et les deux derniers** exploitent à la fois **Ésane et Lifi** (Cf. § *11.3*) **;** le titre précise la campagne sur laquelle porte les tableaux (systématiquement la dernière disponible) ; à un instant donné les tableaux **ne se réfèrent donc pas forcément tous à une même campagne** en raison de délais de mise à jour différents selon les sources :

- Caractéristiques des secteurs marchands par activité en N ;
- Dispersion de la valeur ajoutée des entreprises en N ;
- Chiffre d'affaires des entreprises selon l'activité en N ; (il s'agit d'un tableau croisant des secteurs et des branches) ;
- Bilan des entreprises marchandes en N ;
- Compte de résultats des entreprises marchandes, hors agriculture et secteurs financiers en N ;
- Caractéristiques des secteurs marchands selon la taille des entreprises en N ; (en nombre de salariés avec 2 seuils 10 et 250)
- Ratios des secteurs marchands en N.

Les deux tableaux issus à la fois d'Ésane et de Lifi sont les suivants :

- Caractéristiques des entreprises par catégorie en N (il s'agit des catégories issues du décret 1354-2008 qu'on rappelle par exemple au § *11.4*)
- Caractéristiques des secteurs marchands selon l'origine du groupe en N

La deuxième séquence (Par exemple pour le commerce : *Secteurs/"commerce"/"commerce de détail"/France /Données/Chiffres clés*)[369] donne accès aux **trois tableaux** suivants **issus d'Ésane**[370] (et de même pour chacun des grands secteurs) :

- Caractéristiques du commerce de détail par activité en N ;
- Ratios du commerce de détail en N ;
- Caractéristiques du commerce de détail selon la taille des entreprises en N ;

9.4.3.2 Les chiffres détaillés

On accède aux résultats détaillés d'Ésane par la séquence :
Entreprises/Caractéristiques des entreprises/France/Données/chiffres détaillés et il faut alors choisir comme résultat de la recherche « **Caractéristiques comptables financières et d'emploi des entreprises en N** »[371]. Contrairement aux chiffres clés, la source (ici Ésane) figure sous l'intitulé. On se retrouve alors sur la page de « sommaire des résultats d'Ésane », laquelle fait apparaître deux onglets intitulés « **PRÉSENTATION** » et « **DOCUMENTATION** ».

L'onglet « DOCUMENTATION » renvoie comme son nom l'indique sur une présentation générale d'Ésane et sur le présent Insee Méthodes ; il pointe également sur la NAF puisque tous les résultats sont articulés sur la nomenclature d'activités française officielle ; et enfin il permet de télécharger un document qui donne **la liste et la définition de toutes les variables publiées** d'Ésane (31 pages).

L'onglet « PRÉSENTATION » met en évidence **trois produits de diffusion** sous forme de « téléchargement de tableaux » en format Excel :

1) **Résultats sur les entreprises** : cette rubrique se décompose elle-même en six sous-rubriques détaillées au § *9.4.3.2.A*. Tous les tableaux de résultats croisent des variables en colonne avec une rubrique de la nomenclature (couplé quelquefois à une tranche de taille) en ligne. Cette ligne a trait aux secteurs d'entreprises correspondant aux intitulés de nomenclature considérés. Rappelons qu'un secteur est l'ensemble des entreprises dont l'activité principale (code APE) appartient au niveau de nomenclature défini par le secteur ; ce niveau peut être à un, deux, trois, quatre ou cinq caractères (correspondant aux sections, divisions, groupes, classes ou sous-classes de la nomenclature d'activités).

[368] *https://www.insee.fr/fr/statistiques?taille=20&debut=0&theme=39&categorie=4&geo=FRANCE-1*
[369] *https://www.insee.fr/fr/statistiques?debut=0&idprec=2015133&theme=51&categorie=4&geo=FRANCE-1*
[370] Toujours parmi d'autres intitulés de tableaux qui, eux, ne proviennent pas d'Ésane.
[371] *https://www.insee.fr/fr/statistiques/2384048* pour l'année N=2014 par exemple.

2) **Les branches d'entreprises** : cette rubrique est décrite au § *9.4.3.2.B*. Elle recouvre deux tableaux dont les lignes se rapportent cette fois-ci à des branches de la nomenclature. Rappelons que les branches sont définies conceptuellement comme l'ensemble des parties d'entreprises qui contribuent à une activité donnée, que ce soit l'activité principale ou secondaire de l'entreprise[372].

L'ensemble des 14 tableaux publiés dans le cadre des rubriques 1 et 2, permettent de publier 152 variables pour quelque 3292 lignes, soit environ **500 000 cases** ; certes un certain nombre ne sont pas disponibles (ce serait ce que parce que Ésane ne couvre que 666 sous-classes) ou sont sous secret statistique (Cf. § *9.5*). Mais cela donne une idée du nombre de contrôles qui doivent être effectués pour assurer une bonne qualité de la diffusion.

3) **Alisse** : cette rubrique (Cf. § *9.4.3.2.C*) pointe en réalité non plus sur des tableaux tout faits mais sur une véritable base de données dans laquelle l'utilisateur définit lui-même les croisements de variables et de secteurs qui l'intéressent. Il peut aussi prendre en compte plusieurs années mais se référant bien sûr aux mêmes unités statistiques. Comme l'unité statistique a changé il y a en réalité deux bases, une en unités légales avec des données de 2008 à 2013 et une en entreprises à compter de 2012. Attention que le champ global de ces bases est limité aux **secteurs principalement marchands non agricoles et non financiers** (voir explicitation dans le § 9.4.3.3) ; les changements de champ explicités aux § *4.6* à *4.8* d'une année sur l'autre ne sont pas gérés. On met à chaque fois la dernière version disponible. La comparaison à champ constant est en revanche possible sur deux années consécutives dans la rubrique suivante.

4) Une **quatrième rubrique** (Cf. § *9.4.3.2.D*) vient donc compléter ces trois premières : celle qui permet de comparer l'année N à l'année N-1 sur le même champ, intitulée « **données N-1 pour comparaison** ». Elle donne accès à un zip qui regroupe les 14 fichiers Excel identiques à ceux des deux rubriques 1 et 2 précédentes, mais portant sur l'année N-1, avec le même champ et la même méthodologie qu'en N pour assurer la comparabilité.

Enfin, Ésane ne donne pas lieu qu'à des tableaux de données agrégées mais permet aussi de construire des « fiches descriptives » de grands secteurs de la nomenclature. Ces **« fiches sectorielles** », qui constituent un **cinquième produit** issu d'Ésane (et d'autres sources, donc non accessible via le sommaire d'Ésane), sont décrites dans le § *9.4.3.2.E* et sont publiées également sur papier dans le cadre de l'Insee Références « les entreprises en France ».

9.4.3.2.A. Résultats sur les entreprises

Ces résultats sont présentés sous forme de six sous-rubriques correspondant aux six thèmes suivants :

1) Principales caractéristiques ;
2) Bilan (actif brut, amortissements, passif) ;
3) Compte de résultats ;
4) Emploi ;
5) Investissement ;
6) Ratios comptables.

A chaque thème sont associés **deux tableaux Excel** qui diffèrent par le niveau de nomenclature de secteurs utilisé :

- le premier regroupe tous les résultats en **cinq niveaux de nomenclature imbriqués** ; le § *4.4.1* explicite les 666 sous-classes (sur 732) qui sont suivies dans Ésane et que l'on peut donc retrouver dans ces tableaux :
  - Niveau A21 : sections ;
  - Niveau A88 : divisions ;
  - Niveau A272 : groupes ;
  - Niveau A615 : classes ;
  - Niveau A732 : sous-classes
- le second présente les résultats selon le **croisement d'un niveau de nomenclature avec quatre tranches d'effectifs salariés** (**mesurées en ETP**) :
  - de 0 à 9 salariés ;
  - de 10 à 19 salariés ;
  - de 20 à 249 salariés ;

[372] On peut se référer au schéma du § *2.7.3.3* pour une meilleure compréhension de ce concept.

- plus de 250 salariés.

Les niveaux de nomenclature considérés vont de A10 (nomenclature synthétique en 11 postes) à A 272 (groupes). On ne va pas en deçà du groupe de la Naf quand on croise avec les tranches de taille.

Comme on l'a rappelé dans l'*Avant propos* et dans les réglages des contrôles du *chapitre 2*, ces deux types de tableaux présentent chacun l'un des niveaux le plus fin dont les agrégats correspondants sont contrôlés et validés par les gestionnaires d'Ésane et donc les plus fins que l'on peut publier : la sous-classe et le croisement groupe x tranches de taille.

Si l'on souhaite des résultats concernant un niveau de nomenclature donné, on a donc intérêt à « filtrer » ce niveau en utilisant le filtre d'Excel. Tous les tableaux fournis comprennent la variable **NIVEAU** (qui permet de filtrer un niveau donné) ainsi que le **code** de chaque secteur, son **intitulé en clair** et le **nombre d'entreprises** du secteur (en plus des variables indiquées ci-dessous).

1) Principales caractéristiques[373]

En plus des deux tableaux précédents, cette rubrique commence par un **troisième tableau** extrêmement **synthétique en nomenclature dite A10, et comprenant un onglet par année** (depuis 2012) : 10 postes plus bien entendu le niveau « tous secteurs ». On y a ajouté une 12e ligne spécifique qui distingue l'artisanat commercial[374] au sein de l'industrie manufacturière de façon à pouvoir la regrouper avec les activités commerciales. Ces données sont utiles pour tout utilisateur désirant obtenir des données de cadrage de premier niveau par grand secteur du système productif, sur **six variables** :

- les effectifs salariés en équivalents temps pleins (ETP)[375] ;
- le chiffre d'affaires hors taxes ;
- la valeur ajoutée (y compris autres produits et autres charges) ;
- les frais de personnel ;
- l'excédent brut d'exploitation ;
- les investissements corporels bruts (hors apports).

Les deux autres tableaux (niveau sous-classe et groupe x taille) comprennent six variables supplémentaires :

- les effectifs salariés au 31 décembre ; (dont on rappelle, Cf. § *7.1.2*, qu'en dépit de l'intitulé il s'agit d'une **estimation de l'emploi au sens de la liasse fiscale**, moyenne des effectifs en fin de trimestre) ;
- les effectifs occupés (qui comprennent en plus les non salariés) ;
- le chiffre d'affaires à l'export ;
- la capacité d'autofinancement ;
- le résultat courant avant impôts ;
- le résultat net comptable.

2) Bilan (actif brut, amortissements, passif)[376]

Les deux tableaux fournis nécessitent une bonne connaissance de la comptabilité d'entreprises pour en tirer profit. Ils comprennent chacun 3 onglets « actif brut » actif amorti » et « passif » qui contiennent respectivement **23, 18 et 18 variables** qu'il serait un peu fastidieux de lister ici.

3) Compte de résultats[377]

Les deux tableaux ne comportent qu'un seul onglet qui contient **44 variables comptables** qui nécessitent là encore une bonne connaissance de la comptabilité d'entreprises.

4) Emploi[378]

Les deux tableaux comprennent **20 variables** réparties en six ensembles :

- les effectifs (au 31/12[379], en ETP, prêté) ;

[373] *https://www.insee.fr/fr/statistiques/2384021?sommaire=2384048* pour les résultats de l'année 2014.

[374] L'artisanat commercial regroupe les entreprises ayant pour activité la charcuterie, la boulangerie, la boulangerie-pâtisserie, la pâtisserie, ou la cuisson de produits de boulangerie. Stricto sensu, il s'agit d'activités de transformation donc industrielles ; mais ces activités sont tellement inséparables de la vente de ces mêmes produits une fois fabriqués que la plupart des utilisateurs préfèrent les regrouper avec le commerce.

[375] Pour une définition précise, voir *l'annexe 2*.

[376] *https://www.insee.fr/fr/statistiques/2384024?sommaire=2384048* pour les résultats de l'année 2014.

[377] *https://www.insee.fr/fr/statistiques/2384027?sommaire=2384048* pour les résultats de l'année 2014.

[378] *https://www.insee.fr/fr/statistiques/2384030?sommaire=2384048* pour les résultats de l'année 2014.

[379] Même remarque que ci-dessus ; il s'agit en fait de la moyenne des effectifs de fin de trimestre et non d'un véritable emploi au 31/12 ; Cf. § *7.1.2*.

- la ventilation des effectifs ETP par temps de travail (temps plein, temps partiel) ;
- la ventilation des effectifs ETP par CSP (plus apprentis et emplois aidés) ;
- le personnel non salarié de l'entreprise, personnel intérimaire et personnel prêté à l'entreprise ;
- la part des entreprises employant du personnel intérimaire ;
- les frais de personnels (distinction salaires et charges patronales) et le coût du personnel extérieur.

5) Investissement[380]

Les deux tableaux fournissent 13 variables regroupées en trois ensembles :

- les investissements incorporels et financiers (bruts y. c. apports) ;
- les investissements corporels avec la ventilation selon le type (terrains, constructions, installations, matériels de transport etc.) ;
- les investissements corporels hors apports (bruts et nets des amortissements).

6) Ratios comptables[381]

Ésane permet de calculer, entre autres, 14 ratios comptables (regroupés en organisation et débouchés de la production / partage de la VA / ratios de rentabilité / Ratios d'endettement, de solvabilité et de liquidité / délais de paiement / ratios d'investissements) et dont on donne ici la définition[382] :

- le taux d'exportation [= Chiffre d'affaires à l'export / Chiffre d'affaires] ;
- le taux de valeur ajoutée [= Valeur ajoutée HT / Chiffre d'affaires ; mesure le taux d'intégration de la société dans le processus de production ainsi que le poids des charges externes] ;
- la productivité apparente du travail par tête ; [= Chiffre d'affaires / Effectif salarié ETP ; ne tient pas compte des autres facteurs de production (capital, consommations intermédiaires)] ;
- l'intensité capitalistique [= Immobilisation corporelles / Effectif salarié au 31/12] ;
- la part de frais de personnels ;[= Frais de personnels / Valeur ajoutée brute au coût des facteurs (VABCF)]
- le taux de marge [= Excédent brut d'exploitation / Chiffre d'affaires] ;
- la rentabilité économique [= Excédent brut d'exploitation / (Immobilisations corporelles + Incorporelles + Besoin en fond de roulement)] ;
- la rentabilité financière [= Résultat net comptable / Capitaux propres] ;
- le levier financier [= Emprunts et dettes assimilées / Capitaux propres et autres fonds propres] ;
- le taux de prélèvement financier [= Intérêts et charges assimilées / Excédent brut d'exploitation] ;
- les délais de paiements des fournisseurs en jours [= Total des dettes fournisseurs (sur l'ensemble du secteur y.c. comptes rattachés) / (Total des achats et charges externes annuels TTC / 360)] ;
- les délais des paiements des clients (sous réserves) en jours [= Total des créances clients / (Total des CA annuels TTC/360)] ;
- le taux d'investissement [= Investissements corporels (hors apports éventuels d'autres sociétés) / Valeur ajoutée HT] ;
- le taux d'auto financement [= capacité d'autofinancement / investissements corporels bruts hors apports].

9.4.3.2.B. Les branches d'entreprises

Les deux tableaux correspondants donnent les chiffres d'affaires **en branche** et non en secteur. Ils ne sont donc pas comparables aux autres tableaux et résultats.

Le premier donne, pour chaque sous-classe, le montant du chiffre d'affaires total de la « branche » correspondant à la sous-classe ; l'autre regroupe les branches au niveau groupe de la nomenclature mais les ventile selon les quatre classes de taille des entreprises qu'on a vues plus haut.

Rappelons qu'un « secteur » S est formé de toutes les unités légales ayant l'activité principale S alors qu'une « branche d'activité B » décrit de façon conceptuelle une « activité » B, que celle-ci soit réalisée à titre principal ou à titre secondaire par des entreprises. Participent ainsi à une branche B toutes les entreprises qui fabriquent des produits (ou rendent des services) correspondant à la sous-

[380] *https://www.insee.fr/fr/statistiques/2384033?sommaire=2384048* pour les résultats de l'année 2014.
[381] *https://www.insee.fr/fr/statistiques/2384036?sommaire=2384048* pour les résultats de l'année 2014.
[382] S'agissant de ratios de la forme A/B se rapportant à N entreprises, ces ratios sont calculés comme $\Sigma_N A / \Sigma_N B$.

classe B, quelle que soit la part, même très faible, de ces produits (ou services) dans leur chiffre d'affaires. La disponibilité d'un produit correspond à la production de la branche et non à celle du secteur correspondant.

Les données publiées ici sont des données de statistiques d'entreprises pures. Elles sont fondées sur des déclarations de chiffres d'affaires donc sur un **concept de production vendue**. Elles se distinguent donc **du concept de branche en comptabilité nationale**, plus complet, qui correspond au concept de production tout court et intègrent donc diverses corrections, dont en particulier **une estimation de la variation des stocks** qui n'est évidemment pas incluse ici. C'est ce qui explique que les chiffres d'affaires en branches publiés par la comptabilité nationale et par Ésane ne sont **pas égaux.**

9.4.3.2.C. Alisse[383]

Alisse est probablement le produit de diffusion le plus sophistiqué, **à destination des experts** sur les statistiques des entreprises ; il n'est **d'ailleurs pas limité aux résultats d'Ésane**. Il ne s'agit plus pour l'utilisateur de consulter des tableaux tout faits mais d'accéder à une véritable **base de données multi-sources** et de **confectionner ses tableaux sur mesure** à partir de toutes les variables disponibles **d'Ésane ou de six autres sources**[384].

Comme Ésane, Alisse a basculé en entreprises depuis 2012 et les **données antérieures à 2012, en unités légales,** ne sont donc **plus comparables avec les données en entreprises.** Il y a donc aujourd'hui deux bases de données :

- l'une **en unités légales**[385] qui fournit des séries « mortes » (elles ne sont plus mises à jour) sur les unités légales de 2008 à 2013 ;
- l'autre **en entreprises**[386], qui fournit des données qui continuent d'être mises à jour et qui démarrent en 2012.

Le champ d'Alisse est constitué des **secteurs d'entreprises principalement marchands, non agricoles et non financiers**. Le champ d'Ésane n'est donc **pas intégralement disponible dans Alisse** :

Les secteurs couverts par Alisse ne comprennent ainsi ni les secteurs 02.20Z et 64.20Z, ni la division 66 (activités auxiliaires de services financiers et d'assurance) ni les sections P (enseignement) et Q (santé humaine et action sociale).

Les données Ésane proposées sont centrées sur la comptabilité d'entreprise et nécessitent une certaine familiarité a priori avec ce type de variables puisque l'utilisateur devra sélectionner celles qui l'intéressent, présentées souvent sous forme de ratios, au choix sur **1 ou 2 années.** Il peut soit créer en ligne le tableau qu'il souhaite, soit télécharger les bases disponibles.

S'agissant d'Ésane, **plus de 200 variables**[387] sont proposées selon les thèmes suivants concernant les caractéristiques des entreprises :

- informations fiscales ;
- informations sur l'emploi ;
- principaux ratios d'analyse financière.

Les niveaux de nomenclature d'activité sont paramétrables et le niveau le plus fin est identique à ceux des données détaillées (sous-classes, groupes, divisions, sections et le niveau dit A10).

9.4.3.2.D. Données N-1 pour comparaison

Comme on l'a rappelé plus haut et expliqué aux § *4.7* et *4.8*, ce n'est qu'à compter de la campagne 2012 qu'on a publié à la fois en entreprises et sur des champs et avec une méthodologie comparable par maillons successifs de deux années consécutives. En même temps que les résultats de l'année N, on peut trouver, calculés sur le même champ et avec une méthodologie identique les mêmes 14

[383] Alisse : Accès en Ligne aux Statistiques Structurelles d'Entreprises.
[384] En plus d'Ésane, Alisse regroupe les données de six autres sources : les déclarations annuelles de données sociales (DADS), l'enquête sur les liaisons financières (dite Lifi), les déclarations mensuelles de mouvements de main d'œuvre (enquête dite DMMO de la Dares), l'enquête sur la recherche et le développement (du ministère de l'enseignement supérieur et de la recherche), les déclarations de TVA (déclarations dites CA3 de la DGFiP) et enfin les données du commerce extérieur (de la DGDDI).
[385] *http://www.alisse2.insee.fr/SelectionMesureT1.jsp?item=ACTENT*
[386] *http://www.alisse2.insee.fr/SelectionMesureT1.jsp?item=ACTEND*
[387] Il y a donc des variables qui ne sont accessibles que dans Alisse.

tableaux que ceux qui viennent d'être décrits plus haut. Ils sont simplement regroupés dans un Zip qui permet d'accéder à tous les tableaux à la fois.

9.4.3.2.E. Résultats sectoriels publiés dans Insee Référence

Ésane est la **principale source qui alimente** chaque année la partie « fiche » du numéro des « entreprises en France » dans la collection « Insee Références »[388]. L'édition de l'année N décrit la situation de l'année N-2. Qu'il s'agisse de fiches « thématiques » ou de fiches « sectorielles », elles sont constituées d'une page de texte et d'une page de graphiques et tableaux. Elles visent à présenter une **synthèse** de la situation structurelle du tissu économique français, sur le thème en question en présentant **les chiffres clés** sur le thème ou le secteur concerné, **sur le champ des secteurs « principalement marchands, non agricoles et non financiers** ». Ce champ est le même que celui d'Alisse (qu'on a vu au § *9.4.3.3*) et que celui de la publication Eurostat ci-dessous (§ 9.4.5).

Leur nombre varie légèrement d'une année sur l'autre mais il y a une **trentaine de fiches thématiques** regroupées en quatre grands domaines (structure du système productif / activité emploi et coûts salariaux / mondialisation compétitivité et innovation / énergie et développement durable) et une **quinzaine de fiches sectorielles** différenciant les cinq grands secteurs d'Ésane (industrie, construction, commerce, transports et entreposage, services marchands) plus l'ensemble des secteurs.

Comme tous les numéros de cette collection, « les entreprises en France » est accessible sous forme papier ou téléchargeable en format PDF sur le site de l'Insee. Chaque fiche est également consultable ou téléchargeable séparément.

9.4.4 La création du fichier de données individuelles

Comme on l'a mentionné plus haut au § *9.3.2*, on fabrique également un fichier de données individuelles, le « **Fichier Approché des Résultat d'Ésane (FARE)** » qui est mis à disposition, au travers du centre d'accès sécurisé distant (CASD), de tout utilisateur externe (membre de la statistique publique ou chercheur) dont la demande a été agréée par le comité du Secret.

Ce (gros) fichier contient le résultat de l'ensemble des entreprises du champ Ésane dont on a reçu des liasses fiscales ou qui ont été imputées actives. On a mentionné au § 9.3.2 les précautions à prendre dans l'utilisation de ce fichier dont le contenu précis est donné aux utilisateurs qui ont obtenu l'autorisation d'accès.

Le fichier de l'année 2009 contenait ainsi par exemple 2 889 707 unités légales (lignes) et 178 (variables). À compter des résultats de l'année 2013, le fichier contient **à la fois** les données concernant les unités légales (UL) et les entreprises profilées issues du profilage des groupes (Cf. *chapitre 11*) (y compris les UL qui les composent) Une variable indicatrice permet alors aux utilisateurs de choisir de ne travailler qu'en UL ou qu'en "entreprises" de façon à éviter les doubles-comptes.

9.4.5 La publication Eurostat

Comme on l'a mentionné à maintes reprises, le système Ésane sert également de fondement aux statistiques dites « SSE[389] » publiées par Eurostat sur son site[390] conformément au règlement SBS. Il est toutefois important de savoir que **leur champ n'est pas le même** que celui des données publiées sur le site de l'Insee. Les SSE européennes ne couvrent que l'industrie, la construction, le commerce et les services, c'est-à-dire ce qu'on appelle **les secteurs principalement marchands, non agricoles et non financiers.** Par rapport à ce que nous avons décrit au § *9.4.3* et au champ Ésane décrit au § *4.4.1*, **les parties manquantes correspondent :**

- au secteur 02.20Z (exploitation forestière) qui est en agriculture ;
- à la sous-classe 64.20Z (holdings) et la division 66 (activités auxiliaires de services financiers et d'assurance) qui est dans la section K des activités financières et d'assurance ;
- aux parties marchandes des sections P (enseignement) et Q (santé humaine et action sociale).

[388] *https://www.insee.fr/fr/statistiques/2497179* pour l'édition 2016.
[389] Statistiques Structurelles sur les Entreprises.
[390] *http://epp.eurostat.ec.europa.eu/portal/page/portal/european_business/introduction*

Les totaux France entière ne sont donc absolument **pas comparables avec ceux des tableaux du § 9.4.3**. Sur les mêmes champs, les chiffres sont évidemment les mêmes que les chiffres publiés par l'Insee.

En revanche, comme on l'a mentionné plus haut, **ce champ est aussi celui utilisé** dans l'Insee Références « **les entreprises en France** » publié régulièrement tous les ans depuis 2012.

9.4.6 La publication de l'EAP[391]

Comme on l'a mentionné dans le chapitre 5 (Cf. § 5.2), l'enquête annuelle de production dans l'industrie sert à mesurer la production industrielle (vendue) et par là-même à répondre au règlement européen **Prodcom**[392] (selon la nomenclature du même nom dite « Prodcom ») ; elle bénéficie aussi du coup d'une publication directe particulière sur le site de l'Insee puisque les pays publient généralement aussi, sur leur propre site, les éléments qu'ils envoient à la Commission. Les délais de publication sont plus courts que ceux d'Ésane dans son ensemble : les données de l'année N sont en principe mises en ligne en **décembre N+1**.

De même que pour les publications Ésane, les deux tableaux fournis permettent directement de comparer les données de **l'année N avec celles de l'année N-1**. En revanche, dans le cas de l'EAP, s'agissant de donner une estimation de la production française quelquefois même en unités physiques, il n'y a **pas de « consolidation** » qui soit faite ; en d'autres termes, les productions intermédiaires ou intégrées sont comptabilisées dans la « production » et l'unité statistique de base reste donc bien **l'unité légale**[393].

Les deux tableaux fournis croisent comme toujours un niveau de nomenclature avec un certain nombre de variables ; mais dans le cas de l'EAP, il s'agit de **nomenclatures de « produits »** et non d'activités.

- Le premier tableau dit en « CPF 4-5 » fournit 2 variables : **le nombre d'unités légales** qui produisent le produit répertorié dans la ligne considérée et **les facturations correspondantes**. Un « produit » (une ligne) de la CPF – 4 désigne en général le résultat de l'activité de la classe NAF désignée par le même code[394] ; dans le cas où la NAF comprend plusieurs sous-classes dans la même classe, il faut 5 caractères pour identifier les produits de la sous-classe. On parle alors de CPF-5. D'où le nom de CPF 4-5 pour la résultante. Les chiffres sont fournis pour l'année N et l'année N-1. Bien entendu, les unités légales qui produisent les produits ainsi répertoriés peuvent le faire aussi bien au titre de leur activité principale que secondaire.
- Le second tableau dit au niveau « Prodfra[395] » est à la fois plus complexe et beaucoup plus détaillé ; il comprend – lorsque cela a du sens pour le produit considéré - 2 variables supplémentaires : le volume de la production en **quantités physiques vendues** et **l'unité de volume considérée** (litres, kilos, tonnes, mètres cubes, ou nombre de pièces, etc.). La nomenclature Prodfra est une nomenclature de produits extrêmement fine (dont le code est à 10 positions), incluse dans Prodcom. Comme Prodcom, elle varie à la marge tous les ans. Elle compte encore plus de **4200 positions en 2016**. Elle comprend aussi ce qu'on appelle des « **services industriels** » correspondant aux activités de sous-traitance, de réparation, d'installation etc. qui sont repérés dans le tableau Dans ces cas, la facturation garde du sens, mais la variable « quantité » n'en a pas et n'est pas remplie. Bien entendu, à un tel niveau de finesse, un certain nombre de lignes sont sous secret statistique.

La variable « facturation » de l'EAP correspond bien évidemment à la variable « chiffre d'affaires en branche » d'Ésane lorsqu'on considère le même niveau de nomenclature en classe ou sous-classe. Elle en diffère cependant numériquement (quelquefois même de 5 à 10 %, voire plus dans des cas extrêmes) pour deux raisons principales

[391] https://www.insee.fr/fr/statistiques/2503588 par exemple pour les résultats de l'année 2015.
[392] http://ec.europa.eu/eurostat/web/prodcom/data/excel-files-nace-rev.2?p_l_id=120466&p_v_l_s_g_id=0
[393] Plus précisément, le futur règlement cadre européen FRIBS (qui devrait être adopté vers 2019 ; il est présenté au § 11.2) devrait préciser que l'unité statistique sous-jacente est « l'unité d'activité économique » (UAE) (en anglais KAU Kind of Activity Unit) c'est-à-dire la fraction d'entreprise ayant une activité économique de niveau classe de la Naf, pour laquelle on dispose d'un compte d'exploitation. Pour la France, l'unité légale en est une excellente approximation.
[394] On dit que la nomenclature d'activités (NAF) et la nomenclature de produits (CPF) sont en correspondance ; par exemple la NAF 2052 est la fabrication de colles et la CPF 2052 est le produit « colles ».
[395] Pour production française par opposition à « Prodcom » « production communautaire ».

- Il n'y a pas de "consolidation" dans l'EAP : le chiffre d'affaires correspondant à une production intermédiaire au sein d'un groupe sera pris en compte dans l'EAP alors qu'il disparaîtra dans le chiffre d'affaire (et donc la branche) consolidé du groupe.
- Un agrégat statistique est toujours fonction de la façon dont on estime les non-réponses. Or, chacune des deux chaînes de production a son propre mode de redressement de la non-réponse : dans un cas on raisonne en "secteurs", donc un chiffre manquant sera estimé par l'évolution des entreprises du secteur d'appartenance de l'entreprise ; dans l'autre cas on raisonne en "produits", donc un chiffre manquant sera estimé par l'évolution des unités légales qui fabriquent le même produit.

Le mode de calcul des agrégats est également différent (estimateur composite dans le cas d'Ésane). A l'arrivée, les résultats sont évidemment différents. C'est donc un cas où il a été malheureusement impossible d'atteindre l'un des objectifs forts de Resane au lancement du projet qui était de ne plus publier qu'un seul chiffre par concept, Cf. *Avant-propos*.

9.5 La pose du secret

9.5.1 Problématique

La règle de base pour pouvoir publier un agrégat (une « case » d'un des tableaux qu'on a vus au § 9.4.3), c'est que la connaissance de cet agrégat ne doit pas permettre de « deviner » la réponse individuelle d'une entreprise, sauf si celle-ci a donné son accord pour rendre publique cette information. Lorsque ces conditions ne sont pas remplies, les statisticiens mettent la case en question « sous secret statistique ».

La gestion du secret statistique requiert donc le comptage du nombre d'entreprises qui participent à un agrégat et le calcul de la contribution de chaque unité statistique à l'agrégat, ceci indépendamment de l'outil de traitement du secret que l'on choisit. Le logiciel préconisé par Eurostat, τ-Argus, que l'Insee a choisi d'utiliser[396] ne déroge pas à ces besoins. Celui-ci doit donc disposer en entrée de données individuelles lui permettant de calculer le nombre d'entreprises qui participent aux agrégats que l'on souhaite diffuser et leurs contributions individuelles, afin de poser le **secret dit « primaire »** selon les règles classiques : **règle des trois unités** (trois unités statistiques au moins doivent intervenir dans le calcul d'une case de façon que la connaissance de l'une d'entre elles ne puisse pas permettre de reconstituer la valeur de l'autre, s'il n'y en avait que deux), et **règle de dominance des 85%** (aucune unité ne doit représenter à elle seule 85 % du total d'une case de façon là encore que la connaissance de la case ne suffise pas à « deviner » la valeur de la « grosse » unité ; on retient empiriquement un seuil à 85 %). Le logiciel calcule alors le secret dit « **secondaire** » en s'appuyant sur l'arbre d'agrégation des nomenclatures de façon à assurer que si une case d'un niveau N est sous secret, on ne peut pas la recalculer très simplement par différence, en partant du total de niveau N+1 auquel elle participe, si elle est la seule dans son cas. Lorsque c'est le cas, le logiciel met sous secret une case supplémentaire ; c'est ce que l'on appelle le secret secondaire (le plus complexe à calculer car il faut choisir astucieusement la case à masquer).

Le problème de la pose du secret dans Ésane vient du fait que les estimateurs par différence **ne sont pas construits à partir d'une matrice de données individuelles rectangulaire**, qui est l'intrant normal de τ-Argus. L'estimateur du § *9.2.1* se décompose, comme on l'a vu, en sommes sur des populations différentes et ne s'appuie pas sur un fichier de données individuelles[397] grâce auquel le logiciel τ-Argus pourrait calculer le nombre d'entreprises et la part de celles-ci dans l'agrégat.

Pour résoudre ce problème, on a recours à l'utilisation **d'une variable « ombre** » sur laquelle on calculera un estimateur « classique » : le but est de calculer une structure de secret sur cette nouvelle variable pour laquelle nous disposons de suffisamment d'information pour poser le secret à partir du logiciel et qui soit très proche de l'estimateur composite retenu dans Ésane. La structure de secret des tableaux que l'on souhaite publier sera alors calquée sur celle des tableaux élaborés avec l'estimateur « ombre ». Cela ne sera pas rigoureusement exact et pourrait dans certains cas amener à des erreurs (par exemple lorsque le nombre d'entreprises d'un secteur selon l'estimateur classique est très

[396] C'est le département de la méthodologie statistique qui se charge de cette opération de détermination des agrégats sous secret à l'aide de ce logiciel (se prononce tau-argus du nom de la lettre grecque τ).

[397] Si on voulait exprimer la formule (1) des estimateurs composites (vue plus haut) sous forme de la somme d'une variable individuelle, celle-ci devrait être calculée pour chaque agrégat et chaque secteur et dépendrait pour chaque entreprise des secteurs au lancement et observés dans l'Esa, ainsi que de la réconciliation. Ceci implique de créer autant de fichiers avec l'ensemble des variables fiscales que de secteurs de diffusion. La pose du secret primaire sur chacun de ces fichiers serait alors possible, mais pas celle du secret secondaire du fait de l'absence de liaison entre les fichiers.

différent de celui obtenu selon l'estimateur composite). Ces cas devraient alors être examinés individuellement.

Nous proposons donc de traiter la confidentialité selon un **estimateur plus simple** permettant de disposer des données individuelles nécessaires à l'application des règles classiques de confidentialité. La structure de secret qui sera obtenue sur cet estimateur sera appliquée comme un calque sur les tableaux construits à partir de l'estimateur composite.

9.5.2 L'estimateur utilisé pour poser le secret pour les variables fiscales et pour un secteur X donné

Cet estimateur se décompose en deux sommes sur des populations disjointes. D'une part, il y a les répondants de l'échantillon tiré dans le sous champ Ésane 1 (la première somme), et d'autre part les entreprises des champs 2 à 5.

$$\hat{Y}_{\text{Secret}}(X)=\underbrace{\sum_{i\in R} w_i Y_i^{\text{Re}di} 1\text{I}_{i\in\text{champ}} 1\text{I}_{\text{APEenq}=X}(i)}_{1}+\underbrace{\sum_{i\in \tilde{U}_2\cup\tilde{U}_3} Y_i^{IEG} 1\text{I}_{\text{APErep}=X}(i)}_{2}$$

[*Formule (7)*] :

Rappel :
1) Comme pour les estimateurs du § *9.2*, le poids w_i de la formule ci-dessus et dans tout ce § 9.5 est le poids final après traitement de la non-réponse totale, winsorisation et calage (cf. § *5.6.2*).
2) On a $Y_i^{\text{Re}di}$ = Y_i^{IEG} pour toute variable Y qui n'est présente que dans la source fiscale et sur laquelle le processus Rédi n'a aucun effet.

La première partie de cet estimateur est simplement l'estimateur calé de la variable Y sur l'échantillon. Quant à la seconde partie, on la retrouve à l'identique dans l'estimateur composite du § *9.2.1* (partie ❷). Nous avons donc un estimateur qui possède la **même espérance** que l'estimateur composite. Il possède par contre l'avantage de permettre de disposer des données individuelles nécessaires au traitement de la confidentialité.
La variable « ombre » est définie par $Y^{\text{Re}di}$ sur l'échantillon et Y^{IEG} sur les sous champs 2 à 5.

Nous avons déjà évoqué le fait que d'utiliser pour la pose du secret un estimateur différent de celui qui sera ventilé dans les tableaux pouvait amener des erreurs. On peut toutefois faire le raisonnement que tant que l'on ne descend pas à des niveaux de diffusion trop fins (on rappelle que l'on ne diffusera pas par exemple de croisement sous-classes∗taille), les secteurs soumis au secret sont le plus souvent des secteurs très concentrés, type monopole. Il s'agit de grandes entreprises pour lesquelles l'information Esa collectée est quasi complète, voire exhaustive, et la gestion de la confidentialité n'est alors pas perturbée par l'utilisation d'un estimateur différent.

9.5.3 L'estimateur utilisé pour poser le secret des variables fiscales ventilées selon une variable sectorielle de l'Esa

On s'intéresse tout d'abord au cas le plus classique des branches d'activités. On cherche, par exemple, à estimer le chiffre d'affaires total d'une branche b donnée, pour les entreprises du sous champ Ésane 1 appartenant à un secteur X donné, c'est-à-dire la quantité :

$$\sum_{i\in U_1} \text{CA}(\text{b,i})' \quad 1\text{I}_{i\in\text{champ}} 1\text{I}_{\text{APE}=X}(i)$$

L'estimateur composite retenu dans Ésane et donné par la formule (5) utilise une clef de répartition du chiffre d'affaires du secteur X en branche estimée à partir de l'Esa.
Cette clé de répartition (*) provient de l'Esa, et porte donc uniquement sur le sous champ Ésane 1. Les statistiques sectorielles ainsi ventilées sont par conséquent celles relatives au seul sous champ Ésane 1.

$$clé(b,X) = \frac{\sum_{i \in R} \mathrm{w}_i CA^{ESA}(b,i)\, 1\!\!\mathrm{I}_{\mathrm{APEenq=X}}(\mathrm{i})}{\sum_{i \in R} \mathrm{w}_i CA^{ESA}(i)\, 1\!\!\mathrm{I}_{\mathrm{APEenq=X}}(\mathrm{i})} \; (*)$$

Comme précédemment la présence de l'estimateur par différence ne permet pas une gestion directe de la confidentialité dans τ-Argus.

Par contre, puisque les statistiques sectorielles ainsi ventilées sont celles relatives au seul sous champ Ésane 1, nous disposons de données individuelles en branches pour toutes les entreprises. On utilisera donc à nouveau l'estimateur calé suivant pour estimer le chiffre d'affaires total de la branche b pour les entreprises du secteur X (estimateur qui reprend la forme de l'estimateur choisi pour le traitement de la confidentialité de variables fiscales dans un secteur donné) :

$$\boxed{\hat{Y}_{\mathrm{Secret}}(X) = \sum_{i \in R} w_i CA^{ESA}(b,i)\, 1\mathrm{I}_{\mathrm{APEenq}=X}(i)}$$ [*Formule (8)*]

Cet estimateur possède une espérance proche de celle de l'estimateur composite retenu dans Ésane, tout en ayant une forme simple permettant une gestion de la confidentialité par τ-Argus. Il sera donc **utilisé pour « confidentialiser »** un tableau ventilant l'estimateur composite du chiffre d'affaires total de la branche b pour les entreprises du secteur X.

Cette technique est généralisable à toute statistique résultant de la ventilation d'une estimation sectorielle, obtenue sur une variable comptable, selon une clé de répartition issue de l'Esa.

9.5.4 L'estimateur utilisé pour poser le secret pour une variable uniquement présente dans l'Esa

Une statistique s'appuyant sur une variable uniquement présente dans l'Esa ne porte par construction que sur le sous champ Ésane 1.
Pour une variable Y donnée, l'estimateur pondéré classique du total de cette variable sur le secteur X est :

$$\hat{Y}_{HT}^{Ech}(X) = \sum_{i \in R} w_i Y_i^{\mathrm{ESA}}\, 1\mathrm{I}_{\mathrm{APEenq}=X}(i) \; .$$

Cependant, l'estimateur final utilisé, donné par la formule (6) du § *9.2.6*, est obtenu en faisant appel à une variable auxiliaire A qui varie selon la nature de la variable Y considérée, afin de respecter les structures observées dans l'enquête :

Ces ajustements effectués sur les estimateurs ne sont en fait que des homothéties réalisées sur les agrégats et non sur les données individuelles. Ceci ne perturbe donc pas la pose du secret.
Ainsi, sur les variables uniquement présentes dans l'Esa, nous prendrons :

$$\boxed{\hat{Y}_{\mathrm{Secret}}(X) = \hat{Y}_{HT}(X)}$$ [*Formule (9)*]

Le calcul de cet estimateur ne pose pas de problème. L'estimateur utilisé est un estimateur classique par calage. Pour ces variables, l'ensemble des données nécessaires à la pose du secret **est parfaitement connu**. On peut en effet compter facilement le nombre d'entreprises qui participent à un agrégat (cellule d'un tableau) et les contributions de celles-ci à l'agrégat.

9.6 Les précautions à prendre dans l'analyse des résultats en évolution [2.7],

9.6.1 Présentation des différents types d'évolutions possibles

Toutes les formules [(1) à (6)] qui ont été données dans le § *9.2* se réfèrent à des agrégats **en niveau**. Ceux-ci permettent de connaître la valeur des caractéristiques (le nombre d'entreprises, le chiffre d'affaires, etc.) ou de ratios comptables (chiffre d'affaires par salarié, taux de marge brute, investissement/VA, etc.) pour une année donnée et pour les différents secteurs d'activité de

l'économie française. De fait Ésane fournit essentiellement des renseignements sur la « structure » de l'appareil productif.

Il n'en demeure pas moins que beaucoup d'utilisateurs s'intéressent non pas à l'analyse statique du tissu productif mais à sa déformation ou son évolution dans le temps. Ils doivent alors s'intéresser à des agrégats **en évolution** qui permettent de connaître l'évolution économique d'un secteur pour deux années successives.

Cependant l'explication ou l'interprétation d'une évolution **n'est pas chose aisée**. Elle peut masquer des réalités économiques diverses.

> Exemple : d'après les résultats Ésane, entre 2008 et 2009, le chiffre d'affaires du secteur « Stockage et Entreposage » (521) a **baissé de 0,7 %.** Dans le même temps, le chiffre d'affaires des entreprises pérennes (présentes dans le secteur en 2008 et 2009) a **progressé de 2,8 %.** Les entreprises pérennes ne représentent toutefois que les trois quarts du chiffre d'affaires du secteur, et génèrent donc 2,1 points dans l'évolution sectorielle. A l'inverse, les mouvements de champ pèsent pour - 3,5 points (- 8 points pour les entreprises qui ont changé de secteur, et 4,5 points pour celles qui sont entrées dans le secteur). Ainsi l'évolution observée indique plus **une modification de classement sectoriel qu'une baisse de l'activité des entreprises qui composent ce secteur**. A l'inverse, dans le secteur « Construction de bâtiments résidentiels et non résidentiels » (412) qui a connu une baisse de 18 %, ce sont bien les entreprises pérennes du secteur qui ont vu leur activité diminuer (-18,4 %).

Le classement sectoriel des entreprises étant connu par le biais d'enquêtes, l'analyse des résultats en évolution s'appuiera nécessairement sur les résultats observés sur l'échantillon de l'enquête structurelle. C'est sur ce point que résident les principales difficultés méthodologiques du calcul des résultats en évolution que nous présenterons par la suite.

9.6.1.1 L'évolution apparente

Il s'agit d'un simple rapport des agrégats en niveau N / N-1.
La formule de cette évolution est la suivante :

$$\Delta_{N/N-1} = \frac{\left(X_N^s - X_{N-1}^s\right)}{X_{N-1}^s}$$ où X_N^S est l'agrégat de la variable X sur le secteur S et pour l'année N.

9.6.1.2 L'évolution économique « en comparable »

L'objectif de l'évolution en comparable est de neutraliser **les effets des restructurations** en prenant en compte les enveloppes de restructuration (cf. *chapitre 3* de ce document), c'est-à-dire en prenant en compte des évolutions à champ sectoriel constant donc comparable.

Les entreprises prises en compte pour le calcul de l'agrégat N sur le secteur S en comparable sont :

- les entreprises actives du secteur S en N qui n'appartiennent pas à une enveloppe de restructuration N ;
- les entreprises actives qui appartiennent à une enveloppe de restructuration, enveloppe classée dans le secteur S en N.

Les entreprises prises en compte pour le calcul de l'agrégat N-1 sur secteur S en comparable sont :

- les entreprises actives du secteur S en N-1 qui n'appartiennent pas à une enveloppe de restructuration en N ;
- les entreprises actives en N-1 qui appartiennent à une enveloppe de restructuration, enveloppe classée dans le secteur S en N (l'information sur ces entreprises est consolidée sur le contour de l'enveloppe et est donc différente de la somme des entreprises pour les caractéristiques non additives. La somme obtenue sur les unités légales est pondérée par le coefficient d'agrégation ; Cf. § *3.1.2* et *3.2.2.2*).

L'exemple suivant permet d'illustrer la distinction entre l'évolution apparente et l'évolution en comparable.

> Imaginons que les entreprises 2 et 3 fusionnent entre N-1 et N pour donner l'entreprise 5 en N (le raisonnement serait strictement identique si l'entreprise 3 absorbait l'entreprise 2). Les entreprises 2 et 3 étaient respectivement classées dans les secteurs A et B en N-1 ; l'entreprise 5 est classée en B en N.

Entreprise	Secteur (N-1)	Secteur N	Chiffre d'affaires N-1	Chiffre d'affaires N
Ent. 1	A	A	1000	1100
Ent. 2	A		2000	
Ent. 3	B		500	
Ent. 4	B	B	1500	1600
Ent. 5		B		3000

On est amené à créer ainsi une enveloppe de restructuration qui aboutit au tableau suivant :

Entreprise	Secteur N-1	Secteur N	Chiffre d'affaires N-1	Chiffre d'affaires N
Ent. 1	A	A	1000	1100
Enveloppe 1	B	B	2500	3000
Ent. 4	B	B	1500	1600

L'évolution **apparente** des secteurs A et B donne :

Secteur A = (1100-3000)/3000= - 66%
Secteur B = (4600-2000)/2000= + 130%

On retrouve le fait bien connu qu'en cas de restructuration, les variations sectorielles des secteurs concernés sont fortement affectées alors **qu'en comparable** on obtient :

Secteur A = (1100-1000)/1000=10%
Secteur B = (4600-4000)/4000=+ 15%

On retrouve des évolutions plus « normales » car on travaille à champ comparable entre N et N-1.

L'Insee **ne publie pas pour l'instant** ces évolutions en comparable mais cela reste envisagé pour l'avenir.

9.6.2 La décomposition de l'évolution en comparable par population

Plus qu'une mesure classique de l'évolution d'un agrégat entre deux années, il s'agit **d'analyser les facteurs expliquant cette dernière,** en identifiant différentes populations dans le contexte d'Ésane (c'est-à-dire reposant simultanément sur des sources administratives et des enquêtes).

Ce travail d'analyse **n'a pas encore donné lieu à une diffusion systématique par l'Insee**. Il est en revanche fréquemment utilisé lorsque l'on fait des études longitudinales d'un secteur précis ou pour contrôler des résultats a priori surprenants. Il s'agit d'une ***possibilité nouvelle offerte par Ésane**** qui nous semble donc utile à connaître.

9.6.2.1 Les principes généraux

Compte tenu des remarques précédentes du § *9.6.1*, on s'intéresse à l'interprétation de l'évolution « en comparable » car le premier phénomène qui vient perturber l'évolution apparente est celui des restructurations dont on peut s'affranchir en prenant en compte les enveloppes. Les **quatre facteurs** d'évolution mis en évidence, et les catégories d'entreprises qui en découlent sont les suivants :

1. La démographie d'entreprises (créations et cessations) :

   On cherchera à mesurer par exemple l'impact de la disparition d'entreprises industrielles - extraction, textile…- ou de la création de Start up, ou de l'immatriculation des panneaux solaires, ou la création du statut d'auto-entrepreneurs, etc., dans l'évolution de leur secteur.

   On cherche donc à isoler :

   - **Les entreprises naissantes** (créées au cours de l'exercice N) : on s'intéressera aux entreprises créées ou réactivées, mais on ne prendra pas en compte les conséquences de restructuration (le cas de l'entreprise 5 dans l'exemple précédent ; Cf. 2.), ni les « entrées de champ » qui relèvent d'un autre phénomène (Cf. 3.).
   - **Les entreprises cessées** (entreprises cessées au cours de l'exercice N-1) : même chose on ne prend pas en compte les restructurations (cas des entreprises 2 et 3 de l'exemple précédent) ni les sorties de champ.

   Une entreprise cessée *en cours* d'année N sera considérée comme pérenne (même si elle n'a pas 12 mois d'activité pour l'année N) ; il en va de même pour les entreprises créées en N-1 et ayant encore une activité en N.

2. Les restructurations d'entreprises:

   Les entreprises[398] **en restructuration** voient leur comportement économique modifié, ce qui impacte nécessairement les agrégats sectoriels. On retiendra donc globalement pour notre analyse le contour de ces restructurations, appelé enveloppe (Cf. *Chapitre 3*). Dans l'exemple précédent, si on s'intéresse au secteur B, il s'agit de l'évolution de l'enveloppe 1, comprenant les entreprises 2 et 3 en N-1 et l'entreprise 5 en N ; on notera que l'entreprise 2 était classée en A mais compte tenu de sa participation à une fusion, on en a besoin pour interpréter l'évolution en comparable du secteur B.

3. Les modifications de classement sectoriel :

   On cherche ici à mesurer **les entrées et sorties de champ** : on entend par champ le champ d'observation auquel on s'intéresse (et pas le champ Ésane complet). Si l'on considère un groupe donné de la NAF, on considérera les mouvements sectoriels de et vers le groupe comme des sorties et entrées de champ. Il s'agit donc ici d'identifier les entreprises ayant un classement différent en N-1 et N, soit :

   - Les entreprises en entrée de champ (entreprise active en N dans le secteur et active en N-1 mais dans un autre secteur, les deux secteurs étant compris dans le champ des enquêtes).
   - Les entreprises en sortie de champ (entreprise active en N-1 dans le secteur et active en N mais dans un autre secteur, les deux secteurs étant compris dans le champ des enquêtes).

4. L'évolution des entreprises pérennes :

   Il s'agit des entreprises **ayant connu une activité** au cours **des deux millésimes considérés**, et n'ayant pas connu de modification de classement.

   Le statut de pérenne est lié à l'activité de l'unité. Ainsi, une unité qui cesse au cours de l'exercice N mais qui a été active une partie de l'exercice sera considérée comme pérenne. Les durées d'exercice N et N-1 seront certes différentes mais l'unité est prise en compte dans les pérennes. Idem pour une unité créée en cours d'exercice N-1.

Eu égard à un secteur S donné, les quatre catégories précédentes d'entreprises **forment une partition (P)** des entreprises du secteur S à l'année N. Mais le classement d'une unité donnée dans une de ces catégories **dépend du niveau d'agrégation** auquel on se place. Une entreprise qui change d'APE au « niveau groupe » entre N et N-1, passant du 2830Z au 2841Z par exemple, sera considérée comme :

- pérenne si on calcule des évolutions au niveau division de la NAF. Dans l'exemple, elle sera pérenne dans la division 28 ;
- en entrée de champ pour son groupe d'arrivée et en sortie de champ de son groupe de départ si on calcule des évolutions au niveau groupe de la NAF. Dans l'exemple, elle sera donc considérée en entrée du champ 284 et en sortie du champ 283.

On veut maintenant **estimer** la part imputable à chacune des catégories. Les deux dernières catégories reposent sur la pérennité ou la modification du classement sectoriel entre N-1 et N. Leur repérage implique **d'avoir réellement observé le classement sectoriel en N et en N-1**. Ainsi, seules les données des entreprises ayant réellement **répondu à l'enquête** deux années successives pourront être mobilisées pour les estimations, puisque c'est uniquement pour ces deux populations que l'on dispose de cette information sur deux années consécutives. Cette décision nous pousse donc à **exclure de notre analyse les unités entrantes ou sortantes de l'échantillon de l'enquête en N** (cf. schéma page suivante et § *5.1.4*). Pour ces unités en effet nous ne disposons au mieux que d'une information sur une des deux années. En effet, lorsqu'on observe au cours de l'année N une modification de classement, celle-ci peut être intervenue réellement beaucoup plus tôt, si l'entreprise n'a pas été interrogée depuis longtemps.

Comme précisé dans le § *5.2.2*, le tirage de la partie échantillonnée de l'échantillon de l'EAP (les petites entreprises industrielles) ne comprend pas de partie conservée jusqu'à la campagne 2013 comprise (puisqu'on est en rotatif pur), et ne permet donc pas de connaître les modifications sectorielles de l'année pour les entreprises de cette partie appartenant au champ de l'industrie au lancement, alors que l'Esa permet de les connaître pour la partie conservée de l'échantillon.

[398] Une entreprise absorbée par une autre sera donc considérée comme faisant partie d'une restructuration bien qu'elle soit cessée en N. De la même façon une entreprise issue d'une scission totale sera considérée comme en restructuration et non naissante.

On limitera donc l'analyse aux secteurs couverts par l'Esa et la partie exhaustive de l'EAP[399], avec une possible sous-estimation, puisque les éventuelles petites entreprises venant de l'industrie ne seront pas prises en compte.

9.6.2.2 La décomposition de l'évolution en comparable

La décomposition de l'évolution économique selon la partition (P) peut donc s'écrire de la façon suivante :

$$\boxed{X_N^S = N_N^S + R_N^S + EC_N^S + P_N^S} \quad \textit{[Formule (10)]}$$

avec :

$$\begin{cases} N_N^S = \text{agrégat des naissantes} \\ R_N^S = \text{agrégat des entreprises en restructuration} \\ EC_N^S = \text{agrégat des entrées de champ en N} \\ P_N^S = \text{agrégat des entreprises pérennes} \end{cases}$$

$$X_{N-1}^S = M_{N-1}^S + R_{N-1}^S + SC_{N-1}^S + P_{N-1}^S$$

avec :

$$\begin{cases} M_{N-1} = \text{agrégat des cessées (M comme mortes)} \\ R_{N-1} = \text{agrégat N-1 des entreprises en restructuration} \\ SC_{N-1} = \text{agrégat N-1 des sorties de champ en N} \\ P_{N-1} = \text{agrégat N-1 des entreprises pérennes} \end{cases}$$

D'où l'on peut déduire :

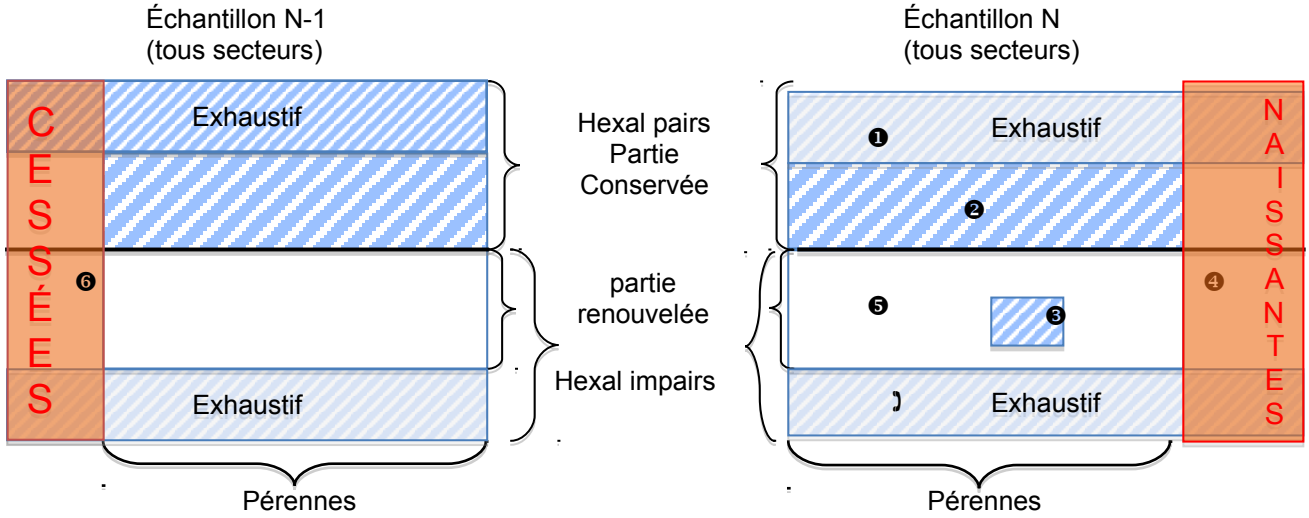
$$\Delta_{N/N-1} = \frac{(X_N^S - X_{N-1}^S)}{X_{N-1}^S}, \quad , = \frac{(N_N^S - M_{N-1}^S)}{X_{N-1}^S} + \frac{(R_N^S - R_{N-1}^S)}{X_{N-1}^S} + \frac{(EC_N^S - SC_{N-1}^S)}{X_{N-1}^S} + \frac{(P_N^S - P_{N-1}^S)}{X_{N-1}^S}$$

Soit [Formule (11)] :

$$\boxed{\Delta_{N/N-1} = \Delta_D^1 + \Delta_R^2 + \Delta_C^3 + \Delta_P^4} \quad \text{(D : démographie ; R : restructuration ; C : champ ; P : pérennité)}$$

L'estimation de cette décomposition est complexe, car il est difficile de définir la notion de «pérennes» lorsque l'on utilise les résultats d'un échantillon pour calculer des agrégats. De plus les **4 facteurs d'évolution** ci-dessus peuvent se retrouver dans les **5 zones de la partition** de l'échantillon N :

Le schéma ci-dessous permet de bien comprendre de quoi on dispose :



[399] Toutefois dans ce cas, les formules qui suivent seraient à adapter du fait que l'on a que l'exhaustif.

On se souvient (Cf. § 5.1.4) que l'échantillon est globalement constitué de deux parties : les entreprises de numéro hexal pair et les entreprises de numéro hexal impair. Hormis la partie exhaustive qui par définition est conservée chaque année, cette dichotomie sert à renouveler l'échantillon par moitié. Il y a donc une moitié de l'échantillon (hachurée en bleu ci-dessus) qui est conservée entre N-1 et N, et une autre moitié qui est renouvelée. Mais comme on effectue un tirage avec remise, dans la partie renouvelée de N, on peut ré-interroger des entreprises qui étaient déjà interrogées en N-1 (c'est le petit rectangle ❸ hachuré dans la partie renouvelée de l'échantillon de N ; on avait vu au § 5.1.4 qu'elle comprend chaque année environ 3000 entreprises). Ces entreprises seront donc prises en compte pour le calcul des résultats en évolution.

9.6.2.3 La contribution des différentes sous populations à considérer

On souhaite utiliser des informations réellement fournies par les entreprises. On ne retiendra donc que les entreprises **ayant réellement répondu à l'enquête deux années de suite** pour les pérennes, et ayant répondu pour l'année concernée pour les cessées (année N-1) et les créées (année N). A noter que, dans les populations des créées et des cessées, on observe un plus grand taux de non-réponse, y compris parmi les entreprises appartenant à la strate exhaustive ; les entreprises appartenant à la strate exhaustive non répondantes en N mais répondantes en N-1 faisant l'objet d'extrapolation **individuelle** (Cf. 5.6.2.1), on propose de les assimiler à des répondantes et donc **de les conserver** afin d'utiliser un nombre suffisant d'entreprises (en particulier pour les cessées). Ceci accroît un peu la population sur laquelle repose les estimations mais complexifie un peu les formules.

On pourra utiliser pour les estimations:

- Les répondantes pour les deux années de la partie exhaustive de l'échantillon (zone ❶) ;
- Les répondantes pour les deux années de la partie conservée de l'échantillon (pour la partie non exhaustive ; zone ❷) ainsi que celles appartenant à la partie renouvelée, mais tirées dans les deux échantillons ; zone ❸ ;
- Les entreprises naissantes répondantes ou les entreprises extrapolées appartenant à la strate exhaustive de l'échantillon des naissantes zone ❹ ;
- Les entreprises cessées en cours d'année N-1 répondantes de l'échantillon N-1 (zone ❻), et les extrapolées appartenant à la strate exhaustive.

[Soit environ 80 à 90 mille entreprises, à raison d'un taux moyen de réponse (avec les hors champ) autour de 70% appliqué à l'exhaustif (≈75000), un demi échantillon renouvelé (≈ 42000) et les 3000 tirées à nouveau].

En revanche, on ne pourra pas utiliser les entreprises de la partie renouvelée de l'échantillon, c'est-à-dire la zone ❺ du graphique.

On voit donc que pour estimer les différentes composantes de la formule 11, il faudra utiliser des estimateurs calculés sur des champs différents ; on fera appel aux indicatrices suivantes :

$$\mathbb{I}^{N}_{i\in naissante,s} = \begin{cases} 1 \text{ si l'unité i n'est pas en restructuration et est naissante et répondante dans le secteur S en N} \\ 0 \text{ sinon} \end{cases}$$

$$\mathbb{I}^{N}_{i\in cessée,s} = \begin{cases} 1 \text{ si l'unité i est répondante et cessée dans le secteur S en N} - 1 \\ 0 \text{ sinon} \end{cases}$$

$$\mathbb{I}^{N}_{i\in restruct,s} = \begin{cases} 1 \text{ si l'unité i appartient à une enveloppe de restructuration de l'année N du secteur S} \\ 0 \text{ sinon} \end{cases}$$

(à noter que toutes les unités légales en restructuration sont mises dans une enveloppe et sont dans la strate exhaustive en N (Cf. § 3.2.1)

$$\mathbb{I}^{N}_{i\in entrée_champ,s} = \begin{cases} 1 \text{ si l'unité i est répondante en N, n'est pas en restructuration et est dans le secteur S en N et dans un autre en N} - 1 \\ 0 \text{ sinon} \end{cases}$$

$$\mathbb{1}^{N}_{i\in sortie_champ,s} = \begin{cases} 1 \text{ si l'unité i n'est pas dans une enveloppe de restructuration de N, est} \\ \text{répondante dans le secteur S en N} - 1 \text{ et dans un autre en N} \\ 0 \text{ sinon} \end{cases}$$

$$\mathbb{1}^{N}_{i\in pérenne,s} = \begin{cases} 1 \text{ si l'unité i n'est pas en restructuration, est répondante en N et N} - 1 \text{ et est} \\ \text{pérenne dans le secteur S} \\ 0 \text{ sinon} \end{cases}$$

1. La notion de « naissante » peut être estimée sur :
   - la partie exhaustive de l'échantillon ; (hachuré bleu et blanc ❶, y compris la partie pointillée rouge des naissantes dans l'exhaustif)
   - l'échantillon des naissantes tiré en N dans la partie non exhaustive de l'échantillon (en pointillé rouge sur le schéma ; ❹).

Un estimateur naturel de la contribution à l'agrégat X des entreprises naissantes, utilisant les poids après calage W_i de l'unité i dans l'échantillon de l'année N, est donc :

$$\hat{N}_S^N = \sum_{i\in exh} X_i^N * 1I_{i\in naissante,s} + \sum_{i\in ech} w_i^N * X_i^N * 1I_{i\in naissante,s} * 1I_{i=répondante_{en_N}}$$

2. La notion de cessée peut être estimée sur :
   - la totalité de l'échantillon N-1.

De la même façon les cessées (M comme mortes) seront estimées sur l'échantillon N-1 par :

$$\hat{M}_S^{N-1} = \sum_{i\in exh} X_i^{N-1} * 1I_{i\in cessées,s} + \sum_{i\in ech} w_i^{N-1} * X_i^{N-1} * 1I_{i\in cessées,s} * 1I_{i=répondante_{en_N}-1}$$

3. La notion de restructuration peut être estimée sur :
   - la totalité de l'échantillon N.

L'estimateur naturel des entreprises qui appartiennent à une enveloppe de restructuration en N (qui appartiennent toutes à l'exhaustif) est donc :

$$\hat{R}_s^N = \sum_{i\in exh} X_i^N * 1I^N_{i\in restruct,s}$$

4. Les notions de pérennité, d'entrée et de sortie de champ ne peuvent être calculées que si on dispose de deux années d'observation pour pouvoir caractériser ainsi ces unités, donc sur :
   - la partie exhaustive de l'échantillon ; (hachuré bleu et blanc)
   - la partie conservée de l'échantillon ; (hachuré en noir)
   - les entreprises de la partie renouvelée ayant été tirées les deux années. (petit rectangle hachuré en noir)

On appellera Cons_large la réunion des zones ❷ et ❸ c'est-à-dire de la partie conservée stricto sensu de l'échantillon et de la partie renouvelée mais qui a été tirée à nouveau et qui était donc déjà présente l'année N-1 (soit la zone hachurée en noir sur le schéma).

Prenons d'abord par exemple le cas des entrées de champ. Celles-ci peuvent se produire dans toutes les zones de l'échantillon sauf la zone ❹ des naissantes, c'est-à-dire parmi les exhaustives répondantes, les exhaustives extrapolées (non répondantes), l'échantillon conservé large et l'échantillon renouvelé. L'estimateur naturel s'écrira donc :

$$\hat{EC}_s^N = \sum_{i\in exh_rep} X_i^N * \mathbb{1}_{i\in entchamp,s} + \sum_{i\in exh_nonrep} X_i^N * \mathbb{1}_{i\in entchamp,s} + \sum_{i\in cons_large} w_i^N * X_i^N * \mathbb{1}_{i\in entchamp,s} + \sum_{i\in rnvl} w_i^N * X_i^N * \mathbb{1}_{i\in entchamp,s}$$

Dans cette expression de la forme A+B+C+D, les quantités B et D sont inconnues : par définition, dans les exhaustives non répondantes, du fait qu'elles ne sont pas répondantes, on ne sait pas celles qui sont en entrée du secteur S. Mais on peut faire l'**hypothèse** que la part des entrées dans S dans les non répondantes (soit B dans le total des non répondantes) est la même que la part des entrées dans S dans les répondantes, c'est-à-dire que :

$$\frac{B}{\sum_{i \in exh_{nonrep}} X_i^N} = \frac{A}{\sum_{i \in exh_{rep}} X_i^N}$$

De la même façon, toujours par définition, on ne connaît pas la part des entrées de champ de la partie renouvelée de l'échantillon puisqu'on ne connaît pas leur secteur en N-1. Mais on peut faire **l'hypothèse** que leur part dans l'échantillon renouvelé est la même que la part des entrées de champ dans l'échantillon conservé, c'est-à-dire que :

$$\frac{D}{\sum_{i \in rnvl} w_i^N * X_i^N} = \frac{C}{\sum_{i \in cons_l \arg e} w_i^N * X_i^N}$$

On peut donc réécrire l'estimateur des entrées de champ sous la forme :

$$E\hat{C}_s^N = A * \left(1 + \frac{\sum_{i \in exh_{nonrep}} X_i^N}{\sum_{i \in exh_{rep}} X_i^N}\right) + C * \left(1 + \frac{\sum_{i \in rnvl} w_i^N * X_i^N}{\sum_{i \in cons_l \arg e} w_i^N * X_i^N}\right)$$

On peut donc définir deux coefficients **calculés** l'un sur les exhaustives Kexh_s^N , l'autre sur l'échantillon Kech_s^N , donc sur les parties connues pour lesquelles on dispose des 2 années d'observation, ayant les valeurs suivantes :

$$Kexh_s^N = \frac{\sum_{i \in exh_{rep}} X_i^N + \sum_{i \in exh_{nonrep}} X_i^N}{\sum_{i \in exh_{rep}} X_i^N} \quad \text{et} \quad Kech_s^N = \frac{\sum_{i \in cons_l \arg e} w_i^N * X_i^N + \sum_{i \in rnvl} w_i^N * X_i^N}{\sum_{i \in cons_l \arg e} w_i^N * X_i^N}$$

de sorte que l'estimateur des entrées de champ s'écrit finalement :

$$\boxed{E\hat{C}_s^N = \mathrm{Kexh}_s^N * \sum_{i \in \mathrm{exh_rep}} X_i^N * 1\mathrm{I}_{i \in \mathrm{entchamp}_s} + \mathrm{Kech}_s^N * \sum_{i \in \mathrm{cons_large}} w_i^N * X_i^N * 1\mathrm{I}_{i \in \mathrm{entchamp},s}}$$

On démontrerait de la même façon que l'estimateur des sorties de champ s'écrit sous la forme :

$$\boxed{S\hat{C}_s^{N-1} = Kexh_s^{N-1} * \sum_{i \in exh_{rep}} X_i^{N-1} * 1I_{¿sortiechamp,s}^N + Kech_s^{N-1} * \sum_{i \in cons_l \arg e} w_i^{N-1} * X_i^{N-1} * 1I_{i \in sortiechamp,s}^N}$$

(Attention que $S\hat{C}_s^{N-1}$ est un agrégat (N-1) des valeurs de X, mais les indicatrices sont mesurées par rapport à la situation en N. Les sommes portent bien sur la situation des entreprises en N).

Enfin, pour les pérennes qui peuvent se décomposer comme les deux précédentes sur 4 sous populations de l'échantillon, on prendra l'estimateur suivant que l'on calculerait de la même façon que ci-dessus :

$$\boxed{\hat{P}_s^N = Kexh_s^N * \sum_{i \in exh_{rep}} X_i^N * 1I_{i \in perenne,s} + Kech_s^N * \sum_{i \in cons_l \arg e} w_i^N * X_i^N * 1I_{i \in perenne,s}}$$

9.6.2.4 Évolution, poids et contribution

À partir des formules ci-dessus, on pourra calculer toutes les composantes de la formule 11 et en particulier :

- le poids d'une catégorie en niveau : le poids des pérennes en N-1 : $\frac{P^s_{N-1}}{X^s_{N-1}}$
- l'évolution de certaines catégories ; par exemple pour les pérennes :

$$\Delta P_{N/N-1} = \frac{\left(P^s_N - P^s_{N-1}\right)}{P^s_{N-1}}$$

- la « contribution » de chacun des éléments par exemple pour les pérennes : $CP_{N/N-1} = \frac{\left(P^s_N - P^s_{N-1}\right)}{X^s_N} = \Delta^4_P$ (de la formule 11)
- ou encore le poids dans l'évolution : $\frac{\left(P^s_N - P^s_{N-1}\right)}{\left(X^s_N - X^s_{N-1}\right)}$

9.6.2.5 Les précautions à prendre dans l'interprétation des résultats

Les formules précédentes ont été programmées en interne à titre expérimental sur les résultats de 2010 en comparaison à ceux de 2009. Les travaux ont été menés sur deux variables (chiffre d'affaires, valeur ajoutée) et au niveau groupe (3 premiers chiffres) de la NAF (soit 110 groupes car on a exclu l'industrie pour laquelle les formules seraient légèrement différentes puisqu'il n'y avait pas de renouvellement par moitié de l'échantillon sur ces secteurs couverts par l'EAP en 2010).

Écarts entre évolution apparente et évolution décomposée estimée :

Comme on l'a vu dans le § *9.6.1*, il y a deux raisons pour lesquelles les évolutions décomposées estimées **peuvent s'éloigner** de l'évolution « apparente » calculable sur les niveaux publiés : le passage à une évolution en comparable par élimination du phénomène des restructurations puis l'estimation proprement dite par l'utilisation des formules ci-dessus. Comme seul l'échantillon permet de mesurer les entrées et sorties de champ, et donc les pérennes, on ne se sert pas du tout des sources administratives (qui donneraient un poids beaucoup trop important aux pérennes). Les niveaux publiés (fondés sur les estimateurs composites) concernent un champ beaucoup plus large et les évolutions estimées pour la décomposition ne prennent en compte **qu'une partie de l'échantillon**.

Les calculs ont cependant montré que dans un assez grand nombre de cas, la décomposition permettait d'expliciter une tendance et d'en fournir les facteurs explicatifs, à condition de respecter les **principes de précaution suivants** :

- l'évolution décomposée doit être de même signe que l'évolution calculée d'après les agrégats publiés, ou l'écart entre les évolutions doit être inférieur à 1 point ;
- l'écart entre ces évolutions ne doit pas être trop important (5 points par exemple) ;
- le rapport des deux évolutions doit être compris entre ½ et 3/2.

Sous ces conditions, on pouvait utiliser l'information pour 70 % des groupes. Pour 40 % des groupes l'écart entre les évolutions était inférieur à 1 point, pour 20 % il était inférieur à 2 points et pour 24 %, l'écart était compris entre 2 et 5 points.

Nombre d'unités participant aux différents éléments des évolutions

Selon les secteurs, l'analyse finale de l'évolution pouvait porter sur assez peu d'unités.
Ainsi, l'analyse de la décomposition des évolutions reposait sur 52 000 entreprises (hors industrie), dont 15 500 de la partie conservée non exhaustive de l'échantillon. Cette situation cache des inégalités sectorielles et peut donc poser des problèmes de fiabilité des résultats. Ainsi, pour 20 groupes, il y avait moins de 10 entreprises de l'échantillon (partie sondée) qui participaient au calcul et pour 32 d'entre eux, il y avait moins de 30 entreprises.
Il est cependant difficile de fixer un seuil raisonnable en dessous duquel on n'accorderait plus de crédit à l'analyse, car le poids de l'exhaustif d'une part, et le poids moyen des entreprises interrogées d'autre part varient fortement d'un secteur à l'autre. Ainsi les secteurs des transports en commun ont très peu d'unités, mais qui sont toutes interrogées exhaustivement.

Présentation d'un résultat :

Ci-joint un exemple des résultats de cette simulation numérique :

Tableau 6 : Décomposition de l'évolution 2010/2009 du CA et de la VA pour le secteur « Intermédiaires du commerce de gros» (461)

Variable	Agrégat 2010	Évolution en comparable	Évolution décomposée Estimée	Les différentes sous-populations				Évolution de la population des pérennes
				Pérennes	Restructu Rations	Créations Cessations	Entrées et sorties de champ	
CA	122 112 711	4,13 %	3,90 %	6,44 %	- 0,02 %	- 0,47 %	- 2,05 %	7,28 %
Va	10 436 913	- 5,43 %	- 5,29%	2,65 %	- 0,04 %	- 0,62 %	- 7,29 %	3,38 %

Source : Ésane 2009 et 2010

Dans l'exemple ci-dessus, les évolutions « en comparable » et décomposée estimée sont relativement proches. L'évolution du chiffre d'affaires est expliquée principalement par les pérennes, alors que le secteur connaît quelques sorties de champ.
En revanche, les mouvements sectoriels expliquent la baisse de la valeur ajoutée du secteur, alors même que la valeur ajoutée des entreprises pérennes augmente.

10. Présentation de la phase de vérification des données et de validation des agrégats

Cette phase, qui se situe **une fois le travail de « contrôle » des données individuelles par les gestionnaires terminé**, vise principalement deux objectifs :

1. le premier se place dans la mise en œuvre d'une ***démarche qualité**** comme on le ferait pour n'importe quel process industriel mais qui était ***innovante pour l'Insee à sa mise en place****. Cette phase doit en effet permettre à une catégorie spécifique d'agents qu'on nomme « valideurs » pour la circonstance, de **vérifier la conformité des procédures** appliquées par les gestionnaires. Les grands principes de cette phase seront présentés dans le § *10.1* et la mise en œuvre concrète au § *10.2*. Elle permet de contrôler la pertinence des procédures définies et de vérifier que les gestionnaires comprennent et mettent bien en application les différentes consignes de contrôle qui leur ont été données. Dans le cas contraire, des formations complémentaires sont mises en place afin de rectifier les écarts. L'*activité Ésane* de vérification finale des dossiers est idéale pour constituer une typologie des problèmes rencontrés par les « *valideurs* », illustrée par des exemples réels. Cette typologie permet d'enrichir les formations de cas plus précis. Véritable complément du bilan de la validation des agrégats, **cette typologie des interventions des « *valideurs* »** qui évolue chaque année fournit une photographie de la situation à laquelle font face les gestionnaires.
2. le second, assez traditionnel pour tout processus de production statistique, est de **s'assurer de la vraisemblance et de la qualité des agrégats diffusés** ; la nécessité de ce contrôle final (malgré tous les contrôles antérieurs), portant en fait à la fois sur les données individuelles et agrégées, fait l'objet du § *10.3* et son contenu qui s'appuie sur ***une approche originale**** des § *10.4* et *10.5*.

Ces phases de contrôles d'abord puis de vérifications des contrôles sont susceptibles d'intervenir quatre fois durant une campagne Ésane[400] :

- durant l'été, à l'issue des premiers contrôles du questionnaire Esa. Ces contrôles concernent essentiellement le chiffre d'affaires et sa ventilation en branches ;
- en début d'automne à l'issue des contrôles des liasses fiscales « anticipées » sur les variables qui font partie de la diffusion provisoire (chiffre d'affaires, valeur ajoutée, EBE, résultat net comptable, investissement) ;
- en fin de campagne sur l'ensemble des données des liasses fiscales ;
- en fin de campagne sur les données sectorielles de l'enquête Esa.

La vérification peut évidemment donner lieu à un contact de l'entreprise le cas échéant.

10.1 Les grands principes de cette phase

La vérification des dossiers est réalisée sur l'initiative des responsables d'équipe de gestionnaires par eux-mêmes et les personnes habilitées pour vérifier :

- d'une part qu'ils auraient traité de la même façon que ne l'a fait le gestionnaire qui a été chargé du dossier ;
- d'autre part que les commentaires laissés par les gestionnaires au moment du contrôle sont compréhensibles et seront utilisables par d'autres utilisateurs des données (et notamment les personnes en charge d'études qui pourront le cas échéant expliquer une évolution surprenante au niveau d'un secteur par le comportement atypique d'une grosse unité, comportement qui aura été expliqué dans le commentaire).

Cette vérification peut donner lieu à des corrections de données ou à de simples observations.

La vérification correspond à deux objectifs principaux :

i. S'assurer que dans les dossiers contrôlés par les gestionnaires **qui ont une influence très forte** sur les agrégats (selon la procédure de macro contrôles décrite au § *2.7*, mesurée par la

[400] La modularité du contrôle des sources est donc également respectée au niveau de cette vérification.

formule du QEG définie au § 2.8) il ne subsiste pas de grosses erreurs qu'il faut absolument éviter. On qualifie ces dossiers **« d'incontournables »**.

ii. Détecter des erreurs éventuelles **pour un échantillon aléatoire complémentaire** de dossiers tiré dans les entreprises contrôlées en vérifiant s'il existe des erreurs systématiques ou tout au moins fréquentes qui proviendraient d'une inadéquation des procédures mises en place ou de la non-compréhension de certaines consignes de traitement par les gestionnaires et qui justifieront la mise en place de mesures correctives pour l'année suivante en général (via une formation, une réécriture des consignes ou des procédures de référence). Les dossiers correspondant à cette population sont appelés « **vérifiables** ». Une fois que la *sélection* « raisonnée » ou « aléatoire » est faite, les dossiers de cette sélection sont dits « **à vérifier** ».

Ces vérifications pourront aussi être l'occasion de **contrôler**, dans le cadre :

- **de créations** (respectivement cessations) **très importantes**, **qu'il s'agit de vraies créations** (respectivement cessations) et non d'une restructuration mal identifiée. Les créations (respectivement cessations) de taille importante sont en effet rares ou alors dûment référencées dans la presse ;
- **d'évolutions importantes sur des entreprises pérennes**, **qu'elles ne cachent pas**, là encore, **des restructurations** juridiques non détectées. Elles peuvent correspondre à des redéfinitions des flux au sein de groupes d'entreprises. Par exemple, une centralisation des achats ou des exportations au sein d'un groupe peut conduire à des modifications importantes des flux ;
- **de changements soudains en termes d'activité principale** ou de branches pour des entreprises importantes, **qu'il ne s'agit pas d'erreurs de déclaration**. En effet, là encore, l'inertie du monde réel fait que ceci est a priori suspect. Changer d'activité suppose d'acquérir de nouveaux moyens de production, de former un nombre important de salariés, de conquérir des marchés. En revanche, une restructuration avec apports partiels d'actifs – qui ne correspond qu'à un changement de structure financière et sociale – est plus rapide. Cela peut également cacher une erreur passée, soit de la part de l'Insee, soit de la part de l'entreprise ;
- **des choix méthodologiques faits, que l'impact** de certains redressements automatiques ou traitements de la non-réponse totale **est conforme** à ce que l'on pouvait en attendre. Ces contrôles sont alors réalisés par un membre du DMS[401].

En termes d'organisation du travail, la vérification peut se faire en parallèle du travail de contrôle des agents éventuellement sur la même phase de contrôle élémentaire mais pas sur les mêmes dossiers. Autrement dit, il est inutile d'attendre que l'ensemble des dossiers à traiter pour un processus de contrôle donné le soit avant de commencer la phase de vérification. Cette phase peut commencer dès qu'un minimum de dossiers aura été validé par les gestionnaires.

10.2 La mise en œuvre de la démarche qualité

Comme vu précédemment, les entreprises à vérifier se répartissent en deux sous-populations :

- les entreprises à vérifier par le service Ésane de Nantes (§ *10.2.1*) ;
- les entreprises à vérifier pour des questions méthodologiques (par la division Ésane ou le DMS) (§ *10.2.2*).

10.2.1 Les entreprises à vérifier par le service Ésane (à Nantes)

Sont susceptibles d'être vérifiées par le service Ésane de Nantes, toutes unités *(entreprise ou enveloppe)* **dont la réponse a été validée ou corrigée par les gestionnaires et n'ayant pas encore été sélectionnée dans une vague de vérification antérieure.**

Les entreprises à vérifier par un gestionnaire se classent en 3 grandes catégories :

i. les entreprises pérennes[402] (y compris les enveloppes de restructuration) ;
ii. les créations et cessations ;
iii. les entreprises changeant d'APE.

[401] Département des Méthodes Statistiques de la direction de la méthodologie et de la coordination statistique et internationale.
[402] Il s'agit des entreprises présentes en N-1 et N et sans changement d'APE.

Parmi ces trois grandes catégories, une population particulière qui sera à vérifier d'office est distinguée. Il s'agit, comme on l'a vu ci-dessus, des dossiers jugés « **incontournables** » ayant une influence forte[403] sur les agrégats. Dans les faits, le seuil des dossiers « incontournables » est défini en commun par le service Ésane de Nantes et la division Ésane parisienne en fonction des ressources disponibles pour la vérification. **Toutes les unités jugées « incontournables » sont vérifiées.**

Pour les autres unités potentiellement vérifiables, on procède par **sondage**. Le volume d'unités à vérifier dépend de la capacité de chaque division sectorielle.

C'est le chef de division nantais qui décide in fine de la taille de cet échantillon. Il définit pour chaque strate de tirage, **le nombre d'unités à vérifier en plus des dossiers « incontournables »**.
Les strates de tirage sont obtenues en croisant :

- la catégorie d'entreprise à vérifier (pérennes, changeant d'APE ou créées ou cessées) ;
- le type de gestionnaire[404] ayant contrôlé le dossier.

Pour prendre sa décision quant à la taille de l'échantillon dans chacune des strates, le chef de division dispose pour chaque strate de tirage :

- du nombre de dossiers vérifiables ;
- du nombre de dossiers incontournables (et qui sont donc à vérifier d'office).

Le tirage de l'échantillon ne tient donc pas du tout compte de l'influence des dossiers (qui, elle, est prise en compte dans les dossiers « incontournables »). Il s'agit d'un sondage aléatoire simple stratifié. Afin d'avoir une représentativité sectorielle, la stratification définitive de tirage prend en compte le groupe (trois premiers chiffres de la NACE) de l'APE des unités à vérifier.

La définition du nombre exact de dossiers à tirer par groupe d'APE s'inspire de la méthode de Cox. L'algorithme en est le suivant :

- Soit n, le nombre de dossiers à tirer pour une strate donnée. Ce nombre est défini par le chef de division ;
- Soit N, le nombre de dossiers vérifiables (et non incontournables) de la strate ;
- Soit N_{APE} le nombre total de dossiers vérifiables (et non incontournables) du croisement strate*groupe APE. On a donc $N = \sum_{APE \in strate} N_{APE}$
- On pose au départ : reste = 0

Pour chaque APE de la strate on définit le nombre de dossiers à »à vérifier » n_{APE} par:
$n_{APE} = partie\ entère(\frac{n}{N} * N_{APE} + reste)$
et Reste= $\frac{n}{N} * N_{APE} + reste$ -n_{APE}
On passe à l'APE suivante de la strate.

Exemple :

Pour la division BTP, il y a 1000 entreprises pérennes « non incontournables » qui ont été traitées par un gestionnaire sectoriel (donc N = 1000 dossiers vérifiables avec la décomposition N_{APE} ci-dessous).
Le chef de division souhaite faire vérifier 130 dossiers (taille de l'échantillon n = 130).
Au final il aura la répartition sectorielle n_{APE} à vérifier suivante :

groupe APE du BTP	Nombre N_{APE} de dossiers vérifiables	Nombre n_{APE} de dossiers à vérifier	reste
411	120	15 = 120 * 130/1000	0,6
412	230	30= PE(230*130/1000+ 0,6)	0,5
421	150	20	0
422	70	9	0,1
429	30	4	0
431	90	11	0,7
432	120	16	0,3
433	80	10	0,7
439	110	15	0
Total	1000	130	

[403] Cette influence se mesure comme on l'a vu par la formule du QEG définie au § 2.8.
[404] Comme vu précédemment, les trois types à distinguer à ce niveau sont les gestionnaires sectoriels, les gestionnaires responsables des très grandes entreprises et les chefs de division eux-mêmes.

10.2.2 Les entreprises à vérifier par les méthodologues (parisiens)

Ce champ correspond aux cas pour lesquels on pourrait être amené à **remettre en question les valeurs redressées de façon automatique** qui servent au calcul des agrégats parce qu'on est dans un cas tellement important que la méthodologie standard n'est pas applicable directement et doit être **réexaminée par un expert**. Il s'agit par exemple des cas ci-après :

- imputation (totale ou partielle sur une variable d'intérêt) sur l'année courante n. En effet, si la méthode d'imputation est trop inadaptée, on peut faire une estimation « sur mesure » dans des cas d'exception. Par exemple, si une très grosse entreprise ne répond pas à l'Esa, et a un comportement dont on pense pour des raisons économiques qu'il est différent du reste du secteur, alors on pourrait par exemple essayer d'utiliser à titre exceptionnel les informations TVA ou une autre source ; en effet pour de telles entreprises des redressements qui utilisent le profil moyen du secteur sont inadaptés. Il faut donc repérer les entreprises influentes qui ont fait l'objet d'une imputation totale ou partielle, sur l'année N, en utilisant la contribution d'un macro contrôle de type Différence utilisée pour la validation des données ;
- pour l'Esa, les entreprises appartenant à la partie renouvelée de l'échantillon en *N*, et pour lesquelles on peut calculer une contribution fictive à l'évolution *N/N-1* qui est importante. Pour cette contribution fictive qui n'est pas calculée durant la phase de contrôle des données, on impute les données *N-1* par la moyenne de strate. On peut alors rencontrer le cas des entreprises mal classées dans la base de sondage en termes de taille, et dont l'inclusion soudaine en *N* va conduire à un « saut » dans les estimations dont on peut penser qu'il est excessif. Le méthodologue devrait dans ces cas-là diminuer le poids, mais de façon raisonnée (cf. traitement de type winsorisation présentés au § *5.6.2.3*) ;
- pour l'Esa, les entreprises dont le poids après calage les rend influentes mais qui sont « normales » (pas de micro-contrôles en anomalie). *Le méthodologue devrait dans ces cas-là diminuer leur poids, mais de façon raisonnée.*

Les vérifications d'ordre méthodologique décrites ci-dessus sont examinées **lors des procédures de traitement de la non-réponse globale par le DMS** ou ultérieurement lors de la validation des agrégats par la division Ésane (Cf. § *10.3 et suivants*) s'il apparaît qu'une des causes d'anormalité d'un agrégat est lié au poids trop important d'une unité.

10.3 La nécessité d'un contrôle final des données individuelles et des agrégats

La démarche qualité décrite précédemment a été particulièrement utile lors de la mise en place d'Ésane au cours des premières campagnes. On constate bien évidemment, au vu de l'expérience accumulée, que **les contrôles sont de mieux en mieux assurés par les gestionnaires** mais la démarche **garde évidemment toute sa pertinence** pour l'accompagnement des nouveaux gestionnaires.

On pourrait penser que cette phase de validation des données individuelles **est suffisante** pour garantir que les agrégats calculés in fine et publiés soient de bonne qualité. Pour de multiples raisons, explicitées ci-dessous, **il n'en est rien** et il est absolument nécessaire de **développer des contrôles après la phase de contrôle (micros et macros) des gestionnaires nantais ;** en effet :

1. les agrégats finaux diffusés (que l'on a vus au *chapitre 9*) sont **différents des agrégats de production** utilisés dans les contrôles, même si dans la pratique ils s'en rapprochent le plus souvent. On se souvient que ces agrégats de production sont obtenus par simple sommation ou utilisation d'un estimateur pondéré classique dans le cadre de l'enquête. Les valeurs de ces deux estimateurs sont forcément différentes de celles obtenues par la méthode de l'estimateur composite (qui est sans biais) simplement parce que **les formules de calcul ne sont pas les mêmes** (cf. *§ 9.2*) ;
2. non seulement les formules de calcul ne sont pas les mêmes, mais dans le cas de l'Esa, **l'extrapolation de la non-réponse en production est différente** de celle faite pour la diffusion des résultats, pour la partie échantillonnée : "imputation" (le plus souvent à partir des données N-1) en production contre "repondération" en diffusion (Cf. § *5.5*) ;
3. les agrégats calculés reposent non seulement sur les données individuelles collectées et vérifiées mais aussi **sur ces estimations des données manquantes** qui, elles, **ne peuvent pas être contrôlées par les gestionnaires** (Cf. plusieurs paragraphes consacrés aux

redressements automatiques : *5.2.4* pour l'EAP, *5.4.2.2* pour l'ESA, *6.2.10.2* pour les liasses etc.).

4. On se souvient que le travail de contrôle des gestionnaires se termine par un gros **traitement automatique** (qu'on appelle le « Rédi balai » Cf. § *8.6*) qui n'est **plus contrôlé ensuite par des gestionnaires** ; ce programme peut créer des branches d'activités fictives pour les entreprises qui ne sont pas interrogées en enquête et il est donc nécessaire de s'assurer de la cohérence interne des données à l'issue de ce traitement.
5. enfin et peut-être surtout, pour des raisons qui ont été explicitées dans le *chapitre 9*, les estimateurs retenus, bien que sans biais, **ne sont pas linéaires** et **ne conservent pas** nécessairement les propriétés « **naturelles** » des variables que l'on souhaite retrouver en diffusion, **même si celles-ci les possèdent au niveau individuel** (le signe d'un solde ou d'une variable ; les relations d'ordre[405] entre variables, etc.) ; cet **inconvénient** des estimateurs justifierait à lui seul cette phase de contrôle.

En pratique, les utilisateurs internes qui utilisent les données pour des études ou pour la comptabilité nationale (responsables secteurs-produits du DSS et comptables nationaux du DCN) participent à la détection des problèmes ; de même les contrôles effectués par Eurostat sur les données qu'on leur envoie permettent aussi de mettre le doigt sur des problèmes ; et il en subsiste toujours, ne serait-ce qu'à cause du nombre de données à vérifier[406]. L'expérience a montré qu'en parallèle du contrôle des agrégats, qui oblige à rechercher la ou les entreprises responsables d'une anomalie, il était rentable d'exercer aussi des contrôles systématiques simples sur données individuelles.

Pour toutes ces raisons, il était donc important de rajouter une véritable **étape finale de validation et de contrôle** des données à la fois **individuelles** et **agrégées**. Ces contrôles ont été mis en place petit à petit, par les cadres de la direction générale et du service Ésane à Nantes chargés de ce contrôle. Chaque problème, une fois détecté et identifié, a donné lieu ensuite à un contrôle systématique.

Conformément à la responsabilité du Service Ésane de Nantes sur la validation des données issues du processus de production (qui passe par la pertinence des procédures, leur bonne application et un contrôle sur la vraisemblance des données produites), le contrôle de données individuelles (§ *10.4*) est plutôt assuré par Nantes et celui de données agrégées (§ *10.5*) par la division Ésane à Paris.

La philosophie générale de ces contrôles est toujours la même : à partir d'une anomalie suspectée, on cherche la ou les entreprises principalement responsables de l'anomalie ; on regarde alors si celle-ci a été **déjà repérée et justifiée par un gestionnaire** au préalable. Lorsque ce n'est pas le cas, on cherche à identifier l'erreur (le plus souvent une estimation de non-réponse ou un redressement automatique inapproprié, un décalage de saisie, un programme qui s'est mal exécuté sans qu'on s'en aperçoive etc.) **et on corrige**. En cas de doute, on demande à un gestionnaire de contacter l'entreprise pour lever le doute.

Il va de soi que cette phase de contrôle est itérative : on **relance un calcul complet** pour tenir compte des premières corrections qui ont été effectuées ; la correction des premières "grosses" erreurs, en fait apparaître ensuite de nouvelles qui étaient jusque là masquées ; le processus ne s'arrête que lorsque toutes les anomalies apparentes sont expliquées ou corrigées.

10.4 Le contrôle sur données individuelles

Le principe des contrôles est toujours le même : les programmes permettent de lister les entreprises qui sont dans tel ou tel cas. C'est l'indice d'une anomalie. Mais l'important est surtout de corriger celles qui ont un réel impact sur les agrégats et qui seront détectées par la phase "d'output editing" qu'on verra au paragraphe suivant. Il est en effet impossible, pour des raisons de charge de travail, de contrôler toutes les données individuelles suspectes, car il faut regarder à chaque fois si l'anomalie a été ou non repérée et déjà validée par un gestionnaire ; c'est pourquoi cette phase se concentre sur le contrôle des anomalies qui sont les plus importantes et qui empêchent le bon fonctionnement du contrôle des agrégats. Les principaux contrôles effectués (par ordre d'importance décroissante) sont les suivants.

[405] Id est le fait qu'une variable « dont » doit rester inférieure à l'ensemble dont elle est issue.
[406] On rappelle que l'on parle de milliards de données individuelles et de centaines de milliers d'agrégats.

10.4.1 Valeurs douteuses en sortie de Rédi

Comme on l'a vu au chapitre 8, c'est la liasse fiscale qui est prioritaire pour le chiffre d'affaires et l'enquête pour sa ventilation (lorsque l'entreprise répond à l'enquête) ; lorsque l'une ou l'autre est absente cet arbitrage "normal" est modifié et cela peut entraîner des aberrations lorsque la réponse à l'Esa (ou son imputation) est incohérente avec celle de la liasse fiscale (ou de son imputation). On cherche donc à repérer ces cas, d'autant qu'ils peuvent être **amplifiés par le poids de sondage**[407]**,** en s'intéressant plus spécifiquement à la valeur ajoutée[408] (VA) ou à des variables liées (chiffre d'affaires ou taux de marge).

Pour repérer ces anomalies, on peut s'intéresser par exemple au fait que la valeur arbitrée d'une variable après Rédi soit très différente de la même variable issue de la liasse fiscale. On effectue ainsi les comparaisons suivantes :

- valeur ajoutée post Rédi comparée à la VA issue des liasses fiscales[409];
- on calcule l'écart entre la valeur ajoutée post rédi et la valeur ajoutée issue des liasses fiscales, à la fois en N et N-1 : s'il a beaucoup varié entre N et N-1, cela signifie que le traitement effectué lors de Rédi est différent de celui fait l'année précédente. Ceci permet d'identifier des incohérences temporelles qui s'expliquent par le fait que l'étape Rédi des traitements dans Ésane ne tient pas compte des évolutions annuelles des variables, ce qui peut provoquer ces ruptures[410] ;

Ces deux contrôles sont systématiques et les plus utilisés. On regarde également les plus gros écarts liés aux comparaisons suivantes :

- taux de marge de l'entreprise avant et après Rédi ;
- taux de valeur ajoutée avant et après Rédi.

De même l'arbitrage sur l'emploi peut conduire à des incohérences entre effectifs au 31/12 et effectifs ETP. On regardera donc les unités telles que :

- l'effectif ETP après Rédi est vide ou nul alors que celui au 31/12 ne l'est pas ;
- l'effectif ETP après Rédi est nul en N alors qu'il ne l'est pas en N-1.

10.4.2 Le cas des entreprises profilées issues des grands groupes

Comme on le verra au *chapitre 11*, à compter de 2012, les grands groupes donnent lieu à un traitement particulier (appelé « profilage du groupe ») conduisant à considérer comme "entreprises" (qu'on appelle du coup "profilées") des ensembles d'unités légales filiales de ces grands groupes, dont la réunion forme un "acteur économique" correspondant à la définition économique d'une entreprise. Les données correspondantes (consolidées Cf. § *11.5.2*) sont alors collectées directement auprès de ces grands groupes par des membres d'une division particulière de la direction générale à Paris (les « profileurs »), et réinjectés ensuite dans la chaîne de production Ésane. Par ailleurs, pour les seuls besoins de la comptabilité nationale, afin de ne pas introduire de rupture dans les séries, on continue de fournir aux comptables nationaux un fichier des résultats d'Ésane en unités légales.

---

[407] La configuration typique est la suivante : l'année N, la liasse fiscale d'une entreprise, qui rentre dans le champ de l'Esa en N et qui y répond, est absente et son CA fiscal mal imputé, car on ne dispose pas non plus de la liasse en N-1. On a vu alors au § *8.3.3.1* que Rédi prend le chiffre d'affaires de l'enquête. Si celui-ci est beaucoup plus fort que le CA fiscal imputé, cela va avoir pour effet d'accroître énormément la VA de cette entreprise. Si on a la malchance que cette unité soit tirée avec un gros poids, lui-même gonflé par la procédure de calage (Cf. § *5.6.2.4*) (cela peut aller jusqu'à plus de 1000), on peut aboutir à des aberrations (impacts de plusieurs milliards d'euros) qu'il faut corriger avant même de se lancer dans le contrôle des agrégats, car elles masquent tout le reste.

[408] Comme on l'a rappelé dans l'*avant-propos*, un des grands objectifs d'Ésane est de calculer la part de PIB attribuable aux entreprises. C'est pourquoi la VA est peut-être la variable la plus importante et la plus contrôlée.

[409] Ce contrôle permet par exemple de détecter le cas typique de la note précédente.

[410] Par exemple Rédi peut retenir la liasse fiscale comme source du CA l'année N-1 puis choisir l'Esa l'année suivante alors que ces deux grandeurs peuvent être très différentes. C'est notamment le cas lorsque la liasse fiscale a fait l'objet d'une imputation l'une des deux années. En effet cela signifie que le CA ESA en N est très éloigné du CA de la liasse en N-1 ce qui n'a pas forcément raison d'être sauf si on est dans le cas d'une restructuration. C'est donc bien une conjonction d'éléments qui fait que l'on arrive à quelque chose d'aberrant et pas uniquement le fait de changer de source dans REDI.

Plusieurs programmes ont été élaborés afin d'effectuer des contrôles sur ces entreprises issues du profilage et les unités légales qui les composent. En effet, plusieurs caractéristiques doivent être vérifiées pour que les données soient valides :

- la valeur ajoutée d'une entreprise profilée doit être égale à la somme de celle des unités légales qui la composent (même critère pour l'investissement) ;
- le chiffre d'affaires des unités légales faisant partie du contour d'une profilée doit être vraisemblable en niveau et en évolution.

Les contrôles permettent de s'assurer que ces critères sont respectés. Si ce n'est pas le cas, on revient vers les « profileurs » afin de comprendre pourquoi les données ne satisfont pas aux contrôles.

10.4.3 Recherche d'autres types d'anomalies

Les programmes de contrôle développés permettent de repérer des cas de **non-respect de certaines règles comptables** (du type total différent de la somme de ses composants, actif du bilan différent du passif, chiffre d'affaires à l'export supérieur au chiffre d'affaires total, investissement net hors apports supérieur à l'investissement brut hors apports, etc.).

Ils permettent également de repérer les entreprises dont le **chiffre d'affaires est fortement négatif**, ce qui traduit souvent un remplissage erroné de la liasse fiscale, voire **exprimé dans la mauvaise unité** (euros ou millions d'euros à la place de k€) par comparaison avec l'année précédente.

10.5 Le contrôle des agrégats

La plupart des utilisateurs s'intéressent à la fois **au total** d'une variable sur un secteur (ou une branche) ainsi très souvent qu'**au taux d'évolution annuelle** correspondant. Comme on l'a déjà mentionné, certaines variables comme le chiffre d'affaires, la valeur ajoutée ou encore l'emploi sont plus particulièrement centrales pour la plupart d'entre eux.
Les outils de contrôle qui ont été développés s'intéressent donc à la fois à l'analyse des taux de variation qu'on détaillera en premier au § *10.5.1* (parce que c'est elle qui permet le mieux de détecter des anomalies) et à la vraisemblance des totaux qu'on étudiera au § *10.5.2*.

10.5.1 L'analyse des taux de variation par « output editing[411] »

L'idée est de s 'appuyer sur les deux types de tableaux suivants :

- Un premier (global) permet avant tout « d'expliquer » le taux de croissance des variables que l'on veut contrôler **à partir du total tous secteurs, en descendant l'arbre d'agrégation** jusqu'au **niveau groupe** de la NAF et en calculant la contribution de chaque niveau dans la variation du niveau supérieur. On peut évidemment classer les secteurs par ordre de contribution décroissante ; ainsi, les fichiers Excel d'aide à l'analyse permettent d'identifier les secteurs à l'origine de l'évolution des principales variables produites par Ésane successivement aux niveaux A17 (sections), A38 et A272 (groupes) de la Naf. Ils permettent également d'analyser les **révisions** intervenues **entre deux publications successives** d'Ésane.
- Lorsqu'on a détecté un ou plusieurs secteurs, au niveau groupe, dont la contribution à l'évolution totale est suspecte, un deuxième type de tableaux (détaillés) permet de **classer les entreprises** selon leur contribution à l'évolution du groupe et ainsi de détecter rapidement si une unité est plus particulièrement responsable de cette évolution. Ces contributions sont différentes de celles des macro-contrôles (décrits au § *2.7*) qui cherchaient à mettre en évidence des unités ayant un poids économique important et un comportement atypique. Ces nouvelles contributions de la phase d'*output editing* se basent sur **l'agrégat diffusé** (différent des agrégats de production) et cherchent à identifier les unités les plus influentes dans l'évolution de cet agrégat, i.e. en tenant compte de **la formule spécifique de l'estimateur par différence en N et N-1**. Ainsi, sont par exemple mises en évidence à cette étape les unités :
    - présentant un poids calé important et changeant d'APE ou fortement affectées par Redi (en effet, dans une telle configuration, leur contribution aux agrégats par

[411] De façon générale, le « *data* editing » désigne en anglais l'activité consistant à détecter puis à corriger les incohérences « logiques » qui se sont introduites dans un processus de traitement de données. On a déjà vu les macro-contrôles qui en font partie (qu'on appelle aussi « *selective* editing ») qui permettent de contrôler les données au fil de l'eau. L'« *output* editing » renvoie à des techniques analysant ce qui « sort » du process en fin de course, pour nous les agrégats d'Ésane.

différence sera importante au vu de la formule de l'estimateur composite, cf. formule 2 du § 9.2.) ;

- ou présentes dans l'échantillon en N et pas en N-1 (ou réciproquement) qui voient de ce fait leur contribution fortement évoluer entre les 2 années à cause de la prise en compte du poids calé une seule des deux années.

Certaines unités mises en évidence durant la phase d'*output editing* ont bien sûr déjà pu être contrôlées durant la phase de *selective editing* et dans ce cas, aucun travail complémentaire n'est nécessaire.

Plus précisément les sorties éditées comprennent :

- **une page de synthèse** des résultats au niveau A17 présentant les évolutions du nombre d'unités légales (ou d'entreprises), des effectifs ETP, des effectifs au 31/12, du chiffre d'affaires, de la valeur ajoutée, de l'EBE, de la masse salariale et de l'investissement corporel hors apport ;
- **une page d'analyse des révisions** au niveau A17 entre la publication semi-définitive et la version définitive d'Esane. Cette analyse porte sur les niveaux et sur les évolutions et concerne toutes les variables citées ci-dessus ;
- plusieurs pages **d'analyse de la croissance** des variables principales. Ici l'objectif est de s'assurer qu'on est capable d'expliquer l'évolution globale de l'économie à partir de l'analyse de la contribution des secteurs d'activité qui la compose.
- pour la valeur ajoutée et les salaires une page **d'analyse fine des révisions** au niveau groupe est également disponible : on y trouve les révisions en niveau et en points de croissance. ;
- une page fournit simplement **les résultats sectoriels du niveau division jusqu'au niveau groupe** pour les principales variables issues d'Esane (les mêmes que pour la page synthèse). L'objectif ici est d'effectuer une analyse de type « top down » c'est-à-dire d'identifier le(s) secteur(s) problématique au niveau A17 dans la feuille de synthèse puis de creuser à des niveaux de plus en plus fin jusqu'à repérer le(s) secteur(s) à l'origine des résultats macro-économiques ;
- une page propose une décomposition du compte des entreprises sous la forme de **soldes intermédiaires de gestions**. L'utilisateur peut choisir le secteur qui l'intéresse dans un menu déroulant y compris au niveau 5 caractères de la Naf (niveau dit « APE »).

Le même type de fichier existe pour l'analyse des secteurs de l'économie par tranches de taille.

10.5.1.1 Un exemple de tableau global d'analyse de la croissance

Pour une variable Y, le tableau fournit pour chaque secteur au niveau groupe de la Naf : les montants (en millions d'€) pour les années N (Y^N) et N-1 (Y^{N-1}), l'évolution annuelle entre N et N-1 en pourcentage et la contribution du secteur à l'évolution de l'ensemble de l'économie ; compte tenu de ce que $Y^N = \Sigma Y_s^N$ où Y_s^N désigne le total de Y sur le secteur s à l'année N, il est facile de vérifier que :

$\frac{\Delta Y}{Y^{N\text{-}1}} = \sum_s \frac{\Delta Y_s}{Y^{N\text{-}1}}$. La contribution de chaque secteur au taux de variation du total vaut donc $\frac{\Delta Y_s}{Y^{N\text{-}1}}$

Ensuite, on cherche à sélectionner les principaux secteurs à valider. Pour cela :

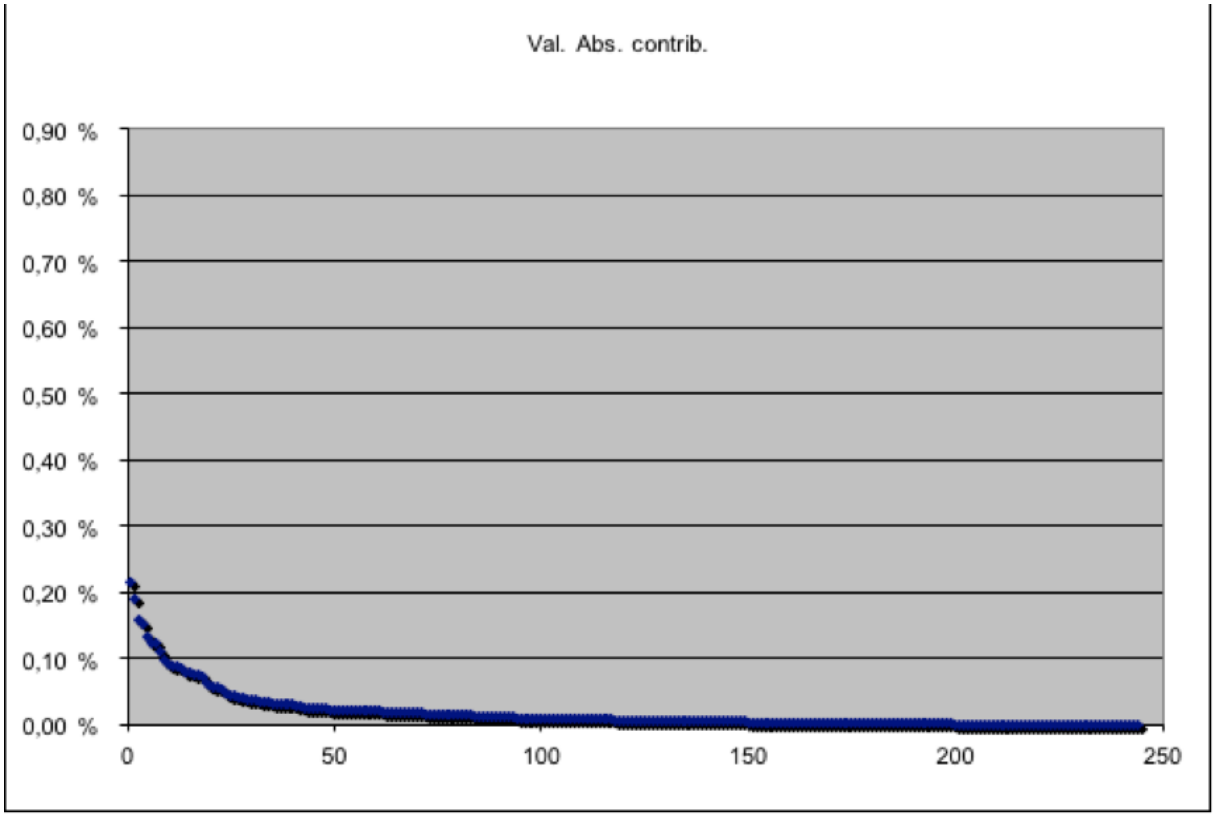
- on calcule d'abord pour chaque secteur, la **valeur absolue** de sa contribution à la croissance,
- ensuite on classe les secteurs par contribution absolue décroissante,
- enfin on attribue un niveau de priorité des contrôles (de N1 à N4, correspondant aux quartiles de la distribution), les secteurs les plus prioritaires devant nécessairement faire l'objet d'une vérification. Idéalement, on cherche à contrôler les secteurs qui expliquent 75% de la somme des contributions absolues à la croissance, soit les priorités N1 à N3).

Un exemple d'un tel tableau est reproduit ci-après dans le tableau 1:

Tableau 1 : Analyse de la croissance globale (pour la variable « Valeur ajoutée (R003)») (extrait)

Valeur ajoutée		R003		Milliards d'€					
secteur	libellé	N-1	N	évol.	contrib	Val. Abs. contrib.	part de l'évol. aboslue	part cumulée de l'évol.	Priorité de contrôle
total	total	1 058	1 082	2,2 %	2,22 %	4,48 %	100,00 %	100,00 %	classement
H491	Transport ferroviaire interurbain de voyageurs	11,30	9,03	-20,1 %	-0,21 %	0,21 %	4,79 %	4,79 %	N1
H522	Services auxiliaires des transports	22,55	24,58	9,0 %	0,19 %	0,19 %	4,27 %	9,05 %	N1
C212	Fabrication de préparations pharmaceutiques	9,20	10,89	18,4 %	0,16 %	0,16 %	3,56 %	12,62 %	N1
C291	Construction de véhicules automobiles	9,59	11,21	16,9 %	0,15 %	0,15 %	3,41 %	16,03 %	N1
C192	Raffinage du pétrole	1,59	3,00	88,5 %	0,13 %	0,13 %	2,97 %	19,00 %	N1
C201	Fabrication de produits chimiques de base, de produits azotés et	7,40	8,72	17,8 %	0,12 %	0,12 %	2,78 %	21,78 %	N1
N782	Activités des agences de travail temporaire	21,99	23,28	5,9 %	0,12 %	0,12 %	2,72 %	24,50 %	N1
M701	Activités des sièges sociaux	10,81	11,98	10,7 %	0,11 %	0,11 %	2,45 %	26,95 %	N2
J620	Programmation, conseil et autres activités informatiques	28,22	29,28	3,7 %	0,10 %	0,10 %	2,22 %	29,17 %	N2
M702	Conseil de gestion	14,33	15,28	6,7 %	0,09 %	0,09 %	2,01 %	31,19 %	N2
C303	Construction aéronautique et spatiale	11,88	12,83	8,0 %	0,09 %	0,09 %	2,00 %	33,19 %	N2
J612	Télécommunications sans fil	5,35	6,30	17,7 %	0,09 %	0,09 %	2,00 %	35,19 %	N2
F439	Autres travaux de construction spécialisés	18,54	17,63	-4,9 %	-0,09 %	0,09 %	1,93 %	37,12 %	N2
G464	Commerce de gros de biens domestiques	24,84	25,68	3,4 %	0,08 %	0,08 %	1,77 %	38,89 %	N2
D351	Production, transport et distribution d'électricité	24,82	23,99	-3,4 %	-0,08 %	0,08 %	1,76 %	40,65 %	N2
K663	Services de gestion de fonds	5,16	5,96	15,6 %	0,08 %	0,08 %	1,70 %	42,34 %	N2
F433	Travaux de finition	18,29	17,50	-4,3 %	-0,08 %	0,08 %	1,68 %	44,02 %	N2
Q862	Activité des médecins et des dentistes	28,49	29,25	2,6 %	0,07 %	0,07 %	1,59 %	45,61 %	N2
G467	Autres commerces de gros spécialisés	22,62	23,30	3,0 %	0,06 %	0,06 %	1,43 %	47,04 %	N2
H511	Transports aériens de passagers	5,23	5,87	12,2 %	0,06 %	0,06 %	1,35 %	48,39 %	N2
Q869	Autres activités pour la santé humaine	17,52	18,13	3,4 %	0,06 %	0,06 %	1,27 %	49,66 %	N2
G451	Commerce de véhicules automobiles	8,65	9,23	6,8 %	0,06 %	0,06 %	1,23 %	50,90 %	N3
M691	Activités juridiques	15,12	15,69	3,8 %	0,05 %	0,05 %	1,20 %	52,09 %	N3
J582	Édition de logiciels	6,15	6,65	8,2 %	0,05 %	0,05 %	1,06 %	53,16 %	N3
G479	Commerce de détail hors magasin, éventaires ou marchés	3,13	3,61	15,3 %	0,05 %	0,05 %	1,01 %	54,16 %	N3
N773	Location et location-bail d'autres machines, équipements et biens	6,45	6,91	7,1 %	0,04 %	0,04 %	0,97 %	55,13 %	N3
C332	Installation de machines et d'équipements industriels	5,32	4,88	-8,4 %	-0,04 %	0,04 %	0,94 %	56,08 %	N3

La VA totale croit de 2,2 % de 1058 G€ à 1082 G€. On voit que le groupe 49.1[412] (Transport ferroviaire interurbain de voyageurs) diminue de 20,1 %. Il contribue à la hausse de 2,2 % pour -0,21%. Si on prend toutes les contributions en valeur absolue, on voit que la somme des contributions ne fait plus 2,2 % mais 4,48 % (parce qu'il y a beaucoup de contributions positives et négatives qui se compensent). Le graphique ci-dessous permet de repérer visuellement la contribution de l'ensemble des secteurs.



[412] Qui appartient à la section H "Transports et entreposage " d'où le H491 du tableau…

Un tel tableau, qui met en évidence à la fois les secteurs qui évoluent le plus et ceux qui contribuent le plus à la croissance globale, permet de cibler très vite en priorité les secteurs pouvant poser problème. L'expérience montre qu'au début de cette phase d'analyse et de validation, les secteurs semblant contribuer le plus fortement à l'évolution du chiffre d'affaires ou de la valeur ajoutée des entreprises **cachent très souvent des traitements erronés** plutôt que de réelles évolutions économiques. L'approche top down permet de repérer rapidement les plus gros problèmes. Reste à trouver ensuite quelles sont les entreprises responsables ; c'est l'objet du second type de tableau.

10.5.1.2 Un exemple de tableau détaillé d'analyse de la croissance

Il s'agit maintenant de passer **du niveau groupe aux entreprises** qui sont responsables de cette évolution. Le principe est le même que précédemment : on va calculer la contribution de chaque entreprise dans le calcul de la variation de l'agrégat de niveau groupe qui nous intéresse. La seule difficulté est qu'il est plus difficile d'exhiber littéralement la formule exacte de la contribution de chaque entreprise puisque la formule (1 bis) du § *9.2.1* (qu'on reprendra au paragraphe 10.5.2) montre que les entreprises y contribuent différemment selon qu'elles sont dans le sous-champ 1 ou pas, pérennes ou pas, dans l'exhaustif ou pas, etc. Y participent toutes les entreprises qui ont appartenu au groupe G en APE de lancement ou en cours d'enquête, dans l'un quelconque des sous-champs Ésane, soit à l'année N-1 soit à l'année N. Mais il est possible numériquement de calculer exactement la contribution de chaque entreprise. On obtient alors un tableau du type suivant (tableau détaillé) :

Tableau 2[413] : analyse de la contribution des entreprises à la croissance d'une variable au niveau groupe

les 100 plus fortes contributions (absolues) à la croissance du chiffre d'affaires (R310) secteur 472															
	(Commerce de détail alimentaire en magasin spécialisé)														
Évolution secteur	3,50%														
total des 100 plus grosses	8,40%														
siren	poids N	poids N-1	APE Lancmt N	APE Diffus. N	APE Lancmt N-1	APE Diffus. N-1	contrib ind_evol R003	agregat_R310 N-1	ctr_niv R310 N-1	Rédi R310 N-1	R310 Initial N-1	agregat_R310 N	ctr_niv R310 N	Rédi R310 N	R310 Initial N
siren1	3059		472	472	472	472	8,3%	18 041 962	915	915	915	18 694 040	1 553 662	608	100
siren2		169	472	463	472	463	2,8%	18 041 962	-524 500	3 114	3 114	18 694 040	5 823	5 823	5 823
siren3	643		472	101	472	472	-1,1%	18 041 962	321	321	321	18 694 040	-207 085	323	323
siren4		233	472	472	472	101	1,1%	18 041 962	-202 892	874	874	18 694 040	839	839	839
siren5	272		472	101	472	472	-0,9%	18 041 962	670	670	670	18 694 040	-170 920	630	630
siren6		419			476	472	-0,9%	18 041 962	159 583	381	381	18 694 040	0		
siren7	18	12	472	472	472	472	-0,8%	18 041 962	-352 042	4 206	36 740	18 694 040	-504 881	4 105	33 646
siren8	176	580	478	478	478	472	-0,8%	18 041 962	144 959	250	250	18 694 040	0	252	252
siren9	64	72	472	472	472	479	0,7%	18 041 962	-137 336	1 926	1 926	18 694 040	2 174	2 173	2 173
siren10		225	472	472	472	101	0,7%	18 041 962	-127 076	567	567	18 694 040	570	570	570
siren11	414	0	478	472	478	472	0,6%	18 041 962	0	205	205	18 694 040	106 280	257	257
siren12	277		472	478	478	478	-0,6%	18 041 962	0	484	484	18 694 040	-103 693	375	375
siren13	17	26	472	472	472	471	0,5%	18 041 962	-97 966	3 877	3 877	18 694 040	3 889	3 889	3 889
siren14		193	472	472	472	101	0,5%	18 041 962	-100 159	521	521	18 694 040	475	475	475
siren15		570	472	472	472	101	0,5%	18 041 962	-98 286	173	173	18 694 040	174	174	174
siren16	182		472	101	472	472	-0,5%	18 041 962	532	532	532	18 694 040	-96 907	536	536
siren17		333	472	472	472	101	0,5%	18 041 962	-96 191	290	290	18 694 040	276	276	276
siren18	0	204	472	472	472	478	0,5%	18 041 962	-92 984	457	457	18 694 040	410	410	410
siren19	367		472	101	101	101	-0,5%	18 041 962	0	246	246	18 694 040	-90 117	246	246
siren20		407			476	472	-0,5%	18 041 962	90 088	222	222	18 694 040	0		
siren21	114		478	472	472	472	0,5%	18 041 962	879	879	879	18 694 040	90 145	789	789
siren22		205	472	472	472	478	0,4%	18 041 962	-82 725	405	405	18 694 040	408	408	408
siren23	179		472	101	472	472	-0,4%	18 041 962	48	48	48	18 694 040	-82 390	463	463
siren24	182		476	472	472	472	0,4%	18 041 962	434	434	434	18 694 040	82 470	453	453
siren25		148	472	472	472	101	0,4%	18 041 962	-79 476	541	541	18 694 040	559	559	559
siren26	18		471	472	472	472	0,4%	18 041 962	5 048	5 048	5 048	18 694 040	84 731	4 630	4 630
siren27			472	472	472	463	-0,4%	18 041 962	79 140	79 140	79 140	18 694 040	38	38	38
siren28		159	478	478	478	472	-0,4%	18 041 962	79 015	498	498	18 694 040	0	403	403
siren29		19	472	472	472	463	0,4%	18 041 962	-74 012	4 172	4 172	18 694 040	3 919	3 919	3 919
siren30		50	561	561	561	472	-0,4%	18 041 962	77 743	1 560	1 560	18 694 040	0	1 569	1 569
siren31	1	1	471	471	471	472	-0,4%	18 041 962	77 007	77 007	77 007	18 694 040	0	82 473	82 473

[413] Dans la réalité, le tableau comprend bien entendu les vrais numéros Siren, la raison sociale des entreprises et des variables de gestion et divers ratios supplémentaires. Il donne les 100 premières contributions [tronquées ici à 31 qui correspond à la première entreprise de l'exhaustif (poids égal à 1)].

Le tableau reproduit l'analyse de la croissance du chiffre d'affaires (variable « R310 ») sur le secteur du commerce de détail alimentaire en magasin spécialisé (47.2). On observe que celui-ci croît de 3,5 % et que les 100 premières contributions à cette variation globale le feraient même croître de 8,4 %.
Le tableau donne dans les deux premières colonnes le poids des unités légales. Dans un tel secteur très peu concentré, on s'aperçoit que les plus grosses contributions viennent d'entreprises échantillonnées dont les poids figurent en colonne 2 et 3 ; la première entreprise de l'exhaustif (poids = 1) n'apparaît qu'en 31e position. Quand le poids n'est pas rempli c'est simplement que l'entreprise n'a pas été tirée dans l'échantillon ; les poids nuls sont des non-réponses traitées par « repondération » (Cf. § 5.6.2.2). Les Siren7 et Siren8 sont des entreprises pérennes entre N et N-1 bien que ne faisant pas partie de l'exhaustif. Siren1 est entrante ; Siren2 est sortante.
Les 4 colonnes suivantes indiquent le secteur d'appartenance de l'entreprise en N-1 ou en N, en APE de lancement ou de diffusion (c'est-à-dire après prise en compte de la réponse de l'entreprise à l'enquête). Par définition le secteur 47.2 figure au moins dans l'une de ces 4 colonnes.
La 8ème colonne qui donne la contribution de l'entreprise à la variation de chiffre d'affaires du secteur est bien évidemment **la colonne cruciale**. On voit ici que le Siren1 fait augmenter à lui seul le CA du secteur de 8,3%. Pour un secteur très dispersé, le plus vraisemblable est qu'il s'agisse d'une erreur. De fait on constate que cette entreprise est interrogée en N avec un chiffre d'affaires initial de 100, puis de 608 après Rédi. Mais, certainement par un concours de circonstances dans le calage, elle se retrouve avec un poids de 3059. Du coup elle « pèse » 1 553 662 dans l'agrégat composite N [en appliquant la formule (1) du § 9.2.1 qui est d'ailleurs rappelée ci-dessous[414]]. Elle n'était pas interrogée en N-1 (poids non rempli) mais on avait bien sûr sa liasse. Comme elle n'était pas interrogée, son chiffre d'affaires initial, après Rédi et sa contribution à l'agrégat final étaient les mêmes : 915[415]. Pour finir, tout se passe, au niveau des agrégats, comme si son chiffre d'affaires avait augmenté de 915 à 1 553 662. Elle « tire » donc l'agrégat vers le haut, à cause de son poids ; si le traitement opéré par Rédi (redressement de la liasse fiscale en se fondant sur l'enquête) est correct, l'estimateur reste sans biais mais le fait d'avoir tiré cette entreprise dans l'échantillon fait grandement augmenter sa variance. Conformément à ce qu'on explique dans le § 10.2.2, les méthodologues seront certainement amenés à diminuer le poids de cette entreprise qui est excessif par une forme de "winsorisation"[416].

10.5.2 L'analyse de la vraisemblance des agrégats

L'analyse précédente est particulièrement bien adaptée pour détecter les secteurs, puis les unités au sein de ces secteurs, qui posent probablement problème. En revanche, lorsque ce n'est pas une entreprise aberrante unique qui est la cause du problème, ces tableaux ne suffisent pas toujours à l'identifier (ce peut être un traitement en production qui s'est mal déroulé, une estimation de valeurs manquantes qui se passe mal dans un secteur donné, etc.). C'est pourquoi on a mis en place, en parallèle des tableaux précédents, une analyse des agrégats en fonction des sous-populations de la formule (1 bis) du § 9.2.1 auxquelles appartiennent les entreprises qui contribuent à un agrégat (Cf. § 10.5.2.1) ou en fonction du type de processus qui affectent les données brutes au sein du système d'information Ésane (changement d'APE, réconciliation enquête/liasse) (Cf. § 10.5.2.2).

10.5.2.1 Analyse de l'impact de chaque sous-population sur un agrégat

On repart de la formule (1 bis) du § 9.2.1 donnant la valeur de l'agrégat composite d'une variable Y pour un groupe G (SC1 désigne le sous-champ 1, Dom les départements d'outre-mer, Champ 2-5 les sous-champ 2, 3, 5 et EAP0 les entreprises industrielles de moins de 20 salariés jusqu'en 2013[417]) :

$$\widehat{Y}_{\text{compo}}(G) = \widehat{Y}^{\text{SC1}}_{\text{compo}}(G) + \widehat{Y}^{\text{Sous-Champ 2_5}}_{\text{Rédi}}(G) + \widehat{Y}^{\text{Dom}}_{\text{Rédi}}(G) + \widehat{Y}^{\text{EAP0}}_{\text{Rédi}}(G) \qquad \text{(formule 1 bis)}$$

L'idée est d'évaluer dans quelle mesure cette valeur reflète celle que l'on aurait obtenue **en se limitant aux plus grosses entreprises** de chaque secteur, c'est-à-dire aux entreprises de l'exhaustif (**Exh**). Lorsque le secteur est fortement diversifié ou composé de nombreuses sociétés de petite taille, cette décomposition permet de repérer d'éventuels cas aberrants.

On repart donc de la valeur de $\widehat{Y}^{\text{SC1}}_{\text{compo}}(G)$ (formule 1) :

[414] Mais que le lecteur ne peut pas vérifier exactement car toutes les cases sont arrondies y compris les poids.
[415] D'après la formule (1) sa contribution est Y_i^{fiscal}.
[416] Ce cas étant tiré d'un cas réel, en l'espèce, le poids a été divisé par deux, ce qui permettait de retomber sur une variation cohérente avec la variation de chiffre d'affaires indiquée par les fichiers de TVA pour ce secteur et avec la consommation des ménages pour l'équilibre ressources-emploi correspondant. On retrouve là l'intérêt de la participation des comptables nationaux à la vérification des agrégats, mentionné dans le § 10.3.
[417] Cf. § 5.2.1 et 5.2.2

$$\widehat{Y}^{SC1}_{compo}(G) = \sum_{i \in U_1 \setminus HC\ externes} Y_i^{fiscal} \mathbb{I}_{groupe_rep=G}(i) + \sum_{i \in R} w_i Y_i^{Rédi} \mathbb{I}_{groupe_diff=G}(i) - \sum_{i \in R \oplus HC\ source\ Esa} w_i Y_i^{fiscal} \mathbb{I}_{groupe_rep=G}(i)$$

dans laquelle on va séparer les entreprises de l'exhaustif ($w_i = 1$) du reste, c'est-à-dire des entreprises dites échantillonnées :

$$\widehat{Y}^{Exh}_{compo} = \sum_{i \in Exh} Y_i^{Rédi} \mathbb{I}_{groupe_diff=G}(i)$$

[Quand on se limite à l'exhaustif, R (deuxième somme) désigne à la fois les répondantes et les non répondantes de l'exhaustif (qui sont extrapolées) (Cf. § 9.2) ; comme une entreprise de l'exhaustif est soit dans R, soit HC source externe, soit HC source Esa, U1\HC externes est égal à R⊕HC source Esa ; le premier et le dernier terme de la formule (1) s'éliminent donc ; finalement, il reste la somme ci-dessus pour l'ensemble des entreprises de l'exhaustif, correspondant à la deuxième somme].

Cela va former la **colonne [1] du tableau**.

Le complémentaire de l'exhaustif étant donc « la partie échantillonnée » (**Ech**), il est évidemment tentant de nommer le reste $\widehat{Y}^{Ech}_{compo}(G)$, même si la première somme de la formule (1) :

$$\sum_{i \in U_1 \setminus Exh \oplus HC\ externes} Y_i^{fiscal} \mathbb{I}_{groupe_rep=G}(i)$$

contient bien plus de termes que l'échantillon ne contient d'entreprises. On va toutefois mettre à part les entreprises de l'échantillon qui étaient dans G dans le répertoire et qui sont sorties du champ de l'Esa par la réponse à l'enquête et qu'on va appeler $\widehat{Y}^{Sor_champ}_{fiscal}(G)$, de façon à mettre dans la même somme toutes les entreprises répondantes. On va donc avoir :

$$\widehat{Y}^{SC1}_{compo}(G) = \widehat{Y}^{Exh}_{compo}(G) + \widehat{Y}^{Ech}_{compo}(G) + \widehat{Y}^{Sor_champ}_{fiscal}$$

avec :

$$\widehat{Y}^{Ech}_{compo} = \sum_{i \in U_1 \setminus Exh \oplus HC\ externes} Y_i^{fiscal} \mathbb{I}_{groupe_rep=G}(i) + \sum_{i \in R} [w_i Y_i^{Rédi} \mathbb{I}_{groupe_diff=G}(i) - w_i Y_i^{fiscal} \mathbb{I}_{groupe_rep=G}(i)]$$

et : $\widehat{Y}^{Sor_champ}_{fiscal}(G) = - \sum_{i \in HC\ source\ Esa} w_i Y_i^{fiscal} \mathbb{I}_{groupe_rep=G}(i)$

Au final on décompose l'estimateur composite $\widehat{Y}_{compo}(G)$ sous la forme d'une somme de six termes :

$$\widehat{Y}_{compo}(G) = \widehat{Y}^{Exh}_{compo}(G) + \widehat{Y}^{Ech}_{compo}(G) + \widehat{Y}^{Sous-Champ\ 2_5}_{Rédi}(G) + \widehat{Y}^{Dom}_{Rédi}(G) + \widehat{Y}^{EAP0}_{Rédi}(G) + \widehat{Y}^{Sor_champ}_{fiscal}(G)$$

= [1] + [2] + [3] + [4] + [5] + [6]

C'est ce que l'on retrouve exactement dans le tableau suivant (dans lequel la colonne "EAP0" est à blanc, car les chiffres proviennent d'une campagne postérieure à 2013 dont on a vu au § 5.2.2 que c'était la dernière campagne pour laquelle ce champ spécifique a été utilisé dans le calcul des agrégats) :

Tableau 3 : décomposition de la valeur ajoutée selon les diverses sous-populations d'entreprises (**extrait**)

Variable	Valeur ajoutée R003							
	unité	milliers d'€						
	année	N						
Secteur		Agrégat Esane =[1]+[2]+[3]+[4]+[5]+[6]	Contribution des entreprises du sous-champ 1		Entreprises des sous-champs 2 à 5 [3]	Entreprises des DOM [4]	Entreprises industrielles < 20 salariés [5]	Entreprises passées du champs de l'Enquête aux sous-champs 2 à 5 [6]
			Partie exhaustive de l'échantillon [1]	Autres unités du sous-champ 1 [2]				
tot	total	1 080 610 099	701 924 351	261 990 306	#########	########	-	1 039 232
C325	Fabrication d'instruments et de fournitures à usage médical et dentaire	566 234	516 277	46 801	25	3 131	-	.
C329	Activités manufacturières n.c.a.	7 847 018	6 315 470	1 454 631	1 986	72 964	-	1 968
C331	Réparation d'ouvrages en métaux, de machines et d'équipements	4 876 287	3 951 080	886 122	3 683	35 402	-	.
C332	Installation de machines et d'équipements industriels	23 990 049	21 448 070	2 013 004	334 292	179 198	-	15 484
D351	Production, transport et distribution d'électricité	2 758 844	2 751 940	1 337	1 357	4 210	-	.
D352	Production et distribution de combustibles gazeux	1 918 917	1 867 433	47 870	.	3 614	-	.
D353	Production et distribution de vapeur et d'air conditionné	3 428 300	2 560 237	21 273	776 777	70 014	-	.
E360	Captage, traitement et distribution d'eau	1 196 277	675 123	156 184	343 319	21 650	-	.
E370	Collecte et traitement des eaux usées	2 269 549	2 093 675	86 918	-184	90 273	-	-1 133
E381	Collecte des déchets	2 089 635	1 920 757	117 709	-2	51 170	-	.
E382	Traitement et élimination des déchets	1 581 623	1 439 394	102 045	276	39 451	-	458
E383	Récupération	465 326	373 128	86 944	.	5 254	-	.
E390	Dépollution et autres services de gestion des déchets	3 266 214	1 754 290	1 461 191	1 149	49 548	-	36
F411	Promotion immobilière	6 903 650	4 032 266	2 571 473	17 585	282 326	-	.
F412	Construction de bâtiments résidentiels et non résidentiels	5 008 156	4 640 716	321 398	133	45 908	-	.
F421	Construction de routes et de voies ferrées	3 795 894	3 031 642	657 649	5	106 599	-	.
F422	Construction de réseaux et de	1 182 248	778 260	371 752	5 574	26 662	-	.
F429	Construction d'autres ouvrages de génie civil	4 331 642	1 744 117	2 379 806	4 371	203 348	-	.
F431	Démolition et préparation des	22 754 177	10 440 559	11 989 555	35 877	288 186	-	.
F432	Travaux d'installation électrique, plomberie et autres travaux d'installation	17 495 831	2 888 236	14 406 426	25 972	175 198	-	.
F433	Travaux de finition	17 626 581	4 319 814	12 952 768	18 092	335 907	-	.
F439	Autres travaux de construction spécialisés	9 231 090	7 031 279	1 889 699	674	309 439	-	.
G451	Commerce de véhicules	6 200 639	638 188	5 447 951	3 656	110 845	-	.
G452	Entretien et réparation de véhicules automobiles	4 143 026	2 659 528	1 341 405	32	142 060	-	.

Ce tableau (partiel ; on n'a pas recopié toutes les lignes, car il y en aurait 244) permet de constater que :

- Les grandes entreprises expliquent en moyenne les 2/3 du résultat final.
- Mais ce n'est évidemment pas vrai pour tous les secteurs et le tableau met directement en évidence les secteurs très concentrés (sur l'extrait ci-dessus D351 et D352, production et transport d'électricité ou de gaz, E381 et E382 collecte ou traitement des déchets, F421 construction de routes et de voies ferrées, etc.) ; mais on voit aussi des secteurs où à l'inverse les petites entreprises pèsent bien davantage (F433 travaux de finition, G452 réparation de véhicules automobiles etc..).
- Les entreprises des sous-champs 2 à 5 comme celles des DOM pèsent généralement peu.
- Les sorties de champ sont marginales.

La « structure » d'un tel tableau est évidemment très stable d'une année sur l'autre ; la comparaison de tels tableaux d'une année sur l'autre (on peut évidemment calculer les évolutions N/N-1) permet d'identifier très rapidement des secteurs et des sous-populations où des événements exceptionnels ont pu se produire.

10.5.2.2 Analyse de l'impact des processus sur un agrégat

La formule (2) qu'on a vue au § 9.2 pour l'estimation relative au sous champ 1 et qu'on rappelle ici, permet de décomposer les agrégats d'une autre manière en isolant les effets changements d'APE et réconciliation des données par le processus Rédi :

$$+\sum_{i\in R} w_i Y_i^{Redi} * \left[\mathbb{I}_{groupe_diff=G}(i) - \mathbb{I}_{groupe\text{-}reg=G}(i)\right]$$

Effet changements d'APE

$$+\sum_{i\in R} w_i \mathbb{I}_{groupe_reg=G}(i) * \left[Y_i^{Redi} - Y_i^{fiscal}\right]$$

Effet Redi

$$-\sum_{i\in HC\ source\ Esa} w_i Y_i^{fiscal} \mathbb{I}_{groupe_rep=G}(i)$$

Effet sorties de champ Esane

formule qu'on peut écrire avec des notations évidentes :

$$\hat{Y}_{compo}^{SC1} = \hat{Y}_{fiscal}^{SC1} + \hat{Y}_{chgmt-APE}^{SC1} + \hat{Y}_{Rédi}^{SC1} + \hat{Y}_{Sortie-Chmp_Esane}^{SC1}$$

Les autres composantes de l'estimateur composite, par définition, ne peuvent être affectées par un changement d'APE dû à l'enquête puisque les entreprises correspondantes ne sont pas interrogées ; mais elles sont passées par le programme Rédi-balai qui a pu modifier leurs caractéristiques de liasses initiales. Comme pour le sous-champ 1, on peut calculer cet effet « Rédi[418] » en faisant explicitement apparaître sa variation par rapport au niveau initial des liasses pour chaque composante sous-champ 2 à 5, Dom ou EAP0 :

$$\hat{Y}_{Rédi}^{champ} = \sum_{i\in champ} Y_i^{fiscal} + \left(\sum_{i\in champ} Y_i^{Rédi\text{-}balai} - \sum_{i\in champ} Y_i^{fiscal}\right)$$

soit

$$\hat{Y}_{Rédi}^{champ} = \hat{Y}_{fiscal}^{champ} + \hat{Y}_{Eff-Rédi}^{champ}$$ pour champ = sous-champ 2_5, Dom et EAP0.

Il suffit alors de regrouper les termes formés par le niveau des liasses fiscales et les termes dus à l'effet Rédi pour aboutir à

$$\hat{Y}_{compo}(G) = \hat{Y}_{fiscal}(G) + \hat{Y}_{Eff\text{-}Rédi}(G) + \hat{Y}_{Eff-chgmt_APE}^{SC1} + \hat{Y}_{Sortie-Chmp_Esane}^{SC1}$$

= [1] + [2] + [3] + [4]

Les deux premiers termes sont calculés sur tout le champ d'Ésane, les deux derniers sur le seul sous-champ 1. Un exemple en est donné dans le tableau 4 ci-dessous[419] ; bien entendu dans la réalité on peut aussi regarder ce tableau en variation de N par rapport à N-1, avec d'autres niveaux de nomenclature, en comparant les agrégats du définitif et des préliminaires, etc.

[418] Ce faisant, on regroupe un effet Rédi pur avec les effets de Rédi balai mais dont les principes sont, à peu de choses près, les mêmes.

[419] Ces tableaux sont utilisés pour contrôler la vraisemblance des agrégats. Mais en fin de processus, ce qui est le cas du tableau ci-dessus, ils donnent une estimation intéressante des effets Rédi et changement d'APE « en soi », utile par exemple aux comptables nationaux pour comprendre les évolutions sectorielles.

Tableau 4 : Analyse de l'impact des processus sur un agrégat (ici valeur ajoutée en nomenclature de sections)

Décomposition des agrégats composites en millions d'€
année en cours : N
Variable Valeur ajoutée

		N				
		Niveau source fiscale	Correction APE	Correction REDI	sorties du champ Ésane	**Agrégat composite**
		[1]	[2]	[3]	[4]	
	Total tous secteurs	1 078 039,8	364,4	3 272,5	-290,3	**1 081 386,4**
	Activités marchandes non agricoles et non financières	1 061 042,3	214,4	3 272,5	-290,3	**1 064 238,9**
A5.BE	**Industrie**	262 833,7	-3 789,6	85,7	-66,8	**259 063,0**
A17.DE	Industries extractives, énergie, eau, gestion des déchets et dépollution	41 527,9	114,0	329,7	-15,4	**41 956,1**
A17.C1	Fabrication de denrées alimentaires, de boissons et de produits à base de tabac	38 610,1	865,0	202,9	-4,4	**39 673,6**
A17.C2	Cokéfaction et raffinage	10 393,7	-7 396,3	0,0		**2 997,4**
A17.C3	Fabrication d'équipements électriques, électroniques, informatiques ; fabrication de machines	31 571,7	1 208,9	-613,1	-6,6	**32 160,9**
A17.C4	Fabrication de matériels de transport	33 690,2	-102,6	-7,3	-0,8	**33 579,5**
A17.C5	Fabrication d'autres produits industriels	107 040,2	1 521,3	173,5	-39,6	**108 695,4**
A5.FZ	**Construction**	81 267,3	686,8	420,8	-10,5	**82 364,4**
A5.GU	**Services principalement marchands**	667 449,3	3 168,6	2 759,2	-213,0	**673 164,1**
A17.G	Commerce ; réparation d'automobiles et de motocycles	206 878,5	2 512,1	135,3	-62,0	**209 463,9**
A17.H	Transports et entreposage	85 837,3	-86,3	-154,3	-2,3	**85 594,4**
A17.I	Hébergement et restauration	38 182,2	-183,7	136,8	-22,1	**38 113,2**
A17.J	Information et communication	81 663,0	774,2	269,8	-7,5	**82 699,4**
A17.K	Activités financières et d'assurance	16 997,6	150,0	0,0		**17 147,5**
A17.LZ	Activités immobilières	42 209,5	-364,9	1 133,7	-46,0	**42 932,4**
A17.MN	Activités scientifiques et techniques ; services administratifs et de soutien	176 014,3	-819,8	1 121,8	-58,6	**176 257,8**
A17.RU	Autres services	19 667,0	1 186,9	116,1	-14,5	**20 955,5**
A5.OQ	**Services principalement non marchands (*)**	66 489,5	298,7	6,8		**66 794,9**

Sur la valeur ajoutée totale, l'effet Rédi n'est que de l'ordre de 3 pour mille, ce qui est logique puisque pour la VA ou le CA, la source de référence est la liasse fiscale et l'effet Rédi n'intervient que lorsque celle-ci n'est pas retenue.
Les changements d'APE ont un effet global encore dix fois plus faible. Mais sur des secteurs particuliers (exemple ici la cokéfaction et le raffinage C2) on peut avoir des effets beaucoup plus importants, ceux-ci se compensant entre les secteurs (ici, c'est la restructuration d'un groupe qui explique ce transfert de valeur ajoutée que l'on retrouve dans un autre secteur ; le tableau, qui se décline ensuite à tous les niveaux de la nomenclature, permet de vérifier que cet effet a été correctement traité).
L'effet des sorties du champ Esane est assez logiquement faible, quel que soit le niveau de nomenclature considéré.

11. Resane « phase 2 » : la prise en compte des groupes et de l'entreprise au sens économique dans le système de production Ésane

La plupart des noms du CAC40 et plus généralement d'entreprises qui font quotidiennement la une des journaux sont en réalité des noms de « groupes[420] ». Mais les groupes, qui forment les deux tiers de la valeur ajoutée des entreprises[421] - c'est dire leur importance -, n'ont pas de véritable existence juridique et ne constituent pas des personnes morales. C'est la raison pour laquelle, jusqu'à des années récentes, la statistique publique les ignorait en pratique et ne s'appuyait dans la plupart des pays que sur les unités légales, même si de plus en plus de textes légaux ou réglementaires y faisaient référence.

À partir du constat qu'**il n'était plus pertinent de nos jours de se baser uniquement sur les unités légales** à la fois pour répondre aux différents **textes** français et/ou européens ayant trait aux statistiques d'entreprises[422], et pour produire des statistiques d'entreprises **plus signifiantes**, il a été décidé, dès l'expression des besoins du projet Resane, de prendre en compte les groupes en utilisant le concept **d'entreprises « profilées** », définies **comme acteurs économiques,** à partir des groupes. ***Il s'agit évidemment d'une innovation majeure*.***

Le § *11.1* revient ainsi sur **l'historique des circonstances et des raisons** qui ont conduit à cette décision.

Le § *11.2* vise à **définir** ce que sont ces nouvelles unités statistiques, **entreprises en général** et **entreprises profilées nationales issues des groupes,** bases des futures statistiques structurelles d'entreprises françaises et européennes. Elles sont d'ailleurs vouées à être prises en compte notamment dans le processus de production d'Ésane, mais pas uniquement ; leur prise en compte devrait se généraliser à l'ensemble des membres du SSE[423] ainsi que pour d'autres statistiques que les seules statistiques structurelles.

Le § *11.3* présente de façon très succincte **le système LIFI** qui permet de recenser aujourd'hui la plupart des groupes présents en France.

Le § *11.4* explicite ensuite la **stratégie** que l'Insee envisage pour profiler l'ensemble des groupes français, avec une répartition des groupes en **trois cibles différentes dont la prise en compte sera échelonnée dans le temps** (cibles dites 1,2,3 en fonction de la date de leur prise en compte).

Le § *11.5* explique la façon dont on profile les groupes **de la cible dite 1, c'est-à-dire les plus grands groupes** et le § *11.6* **les traitements** qu'on leur applique dans la chaîne Ésane. **Les cibles 2 et 3** font, de leur côté, l'objet d'un nouveau projet de l'Insee dont la mise en production devrait se situer en 2018 et dont le traitement n'est pratiquement pas décrit ici.

Enfin, pour utiliser, gérer et différencier toutes ces nouvelles unités – groupes, entreprises profilées, unités légales – il a fallu construire un nouveau **répertoire statistique,** dénommé **SIRUS** (pour Système d'Identification au Répertoire, des Unités Statistiques) qui avait donc été prévu dès l'origine du programme Resane. Il a été mis en production à fin 2012 et fait l'objet du § *11.7*.

11.1 Pourquoi profiler ?

La nécessité de profiler certains groupes est apparue dès la fin des années 90, à l'occasion d'une restructuration juridique des deux grands constructeurs automobiles français qui avaient décidé, à peu près à la même date, de donner la personnalité morale à chacun de leurs établissements de production. Ainsi de nouveaux flux monétaires, sans réelle consistance économique, sont apparus entre les unités légales de ces groupes, correspondant aux flux physiques des composants des

[420] Un groupe d'unités légales, au sens statistique, est un ensemble d'unités légales, le plus souvent liées entre elles par des liens financiers, de telle sorte qu'elles sont toutes contrôlées à plus de 50 %, directement ou indirectement, par l'une d'entre elles (appelée tête de groupe), cette dernière n'étant pas contrôlée par une autre. La tête de groupe peut être située à l'étranger.
[421] Voir par exemple l'Insee Première n°1399
[422] Le règlement unités statistiques 696/93, le nouveau règlement répertoires (177/2008) et le décret français 2008-1354 relatif aux catégories d'entreprises dont on rappelle un peu plus loin les définitions.
[423] SSE : Système Statistique d'Entreprises.

voitures. Dans cette restructuration, les unités légales de production louaient la main d'œuvre et les machines à certaines unités légales du même groupe et elles vendaient leur production à d'autres unités légales du groupe en charge de la commercialisation. Toutes ces nouvelles unités légales n'avaient pas de réelle autonomie de production, ni de décision, les décisions continuant bien évidemment d'être prises au niveau du groupe. Elles ne satisfaisaient donc pas aux critères de définition classiques d'une « entreprise » et n'étaient pas des « acteurs économiques » à part entière. La restructuration mentionnée ci-dessus conduisait à **multiplier « fictivement » par près de trois** le chiffre d'affaires, la sous-traitance, etc., de la filière automobile française alors qu'en réalité rien n'avait changé dans les capacités de production du fait de cette restructuration, de sorte **qu'il n'était pas possible de ne pas retraiter** ces nouvelles données pour se ramener à une situation économiquement significative.

Ce genre de décision, éclatement d'une unité légale en plusieurs morceaux – ou la décision inverse consistant à fusionner plusieurs unités légales au sein d'une nouvelle entité – peut se produire à tout instant au sein des grands groupes. L'assimilation de l'unité légale à l'entreprise n'est donc plus pertinente dans ce genre de cas où c'est au niveau du groupe que se trouve l'autonomie de décision et la gestion de l'ensemble des facteurs de production et non plus au niveau des unités légales. Le fait de continuer à raisonner en unités légales **perturbe non seulement les niveaux des agrégats mais également leur répartition sectorielle**.

D'autres cas particuliers du même genre, ont ainsi amené l'Insee à expérimenter progressivement, à compter des années 2000, le profilage relativement à 5 groupes souvent qualifiés « **d'historiques** », pour lesquels on ne prenait pas en compte les unités légales qui les constituaient, mais on rentrait dans le système Suse (Cf. *Avant-propos*) des comptes consolidés directement obtenus auprès de ces groupes, comme s'il s'agissait d'une liasse fiscale qu'on avait obtenue de la DGFiP.

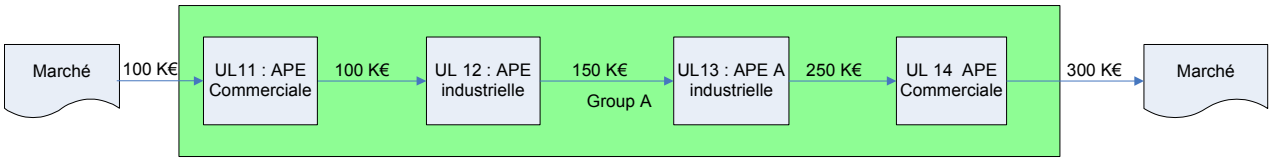
On peut illustrer la démarche suivie par l'exemple fictif qui suit. Soit l'unité légale indépendante suivante :

Cette unité légale fait l'année N un chiffre d'affaires de 300 K€ et crée une valeur ajoutée de 200 K€.

Imaginons que pour des raisons fiscales les dirigeants de cette unité légale décident de créer en N+1 une structure de groupe dans laquelle il y a autant d'unités légales qu'il y a d'activités différentes dans le processus de production de l'unité légale initiale. Il se crée donc :

- une unité légale (UL11 sur le graphique ci-dessous) en charge de l'approvisionnement en matières premières qui revend l'ensemble de ses achats au prix coûtant à une première société industrielle (UL12) ;
- cette dernière est un sous-traitant exclusif d'une autre unité légale industrielle (UL13) du groupe dont l'activité est le cœur de métier du groupe et qui correspond à l'APE de l'unité légale ci-dessus (APE=A) ;
- l'unité légale UL13 vend l'ensemble de sa production à une société commerciale du groupe (UL14) qui se charge de commercialiser les produits sur le marché.

Si on fait l'hypothèse simplificatrice qu'entre les deux années il n'y a aucune modification de l'activité économique de la société on peut aboutir au résultat suivant :



Le fait de raisonner en unité légale aurait alors pour conséquence de :

- faire passer le chiffre d'affaires de 300 lorsque l'unité légale était indépendante à 800 à cause des flux intra-groupe sans qu'il y ait de modification de l'activité économique. Ce dernier point est d'ailleurs observable au niveau de la valeur ajoutée qui ne change pas entre les deux années. La valeur ajoutée est en effet une variable additive qui est censée être indépendante de la structure des unités (mais ce n'est plus vrai si on s'intéresse à la ventilation de la valeur ajoutée par secteur) ;
- faire apparaître de nouvelles activités (une industrielle et deux commerciales) ;
- modifier les répartitions sectorielles du chiffre d'affaires mais aussi de la valeur ajoutée.

Au contraire, le fait de considérer « le groupe dans son ensemble » comme étant « une entreprise » permet de retomber sur la situation précédente qui, par hypothèse, n'a pas changé. C'est tout l'intérêt du profilage qui permet de se conformer à une vision économique de l'entreprise. C'est ce qui était fait au niveau des 5 groupes historiques et qui a présidé à la généralisation de la démarche dans la 2e moitié des années 2000.

En conclusion, cet exemple permet de mettre en évidence des modifications qui apparaîtront au moment du profilage des groupes :

- **disparition des activités secondaires**, internes au groupe ;
- **réduction du chiffre d'affaires total** (et des variables non additives en général) à cause de la suppression des flux intra-groupes ; ceci est légitime dans la mesure où ces derniers ne respectent pas les règles du marché ; on suivra donc désormais des **comptes « consolidés » ;**
- **conservation du niveau total de la valeur ajoutée** (seule sa ventilation sectorielle sera modifiée).

11.2 Qu'est ce qu'une entreprise et une entreprise profilée ?

Les réflexions ci-dessus ont été conduites en France en 2006 et 2007 dans le cadre d'un groupe de travail du Cnis[424] qui a amené l'Insee à prendre en 2008 **la décision de modifier à terme la définition des entreprises** dans le système statistique public français.

Mais elles sont également présentes **au niveau européen**, sous l'égide d'Eurostat, qui a entrepris une vaste refonte de l'ensemble des règlements qui régissent la statistique d'entreprise européenne, sous la forme d'un futur règlement cadre (dit « FRIBS[425] ») et lancé un groupe de travail en réseau (ESSnet) sur le sujet du profilage[426] et de la définition de « l'entreprise ».

A cette occasion, Eurostat a pris conscience de ce que la définition de « l'entreprise », telle qu'elle figure dans le règlement initial sur les unités statistiques de 1993, était appliquée de façon souvent biaisée par nombre de pays ; rappelons cette définition issue du règlement (N°696) de 1993 qui a été ensuite **reprise dans la LME**[427] française en 2008 pour définir les catégories d'entreprises :

L'entreprise correspond à la plus petite combinaison d'unités légales qui constitue une unité organisationnelle de production de biens et de services jouissant d'une certaine autonomie de décision, notamment pour l'affectation de ses ressources courantes. Une entreprise exerce une ou plusieurs activités dans un ou plusieurs lieux. Une entreprise peut correspondre à une seule unité légale.

La plupart des pays (donc y compris la France jusqu'aux années 2010, si on excepte le cas des cinq groupes historiques) ont :

- considéré que « la plus petite combinaison d'unités légales » était l'unité légale elle-même ;
- généralisé la dernière phrase en assimilant systématiquement entreprise et unité légale,

modifiant par là même l'esprit initial de la définition.

Dans le cadre de FRIBS, Eurostat s'emploie donc désormais à restituer l'esprit de la définition d'origine c'est-à-dire à demander aux pays de vérifier que les unités légales jouissent d'une « certaine autonomie de décision » lorsqu'elles sont assimilées à des entreprises. Lorsqu'elles appartiennent à des groupes, cela revient de facto à modifier les unités utilisées jusqu'à présent et à pratiquer le « profilage » des grands groupes, c'est-à-dire leur découpage (éventuel) en entreprises.

Le respect de la définition implique donc désormais qu'une entreprise devrait correspondre :

- soit à une seule unité légale non contrôlée par une autre unité légale ;
- soit à un groupe d'entreprises en tant que combinaison d'unités légales sous un contrôle commun ;
- soit à une partie autonome d'un groupe d'entreprises lorsque le groupe est constitué de plusieurs « unités organisationnelles » relativement autonomes.

[424] Voir le rapport du Cnis n° 107 : Statistiques structurelles fondées sur les groupes d'entreprises et leurs sous-groupes

[425] Framework Regulation Integrating Business Statistics. Ce règlement cadre prolonge une réflexion entamée dans le cadre du programme "MEETS" (Modernisation of European Enterprise and Trade Statistics), adopté en décembre 2008 par le Parlement européen et le Conseil.

[426] L'Essnet profilage s'est déroulé de fin 2009 à fin 2013.

[427] Loi de Modernisation de l'Économie du 4 août 2008.

En tant que statisticiens nationaux (en l'occurrence français), chargés d'élaborer des statistiques nationales, on est obligé de considérer comme entreprise « française » un ensemble d'unités légales « françaises ». Lorsqu'on a affaire à des groupes internationaux et lorsque « l'unité organisationnelle productrice de biens et services, jouissant d'une certaine autonomie » dépasse les frontières nationales (ce qui est bien entendu fréquemment le cas), on est alors obligé de se limiter à la trace française de cette unité organisationnelle. On entend par « trace française » la simple **présence sur le sol national de l'activité**, tout à fait indépendamment de la nationalité du ou des propriétaires (actionnaires par exemple).

Une entreprise « française » (à la base des statistiques « françaises ») sera donc au final :

- soit une unité légale (française), indépendante (hors groupe).
- soit la trace française d'un groupe d'unités légales entier ;
- soit une partie de la trace française d'un groupe d'unités légales ;

Les deux derniers types d'entreprise cités ci-dessus sont plus communément appelés « **entreprises profilées** » au sens où ils sont le résultat du « profilage » de ces groupes, **le profilage** d'un groupe étant par définition l'opération de **délimitation d'une ou plusieurs** « entreprises »[428] en son sein (plusieurs lorsque celui-ci est constitué de plusieurs « unités opérationnelles » au sens de la définition ci-dessus).

On notera que dans cette restriction au sol national, **on peut perdre une partie de « l'autonomie** » censée caractériser l'entreprise en tant qu'acteur économique. On touche là, de fait, à la contradiction fondamentale de l'économie mondialisée actuelle : il n'est **pas possible** de décrire fidèlement son fonctionnement, **en se basant exclusivement sur des « acteurs nationaux »**. Les ensembles (nationaux) que l'on considère à la fin n'ont pas toujours le caractère d'autonomie que l'on souhaiterait et qui caractérise normalement une « entreprise ». Mais seul un niveau supra national (européen, voire mondial) permettrait de ne prendre en compte que de véritables acteurs économiques indépendants.

11.3 Qu'est ce qu'un groupe ? Le système Lifi

La constitution des entreprises se conformant réellement au règlement européen se base donc sur les « **groupes** » d'unités légales. Comme on l'a vu en note de bas de page en introduction du chapitre, au sens statistique, **un groupe** est **un ensemble d'unités légales**, le plus souvent liées entre elles par des liens financiers, de telle sorte qu'elles sont toutes contrôlées à plus de 50 %, directement ou indirectement, par l'une d'entre elles (appelée tête de groupe), cette dernière n'étant pas contrôlée par une autre. La tête de groupe peut être située à l'étranger. Pour les appréhender, il faut donc connaître tous les groupes présents dans un pays donné. Déjà dans le cadre de MEETS[429], Eurostat avait ainsi lancé un projet européen visant à la construction d'un répertoire européen des groupes (l'EGR pour European Group Register) contenant au moins tous les grands groupes présents dans au moins un pays de l'Union. Ce répertoire est la synthèse de la connaissance des groupes issus de **bases de données privées** et de **celle des instituts de statistique** pour les pays qui ont une connaissance en propre de leurs groupes.

C'est le cas de la France qui a développé un système d'information sur les groupes depuis le début des années 80. Ce système est connu sous le nom de « **LIFI » pour « Llaisons FInancières »** du nom de **l'enquête** qui a permis à l'Insee, **de 1981 à 2011**, de connaître les contours de groupes (c'est-à-dire la liste des unités légales contrôlées par la tête de groupe), à partir de la détention d'un portefeuille d'actions dont on leur demandait la composition. Il s'agissait d'une enquête « **à seuil** » dans laquelle n'étaient interrogées que les sociétés d'une certaine taille ou disposant d'un portefeuille important.

Ces contours sont désormais connus à partir **des déclarations exhaustives de détention de capital** provenant à la fois de la Banque de France, à compter des données **2012**, et de la DGFiP (ainsi que de sources de données privées), à compter des données **2015**.

La description précise de ce système d'information serait assez complexe et n'entre pas dans le champ de cet ouvrage. Ce qui importe ici est de savoir qu'il existe et peut être considéré désormais comme **exhaustif**. Depuis 2011, la prise en compte de sources nouvelles a fait passer le nombre de

[428] La dénomination exacte devrait donc plutôt être « entreprises issues d'un profilage » mais on emploie ce raccourci sémantique.

[429] MEETS pour Modernisation of European Enterprise and Trade Statistics était le programme qui a précédé FRIBS, adopté en décembre 2008.

groupes ayant au moins une filiale en France présents dans LIFI de 50 000 à **85 000 en 2014** et plus de **100 000 en 2015**, sachant qu'il se crée plusieurs milliers de groupes nouveaux chaque année.

Ce système a désormais (à compter de 2013) la dimension d'un **répertoire,** au sens où **chaque groupe est muni d'un identifiant** et où l'on sait définir la continuité d'un groupe d'une année sur l'autre même si son contour bouge un peu (Cf. le § *11.7.3* qui présente la définition de cette continuité).

11.4 La stratégie de profilage décidée par l'Insee

Elle s'appuie sur les « catégories d'entreprises » définies par le décret (1354-2008) d'application de la Loi de Modernisation de l'Économie (LME) que l'on rappelle ci-dessous :

- les **micro-entreprises** emploient moins de 10 personnes et ni leur chiffre d'affaires annuel ni leur total de bilan n'excède 2 millions d'euros ;
- Les **PME** (petites et moyennes entreprises) emploient moins de 250 personnes, et [leur chiffre d'affaires annuel est inférieur à 50 millions d'euros ou leur total de bilan à 43 millions d'euros] ;
- Les **ETI** (entreprises de taille intermédiaire) emploient moins de 5000 personnes, et [leur chiffre d'affaires annuel est inférieur à 1,5 milliard d'euros ou leur total de bilan à 2 milliards d'euros] ;
- Les **grandes entreprises** sont les entreprises non classées précédemment.

Les PME, ETI et grandes entreprises constituent une partition des entreprises. Les micro-entreprises sont incluses dans les PME.

Afin de définir le périmètre des entreprises françaises au sein des groupes, trois **cibles**[430] **de groupes** d'au moins deux filiales[431] ont été identifiées pour constituer ces entreprises ; chacune d'entre elle fait ou fera l'objet de **traitements différents quant au profilage des groupes** correspondants. Schématiquement :

- Cible 1 : les groupes de cette cible sont **susceptibles d'être découpés** en plusieurs entreprises profilées pour lesquelles ils fournissent, **depuis l'année de constat 2012** et pour chaque entreprise, l'équivalent de liasses fiscales et de réponse à l'Esa. Ces informations sont alors rentrées dans le système Ésane, à l'instar de celles des unités légales. Le mode de découpage et la collecte des informations font l'objet du § *11.5* ; leur traitement dans Ésane du § *11.6*.
- Cibles 2 et 3 : **tous les autres groupes**[432] seront **assimilés à des entreprises uniques** (un groupe = une entreprise). Les comptes consolidés (Cf. fin du § *11.1*) correspondants et la réponse consolidée à l'Esa seront **estimés**[433] **à compter de l'année de constat 2017** par un algorithme automatique, à partir des liasses fiscales individuelles des filiales et des réponses individuelles des unités légales du contour**.** Celui-ci n'étant pas encore définitivement adopté**,** il n'est pas décrit ici.
- Cible 3 : l'estimation d'un compte consolidé à partir des liasses fiscales des unités légales du contour est fragile sans information complémentaire sur les flux internes au groupe et l'imprécision est proportionnelle à l'importance du groupe et à son nombre de filiales. Pour les groupes les plus importants ou les plus complexes, **une enquête complémentaire périodique,** programmée **à compter de 2019**, permettra d'affiner les flux internes à consolider dans l'algorithme précédent. Cette enquête permettrait aussi bien sûr de collecter directement des comptes consolidés sur un périmètre français, si ces groupes en possèdent.

Plus précisément, la composition des cibles est donnée ci-après ; la limite entre les cibles 2 et 3 sera définie par le champ de la future enquête auprès de la cible 3 et **reste susceptible d'évolutions**. Les

[430] Comme on l'a vu, elles sont numérotées **dans l'ordre de leur prise en compte programmée dans Ésane** et non dans l'ordre décroissant de l'importance des groupes pris en compte : les groupes de la cible 3 sont plus gros que ceux de la cible 2.

[431] De nombreux groupes étrangers – voire quelques français d'ailleurs – n'ont qu'une seule filiale en France. Ils sont présents dans Lifi qui recense tous les groupes présents sur le territoire national. Mais pour Ésane, ils se comportent comme une UL indépendante pour laquelle on dispose de la liasse fiscale et de sa réponse à l'Esa si ses caractéristiques la font rentrer dans le champ. Seuls les groupes d'au moins deux filiales créent donc problème.

[432] Plus précisément, comme on l'a vu, leur trace française.

[433] Sauf dans quelques cas où ces groupes élaborent des comptes consolidés et les communiquent à la Banque de France qui peut alors nous les transmettre.

comptages donnent des ordres de grandeur, mais il faut garder à l'esprit que le nombre de groupes a beaucoup augmenté dans Lifi entre 2011 et 2014 du fait de la prise en compte de nouvelles sources d'information (Cf. § précédent) ; par ailleurs il se crée entre cinq et dix mille groupes par an ; bien entendu un certain nombre de groupes disparaissent aussi chaque année.

- **La cible 1** est formée des **plus grands groupes** présents en France ou des plus complexes[434]. Elle contient une cinquantaine de groupes[435]. Il s'agit, à peu de choses près, de groupes de plus de 10 000 salariés en France. Ils sont profilés « **manuellement** » (c'est-à-dire à la suite de rencontres en face à face avec des représentants des groupes) par les membres d'une division spécialisée de l'Insee[436], qu'on appelle des « profileurs ». Seuls les groupes de la cible 1 **peuvent être multi-entreprises**. Les contacts privilégiés entre ces groupes et les profileurs Insee permettent en effet de discuter de l'organisation interne du groupe et de l'existence « d'une structure formelle » en charge de la gouvernance des parties des groupes qui formeront les entreprises profilées en leur sein. Ces rencontres permettent également d'obtenir :
    - les éléments nécessaires à l'estimation de l'équivalent des liasses fiscales (« consolidées ») des entreprises profilées ;
    - leurs réponses à l'Esa ou à l'EAP ;
    - ainsi que les réponses à un ensemble d'enquêtes à définir avec eux.
- **La cible 2** serait formée des groupes **de petite taille ou simples,** c'est-à-dire des **groupes PME** ainsi que des **groupes ETI** qui ne sont constitués **que de 2 unités légales** (en France). Avec cette définition, elle contient (chiffres de 2014) **63 555** groupes dont **40 475** ne sont composés que de **deux unités légales** et parmi elles **584** ETI. Ces groupes sont **assimilés, par hypothèse, à une seule entreprise**. L'idée sous-jacente est que ces groupes sont en général assez simples, avec un petit nombre de filiales (beaucoup n'en ont que deux) de sorte que la consolidation de leur chiffre d'affaires de façon algorithmique, à partir de l'activité des filiales (connue par leur code APE), devrait approcher la réalité d'assez près, le nombre et la direction possible des flux internes étant assez limités. Dans beaucoup de cas, il n'y a qu'une seule activité productrice et la ou les autres filiales sont des unités auxiliaires relativement simples à consolider avec l'unité qui porte l'activité. Étant donné leur nombre et le coût des discussions en face à face, il est de toutes façons exclu de les contacter individuellement. Une vérification de terrain a d'ailleurs permis de constater qu'ils se vivaient complètement comme des « entreprises » et n'avaient d'ailleurs pas toujours conscience de constituer des « groupes » de sorte qu'ils possèdent rarement des comptes consolidés. Ils feront l'objet d'une **consolidation « automatique »** algorithmique, **sans collecte d'information complémentaire**.
- **La cible 3** correspondrait aux groupes restants, c'est-à-dire aux **groupes ETI d'au moins trois unités légales** et aux **grands groupes qui n'ont pas pu être traités en face à face**. Elle contient **3855** groupes en 2014. Leur nombre exclut là encore des rencontres en face à face ; en revanche, les hypothèses sous tendant l'estimation des flux internes à partir de l'activité (APE) des filiales sont beaucoup plus fragiles, le nombre de combinaisons étant beaucoup plus important. L'application de l'algorithme défini pour la cible 2 donne toujours une estimation d'un compte consolidé pour ces groupes ; elle est simplement de moins bonne qualité, mais a priori meilleure que si on ne faisait rien[437]. On envisage donc **une enquête complémentaire** qui permettrait une meilleure estimation des flux internes, et donc des comptes consolidés correspondants, ainsi que le traitement des enquêtes ESA et EAP. Lorsqu'ils existent sur le périmètre national, on collectera aussi bien entendu directement leurs comptes consolidés. Enfin, la disponibilité d'une base de comptes consolidés à la Banque de France ouvre des possibilités de simplifier la collecte de données pour certains groupes.

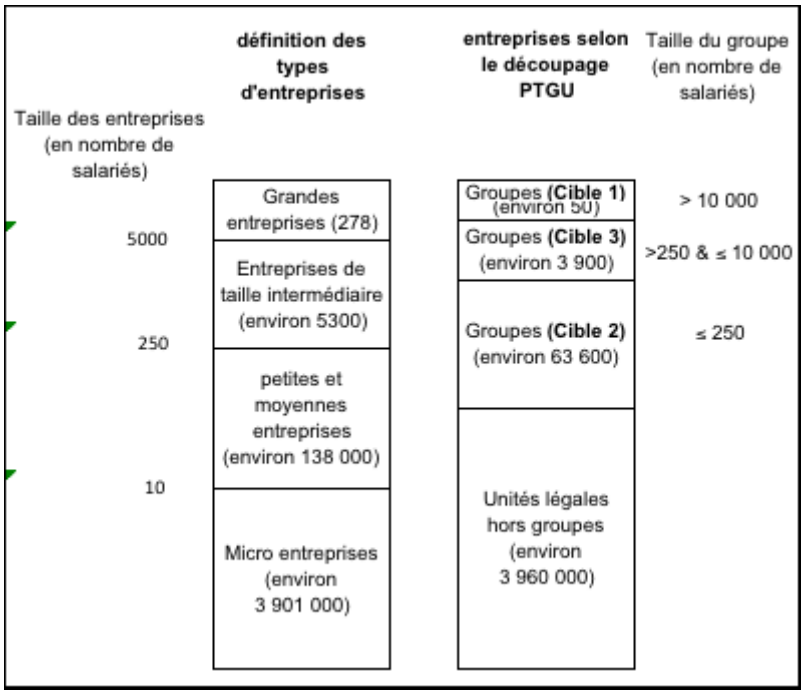
[434] On entend par « complexe » par opposition à « simple », des groupes ayant de nombreuses filiales et de multiples activités.
[435] À fin 2016 (pour les résultats de 2015), les discussions étaient achevées avec 51 groupes correspondant à 116 entreprises profilées et quelque 4149 filiales en France. Voir tableau 3 du § *11.6.6*.
[436] Il s'agit de la division PTGU (Profilage et Traitement des Grandes Unités).
[437] Une comparaison entre l'algorithme automatique de consolidation (version provisoire) et les comptes consolidés de la Banque de France, sur le champ partiel sur lequel elle était possible, a montré plutôt une sous-consolidation avec l'algorithme car par défaut, lorsqu'on a des doutes sur l'existence d'un flux a consolidation, on ne fait rien. Si on ne « sur-consolide » pas, on ferait donc nécessairement mieux qu'aujourd'hui où on ne consolide pas du tout.

Pour retrouver la totalité des groupes présents en France, il faut ajouter ceux qui n'ont qu'une filiale[438] en France (soit 13 267 groupes) et qu'on suit alors directement en tant qu'unités légales. On arrive alors au total de **80 727 groupes actifs en 2014,** et dans le champ « Insee » (c'est-à-dire non financiers, non agricoles, et dont le chiffre d'affaires est strictement positif). Si on rajoute les 3424 groupes « financiers », dont la collecte d'informations est du ressort de la Banque de France mais qui sont dans Lifi, on arrive à un total de **84 151 groupes en 2014**.

Le graphique suivant (tableau 1, toujours fondé sur les chiffres 2014) permet de faire le lien entre les types d'entreprises définis par le décret 2008-1354 du 18 décembre 2008 sur les catégories d'entreprises et la décomposition qui a été faite en termes de profilage.

Dans ce graphique ne figurent que des critères de taille d'entreprises en termes de salariés par pur souci de simplification. En réalité il y a aussi des critères de chiffre d'affaires et de total de bilan pour caractériser les PME, ETI ou grandes entreprises.

Tableau 1 : Correspondance entre les cibles et les catégories des groupes correspondants, exprimée en fonction de la taille :



Source : Lifi 2014 ;
On assimile groupes et entreprises, sauf pour les 50 groupes de la cible 1 qui donnent lieu à 118 "entreprises profilées".

Le tableau 2 suivant dénombre les **entreprises** (en tant que « unités statistiques ») selon leur **catégorie**, leur **taille en termes d'effectifs salariés** (en cinq tranches), et leur **origine** (groupes ou unités légales indépendantes), **en 2014**. Il permet de comprendre précisément la répartition entre les entreprises au sens « groupes » et celles au sens « unités légales indépendantes » (attention en revanche, ce tableau ne donne pas le nombre de filiales composant les groupes) ; dans le cas où l'entreprise provient d'un groupe (elle est égale au groupe ou provient du découpage d'un groupe de la cible 1) on a rajouté la **nationalité** du groupe, car on distingue souvent les groupes selon qu'ils sont étrangers, franco-français, ou français multinationaux. La nationalité d'un groupe est définie dans Lifi[439].

[438] Voire 0 ! Il y a en effet quelques groupes qui ne sont présents qu'au travers de succursales c'est-à-dire d'établissements dépendants d'unités légales étrangères. Ils peuvent donc avoir une « activité » en France mais n'ont pas véritablement « de filiales » (au sens unité légale) en France. Ils n'ont qu'une unité locale.

[439] C'est le plus souvent celle de l'unité légale « tête de groupe ». Cependant lorsque la tête de groupe a été implantée dans un pays exclusivement pour des raisons fiscales, on prend la nationalité de l'unité légale qui abrite les instances dirigeantes du groupe.

Tableau 2 : Ventilation du nombre total d'entreprises selon leur catégorie, leur tranche d'effectifs, et leur "origine" unité légale indépendante ou groupe en distinguant la nationalité du groupe.

Catégorie d'entreprise	Origine des entreprises	Tranches d'effectifs					TOTAL
		Sans salarié	De 1 à 9 salariés	De 10 à 249 salariés	De 250 à 4999 salariés	Plus de 5000 salariés	
A - MIC	Groupes étrangers	900	3 311	0	0	0	**4 211**
	Groupes français multinationaux	516	822	0	0	0	**1 338**
	Groupes franco-français	2 613	17 740	0	0	0	**20353**
	Unités indépendantes	2 869 208	1 006 245	0	0	0	**3 875 453**
Total MIC		**2 873 237**	**1 028 118**	**0**	**0**	**0**	**3 901 355**
B - PME	Groupes étrangers	243	752	5 424	0	0	**6 419**
	Groupes français multinationaux	173	234	3 462	0	0	**3 869**
	Groupes franco-français	316	2 995	40 582	0	0	**43 893**
	Unités indépendantes	2 290	6 753	74 823	0	0	**83 866**
Total PME		**3 022**	**10 734**	**124 291**	**0**	**0**	**138 047**
C - ETI	Groupes étrangers	3	7	350	930	0	**1 290**
	Groupes français multinationaux	18	5	323	1 250	0	**1 596**
	Groupes franco-français	10	6	528	1 405	0	**1 949**
	Unités indépendantes	37	10	157	282	0	**486**
Total ETI		**68**	**28**	**1 358**	**3 867**	**0**	**5 321**
D - GE	Groupes étrangers	0	0	0	29	46	**75**
	Groupes français multinationaux	0	0	0	33	155	**188**
	Groupes franco-français	s	s	0	s	10	**s**
	Unités indépendantes	s	s	0	s	0	**s**
Total GE		**s**	**s**	**0**	**65**	**211**	**278**
Toutes catégories	Groupes étrangers	**1146**	**4 070**	**5 774**	**959**	**46**	**11 995**
	Groupes français multinationaux	**707**	**1 061**	**3 785**	**1 283**	**155**	**6 991**
	Groupes franco-français	**2 939**	**20 741**	**41 110**	**1 407**	**10**	**66 207**
	Unités indépendantes	**2 871 536**	**1 013 009**	**74 980**	**283**	**0**	**3 959 808**
TOTAL		**2 876 328**	**1 038 881**	**125 649**	**3 932**	**211**	**4 045 001**

Source : Lifi 2014

Champ : hors entreprises dont le chiffre d'affaire est négatif ou nul, et hors entreprises des secteurs agricoles et administratifs (*mais y compris le secteur financier*) ; les entreprises étant ici assimilées aux groupes.

S : secret statistique

11.5 Le traitement des entreprises profilées de la cible 1 par les profileurs

Comme on l'a vu au § 11.4, les groupes de la cible 1 sont traités dans Ésane par les « profileurs », experts de l'Insee spécialisés pour ce faire, depuis environ 2010. Les premiers résultats publiés portent sur l'année 2012, car les négociations avec ces groupes n'ont évidemment pas été instantanées.
Plus précisément, il s'agit pour eux de :

- **définir le périmètre** (en termes d'unités légales et donc limité au territoire français) des entreprises profilées au sein des groupes (cf. § 11.5.1) ;
- collecter les paramètres nécessaires pour **reconstituer des comptes « consolidés »** correspondant au périmètre des entreprises profilées précédemment définies (cf. § 11.5.2 et 11.5.3) ;
- **collecter les réponses aux enquêtes Esa et EAP**, toujours correspondant au périmètre des entreprises profilées et donc cohérentes avec les comptes ci-dessus (cf. § 11.5.4).

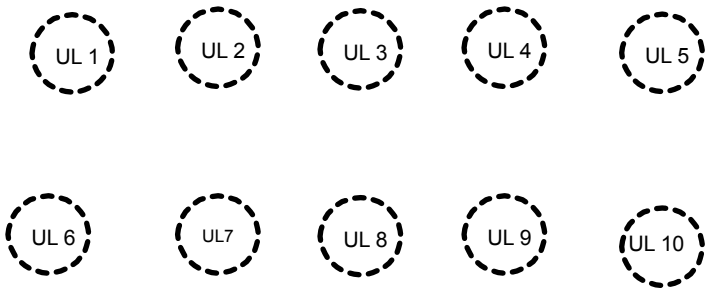
Ces informations sont stockées et traitées dans une base de données spécifique avant d'être envoyées à Ésane (cf. § 11.6).

11.5.1 La création des périmètres des entreprises profilées

La création des entreprises profilées se fait en deux étapes :

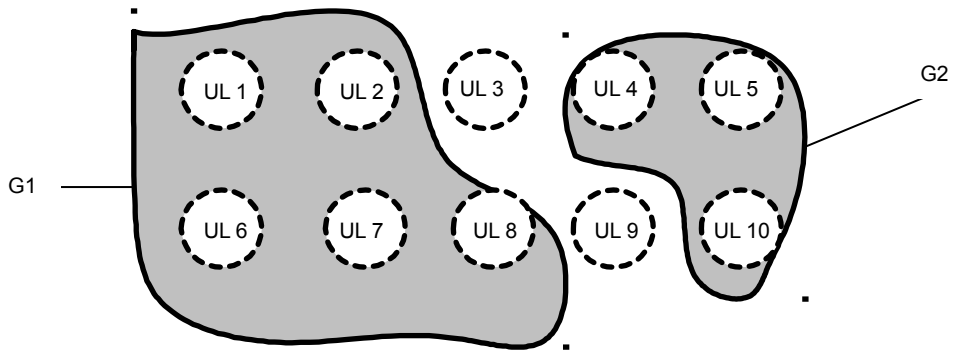
- en relation directe avec le groupe, il faut **définir le nombre d'entreprises** (au sens économique du § 11.2) qui composent le groupe. La définition des entreprises dépend bien entendu aussi de la possibilité qu'aura le groupe de fournir des informations fiscales et statistiques sur ces nouvelles unités car il ne servirait à rien de définir des « entités » théoriques sur lesquelles le groupe serait incapable de fournir une information comptable ou sectorielle détaillée. C'est souvent tout l'objet de la négociation avec le groupe ; parallèlement, il n'est guère pensable qu'une entité ait une existence économique réelle sans que des données correspondantes ne décrivent son activité globale ; mais la difficulté principale vient de ce que nous nous intéressons in fine à la trace française de cette activité, qui n'est pas souvent spontanément disponible au sein des groupes. La définition des entreprises dépend donc **de son système d'information** et des capacités du groupe à **l'adapter à nos demandes**.
- dans un second temps, il convient **d'associer chacune des unités légales**[440] qui composent le groupe sur son périmètre français **à l'une des entreprises profilées créées**.

Prenons l'exemple d'une population de 10 unités légales (UL) de SIRENE :



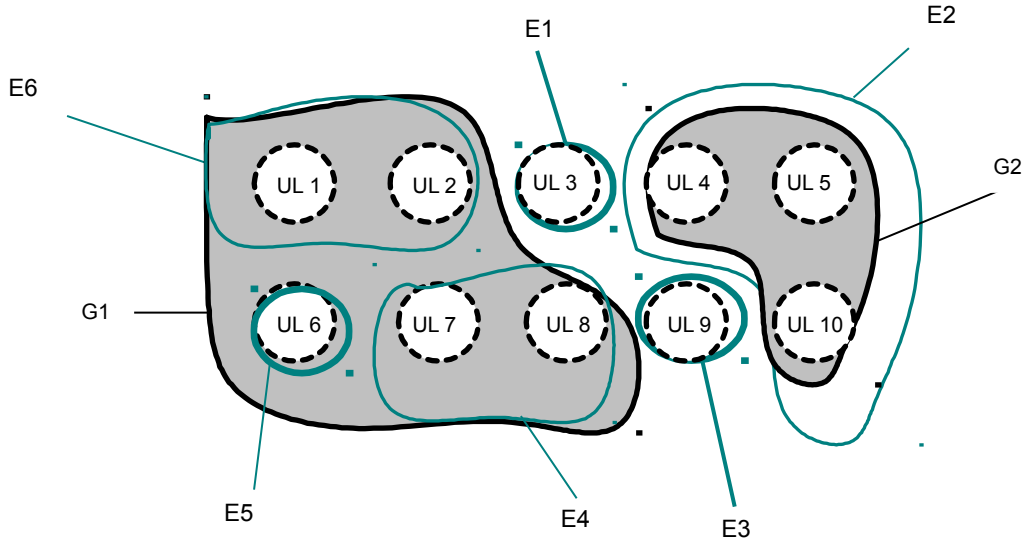
[440] Ou « partie » d'unités légales. En effet une unité légale qui joue un rôle transverse pour le groupe (par exemple elle abrite tous les salariés du groupe pour la paye) peut, exceptionnellement, être « partagée » entre plusieurs entreprises profilées.

Supposons que plusieurs d'entre elles soient détenues par des groupes de sociétés G1 (cible 1) et G2 (cible 2) :



Par principe, comme on l'a vu (§ *11.4*), tout groupe de la cible 2 forme une entreprise. En revanche pour le groupe G1 de la cible 1, le profileur en charge du dossier s'entend avec le groupe pour créer (dans l'exemple) 3 entreprises. Ainsi, UL1 et UL2 forment une entreprise, de même que les UL7 et UL8 ; l'UL 6 constitue pour sa part une entreprise à part entière.

Après avoir profilé les groupes, la population initiale des *10 unités légales* devient alors une population de *6 entreprises* comme le décrit la figure ci-dessous (les 2 unités légales indépendantes, les 3 entreprises profilées issues du profilage manuel du grand groupe et le petit groupe G2 qui a été assimilé à une entreprise):



11.5.2 Les différents types de consolidation des données fiscales envisagés

Une fois les périmètres des entreprises créés, le deuxième travail du profileur est d'obtenir du groupe les éléments permettant d'établir **l'équivalent d'une liasse fiscale pour l'entreprise profilée**. Comme on ne veut plus prendre en compte les flux entre les unités légales du périmètre retenu, il s'agira de « **comptes consolidés** » sur ce périmètre. La difficulté principale réside là dans le fait que la liasse fiscale est conçue pour être compatible et cohérente avec le plan comptable général (PCG) français alors que la plupart de comptes consolidés des groupes sont élaborés en respectant les normes comptables dites « IFRS[441] » qui, elles, ne le sont pas.

Il y a donc **deux principales méthodes de consolidation** pour reconstituer des comptes consolidés des entreprises profilées, que l'on peut ensuite intégrer à Ésane : soit on part « du bas » (on parle de méthode « bottom-up ») c'est-à-dire des liasses fiscales des unités légales du périmètre que l'on consolide, soit on part « du haut » (on parle de méthode « top-down ») c'est-à-dire des comptes déjà consolidés au niveau du groupe et il faut d'une part les déconsolider au niveau de chaque entreprise profilée (seul le groupe peut le faire ; c'est l'objet de la discussion avec lui), et d'autre part passer des concepts IFRS aux concepts et aux variables de la liasse fiscale actuelle (typiquement le travail du

[441] International Financial Reporting Standards. Ces normes comptables, élaborées par l'International accounting standards board (IASB), ont été rendues obligatoires par la Commission européenne, via un règlement de 2002, pour tous les groupes faisant appel à l'épargne publique. De fait, la plupart des grands groupes les ont adoptées.

profileur). [Il existe également une norme de consolidation « française, mais on la rencontre très rarement dans les grands groupes[442]]. Plus précisément,

- Dans la méthode **bottom-up**, il s'agit de cumuler les liasses des unités légales du périmètre de l'entreprise profilée et d'en retirer les flux internes à l'entreprise, généralement appelés « intercos ». Ces flux doivent bien évidemment nous être communiqués par le groupe qui doit donc les calculer pour nous. Ils permettent ensuite au profileur de reconstituer une liasse fiscale « consolidée ».
- Dans la méthode **top-down,** le groupe fournit pour chaque entreprise des informations comptables en normes IFRS (une centaine de variables), tirés directement des comptes consolidés du groupe. Dans ce cas, l'inconvénient est que les caractéristiques de la liasse fiscale (plus de 200) correspondant à l'entreprise profilée, que l'on doit reconstituer pour la rentrer dans le système Ésane, doivent être « estimées » à partir de ces données et des cumuls des caractéristiques des UL des périmètres des entreprises, de façon à garantir le maintien de la valeur ajoutée telle qu'elle figure dans les liasses fiscales des unités légales du contour (la valeur ajoutée étant une variable dite « additive », elle n'est pas impactée par la consolidation).

Le choix entre les deux méthodes est fait en collaboration entre le groupe et les profileurs, en fonction du type d'information qui est le plus facile à mobiliser pour le groupe (souvent fonction du nombre d'unités légales concernées) : soit la mesure des flux intra-entreprises profilées, soit les variables IFRS relatives à ces entreprises. Dans quasiment tous les cas, **un paramétrage ad hoc du logiciel de consolidation** du groupe est nécessaire, car il est exceptionnel que les segments opérationnels correspondent aux frontières nationales[443].

Lorsque le choix inverse n'est pas explicitement requis par le groupe, **la méthode bottom-up est privilégiée** car elle est moins coûteuse à gérer puisque les données sont directement au format du plan comptable général, comme les liasses des unités légales ; de plus elle conserve formellement la valeur ajoutée que l'on peut calculer sur les unités légales alors que le calcul de la valeur ajoutée à partir des normes IFRS est beaucoup plus délicat.

Les paramètres de consolidation (flux intra-entreprise ou données fiscales au format IFRS) sont stockés dans une base de données ad hoc, la **BCE**[444]. Des programmes de consolidation, adaptés si nécessaire à chaque cas particulier par les profileurs, permettent ensuite de **créer les variables de la liasse** pour ces entreprises profilées, à partir de ces paramètres et des liasses fiscales des unités légales des périmètres qui sont déjà présentes dans le système Ésane, puisque livrées par la DGFiP. Elles sont chargées dans le système Ésane à la fin du processus (Cf. § *11.6.3*).

Les spécifications de ces programmes de consolidation, très techniques, n'ont pas véritablement leur place ici car elles sont relativement exogènes au système Ésane lui-même et alourdiraient considérablement l'ouvrage[445]. Il suffit de retenir qu'**on sait reconstituer**, sur le périmètre des entreprises profilées, l'équivalent des **variables comptables** que l'on récupère habituellement au travers des liasses fiscales.

[442] La règle précise est de consolider « aux normes françaises », si le groupe ne fait pas appel à l'épargne publique et s'il a franchi certains seuils de taille: "Le règlement CRC 99-02 du Comité de la réglementation comptable, relatif aux comptes consolidés, homologué par arrêté du 22 juin 1999, expose les Règles et méthodes relatives aux comptes consolidés (RMCC) qui doivent être appliquées obligatoirement par les sociétés commerciales et les entreprises publiques soumises à l'obligation d'établir des comptes consolidés."
Si le groupe fait appel à l'épargne publique (cotation de titres ou obligations sur des marchés réglementés), alors la consolidation répond à des règles européennes et est en IFRS ; les groupes peuvent néanmoins opter pour la consolidation IFRS et dans ce cas sont dispensés de la consolidation aux normes françaises. Au final, seuls quelques très rares groupes de la cible 1 ont adopté cette méthode « française » de consolidation (cohérente avec le PCG) qui ne nécessite que quelques adaptations supplémentaires de la part des profileurs.

[443] Les « segments opérationnels » ont été introduits par la norme dite IFRS 8 dont l'emploi est obligatoire pour les groupes cotés. Cette norme impose au groupe de publier des informations financières et descriptives sur les segments qu'il a définis. S'ils respectaient les frontières nationales, on aurait donc directement de l'information sur la trace nationale du groupe directement dans la communication obligatoire du groupe, ce qui est très rarement le cas.

[444] Base de Création des Entreprises. Cf. le § *11.7.2* qui explicite les relations entre les bases de données.

[445] Les notes correspondantes, internes à l'Insee, dépassent la centaine de pages en raison du très grand nombre de variables concernées.

11.5.3 Le cas des co-entreprises

On appelle co-entreprises (ou le plus souvent JV pour joint venture en anglais) des unités légales qui sont détenues à parts égales par plusieurs groupes (le plus souvent deux à 50/50 mais des détentions 1/3 1/3 1/3 sont aussi possibles). Formellement, ces co-entreprises ne font pas partie du contour des groupes puis qu'aucun ne les « contrôle », par définition. Toutefois lorsqu'on profile des groupes engagés dans des co-entreprises, les profileurs collectent et contrôlent aussi les comptes de ces co-entreprises afin de s'assurer qu'il n'y a ni omission ni double-compte avec les unités légales indépendantes. Des règles spécifiques de consolidation leur étaient applicables jusqu'en 2014.

Jusqu'au 1er janvier 2014 en effet, les comptes des coentreprises pouvaient être consolidés selon la méthode de « l'intégration proportionnelle » (qui permet au groupe partenaire de la coentreprise d'afficher une part proportionnelle à sa participation des éléments constituant le patrimoine et le résultat de l'entreprise après élimination des opérations réciproques). Le 1er janvier 2014, est entrée en vigueur la norme IFRS 11 qui met fin à la possibilité pour les groupes établissant leurs comptes en normes internationales de choisir l'intégration proportionnelle et impose la méthode de la mise en équivalence pour la comptabilisation des participations dans les coentreprises. Ceci a conduit de nombreux groupes à **prendre le contrôle de leurs coentreprises ou au contraire à s'en dégager**. Le nombre de coentreprises gérées par la division a donc fortement diminué (Cf. Tableau 3 ci-après).

11.5.4 La collecte des réponses aux enquêtes Esa ou EAP

Afin de pouvoir traiter dans Ésane les entreprises profilées comme les entreprises constituées par des unités légales indépendantes, il faut non seulement disposer des variables comptables mais également des variables complémentaires collectées au travers des enquêtes Esa ou EAP ; ce d'autant plus que la taille des entreprises profilées les place systématiquement dans la strate exhaustive de ces enquêtes. La troisième mission du profileur est donc de **collecter également ces informations** auprès des groupes.

Bien que le profilage conduise très fréquemment à un **allègement de la charge statistique très fort**, puisqu'il n'est pas rare de remplacer plusieurs dizaines (voire centaines) de questionnaires adressés individuellement aux unités légales du groupe par quelques unités (un par entreprise profilée), c'est souvent la partie la plus difficile de la négociation avec les groupes. En effet, la variable centrale de ces enquêtes est **la ventilation fine du chiffre d'affaires** selon l'ensemble des activités de l'entreprise, et, selon les secteurs, on a besoin de descendre fréquemment au niveau fin des produits. L'EAP dans l'industrie sert par exemple, comme on l'a vu, à répondre au règlement européen Prodcom qui comprend environ 3800[446] produits. Or nos correspondants dans les groupes sont le plus souvent des experts en comptabilité et en consolidation (le plus souvent mondiale) ; ils ne sont pas nécessairement au courant de la décomposition fine du chiffre d'affaires français en produits.

Comme pour la consolidation, et c'est tout l'intérêt d'une relation personnalisée avec les groupes, il y a donc **diverses stratégies d'interrogation** qui se mettent en place :

- Dans certains cas, les groupes sont en mesure de répondre directement à l'enquête au niveau de détail demandé et il n'y a donc aucun problème.
- Dans d'autres, ils préfèrent – malgré la lourdeur de la charge de réponse correspondante – continuer de répondre au niveau des unités légales. Il appartient alors au profileur de retirer des réponses spontanées des unités légales, les produits qui sont autoconsommés au sein des entreprises profilées afin de retomber sur leur chiffre d'affaires « consolidé ». Il le fait bien évidemment avec l'aide du groupe (qui donc ne fait pas lui-même ce calcul mais accepte en général de « valider » le résultat).
- Dans d'autres enfin, on met en place des stratégies d'interrogation « mixtes » : par exemple le groupe remplit un questionnaire Esa pour ses activités commerciales ou de service mais on garde une collecte au niveau des unités légales industrielles. Ou bien le groupe nous donne des totaux intermédiaires au niveau sous-classe de la nomenclature qui permet au profileur de recaler les ventilations fines par produit. Ou bien encore, il accepte de faire le travail détaillé sur une année « de base » sur laquelle on se calera pendant plusieurs années, en validant auprès du groupe qu'il n'y a pas eu de changement important dans la structure d'autoconsommation en produits.

[446] En 2014. Cette liste est revue chaque année, généralement dans le sens de la diminution.

Quoi qu'il en soit, ce n'est que lorsque les profileurs ont réussi avec le groupe cette triple mission (définition des périmètres, collecte des comptes consolidés et collecte des enquêtes) que l'on considère que le groupe est véritablement « profilé ». En d'autres termes, tel qu'il est mis en œuvre à l'Insee, le profilage comprend non seulement le **découpage** des groupes en entreprises profilées mais aussi la **collecte** des informations nécessaires pour être traitées et intégrées au système d'information Ésane[447].

De plus, on impose **une année de collecte en double** (c'est-à-dire à la fois en unités légales et selon les nouveaux périmètres) de façon à vérifier pragmatiquement la capacité du groupe à répondre sur ces nouveaux périmètres. Cette collecte en double permettra également aux comptables nationaux d'estimer la « marche d'escalier » correspondant au passage de l'ancien au nouveau système.

11.5.5 La collecte des réponses aux autres enquêtes de la statistique publique

Se pose enfin la question de la réponse aux « autres enquêtes » du système statistique public (enquêtes infra-annuelles ; enquêtes thématiques). Comme on l'a vu au § *11.2*, ces nouvelles unités statistiques ont vocation à remplacer progressivement les unités légales **lorsque la problématique étudiée le justifie**. Le découpage en « entreprises » est a priori bien adapté pour l'étude des questions économiques autour de la fonction de production. « L'unité légale » reste adaptée aux problèmes d'emploi et de droit du travail qui s'applique à des entités bien définies juridiquement. Le « groupe entier » est généralement l'unité statistique adéquate pour les problèmes de financement ou les dépenses de recherche. C'est **au statisticien responsable d'enquête qu'il appartient de définir quelle doit être l'unité statistique appropriée à l'objet de son enquête**[448].

Il faut bien distinguer à ce sujet trois concepts différents :

1. L'unité statistique : elle définit le champ des données qui seront prises en compte dans les agrégats diffusés au titre des résultats de l'enquête.
2. L'unité de collecte : elle définit le champ des données élémentaires qui sont collectées dans l'enquête.
3. L'unité interrogée : c'est l'unité à qui on envoie les questionnaires et qui répond à l'enquête.

Bien entendu, le plus souvent, ces trois unités sont confondues. Mais dans le cas d'un groupe profilé, ces trois unités (ou deux parmi les trois) **peuvent être distinctes** : on peut très bien imaginer qu'un même correspondant au niveau du groupe (unité interrogée) réponde pour chaque unité légale du groupe (unité de collecte) et qu'ensuite le statisticien regroupe les réponses au niveau de l'entreprise (unité statistique) (selon un mode de consolidation propre aux questions posées dans l'enquête).

Les dispositifs informatiques de gestion des enquêtes s'adaptent peu à peu à ces distinctions. Le contact direct avec les grands groupes du responsable d'une enquête, à l'occasion du profilage, permet alors de débattre avec eux, et compte tenu de leur organisation, de la meilleure façon de les interroger, qui n'est pas nécessairement la même d'un groupe à l'autre.

11.6 La prise en compte des entreprises de la cible 1 dans le système Ésane

11.6.1 La gestion des périmètres dans Ocsane

Les unités légales appartenant à une entreprise profilée doivent être placées **hors champ Ésane** tel qu'il a été défini au *chapitre 4*, pour éviter qu'elles ne soient prises en compte en double dans le calcul des agrégats et qu'elles soient traitées, en parallèle des profileurs, par les gestionnaires de l'Esa ou de l'EAP. Ceci est géré au niveau du répertoire Ocsane qu'on a présenté au § *1.2.1.2*. Dans le cas présent, Ocsane enregistre les contours des entreprises profilées et permet de piloter finement la collecte des enquêtes selon que les profileurs ont, ou pas, besoin de continuer à obtenir des réponses

[447] La définition qui en est donnée par Eurostat est la suivante : "*a method to analyse the legal, operational and accounting structure of an enterprise group at national and world level, in order to establish the statistical units within that group, their links, and the most efficient structures for the collection of statistical data*"

[448] On peut rappeler à ce sujet (comme on l'a précisé en note au § *9.4.6*) que le règlement FRIBS adoptera comme unité statistique pour les statistiques de court terme l'UAE (Unité d'Activité Économique, KAU pour kind of activity unit en anglais), dont l'unité légale est une très bonne approximation en France puisqu'on doit disposer d'un compte d'exploitation au niveau de l'UAE.

au niveau des unités légales (voir ce qu'on a dit ci-dessus au § 11.5.4 des stratégies d'interrogation en Esa-EAP ou au § 11.5.5 de la distinction unité statistique / unité de collecte : la chaîne Ésane est une des rares applications qui permette désormais d'opérer ces distinctions assez facilement).

Dans la chaîne Ésane on attend en tout cas une réponse finale (liasse fiscale, questionnaire, données d'emploi) **au niveau des entreprises profilées** (unités statistiques) et non au niveau des unités légales qui les composent, même si ce sont des éléments des unités légales (liasses fiscales et données d'emploi) (unités de collecte) qui permettent d'établir la valeur des caractéristiques au niveau de l'entreprise profilée.

11.6.2 La prise en compte des questionnaires Esa ou EAP

Qu'ils résultent d'une réponse directe des groupes ou d'une reconstitution par les profileurs, les questionnaires finaux des entreprises profilées subissent ensuite les mêmes contrôles et redressements que ceux des unités légales indépendantes comme on l'a vu dans les chapitres 2 et 5. Ceux-ci sont assurés par le profileur en charge du groupe correspondant.

11.6.3 L'intégration des liasses fiscales calculées

Dans un premier temps, les liasses fiscales des entreprises profilées sont calculées **en dehors de la chaîne de production Ésane**. Ce calcul est fait par les profileurs, comme on l'a vu au § 11.5.2, à partir des liasses fiscales des unités légales qui les composent et des paramètres de consolidation récupérés par les profileurs selon la méthode de consolidation adaptée au groupe (Bottom-Up ou Top-Down), grâce à des programmes ad-hoc mis à leur disposition.

Comme on souhaite disposer de liasses fiscales pour les entreprises profilées après chaque livraison de la DGFiP, **il est nécessaire de recalculer la liasse des entreprises après chaque livraison**. En effet, les liasses fiscales des unités légales des contours, absentes d'une livraison donnée de la DGFiP, sont extrapolées afin d'avoir un périmètre complet pour le calcul de la liasse de l'entreprise profilée. Toutefois si par la suite on obtient une liasse pour les unités absentes, il faut remplacer la liasse extrapolée par la « vraie » liasse reçue et donc relancer la consolidation

Les liasses consolidées finalement reconstituées sont ensuite introduites dans les bases Ésane à la fin du mois d'octobre, à l'instar de ce qui est fait pour les liasses fiscales en provenance de la DGFiP. Elles subissent donc les contrôles et redressements des chaînes amont et aval. Si le pro forma de la liasse fiscale de l'entreprise profilée doit être contrôlé par un gestionnaire, c'est le profileur qui s'en charge.

11.6.4 Le calcul des données d'emploi

Les données d'emploi des entreprises profilées sont calculées directement en sommant les données Clap des unités légales des périmètres. Les caractéristiques d'emploi sont en effet des données additives.

Ces calculs sont déjà implémentés dans la chaîne de production Ésane actuelle.

11.6.5 Une réconciliation traditionnelle

La liasse, le questionnaire et les données d'emploi des entreprises profilées sont réconciliés selon les mêmes règles que celles qui s'appliquent aux unités légales.

Si des contrôles manuels doivent être faits, ils sont réalisés par les profileurs.

11.6.6 2012 : année de transition et première année de prise en compte des résultats de la cible 1.

2012 a été la dernière année où les statistiques structurelles ont été calculées et publiées au départ en considérant systématiquement **les unités légales**[449] **comme des entreprises**. En même temps c'était la première année de collecte en double pour les groupes de la cible 1 pour lesquels le

[449] On a déjà mentionné toutefois au § 6.1.1 ainsi qu'au § 11.1 le cas des 5 entreprises profilées « **historiques** » qui étaient traitées dans Suse d'abord puis dans les premières campagnes d'Ésane comme s'il s'agissait d'unités légales dont les caractéristiques étaient calculées en dehors de ces systèmes d'information et intégrées ensuite telles quelles.

profilage était achevé. Les résultats de l'année 2012 **ont donc été publiés une première fois** en se basant **uniquement sur les unités légales**.

La campagne 2013 a ensuite été la première campagne au cours de laquelle les agrégats d'Ésane ont été calculés **dès le départ** en utilisant les **résultats des entreprises profilées** au sein des groupes de la cible 1 dont le profilage était achevé en 2014, c'est-à-dire qui avaient passé avec succès les 3 étapes présentées au § *11.5* et dont les résultats ont été ensuite intégrés dans la chaîne Ésane conformément à ce qui est exposé dans ce § *11.6*.

A la date du calcul (début 2015), 44 groupes étaient dans ce cas ; ils représentaient environ la moitié de la VA des grands groupes (soit 16% de la VA totale des entreprises). Ils avaient donné lieu à **109** entreprises profilées en leur sein et comptabilisaient 4224 unités légales. Cela représente 2,4 entreprises profilées par groupe en moyenne, **certains n'étant pas découpés du tout**, et le maximum étant de **6 entreprises profilées** au sein d'un même groupe.

Comme on l'a mentionné à la fin du § *11.5.4*, on a utilisé ensuite l'année 2012 de collecte « en double » pour recalculer **ex post** quels auraient été les résultats 2012 (chiffres d'affaires, EBE, valeur ajoutée, etc.) si on les avait calculés en s'appuyant cette fois-ci sur les valeurs des variables caractérisant les 109 entreprises profilées. Les résultats de 2012 **ont donc été publiés une seconde fois** en tenant compte des résultats **de ces 109 entreprises profilées** à la place des 4224 unités légales correspondantes. Cela a permis aux utilisateurs de constater l'impact du profilage et de calculer des évolutions 2013/2012 significatives c'est-à-dire se rapportant aux mêmes 109 entreprises en 2012.

Le nombre d'entreprises profilées et de groupes de la cible 1 évolue légèrement chaque année car :

- la durée d'obtention d'une année de collecte en double à la fois d'une liasse consolidée et d'une enquête (ESA ou EAP) pour chaque entreprise profilée nécessite un paramétrage spécifique du logiciel de consolidation du groupe et **varie d'un groupe à l'autre** ;
- les **restructurations** des groupes peuvent **modifier les contours et le nombre d'entreprises profilées** en leur sein ; il est fréquent qu'un groupe se défasse ou au contraire acquière une unité opérationnelle nouvelle (via des OPA[450] par exemple) ;
- de nouveaux groupes vont être profilés dans les années à venir pour accroître un peu la taille de la cible 1.

Le tableau 3 ci-dessous résume l'évolution de la cible 1 entre les campagnes 2012 et 2016 : que ce soit sur le nombre d'entreprises profilées ou sur celui des filiales, il montre bien les constantes restructurations qui affectent les grands groupes.

Tableau 3 : évolution des nombres de groupes de la cible 1, d'entreprises profilées et de filiales correspondantes :

Année	Nombre de groupes	Nombre d'entreprises profilées				Nombre d'UL filiales actives	Nombre de Co-entreprises
		Total	dont				
			Pérennes	Entrantes	Sortantes		
2012	**44**	109	109				
2013	44	109	109				29
2014	44	109	109			4224	7
2015	**51**	119	105	14	4	3922	1
2016	51	116	113	3	6	4149	1

11.7 Le répertoire SIRUS

11.7.1 Les principales fonctions du répertoire *[1.2], [1.12]*

Comme on l'a mentionné dans le paragraphe introductif du *chapitre 11*, le profilage des groupes et la prise en compte de nouvelles unités statistiques, les entreprises « profilées », a nécessité la mise en place d'un nouveau répertoire, dans le cadre du projet Resane. Intitulé **Sirus,** pour « système d'identification au répertoire des unités statistiques », il est désormais **le répertoire statistique de toutes les unités statistiques** dont **la statistique publique** peut avoir besoin, en particulier bien sûr des entreprises définies au sens économique dont on vient de parler (voir définitions ci-dessus), mais aussi des groupes eux-mêmes qui peuvent être des unités d'observation en tant que telles et bien sûr

[450] Offre Publique d'Achat (en bourse).

des unités légales qui restent l'unité statistique pertinente pour un certain nombre de questions (droit du travail, emploi, etc.).

En cela, Sirus diffère doublement de Sirene (système informatique pour le répertoire des entreprises et de leurs établissements), qui est un répertoire **inter-administratif,** exclusivement **d'unités légales**.

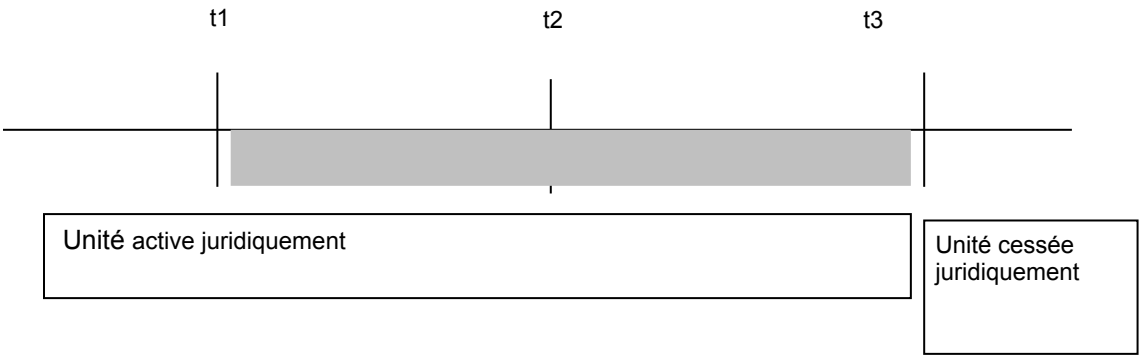
Les **objectif**s de Sirus sont nombreux. Ils n'ont été intégralement opérationnels qu'à partir de la campagne 2012. On en explicite un certain nombre ci-dessous :

1. Tout d'abord, il permet de **recenser** à la fois **les groupes, les entreprises** (au sens économique) **et les unités légales et établissements composant ces entreprises**. À terme, d'ailleurs, il enregistrera aussi **les réseaux** (type réseau « d'enseignes » comme Leclerc). Cela signifie qu'il dispose **d'identifiants** caractérisant chacune de ces entités (Cf. § *11.7.3*).
2. Sirus enregistre également **les liens** qui unissent ces différentes unités statistiques. Cela signifie qu'on connaît la liste des entreprises profilées qui composent un groupe, la liste des unités légales qui composent une entreprise, et la liste des établissements appartenant à une unité légale donnée. Par transitivité, on connaît bien entendu toutes les relations d'appartenance entre ces différentes entités.
3. Pour toutes ces unités, Sirus stocke également **des caractéristiques statistiques** comme le chiffre d'affaires, le classement sectoriel (APE), l'effectif salarié, grâce à des mises à jour provenant d'**une multitude de sources** : Sirene, Lifi, BCE, Clap (effectifs), Ésane (chiffre d'affaires, branches), les résultats d'enquêtes (Esa, EAP), etc. En cas de divergence, Sirus arbitre « la valeur » à retenir. Des **règles automatiques de priorité entre sources** ont été mises en place pour chaque caractéristique. Ainsi, l'Esa est la source prioritaire pour l'APE, Clap pour les données d'emploi, etc.
4. En tant que répertoire statistique, Sirus est la **base de sondage** de toutes les enquêtes entreprises de la statistique publique. Sa mise en production a d'ailleurs permis de centraliser auprès du département des méthodes statistiques le tirage de la totalité des échantillons.
5. Sirus permet de **mettre à disposition** des statisticiens d'entreprises **des référentiels de populations de référence**. Ainsi, une entreprise, à un instant donné et pour une période de référence donnée, se verra affectée à la même population de référence dans toutes les enquêtes et y aura les mêmes caractéristiques.
6. Sirus met à disposition des informations statistiques nouvelles et exclusivement accessibles par ce répertoire ; en particulier **le classement en quatre catégories,** rappelé au début du § *11.4*, [micro-entreprises, petites et moyennes entreprises (PME), entreprises de taille intermédiaire (ETI), grandes entreprises] issu du décret 2008-1354 du 18 décembre 2008 d'application de la loi française de modernisation de l'économie (LME) qui a défini les « catégories d'entreprises ». Cette information est calculée dans le répertoire statistique lui-même grâce aux caractéristiques mentionnées au point 3 et diffusée ensuite[451], via le site Insee.fr ou via la diffusion des listes Sirene.
7. En outre, **Sirus permet d'enregistrer la charge** *[1.5]* de réponse aux enquêtes statistiques, c'est-à-dire le temps passé par les entreprises à remplir les questionnaires des enquêtes statistiques, désormais systématiquement demandée dans les enquêtes. Ceci permet d'une part de connaitre pour chaque unité statistique de SIRUS la charge statistique qui pèse réellement sur elle et d'autre part, la mise en œuvre de nouvelles méthodes pour la coordination d'échantillons *[2,14]*, afin de **mieux répartir cette charge** sur l'ensemble des entreprises.
8. Enfin, dernier objectif, **Sirus gère une « cessation statistique » pour les unités**. Ainsi, il permet de distinguer une unité ayant une activité économique d'une unité active juridiquement mais sans consistance économique (ces unités sont communément appelées aujourd'hui de « **fausses actives** » dans Sirene). Le schéma ci-dessous montre l'intérêt de la cessation statistique :

[451] A noter que seule la partition en PME, ETI et grande entreprise est diffusée. La distinction des micro-entreprises au sein des PME n'est pas rendue publique en raison des contestations auxquelles elle pourrait donner lieu quant à l'appréciation précise du nombre de salariés ETP autour de 10.

a) Situation passée dans SIRUS (et SIRENE)

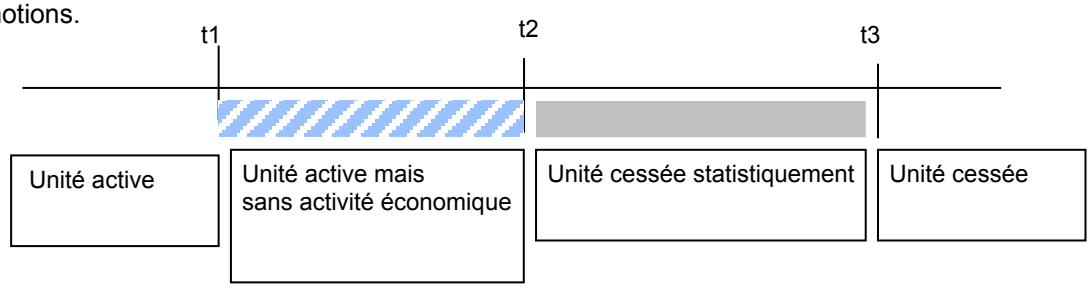
Prenons une entreprise qui cesse son activité économique à une date t1[452]. En t2, elle cesse également d'employer des salariés. En t3, elle signale sa cessation juridique (que Sirene et donc SIRUS enregistrent). Dans notre exemple, on suppose (et ceci n'est pas aberrant dans les faits) que t3=t1+ 5ans.



Pendant 5 ans, l'unité n'a donc plus d'activité mais est considérée comme active par Sirene. L'unité est donc une « fausse active » durant ces 5 ans ; 5 ans au cours desquels elle est non-répondante aux enquêtes statistiques et peut être extrapolée à tort.

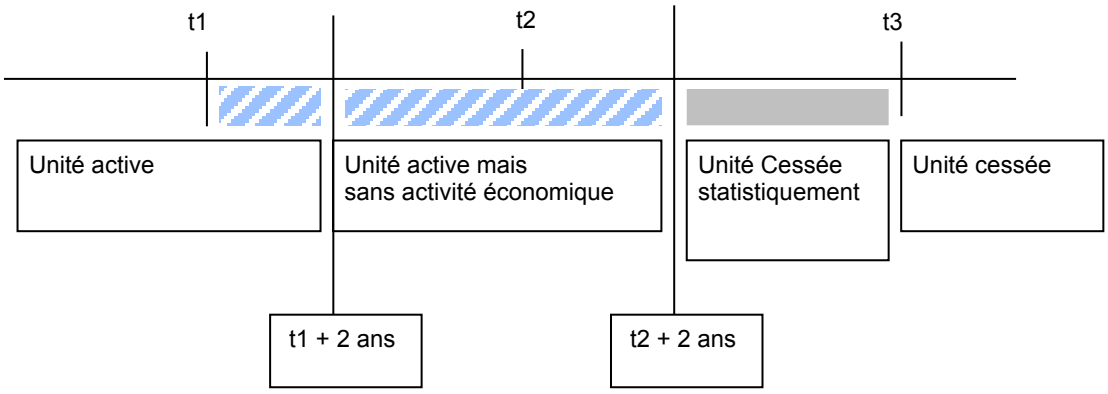
b) Situation idéale

On souhaiterait savoir que l'unité n'a plus d'activité économique depuis t1, et est cessée statistiquement depuis t2. **Toute unité employeuse sera considérée comme active** mais certaines unités employeuses n'ont plus d'activité économique. En effet, la présence d'effectifs n'est pas garante d'une activité économique. Une unité peut continuer à employer des personnes pour gérer la liquidation économique par exemple. Sirus gère ces deux notions.



c) Situation nouvelle dans SIRUS

Dans les faits, il est impossible de mesurer immédiatement l'arrêt de l'activité économique : t1, t2 et t3 sont inconnus. Il a donc fallu établir des règles pour considérer une unité sans activité économique. Ainsi, avec les conventions retenues, **si l'unité ne fait pas de déclaration fiscale pendant 2 ans, SIRUS considère qu'elle n'a plus d'activité économique**. Deux ans après l'arrêt de données d'emploi, SIRUS **cesse statistiquement l'unité**.

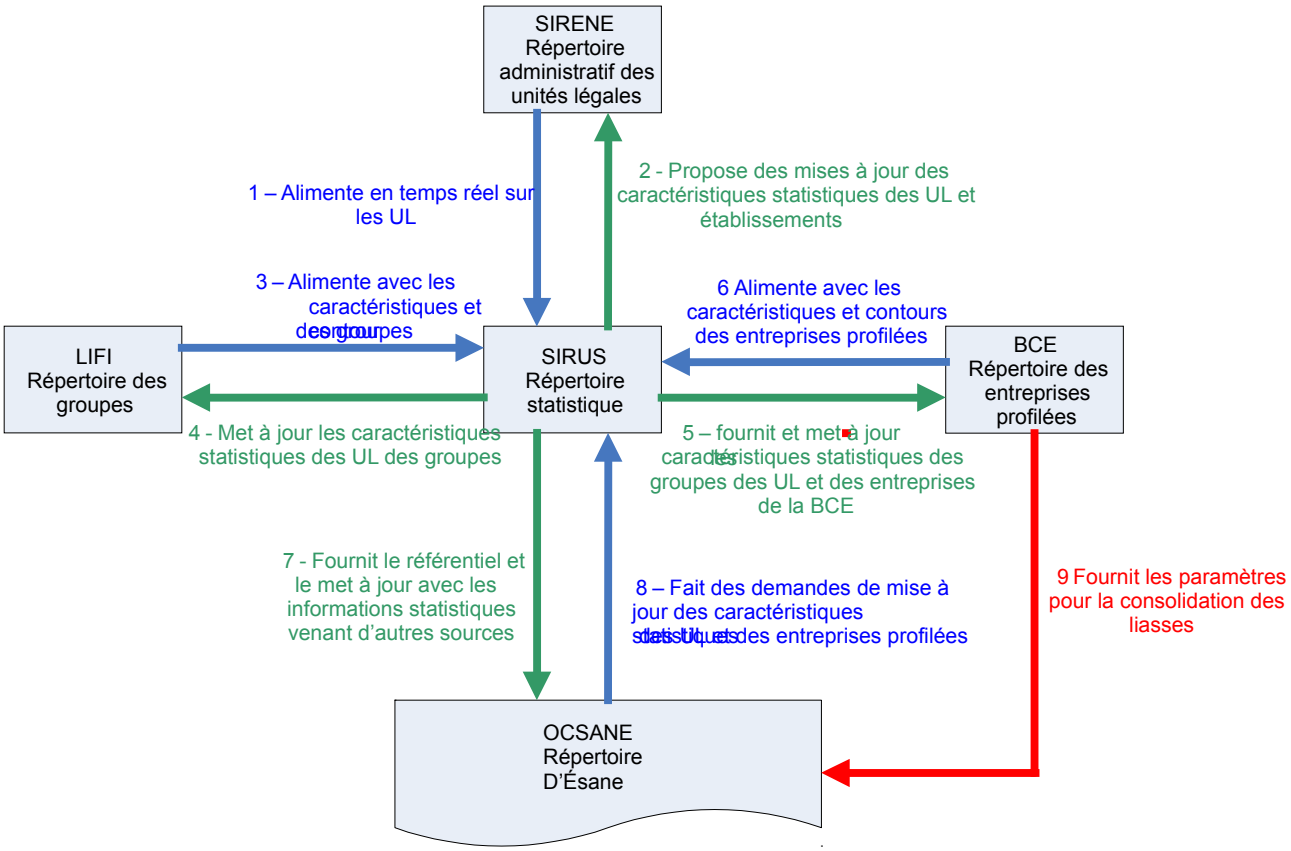


Bien évidemment en pratique, il est possible que t1 = t2 (elle cesse à la fois son activité économique et d'être employeuse).

[452] Cette date est bien entendu inconnue des statisticiens.

11.7.2 Les liens de Sirus avec les autres systèmes d'information

Le schéma ci-dessous synthétise les relations entre les trois bases de données spécialisées sur un type d'unités statistiques (Sirene, Lifi et la BCE), le répertoire interne d'Ésane (Ocsane) et Sirus :



1 Les flux de SIRENE vers SIRUS
SIRENE alimente quotidiennement SIRUS avec les mises à jour qu'il enregistre chaque jour sur les unités légales et leurs établissements, y compris les créations de nouvelles unités.

2 Les flux de SIRUS vers SIRENE
En contrepartie, SIRUS alimente SIRENE hebdomadairement avec les mises à jour des variables statistiques qu'il a faites durant la semaine pour les unités légales et les établissements. SIRUS remplace SIRENE dans son rôle de correspondant privilégié des opérations statistiques. Ainsi, par exemple, Clap (cf. *chapitre 7*) va mettre à jour les effectifs des unités légales et des établissements dans SIRUS, et c'est SIRUS qui met ensuite à jour SIRENE. De même, les demandes de mise à jour d'APE en provenance d'Ésane sont d'abord validées par SIRUS avant d'être transmises à SIRENE. Il n'y a plus de lien direct entre SIRENE et les applications statistiques du système statistique public.

3 Les flux de Lifi vers SIRUS
Deux fois par an, Lifi alimente SIRUS avec le contour des groupes en unités légales et les caractéristiques statistiques des groupes :
- en octobre N avec la version provisoire des contours N-1 ;
- en mars N+1 avec la version définitive des contours N-1.

4 Les flux de SIRUS vers Lifi
SIRUS initialise Lifi avec les caractéristiques des unités légales qui intéressent Lifi. Par la suite, SIRUS alimente Lifi hebdomadairement avec toutes les mises à jour des variables statistiques qu'il a faites durant la semaine pour les unités légales qui intéressent Lifi.
En revanche, seul Lifi alimente SIRUS pour les groupes. SIRUS ne renverra donc aucune mise à jour à Lifi sur les groupes.
Concernant les groupes, SIRUS ne joue qu'un rôle de boite aux lettres. Il permet uniquement de diffuser à ses partenaires la liste des groupes, leurs caractéristiques et contours en unités légales mais ne gère pas de mises à jour.

5 Les flux de SIRUS vers la BCE
Trois fois par an, SIRUS alimente la BCE de l'année N, en octobre N avec les contours provisoires N-1 des groupes, les caractéristiques statistiques des groupes et des unités légales qu'ils contiennent ; en mars N+1 avec les contours définitifs N-1 et en janvier N+2 avec les provisoires N.
Par la suite, SIRUS alimente la BCE hebdomadairement avec toutes les mises à jour des variables statistiques qu'il a faites durant la semaine pour les unités légales et les entreprises profilées qui intéressent la BCE.
Ainsi, par exemple, si la réponse à l'Esa conduit à modifier l'APE d'une entreprise profilée, cette mise à jour est faite dans un premier temps dans SIRUS qui renvoie ensuite à la BCE cette nouvelle APE.
De la même façon, lorsque Lifi envoie ses contours définitifs, SIRUS met à jour la BCE avec ces contours définitifs. Il n'y a aucun lien direct entre Lifi et la BCE.

6 Les flux de la BCE vers SIRUS
La BCE envoie à SIRUS hebdomadairement toutes les mises à jour qui concernent les entreprises profilées. Ces mises à jour peuvent concerner des caractéristiques statistiques des entreprises profilées ou la définition du périmètre des entreprises profilées en unités légales.
Un profileur peut également demander à SIRUS de lui fournir des informations sur des unités légales qui ne sont pas dans les contours des groupes de Lifi[453] et ainsi mettre à jour la BCE. Enfin, un profileur peut aussi faire des demandes de mise à jour à SIRUS pour l'état et le code APE des unités légales. Ces demandes peuvent toutefois être refusées par SIRUS. En effet, si la BCE est considérée comme la source prioritaire pour toute variable concernant les entreprises profilées, elle ne l'est pas pour les unités légales.
À partir du 2e exercice[454] où une entreprise profilée est présente dans la BCE, la BCE fournit également à SIRUS des informations sur les prédécesseurs et les successeurs de cette entreprise profilée. Les périmètres des entreprises en unités légales étant sujets à modifications d'une année sur l'autre, il est nécessaire de suivre la continuité des périmètres pour les études longitudinales ou les contrôles statistiques temporels par exemple, ce qui implique de définir la notion de continuité (Cf. le § suivant *11.7.3*).

7 Les flux de SIRUS vers Ésane
SIRUS fournit chaque année en octobre N à Ocsane le référentiel de la campagne Ésane sur l'exercice N. Par la suite, SIRUS alimente Ésane hebdomadairement avec toutes les mises à jour des variables statistiques qu'il a faites durant la semaine pour les unités légales et les entreprises profilées qui intéressent Ésane.
SIRUS alimente notamment Ocsane lorsqu'il calcule la cessation statistique (les unités statistiquement cessées ne doivent pas être extrapolées).

8 Les flux d'Ésane vers SIRUS
Ésane fait des demandes de mise à jour pour les caractéristiques des unités légales et entreprises profilées du champ Ésane.
Ces demandes de mise à jour concernent notamment l'APE, le chiffre d'affaires, la ventilation en branches du chiffre d'affaires (où Ésane est la source prioritaire), et l'état des unités (où Ésane est la 2e source prioritaire après SIRENE).

9 Les flux de la BCE vers Ésane
La BCE met à disposition d'Ésane les informations sur la collecte des enquêtes dans la phase de préparation de la collecte de façon à pouvoir mettre en œuvre les stratégies de collecte présentées au § *11.4.3* ainsi que le pro forma de la liasse des entreprises profilées afin qu'elles puissent être prises en compte dans Ésane, en fin de campagne.

[453] Ceci vient du fait que Lifi a structurellement un exercice de retard par rapport au profilage, qui en octobre N met au point les périmètres N des entreprises profilées. Il est donc possible que des unités légales qui sont entrées dans le groupe au cours de l'exercice N (et qui ne sont donc pas connues de Lifi N-1) soient consolidées dans une entreprise profilée.
[454] 2013 pour les fameuses 109 entreprises qui ont été introduites pour le calcul de l'année 2013. En effet, comme on l'a vu au § *11.6.6*, on a aussi leurs résultats sur 2012. On a donc déjà sur 2013 l'information sur la continuité de ces entreprises entre 2012 et 2013 et sur leurs éventuelles successeuses en 2013 lorsqu'il n'y a pas continuité entre 2012 et 2103 (Cf. § *11.7.3* pour la définition de ces notions).

11.7.3 La continuité des groupes et des entreprises

Affecter tous les ans un identifiant aux groupes ou aux entreprises profilées implique que l'on est capable de dire systématiquement si l'on a affaire à la même entité ou pas, deux années de suite. C'est par exemple être capable de dire si Arcelor-Mittal est en fait la suite du groupe Mittal, celle du groupe Arcelor ou s'il s'agit d'un groupe nouveau qui doit donc bénéficier d'un nouvel identifiant. Si on veut affecter un identifiant unique à un groupe, on ne pourra pas accepter qu'il soit à la fois la continuation de Mittal et celle de Arcelor ; il faudra choisir.

Sur l'exemple des groupes, on va donner ici **le principe qui est mis en œuvre par l'Insee** pour trancher ces cas aussi bien au sein du système d'information Lifi pour les groupes que dans la BCE pour les entreprises.

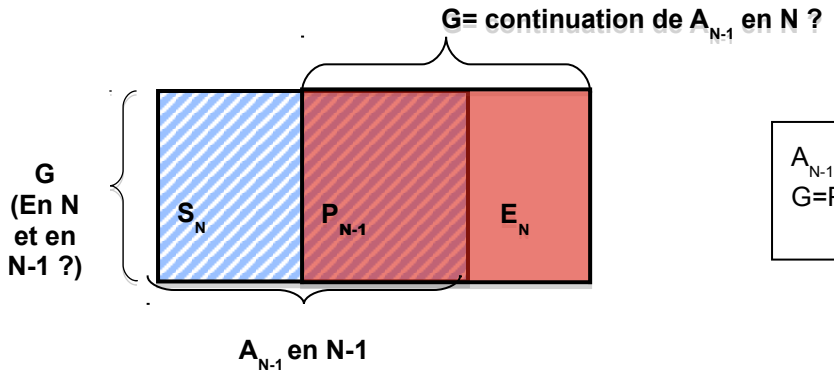
Au 31/12 de l'année N, un groupe G est composé d'un certain nombre d'unités légales. Il n'a pas encore d'identifiant puisqu'on ne sait pas encore s'il est ou non la continuation d'un groupe existant en N-1.

Appelons A_{N-1} un groupe existant en N-1 qui pourrait être un « antécédent » de G. Les unités légales qui composent A_{N-1} en N-1 peuvent être devenues en N indépendantes ou se retrouver dans de nombreux autres groupes (suite aux restructurations, achats et ventes que peuvent faire les groupes) ; et réciproquement, G a pu acquérir de nouvelles filiales qui étaient auparavant indépendantes ou qui appartenaient à d'autres groupes. En revanche, on sait que tous les groupes qui composent l'ensemble des { A_{N-1}} possèdent **au moins une unité légale en commun avec G.**

Par rapport à G, tout groupe A_{N-1} est donc constitué (voir graphique ci-dessous) :

- d'unité légales qui sont encore présentes dans le groupe G en N, (par hypothèse il y en au moins une) ; soit {P_{N-1}} cet ensemble d'unités légales (P comme « pérennes ») ;
- d'autres qui sont « sorties » en N, soit {S_N} cet ensemble ; (cet ensemble peut être vide) ;
- d'autres enfin qui sont "entrées » en N, soit {E_N} cet ensemble ; (cet ensemble peut être vide).

Dire que G est la « continuation » d'un des groupes A_{N-1}, c'est dire qu'ils se ressemblent suffisamment donc que l'on retrouve beaucoup d'éléments de G dans A_{N-1}. On va donc assez naturellement considérer l'ensemble des {P_{N-1}} (on sait que ce n'est pas l'ensemble vide sauf si le groupe se crée à partir d'UL indépendantes) et **prendre le plus gros en N**. Le plus gros en quel sens ? On pourrait sûrement considérer plusieurs variables de taille mais on a retenu l'emploi car intuitivement on a envie de dire que deux groupes sont la continuation l'un de l'autre s'il y a les mêmes salariés avant et après. On sélectionne donc le groupe A_{N-1} qui correspond au P_{N-1} maximum mesuré sur l'année N ; soient P_{N-1}^{Max} cet ensemble d'unités légales pérennes et A_{N-1}^{Max} le groupe auquel il appartient en N-1. Pour que G et A_{N-1}^{Max} soient bien en continuité l'un de l'autre on veut que P_{N-1}^{Max} **représente à la fois la majorité des salariés de** A_{N-1}^{Max} **en N-1 et de G en N**.



$A_{N-1} = S_N + P_{N-1}$
$G = P_{N-1} + E_N$

Il est facile de voir que cette condition équivaut à :

- en N-1 : $P_{N-1}^{Max} \subset A_{N-1}^{1}$ et effectif$_{N-1}$ (P_{N-1}^{Max})>effectif$_{N-1}$ (S_N^1)
- en N : $P_{N-1}^{Max} \subset G$ et effectif$_N$ (P_{N-1}^{Max})>effectif$_N$ (E_N^1)

Ou de façon équivalente :

- En N-1 : effectif(A_{N-1}^{Max}) ≥ effectif (P_{N-1}^{Max}) **> 50%** effectif(A_{N-1}^{Max}) (supérieur strict)
- En N : effectif G ≥ effectif (P_{N-1}^{Max}) **> 50%** effectif G (supérieur strict)

1. **Si tel est le cas, on considère que G est la continuation de A_{N-1}^{Max} . On affecte alors en N à G, l'identifiant du groupe A_{N-1}^{Max} en N-1.**
   Comme on a pris « le » plus gros sous ensemble de G en N[455], on est sûr que seul P_{N-1}^{Max} peut être majoritaire au sein de G en N. S'il existe il est donc unique. Il suffit de vérifier s'il est aussi majoritaire (strict) dans A_{N-1}^{Max} en N-1 pour savoir si G est en continuité avec A_{N-1}^{Max} .
2. **Si P_{N-1}^{Max} représente la majorité des salariés de G en N** (mais pas nécessairement celle de A_{N-1}^{Max} en N-1), on dit que A_{N-1}^{Max} et G sont en **continuité arrière**[456] [ou que G est un des successeurs de A_{N-1}^{Max} ou que A_{N-1}^{Max} est l'affluent majoritaire de G, (apportant la majorité de l'effectif de G)].
3. **Si P_{N-1}^{Max} représente la majorité des salariés de A_{N-1}^{Max} en N-1** (mais pas nécessairement celle de G en N), on dit que A_{N-1}^{Max} et G sont en **continuité avant** [ou que A_{N-1}^{Max} est un des prédécesseurs de G ou que G est l'héritier majoritaire de A_{N-1}^{Max} , (héritant de la majorité de l'effectif de A_{N-1}^{Max})].
4. Enfin si P_{N-1}^{Max} ne représente ni la majorité des salariés de A_{N-1}^{Max} en N-1 ni celle de G en N, on dit que **G est une création stricte et on lui affecte de toutes façons un nouvel identifiant.**

On aura remarqué que bien entendu il y aura continuité si et seulement si A1 et G sont à la fois en continuité avant et arrière. **Dans tous les autres cas, on affecte un nouvel identifiant au groupe G**.

Dans la pratique, l'algorithme est un petit peu plus complexe car :

1. On a considéré jusqu'ici qu'une unité légale appartenait ou pas à un groupe, donc que ses effectifs comptaient pour 100 % ou 0 % dans les effectifs du groupe. Mais dans le cas de joint venture pour les groupes, ou d'unités légales « partagées » entre entreprises profilées pour la continuité des entreprises profilées[457], il faut tenir compte d'un coefficient (correspondant à une consolidation proportionnelle d'une joint-venture ou au coefficient de partage d'une unité légale entre plusieurs unités légales). Mais cela ne change pas le raisonnement global (même si cela le complexifie un peu : en cas d'unité légale appartenant pour 50 % à deux groupes ou deux entreprises différentes, on peut trouver des cas où ces deux entités sont en continuité avec le même groupe ou la même entreprise de départ ; il faut alors choisir le P_{N-1} qui assure la plus grosse majorité en pourcentage).
2. Il y a plusieurs notions d'emploi dans Ésane. On a retenu l'emploi au 31/12 fourni par Clap qui est celui qui a le moins de valeurs manquantes.

Pour finir, on voit que ce critère assez simple **permet d'attribuer des identifiants et de gérer une notion de continuité** de groupes ou d'entreprises profilées dans le temps.

[455] Il faut absolument qu'il n'y en ait qu'un sans quoi on pourrait avoir des ex æquo : un groupe formé de deux filiales, qui proviennent de deux groupes différents l'année précédente et qui ont le même nombre de salariés, ne peut être en continuité avec aucun de ces deux groupes puisque chaque filiale ne fait que 50 % du total.
[456] Au sens où on peut remonter de G à A_1, mais on ne peut pas descendre car A_1 a plusieurs successeurs.
[457] Voir la définition donnée au § *11.2* ; de tels cas sont rares mais peuvent se produire. Par exemple si une unité légale sert à gérer les effectifs d'un groupe qui comprend 2 entreprises profilées en les mettant à disposition des deux EP, on pourra dire que l'unité légale est affectée partiellement aux deux EP au pro rata des effectifs qu'elle gère pour leur compte.

Nota : en revanche, on considère implicitement que les unités légales ont des contours fixes dans le temps alors qu'on a vu au *chapitre 3* qu'on était amené à prendre en compte des restructurations entre unités légales (fusions, scissions, transferts d'actifs entre unités légales). Il y a un certain nombre de règles qui permettent de déterminer dans Sirene à quelles conditions une unité légale garde ou non le même identifiant après une restructuration qu'une autre unité légale qui a participé à la restructuration. Pour mesurer le caractère « pérenne » d'une unité légale, l'algorithme de continuité des groupes s'appuie sur les règles de Sirene et ne considère qu'une UL A est pérenne entre N-1 et N que si elle garde le même numéro d'identification dans Sirene.

12. Déroulement chronologique d'une campagne

Les tableaux qui suivent visent à faire **le lien** entre les **étapes concrètes d'une campagne** Ésane et les **parties de l'Insee Méthodes Ésane** qui décrivent l'étape en question sur le plan méthodologique. Ils donnent une idée de **la complexité et du nombre d'étapes** qui se déroulent souvent simultanément. Ils servent aussi à montrer qu'il y a bien un déroulement logique dans le temps qui permet d'effectuer l'ensemble des traitements qui ont été présentés, alors qu'on peut quelquefois se demander comment et dans quel ordre effectuer tous ces traitements. On voit aussi voir grâce aux dates de début et de fin que de nombreuses étapes se chevauchent.

Ils montrent aussi qu'en plus des étapes mettant en œuvre des programmes informatiques, il y a de nombreuses étapes de travail humain et organisationnel. Certaines de ces étapes purement organisationnelles, comme l'envoi des questionnaires, n'ont évidemment pas de « pendant méthodologique ».

Les tableaux figurent les dates programmées pour la campagne 2015 sur l'année de constat 2014. Le calendrier est évidemment réajusté chaque année en fonction du calendrier de l'année et du déroulement de la campagne précédente.

	Les étapes d'une campagne Esane pour une année N donnée.	**N =**	**2014**	
Numéro d'opération	**OPERATIONS**	**Date de**		**Paragraphe correspondant de l'Insee Méthode**
		début	fin	
	1. PREPARATION DE LA CAMPAGNE (échanges BCE - Sirus - Ocsane)			
	1.1. Contours des profilées			**Chapitre 11**
1	**1.1.1. Détermination de la liste des entreprises pour l'année N (cibles 1,2,3)**			**11.3**
2	**1.1.2. Liste des UL du contour des entreprises profilées pour l'année N**		**15-Oct-14**	**11.5.1**
3	**1.1.3. Mise à jour du contour des profilées de l'année N-1 (à discuter)**			**11.6.2**
	1.2. Actualisation des données utiles au lancement d'Ocsane		**20-Nov-2014**	
4	**1.2.1. Alimentation de Sirus par la BCE**			
5	**1.2.2. Mise à jour de Sirus**			**11.6.2**
6	**1.2.3. Alimentation d'Ocsane par Sirus (y compris évol. de**	**10-Oct-14**	**30-Oct-14**	**4.1**
7	**1.2.4 Intégration du contour des profilées**	**5-Nov-14**	**20-Nov-14**	
	1.3. Constitution des bases de sondage	**15-Nov-14**	**30-Nov-2014**	
	1.3.1.Initialisation d'Ocsane	**15-Nov-14**	**15-Mar-14**	
8	1.3.1.1. Métropole	15-Nov-14	20-Nov-14	
9	1.3.1.2. Dom	1-Mar-14	15-Mar-14	
10	**1.3.2. Constitution/livraison de la base de sondage de l'EAP (au DSCT)**	**20-Nov-14**	**30-Nov-14**	**5.2.2**
11	**1.3.3. Constitution/livraison de la base de sondage de l'Esa (au DRISS)**	**20-Nov-14**	**30-Nov-14**	**5.1.2 à 5.1.4 ;**
	1.4. Lancement de la campagne (dont tirages d'échantillon)	**15-Feb-15**	**30-janv.-15**	
12	**1.4.1 Chargement des échantillons de l'Esa (UL et profilées)**	**15-Feb-15**	**30-Mar-15**	
13	**1.4.2. Marquage des unités du sous-champ 1 (UL, profilées cible 1, 2 et 3)**	**1-Nov-15**	**30-Jan-15**	**4.3**
14	**1.4.3. Mise à jour des Tables utiles au service esane (TGE, correspondant, codif autom…)**	**01-janv.-15**	**30-janv.-15**	

	2. CAMPAGNE DE L'ESA (de la collecte au traitement des données)	mai-14	août-15	
	2.1. Préparation des questionnaires	**02-mai-14**	**30-janv.-15**	
15	**2.1.1. Mise à jour des questionnaires avec les MOA (tous les deux ans)**	**02-mai-14**	**30-juin-14**	
16	**2.1.2. Préparation de l'envoi des questionnaires (prestataire Seven)**	**01-déc.-14**	**01-fevrier-15**	
17	**2.1.3. Préparation de la saisie des questionnaires (prestataire SPS)**	**01-oct.-14**	**30-janv.-15**	
	2.2. Réalisation de l'enquête (envois et saisie)	**01-fev-15**	**30-nov.-15**	
	2.2.1. Envois des questionnaires par vagues	01-fev-15	**01-août-15**	
18	2.2.1.1. Envois des questionnaires des vague n°1 (15 000) et n°2 (90 000)	01-fev-15	30-avr.-15	
19	2.2.1.2. Envois des vague 3 (7000) vague 4 (1000)	01-mai-15	01-août-15	
20	2.2.1.3. Envois des questionnaires de profilées cibles 2 et 3 ?			
21	2.2.1.4. Envois des questionnaires aux profilées de la cible 1	1-Fev-15	30-avr.-15	
	2.2.2. Saisie des questionnaires			
22	2.2.2.1. Saisie des questionnaires (y compris des profilées)	01-fev-15	30-nov.-15	
23	2.2.2.2. Saisie manuelle des questionnaires du secteur de l'informatique	15-avr.-15	15-oct.-15	
24	2.2.2.3. Intégration des données par le service Esane (saisies + images)	01-fev-15	30-nov.-15	
25	2.2.2.4. Contrôle de la saisie par le service Esane (sur un échantillon)	01-avr.-15	30-avr.-15	
26	**2.2.3. Relances, mises en demeure et CNR**	**01-avr.-15**	**30-août-15**	
	2.3. Contrôle des résultats	**15-fev-15**	**15-mars-16**	**Chapitre 2 et 3**
	2.3.1 Codification des branches et Contrôle de l'APE (Processus PSP1)	**15-févr.-15**	**30-août-15**	**5.4.2 et 5.4.3**
27	2.3.1.1 premiers traitements PSP1 (avec CQR simplifié sans macros controles)	**15-févr.-15**	**01-mai-15**	"
28	2.3.1.2 Suite des traitements PSP1 (avec CQR complet avec macro-contrôle)	**01-mai-15**	**30-août-15**	"
	2.3.2 Contrôle Ventes Branches (processus PSP2)	**15-mai-15**	**15-oct.-15**	"
29	2.3.2.1 Tests	15-mai-15	01-juin-15	"
30	2.3.2.2 Production	01-juin-15	15-oct.-15	"
	2.3.3. CCP : contrôle des caractéristiques particulieres (PSP3)	**01-oct.-15**	**15-mars-16**	"
31	2.3.3.1 Tests	**01-oct.-15**	**15-oct.-15**	"
32	2.3.3.2 Production	**15-oct.-15**	**15-mars-16**	"

	Les étapes d'une campagne Esane pour une année N donnée.	N =	2014	
Numéro d'opération	OPERATIONS	Date de		Paragraphe correspondant de l'Insee Méthode
		début	fin	
	3. CAMPAGNE IEG (liasses fiscales)	01-avr.-15	15-mars-16	
	3.1. Réception - chargement des fichiers DGFiP	01-avr.-15	30-mai-16	Intro. Chap. 6
33	3.1.1. Préparation du chargement des liasses (nouvelles variables, formats…)	01-avr.-15	15-mai-15	"
34	3.1.2. Réception intégration des liasses "anticipées"	25-juin-15	25-juil.-15	"
35	3.1.3. Réception et intégration des liasses "intermédiaires"	15-sept.-15	15-oct.-15	"
36	3.1.4. Réception et intégration des liasses "normales"	15-oct.-15	30-oct.-15	"
37	3.1.5. Réception et mise en forme des liasses "complémentaires"	15-mai-16	30-mai-16	"
	3.2. Traitements en vue de la production des résultats semi-définitifs (PSCP)	20-juil.-15	30-sept.-15	Chapitre 2, 3 & 6
38	3.2.1 Tests	20-juil.-15	20-août-15	"
39	3.2.2 Production	20-août-15	30-sept.-15	"
	3.3. Traitements en vue de la production des résultats définitifs (PSCD)	01-oct.-15	15-mars-16	Chapitre 2, 3 & 6
40	3.3.1 Tests	01-oct.-15	15-oct.-15	"
41	3.3.2 Production	15-oct.-15	15-mars-16	"
	4. RÉCONCILIATION des données (REDI)	30-sept.-15	30-mars-16	Chapitre 8
	4.1. Traitements en vue de la production des résultats semi-définitifs	30-sept.-15	01-nov.-15	
42	4.1.1 Passage automatique de Redi (Redi balai)		01-nov.-15	8.6
	4.2. Traitements en vue de la production des résultats définitifs	30-nov.-15	30-mars-16	
	4.2.1 Intégration des données d'emploi	30-nov.-15	15-mars-16	7.2.1.2 7.3 et 7.4
43	4.2.1.1. Tests	01-oct.-15	01-nov.-15	
44	4.2.1.2. Production	01-nov.-15	20-mars-16	
45	4.2.2 Passage automatique de Redi (Redi balai) sur les données finales	20-mars-15	30-mars-16	8.6
	5. DIFFUSION des résultats	30-sept.-15	15-juil.-16	
	5.1. Résultats préliminaires	01-oct.-15	.	9.4.2
	5.1.1. Production d'une estimation de l'année N	01-oct.-15	15-oct.-15	
46	5.1.1.1. Calculs de 6 variables par exploitation des liasses anticipées et imputation des liasses manquantes	01-oct.-15	10-oct.-15	9.4.2
47	5.1.1.2. Contrôle qualité des résutlats et rectifications si nécessaire	10-oct.-15	15-oct.-15	10.3 et 10.4
48	5.1.1.3. Production du fichier de diffusion	01-oct.-15	15-oct.-15	
	5.1.2. Diffusion vers Eurostat	15-oct.-15	30-oct.-15	9.4.2
49	5.1.2.1. Mises en forme spécifiques	15-oct.-15	20-oct.-15	
50	5.1.2.2. Contrôles spécifiques à Eurostat	20-oct.-15	25-oct.-15	
51	5.1.2.3. Envoi des résultats	25-oct.-15	30-oct.-15	
	5.1.3. Diffusion sur Insee.fr	01-nov.-15	31-déc.-15	

	Les étapes d'une campagne Esane pour une année N donnée.	N =	2014	
Numéro d'opération	OPERATIONS	Date de		Paragraphe correspondant de l'Insee Méthode
		début	fin	
	5. DIFFUSION des résultats (suite)	30-sept.-15	15-juil.-16	
	5.2. Résultats semi-définitifs	30-sept.-15	28-févr.-16	
	5.2.1. Production des fichiers de diffusion	**30-sept.-15**	**30-janv.-16**	
55	5.2.1.1. Extraction des données de production vers la diffusion	03-nov.-15	05-nov.-15	
56	5.2.1.2. Exécution de traitements particuliers (remise à champ comparable etc.)	05-nov.-15	15-nov.-15	4.6.3.2
57	5.2.1.3. Exécution des programmes de calcul des résultats Esane	15-nov.-15	09-janv.-16	5.6.2 9.2 et 9.5
58	5.2.1.4. Fichiers Fare, agrifin…	16-déc.-15	30-janv.-16	
	5.2.2. Contrôles/ correction des données	**05-nov.-15**	**09-janv.-16**	
59	5.2.2.1. Contrôles sur données individuelles	05-nov.-15	15-nov.-15	
60	5.2.2.2. Contrôles sur les agrégats	20-nov.-15	16-déc.-15	
61	5.2.2.3. Modification des données	20-nov.-15	16-déc.-15	
62	5.2.2.4. Nouvelle exécution des chaînes de diffusion avant livraison au DCN	16-déc.-15	22-déc.-15	
63	5.2.2.5. Eventuelles modifications des données suite aux échanges avec le DCN	22-déc.-15	06-janv.-16	
64	5.2.2.6. Eventuelle dernière exécution des chaînes de diffusion	06-janv.-16	09-janv.-16	
	5.2.3. Livraison des résultats semi-définitifs aux DCN + DSS	**22-déc.-15**	**28-févr.-16**	
65	5.2.3.1. Mise à disposition des résultats semi-définitifs N et N-1, sur le champ N-1 et méthode N	22-déc.-15	10-janv.-16	
66	5.2.3.2. Diffusion d'une note décrivant les données	22-déc.-15	15-janv.-16	
	5.2.4. Publication des résultats (pas encore fait)			
67	5.2.4.1. Mise en forme spécifique des données (dont délais de paiement)			
68	5.2.4.2. Rédaction d'un 2 pages			
69	5.2.4.3. Echanges/validation avec le DOE		28-févr.-16	

	Les étapes d'une campagne Esane pour une année N donnée.	N =	2014	
Numéro d'opération	OPERATIONS	Date de		Paragraphe correspondant de l'Insee Méthode
		début	fin	
	5. DIFFUSION des résultats (suite)	30-sept.-15	15-juil.-16	
	5.3. Résultats définitifs	01-avr.-16	31-mars-17	
	5.3.1. Production des fichiers de diffusion	**08-avr.-16**	**30-juil.-16**	
70	5.3.1.1. Extraction des données de production vers la diffusion	08-avr.-16	09-avr.-16	
71	5.3.1.2. Exécution de traitements particuliers (remise à champ comparable etc.)	09-avr.-16	14-avr.-16	4.6.3.2
72	5.3.1.3. Exécution des programmes de calcul des résultats Esane	14-avr.-16	01-juin-16	9.2 et 9.5
73	5.3.1.4. Fichiers Fare, VAC, ESA-EAP	15-mai-16	15-juil.-16	9.4.4
	5.3.2. Contrôles/ correction des données	**09-avr.-16**	**31-mai-16**	
74	5.3.2.1. Contrôles sur données individuelles	09-avr.-16	14-mai-16	
75	5.3.2.2. Premiers Contrôles sur les agrégats	18-avr.-16	24-avr.-16	10.3 et 10.4
76	5.3.2.3. Modification des données	18-avr.-16	24-avr.-16	9.5
77	5.3.2.4. Nouvelle exécution des chaînes de diffusion	16-mai-16	20-mai-16	
78	5.3.2.5. Derniers contrôles sur les agrégats (et corrections si nécessaires)	20-mai-16	31-mai-16	10.3 et 10.4
	5.3.3. Livraison des résultats définitifs aux DCN + DSS	**01-juin-16**	**30-juin-16**	
79	5.3.3.1. Mise à disposition des résultats définitifs N et N-1, sur le champ N-1 et méthode N	01-juin-16	15-juin-16	
80	5.3.3.2. Diffusion d'une note décrivant les données	15-juin-16	30-juin-16	
	5.3.4. Diffusion vers Eurostat	**01-juin-16**	**31-mars-17**	**9.4.5**
81	5.3.4.1. Mises en forme spécifiques	01-juin-16	15-juin-16	"
82	5.3.4.2. Contrôles spécifiques à Eurostat	15-juin-16	25-juin-16	"
83	5.3.4.3. Envoi des résultats	25-juin-16	28-juin-16	"
84	5.3.4.4. Rapports qualité pour Eurostat	01-mars-17	31-mars-17	"
	5.3.5. Publication des résultats	**01-juin-16**	**30-juil.-16**	**9.4**
85	5.3.5.1. Mise en forme spécifique des données (dont délais de paiement)	01-juin-16	15-juin-16	
86	5.3.5.2. Echanges / vérification avec le DOE	15-juin-16	30-juin-16	
87	5.3.5.3. Rédaction d'un 4 pages (à venir)	30-juin-16	30-juil.-16	
88	5.3.5.4. Mise en ligne sur Insee.fr (principales caractéristiques, données détaillées, Alisse)	30-juin-16	07-juil.-16	9.4.3
89	5.3.5.5. Mise en ligne sur Insee.fr des fiches sectorielles	30-juin-16	30-juil.-16	9.4.3.3
90	5.3.5.6. Mise sous GEN des fichiers Fare, VAC, ESA-EAP	01-juil.-16	30-juil.-16	9.4.4

Annexes

13. Annexe 1 : présentation générale du calcul de l'APE[458].

13.1 Principe général de détermination du code APE

Le code d'activité principale exercée (APE) **initial** d'une unité légale est codé dans le répertoire « Sirene » à partir de la déclaration des activités qu'est susceptible d'avoir l'unité, au dire de son créateur au moment de la création de l'entreprise dans un centre de formalités des entreprises (CFE).

Elle est ensuite éventuellement **mise à jour**,

- soit à la demande de l'entreprise si elle estime que la valeur initiale est erronée,
- soit à la suite d'une « enquête d'amélioration du répertoire » qui est un acte administratif prévu dans le décret Sirene,
- soit à la suite d'une des deux enquêtes statistiques d'Ésane (Esa ou EAP). Le code APE de l'entreprise est alors déterminé en théorie comme étant l'activité de l'entreprise **qui génère le plus de valeur ajoutée**.

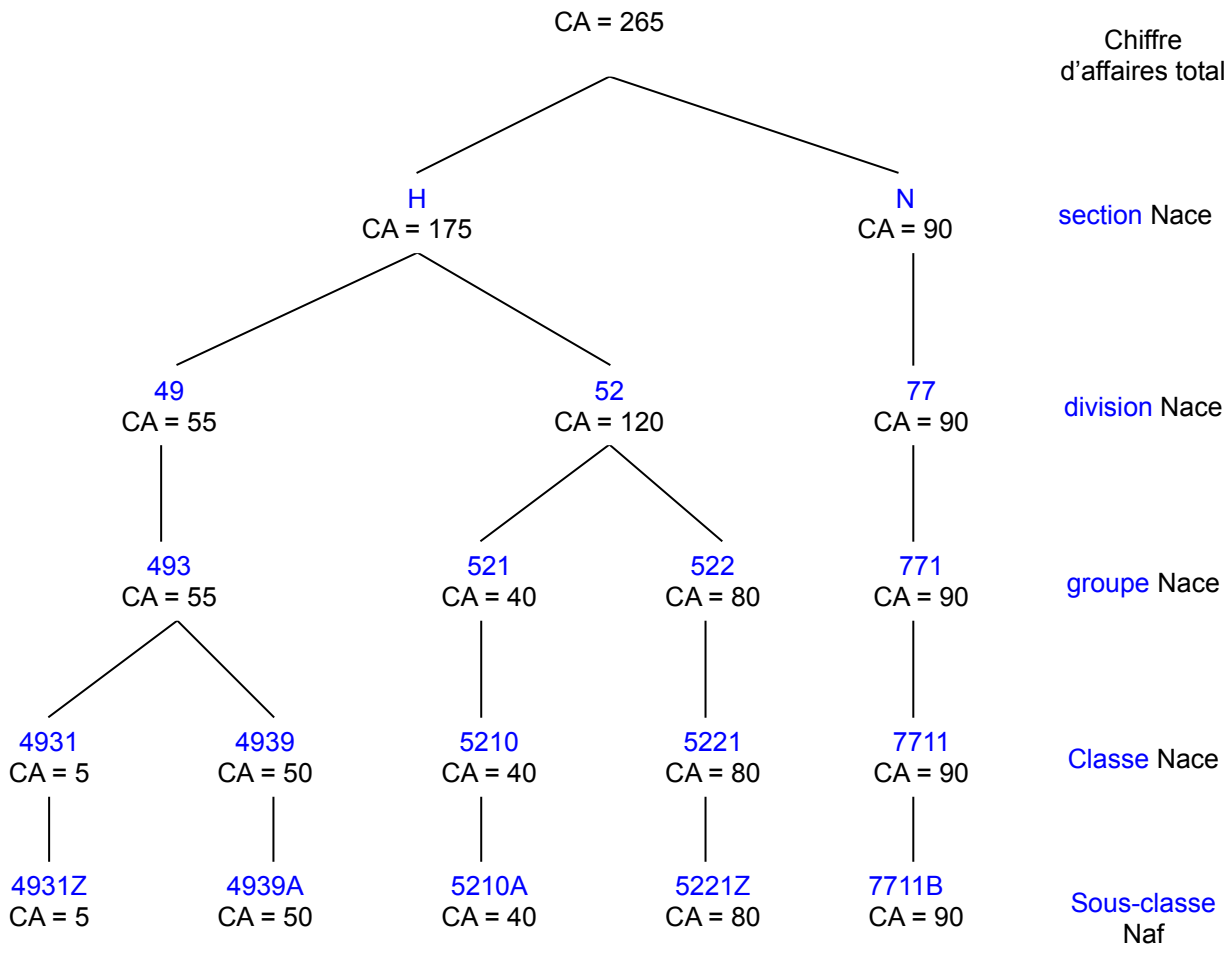
En pratique, comme on ne connaît pas la valeur ajoutée dégagée par chaque activité, **on utilise comme proxy la valeur de la production correspondante** c'est-à-dire le plus souvent **le chiffre d'affaires**, partant de ce qu'il existe une forme de proportionnalité entre production et valeur ajoutée. Toutefois, pour les activités commerciales, la valeur de la production correspond à « la marge commerciale », inconnue également, et qui doit donc être estimée à partir du chiffre d'affaires. C'est la raison principale pour laquelle les activités commerciales ne sont pas traitées comme les autres activités productives de biens ou de services (il y a d'autres raisons liées à la forme particulière de la nomenclature dans ce secteur).

On part donc de la ventilation du chiffre d'affaires en ses diverses composantes qui constituent les « branches » de l'entreprise, ventilation collectée par voie d'enquête dans l'Esa ou l'EAP. La première étape consiste à agréger le chiffre d'affaires de l'entreprise à tous les niveaux de la Naf (Nomenclature d'activité française) en partant du niveau le plus fin de la Naf (sous-classe) jusqu'à atteindre le niveau le plus agrégé (le chiffre d'affaires total de l'entreprise). Ensuite, on détermine l'APE de l'entreprise **en partant du niveau le plus agrégé** et en recherchant **à chaque nœud de la Nace**, le nœud inférieur **qui présente le chiffre d'affaires le plus élevé** (rappelons que la Naf est la déclinaison française, à 5 chiffres, de la « Nace », nomenclature d'activités économiques de la communauté européenne, qui en constitue tous les niveaux supérieurs).

Prenons un exemple : une entreprise ventile son chiffre d'affaires en cinq branches : 4931Z (5M€), 4939A (50M€), 5210A (40M€), 5221Z (80M€) et 7711B (90M€).

Dans un premier temps, on agrège le chiffre d'affaires à tous les niveaux de la nomenclature d'activité. Cela nous donne l'arborescence suivante :

[458] Cet algorithme est décrit dans l'ouvrage « Nomenclature d'Activité Française » (décret n°2007-1888 du 26 décembre 2007) des pages 17 à 23. Il est resté le même quand on est passé des EAE à Ésane.

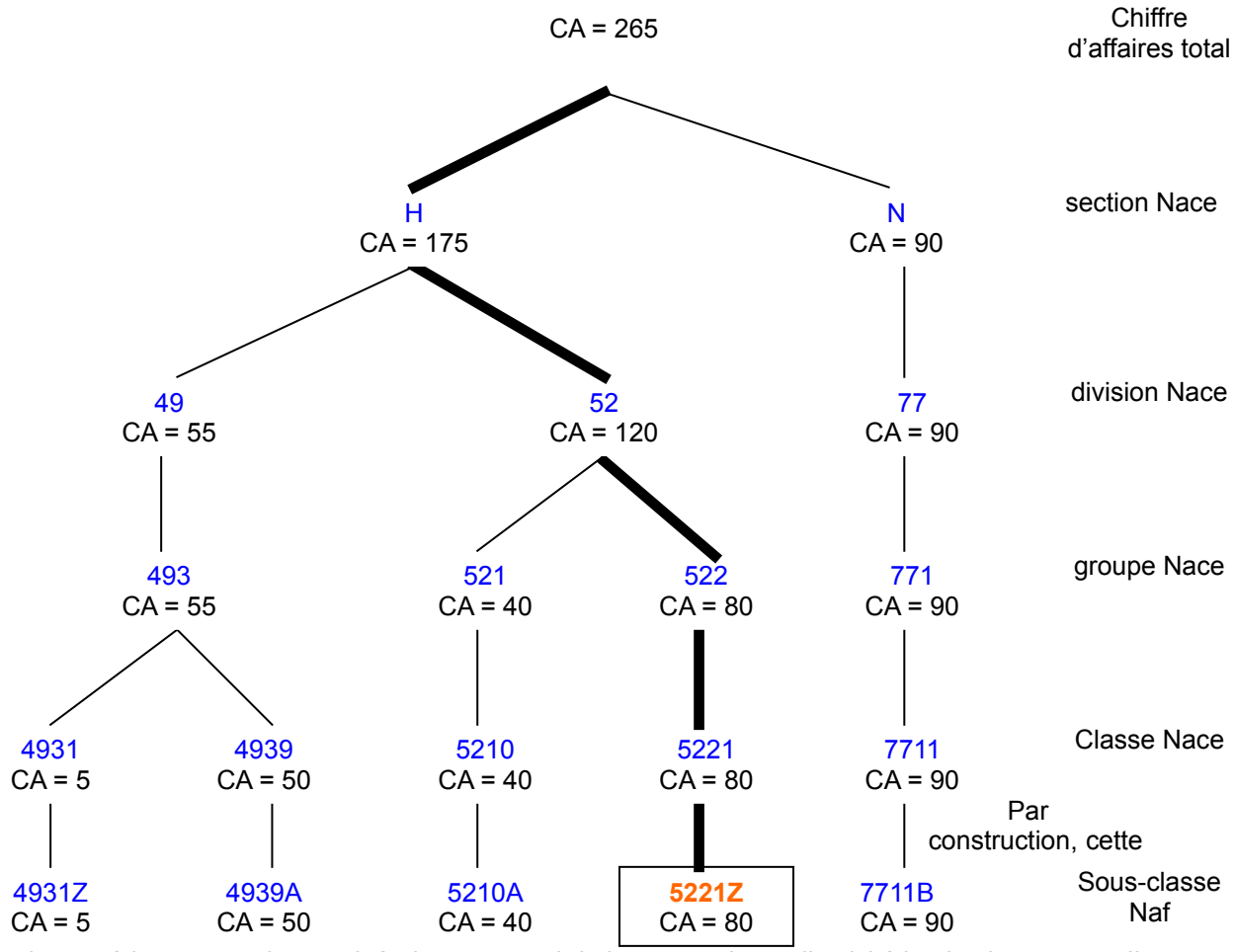


49.31Z Transports urbains et suburbains de voyageurs ;
49.39A Transports routiers réguliers de voyageurs ;
52.10A Entreposage et stockage frigorifique ;
52.21Z Services auxiliaires des transports terrestres ;
77.11B Location de longue durée de voitures et de véhicules automobiles légers

Ensuite, pour déterminer l'APE de l'entreprise, **on part du niveau le plus agrégé** et on recherche à chaque nœud, le nœud inférieur qui présente le chiffre d'affaires le plus élevé. Le chiffre d'affaires de la section H (175M€) étant plus élevé que celui de la section N (90M€), le code APE de l'entreprise devra appartenir à la section H. On itère ensuite ce raisonnement à chaque nœud, en descendant l'arbre. Au final, dans l'exemple ci-dessus, l'APE de l'entreprise est **5221Z**.

Remarque : Il faut également préciser qu'on ne tient pas compte dans cette arborescence des éventuelles activités « ***auxiliaires*** » de l'unité légale. Il existe une liste d'activités considérées comme auxiliaires[459] dès l'instant que l'unité légale dont on cherche le code APE a un chiffre d'affaires non nul dans au moins une activité ordinaire (c'est-à-dire différente de cette liste).

[459] Pour les unités légales industrielles au départ, il s'agit de 6420Z", "7010Z", "8110Z", "8411Z", "9499Z", "5210A", "5210B", "4941A","4941B", "8559A", "6203Z", "6311Z" ou toutes les classes de la division 46. Pour les unités légales non industrielles, il s'agit de "6420Z", "7010Z", "7830Z", "6820B.



façon de procéder permet de retenir à chaque nœud de la nomenclature l'activité la plus importante. Il est facile de voir **qu'elle n'aboutit pas nécessairement en revanche sur l'activité la plus importante au niveau sous-classe** : dans l'exemple ci-dessus l'activité la plus importante eût été celle de code 77.11B, qui n'est pas in fine le code APE qui sera retenu pour cette entreprise.

Un avantage de cette méthode est qu'elle assure la cohérence de la codification, quel que soit le niveau de détail de la nomenclature utilisée (si les données de branches ne sont disponibles qu'au niveau division, la division retenue pour l'APE sera celle de la sous-classe sélectionnée sur la base de données disponibles au niveau le plus détaillé de la nomenclature).

Le principe qui vient d'être exposé est **général pour toutes les activités**. Toutefois, comme on l'a mentionné plus haut, pour les entreprises du commerce, le calcul de l'APE présente des spécificités qui nécessitent des développements particuliers et des variables auxiliaires.

13.2 Les spécificités liées au commerce

Lorsque l'on détermine l'APE des entreprises du commerce, la règle consistant à rechercher à chaque nœud, le nœud inférieur qui présente le chiffre d'affaires le plus élevé n'est pas toujours appliquée. Il existe en effet deux exceptions qui se situent au niveau des divisions « commerce de gros » (46) et « commerce de détail » (47).

13.2.1 Le cas particulier du commerce de gros

Quand on se trouve au niveau de la division « commerce de gros », on procède de la manière suivante : dans un premier temps, on détermine si l'entreprise doit être classée en *intermédiaire* du commerce de gros (Nace 461) ou en *commerce de gros* (spécialisé ou non spécialisé) (Nace 462 à 469). Pour cela, on compare le chiffre d'affaires du groupe **« 461 »** au **total du chiffre d'affaires des groupes « 462 » à « 469 »** et non au chiffre d'affaires de chacun de ces groupes :

- Si l'entreprise est classée en intermédiaire du commerce de gros (461), on utilise ensuite la méthode présentée dans le paragraphe précédent pour déterminer la classe puis la sous-classe.

- Si en revanche l'entreprise est classée en commerce de gros :
    - la **distinction** entre *spécialisé* et *non spécialisé* dépend du **nombre de classes Nace de taille minimale**[460] représentées dans les ventes et non de la comparaison des chiffres d'affaires au niveau groupe.
    - Lorsque l'entreprise a été classée en commerce de gros **spécialisé ou non spécialisé,** la détermination de la classe puis de la sous-classe se fait **selon la méthode générale** décrite dans le premier paragraphe.

13.2.2 Le cas particulier du commerce de détail

La deuxième exception se trouve au niveau de la division « commerce de détail » pour laquelle on va suivre un cheminement similaire. On détermine d'abord si l'entreprise doit être classée en commerce de détail *en magasin* ou *hors magasin* :

- Si elle est classée en « **hors magasin** », la détermination du groupe, de la classe et de la sous-classe se fait **selon la méthode générale** décrite dans le premier paragraphe.
- Par contre, si l'entreprise est classée en commerce de détail **en magasin** :
    - la distinction entre *spécialisé* et *non spécialisé* dépend, comme pour le commerce de gros, du **nombre de classes Nace** représentées dans les ventes.
    - Lorsque l'entreprise a été classée en commerce de détail **spécialisé ou non** spécialisé, la détermination de la classe puis de la sous-classe se fait selon la **méthode générale** décrite dans le premier paragraphe.

13.2.3 La détermination de la ventilation du chiffre d'affaires en branches.

La ventilation du chiffre d'affaires en branches constitue la matière première pour déterminer l'APE d'une entreprise. Or dans l'Esa Commerce, le chiffre d'affaires est collecté par « produits » (dans une nomenclature fine à 7 caractères : viandes / volailles / produits surgelés / fruits et légumes frais / poissons / produits laitiers, etc.) et pour les entreprises du commerce de détail, également par « forme de vente » (vente en magasin / vente à distance / vente sur éventaires et marchés / par démarchage, etc.).

Une première étape, avant le calcul de l'APE proprement dit, va donc consister à déterminer la ventilation du chiffre d'affaires en **branches** à partir de la ventilation du chiffre d'affaires **par produit et par forme de vente :**

- Pour les entreprises du *commerce de détail*, on détermine la ventilation du chiffre d'affaires en branches à partir de la ventilation du chiffre d'affaires par forme de vente.
- Pour *toutes les autres* entreprises, on utilise la ventilation du chiffre d'affaires par produit.

De plus, **pour pouvoir comparer** des activités « hors commerce » aux chiffres d'affaires issus du commerce, il est nécessaire de prendre en compte l'équivalent de la production du commerce, autrement dit **les marges commerciales**. Ces marges, puisqu'elles ne sont pas collectées dans l'Esa, sont estimées de façon assez grossière à partir du chiffre d'affaires par produit, en appliquant un pourcentage à ce chiffre d'affaires – variable selon les produits. Le résultat de cette estimation qui est donc une estimation de la production commerciale de l'entreprise a la dimension d'un « **chiffre d'affaires pondéré** ». C'est lui qui est ensuite utilisé pour appliquer le principe général à une entreprise qui mixte des activités commerciales et productives.

D'autres spécificités du calcul **n'ont pas été détaillées ici** : par exemple :

- on ne retient **pas les ventes de carburant** dans le chiffre d'affaires du commerce de détail ;
- dans le calcul de l'APE au niveau sous-classe **intervient également la surface des magasins**, qui est une des variables collectée dans le cadre de l'Esa ; un magasin est généralement un établissement d'une entreprise ; il suffit que la surface de l'un des magasins dépasse certains seuils pour modifier le code APE de l'entreprise ;
- dans le commerce de détail non spécialisé, le code APE change selon que la part de l'alimentaire dans le chiffre d'affaires total dépasse ou pas le seuil de 35 % ;
- etc.

Malgré la simplicité du principe général, le calcul détaillé du code APE est in fine **assez complexe** à cause de la spécificité de la nomenclature des activités commerciales.

[460] Ne sont prises en compte dans le décompte que les activités qui représentent au moins 5% du chiffre d'affaires

14. Annexe 2[461] : les traitements DADS ; postes annexes et non annexes ; volume de travail ; notion d'ETP

14.1 La définition des postes et le concept de postes annexes

La Déclaration Annuelle de Données Sociales (DADS) est une formalité déclarative que doit accomplir toute unité légale employant des salariés. Tous les employeurs sont tenus, annuellement et pour chaque établissement, de communiquer grâce à cette DADS aux organismes de Sécurité Sociale d'une part, à l'administration fiscale d'autre part, la masse des traitements qu'ils ont versés, les effectifs employés et une liste nominative de leurs salariés indiquant pour chacun le montant des rémunérations salariales perçues. L'Insee est destinataire officiel de la DADS, qui lui est transmise dans le but d'élaborer des statistiques sur l'emploi et les salaires et ce, depuis 1950. L'information statistique déclarative de base est donc la période d'emploi de chaque salarié, dans un établissement donné d'une unité légale[462].

Pour assurer la production de statistiques sur l'emploi et les revenus à partir des DADS, l'Insee procède à différents traitements informatiques et manuels (appariement avec le répertoire Sirene afin d'enrichir le fichier de base, contrôle et codification, vérification de l'exhaustivité et de la cohérence annuelle, redressements divers).

Le premier traitement qui nous intéresse est la confection de **« postes** » à partir de l'information sur les périodes : **un poste correspond au cumul** des périodes d'emploi et des rémunérations **d'un salarié** dans un **même établissement**, que ces lignes correspondent à des périodes d'emploi distinctes (exemple : emploi saisonnier) ou à des périodes d'emploi identiques (exemple : une ligne prime et une ligne salaire). Le poste est donc identifié par le couple (Siret, Nir).

Dans les exploitations standard des données DADS, on cherche à ne retenir que les « vrais » emplois, ceux qui contribuent à l'appareil productif. Pour déterminer ces emplois, on utilise des seuils **en deçà desquels une période sera marquée comme annexe**[463] d'un point de vue statistique. Une période est donc considérée comme annexe si le volume de travail qui lui est associé est trop faible[464].

On examine ensuite **les postes eux-mêmes** (qui peuvent être constitués de plusieurs périodes de travail) pour déterminer s'ils sont eux-mêmes annexes.

Les seuils utilisés portent sur 3 variables observées sur chaque période de travail : la durée du poste en jours **(DUREE)**, le nombre d'heures salariées **(NBHEUR)**, et le salaire net **(S_NET**).

Dans le cas général, **une période** est considérée comme **non annexe**, **si l'une des deux** conditions suivantes est réalisée :

- DUREE > 30 et NBHEUR > 120 et NBHEUR/DUREE > 1,5
- S_NET > 3 * SMIC mensuels[465]

En d'autres termes si un salarié a reçu au cours d'une période d'emploi plus de 3 SMIC, on considère qu'il s'agit d'une période d'emploi non annexe, sans considération de durée.

[461] Les informations figurant dans cette annexe sont tirées d'un guide méthodologique des DADS, interne à l'Insee.

[462] Dans la pratique c'est un peu plus complexe que cela car certaines grandes entreprises regroupent dans leur déclaration les salariés dans un même établissement et le rôle des applications Clap et DADS est justement d'opérer un « dégroupement » de ces salariés pour les réaffecter dans l'établissement où il est le plus probable qu'ils soient. Ésane est peu concerné par cette opération de dégroupement qu'on ne décrit pas ici puisqu'on ne travaille que sur des emplois non localisés et au niveau de l'unité légale complète.

[463] Cette distinction entre postes annexes et non annexes a été introduite à la fin des années 90 dans les DADS.

[464] Ces seuils ont aussi pour objectif d'éliminer des rémunérations supplémentaires de type primes gérées par des associations (dans la fonction publique hospitalière et pour des collectivités territoriales) ou certaines périodes particulières qui ne correspondent pas à de l'emploi.

[465] Le smic mensuel moyen était de 1160 € en 2010.

Pour l'autre condition, il faut avoir travaillé à la fois plus de 30 jours, plus de 120 heures et au moins une heure et demie par jour en moyenne. Une période de 120 jours (presque 6 mois calendaires) où un salarié n'aurait travaillé qu'une heure par jour, vérifierait bien les deux premières conditions mais pas la troisième et serait donc considérée comme une période d'emploi « annexe ».

On examinera plus loin 2 cas particuliers où l'on peut définir une période non annexe en dehors du cas général.

Un **poste est non annexe** s'il vérifie **les mêmes conditions** (relatives aux variables du poste) ou s'il est constitué d'au moins une période non annexe (pour tenir compte des 2 cas particuliers).

Les deux cas particuliers sont les suivants :

1. Si le nombre d'heures est nul mais que la durée d'emploi est supérieure à 30 jours, on considère qu'il s'agit d'une erreur de déclaration et la période est aussi déclarée non annexe.

2. Si la période débute en décembre et se termine le 31/12 (embauches de fin d'année pour les achats de Noël par exemple), la durée ne peut pas être supérieure à 30. Mais comme la déclaration est annuelle, on ne sait pas ce qui se passe après le 31. Au bénéfice du doute, on qualifie de non annexe une période qui vérifie l'une des conditions suivantes (mesurées donc sur le mois de décembre) :
    - DUREE > 7,5 et NBHEUR > 30 et (toujours) NBHEUR / DUREE >1,5
    - S_NET > 0,75 SMIC mensuel

Avec les seuils choisis, en 2010, la part de postes annexes est de 27 **%, donc relativement importante** en effectifs (mais bien évidemment beaucoup plus faible en équivalents-temps plein) : il y a 43,6 millions de postes, dont 12,1 millions sont annexes. Rappelons que Ésane s'appuie sur la source Clap et que **les postes annexes ne sont pas retenus dans Clap pour déterminer les effectifs au 31/12.**
La répartition de ces postes annexes n'est pas homogène selon la profession, le secteur d'activité, l'âge ou le sexe.

Les secteurs d'activité qui connaissent le moins de postes annexes (moins de 10 %) sont des secteurs de l'industrie (hors la fabrication de denrées alimentaires, de boissons et de tabac) et de la construction. Au contraire, les secteurs connaissant une forte part de postes annexes (plus de 40 %) sont plutôt des activités de services, les activités d'arts, spectacles et activités récréatives mais aussi les activités spécialisées, scientifiques et techniques (publicité…) les activités de services administratifs et de soutien (dont intérim) et les activités d'édition, d'audiovisuel et de diffusion.

14.2 Le volume de travail et le concept d'emploi ETP

La qualification d'un poste de travail selon le taux d'activité est appelée « condition d'emploi ». Depuis 2001, la réglementation **sociale considère comme à temps complet** toute personne **effectuant la durée légale ou conventionnelle de l'entreprise**. Il s'agit d'une notion différente du taux d'activité « déclaré » sur le contrat de travail du salarié. Cette durée légale ou conventionnelle est propre à l'entreprise et donc a priori inconnue. En l'absence d'un critère contrôlable, on définit donc dans les DADS la condition d'emploi d'un salarié **de manière statistique** :

Sont considérés **comme étant à temps complet** les postes **dépassant un certain nombre d'heures par jour.**

Un seuil en nombres d'heures par jour est calculé tous les ans pour toutes les strates secteur (A38) x taille de l'établissement : ce seuil est fonction des distributions des nombres d'heures par jour des périodes déclarées en temps partiel (décroissante) et des périodes déclarées en temps complet (croissante). Les postes au-dessus de ce seuil sont déclarés « **à temps plein** » (ou temps complet) et les postes au-dessous sont déclarés « **à temps partiel** ».

La **durée d'emploi du poste** est le nombre **de jours** correspondant à une ou plusieurs périodes d'emploi. Elle est comprise entre 1 et 360. Très exceptionnellement - dans moins de 1 cas sur 1000 -

la durée peut être comprise entre 361 et 390 : cela correspond à des cas de salariés qui étaient en décalage de paie et qui ne le sont plus et qui ont travaillé une année complète.

La notion d'**année-travail** (AT) rapporte la durée d'emploi effective à la durée d'emploi maximale, 360 jours[466]. Ainsi, un salarié présent du 1er janvier au 31 décembre aura un volume de travail égal à une année travail, tandis qu'un autre salarié, présent du 1er janvier au 31 mars aura un volume de travail égal à 0,25 AT (= 3 * 30 / 360). Cette notion permet de calculer par exemple des **effectifs proratisés par la durée d'emploi.**

Mais cette approche du volume de travail a l'inconvénient de ne pas tenir compte **du taux d'activité** du salarié : en effet, que le premier salarié travaille à mi-temps ou à plein temps, il comptera pour 1. C'est pour cette raison que **l'on privilégie la notion d'année-travail sur le champ restreint aux postes à temps complet** (CPFD = 'C'). La notion suivante d'équivalent temps plein remédie à cet inconvénient.

Évaluation du volume de travail sur la base du nombre d'heures

La notion d'**équivalent temps plein** (ETP) correspond au volume de travail annuel exprimé en heures et rapporté à un horaire annuel de référence : ETP = min (nombre d'heures / référence ; 1)[467].

Pour construire la référence, on considère la taille de l'entreprise donnée par Sirene :

- pour les unités légales dont la taille est supérieure ou égale à 1 000 et hors intérim (APEN≠7820Z), **la référence est le troisième quartile** (C75) de la distribution du nombre d'heures des postes non annexes du Siren, i.e. **de l'unité légale** (en d'autres termes le seuil est défini de façon interne à l'unité légale) ;
- pour les unités légales dont la taille est inférieure à 1 000 ou du secteur de l'intérim, la **référence est donnée par le troisième quartile** (C75) de la distribution du nombre d'heures **des croisements APEN * taille**, en distinguant les entreprises de 20 salariés ou moins, et celles de plus de 20 et moins de 1 000 salariés. [Les périodes servant au calcul ont une durée de 360 jours, sont sans anomalie, ont une PCS codée et un nombre d'heures positif].

Un équivalent temps plein est ainsi calculé pour chaque poste de travail, sauf pour les postes du secteur de l'intérim (7820Z). Néanmoins, une référence est calculée afin d'estimer l'effectif des permanents de l'agence d'intérim.

Dans les données « DADS » communiquées à Ésane, la variable ETP calculée est ainsi une variable statistique continue, calculée par rapport à une référence « statistique » et différente du taux d'activité déclaré.

[466] Dans les DADS, conventionnellement, tous les mois ont une durée de 30 jours et une année fait donc 360 jours.

[467] Si le nombre d'heures d'un poste est supérieur à la référence, le Min vaut 1 et la variable ETP ne peut donc jamais prendre une valeur plus grande que 1.

15. Annexe 3 : Bibliographie[468]

[1] GENERALITES

- *[1.1] First methodological studies for the redesigning of French business statistics*, Brion Ph., UN/ECE Work Session on Statistical Data Editing, Bonn, 2006.
- *[1.2]* "Improving the quality of business registers", UN ECE, paper; A. Husseini-Skalitz, ECE/CES/GE 42/2007/3
- *[1.3] Les EAP* : Jean-Paul Lachize ; document de travail de la DSE N° E2011 / 06 d'août 2011
- *[1.4] Le rôle des estimateurs dans la production de statistiques à partir de données administratives et de données d'enquêtes : l'exemple des statistiques structurelles d'entreprises françaises*, Brion Ph., Conférence jointe SSC/SFDS, Ottawa, 2008.
- *[1.5]* "French methods and practices regarding the statistical burden", paper presented by France, DGINS 2008 Conference, Vilnius
- *[1.6] "Redesigning French structural business statistics: how can the response burden on companies be lessened",* Béguin J-M, UN ECE, paper CES/2009/35.
- *[1.7] Reengineering French structural business statistics : an overview,* Depoutot R., paper presented at the Q2010 conference, Helsinki.
- *[1.8] Reengineering French structural business statistics - an extended use of administrative data*, Chami S., paper presented at the Q2010 conference, Helsinki.
- *[1.9] Reengineering French structural business statistics : redesign of the annual survey*, Haag 2008 à 2011., paper presented at the Q2010 conference, Helsinki.
- *[1.10] Ésane, le dispositif rénové de production des statistiques structurelles d'entreprises,* Brion Ph, Courrier des statistiques n°130, 02/05/2011.
- *[1.11]* The LSE : a standardized specification language for statisticians, A. Bouichet, 2008 à 2011. Haag, S. Chami, New Techniques and Technologies for Statistics (NTTS) Conference 2011.
- *[1.12] Construction of a statistical database linked to the interadministrative business directory of the National Enterprise and Establishment Register Database: registering samples to measure the response burden,* Husseini-Skalitz A., UN/ECE Group of expert of Business register Paris, 2012.
- *[1.13] Bilan de RESANE phase 1, Beguin J-M, Bouichet A., Haag 2008 à 2011. Hentzgen J., Minodier F.,* document de travail du Secrétariat Général N° C2012/02 d'octobre 2012.

[2] SONDAGE, ESTIMATION

- *[2.1] Redesigning the French structural business statistics, using more administrative data*, Brion Ph. paper presented at the ICESIII Conference Montreal 2007.
- *[2.2] The future system of French structural business statistics: the role of the estimates*, Brion Ph., UN/ECE Work Session on Statistical Data Editing, Vienna, 2008.
- *[2.3] Le rôle des estimateurs dans la production de statistiques à partir de données administratives et de données d'enquêtes : l'exemple des statistiques structurelles d'entreprises françaises*, Brion Ph., Conférence jointe SSC/SFDS, Ottawa, 2008.
- *[2.4] L'utilisation combinée de données d'enquête et de données administratives pour la production des statistiques structurelles d'entreprises*, Brion Ph., papier présenté aux JMS de l'Insee 2009.
- *[2.5] Le plan de sondage de l'ESA (enquête sectorielle annuelle du futur dispositif Ésane)*, Bauer P., Brilhaut G., Gros E. papier présenté aux JMS de l'Insee 2009.
- *[2.6] Winsorisation dans les enquêtes annuelles auprès des entreprises franç*aises, Brion Ph., Guggemos F., 2010, « Pratiques et méthodes de sondage », Dunod, actes du 6ème colloque francophone sur les sondages, Tanger.
- *[2.7] Diffuser des taux d'évolution entre l'année N et N-1, c'est bien, décomposer cette évolution en différents facteurs explicatifs, c'est mieux !,* Koumarianos H., papier présenté aux JMS de l'Insee 2012.

[468] La bibliographie intégrale a été conservée mais les renvois aux documents purement internes à l'Insee ont été supprimés au sein de l'ouvrage.

- *[2.8] Le kit de diffusion Ésane ou comment diffuser des agrégats à des niveaux fins tout En alertant l'utilisateur sur la pertinence des chiffres publiés*, Seng J., papier présenté aux JMS de l'Insee 2012.
- *[2.9] Ésane ou les malheurs de l'estimation composite : comment gérer les valeurs négatives d'estimateurs par différence ?*, Gros E. papier présenté aux JMS de l'Insee 2012.
- *[2.10] The ÉSANE estimation kit or how to get more detailed results with quality indicators,* Seng J.*,* paper presented at the Q2012 conference, Athens.
- *[2.11] A first assessment of the French system of production of structural business statistics combining administrative data and survey data*, Brion Ph., paper presented in the ICES IV 2012, Montreal 2012.
- *[2.12] Les estimateurs Ésane pour les nuls*, Koumarianos H., document de travail DSE n°E2013/01.
- *[2.13*] *Missing data treatment in administrative fiscal sources for the French Structural Business Statistics production system*, Deroyon Thomas, a paper presented in EESW13, Nuremberg, sept 2013.
- *[2.14] La coordination d'échantillons d'enquêtes auprès des entreprises mise en place à l'Insee, Guggemos F., Sautory O.,* papier présenté aux JMS de l'Insee 2012.
- *[2.15] The unit problem: A first assessment of the impact of profiling on sampling,* Ronan Le Gleut, Anaïs Levieil, Élodie Martal, Thomas Merly-Alpa, paper presented in the ICES V, 2016, Geneva 2016.
- *[2.16] Statistical estimators using jointly administrative and survey data to produce French structural business statistics,* Brion P. and Gros E., Journal of Official Statistics (JOS), Vol. 31, No. 4, 2015, pp. 589–609.

[3] CONTROLE DES DONNEES

- *[3.1] Vérification sélective des données pour le futur système français de statistiques structurelles d'entreprises*, Brihault, G., Brion, Ph., 2008, « Méthodes de sondage », Dunod, actes du 5ème colloque francophone sur les sondages, Marseille.
- *[3.2] The implementation of the new system of French structural business statistics, Brion Ph. UN/ECE Work Session on Statistical Data Editing, Neuchâtel, 2009.*
- *[3.3] Setting cut off scores for selective editing in structural business statistics: an automatic procedure using simulation study*, Gros E., UN/ECE Work Session on Statistical Data Editing, Neuchâtel, 2009
- *[3.4] Une procédure de détermination automatique des seuils de macro contrôles par simulation sur données passées*, E. Gros, 2010, « Pratiques et méthodes de sondage », Dunod, actes du 6ème colloque francophone sur les sondages, Tanger".
- *[3.5] Quality improvement of individual data and statistical outputs based on combined use of administrative and survey data*, Gros E., UN/ECE Work Session on Statistical Data Editing, Ljubljana, 2011.
- *[3.6] First elements relative to the data editing strategy used for the new system of French structural business statistics*, Brion Ph.*,* UN/ECE Work Session on Statistical Data Editing, Ljubljana, 2011.
- *[3.7] Ésane : À la recherche d'une cohérence maximale des données multi-sources sur les entreprises par le biais de micro et macro contrôles*, Haag 2008 à 2011., papier présenté aux JMS de l'Insee 2012.
- *[3.8] Assessment and improvement of the selective editing process in ÉSANE (French SBS),* Gros E., UN/ECE Work Session on Statistical Data Editing, Oslo, 2012.
- *[3.9] Methodological questions raised by the combined use of administrative and survey data for the French structural business statistics*, Brion Ph, UN/ECE Work Session on Statistical Data Editing, Oslo, 2012.
- *[3.10] First assessment of the combined use of administrative and survey data in the new system of French structural business statistics,* Gros, E., a paper presented in the ICESIV 2012 Montreal 2012.
- *[3.11] Le dispositif Ésane, ou comment l'utilisation combinée de données administratives et de données d'enquête permet d'améliorer la qualité des données individuelles et des statistique,* Gros E., *7e colloque francophone de sondage 2012.*

16. Annexe 4 : Glossaire

A

ACOSS	Agence Centrale des Organismes de Sécurité Sociale (assure la tutelle des Urssaf)
ALISSE	Accès en Ligne aux Statistiques Structurelles d'Entreprises (Base de données de diffusion accessible sur le site de l'Insee)
APE	Activité Principale Exercée (Code attribué à toutes les unités légales par l'Insee dans Sirene)

B

BALO	Bulletin des annonces légales obligatoires
BCE	Base de création d'entreprises (Base de données en liaison avec le système Ésane)
BODACC	Bulletin officiel des annonces civiles et commerciales
BRC	Bordereau de recouvrement des cotisations (pour l'Urssaf)

C

CA	Chiffre d'affaires
CA3	Déclaration mensuelle ou trimestrielle de chiffre d'affaires aux services fiscaux
CES	Centre d'Enquêtes Statistiques du Sessi à Caen. Au 1er janvier 2011, le CES est devenu SSNE : Service de Statistiques Nationales d'Entreprises
CFE	Centre de formalité d'entreprises : c'est dans un de ces centres que les créateurs d'entreprises doivent enregistrer l'unité légale qu'ils créent
Citrus	Coordination des Informations et des Traitements sur les Restructurations d'Unités Statistiques : il s'agit d'une base de données interne à l'Insee qui permet d'enregistrer toutes les « restructurations » d'unités légales (scission, absorption, transfert d'actifs)
CLAP	Connaissance locale de l'appareil productif (système d'information interne à l'Insee)
CNAM	Caisse nationale d'assurance maladie
CNAV	Caisse nationale d'assurance vieillesse
CNIN	Centre national Informatique de Nantes (de l'Insee)
CNIS	Conseil national de l'information statistique
CNR	Constat de non-réponse (phase de la procédure de contentieux)

D

DADS	Déclaration annuelle de données sociales
DAP	Département applications et projets (de l'Insee)
DARES	Direction de l'animation de la recherche, des études et des statistiques au ministère du travail, de l'emploi et de la formation professionnelle.
DCN	Département des comptes nationaux (de l'Insee)
DERA	Département de l'emploi et des revenus d'activité (de l'Insee)
DESE	Direction des études et synthèses économiques (de l'Insee)
DGCIS	Direction générale de la compétitivité, de l'industrie et des services (Ministère de l'économie et des finances)
DGDDI	Direction générale des douanes et des droits indirects
DGFiP	Direction générale des finances publiques (ministère de l'économie et des finances)
DGI	Direction générale des impôts (ministère de l'économie et des finances)
DILA	Direction de l'information légale et administrative (rattachée au Premier Ministre ; elle publie entre autres le Journal Officiel
DMS	Département des méthodes statistiques (de l'Insee)
DMMO	Déclaration mensuelle de mouvements de main d'œuvre
DOE	Département de l'offre éditoriale (de l'Insee)
DSE	Direction des statistiques d'entreprises (de l'Insee)
DRISS	Département répertoires infrastructures et statistiques structurelles (de l'Insee)
DSCT	Département statistiques de court terme (de l'Insee)

DSDS	Direction des statistiques démographiques et sociales (Insee DG)
DSN	Déclaration sociale nominative
DSS	Département des synthèses sectorielles (de l'Insee)

E

EAE	Enquête annuelle d'entreprise (ancienne enquête précédent le système Ésane)
EAP	Enquête annuelle de production industrielle (fait partie du dispositif Ésane)
EAR	Enquête Amélioration du Répertoire (SIRENE)
EBE	Excédent brut d'exploitation : un des postes comptables du PCG
EPURE	Extension du projet Urssaf sur les revenus et l'emploi
EPS	Équipe de Projet Statistique
ÉSANE	Élaboration statistique structurelle d'entreprise
ETP	Équivalent temps plein
Eurostat	Nom de la direction de la Commission européenne en charge de la coordination des statistiques européennes

F

FUL	Fichier unique de lancement d'enquêtes industrie

G

GEN	Global Enterprise (un groupe mondialisé est découpé en "GEN")

I

IAA	Industries Agricoles et Alimentaires
IEG	Information économiques générales (un des sous-systèmes de Ésane)
IHM	Interface homme machine (désigne la succession d'écrans que voit un utilisateur dans une application donnée)

L

LIFI	Enquête (Insee) sur les liaisons financières entre sociétés
LME	Loi de modernisation de l'économie, entrée en vigueur en 2009

M

MOA	Maîtrise d'ouvrage
MOE	Maîtrise d'œuvre
MSA	Mutualité sociale agricole

N

NACE	Nomenclature d'activités de la communauté européenne
NAF	Nomenclature d'Activités française. Il en existe 3 versions : Révision 1 (1993) ; Révision 1.1 (2002) et révision 2 (2008). Les quatre premiers caractères de la NAF sont identiques à ceux de la NACE

O

Ocsane	Outil de coordination des statistiques annuelles d'entreprises (répertoire interne de Ésane)

P

PCG	Plan comptable général (standard français pour l'élaboration des comptes d'entreprises)

PIB	Produit Intérieur brut : calculé à l'Insee par les comptables nationaux
PRODCOM	« Production communautaire » : règlement européen qui détermine annuellement une liste de produits dont la production annuelle totale doit être fournie à Eurostat
Profileur	Agent de la division PTGU chargé du profilage en face à face des grands groupes.
PSCD	Production de statistiques pour les comptes détaillées (sous-processus de Ésane)
PSCP	Production de statistiques pour les comptes particuliers (sous-processus de Ésane)
PTGU	Division « profilage et traitement des grandes unités » (au sein de la DSE)

R

RCS	Registre du commerce et des sociétés (tenu par les greffes des tribunaux de commerce)
REDI	Réconciliation des données individuelles (sous-processus de Ésane)
RESANE	Refonte des Statistiques Annuelles d'Entreprises : projet ayant donné lieu à la construction du système d'information Ésane
RSP	Responsables Secteur produits (agents de l'Insee à la direction générale en charge de certaines étapes de la comptabilité nationale

S

SBS	Structural Business Statistics (Règlement européen sur les statistiques structurelles N° 58/97)
SCI	Société civile immobilière
SCM	Société civile de moyens
SOeS	Service de l'Observation et des Statistiques (ministère de l'environnement)
SESSI	Service des statistiques industrielles, ministère de l'industrie (avant sa fusion avec l'Insee)
SIASP	Système d'information sur les agents des services publics (concerne l'emploi et les rémunérations)
SIE	Système intermédiaire d'entreprise (système d'information interne à l'Insee qui précédait Ésane
SIRENE	Système informatisé du répertoire des entreprises et des établissements (répertoire des unités légales géré par l'Insee avec l'aide des CFE)
SIRUS	Système d'identification au répertoire des unités statistiques (nom du répertoire statistique développé par l'Insee dans le cadre d'Ésane)
SSE	Système Statistique d'Entreprise
SSM	Services statistiques des ministères
SSNE	Service des statistiques nationales d'entreprise (Caen)
SSP	Service de la statistique et de la prospective (ministère de l'agriculture et de la pêche)
STS	Règlement européen sur les statistiques de court terme
SUSE	Système unifié des statistiques d'entreprises (système d'information de l'Insee permettant d'enregistrer les déclarations fiscales avant Ésane)

T

TEN	Truncated enterprise (il s'agit de la trace, sur un pays donné, d'une GEN (voir ce mot)

U

URSSAF	Unions de recouvrement des cotisations de sécurité sociale et d'allocations familiales
UNEDIC	Union Nationale pour l'Emploi Dans l'Industrie, le Commerce et l'agriculture

V

VA	Valeur ajoutée.