

BIEN COMPRENDRE LA DÉCLARATION SOCIALE NOMINATIVE POUR MIEUX MESURER

Catherine Renne*

L'Insee établit des statistiques sur l'emploi et les salaires depuis plus d'un demi-siècle à partir de différentes déclarations administratives. La mise en place de la déclaration sociale nominative n'a pas simplement consisté à fusionner les déclarations administratives existantes. Elle a également conduit à rationaliser et simplifier les données d'une part, et à définir rigoureusement le cadre dans lequel s'effectuent les échanges d'informations entre les entreprises et les organismes destinataires de la DSN, d'autre part. Elle a obligé l'institut à mener une refonte de grande ampleur de son système d'information en même temps qu'elle est venue questionner l'organisation des travaux. Les contrôles réalisés en amont, à l'occasion de la collecte des données, assurent en contrepartie de recevoir les déclarations avec un certain degré de qualité et donc avec une certaine valeur ajoutée. Pour tirer le meilleur parti de la DSN, le statisticien doit mieux connaître l'étape de collecte et de contrôle et s'impliquer davantage dans les nouvelles instances de gouvernance de la DSN. Même si la priorité à court terme a été d'intégrer la DSN dans son système d'information, la prochaine étape va consister pour l'Insee à construire un système suffisamment souple et modulaire qui permettra d'exploiter toutes les potentialités de cette nouvelle source d'information.

📍 LA DSN, UNE SOURCE INCONTOURNABLE POUR LE STATISTICIEN

Déclaration unifiée de cotisations sociales, déclaration mensuelle de mouvements de main-d'œuvre, relevé mensuel de missions, déclaration annuelle de données sociales... Toutes ces déclarations administratives auxquelles étaient soumis encore récemment la plupart des employeurs sont désormais remplacées par la déclaration sociale nominative (DSN) (voir l'article d'E. Humbert-Bottin dans le même numéro).

Mise en place progressivement depuis 2013, la DSN est bien plus qu'une nouvelle déclaration administrative, elle institue une nouvelle logique déclarative entre les entreprises et les organismes destinataires (Buhl, 2014). De fait, la DSN n'a pas vocation à servir un besoin spécifique, mais au contraire à couvrir différents usages.

Avant la DSN, les employeurs transmettaient des données parfois redondantes à une diversité d'acteurs avec des échéances et des formats différents. Désormais, ils transmettent une seule fois chaque mois les informations issues de leur logiciel de paie à un point

* Cheffe-adjointe du département de l'emploi et des revenus d'activité, Insee

de dépôt unique. Ces informations sont contrôlées puis redistribuées aux différents organismes destinataires, charge à eux de les transformer pour les mettre au format qui leur convient. D'une certaine façon, la DSN transfère une partie de la charge des entreprises vers les systèmes d'information en aval.

Pour l'Insee, la généralisation de la DSN se traduit progressivement par le remplacement des principales sources qui alimentent son système d'information sur l'emploi et les revenus d'activité (Siera). Dit autrement, la DSN devient « LA » source d'information majeure et incontournable et il est essentiel que les statisticiens en comprennent et maîtrisent les différents aspects de son traitement.

Lorsqu'on parle du traitement d'une déclaration administrative, on pense en général aux traitements statistiques qui sont mis en œuvre, notamment à l'Insee. Mais la phase amont de collecte et de contrôle de l'information qui, elle, relève largement d'acteurs extérieurs est tout aussi importante. Elle est souvent méconnue des utilisateurs finaux alors que c'est précisément cette étape qui leur assure de recevoir les déclarations avec un certain degré de qualité et donc avec une certaine valeur ajoutée.

Cette problématique n'est pas spécifique à la DSN et est largement généralisable à toute autre déclaration administrative. L'objet de cet article est de décrire, à travers l'exemple illustratif de la DSN, les étapes de gestion de la qualité d'un processus de traitement d'une déclaration administrative.

RATIONALISER ET SIMPLIFIER : DEUX OBJECTIFS AFFICHÉS DE LA DSN

Avec la DSN, les entreprises transmettent trois fois moins de données que lorsqu'elles remplissaient leur déclaration annuelle de données sociales. En adoptant le principe du « Dites-le nous une fois », la DSN limite en effet les données demandées aux seules informations élémentaires qui ne peuvent être obtenues par un autre moyen. Par exemple, depuis janvier 2018, les entreprises n'ont plus obligation de fournir leur effectif salarié de fin de période, celui-ci pouvant être recalculé directement par les organismes destinataires à partir des informations individuelles transmises sur les salariés.

La rationalisation des données a permis également la mutualisation de l'information collectée pour un grand nombre d'organismes. D'une part, la DSN ne contient plus de rubriques spécifiques à tel ou tel organisme comme c'était le cas dans la déclaration annuelle de données sociales. D'autre part, le fait de rassembler toutes les informations dans une déclaration unique garantit la cohérence de l'information destinée à des usages différents. Par exemple, avec la DSN, on est sûr que les employeurs et les salariés sont identifiés de la même façon quel que soit l'organisme destinataire de l'information.

Cet objectif de rationalisation n'aurait pas pu être atteint sans une définition rigoureuse du cadre dans lequel s'effectuent les échanges d'informations entre les entreprises et les organismes destinataires de la DSN. Ce cadre appelé également norme d'échanges¹ définit le modèle d'information, le modèle de message et la cinématique des échanges. Leur description figure dans un document appelé « cahier technique » qui est mis à jour chaque année et versionné. Le modèle d'information contient la définition des différentes rubriques à renseigner, précise leur sémantique, c'est-à-dire comment elles sont organisées et à quoi elles sont rattachées (l'entreprise, le contrat, le salarié ?...) et vérifie que les données ne sont pas demandées plusieurs fois. Par ailleurs, lorsque les employeurs transmettent

1. Dans le cas de la DSN, il s'agit de la norme NEODeS : Norme d'Échange Optimisée des Données Sociales.

Encadré 1. La cinématique des échanges de la DSN



leur DSN, ils n'envoient pas les informations « en vrac » mais selon un modèle de message qui place les informations dans un certain ordre et selon une certaine structure.

La cinématique des échanges décrit, quant à elle, le processus d'échange des données entre les différents organismes partenaires. De façon très schématique, le processus de collecte de la DSN s'appuie sur deux portails de déclaration : [le portail net-entreprise.fr](http://leportailnet-entreprise.fr) pour l'ensemble des entreprises et le portail msa.fr uniquement destiné aux employeurs du régime agricole. Les informations transmises dans les DSN mensuelles sont d'abord réceptionnées et contrôlées par l'Accoss² qui les transmet à la Caisse nationale d'assurance vieillesse (Cnav) d'une part, et les filtre pour distribuer un extrait du message aux organismes d'assurance, de prévoyance et de mutuelle, d'autre part. La Cnav certifie l'identification des salariés (Nir) grâce à son système national de gestion des identifiants (SNGI), stocke les DSN pour ensuite les redistribuer de façon dématérialisée aux organismes de protection sociale et aux organismes publics, dont l'Insee (**encadré 1**).

2. Agence centrale des organismes de sécurité sociale.

📌 S'IMPLIQUER DANS LA GOUVERNANCE DE LA QUALITÉ DE LA COLLECTE

L'utilisation de déclarations ou de sources administratives à des fins statistiques est préconisée par le code de bonnes pratiques de la statistique européenne dans le but d'alléger la charge statistique des déclarants. En France, ceci est rendu possible par la loi n° 51-711 de 1951 relative à l'obligation, à la coordination et au secret en matière de statistique et a été réaffirmé récemment par la loi n° 2016-1321 de 2016 pour une République numérique.

Si la loi facilite l'accès de l'Insee aux déclarations administratives, l'obtention des données se fait en général après un long temps d'échanges, de préparation et de formalisation avec l'organisme concentrateur. Cette phase préliminaire du processus ne doit pas être sous-estimée, car elle conditionne les étapes ultérieures du processus et l'utilisation que l'on pourra faire des données. Elle doit l'être d'autant moins que les usages statistiques de la DSN ne sont pas de même nature que les usages administratifs en fonction desquels le nouveau système a été conçu en priorité.

S'agissant de la DSN, l'Insee a conclu une convention tripartite en 2016 avec la Cnav et le Gip-MDS³, qui précise les modalités de livraison (date de livraison, contenu et format des fichiers, etc.). Un contrat de service a également été mis en place avec la Cnav dont le contenu précise de façon plus technique les modalités d'échanges des fichiers. Enfin, les traitements envisagés par l'Insee et les modalités de conservation des données ont fait l'objet d'une déclaration auprès de la Cnil.

Mais la bonne utilisation de la DSN nécessite également un suivi attentif des évolutions du cahier technique. Celui-ci est susceptible d'évoluer chaque année pour intégrer notamment de nouvelles rubriques. C'est le cas de la version 2019 du cahier technique qui intégrera de nouvelles rubriques spécifiques aux salariés de la fonction publique ou de la version 2020 qui intégrera quant à elle la Déclaration Obligatoire d'Emploi des Travailleurs Handicapés (DOETH). L'Insee doit alors signaler au Gip-MDS s'il souhaite ou pas recevoir ces nouvelles rubriques.

L'Insee est en effet régulièrement informé des modifications des rubriques de la norme via les instances de gouvernance, notamment le comité de normalisation des données sociales dont il est membre. Il a également des échanges réguliers avec le Gip-MDS qui assure la maîtrise d'ouvrage opérationnelle du projet DSN. L'institut peut, au même titre que les autres organismes, faire remonter les anomalies ou erreurs qu'il pourrait observer dans les données afin d'améliorer leur qualité. D'une certaine façon, le système DSN fonctionne comme une boucle vertueuse, dans laquelle l'institut doit prendre progressivement toute sa place.

📌 UNE COLLECTE TRÈS CONTRÔLÉE

L'utilisation de déclarations administratives n'est pas toujours aisée pour le statisticien, car ce dernier n'est pas directement à l'origine des concepts sous-jacents ou de la collecte. Il se doit pourtant de comprendre les différentes étapes de la fabrication des données qu'il reçoit afin de juger de leur qualité, d'une part, et d'éviter de reproduire inutilement des traitements qui auraient déjà été réalisés en amont, d'autre part.

Lorsque l'Insee réceptionne chaque mois les déclarations sociales nominatives, celles-ci ont déjà fait l'objet d'une longue série de vérifications automatiques destinées à sécuriser le processus déclaratif. Cette notion de contrôle est à ce point essentielle dans le dispo-

3. Groupement d'intérêt public « Modernisation des déclarations sociales ».

sitif de collecte que les entreprises peuvent tester, à l'aide d'un outil d'auto-contrôle mis à leur disposition, le fichier contenant leur déclaration avant de le transmettre. Cet outil est cohérent avec celui de la plateforme de déclaration qui met en œuvre par la suite divers types de contrôles.

Des contrôles de structure et de syntaxe permettent de vérifier la conformité d'ensemble des fichiers envoyés. Il s'agit par exemple de vérifier que les variables déclarées respectent bien une certaine longueur et un certain format. Des contrôles de cohérence vérifient, quant à eux, la cohérence entre différentes rubriques. Par exemple, dans les éléments d'adresse, le renseignement du libellé d'une localité oblige à renseigner le code postal associé. Des contrôles sur des référentiels externes vérifient que la valeur prise par certaines données déclarées appartient bien à une nomenclature (catégorie socio-professionnelle, activité de l'établissement, code postal, etc.). Enfin, la validité de tous les numéros SIRET déclarés dans la DSN font l'objet d'une vérification, notamment le SIRET de l'établissement utilisateur et celui du lieu de travail lorsque celui-ci a été déclaré comme un établissement. Le SIRET des déclarants est, quant à lui, contrôlé dès le portail d'identification.

Tous ces contrôles sont bloquants, c'est-à-dire qu'ils entraînent le rejet complet de la déclaration tant que les erreurs ne sont pas corrigées. Ils représentent l'essentiel des vérifications automatiques faites par le processus de collecte de la DSN. Il existe par ailleurs des contrôles non bloquants qui ne sont présents qu'à titre d'information. Il n'était en effet pas envisageable de rendre bloquants tous les contrôles effectués sur les DSN au risque d'empêcher leur transmission. Les contrôles effectués sur la DSN sont uniquement des contrôles de flux ; ils ne permettent pas de vérifier la cohérence des informations avec une autre source ni celle des données transmises d'un mois ou d'une année sur l'autre. Tous ces contrôles sont documentés et décrits précisément dans le cahier technique de la norme d'échanges de la DSN et ont fait l'objet d'une réflexion concertée avec les différents organismes destinataires.

TRANSFORMER LES DONNÉES POUR RÉPONDRE AUX BESOINS DE LA STATISTIQUE

Pour le statisticien, disposer d'une source d'information documentée et contrôlée en amont, comme la déclaration sociale nominative, est plus que précieux. Cependant, son travail ne consiste pas à reproduire simplement l'information qu'il a reçue. Avant de la diffuser, il doit encore procéder à des contrôles et autres traitements afin de transformer cette matière brute en une information statistique pertinente pour l'analyse économique.

Le contrôle d'exhaustivité, souvent appelé de façon abusive « repérage des trous de collecte » fait partie des étapes essentielles du traitement statistique. S'agissant de la DSN, il ne s'agit pas de se substituer au Gip-MDS qui contrôle déjà l'exhaustivité de la collecte des déclarations en relançant les entreprises retardataires. À ce stade, il s'agit de vérifier que l'Insee a bien reçu toutes les déclarations qu'il était supposé recevoir, d'une part et qu'il a bien intégré dans son système d'information toute l'information qu'il a reçue, d'autre part. Toute la difficulté est de définir la notion de « déclarations attendues ».

Une première façon de procéder est d'effectuer des contrôles de vraisemblance « macro » destinés à repérer une éventuelle évolution anormale du nombre de déclarations reçues. En général ce type de contrôles se fait au moment de la réception des données, ce qui permet de réagir rapidement en cas de détection de déclarations manquantes conformément au contrat de service établi avec l'organisme émetteur de la déclaration. Le statisticien peut également effectuer des contrôles plus « micro » en confrontant les déclarations reçues au cours d'une période donnée avec celles de la période précédente ou à un référentiel externe. Avant la mise en place de la DSN, ce référentiel externe était construit

à partir des différentes déclarations administratives reçues à l'Insee. Le remplacement des différentes déclarations par une déclaration unique, la DSN rend progressivement difficile la construction d'un tel référentiel. Il serait en effet paradoxal de construire un référentiel externe à partir des données reçues en DSN alors que c'est précisément l'exhaustivité de cette information que l'on cherche à contrôler. Les contrôles de cohérence intertemporelle seront donc de plus en plus privilégiés au fur et à mesure de la montée en charge de la DSN.

Ensuite, le statisticien va s'assurer de la vraisemblance des variables qu'il juge structurantes pour son système d'information. Dans cette phase du processus de traitement de la déclaration administrative, la détection d'éventuelles anomalies constitue un enjeu important. Plusieurs questions se posent alors : comment distinguer, parmi ces anomalies, celles qui s'avèrent être des erreurs et doivent donc être corrigées ? Dit autrement, si on trouve dans les déclarations sociales des salaires horaires inférieurs au Smic ou à l'opposé des salaires horaires extrêmement élevés, s'agit-il d'une erreur de déclaration ou le reflet de la réalité ? Ces erreurs peuvent-elles être traitées de façon automatique ou nécessitent-elles un traitement manuel ? Toutes les erreurs

Encadré 2. Le contrat de travail : une notion centrale en DSN

Les caractéristiques du contrat :

- libellé d'emploi
- catégorie socio-professionnelle
- nature du contrat (CDD, CDI, intérim...)
- motif de recours (pour les CDD et les missions)
- dispositif de politique publique (contrats aidés)
- établissement utilisateur (pour les contrats d'intérim)
- lieu de travail
- multi-activité du salarié.

Les revenus rattachés au contrat :

- durée de la période d'activité
- éléments de rémunération
- primes, gratifications et indemnités.

La vie du contrat :

- date de début
- date prévisionnelle de fin de contrat (CDD)
- date de fin
- motif de rupture
- arrêts de travail (motif, durée).

Les cotisations sociales et les taxes :

- bases assujetties
- cotisations individuelles
- exonérations
- assujettissement fiscal (taxe d'apprentissage, taxe sur les salaires...).

« Comment distinguer, parmi ces anomalies, celles qui s'avèrent être des erreurs et doivent donc être corrigées ? »

doivent-elles être corrigées et si non, sur quels critères définit-on les priorités de traitement et à quel moment doit-on s'arrêter ? Du fait de la multiplicité des usages parfois antagonistes, il est souvent difficile de fixer un critère d'arrêt unique. On sera en effet plus ou moins exigeant sur le critère d'arrêt selon qu'on souhaitera publier

des données individuelles ou des estimations agrégées pour lesquelles on peut tout à fait accepter qu'elles reposent sur des données individuelles imparfaites. Le contrôle de la vraisemblance des données se poursuit donc en aval de la chaîne de traitement de la déclaration administrative. Dans le domaine de l'emploi et des salaires, le processus de validation des données produites à partir des déclarations administratives est un processus long et itératif qui implique, outre l'Insee, de nombreux services statistiques ministériels (Dares, DGAFP, Drees, DGCL, OED...).

Mais la principale difficulté pour le statisticien lorsqu'il travaille à partir d'une déclaration administrative, c'est que celle-ci n'a pas été conçue pour son usage exclusif. La déclaration peut donc contenir des informations dont il n'a pas besoin et surtout ne pas reposer sur ses unités statistiques de référence. Dans le domaine de l'emploi et des revenus d'activité par exemple, on va s'intéresser à des notions telles que le poste de travail ou les périodes d'activité. La DSN véhicule quant à elle les informations relatives à des contrats de travail et des périodes de rémunération, d'une part et à des signalements d'événements comme les arrêts de travail, d'autre part (**encadré 2**).

UN CHANGEMENT D'ÉCHELLE ET DE TEMPORALITÉ DE LA PRODUCTION STATISTIQUE

Avec la mise en place de la DSN, c'est bien plus que le remplacement progressif des déclarations administratives qui alimentaient le système d'information de l'Insee qui est à l'œuvre. L'institut doit désormais se donner les moyens de traiter chaque mois un volume d'information équivalent à celui d'une déclaration annuelle de données sociales. La volumétrie des données devient un enjeu important ; il s'agit en effet de stocker et traiter chaque mois un peu moins d'une centaine de gigaoctets de données, soit plus d'un téraoctet par an. Vu sous cet angle, la DSN aurait pu être considérée uniquement par l'Insee comme une contrainte technique et traitée comme telle, ce qui aurait été une erreur.

Sans nul doute, la DSN vient bouleverser les systèmes d'information construits depuis longtemps autour des différentes déclarations administratives. Pour l'Insee, il y a là une formidable opportunité de repenser et documenter ses processus de production afin de gagner en efficacité et en cohérence. Même si la priorité à court terme a été d'intégrer la DSN, il faut dès à présent anticiper la demande qui sera adressée à l'Insee à moyen terme, dans le domaine de l'emploi et des revenus d'activité, afin de construire un système suffisamment souple et modulaire qui permettra d'y répondre.

Le fait de disposer de données mensuelles permet de traiter désormais les données au fil de l'eau, ce qui devrait à terme réduire les délais de production. Devoir traiter chaque mois environ 2 millions de déclarations contenant les déclarations sociales nominatives de 20 millions de salariés oblige à se montrer parcimonieux dans les traitements manuels confiés aux gestionnaires. Les nouveaux processus mis en place se fondent donc largement sur un principe de capitalisation afin d'éviter de reproduire chaque mois les mêmes traitements. Par exemple, si un gestionnaire traite une anomalie présente dans la catégorie socio-professionnelle d'un salarié un mois donné, et que l'anomalie subsiste dans la déclaration du mois suivant, le système reproduira de façon automatique le traitement effectué

par le gestionnaire le mois précédent. Les travaux menés sur la qualité de la localisation des emplois, dits de « dégroupement », seront désormais conduits au sein d'un seul processus au sein du Siera. Ils étaient pris en charge jusqu'à présent par différents processus au sein et à l'extérieur du Siera. Enfin, la ré-ingénierie des processus se fera avec la volonté d'améliorer la cohérence entre données macro et données micro, d'une part et entre statistiques en évolution et en niveau, d'autre part.

PIRÉNÉS⁴ : UN PROJET DE RÉINGÉNIERIE AMBITIEUX

L'Insee a décidé de lancer en 2012 un ensemble de projets de ré-ingénierie des processus du Siera, dont le premier, le projet Pirénés, a pour objectif d'accueillir et de traiter la déclaration sociale nominative. Le processus de production a été organisé autour d'applications remplissant chacune un rôle bien précis.

Il s'agit tout d'abord d'être en mesure d'accueillir chaque mois environ 2 millions de fichiers. La première étape pour l'Insee consiste à les intégrer selon son propre modèle de données en ayant au préalable transcrit les données qu'ils contiennent en « langage statistique ». Il peut s'agir d'un simple renommage des variables ou d'une transformation plus complexe. Cette opération dite de « mapping », qui concerne une centaine de variables, permet aux phases ultérieures du processus de ne jamais faire appel aux données sources mais uniquement à ces nouvelles variables. Cette étape du processus s'appuie sur l'application Arc⁵. À l'origine, l'application Arc avait été conçue pour réaliser également un certain nombre de contrôles de conformité mais l'expérience a montré que la quasi-totalité d'entre eux étaient inutiles puisque déjà réalisés par le processus de collecte de la DSN.

Une fois les données intégrées dans le système d'information, des traitements sont effectués sur chaque DSN mensuelle dans l'application Artemis⁶. Il s'agit à ce stade de s'assurer de la qualité de variables telles que l'identification des employeurs ou encore l'identification, le lieu de résidence et la catégorie socio-professionnelle des salariés (PCS). Le principe est de déterminer pour chaque information unitaire s'il y a ou non anomalie. S'agissant de l'identification des employeurs et de la commune de résidence des salariés, les informations déclarées sont comparées à celles figurant dans des référentiels (référentiel d'unités statistiques « Sirus » et répertoire « Sirene », code officiel géographique). Le nombre d'anomalies détectées sur un mois est très faible au regard du nombre de déclarations reçues à savoir de l'ordre d'un pour dix mille pour le SIRET de l'employeur et de l'ordre d'un pour mille pour le code de la commune de résidence du salarié. Pour la catégorie socio-professionnelle (PCS), on compare le code déclaré par l'employeur à celui calculé par l'outil Sicore⁷ à partir notamment du libellé de profession déclaré. Dans 90 % des cas, le code profession déclaré par l'employeur est confirmé par l'outil de codification automatique, ce qui signifie que la majorité des employeurs déclarent des codes et des libellés de profession cohérents. Malgré tout, ce sont environ 1,5 million d'anomalies qui doivent potentiellement être traitées chaque mois, ce qui n'est matériellement pas possible. Des arbitrages coût-qualité ont donc été faits dans un premier temps, qui conduisent à ne traiter que les divergences de codification portant sur les deux premières positions du code. Cela représente tout de même 40 % des anomalies à traiter. Compte tenu de leur nombre encore important, les cas à traiter sont regroupés en grappes homogènes. Le gestionnaire établit alors une règle de traitement au niveau de la grappe qui sera ensuite appliquée de façon auto-

4. Projet informatique de refonte sur l'emploi et les salaires.

5. Accueil Réception Contrôle.

6. Application de reprise et de traitements élémentaires de l'emploi et des salaires.

7. Système Informatique de COdage des Réponses aux Enquêtes.

matique à l'ensemble des cas à traiter dans la grappe. Ces règles sont capitalisées par l'application et seront appliquées de façon automatique lors du traitement des déclarations des mois suivants, ce qui doit conduire à diminuer à terme le nombre de cas à traiter par les gestionnaires. Les gains engendrés par ce processus de capitalisation seront réemployés pour traiter les cas de divergence portant sur les deux dernières positions du code PCS.

« Des arbitrages coût-qualité ont dû être faits et seules les anomalies les plus importantes sont soumises à l'expertise des gestionnaires. »

L'étape suivante consiste à analyser les incohérences internes aux DSN d'une même année entre les variables sur le volume de travail et celles relatives aux éléments de la rémunération, une fois les 12 mois de déclarations agrégés. Là encore, des arbitrages coût-qualité ont dû être faits et seules les

anomalies les plus importantes sont soumises à l'expertise des gestionnaires. Les traitements de détection et de correction des anomalies sont effectués dans l'application Diane⁸. Les méthodes de détection d'anomalies sont parfois relativement complexes et reposent pour certaines sur des modèles économétriques.

À ce stade des traitements, il reste encore à reconstituer l'information selon l'unité statistique d'intérêt (le poste de travail) et de calculer un certain nombre de variables statistiques. C'est l'objet de l'application Thésée⁹. Cette application permet également de produire un fichier sur l'ensemble des salariés c'est-à-dire en incluant les informations issues des systèmes d'information sur l'emploi et les salaires des agents de la fonction publique d'une part, et les salariés des particuliers employeurs d'autre part.

➊ APRÈS DEUX ANS D'UTILISATION, QUEL BILAN ET QUELLES PERSPECTIVES ?

La mise en place de la DSN ouvre des perspectives nouvelles pour la production des statistiques sur l'emploi et les revenus d'activité. Les attentes des utilisateurs en la matière sont importantes, en atteste le rapport du Conseil national de l'information statistique (Cnis) de mars 2016 sur la diversité des formes d'emploi. Il est cependant prématuré de fixer l'échéance à laquelle l'Insee sera en mesure d'exploiter toutes les potentialités de cette nouvelle source d'information. Par ailleurs, il faudra attendre 2022 et l'entrée de tous les employeurs publics en DSN pour avoir une couverture quasi exhaustive du champ des employeurs. À ce stade, il est tout de même possible de dresser un premier bilan.

Comme c'est souvent le cas dans les projets de cette ampleur, les coûts de transition ont été sous-estimés. L'Insee a dû composer avec un projet externe dont il ne maîtrisait pas le calendrier. Le rythme de montée en charge de la DSN a été plus lent qu'initialement prévu et a obligé à accueillir des DADS en nombre important et à faire cohabiter les deux déclarations administratives.

L'Insee a dû apprendre à évoluer dans un environnement qui n'est pas le sien, avec des acteurs habitués à travailler selon un certain formalisme faisant appel à la notion de norme d'échanges et avec des exigences en termes de réactivité. En quelque sorte, la DSN est venue bousculer la temporalité de nos travaux. Pour autant, l'Insee doit continuer à développer les échanges en amont avec ces acteurs afin de profiter pleinement de la boucle vertueuse que le projet externe DSN a su mettre en place. Un mode de travail collaboratif similaire mériterait d'être mis en place, cette fois en aval du processus,

8. Dispositif Informatique pour l'Agrégation et la Normalisation de l'Emploi.

9. Traitement Harmonisé des Salaires et de l'Emploi.

avec les acteurs du Système Statistique Public (SSP) ou du péri-SSP également destinataires de la DSN, avec qui l'Insee partage des préoccupations communes. Par exemple, la question de l'utilisation de la DSN pour alléger le questionnaire de certaines enquêtes statistiques se pose à l'Insee pour l'enquête coût de la main-d'œuvre et structure des salaires (Ecmoss) mais aussi à la Dares¹⁰ pour l'enquête activité et conditions d'emploi de la main-d'œuvre (Acemo), notamment.

En interne à l'Insee, il reste maintenant à mettre en place des processus industrialisés, documentés et sécurisés. Le traitement de la DSN à l'Insee doit s'appuyer sur une démarche qualité, ce qui suppose une description fine des processus, la mise en place d'un cadre d'assurance qualité et une revue régulière des processus. Ceci est d'autant plus important que le projet DSN n'est pas encore terminé qu'un nouveau chantier de simplification voit le jour. Il doit permettre aux organismes délivrant des revenus de remplacement (retraites, allocations chômage, etc.) de le faire à leur tour selon les mêmes principes que la DSN, c'est-à-dire selon une norme d'échanges précise. Dès lors, la démarche mise en place à l'Insee pour le traitement de la DSN à des fins statistiques devrait être facilement généralisable au traitement de cette nouvelle déclaration administrative.

BIBLIOGRAPHIE

Bonnet O., Cordier-Villoing M., Deroyon T., Djiriguian J. et Sakarovitch B., « 5 324 euros de l'heure : outlier ou footballeur ? Méthodes d'apprentissage non supervisé pour la détection d'anomalies : application au cas de la Déclaration sociale nominative », 13^e Journées de méthodologie statistique de l'Insee (JMS 2018), juin 2018.

Buhl J.L., « La Déclaration sociale nominative (DSN) et l'accès aux droits », Regards n° 46, EN3S, 2014.

Gazier B., Picart et C., Minni C., « La diversité des formes d'emploi », Rapport du Cnis, n° 142, juillet 2016.

Millin K., « Reconstitution des mouvements de main-d'œuvre depuis 1993 », Document d'études Dares, n° 221, juin 2018.

Rivière P., « Approche coût-qualité pour l'amélioration des processus de production statistique », Courrier des statistiques, n° 105-106, pp. 65-75, juin 2003.

Rouppert B., « Modélisation du processus de traitement d'une source administrative à des fins statistiques », Document de travail Insee, n° C2005/02, 2005.

Pour en savoir plus sur la DSN :

Site internet du Gip-MDS : www.gip-mds.fr

Site internet de la DSN : <http://www.dsn-info.fr/>

Cahier technique de la norme Neodes : <http://www.dsn-info.fr/dsn-norme-phase3.htm>

10. Direction de l'Animation de la Recherche, des Études et des Statistiques du ministère du travail.