

2.3 – Le contrôle qualité de la saisie

L'objectif du contrôle qualité est de valider le travail de saisie réalisé par le prestataire. Cette validation se fait au niveau de chaque lot de saisie et elle comprend trois phases :

1 - Un contrôle, par l'INSEE, des fichiers de données à leur réception selon des critères de lisibilité, d'exhaustivité, et de conformité.

2 - Un contrôle de lisibilité des Bases images Adresse des Logements (BAL) qui seront utilisées ensuite par l'Insee pour le repérage des logements, le RP servant de base de sondage pour les enquêtes auprès des ménages. Ce contrôle est réalisé par le prestataire de mesure de la qualité.

3 - Un contrôle de la qualité de la saisie sur les fichiers de données. Ce contrôle, pris en charge par l'INSEE, procède par confrontation et analyse des divergences sur la base d'un échantillon saisi en double par le prestataire de mesure de la qualité. Les taux d'erreurs attribuées au prestataire de saisie sont calculés par type de variable, comme le nombre de variables en erreur sur le nombre de variables à saisir sur les questionnaires.

Descriptif des traitements de chacune des phases :

1 – Contrôle des fichiers de données

Dès réception du fichier de données en provenance du prestataire de saisie, celui-ci est automatiquement contrôlé par l'Insee pour vérifier :

- la lisibilité du fichier ;
- son exhaustivité par rapport au fichier de référence issu du flashage ;
- sa conformité au format demandé ;
- la conformité de la structure des données : positionnement et modalités des variables ;
- la conformité des variables calculées (analyse ménage/famille) et de la répartition entre l'exploitation principale et complémentaire.

Tout manquement à l'un de ces critères entraîne le refus du fichier. Il est alors demandé au prestataire de livrer un nouveau fichier dans les 48H.

2 – Contrôle des Bases images Adresse des Logements

Pour chaque Lot de saisie, le prestataire transmet à l'Insee l'ensemble des Bases images des logements contenus dans le LS, obtenues à partir des feuilles de logements (FL) ou des feuilles de logements non enquêtés (FLNE). L'image comprend le cadre d'identification du logement, page 1 (« commencez ici ») pour les FL, et pour les FLNE, le concaténation d'une découpe du cadre A et d'une découpe du cadre B.

Dès réception des BAL d'un LS, un premier contrôle automatique permet de vérifier la cohérence entre les logements contenus dans le fichier des données et les BAL fournies (exhaustivité). Ensuite les BAL sont envoyées au prestataire de mesure de la qualité pour analyse.

Le prestataire analyse les images de la BAL suivant différentes étapes :

Première étape : analyse du volume des images

Le prestataire analyse le volume de chacune des images de la BAL, et fournit les renseignements suivants sur la distribution de ce volume (en Ko) :

- Le volume minimum et le volume maximum ;
- le volume moyen ;

- l'écart-type ;
- les volumes correspondant à chaque décile (10%, 20%, ...) ;
- taux d'images dont le volume est dans la fourchette.

Deuxième étape : tirage d'un échantillon d'images

Le prestataire tire un échantillon de 2 000 images selon la méthode suivante :

- les images sont triées par volume croissant ;
- les 200 premières et les 200 dernières images ainsi triées font partie de l'échantillon ;
- les 1 600 images restantes sont tirées aléatoirement dans le reste des images.

Troisième étape : calcul d'indicateurs pour les images de l'échantillon

Pour chacune des images de la BAL faisant partie de l'échantillon défini ci-dessus, le prestataire calcule la valeur des indicateurs suivants :

Nom de l'indicateur	Anomalie détectée	Calcul de l'indicateur
I 1	Zones sombres	Densité de pixels
I 2	Présence anormale de pixels isolés	Nombre de pixels isolés/ Nombre de pixels informants (hors masque)
I 3	Inclinaison (« skew »)	Inclinaison par rapport à une image de référence, en degrés
I 4	Calage latéral et vertical du document	maximum, en mm, entre le décalage horizontal et le décalage vertical par rapport à une image de référence

Quatrième étape : évaluation visuelle de 20 images minimum

Le prestataire visualise, parmi les images de l'échantillon, 20 images minimum en particulier parmi celles dont les valeurs des indicateurs I 1 à I 4 sont hors des fourchettes ou dont le volume de l'image est le plus faible ou le plus élevé.

Pour chacune de ces images, il fournit les renseignements suivants :

Renseignement	Commentaire
nom de l'image	
volume	
valeur de I 1	
valeur de I 2	
valeur de I 3	
valeur de I 4	
visualisation du fond d'image (correspondant aux renseignements imprimés sur le questionnaire)	0 : le fond d'image n'apparaît pas 1 : le fond d'image apparaît mais est difficile à reconnaître intégralement (décalage trop important, partie manquante, zébrures, etc.) 2 : le fond d'image est visuellement correct
visualisation de la partie manuscrite	0 : aucune information manuscrite n'est visible 1 : il y a des traces d'information manuscrite, probablement incomplète par rapport au questionnaire papier 2 : les informations manuscrites semblent correctement restituées sur l'image
commentaire littéral éventuel	

Dernière étape : rapport sur la mesure de la qualité de la BAL

Ce rapport reprend l'ensemble des éléments de l'analyse avec quelques commentaires si nécessaire. Il permet à l'Insee de valider ou non la qualité de la BAL transmise.

3 – Contrôle des erreurs de saisie

Dès la réception du fichier des données d'un LS, et si les contrôles du fichier ne font paraître aucune anomalie, l'Insee tire un échantillon de contrôle. Cet échantillon couvre à la fois les variables de l'exploitation principale et celles de l'exploitation complémentaire. Il comprend 700 images des questionnaires logements, 3000 images des questionnaires individus réparties entre exploitation principale et complémentaire, 1000 images des questionnaires spécifiques à l'analyse ménage/famille, et 500 images des questionnaires spécifiques à l'échantillon démographique permanent (EDP) qui nécessite la saisie de quelques données spécifiques (voir fiche 2.2).

La liste des logements constituant cet échantillon est envoyée au prestataire de saisie, qui transmet en retour les images des questionnaires correspondants. L'INSEE transmet ensuite ces images au prestataire de mesure de la qualité pour saisie.

Le prestataire de mesure de la qualité réalise la saisie des données contenues dans les images en fonction de leurs caractéristiques (exploitation principale ou complémentaire, analyse ménage/famille, EDP). Ces caractéristiques sont fournies en accompagnement des images. L'exploitation principale correspond à une saisie partielle des BI, alors que l'exploitation complémentaire comprend la saisie de l'ensemble des données du BI.

Les consignes de saisie sont identiques aux consignes fournies au prestataire de saisie, à l'exception des problèmes de qualité d'image. Lorsque la qualité de l'image est insuffisante pour saisir l'information, le contenu de la variable est remplacé par un dièse (#), et lorsqu'un caractère est illisible il peut être remplacé par une étoile (*).

Dès réception par l'Insee du fichier des données saisie de l'échantillon de contrôle, celui-ci est automatiquement contrôlé. Ensuite, ces données sont comparées aux données saisie par le prestataire de saisie, et toutes les divergences sont enregistrées pour être analysées. C'est l'Insee qui réalise l'analyse des divergences en regardant les images des questionnaires concernés. L'objectif de cette analyse est de repérer l'ensemble des erreurs de saisie réalisées par le prestataire de saisie et d'en déduire un taux d'erreurs observé. Cette analyse permet également d'obtenir un taux d'erreurs observé pour le prestataire de mesure de la qualité, afin de vérifier qu'il satisfait aux exigences du cahier des charges.

En fonction de la qualité observée, l'Insee ne réalise pas forcément un contrôle qualité sur la totalité des lots de saisie d'une campagne, en particulier, lorsque le niveau de qualité observé est satisfaisant.

Pour chacune des variables (ou type de variable), un taux maximum d'erreurs admises a été défini initialement par l'Insee et consigné au PAD dans le cahier des charges. Si, sur un lot de saisie, il est constaté que ce taux maximum est dépassé sur l'échantillon de contrôle, le lot est refusé dans son ensemble, et une nouvelle saisie est demandée au prestataire.

Mesures de la qualité des variables saisies pour la campagne 2008

Enquête 2008 - Bilan des contrôles qualité pour quelques LS

Type de Variable	Taux maximum d'erreurs admis	Taux d'erreurs				
		LS01	LS02	LS07	LS12	LS16
Sexe	0,10 %	0,00%	0,00%	0,00%	0,00%	0,10%
Situation principale	0,10 %	0,00%	0,04%	0,00%	0,04%	0,08%
Autres variables avec cases à cocher	1,00 %	0,19%	0,19%	0,22%	0,25%	0,23%
Jour mois et année de naissance	0,10 %	0,11%	0,06%	0,05%	0,03%	0,08%
Autres précasés	1,00 %	0,07%	0,00%	0,27%	0,09%	0,39%
Libellés avec dictionnaire (Communes Pays Nationalité)	1,00 %	0,36%	0,23%	0,29%	0,33%	0,26%
Libellés sans dictionnaire (Autres libellés)	2,00 %	1,00%	0,79%	0,69%	1,21%	0,83%
Analyse Ménage Famille	2,00 %	0,47%	0,00%	0,00%	0,00%	1,36%
Codes à barres	0,10 %	0,00%	0,00%	0,00%	0,00%	0,00%

Exemples d'images

Exemples d'images de la BAL

Pour une FL

Localisation du logement		
N° et voie (ou lieu-dit)		
Complément d'adresse (résidence, immeuble...)		
Commune		
Escalier	Étage	Position par rapport à l'escalier ou n°
Nom de l'occupant principal		


Pour une FLNE

B <i>Identification du logement</i>		
Identifiant		
rang L		
Bâtiment	Escalier	
Étage	Position par rapport à l'escalier ou n°	
Nom de l'occupant principal		

Exemples d'images de la base image de l'échantillon de contrôle (BIE)

Image de type EDP

Nom :		
Prénom :		
Adresse :		
1 Sexe	Masculin <input type="checkbox"/> 1	Féminin <input type="checkbox"/> 2
2 Date et lieu de naissance		
Né(e) le :		
jour	mois	année
à :		
commune		
département	n° DOM	pays pour l'étranger, territoire pour les TOM

Recensement de la population - 2004 Bulletin individuel			
<p>Exemple : DURAS, épouse MAURIV</p> <p>Nom : _____</p> <p>Prénom : _____</p> <p>Adresse : _____</p>		<p>Cadre à remplir par l'agent recenseur</p> <p>_____</p> <p>_____</p> <p>dépt. _____ commune _____</p>	
<p>1. Sexe Masculin <input type="checkbox"/> 1 Féminin <input type="checkbox"/> 2</p>		<p>6. La suite du questionnaire s'adresse aux personnes de 14 ans ou plus.</p>	
<p>2. Date et lieu de naissance</p> <p>Né(e) le : _____</p> <p> jour mois année</p> <p>à : _____</p> <p>_____</p> <p>_____</p> <p>département n° pays pour l'étranger, territoire pour les TOM</p> <p>Si vous êtes né(e) à l'étranger, en quelle année êtes-vous arrivé(e) en France ? _____</p>		<p>7. Vivez-vous en couple ? Oui <input type="checkbox"/> 1 Non <input type="checkbox"/> 2</p>	
<p>3. Quelle est votre nationalité ?</p> <p>• Française</p> <p> - Vous êtes né(e) français(e)..... <input type="checkbox"/> 1</p> <p> - Vous êtes devenu(e) français(e) (par exemple : par naturalisation, par déclaration, à votre majorité) <input type="checkbox"/> 2</p> <p> ↳ Indiquez votre nationalité à la naissance : _____</p> <p>• Étrangère <input type="checkbox"/> 3</p> <p> ↳ Indiquez votre nationalité : _____</p>		<p>8. Quel est votre état matrimonial légal ?</p> <p>• Célibataire (<i>jamais également marié(e)</i>)..... <input type="checkbox"/> 1</p> <p>• Marié(e) (ou séparé(e) mais non divorcé(e)) <input type="checkbox"/> 2</p> <p>• Veuf, veuve <input type="checkbox"/> 3</p> <p>• Divorcé(e) <input type="checkbox"/> 4</p>	
<p>4. Êtes-vous inscrit(e) dans un établissement d'enseignement pour l'année scolaire en cours ? Y compris apprentissage ou études supérieures.</p> <p>Oui <input type="checkbox"/> 1 Non <input type="checkbox"/> 2</p> <p>↳ Si oui, où est situé cet établissement ?</p> <p>• Dans la commune où vous résidez (ou dans le même arrondissement pour Paris, Lyon, Marseille) <input type="checkbox"/> 1</p> <p>• Dans une autre commune (ou un autre arrondissement) <input type="checkbox"/> 2</p> <p> ↳ Indiquez cette autre commune (précisez l'arrondissement) : _____</p> <p>_____</p> <p>commune (et arrondissement pour Paris, Lyon, Marseille) département n° TOM</p>		<p>9. Quel(s) diplôme(s) avez-vous ?</p> <p>• Vous n'avez pas été scolarisé(e)..... <input type="checkbox"/> 01</p> <p>• Aucun diplôme mais scolarité jusqu'en école primaire ou au collège..... <input type="checkbox"/> 02</p> <p>• Aucun diplôme mais scolarité au-delà du collège..... <input type="checkbox"/> 03</p> <p>• CEP (certificat d'études primaires) <input type="checkbox"/> 11</p> <p>• BEPC, brevet élémentaire, brevet des collèges <input type="checkbox"/> 12</p> <p>• CAP, brevet de compagnon..... <input type="checkbox"/> 13</p> <p>• BEP <input type="checkbox"/> 14</p> <p>• Baccalauréat général, brevet supérieur <input type="checkbox"/> 15</p> <p>• Baccalauréat technologique ou professionnel, brevet professionnel ou de technicien, BEA, BEC, BEI, BEH, capacité en droit..... <input type="checkbox"/> 16</p> <p>• Diplôme de 1^{er} cycle universitaire, BTS, DUT, diplôme des professions sociales ou de la santé, d'infirmier(ère)..... <input type="checkbox"/> 17</p> <p>• Diplôme de 2^e ou 3^e cycle universitaire (y compris médecine, pharmacie, dentaire), diplôme d'ingénieur, d'une grande école, doctorat, etc. <input type="checkbox"/> 18</p>	
<p>5. Où habitez-vous le 1^{er} janvier 1999 ? <i>Les enfants nés après cette date ne sont pas concernés.</i></p> <p>• Dans le même logement que maintenant <input type="checkbox"/> 1</p> <p>• Dans un autre logement de la même commune (ou du même arrondissement pour Paris, Lyon, Marseille) ... <input type="checkbox"/> 2</p> <p>• Dans une autre commune (ou un autre arrondissement pour Paris, Lyon, Marseille) <input type="checkbox"/> 3</p> <p> ↳ Indiquez cette autre commune (précisez l'arrondissement) : _____</p> <p>_____</p> <p>_____</p> <p>commune (et arrondissement pour Paris, Lyon, Marseille)</p> <p>département n° pays pour l'étranger, territoire pour les TOM</p>		<p>10. Quelle est votre situation principale ? <i>Ne cochez qu'une seule case.</i></p> <p>• Emploi (salarié ou à votre compte, y compris aide d'une personne dans son travail) ↳ cochez puis passez en 17 <input type="checkbox"/> 1</p> <p>• Apprentissage sous contrat ou stage rémunéré ↳ cochez puis passez en 17 <input type="checkbox"/> 2</p> <p>• Études (élève, étudiant) ou stage non rémunéré <input type="checkbox"/> 3</p> <p>• Chômage (inscrit ou non à l'ANPE) <input type="checkbox"/> 4</p> <p>• Retraite ou préretraite (ancien salarié ou ancien indépendant) <input type="checkbox"/> 5</p> <p>• Femme ou homme au foyer <input type="checkbox"/> 6</p> <p>• Autre situation <input type="checkbox"/> 7</p>	
<p>11. Travaillez-vous actuellement ? <i>Si vous avez un emploi occasionnel ou de très courte durée, ou si vous êtes en apprentissage ou en stage rémunéré, cochez « Oui ». Si vous êtes en congé maladie ou de maternité, cochez « Oui ».</i></p> <p>• Oui ↳ cochez puis passez en 17 <input type="checkbox"/> 1</p> <p>• Non ↳ cochez puis passez en 12 <input type="checkbox"/> 2</p>		<p>Continuez page suivante →</p>	