



MINISTÈRE DU LOGEMENT, DE L'ÉGALITÉ DES TERRITOIRES ET DE LA
RURALITÉ

Comité scientifique de l'observation des loyers

**Prescriptions méthodologiques de collecte, contrôle,
traitement et diffusion des données**

S O M M A I R E

Préambule	3
I. Les principes généraux de la méthode	
I. 1. Les finalités des observatoires locaux	4
I. 2. La méthode	4
II. La détermination de la cible	
II. 1. La grandeur à observer	5
II. 2. Les données à collecter	5
II. 3. Le champ d'observation	6
II. 4. Le périmètre géographique	6
II. 5. La déclaration à la Cnil	6
III. La préparation de la collecte	
III. 1. Le questionnaire	6
III. 2. Le fichier de saisie	6
III. 3. Les sources de données	7
III. 4. Les partenariats de collecte	7
III. 5. Le zonage	7
III. 6. La structure du parc et les objectifs de collecte	8
III. 7. Les modes de collecte	9
IV. La collecte, l'enrichissement et le contrôle des données	
IV. 1. Le guide de l'enquête	9
IV. 2. L'enrichissement des données	10
IV. 3. Le contrôle des données – dispositif général	10
IV. 4. Les contrôles locaux (niveau 1)	12
IV. 5. Les contrôles de conformité de la plateforme de dépôt (niveau 2)	12
IV. 6. Les contrôles de cohérence (niveau 3)	13
IV. 7. Les contrôles nationaux (niveau 4)	16
V. Le traitement et la diffusion des données	
V. 1. Les valeurs aberrantes	17
V. 2. Les effets de grappe	17
V. 3. Les poids initiaux	17
V. 4. La pondération – les marges de calage	18
V. 5. Le contrôle de la pondération – les ajustements	19
V. 6. La précision des médianes	19
V. 7. Les règles de diffusion des données	19
Annexes et documents complémentaires	20

Préambule :

Les prescriptions méthodologiques développées dans ce document s'appliquent à l'ensemble des Observatoires locaux des loyers (OLL) déléguant au Centre National de Traitement (CNT) le traitement des données de loyer récoltées.

Les OLL ayant choisi de réaliser eux-mêmes ces traitements doivent respecter l'ensemble des prescriptions qui s'appliquent au CNT. Néanmoins, afin de garantir l'homogénéité et la qualité des résultats, il est recommandé que l'ensemble des traitements statistiques soit effectué par le centre national de traitement, sous le contrôle du comité scientifique, pour ce qui concerne les OLL qui ne disposent pas de l'expérience suffisante pour les conduire.

Par ailleurs, les méthodes retenues par les observatoires locaux des loyers ayant été étalonnés par le Centre National de l'Information Statistique (CNIS) sont conformes aux prescriptions méthodologiques développées ci-après.

I. Les principes généraux de la méthode

I. 1. Les finalités des observatoires locaux

La mise en place d'observatoires locaux des loyers (OLL) doit permettre d'améliorer la connaissance des loyers pratiqués dans le parc locatif privé et de disposer de données fiables, homogènes et comparables d'un territoire à l'autre.

Les informations diffusées peuvent contribuer à la définition de politiques publiques locales et nationales (PLH, zonages, plafonds Anah, encadrement des loyers...) mais aussi aider les professionnels (administrateurs de biens, promoteurs...) et les particuliers (bailleurs et locataires) dans leurs arbitrages.

I. 2. La méthode

La méthode retenue est pragmatique : elle tient compte de ce qu'il est possible de faire tout en imposant des contrôles à toutes les étapes du processus de production.

Elle peut être résumée en cinq phases :

- **détermination de la cible**
 - grandeur à observer
 - variables à collecter
 - champ d'observation
 - périmètre géographique
- **préparation de la collecte**
 - questionnaire, fichier de saisie
 - recensement des sources, prospection des professionnels
 - choix du zonage
 - structure du parc et objectifs d'enquête
 - modes de collecte : extraction (notices), enquêtes directes en interne ou externalisées
- **collecte, enrichissement et contrôle des données**
 - récupération des extractions de fichiers de professionnels
 - réalisation des enquêtes en gestion directe
 - contrôle et enrichissement des données collectées
 - dépôt des données sur la plateforme de dépôt
 - aller et retour avec le centre national de traitement jusqu'à validation de l'ensemble des données déposées
- **traitement des données**
 - recherche de doublons
 - traitement des grappes
 - pondération
 - calcul de la précision
- **restitution et diffusion des résultats**
 - règles de diffusion
 - publication minimale
 - résultats agrégés
 - bases de données individuelles.

Cette méthode a été validée par un comité scientifique indépendant et composé de cinq experts dans les domaines des statistiques et/ou du logement.

II. La détermination de la cible

II. 1. La grandeur à observer

La variable observée est le loyer mensuel hors charges, ramené généralement au m² de surface pour faciliter les comparaisons. La surface retenue est la surface habitable.

L'observatoire ne mesure donc pas la dépense logement des locataires même si les provisions pour charges sont demandées lors de l'enquête pour s'assurer que le loyer fourni est bien hors charges.

II. 2. Les données à collecter

Selon les finalités, les données à collecter peuvent différer légèrement. Il a été retenu un socle minimal de données à recueillir, selon une périodicité annuelle.

Les données à collecter ont été réparties en trois catégories :

- les obligatoires,
- les obligatoires à terme,
- les facultatives.

Les données à recueillir sont, à la date de rédaction de ces prescriptions, les suivantes :

Année d'enquête

Date de référence (1^{er} janvier par défaut)

Adresse complète du logement

Code Insee, zone

Iris, section cadastrale, parcelle

Type d'habitat (maison, appartement)

Epoque de construction

Année d'achèvement de construction

Type de propriété (mono, copropriété)

Équipement du logement (ascenseur, gardien, interphone, digicode, vidéophone)

Nombre de pièces

Surface habitable

Étage

Annexes (parking, balcon, jardin)

Nombre de WC et de salles d'eau

Type de chauffage

Type de propriétaire

DPE

Type de location (vide, meublé)

Type de gestion

Catégorie de location (libre, conventionnée, défiscalisée)

Date d'entrée du locataire

Première location (oui/non)

Loyer mensuel hors charges à la date d'enquête

Loyer mensuel hors charges à la date de référence

Provisions pour charges à la date d'enquête

(En gras : indispensable en 2014, en normal non gras : indispensable à terme, en italique : optionnel)

II. 3. Le champ d'observation

Le champ d'observation porte sur les logements locatifs non HLM donc à loyer libre mais aussi les conventionnés Anah, les PLS privés... Il comprend les meublés non professionnels et non saisonniers.

L'observation porte sur le stock de logements donc sur tous les logements locatifs privés quelle que soit l'ancienneté de leur locataire, et non sur les seuls logements ayant connu un emménagement récent.

Le champ d'observation est précisé dans une note disponible sur l'extranet du réseau www.observatoiresdesloyers.fr

II. 4. Le périmètre géographique

Le périmètre géographique est libre. Il doit cependant comprendre au minimum quelques milliers de logements locatifs privés pour permettre de sortir des résultats respectant les préconisations de la méthode en particulier sur le seuil de diffusion.

II. 5. La déclaration à la CNIL

Préalablement à la première collecte, une déclaration à la CNIL, dite normale, doit être souscrite par l'observatoire local. Cette déclaration est obligatoire et unique.

Un modèle de déclaration est disponible sur l'extranet du réseau : www.observatoiresdesloyers.fr .

III. La préparation de la collecte

III. 1. Le questionnaire

Il reprend toutes les variables du § II. 1. sur les données à collecter quelles soient obligatoires ou optionnelles. Il comprend de plus des variables destinées à gérer l'enquête et les contrôles : source des données...

Le questionnaire est organisé en quatre parties :

- les données de gestion de l'enquête,
- la description de l'immeuble,
- la description du logement,
- le bail et le loyer.

Le questionnaire est mis à la disposition des OLL mais son utilisation n'est pas obligatoire, la plupart des enquêtes se faisant par transfert électronique ou par téléphone avec un CATI (Computer Assisted Telephone Interview). Il est fourni en annexe 1 de ces prescriptions.

III. 2. Le fichier de saisie

C'est le fichier qui permet de déposer les données sur la plateforme. Il est fourni aux OLL sous format excel et doit être transformé en .csv pour le dépôt. En cas de recours au centre national de traitement, son utilisation et le passage par la plateforme de dépôt sont obligatoires, aucun autre fichier ne sera accepté.

Le contenu du fichier de saisie ne doit pas être modifié : l'ordre, le format et la longueur des variables doivent être respectés. En particulier le séparateur est le point-virgule et les noms de dépôt des fichiers doivent respecter une certaine sémantique :

LxxxxLOTnn.csv

Avec xxxx = code observatoire attribué par le centre national de traitement

nn = numéro de lot dans la collecte de l'année

Quel que soit le mode de collecte (extraction, téléphone...) les données doivent être retranscrites dans ce fichier en respectant les préconisations (cf. infra partie sur les contrôles).

III. 3. Les sources de données

Les données de loyer peuvent être obtenues auprès de différentes sources :

- directes : locataire, bailleur, gestionnaire (administrateur de biens, notaire...)
- indirectes : Adil, collecteur d'Action Logement, CAF...

Il a été décidé de n'exclure aucune source mais de privilégier celles permettant d'obtenir des données en grand nombre et pour un coût modéré : en particulier les professionnels de l'immobilier, qui ont l'obligation de transmettre l'intégralité des données de loyer du parc qu'ils gèrent sur les territoires où il existe un observatoire local des loyers agréé (décret n°2014-890 du 1^{er} août 2014 relatif au plafonnement des honoraires imputables aux locataires et aux modalités de transmission de certaines informations par les professionnels de l'immobilier).

Pour préparer efficacement sa collecte, l'OLL doit identifier la part des différentes sources, en particulier professionnelles, dans le parc de son périmètre et remplir le tableau des sources (annexe 2).

Ce recensement permet entre autre d'estimer la part de la gestion déléguée dans le parc, en l'absence de source externe exhaustive.

III. 4. Les partenariats de collecte

Une fois les sources recensées, l'OLL s'attache à mettre en place des partenariats avec les principales organisations de professionnels et de particuliers : Fnaim, Unis, Snpi, chambre des notaires, Unpi, autres associations de bailleurs... Cette liste n'est pas limitative et doit être adaptée au contexte local. Ainsi la gestion locative par les notaires est plus ou moins développée selon les régions, l'opportunité d'un partenariat avec eux est donc à apprécier localement.

Ces partenariats institutionnels peuvent être complétés par des partenariats de collecte bilatéraux formalisés au moyen d'une convention disponible sur l'extranet du réseau www.observatoiresdesloyers.fr.

III. 5. Le zonage

Le choix d'un bon zonage est essentiel.

Selon l'importance du parc locatif privé, le territoire d'observation est découpé en une ou plusieurs zones. Les zones idéales sont homogènes en termes de niveaux de loyer, toutes choses égales par ailleurs. Elles doivent aussi contenir suffisamment de logements pour respecter les règles de diffusion des données agrégées.

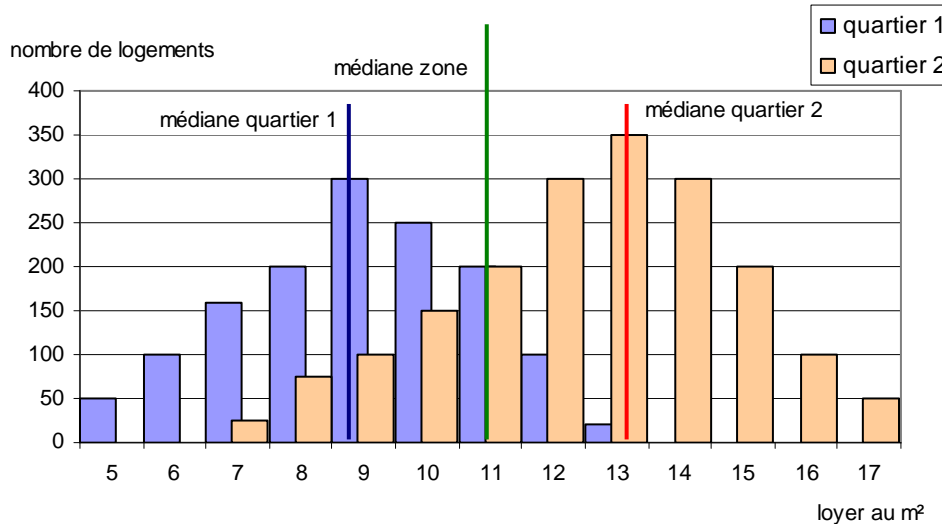
Les zones sont donc rarement des zones administratives préexistantes (PLH, zones d'habitat social...). Elles doivent être spécifiquement conçues pour l'observation des loyers privés.

Il est cependant difficile de les construire en l'absence d'informations fiables individuelles sur les loyers et, dans la plupart des cas, le zonage se fera en deux temps :

- la première année, utilisation d'un zonage a priori (préexistant ou non),
- la deuxième année, révision de ce zonage à l'aide de la première collecte par un traitement économétrique des données.

Une zone ne doit pas regrouper deux quartiers ayant des niveaux de loyer très différents.

Zonage : ce qu'il ne faut pas faire



Dans la pratique les zones sont des regroupements d'iris (ou de sections cadastrales) et/ou de communes, fréquemment non contigus.

Le choix de zones n'est pas neutre car les objectifs d'enquête et les traitements seront faits selon le zonage. De même les données agrégées seront produites selon ces zones ou des regroupements de zones. Par exemple, pour publier des données sur la ville centre de l'agglomération, il ne faut pas définir une zone comprenant une partie de cette ville et d'autres communes. Le zonage final est généralement la juxtaposition de zones infra-communales pour la ville centre et de regroupements de communes pour la périphérie.

III. 6. La structure du parc et les objectifs de collecte

Des tableaux descriptifs de la structure du parc observé sont établis à partir de sources externes, essentiellement le recensement de la population de l'Insee et Filocom. A minima ces tableaux donnent la répartition des logements locatifs privés par zone, type d'habitat (maison ou appartement) et nombre de pièces.

La collecte idéale doit être réalisée avec un échantillon tiré aléatoirement à partir d'une base de sondage de logements locatifs et/ou de locataires). L'OLL ne disposant généralement pas d'une telle base, la collecte doit obéir à des règles visant à limiter les biais statistiques.

L'introduction d'une part de collecte en gestion directe répond à ce souci car rien ne dit que :

- les logements gérés directement ont les mêmes caractéristiques que ceux donnés en gestion,
- les niveaux de loyer pratiqués sont les mêmes, toutes choses égales par ailleurs.

Le centre national de traitement vérifie que les zones sont suffisamment importantes pour établir des statistiques diffusables et établit les objectifs d'enquête de la façon suivante :

- détermination de la répartition gestion directe – gestion déléguée à partir du tableau de recensement des sources et en concertation avec l'OLL,
- choix d'un taux de sondage pour chaque type de gestion en fonction du zonage, le taux de sondage doit permettre de « couvrir » correctement la plus petite zone pour pouvoir diffuser les principaux agrégats statistiques (cf. infra les seuils de diffusion).
- application de façon uniforme du taux de sondage retenu pour déterminer les objectifs par zone, type d'habitat et nombre de pièces,
- fixation d'un nombre minimal de professionnels à collecter en fonction du tableau des sources,
- validation avec l'OLL.

III. 7. Les modes de collecte

L'organisation de la collecte diffère selon qu'elle est réalisée auprès des professionnels ou directement auprès des particuliers (bailleurs ou locataires).

Pour les professionnels :

- si l'observatoire est agréé, la collecte est exhaustive
- sinon l'OLL constituera un fichier de contacts qu'il prospectera pour récupérer l'intégralité des parcs gérés (administrateurs de biens) ou loués (agences immobilières).

Pour les particuliers :

- en raison du coût des enquêtes directes auprès des particuliers, l'OLL fera une collecte auprès d'un échantillon obtenu par tirage aléatoire dans une ou plusieurs bases externes (fichier de coordonnées téléphoniques obtenu auprès d'un prestataire, contacts récupérés auprès de diverses sources (Adil, Unpi, collecteurs d'Action Logement, etc.)
- réalisation des enquêtes directement ou par l'intermédiaire d'un prestataire

L'OLL organisera un suivi de sa collecte et vérifiera en particulier que les objectifs fixés sont bien atteints.

Une source ne doit pas être prépondérante dans la collecte. Un nombre minimal de professionnels à enquêter est ainsi fixé dans les objectifs par le Centre national de traitement. Du côté des enquêtes directes, essayer de ne pas utiliser une source au-delà de sa part (estimée) dans le parc. Ainsi si les Adil donnent des consultations sur 10 % du parc locatif privé dans l'année, la collecte à partir des contacts Adil doit représenter environ 10 % des enquêtes directes.

IV. La collecte, l'enrichissement et le contrôle des données

IV. 1. Le guide de l'enquête

La collecte se fait conformément au guide de l'enquête qui détaille chaque variable du fichier de saisie : définition, modalités, format. Le guide d'enquête est actualisé tous les ans et disponible sur l'extranet du réseau www.observatoiresdesloyers.fr.

L'identifiant logement

Conformément à ce que prévoit le guide d'enquête, l'OLL attribuera à chaque logement enquêté un identifiant. Cet identifiant sera propre au logement et utilisé pour chaque enquête sur ce logement (suivi dans le temps). En cas de sortie du logement de l'observation (démolition...), l'identifiant ne sera pas réattribué à un autre logement. En revanche un logement qui disparaît provisoirement de l'échantillon (vacance...) récupérera le même identifiant à son retour dans la collecte.

A noter que l'identifiant logement attribué par l'OLL n'est en aucun cas celui que le professionnel utilise pour sa gestion.

La surface habitable est définie à l'article R*112-2 du Code de la Construction et de l'Habitation :

"La surface habitable d'un logement est la surface de plancher construite, après déduction des surfaces occupées par les murs, cloisons, marches et cages d'escaliers, gaines, embrasures de portes et de fenêtres ; [le volume habitable correspond au total des surfaces habitables ainsi définies multipliées par les hauteurs sous plafond].

Il n'est pas tenu compte de la superficie des combles non aménagés, caves, sous-sols, remises, garages, terrasses, loggias, balcons, séchoirs extérieurs au logement, vérandas, volumes vitrés prévus à l'article R. 111-10, locaux communs et autres dépendances des logements, ni des parties de locaux d'une hauteur inférieure à 1,80 mètre."*

IV. 2. L'enrichissement des données

Une fois collectées, les références peuvent être insuffisamment renseignées par rapport aux spécifications de la méthodologie. Les principaux cas de non-conformité sont :

- l'absence de l'époque de construction,
- l'adresse non normalisée.

L'observatoire local dispose ou doit se doter d'outils pour compléter les données :

- les fichiers fonciers (dits Majic)
- un outil de normalisation des adresses
- des outils locaux.

IV. 3. Le contrôle des données – dispositif général

Des contrôles sont organisés à plusieurs niveaux du processus de collecte et de traitement des données :

- 1. contrôles locaux lors de la collecte ; ces derniers figurent dans le document « guide d'enquête », ils sont faits au fur et à mesure de la collecte par les enquêteurs eux-mêmes et/ou les superviseurs ; ils concernent la pertinence et la cohérence individuelle de chaque référence collectée ;
- 2. contrôles automatiques lors du dépôt des fichiers de saisie sur la plateforme du centre national de traitement ; ils vérifient le respect des spécifications (nom de fichier, ordre des variables, variables obligatoires...)
- 3. contrôles de cohérence sur chaque lot déposé ; sous forme d'échanges itératifs entre l'observatoire local et le centre national de traitement ; certains des contrôles du point 1 sont repris et il sera demandé le cas échéant à l'observatoire local (OLL) de valider l'anomalie ;
- 4. contrôles préalables au traitement par le centre national de traitement : doublons, effets de grappe..., faits après la fusion des lots de l'OLL et des éventuelles données issues de sources nationales.

Les contrôles doivent être faits dans l'ordre : on ne peut passer à un niveau de contrôle supérieur que si le niveau précédent a été validé.

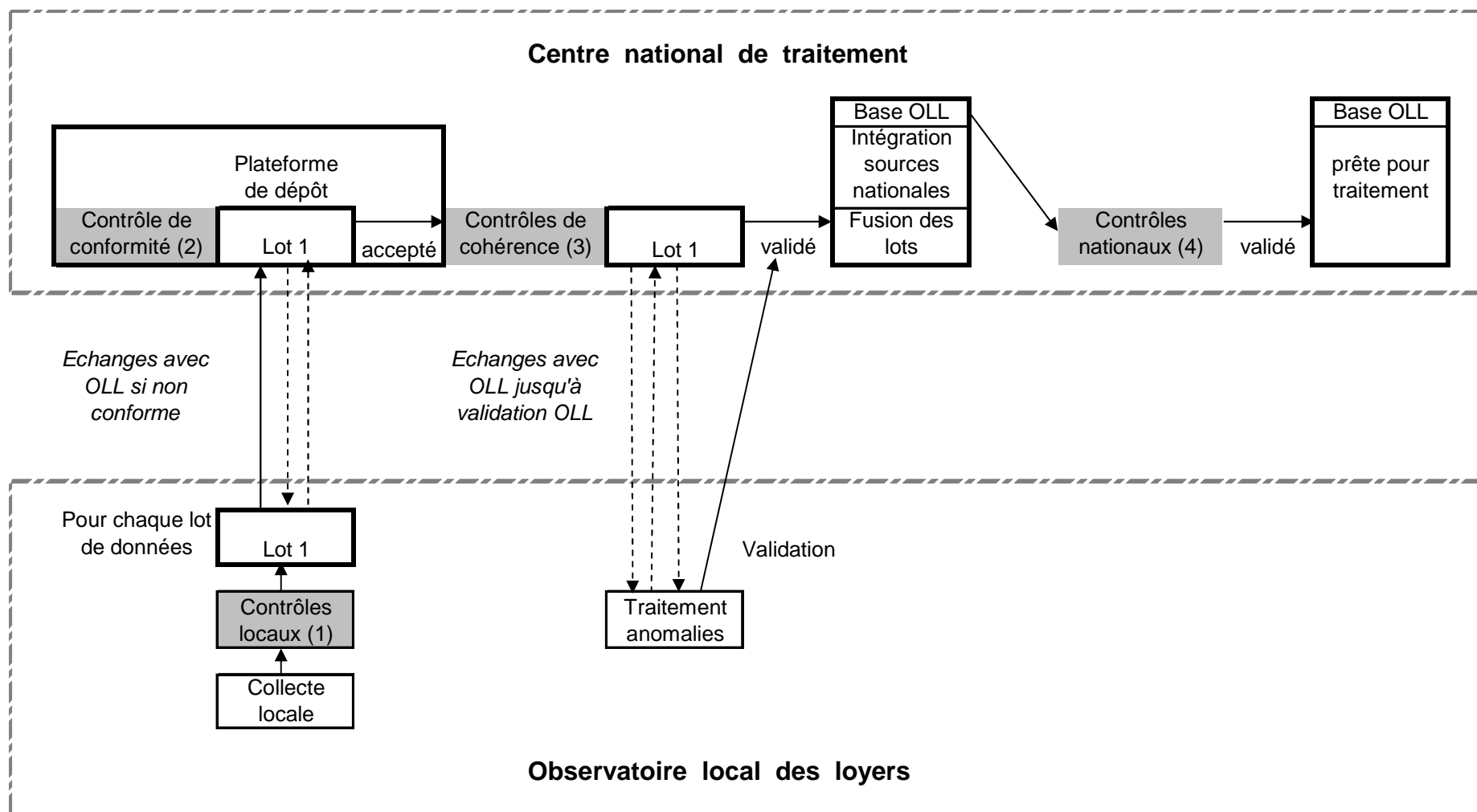
Certains contrôles sont volontairement redondants. Ainsi la plupart des contrôles recommandés au niveau (1) avant dépôt du fichier de données sont refaits par sécurité au niveau (3).

A l'issue des quatre niveaux de contrôle, est obtenue la base « prête pour le traitement » sur laquelle sont appliquées les opérations de la partie V – traitement des données.

Transmission de données des professionnels via un organisme tiers au centre de traitement

Chaque OLL reçoit du centre de traitement l'intégralité des données de son périmètre géographique sans aucune modification. Ces données sont complétées et contrôlées par l'OLL et sont ensuite déposées sur la plateforme selon les mêmes modalités que les autres lots.

Schéma du processus de validation des données collectées



IV. 4. Les contrôles locaux (niveau 1)

Il est fortement recommandé aux observatoires locaux de réaliser l'ensemble des contrôles de la liste ci-dessous pour chaque référence collectée avant envoi du fichier au centre national de traitement. Ces contrôles figurent dans le guide d'enquête. Si l'observatoire ne parvient pas à corriger l'anomalie détectée (complément d'enquête, recours à une base de données externe...), la référence sera retirée du fichier si l'une des variables obligatoires est concernée.

Certains contrôles sont adaptés aux spécificités locales : surface / nombre de pièces, fourchette de loyer au m²...

Voici une liste non exhaustive de vérifications telle qu'elle figure dans le guide d'enquête :

- le nom du lot et de l'observatoire
- le format du fichier (csv)
- le remplissage des variables obligatoires
- l'unicité de l'identifiant logement
- que la modalité existe pour les variables avec un choix de modalités
- que les variables loyers et charges sont bien renseignées (et supérieures à un minimum local...)
- que la variable surface est bien renseignée
- que la commune est dans le champ géographique de l'observatoire
- que la date d'enquête est bien renseignée et dans l'année N
- que le n°, le type de voie, le nom de voie... sont renseignés (liste des types de voie à utiliser disponible sur l'extranet www.observatoiresdesloyers.fr)
- le code postal et l'écriture officielle du nom de la commune, puis du code Insee
- le bon renseignement du zonage (et autre géocodage)
- que le logement n'est pas une maison avec ascenseur ou à l'étage
- qu'un logement construit avant 1975 n'est pas au-dessus du 15^e étage
- qu'un logement construit après 1990 a bien WC et SdB
- qu'un appartement construit après 1975 et au moins au 4^e étage a un ascenseur
- qu'un appartement construit après 1975 a du chauffage
- que si l'enquête est faite auprès d'un professionnel la raison sociale est renseignée
- que si le DPE est A ou B, le logement est construit après 1990
- entre le nombre de pièces et la surface : 1 pièce entre ... et ... m², 2 pièces entre ... et ... m², etc. (bornes à compléter par l'observatoire local)
- le loyer en €/m² à la date d'enquête et au 1^{er} janvier selon le nombre de pièces est compris dans des fourchettes à définir selon le territoire
- que le loyer est compris dans des fourchettes à définir selon le nombre de pièces et le territoire
- les évolutions de loyer, le cas échéant
- que les charges au m² (si renseignées) sont entre 0,05 et 6 €
- que la date d'emménagement est antérieure à la date d'enquête
- que les différences de loyer (surtout si faible écart d'1 ou 2€) entre les 1^{er} janv. N et /1^{er} janv. N-1 ou entre le 1^{er} janv. et la date d'enquête correspondent bien à une augmentation et non à un arrondi pour l'une des dates
- en cas de baisse (notamment au cas où inversion de remplissage des loyers N et N-1).

IV. 5. Les contrôles de conformité de la plateforme de dépôt (niveau 2)

Les contrôles automatiques lors du dépôt d'un fichier par un observatoire local sont les suivants :

1. nom de fichier et extension valides
2. variables dans l'ordre et avec le bon format
3. sur les variables obligatoires ou non et les valeurs qu'elles peuvent prendre :
 - identifiant logement : obligatoire
 - source : obligatoire et compris entre 1 et 7
 - mode de collecte : obligatoire et compris entre 1 et 4
 - date d'enquête JJ MM AA : obligatoire et AA = année en cours
 - adresse : obligatoire pour code postal, commune, Insee et zone avec code postal et Insee différents de 0

- raison sociale du professionnel non obligatoire mais si source = 1 ou 2, alors colonnes X (n° carte gestion) et Z (raison sociale) obligatoires
- type d'habitat : obligatoire et vaut 1 ou 2
- époque de construction : obligatoire et compris entre 1 et 7
- nombre de pièces : obligatoire et compris entre 1 et 20
- surface : obligatoire et compris entre 5 et 500, en unités sans décimales
- type de location : obligatoire et vaut 1 ou 2
- type de gestion : obligatoire et vaut 1 ou 2
- année d'emménagement obligatoire et au plus égale à année en cours
- loyer mensuel N à la date d'enquête (colonne BE) obligatoire et compris entre 80 et 15 000 euros mensuels.

Les noms de colonnes renvoient au fichier de saisie : Fichier de saisie OLL.xls, disponible sur l'extranet du réseau www.observatoiresdesloyers.fr.

Des exemples de messages retournés à l'observatoire local sont donnés ci-après :

- *extension invalide (envoyer un fichier en .csv)*
- *au moins un identifiant logement manquant*
- *au moins un code source manquant*
- *au moins un code source invalide*
- *au moins un mode de collecte manquant*
- ...

Il n'est pas prévu à ce niveau de donner la liste des lignes en anomalies (pour éviter de renvoyer 6000 lignes à un observatoire qui a oublié l'identifiant logement pour tout le fichier par exemple), mais simplement la première ligne sur laquelle se produit l'anomalie.

Un lot ne peut être déposé et donc accessible au centre national de traitement que s'il répond à ces recommandations minimales.

Le centre national de traitement utilise le logiciel SAS V9 pour le contrôle et le traitement des données. Il n'y a pas de préconisation particulière d'outils pour les observatoires locaux, sous réserve que les fichiers déposés soient acceptés par la plateforme.

IV. 6. Les contrôles de cohérence (niveau 3)

Ces contrôles de données se font de façon individuelle, c'est-à-dire observation par observation et en deux phases :

- la première directement à la lecture des données : le contrôle est général et reprend les tests du contrôle de conformité (nom du fichier...) au cas où une mauvaise manipulation aurait été faite après la phase de dépôt. Cette étape comprend aussi la création de variables complémentaires.
- la seconde afin de contrôler les variables plus précisément et assurer ainsi la qualité des données, que cela soit sur le remplissage des données obligatoires, mais aussi les formats, les cohérences ...

Première partie : Chargement des données

1. Des variables complémentaires sont créées automatiquement notamment pour faciliter le traitement statistique : identifiant logement centre national, origine du lot (local, national), date de chargement...
2. Des variables de localisation sont contrôlées (uniformisation des noms de commune selon la nomenclature Insee ou encore vérification des cohérences entre zone/Insee, code postal/nom de commune/Insee)

3. La jonction avec les autres lots déjà déposés de l'observatoire local permet d'effectuer une recherche de doublons sur l'identifiant logement. Cette variable associée à celle du code observatoire assure l'unicité d'un identifiant nécessaire à toute gestion de base de données. Ce contrôle se fait à chaque nouveau dépôt d'un lot avec les identifiants du lot et des lots précédents et une dernière fois dans les contrôles nationaux quand toutes les données d'un observatoire (locales et nationales) auront été fusionnées.

Seconde partie : Contrôle individuel des données

Chaque observation est ensuite contrôlée de façon individuelle. Cette étape permet de s'assurer du bon remplissage du fichier fourni au centre national et de relever les erreurs mais aussi de porter l'attention sur des données qui pourraient paraître douteuses lors du traitement (comme par exemple un doute sur la cohérence entre un loyer et une surface).

Les contrôles indiqués ci-après sont susceptibles d'être modifiés ou complétés à la marge. Ce sont ceux en vigueur pour la collecte 2014.

Contrôles d'ordre général

4. Remplissage des variables obligatoires : toute donnée non renseignée ou non correctement remplie (principalement choix d'une modalité non proposée) est à corriger ou à forcer.
5. Remplissage des variables non obligatoires : toute donnée non correctement remplie (principalement choix d'une modalité non proposée) est à corriger ou à supprimer.
6. Vérification de la date d'enquête : permet entre autres de s'assurer que le dépôt sur la plateforme concerne bien des données de l'année en cours.
7. Vérification que le logement appartient bien au périmètre géographique fourni par l'observatoire.

Ces contrôles ont déjà été faits lors du dépôt du fichier, ils sont refaits à titre de sécurité et s'appliqueront aussi aux données issues de sources nationales.

Cohérences en rapport avec la source de données

8. Si la source est un administrateur de biens ou une agence immobilière, vérification que la raison sociale est renseignée.
9. Si la source est un bailleur en gestion directe ou un locataire en gestion directe, vérification qu'aucune raison sociale n'est remplie et que la gestion est bien codée comme directe.
10. A l'inverse, si la gestion est déléguée, vérification que la source n'est pas un bailleur en gestion directe ou un locataire en gestion directe.
11. Si le mode de collecte est un transfert de fichiers, vérification que la source est un administrateur de biens, une agence immobilière, une personne morale ou un autre professionnel (à défaut, « autres »).
12. Si le mode de collecte est internet, vérification que la source n'est pas un professionnel.

Cohérences en rapport avec l'époque de construction

13. Vérification de la cohérence entre l'époque de construction et l'année de construction (si celle-ci est renseignée).
14. Vérification que la date d'emménagement est supérieure à la borne inférieure de l'époque de construction (test non possible pour la première époque de construction) – de même pour la cohérence entre l'année d'emménagement et l'année de construction (si celle-ci est renseignée).
15. Si la construction de l'appartement date d'avant 1971, vérification que le logement se trouve au plus au 15^e étage (si celui-ci est renseigné).
16. Si la construction du logement date d'après 1990, vérification que le logement possède au moins un WC et une salle de bains (si ceux-ci sont renseignés).
17. Vérification que le nombre de toilettes et de salles de bains est inférieur à 9.
18. Si la construction du logement date d'avant 1946, vérification que le DPE soit compris entre D et G (si celui-ci est renseigné).
19. Si le DPE est A ou B, vérification que l'époque de construction est postérieure à 1990.

20. Si la construction de l'appartement date d'après 1990 et qu'il est situé au moins au 4^e étage, vérification de la présence d'un ascenseur (si celui-ci est renseigné).

Autres cohérences sur la description du logement et de l'immeuble

21. Si le logement est une maison, vérification qu'il n'y a pas d'ascenseur et que le logement se trouve au rez-de-chaussée.
22. Vérification que le nombre de pièces est compris entre 1 et 20.
23. Vérification que la surface est comprise entre 5 et 500.
24. Vérification de la cohérence pour chaque observatoire entre le nombre de pièces et la surface du logement.

Cohérences sur les informations concernant le bail et les loyers

25. Vérification que le locataire a emménagé avant la date d'enquête et après 1920 et au plus dans l'année en cours.
26. Vérification que si le propriétaire est une personne morale, la gestion est codée comme déléguée (par convention).
27. Vérification que les montants des loyers sont compris entre 80 et 15 000 € (pour éliminer les erreurs de virgule...).
28. Vérification que les montants des provisions pour charges (si ceux-ci sont renseignés) sont compris entre 0,05 et 6 €/m² (afin notamment de voir une possible inversion de saisie entre les charges et les loyers).
29. Vérification selon la localisation du logement et le nombre de pièces que le loyer au m² se situe dans une fourchette définie.

Si calcul d'évolution

30. Vérification de l'évolution si les loyers N et N-1 sont renseignés (notamment en cas de locataires stables). Affichage d'une mise en garde en cas de baisse ou de hausse supérieure à l'IRL.
31. Vérification en cas de nouvel emménagement de la variable « première location » (si celle-ci est renseignée) : en cas de première location, impossibilité de loyer au 1^{er} janvier N-1 et nouvel identifiant logement...

Les incohérences détectées seront regroupées dans un document indiquant pour chaque identifiant logement les problèmes rencontrés, ce document est envoyé à l'observatoire local, qui doit corriger ou valider les anomalies. Dans ce dernier cas, un code « forçage » de l'observation est mis en place, la validation ne pourra se faire par défaut : « tout ce qui n'a pas été corrigé est correct ».

Après corrections, l'observatoire local redépose son fichier qui repasse à la moulinette des tests de cohérence jusqu'à ce que toutes les anomalies aient été corrigées. L'observatoire local valide alors le lot.

Comme déjà signalé plus haut, les données issues des sources nationales et extraites pour chaque observatoire local dans un lot spécial par le centre de traitement feront l'objet des contrôles de ce paragraphe et d'une validation par le centre de traitement lui-même.

IV. 7. Les contrôles nationaux (niveau 4)

Les contrôles résiduels à faire à ce niveau sont peu nombreux mais essentiels. Ils sont effectués après fusion dans une même base de toutes les données collectées (localement ou nationalement) pour un observatoire.

1. Intégration des sources nationales

Le centre de traitement complète la collecte locale par les données récupérées auprès des sources nationales. Ces données font l'objet des mêmes contrôles que les données locales.

Le centre de traitement fait à ce niveau un contrôle de cohérence Insee/zone et source/type de gestion et une ultime vérification que les variables obligatoires sont bien renseignées.

2. Contrôle des doublons

La collecte multimode mise en place ne permet pas de garantir que le même logement ne sera pas interrogé deux voire trois fois. Par ailleurs, on ne peut exclure une saisie en double d'un même logement d'une même source, les transferts électroniques étant loin d'être généralisés.

Les doublons peuvent correspondre aux cas suivants :

- une agence et un administrateur de biens,
- une agence et une personne morale,
- une personne morale et un administrateur de biens
- un professionnel et un particulier, bailleur ou locataire
- tous les cas précédents avec une donnée du centre national de traitement.

La méthode retenue permet d'éliminer la plus grande partie des doublons, mais ne garantit pas l'élimination de tous les doublons réels en l'absence de caractéristiques suffisantes des logements.

Les doublons potentiels sont identifiés si correspondance exacte sur les variables suivantes :

- adresse complète
- année d'entrée dans le logement
- type de logement (maison, appartement)
- étage
- nombre de pièces
- loyer mensuel à la date d'enquête.

N.B. :

1. l'étage n'étant pas renseigné dans tous les cas, on ne détecte pas ainsi tous les doublons potentiels. Certains cas seront traités avec les grappes.
2. cela suppose une normalisation des adresses
3. un ensemble de logements correspondant à une personne morale est considéré comme une grappe et non comme des doublons.

Ces logements sont regardés un à un et arbitrés au cas par cas avec la règle de gestion suivante : on retient dans cet ordre la donnée obtenue au niveau national, puis la donnée professionnel local (personne morale puis administrateur de biens puis agence), puis la donnée bailleur privé, puis la donnée CAF, puis la donnée locataire.

Avec cette règle, deux logements avec deux raisons sociales différentes et une année d'entrée qui n'est pas la dernière année (nouvel emménagé pouvant être dans une agence et chez un administrateur de biens) ne sont pas considérés comme des doublons.

Pour les doublons d'origine OLL sans raison sociale renseignée, on prend le premier selon la règle de gestion exposée précédemment. Il peut rester des doublons (exemple double saisie), ils sont traités au mieux au cas par cas.

Les variables non renseignées du logement retenu seront le cas échéant complétées par les informations des doublons éliminés (étage...).

V. Le traitement et la diffusion des données

Une fois constituée, et après les quatre niveaux de contrôle, la base d'un observatoire, dite « prête pour le traitement » fait l'objet d'un redressement.

Ce redressement doit en particulier permettre de tenir compte des taux de sondage et de réponse différents selon les segments de parc, le type de gestion et modes de collecte.

Les redressements sont réalisés à l'aide de la macro CALMAR (CALage sur MARGes) de l'Insee.

V. 1. Les valeurs aberrantes

L'observatoire ou le centre national de traitement n'élimine pas d'observations. Si certaines sont considérées comme aberrantes, elles ont été éliminées lors des contrôles de cohérence et de validation des données.

V. 2. Les effets de grappe

Les grappes sont des accumulations de logements à une même adresse. La collecte n'étant pas exhaustive, sans traitement spécifique, ces logements écraseraient les autres lors du redressement. Le traitement des effets de grappe mis en œuvre vise à réduire la place de ces logements dans l'échantillon tout en les conservant tous.

Les effets de grappe sont détectés ainsi : présence de 4 logements ou plus à la même adresse.

Le traitement des grappes se fait en trois étapes.

1^{er} étape

Une fois détectée, la grappe est décomposée en sous-grappes. Une sous-grappe est un ensemble de logements ayant les mêmes caractéristiques (servant aux marges de pondération, suivantes) :

- même nombre de pièces,
- même type d'habitat (maison, appartement),
- même type de location (vide, meublé),
- même type de gestion,
- même ancienneté du locataire en 2 modalités : moins de 3 ans et 3 ans et plus (à adapter selon l'OLL),
- même adresse donc même zone et même époque de construction.

2^e étape

Pour chaque sous-grappe i , création d'un logement fictif ayant les caractéristiques de sa sous grappe. C'est ce logement fictif qui participera à la pondération finale pour un poids P_i .

3^e étape

Remplacement dans la base pondérée du logement fictif par les n logements de la sous-grappe i avec chacun un poids P_i/n .

V. 3. Les poids initiaux

Les poids initiaux sont calculés à partir de la répartition gestion déléguée / gestion directe. Pour les grappes, les poids initiaux sont calculés uniquement pour les logements fictifs, les logements des grappes ne participant pas directement à la pondération.

V. 4. La pondération – les marges de calage

La technique de calage utilisée est l'option ranking ratio de Calmar. Les poids finaux ne sont pas arrondis.

Les marges de calage sont déterminées à partir des données détail logement du recensement de la population de l'Insee et de Filocom. Ce sont les suivantes :

• **Répartition du parc par type d'habitat, nombre de pièces et zone (RP – Insee)**

avec

type d'habitat

Maisons

Appartements

nombre de pièces

1 pièce

2 pièces

3 pièces

4 pièces

5 pièces et plus

Ce découpage est à adapter à chaque territoire selon la structure du parc : à ramener à 4 pièces et plus si le nombre de 5 pièces et plus est faible pour les appartements, regroupement des maisons de 1 et 2 pièces, voire de 3 pièces...

N.B. : La diffusion des résultats est généralement limitée à 4 pièces et plus dans les observatoires locaux.

• **Répartition du parc selon le type de gestion (Enquête logement 2006 et dires d'expert)**

Gestion déléguée

Gestion directe

Rappel : par convention sont inclus dans la gestion déléguée les investisseurs institutionnels gérant leur parc.

• **Répartition du parc par époque de construction (Filocom)**

avant 1919

1919-1945

1946-1970

1971-1990

1991-2005

après 2005

avec des regroupements éventuels pour les 2 premières et/ou les 2 dernières époques selon le nombre de logements concernés

• **Répartition du parc selon l'ancienneté du locataire dans le logement (Filocom)**

moins d'un an (= mobilité)

1 ou 2 an(s)

3 à 5 ans

6 à 10 ans

plus de 10 ans.

Après pondération, réintégration des logements des grappes avec le poids P_i/n et suppression des i logements fictifs.

V. 5. Le contrôle de la pondération – les ajustements

1. Vérification de la convergence au bout de N itérations ($N < 6$).
2. Contrôle et diagnostic sur les poids obtenus : vérification que le ratio entre poids initial et final est compris entre 0,3 et 3

Sinon, sortie de la liste des logements en dehors de cette fourchette, vérification des variables (données aberrantes), puis vérification que les logements rares sont suffisamment nombreux, sinon, regroupement éventuel de certaines modalités dans les marges et nouvelle pondération.

V. 6. La précision des médianes

La précision des médianes est calculée par la méthode du bootstrap pour les cases élémentaires suivantes : type d'habitat * zone * nombre de pièces * époque de construction.

Cette méthode consiste, pour chaque cellule élémentaire, à tirer aléatoirement plusieurs milliers de sous-échantillons au sein des données collectées, chaque sous-échantillon représentant 80 % de l'échantillon total, à déterminer la médiane de chaque sous-échantillon et à définir la plage de variation de la médiane.

V. 7. Les règles de diffusion des données

Les agrégats statistiques sont diffusables en respectant certaines règles en particulier un seuil minimal d'observations fixé à 50. Tout observatoire agréé doit faire une diffusion annuelle minimale de ses résultats. Des indicateurs de dispersion (quartiles et déciles) ainsi que le nombre d'observations doivent être publiés avec les résultats.

Les données individuelles non anonymisées sont diffusables aux entités prévues dans les textes réglementaires : observatoire local, Anil, chercheurs via un accès sécurisé type CASD.

Les données individuelles anonymisées font l'objet d'une diffusion plus large sous réserve que l'anonymisation soit suffisante pour qu'on ne puisse pas identifier un logement.

Les règles de diffusion et d'anonymisation sont décrites dans une note disponible sur le site du réseau.

ANNEXES

1. Questionnaire
2. Tableau des sources

DOCUMENTS COMPLEMENTAIRES DISPONIBLES SUR L'EXTRANET DU RESEAU

www.observatoiresdesloyers.fr

Champ d'observation
Déclaration à la Cnil
Convention type entre l'observatoire local et le fournisseur de données
Notices d'extraction
Fichier de saisie et de dépôt des données
Guide d'enquête
Liste des types de voie
Règles de diffusion

Annexe 1 – Questionnaire

1. Identifiant logement de l'OLL -----	2. Source 1. Administrateurs de biens 2. Agences immobilières 3. Personnes morales 4. Autres professionnels (notaires...) 5. Bailleurs en gestion directe 6. Locataires en gestion directe 7. Autres sources	3. Mode de collecte 1. Transfert de fichiers 2. Classique (rendez-vous, téléphone, mail...) 3. Internet 4. Données centralisées
4. Date d'enquête -- -- -- Enquête précédente -- -- --	5. Adresse du logement N° ----- B/T/Q -- Type de rue ----- Nom de rue ----- Compl. Adresse ----- CP ----- Commune -----	6. Insee Zone Iris Sec. Cadas. Parcelle ----- -- ----- -- -----
7. Date de référence si autre que le 1 ^{er} janvier -- -- --	8. Si enquête auprès d'un professionnel N° de carte G (ou T) ----- N° de gestion du logement ----- Raison sociale -----	
21. Remarques pour le centre national de traitement -----		

Description de l'immeuble				
31. Type d'habitat 1. individuel 2. collectif --	32. Epoque de construction 1. avt 1919 à défaut 2. 1919-1945 3. avt 1946 4. 1946-1970 5. 1971-1990 6. 1991-2005 7. après 2005 --	33. Année de construction -----	34. Type de propriété 1. mono 2. copropriété --	35. Equipement 1. oui, 2. non Asc. Interphone Vidéophone Gardien Digicode -- -- -- --

Description du logement						
41. Nombre de pièces --	42. Surface ----- m ²	43. Etage --	44. Annexes 1. oui, 2. non Parking, Balcon si inclus dans loyer Jardin -- -- --	45. Nombre de WC SdB -- --	46. Type de chauffage 1. individuel 2. collectif --	47. Type de propriétaire 1. particulier 2. personne morale --

Informations relatives au bail					
50. DPE consommation énergétique lettre --	51. Type de location 1. vide 2. meublé --	52. Type de gestion 1. déléguée 2. directe --	53. Catégorie de location 1. libre 2. conventionnée 3. défiscalisée --	54. Date d'entrée du locataire -----	55. Première location si emménagé depuis le 1^{er} janvier N-1 1. oui, 2. non --
56. Loyer mensuel hors charges à la date d'enquête 1 an avant (si évol) -----, -- € année en cours -----, -- €		57. Loyer mensuel hors charges à la date de référence (1^{er} janvier en général) 1 an avant (si évol) -----, -- € année en cours -----, -- €		58. Provisions mensuelles pour charges à la date d'enquête 1 an avant (si évol) -----, -- € année en cours -----, -- €	

Annexe 2 – Recensement des sources

	OLL :
Professionnels de l'immobilier	
<p>Sur le périmètre de l'observatoire, récupérer la liste des professionnels et indiquer leur nombre.</p> <ul style="list-style-type: none"> - à la préfecture <ul style="list-style-type: none"> nombre cartes G : nombre cartes T : - sur les pages jaunes <ul style="list-style-type: none"> nombre d'agences : nombre d'administrateurs de biens : nombre de notaires : - dans les annuaires <ul style="list-style-type: none"> FNAIM <ul style="list-style-type: none"> nombre d'adhérents : (http://www.fnaim.fr/rechercher-une-agence.html) UNIS <ul style="list-style-type: none"> nombre d'adhérents : (http://www.unis-immo.fr/immo_annuaire_adherents.htm?lang=fr) 	
Administrateurs de biens	
<p>Nombre présents sur le périmètre de l'observatoire (cartes G)*</p> <p>Nombre de lots habitation gérés</p> <p>Nombre participant à la collecte*</p> <p>Nombre de lots habitation gérés</p> <p>Nombre d'enquêtes</p> <p>Mode de collecte (transfert de fichier, RdV...)</p> <p>Variables obligatoires non collectées</p> <p>Variables optionnelles collectées</p>	

<p>Agences immobilières</p> <p>Nombre présentes sur le périmètre de l'observatoire (cartes T)* Nombre de baux d'habitation signés en un an</p> <p>Nombre participant à la collecte* Nombre de baux d'habitation signés en un an</p> <p>Nombre d'enquêtes Mode de collecte (transfert de fichier, RdV...)</p> <p>Variables obligatoires non collectées</p> <p>Variables optionnelles collectées</p>	
<p>Autres professionnels (préciser : notaires...)</p> <p>Nombre présents sur le périmètre de l'observatoire* Nombre de lots d'habitation gérés</p> <p>Nombre participant à la collecte* Nombre de lots habitation gérés</p> <p>Nombre d'enquêtes Mode de collecte (transfert de fichier, RdV...)</p> <p>Variables obligatoires non collectées</p> <p>Variables optionnelles collectées</p>	

Particuliers	
<p>Bailleurs personnes physiques</p> <p>Parc détenu par les bailleurs privés (hors institutionnels) Dont en gestion directe</p> <p>Nombre participant à la collecte Source (consultants Adil...)</p> <p>Nombre d'enquêtes Mode de collecte (téléphone, internet, courrier...)</p> <p>Part de chaque type de gestion (directe, déléguée)</p> <p>Variables obligatoires non collectées</p> <p>Variables optionnelles collectées</p>	
<p>Locataires*</p> <p>Source (pages blanches, consultants Adil, terrain, contacts personnels, autres)</p> <p>Nombre d'enquêtes Mode de collecte (téléphone, internet, courrier...)</p> <p>Part de chaque type de gestion (directe, déléguée)</p> <p>Variables obligatoires non collectées</p> <p>Variables optionnelles collectées</p>	

Autres sources	
CAF Nombre d'allocataires du parc locatif privé du périmètre de l'observatoire* Nombre d'enquêtes Mode de collecte (transfert de fichier, RdV...) Variables obligatoires non collectées Variables optionnelles collectées	
Autre (préciser) Nombre de logements du parc locatif privé concerné Nombre d'enquêtes Mode de collecte (transfert de fichier, RdV...) Variables obligatoires non collectées Variables optionnelles collectées	
Remarques	

* faire un tableau de répartition selon le zonage retenu