

Note sur la restitution des résultats et la diffusion des données des observatoires locaux des loyers

1 – Diffusion de résultats

Règles générales

Pour toute publication de résultats statistiques, les observatoires locaux sont soumis aux règles suivantes fixées par le Comité scientifique :

- tout résultat doit être fondé sur un nombre minimum de 50 observations ;
- tout résultat doit être accompagné de l'indication du nombre d'observations sur lequel il est fondé.

Ces règles s'appliquent autant pour les exploitations effectuées par les observatoires eux-mêmes que pour celles effectuées par le centre national de traitement.

La règle de 50 observations garantit un minimum de précision des résultats, et a fortiori le respect du secret statistique, pour lequel on se contente habituellement de fixer des effectifs minimaux bien moins élevés.

Diffusion minimale

Les observatoires locaux publient des résultats dont le contenu minimal est défini dans une maquette validée par le comité scientifique et respectant les règles définies ci-dessus. Ils doivent présenter, dans le cadre de cette publication minimale, non seulement des caractéristiques de tendance centrale (médianes et/ou moyennes), mais aussi des indicateurs de dispersion des loyers (1^{er} et 3^{ème} quartiles, 1^{er} et 9^{ème} déciles).

Il est en outre recommandé aux observatoires locaux d'appliquer cette dernière règle à l'ensemble de leurs publications.

Pour mémoire, l'annexe décrit les résultats agrégés transmis par le centre national de traitement aux observatoires.

2 – Diffusion des données individuelles

Cette partie reprend, en les développant, des dispositions figurant en annexe de la convention type observatoire-fournisseur de données.

Principes

Données diffusables

Le principe retenu, validé par le comité scientifique, figure dans les conventions observatoires – fournisseurs de données :

« Chaque observatoire local et le centre de traitement national diffusent, aux personnes ou organismes qui lui en font la demande, les données redressées dont ils disposent, à condition que ces données respectent les conditions du secret statistique définies précédemment, c'est-à-dire qu'elles ne permettent pas de repérer les logements. »

Les données diffusables sont donc **des données redressées**, c'est-à-dire des données individuelles collectées, contrôlées, enrichies, anonymisées et traitées par le centre de traitement. Le traitement comprend notamment le calcul d'un coefficient de pondération pour chaque logement.

En revanche, les données brutes restent la propriété de ceux qui les ont fournies. Elles ne peuvent en aucun cas être diffusées. C'est un engagement pris vis-à-vis des fournisseurs, et qui permet de garantir également le respect du secret commercial pour eux.

Obligation de diffusion et non-discrimination

Selon la convention type, les observatoires et le centre de traitement ont l'obligation de diffuser les données redressées à qui en fait la demande, à condition que le demandeur s'engage à respecter des règles d'utilisations définies dans une convention. Aucune discrimination ne peut être faite entre les demandeurs.

Tout refus opposé à une demande doit être motivé. Le cas le plus évident est celui où le demandeur refuserait de s'engager à respecter les conditions d'utilisation, mais ce pourrait être aussi parce que le demandeur, dans le passé, n'aurait pas respecté les règles qu'il avait acceptées.

Respect du secret statistique

Pour garantir le respect du secret statistique, les données individuelles diffusées doivent être anonymisées, ce qui signifie qu'il doit être impossible d'identifier un logement, que ce soit directement ou par recoupement.

En pratique, l'anonymisation peut difficilement être absolue, lorsque l'on souhaite conserver une indication de localisation des logements aussi précise que possible. Savoir dans quelle zone est situé un logement peut en effet, dans certains cas, permettre de l'identifier, par exemple s'il s'agit d'une maison située dans une zone où les maisons sont rares et si l'on connaît ses autres caractéristiques (taille, époque de construction, date de location et loyer). Or une localisation trop vague peut ôter tout intérêt à l'exploitation des données. La procédure d'anonymisation détaillée ci-après est donc le résultat d'un compromis.

Toute diffusion de données individuelles doit être assortie d'un avertissement sur les précautions d'utilisation des données.

Procédure d'anonymisation des données

Toutes les informations concernant l'enquête (gestionnaire de biens en général), demandées pour les contrôles, seront éliminées. Il s'agit essentiellement des variables suivantes : source, mode de collecte, N° de carte de gestion et raison sociale du gestionnaire, numéro de gestion du logement.

Cette anonymisation est toutefois insuffisante en particulier pour les logements rares : maisons, grands logements ou situés dans des zones peu denses en parc locatif privé.

L'information sur la taille (nombre de pièces et surface) est conservée, l'anonymisation est réalisée sur la localisation (adresse) avec une progressivité fonction de la taille du parc locatif de la commune.

Les données de localisation disponibles pour un logement sont : l'adresse complète, l'Iris (essentiellement pour les zonages infra communaux), le zonage défini par l'observatoire et, à terme, la section et la parcelle cadastrale.

L'information sur la localisation conservée dans le fichier anonymisé dépend de l'importance du parc locatif privé de la commune pour chaque type de logement (maison ou appartement).

Quelle que soit la commune, l'adresse complète et la parcelle sont systématiquement supprimées et la zone d'observation des loyers est systématiquement conservée.

- dans les *communes de moins de 1 000 logements locatifs privés*, le code postal, le code Insee, l'iris et la section cadastrale sont supprimés) ;
NB : si dans une zone une seule commune est dans ce cas, on applique la même règle à la 2ème plus petite commune de la zone.
- dans les *communes de 1 000 logements locatifs privés ou plus*, on conserve le code postal et le code Insee. Dans ce cas, si le logement se situe dans un iris (ou une section cadastrale) de plus de 1000 unités, l'IRIS (ou la section) sont conservés.

Cette règle conduit à un degré de précision moindre sur la localisation pour les maisons (peu nombreuses en général) que pour les appartements.

Schéma récapitulatif du traitement de la localisation

Information de localisation figurant dans la base pondérée	Taille du parc locatif privé de la commune par type de logement		
	Inférieur à 1000	Supérieur à 1000	
		Taille du parc locatif de l'Iris (ou section)	
		<1000	≥1000
Iris (ou section)	non	non	oui
Code postal, libellé commune, code Insee	non	oui	oui
Zone	oui	oui	oui

Suivi dans le temps

Le parc locatif privé de référence est, pour l'expérimentation de 2013, celui du Recensement de la population 2009, en distinguant maisons et appartements.

Il est proposé de maintenir ce classement initial des communes sur une période donnée.

Contenu du fichier

Variables issues de la collecte (limitées aux variables obligatoires)

Année d'enquête

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

Variable de localisation selon le résultat de l'anonymisation : Code Insee, zone (d'observation du loyer), iris si connu

Type d'habitat (maison, appartement)

Epoque de construction : avant 1946, 1946-1970, 1971-1990, après 1990

Nombre de pièces : 1, 2, 3, 4 et plus pour les appartements, 1-2, 3, 4, 5 et plus pour les maisons

Surface

Type de location (loi 89 vide, meublé)

Loyer mensuel hors charges à la date d'enquête

Variables complémentaires ajoutées par le centre national de traitement

Code OLL

Poids final

Tranches de surface (< 30 m², 30-50, 50-80, 80-120, ≥ 120)

Zonage en 2 modalités (ville centre, périphérie)

Ancienneté du locataire en 2 (mobile, stable)

Réalisation du traitement d'anonymisation

Outre les traitements décrits ci-dessus, les données anonymisées sont assorties de pondérations calculées en fonction des zonages retenus. Le traitement d'anonymisation des données individuelles est effectué par le centre national de traitement, qui met les données anonymisées à la disposition des observatoires locaux en vue de leur diffusion.

Diffuseur

Les observatoires locaux pour le périmètre qui les concerne et le centre de traitement pour l'ensemble du réseau, sous réserve d'en informer préalablement les observatoires locaux concernés, sont habilités à diffuser les données anonymisées.

Convention de mise à disposition des données

Le bénéficiaire doit s'engager à ne céder les données transmises à quiconque, à ne les utiliser qu'à des fins statistiques et à respecter le secret statistique dans toute publication de résultats.

A cet effet, les observatoires et le centre de traitement doivent s'engager à conclure avec le bénéficiaire une convention conforme à un modèle qui devra leur être fourni et qui devrait notamment contenir une règle de nombre minimum d'observations pour tout résultat publié afin de garantir le respect du secret statistique et la précision des résultats.

Diffusion de données complètes aux chercheurs

A l'instar de ce qui se pratique pour d'autres bases de données, il est souhaitable de permettre aux chercheurs d'accéder aux données complètes, c'est-à-dire comprenant l'adresse précise des logements. Des contacts sont en cours avec le Centre d'accès sécurisé distant aux données (CASD), qui assure cette fonction, pour examiner la faisabilité et les conditions de cette option.

Tarifification

On peut à cet égard distinguer le cas où le fichier diffusé est « standard » et celui où sa production implique un traitement spécifique par l'observatoire local ou par le Centre de traitement national. Dans le second cas, un travail est nécessaire pour l'observatoire ou le centre de traitement, qui représente un coût.

ANNEXE

Résultats agrégés restitués par le centre national de traitement aux observatoires locaux

Des tableaux standards minimaux seront produits par l'Olap pour les observatoires.

NB : Le détail sera fourni même si le nombre de logements est inférieur au seuil retenu pour la diffusion des données, à charge pour l'observatoire de supprimer ou regrouper les cases d'effectif inférieur au seuil.

Contenu des tableaux

Deux séries de tableaux sont prévues : avec médianes et avec moyennes.

1. Médianes (pour l'encadrement des loyers en particulier)

Nombre d'observations

Loyer médian au m²

1^{er} et 3^{ème} quartiles du loyer au m²

Précision (par bootstrap)

2. Moyennes

Nombre d'observations

Nombre de logements de la cellule (pondération)

Surface moyenne

Loyer moyen

Loyer moyen au m²

Croisements prévus pour les moyennes

Tous les croisements de 2 ou 3 des variables suivantes :

- nombre de pièces
- zonage d'observation des loyers ou zonage complémentaire
- type d'habitat
- durée d'occupation
- époque de construction

Exemples de croisements

Nombre de pièces * zone

Type d'habitat * zone

Nombre de pièces * durée d'occupation

Type d'habitat * durée d'occupation

Type d'habitat * nombre de pièces * zone

Nombre de pièces * époque de construction

Type d'habitat * époque de construction

Nombre de pièces * époque de construction * durée d'occupation

Ces variables étant caractérisées par les modalités suivantes :

Variable	Modalités	Remarques
nombre de pièces	1 pièce / 2 pièces / 3 pièces / 4 pièces et plus	si nécessaire (effectif insuffisant) regroupement des 1 et 2 pièces pour les maisons, et si possible les 5 pièces et plus.
type d'habitat	Maisons / Appartements	
époque de construction	avant 1919 / 1919-1945 / 1946-1970 / 1971-1990 / 1991-2005 / après 2005 et avant 1946, 1946-1970, 1971-1990, après 1990	si le nombre de logements anciens est peu élevé : avant 1946 / 1946-1970 / 1971-1990 / 1991-2005 / après 2005
ancienneté du locataire	Moins d'un an (mobilité) / 1 à 2 ans / 3 à 5 ans / 6 à 9 ans / 10 ans et plus	regroupement des deux dernières tranches si nécessaire.
zonage d'observation des loyers	zones définies pour l'expérimentation	regroupements si le nombre d'observations s'avère nettement insuffisant ;
zonage complémentaire	Ville centre / périphérie	