

Ouverture des données publiques dans le champ du logement

Rapport

Synthèse des débats du groupe de travail du
Conseil national de l'Habitat

Table des matières

Présentation du rapport.....	5
Introduction	7
PARTIE I - Contexte et enjeux	9
A) Définition et cadre juridique de l'Open Data ou « ouverture des données publiques »	9
1) Définition.....	9
2) Le cadre juridique de l'ouverture des données publiques	9
3) La directive « Public Sector Information » (PSI)	10
B) Les enjeux de la politique publique en faveur de l'« Open Data »	11
1) La feuille de route numérique du Gouvernement	11
2) Le principe de gratuité de la mise à disposition des données publiques	12
3) La Licence Ouverte/ Open Licence	14
4) Nomination d'un administrateur général des données	14
5) Une logique d'organisation d'ensemble et d'ouverture du système d'information de l'État	15
C) Le débat thématique sur les données logement	17
1) Présentation du cadre du groupe de travail.....	17
2) Présentation des données publiques dans le champ du logement.....	18
3) L'ouverture et la mise à disposition des données logement : un sujet à la fois connexe et lié avec la problématique des statistiques dans ce champ	19
PARTIE II - Les enjeux de l'ouverture des données publiques et les attentes des acteurs du logement	22
A) L'ouverture des données publiques dans le champ du logement : un enjeu démocratique, de connaissance, économique et d'amélioration de l'action publique	22
1) Un enjeu démocratique.....	22
2) Un enjeu de connaissances du secteur.....	22
3) Un enjeu d'amélioration de l'action publique	23
4) Des enjeux économiques	23
B) Les attentes exprimées par les acteurs	24
1) En termes de données souhaitées	24
2) En termes de format et de granularité de mise à disposition de données	26

3) En termes d'ouverture de nouvelles données publiques.....	29
4) En termes d'accessibilité et de visibilité des sources de données mises à disposition	30
PARTIE III – Propositions de pistes de travail	32
A) Améliorer le format de mise à disposition (formats ouverts, interopérables)	32
B) Accroître la granularité des données mises à disposition :	33
C) Prévoir le recensement des données dans le cadre des délégations de compétences ou transfert de compétences	33
D) Mettre à disposition de nouveaux jeux de données en matière de logement.....	34
E) Penser la mise à disposition des données lors de la mise en place du système d'information..	35
F) S'appuyer sur la mission Etalab et l'administrateur général des données pour mettre en œuvre le processus de mise à disposition des données logement.....	36
G) Rendre visible les données existantes et projets de réutilisation des producteurs à travers data.gouv.fr.....	36
Conclusion.....	38
ANNEXES	41
Annexe 1 – Lettre du Président du Conseil National de l'Habitat	42
Annexe 2 – Liste des participants au groupe de travail	44
Annexe 3 – Liste des auditions et entretiens	47
Annexe 4 – Cartographie des données	48

Présentation du rapport

La première partie du rapport dresse le cadre général dans lequel se déroule le débat thématique sur l'ouverture des données publiques (« Open Data») dans le champ du logement.

Après avoir rappelé en quoi consistent l'ouverture et la mise à disposition des données publiques, le rapport en présente le cadre juridique. Il expose la politique d'ouverture des données publiques de l'État qui s'inscrit dans celle, plus large, portant sur l'ensemble de son système d'information et dans laquelle s'inscrit le débat thématique sur les données relatives au logement. Il précise enfin l'organisation du groupe de travail qui a mené ce débat et son articulation avec les autres travaux menés sur les données logement, sur lesquels il s'est appuyé et dont il est complémentaire.

La deuxième partie présente les attentes des acteurs du logement en matière d'ouverture et de mise à disposition de données publiques. En effet, ces données intéressent différents acteurs à des fins diverses telles que le développement de la connaissance du secteur, l'amélioration des politiques publiques, la transparence de l'action publique ou encore, le développement de nouveaux services.

Ces attentes portent à la fois sur les données mises à disposition et sur leur format et leur granularité (ou niveau territorial), qui sont des questions majeures pour une réutilisation efficace et facile de ces données.

La troisième partie présente les différentes propositions du groupe de travail.

Elles portent tout d'abord sur l'amélioration des formats de mise à disposition et l'optimisation de la granularité géographique des données. Elles dressent également des pistes d'amélioration en termes de méthodes pour favoriser cette mise à disposition, notamment en la prévoyant le plus en amont possible de la création des systèmes d'information.

De plus, le rapport propose de mettre à disposition de nouveaux jeux de données en matière de logement et invite les producteurs à donner davantage de visibilité à celles ouvertes ainsi qu'à leurs réutilisations.

Enfin, le rapport inscrit ces préconisations au regard des nouvelles missions confiées à l'administrateur général des données.

Ce rapport se veut ainsi une synthèse des échanges entre les producteurs et les réutilisateurs de données, dans le but de contribuer à l'ouverture des données publiques en matière de logement. Il constitue un point d'étape, dont l'un des principaux intérêts a été de créer les conditions de ces échanges, utiles en eux-mêmes pour renforcer l'appropriation des enjeux de l'ouverture des données publiques, comme la prise en compte des contraintes et questions soulevées.

Introduction

Dans le cadre de la politique du Gouvernement en faveur de l'ouverture des données publiques, il a été décidé de lancer des débats sur quatre thématiques prioritaires parmi lesquelles figure le logement, en y associant les acteurs intéressés.

Le sujet du logement, qui relève de l'essentiel pour chacun, est aussi une question économique et occupe une place importante dans l'action publique. La question de l'ouverture des données publiques sur ce sujet présente donc des enjeux importants.

Le Conseil national de l'habitat (CNH), dont la vocation est d'être un lieu de débat et de concertation sur le logement, s'est vu confier le rôle d'animation des échanges, en lien avec la mission Etalab au sein du Secrétariat général pour la modernisation de l'action publique (SGMAP) et avec l'objectif d'élargir le débat à une diversité d'acteurs¹.

C'est donc un groupe composé à la fois de membres du CNH et d'acteurs de la « communauté de l'Open Data » qui s'est réuni tout au long du premier semestre de l'année 2014. Il a mené un travail d'auditions des principaux producteurs dans le champ du logement et a fait part de ses attentes et de pistes d'évolution.

Ce rapport vise à présenter les principaux enjeux de l'ouverture des données publiques en matière de logement et à faire des propositions d'amélioration quant à leur mise à disposition.

Si la question des données en matière de logement a fait l'objet de nombreuses réflexions², notamment partant du constat de déficits de données à la fois brutes et statistiques, ce rapport s'y intéresse dans une démarche nouvelle et sous l'angle bien spécifique de leur réutilisation. Il ne cherche pas à analyser la qualité des données produites, ni à réfléchir sur les besoins en termes de systèmes d'information. Il se concentre en outre sur les données publiques qui constituent seulement une partie de celles qui existent en matière de logement.

Ces réflexions s'inscrivent de plus dans un contexte marqué par une évolution importante des systèmes d'information dans le champ du logement, avec la construction en cours de nombreux outils, et par des avancées déjà importantes en matière de mises à disposition des données publiques par les différents acteurs concernés.

Ainsi, en synthétisant les échanges qui se sont tenus dans le cadre des réunions du groupe de travail, ce rapport vise à apporter une contribution à la réflexion pour renforcer l'ouverture des données publiques en matière de logement. Après avoir présenté les enjeux de connaissances, démocratiques et économiques, il fait part des attentes exprimées par une diversité d'acteurs en termes de mise à disposition de données existantes dans le champ du logement. Sans prétendre à l'exhaustivité, ni passer

¹ Annexe 1 : Lettre de Michel PIRON, président du CNH, aux membres du CNH

² Cf en particulier : Inspection générale de l'Insee N°1.7.25 - Conseil général de l'Environnement et du Développement Durable N°009075-02 Rapport sur l'organisation du service statistique dans le domaine du logement -

<http://www.ladocumentationfrancaise.fr/var/storage/rapports-publics/144000532/0000.pdf> publié en septembre 2014 ;

Conseil national de l'information statistique - Rapport du groupe de travail « sur le logement et la construction » - Mars 2010

en revue chacune des sources de données, il dresse des propositions d'améliorations qui se veulent à la fois concrètes et précises, mais surtout avec un intérêt méthodologique et d'exemplarité. Il s'agit en effet de réfléchir à une logique d'inscription du champ du logement dans une démarche d'ouverture des données publiques.

PARTIE I - Contexte et enjeux

A) Définition et cadre juridique de l'Open Data ou « ouverture des données publiques »

1) Définition

L'Open Data désigne l'ouverture et le partage de données par leur mise en ligne dans des formats ouverts, en autorisant la réutilisation libre et gratuite par toute personne.

Les données ouvertes peuvent être utilisées par une organisation publique ou privée, à des fins différentes de celles ayant motivé leur création.

Les modalités de mise à disposition doivent respecter le principe de la plus grande liberté de réutilisation des données publiques, en évitant le plus possible les contraintes d'ordre technique, financier, juridique ou autre.

La donnée occupe désormais une place centrale dans l'économie et la société. L'Open Data s'inscrit ainsi dans un mouvement plus vaste de « fluidification » qui appelle une gouvernance de la donnée et une architecture des systèmes d'information concernant toutes les données, qu'elles aient ou non vocation à être ouvertes à tous.

2) Le cadre juridique de l'ouverture des données publiques

a) *La loi n° 78-753 du 17 juillet 1978 dite «loi CADA »*

L'ouverture des données publiques s'inscrit dans le cadre de la loi n° 78-753 du 17 juillet 1978 qui porte, dans son premier titre, sur la liberté d'accès aux documents administratifs et de la réutilisation des informations publiques. Elle reconnaît, à toute personne, le droit d'obtenir communication des documents détenus dans le cadre de sa mission de service public par une administration, quels que soient leur forme ou leurs supports. La loi dite "CADA" crée notamment la Commission d'accès aux documents administratifs. Cette autorité administrative indépendante a pour objectif de faciliter et contrôler l'accès des particuliers aux documents administratifs.

Le droit d'accès et de réutilisation des données publiques concerne les textes, mémorandums, documents et les tableaux ou statistiques produits par l'administration dans le cadre d'une mission de service public.

Ce droit s'exerce à l'égard de toutes les personnes publiques (l'État, les collectivités territoriales et leurs établissements publics) ainsi qu'à l'égard des organismes privés chargés d'une mission de service public.

La loi prévoit toutefois des restrictions au droit d'accès (articles 6 et 7), nécessaires pour préserver divers secrets, tel par exemple celui qui garantit dans l'intérêt des personnes le respect de la vie privée³, le

³ Le cadre juridique d'utilisation des données à caractère personnel et de protection de la vie privée est défini en particulier par la loi du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés. Au niveau du droit

secret industriel et commercial, ou encore celui qui garantit le secret fiscal⁴, ou que pour des raisons de sécurité. Ainsi, l'accès à certaines informations, par exemple les dossiers médicaux, les listes électorales ou les informations environnementales, obéit à des règles particulières, souvent plus restrictives que le régime général.

Modifiée par l'ordonnance du 6 juin 2005 et celle du 29 avril 2009, la loi pose un principe de libre réutilisation des informations ou données contenues dans un document administratif communicable, à l'exception des « documents produits ou reçus par les administrations mentionnées à l'article 1^{er} de la loi CADA dans l'exercice d'une mission de service public à caractère industriel ou commercial » et des documents « sur lesquels des tiers détiennent des droits de propriété intellectuelle » (article 10 de la loi CADA). Cependant, la réutilisation des données est possible « à condition que ces dernières ne soient pas altérées, que leur sens ne soit pas dénaturé et que leurs sources et la date de leur dernière mise à jour soient mentionnées ». De plus, « les informations publiques comportant des données à caractère personnel peuvent faire l'objet d'une réutilisation soit lorsque la personne intéressée y a consenti, soit si l'autorité détentrice est en mesure de les rendre anonymes ou, à défaut d'anonymisation, si une disposition législative ou réglementaire le permet » (articles 12 et 13 de la loi CADA modifiés)⁵.

Enfin, la loi CADA prévoit que la réutilisation d'informations publiques peut donner lieu au versement de redevances, pour tenir compte des coûts de mise à disposition des informations, notamment, le cas échéant, du coût d'un traitement permettant de les rendre anonymes. De plus, lorsque la réutilisation n'est possible qu'après anonymisation des données à caractère personnel, le détenteur de données n'y procède que si cela n'entraîne pas d'efforts disproportionnés (décret n°2005-1755 du 30 décembre 2005).

Le cadre juridique vise donc à permettre à un accès aux données, tout en conciliant ce droit avec la protection d'autres droits.

3) La directive « Public Sector Information » (PSI)

Au niveau européen, le cadre juridique concernant la réutilisation des informations du secteur public est fixé par la directive 2013/37/UE « Public Sector Information » publiée le 26 juin 2013 et modifiant la directive 2003/98/CE (« Directive PSI »).

La Commission européenne a engagé le processus de révision de la directive PSI en décembre 2011 en proposant un projet au Conseil de l'Union européenne et au Parlement européen. Dans le cadre de la procédure de co-décision, après 18 mois de travail et d'échanges, le Conseil et le Parlement ont adopté un nouveau texte pour la directive PSI qui met l'accent sur les avancées récentes de l'Open Data à travers

européen, la directive 95/46/CE traite de la protection des personnes physiques à l'égard des traitements de données à caractère personnel et à la libre circulation de ces données.

⁴Les données nominatives afférentes aux contribuables recueillies par l'administration dans l'exercice de ses missions fiscales sont couvertes par le secret professionnel, conformément aux dispositions de l'article L. 103 du livre des procédures fiscales (LPF). Le secret s'applique à l'ensemble des éléments détenus par l'administration fiscale et notamment aux éléments relatifs aux revenus et au patrimoine des particuliers et aux bénéficiaires et au chiffre d'affaires des entreprises.

⁵ En matière d'informations statistiques, le respect du secret statistique est défini par la loi du 7 juin 1951 et l'article 285 du Traité instituant la communauté européenne, notamment dans un but de protection de la vie privée

l'Europe des 27 et qui propose un cadre juridique harmonisé, propice au développement de l'ouverture et de la réutilisation des données publiques en France comme dans les autres pays européens.

Lors du CIMAP du 17 juillet 2013, le Premier ministre a rappelé dans la décision 17 que « *Le SGMAP (mission Etalab) coordonnera, avec l'ensemble des administrations concernées, la transposition dans le droit national de la directive européenne du 26 juin 2013 concernant la réutilisation des informations du secteur public.* »

Ce travail devra être finalisé en juillet 2015 et permettra notamment d'adapter le dispositif juridique et réglementaire français d'encadrement de l'ouverture des données publiques.

B) Les enjeux de la politique publique en faveur de l'« Open Data »

L'Open Data, ou l'ouverture et le partage des données publiques, représente trois enjeux :

- Un enjeu démocratique : l'ouverture des données publiques doit permettre à l'État de s'organiser afin de rendre des comptes, d'ouvrir le dialogue et transformer sa relation avec les citoyens, notamment en renforçant la transparence de l'action publique et en permettant aux citoyens de construire leurs propres points de vue à partir des données mises à disposition ;
- Un enjeu de modernisation de l'action publique : l'ouverture des données publiques contribue à renforcer l'efficacité de l'action publique, en simplifiant le fonctionnement des administrations et de leurs procédures et en permettant à la puissance publique de mobiliser toutes les ressources de la culture de la donnée : décloisonnement des services, décisions fondées sur les données, pilotage des politiques publiques ;
- Un enjeu économique : l'ouverture des données publiques vise à stimuler l'innovation et la recherche et encourager la création de nouveaux services, utiles aux citoyens et sources d'activité économique.

1) La feuille de route numérique du Gouvernement

La politique engagée par l'État en faveur de l'Open Data est un levier important de la modernisation de l'action publique. La feuille de route du Gouvernement français sur l'ouverture et le partage des données publiques a été publiée en février 2013. Les Comités interministériels pour la modernisation de l'action publique (CIMAP) suivants ont complété cette feuille de route par plusieurs décisions concernant l'Open Data. Le principe de gratuité de la réutilisation des données publiques (décisions 24, 25 et 26 du CIMAP du 18 décembre 2013) a notamment été réaffirmé et étendu, en concertation, aux collectivités, aux délégataires de service public, à la sphère sociale et aux autorités administratives indépendantes.

Enfin, un Vademecum pour l'ouverture des données publiques⁶, a été établi pour faire part aux agents publics des modalités d'ouverture des données. Il a été adressé à toutes les administrations par une circulaire du Premier ministre⁷.

Ainsi, la politique d'ouverture des données publiques se donne pour objectif, au-delà de l'obligation légale de l'administration en termes de droit d'accès, de mettre à disposition des données de manière volontaire, gratuite, et dans des formats faciles à exploiter, afin d'en favoriser la réutilisation pour différents usages.

2) Le principe de gratuité de la mise à disposition des données publiques

Dans le cadre de la politique d'ouverture des données publiques et à la suite du travail sur les redevances initié à l'occasion du rapport Trojette⁸, le CIMAP du 18 décembre 2013 a réaffirmé le principe de gratuité de la réutilisation des données publiques tout en précisant sa doctrine en matière d'exceptions à ce principe⁹. Il a été décidé de ne plus autoriser la création de nouvelle redevance. Il a également été décidé de supprimer des redevances perçues actuellement par les administrations pour la réutilisation de leurs données. Par ailleurs, le Gouvernement a chargé le Secrétariat général pour la modernisation de l'action publique (SGMAP) d'accompagner les opérateurs dont la mission est de produire des données vers de nouveaux modèles économiques compatibles avec le redressement des finances publiques et susceptibles de compenser la suppression totale des redevances (cf. encadré).

Décision n°24:

« Le Gouvernement réaffirme le principe de gratuité de la réutilisation des données publiques et décide de ne plus autoriser la création de nouvelle redevance. »

Décision n°25:

« Sur la base du rapport remis par M. Trojette relatif au modèle économique des redevances de réutilisation des données publiques et après examen des différentes catégories de données publiques soumises à redevance, le Gouvernement décide de supprimer plusieurs redevances. Les redevances supprimées sont celles perçues sur :

- les données relatives aux dotations globales de fonctionnement versées aux collectivités locales par la direction générale des collectivités locales (ministère de l'Intérieur) ;*
- les données comptables brutes gérées par la direction générale des finances publiques – balance des comptes définitifs des collectivités et établissements publics locaux, recensées à compter de l'arrêté définitif des comptes 2013. Les fiches de situation financière des collectivités locales restent consultables sur le site collectivites-locales.gouv.fr.*

⁶ <http://www.modernisation.gouv.fr/sites/default/files/fichiers-attaches/vademecum-ouverture.pdf>

⁷ http://www.modernisation.gouv.fr/sites/default/files/fichiers-attaches/circulaire_ndeg_5677-sg_du_17_septembre_2013.pdf

⁸ Le rapport Trojette relatif à l'évaluation des modèles économiques des redevances portant sur l'utilisation des données publiques est accessible sur le lien suivant : <http://www.modernisation.gouv.fr/laction-publique-se-transforme/en-ouvrant-les-donnees-publiques/remise-du-rapport-trojette-sur-louverture-des-donnees-publiques>

⁹ <http://www.modernisation.gouv.fr/le-sgmap/le-cimap/le-cimap-du-18-decembre-2013>

- le parcellaire agricole de l'Agence de services et de paiement (ministère de l'agriculture, de l'agroalimentaire et de la forêt) ;
 - les données cadastrales de l'Institut national de l'origine et de la qualité (ministère de l'agriculture, de l'agroalimentaire et de la forêt) ;
 - les données brutes diffusées par l'Institut français du cheval et de l'équitation ;
 - les données relatives aux adresses postales des établissements scolaires (ministère de l'éducation nationale) ;
 - les données de l'Office national d'information sur les enseignements et les professions – ONISEP (ministère de l'éducation nationale) ;
 - les données de la CADA ;
 - les redevances sur la réutilisation professionnelle des informations suivantes diffusées par la DILA : ASSOCIATIONS, DEPOTS DES COMPTES DES ASSOCIATIONS, JORF, LEGI, KALI, JADE, CONSTIT, CASS-INCA, CAPP, CNIL, CIRCULAIRES.
- Par ailleurs, le ministère de l'économie et des finances fournira en Open Data, sur une base hebdomadaire, la base des prix des carburants. »

Décision n°26 :

« Le Gouvernement précise sa doctrine en matière d'exceptions au principe de gratuité. Aucune redevance ne saurait être exigée sur des données résultant des missions de service public des administrations générales. Il est demandé aux administrations d'accroître leurs efforts en matière de publication de données leur permettant de mener à bien leurs missions, de motiver leurs décisions, de mesurer leurs activités ou leur bilan. Le site data.gouv.fr est en mesure d'accueillir ces données ; le respect du secret statistique et donc l'anonymisation des informations pouvant éventuellement permettre d'identifier des personnes est un principe central du droit français. Certaines redevances ont été instaurées pour financer cette tâche, qui est indispensable avant ouverture de la donnée. Le Gouvernement demande au SGMAP d'apporter tout le soutien possible aux services publics concernés, afin d'identifier les moyens de réduire ces coûts sans dégrader la qualité de cette anonymisation ;

Enfin, le Gouvernement réaffirme que les opérateurs dont la mission est de produire des données doivent rechercher des modèles économiques leur permettant de faire face à un paysage économique en profonde reconstitution. Conformément aux conclusions du rapport Trojette, il leur demande d'engager, dans les meilleurs délais, avec l'appui du SGMAP et du ministère du Budget, une réflexion sur les évolutions de leurs modèles économiques. Il leur demande de rechercher des modèles stimulant l'innovation autour de leurs données, favorables aux entrepreneurs innovants et soutenables à l'heure de l'économie numérique, de la production de nombreuses données par les citoyens eux-mêmes et des stratégies de plateformes. Afin de mener cette transition, qui vise à maintenir et améliorer la qualité des informations publiques, à garantir la pérennité du service public et à développer des stratégies numériques pérennes, efficaces et innovantes, le Gouvernement demande au SGMAP d'accompagner au mieux ces opérateurs, tel l'IGN (Institut géographique national), le SHOM (Service hydrographique et océanographique de la marine) ou Météo France, en lien avec les ministères de tutelle pour définir

avec eux des trajectoires d'investissement dans l'innovation compatibles avec le redressement des finances publiques. »

3) La Licence Ouverte/ Open Licence¹⁰

Dans le cadre de la politique de l'État en faveur de l'ouverture des données publiques, Etalab a conçu la « Licence Ouverte / Open Licence ». Cette licence, élaborée en concertation avec l'ensemble des acteurs concernés, facilite et encourage la réutilisation des données publiques mises à disposition gratuitement. Depuis novembre 2011, la « Licence Ouverte / Open Licence » s'applique à l'ensemble des réutilisations libres et gratuites de données publiques issues des administrations de l'État et de ses établissements publics administratifs, à l'exclusion de tout autre licence.

La « Licence Ouverte / Open Licence » :

- est libre et gratuite ;
- offre une grande liberté de réutilisation des informations ;
- apporte la sécurité juridique nécessaire aux producteurs et aux réutilisateurs des données publiques ;
- promeut la réutilisation la plus large en autorisant la reproduction, la redistribution, l'adaptation et l'exploitation commerciale des données ;
- s'inscrit dans un contexte international en étant compatible avec les standards des licences Open Data développées à l'étranger et notamment celles du gouvernement britannique (Open Government Licence) ainsi que les autres standards internationaux (ODC-BY, CC-BY 2.0);
- exige une forte transparence de la donnée et de qualité des sources en rendant obligatoire la mention de la paternité.

De plus, elle constitue une opportunité de mutualisation pour les autres données publiques en mettant en place un standard réutilisable par les collectivités territoriales qui souhaiteraient se lancer dans l'ouverture des données publiques.

4) Nomination d'un administrateur général des données

A été institué par un décret du 16 septembre 2014 un administrateur général des données (chief data officer). Autorisé à connaître les données détenues par l'administration de l'État et ses opérateurs, cet administrateur a pour mission :

- d'organiser une meilleure circulation des données dans l'économie comme au sein de l'administration - dans le respect de la vie privée et des différents secrets légaux ;
- de veiller à la production ou à l'acquisition de données essentielles ;
- de lancer des expérimentations pour éclairer la décision publique ;

¹⁰ Le texte détaillé de la licence ouverte/ open licence est téléchargeable gratuitement sur la plateforme data.gouv.fr : https://wiki.data.gouv.fr/wiki/Licence_Ouverte/_Open_Licence

- de diffuser les outils, méthodes et culture de la donnée au sein des administrations et au service de leurs objectifs respectifs.

Henri Verdier, par ailleurs directeur de la mission Etalab en charge de la mise en œuvre de la politique d'Open Data au sein du Secrétariat général à la modernisation de l'action publique, a été nommé administrateur général des données.

La France est le premier État européen à se doter de cette fonction au niveau national. Cette fonction de Chief Data Officer avait auparavant été réservée à des villes (New York, San Francisco, Chicago, Philadelphie, Baltimore...) ou des entreprises (Yahoo, Ogilvy, CitiGroup...) pionnières.

À l'heure où l'utilisation des données est au cœur de la révolution numérique, les informations produites et détenues par l'État dans le cadre de ses missions appellent l'utilisation des data sciences¹¹ pour permettre de mieux informer sur les politiques publiques. En parallèle, la politique d'Open Data conduit à ouvrir une partie toujours plus grande de ces informations dans des formats librement réutilisables par tous.

Placé sous l'autorité du Premier ministre au sein du Secrétariat général pour la modernisation de l'action publique (SGMAP), l'Administrateur général des données est chargé de faire levier sur tout le potentiel des données publiques pour améliorer le service public.

Amélioration de la qualité des données produites par l'État, fluidification de la circulation au sein de l'administration comme à destination des chercheurs, entreprises et citoyens, et production de données clefs comptent parmi ses principaux chantiers.

Parce que la transformation numérique touche les pratiques quotidiennes de l'administration, il a pour mission d'y stimuler la diffusion des nouvelles stratégies d'action fondées sur la donnée. Enfin, il est autorisé à conduire des expérimentations sur l'utilisation des données pour renforcer l'efficacité des politiques publiques, contribuer à la bonne gestion des deniers publics et améliorer la qualité des services rendus aux usagers.

5) Une logique d'organisation d'ensemble et d'ouverture du système d'information de l'État

Le 17 septembre 2014, le Secrétaire d'État à la Réforme de l'État et à la Simplification a annoncé en Conseil des ministres la volonté d'accélérer la transformation numérique de l'État¹².

¹¹ La science des données porte sur l'extraction de connaissances à partir de données permettant de produire des « produits de données ». Ce passage des données aux connaissances nécessite une approche interdisciplinaire. La « science des données » s'appuie sur le traitement statistique de l'information (statistiques mathématiques, statistiques numériques, apprentissage statistique ou machine learning). Les techniques de la science de données touchent la recherche dans de nombreux domaines, dont les sciences biologiques, l'informatique médicale, les soins de santé, les sciences sociales et les sciences humaines.

¹² Compte rendu du Conseil des ministres du 17 septembre 2014 : <http://www.gouvernement.fr/conseil-des-ministres/2014-09-17/le-numerique-pour-transformer-l-etat>

Les infrastructures informatiques, les réseaux de communication, les applications partagées et les systèmes d'information transverses seront dorénavant placés sous la responsabilité des services du Premier ministre.

La finalité est de disposer d'une vision et d'une trajectoire cohérente d'ensemble qui tiennent compte de l'histoire et de l'actualité de chacun des ministères. Il y a très peu de ministères qui offrent seuls un service aux citoyens et aux entreprises. La plupart d'entre eux ont besoin d'autres administrations pour délivrer un service. Or, beaucoup de services numériques sont construits sur l'architecture du décret de 1986 qui prévoyait que chaque ministre était responsable de son SI et qui faisait de l'État une juxtaposition de systèmes d'information.

Le principal point du décret du 1^{er} août 2014¹³ consiste à donner les moyens à la Direction interministérielle des systèmes d'information et de communication (DISIC) d'organiser l'architecture d'ensemble du SI.

Elle s'oriente vers une architecture de plateforme ouverte. L'ouverture servira par exemple à délivrer des services en collaboration avec des collectivités territoriales et des opérateurs de l'État. L'État pourra également s'appuyer par le co-design, sur d'autres acteurs (secteur associatif, startups,...) pour concevoir des applications composites qui assembleront, par le biais d'API¹⁴ sécurisées, des données détenues par différents acteurs. Il ne s'agira pas de transmettre les données et les traitements à gérer, mais de créer des « ponts ».

L'ouverture et la mise à disposition des données publiques participent cependant pleinement à cette démarche de décloisonnement.

¹³ <http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000029337021>

¹⁴ Une **API** (*Application Programmable Interface*, traduisez « *interface de programmation* » ou « *interface pour l'accès programmé aux applications* ») est un ensemble de fonctions permettant d'accéder aux services d'une application, par l'intermédiaire d'un langage de programmation. Grâce aux API, un développeur n'a donc pas à se soucier de la façon dont une application distante fonctionne, ni de la manière dont les fonctions ont été implémentées pour pouvoir l'utiliser dans un programme

C) Le débat thématique sur les données logement

1) Présentation du cadre du groupe de travail

Conformément à sa feuille de route établie le 28 février 2013 en matière d'ouverture et de partage des données publiques, le Gouvernement a décidé de l'organisation de six débats thématiques et ouverts, afin d'identifier et de publier de nouveaux jeux de données, parmi lesquels le logement.

Le CIMAP du 17 juillet 2013 précise quant à lui, la programmation des trois premiers débats thématiques prioritaires, dont le logement. Il a été confié au Conseil national de l'habitat (CNH), qui rassemble l'ensemble des acteurs du logement, le soin de mettre en place, avec l'appui de la mission Etalab, un groupe de travail pluraliste (composé de membres du CNH et de représentants du monde des réutilisateurs de données publiques et de spécialistes de l'Open Data), afin de faire des propositions pour renforcer la mise à disposition de données publiques dans le champ du logement¹⁵.

Un groupe de travail a été installé au début de l'année 2014, associant les membres du CNH volontaires pour y participer et des représentants, proposés par Etalab, du monde des réutilisateurs de données publiques et de spécialistes de l'Open Data.

La diversité des acteurs au sein de ce groupe a été permise par la composition du CNH de :

- représentants des élus et des collectivités locales ;
- représentants des constructeurs, des maîtres d'œuvre et des entreprises du bâtiment, des gestionnaires de logements, des établissements financiers, des organismes d'allocations familiales et du notariat ;
- représentants des usagers ;
- représentants des associations d'insertion et de défense des personnes en situation d'exclusion par le logement ;
- personnalités choisies en raison de leur compétence par le ministre chargé du logement.

Par ailleurs, cette diversité a été renforcée par la présence de certains acteurs issus de la société civile se positionnant comme des réutilisateurs de données logement ainsi que des acteurs promouvant l'Open Data en France.

Cette composition a permis de rassembler les producteurs et les réutilisateurs de données, avec une grande richesse de points de vue¹⁶.

Ce groupe a été chargé de faire des propositions, notamment :

- sur la préparation et l'organisation de ce débat ;
- sur le format à donner à ce débat, relevant plutôt d'un groupe de travail large ;
- sur l'ouverture cohérente des données en matière d'offre de logement ;
- sur les objectifs poursuivis et bénéfiques attendus ;

¹⁵ Cf. Annexe 1

¹⁶ Liste des membres du groupe de travail en annexe 2

- sur les modalités d'ouvertures de ces données ;
- sur les conditions de réutilisation associées.

Géraldine Chalencon, Directrice générale de l'Agence Nationale pour l'Information sur le Logement (ANIL) a été désignée par le Président du CNH pour animer les travaux de ce groupe en lien avec la mission Etalab.

Le groupe de travail a procédé à l'audition des principaux producteurs de données afin d'avoir davantage de visibilité sur les informations produites en matière de logement¹⁷. Quelques acteurs ont également apporté leur contribution, en tant qu'utilisateurs.

Enfin, le groupe a articulé ses réflexions avec les travaux menés en parallèle par la mission commune de l'Inspection générale de l'INSEE (IG-INSEE) et du Conseil général de l'environnement et du développement durable (CGEDD)¹⁸.

2) Présentation des données publiques dans le champ du logement

Les données¹⁹ concernant directement ou indirectement le logement constituent un champ très vaste. Cependant, on peut tenter de définir le champ de ces données sous plusieurs angles.

Différentes formes de données doivent être distinguées, par nature :

- les données brutes : elles peuvent être collectées à des fins de gestion d'un dispositif ou d'une aide notamment, ou concernant un logement particulier, ou bien être obtenues sur la base d'enquêtes dédiées ;
- les données retraitées ou extraites, à partir des données brutes ;
- les données agrégées ou statistiques.

Ces données peuvent concerner, notamment :

- le logement en tant que bâtiment : sa localisation, son terrain, sa production, ses caractéristiques techniques, son entretien, son état, sa performance énergétique, son confort, ses équipements, ;
- le logement en tant que bien économique : ses coûts, sa valeur, les prix (achat, loyer,..), le financement, les aides publiques dont il a bénéficié, l'offre, la demande, la production de biens et services qu'il génère, la vacance,...

¹⁷ Liste des auditions en annexe 3

¹⁸ Inspection générale de l'Insee N°1.7.25 - Conseil général de l'Environnement et du Développement Durable N°009075-02 Rapport sur l'organisation du service statistique dans le domaine du logement - <http://www.ladocumentationfrancaise.fr/var/storage/rapports-publics/144000532/0000.pdf>

¹⁹ Pour la définition d'une donnée, on pourra se référer à celle contenue dans : [Sciences des données : de la logique du premier ordre à la Toile](http://books.openedition.org/cdf/529) -Leçon inaugurale prononcée le jeudi 8 mars 2012 – Serge Abiteboul - <http://books.openedition.org/cdf/529>

- les occupants : leur profil, leurs ressources, leurs conditions d'habitation, les dépenses et les aides dont bénéficient les ménages en lien avec le logement, leurs besoins, leurs attentes, leur parcours,....

Les producteurs de données dans ce domaine sont également très nombreux et divers : le champ couvert par le groupe porte uniquement sur les données publiques, produites par une administration ou reçues dans le cadre d'une mission de service public, tel que défini précédemment.

Le groupe de travail s'est appuyé sur le recensement des données réalisé en amont par la mission IG-INSEE-CGEDD, dans la cadre de l'élaboration de son rapport, produit en annexe de ce rapport (annexe 12)²⁰. Il convient d'ailleurs de souligner que cette annexe constitue en elle-même une source très utile pour améliorer l'accès aux données en matière de logement, en portant à connaissance l'ensemble des sources et en précisant leur contenu. Elle a été réutilisée pour réaliser une cartographie des données, qui a vocation à être mise à disposition et mise à jour²¹.

Le groupe s'est concentré sur le sujet du logement au sens strict, même si des sujets connexes ont pu être abordés à l'occasion des échanges, que ce soit l'aménagement urbain, l'hébergement, les fluides, etc...

3) L'ouverture et la mise à disposition des données logement : un sujet à la fois connexe et lié avec la problématique des statistiques dans ce champ

La volonté de l'État de travailler à l'ouverture des données publiques en matière de logement fait résonance avec le sujet des statistiques en la matière.

Les secteurs du logement, de l'immobilier et de la construction occupent à la fois une place centrale dans la vie des citoyens et dans l'économie et, de fait, dans les politiques publiques. Les enjeux qu'ils présentent appellent à disposer d'une connaissance fine et fiable des évolutions conjoncturelles. Or, force est de constater la faiblesse des données disponibles en la matière, à la fois sur le plan quantitatif et qualitatif. Pour répondre à cette difficulté, un rapport avait été commandé au Conseil national de la statistique en 2009 par les ministres chargé de l'économie et chargé du logement. Le groupe de travail, présidé par Bernard Vorms, Alain Jacquot et Jean-Louis Lhéritier, avait, parmi les recommandations faites dans son rapport publié en 2010²², déjà souligné l'importance de l'ouverture des données publiques en invitant les institutions chargées d'une mission de service public à diffuser les informations qu'ils produisent (données publiques et/ou fichiers individuels anonymisés) et qui sont utiles à la connaissance des conditions de logement des ménages et du fonctionnement des marchés du logement, ou à la conduite des politiques publiques en la matière. Le groupe soulignait également l'utilité, pour le

²⁰ Inspection générale de l'Insee N°1.7.25 - Conseil général de l'Environnement et du Développement Durable N°009075-02 Rapport sur l'organisation du service statistique dans le domaine du logement - précité

²¹ Annexe 4

²² Conseil national de l'information statistique - Rapport du groupe de travail « sur le logement et la construction » -Mars 2010

domaine du logement, des données d'origine fiscale et estimait que chacun devait pouvoir y accéder dès lors qu'elles avaient fait l'objet d'un traitement visant à les rendre anonymes.

Plus récemment, en 2013-2014, en réponse aux mêmes difficultés rencontrées pour définir et évaluer les politiques publiques en matière de logement, une mission conjointe CGEDD/INSEE²³ a été conduite avec pour objectifs de :

- faire des recommandations sur l'organisation cible la plus pertinente du service statistique public dans le domaine du logement ;
- faciliter l'utilisation et la compréhension des statistiques disponibles dans ce domaine, dans le contexte de l'article 29 de la loi du 18 janvier 2013 relative à la mobilisation du foncier public en faveur du logement et au renforcement des obligations de production de logement social ;
- dresser un panorama synthétique des lacunes en matière de production et de mise à disposition des données.

Le rapport, publié en septembre 2014, a également, dans ses préconisations, souligné la nécessité de développer l'accès aux données pour améliorer la connaissance statistique du secteur et plus largement, les travaux d'études et d'expertise.

« Le déficit d'études et d'offres d'expertise en matière de logement, manifeste au regard d'autres domaines, est aussi analysé par la mission comme découlant de l'insuffisante accessibilité de certaines bases de données, notamment les plus riches. Cette capacité d'expertise, singulièrement celle émanant du monde académique et scientifique, est pourtant indispensable pour répondre aux besoins de conseil et d'évaluation de l'Administration, tout autant que pour alimenter de manière indépendante et contradictoire le débat public. Le rapport recommande dans cet esprit que soient développées les possibilités d'accès à ces données, à des conditions économiques compatibles avec le statut non lucratif de la recherche publique, dans un cadre sécurisé. » (extrait de la synthèse du rapport).

Si la question de l'ouverture des données publiques et celle de la connaissance statistique sont bien deux questions distinctes, les conclusions des deux rapports précités montrent que l'un des enjeux de cette ouverture est aussi d'améliorer la connaissance du secteur, pour laquelle les attentes exprimées sont récurrentes et fortes.

Il convient de préciser que la logique d'ouverture des données publiques ne se limite pas à l'enjeu statistique. Elle ne porte pas non plus sur la qualité des données, ni sur leurs modes de traitement. Ainsi, les données disponibles se doivent d'être mises à disposition, même si elles ne sont pas nécessairement exhaustives, en précisant leur complétude. Indirectement, cette logique peut d'ailleurs jouer un effet de levier en termes d'optimisation des contenus, en ce qu'elle encourage ou contribue à enrichir les bases en les rendant visibles.

Cependant, les conditions de réutilisation des données que le réutilisateur s'engage à respecter prévoient que les exploitations qui en sont faites mentionnent la source des données, dans un souci de transparence. De plus, il est le seul responsable de la réutilisation de « l'information » qui en découle. En particulier, elle ne doit pas induire en erreur des tiers quant au contenu de l'information, sa source et sa date de mise

²³ Déjà cité

à jour. L'Open Data s'attache donc à apporter des garanties aux producteurs de données sur les réutilisations de leurs données et veille à la transparence quant aux sources et modes d'exploitation.

Ainsi, le groupe de travail s'est attaché à concentrer sa réflexion sur l'ouverture des données publiques, tout en montrant l'articulation de ce sujet avec les enjeux liés aux statistiques dans le champ du logement.

PARTIE II - Les enjeux de l'ouverture des données publiques et les attentes des acteurs du logement

Le logement est une question qui concerne chacun et qui interagit avec toutes les autres dimensions de la vie d'une personne (vie familiale, emploi, santé, etc.). Il présente des enjeux économiques directs et indirects importants. C'est aussi un sujet qui souffre d'un déficit de connaissances comme précédemment évoqué. Il s'agit donc d'une thématique sur laquelle les attentes sont fortes.

La mission Etalab, créée en février 2011, a d'ailleurs pu le constater depuis ses débuts, à la fois de la part des acteurs économiques, comme des citoyens. Une autre preuve de l'intérêt de cette thématique : les collectivités locales qui se sont engagées dans une politique d'Open Data ont très souvent inscrit le logement parmi leurs priorités.

A) L'ouverture des données publiques dans le champ du logement : un enjeu démocratique, de connaissance, économique et d'amélioration de l'action publique

1) Un enjeu démocratique

Les acteurs qui promeuvent de manière générale l'Open Data, citent très souvent le secteur du logement. Ainsi, par exemple, le site opendata.net en fait l'un de ces thèmes prioritaires. Ils soulignent l'enjeu de la mise à disposition de ces données notamment dans un but de transparence de l'action publique, le logement étant un domaine où les politiques publiques sont importantes, à la fois sur le développement de l'offre (foncier, construction, production de logements sociaux, aides à l'accession ou à l'investissement locatif, ...), sur l'entretien du parc existant ou sur les aides aux personnes. Pour eux, comme pour les associations agissant dans le secteur, la mise à disposition de données doit permettre de mieux évaluer les politiques menées, les résultats atteints et les besoins insatisfaits des ménages qui appelleraient des évolutions de l'action publique. Elle peut aussi contribuer à un meilleur partage des connaissances entre les acteurs publics et la société civile.

2) Un enjeu de connaissances du secteur

Le besoin de renforcer les connaissances dans le domaine du logement a été souligné par de nombreux rapports, comme évoqué précédemment. Les chercheurs qui travaillent sur le logement soulignent le besoin de disposer de données pour améliorer la connaissance. Il est en effet difficile d'approfondir les travaux de recherche sans disposer de données. Par exemple, comment mesurer l'évolution des prix (loyer, achat) sans disposer de données fiables sur le sujet ? Compte tenu à la fois des enjeux pour les conditions de vie des ménages et des impacts budgétaires et économiques, améliorer la connaissance dans ce domaine paraît fondamental. Les travaux de recherches sur le logement peuvent notamment

contribuer à évaluer les politiques publiques et à alimenter les réflexions des pouvoirs publics. Ces derniers sont en effet souvent confrontés à des difficultés pour mesurer l'impact des politiques menées. Ainsi, évaluer l'impact du soutien à l'investissement locatif nécessiterait au préalable de connaître précisément les logements et les investisseurs ayant bénéficié des aides, leur devenir, etc...

3) Un enjeu d'amélioration de l'action publique

Une meilleure connaissance des conditions de logement, des besoins, des mécanismes économiques, des comportements des acteurs, permettrait d'améliorer les politiques publiques dans le champ du logement : c'est ce déficit de connaissances qui a conduit les parlementaires à demander un outil statistique, après avoir souligné dans le débat le fait que sur différents sujets, des chiffres divers et parfois contradictoires coexistent, rendant difficile l'élaboration de politiques publiques. Ainsi, l'article 29 de la loi du 18 janvier 2013 a donné lieu au rapport précité IG-INSEE-CGEDD. Ce rapport, qui porte sur l'information statistique, met en regard le montant consacré à la production de statistiques publiques en matière de logement, de l'ordre de 30 millions et le montant des politiques publiques, de l'ordre de 42 milliards. Les collectivités locales sont également en attente de disposer de données pour les aider à conduire leurs politiques en ayant une connaissance plus fine du logement sur leur territoire. L'ouverture des données publiques peut également favoriser des synergies et complémentarités entre les acteurs publics, dans le sens d'une plus grande efficacité grâce au partage d'informations.

4) Des enjeux économiques

En termes d'enjeux économiques, l'ouverture des données publiques peut concerner différents domaines.

D'une part, en termes de connaissance et dans ce que l'on peut appeler « l'économie classique » du logement, le secteur de la construction, de l'immobilier ou du financement du logement peuvent voir un intérêt dans le fait de mieux connaître les attentes et besoins des ménages, ainsi que les marchés ou l'offre existante pour mieux orienter leur activité.

D'autre part, l'ouverture des données peut également conduire à développer de nouvelles activités et constituer des sources d'innovation. De nombreuses start-up se sont développées dans le secteur du logement en identifiant la réutilisation des données publiques comme un vecteur potentiel d'activité. Les données publiques peuvent permettre à ces acteurs économiques, après en avoir assuré un traitement, d'apporter un niveau d'informations supplémentaires à leurs usagers qui, croisées avec d'autres données, permettent des analyses fines et pertinentes. Les données en matière de logement représentent donc potentiellement une source de développement de nouveaux services pour les usagers. Par exemple, il peut s'agir de croiser des données concernant les services et les logements sur un territoire, ou encore, l'accès aux transports et les logements. À partir de ces données, il est possible d'innover et de créer de nouveaux services, sans qu'il soit possible d'envisager toutes les pistes à ce stade.

Sur ces différents enjeux, le groupe de travail, constitué à la fois de professionnels du secteur, d'associations œuvrant dans le domaine du logement, d'associations de consommateurs ou familiales,

d'entrepreneurs de start-up ayant un lien avec le logement, a exprimé des attentes diverses vis-à-vis des données dont il souhaitait la mise à disposition. Il a également souligné les avancées déjà faites dans ce domaine, qui ne sont cependant pas toujours connues. Et l'un des intérêts de ce groupe a aussi été de permettre de les rendre visibles.

B) Les attentes exprimées par les acteurs

Les échanges ont fait émerger des attentes portant sur les données disponibles, mais aussi sur leur format. De même, les échanges ont souvent porté sur la granularité, ou l'échelle géographique des données mises à disposition. Les acteurs ont fait part de l'utilité d'une localisation la plus précise possible des données mises à disposition, cette attente devant être conciliée avec le respect de la vie privée qui s'impose tout particulièrement dans le domaine du logement. Les producteurs ont également souvent évoqué la question de l'obstacle que peut présenter le coût de la mise à disposition des données notamment lié à la nécessaire anonymisation des données pour assurer le respect de la vie privée.

1) En termes de données souhaitées

a) Des données brutes attendues qui ne sont pas collectées

Les attentes exprimées par les acteurs ont pu porter sur des données qui sont en fait inexistantes sur le logement. Le déficit de données a été déjà souligné dans le cadre de différents rapports.

Ainsi, par exemple, il serait souhaité un référencement des logements adaptés au handicap, en particulier dans le parc social. De même, les loyers dans le parc privé, ou les loyers perçus par les bailleurs privés par logement et déclarés aux impôts. Ou encore une base de tous les logements existants avec leurs grandes caractéristiques et un numéro d'identification.

À ce jour, de nombreuses données ne sont pas collectées, ou très peu.

Il convient cependant de souligner que dans le champ du logement, de nombreuses avancées ont eu lieu et se poursuivent dans la construction de systèmes d'information, permettant d'améliorer le recensement des données existantes. Ainsi, par exemple, le Répertoire du Parc Locatif Social (RPLS) permet aujourd'hui d'avoir une bonne connaissance de ce parc. En termes de perspectives, un système national d'enregistrement des demandes est en cours de construction, un registre national des copropriétés est prévu par la loi ALUR du 24 mars 2014 pour l'accès au logement et un urbanisme rénové (ALUR), ou encore, un outil commun aux ministères du logement et de l'intérieur est en cours de mise en place sur les expulsions locatives (Exploc).

Il convient de distinguer les systèmes de gestion d'un dispositif ou d'une politique qui sont mis en place par l'État et qui peuvent produire des données publiques, des travaux de collecte de données à des fins statistiques (comme les observatoires locaux des loyers). Pour ce qui concerne les systèmes d'information destinés à la gestion d'un dispositif, l'enjeu de connaissance et de mise à disposition n'est pas le premier objectif, mais le système d'information peut ensuite avoir une valeur ajoutée importante en termes de données disponibles, pouvant contribuer à améliorer la connaissance des besoins ou à créer des services aux citoyens. Cependant, il faut prendre en compte le fait que le producteur ne cherche en

principe qu'à recueillir les données dont il a besoin pour son propre usage et dans le cadre du dispositif qu'il gère. Ainsi, pour prendre un exemple de système existant qui est celui de gestion des déclarations de revenus par l'administration fiscale, concernant les loyers, les bailleurs n'ont pas à renseigner par logement les revenus perçus à ce titre, ni à préciser leurs caractéristiques dans la mesure où il ne s'agit pas d'un besoin de l'administration fiscale.

L'État peut cependant, parce qu'il y perçoit un enjeu, organiser un système de collecte des données à partir de la saisie par les acteurs impliqués. Par exemple, l'article 1^{er} de la loi du 12 juillet 2010 portant engagement national pour l'environnement a introduit l'obligation de transmettre les Diagnostics de performance énergétique (DPE), réalisés à l'occasion de la vente ou de la location d'un bâtiment ou d'un logement, à l'ADEME. Ainsi, une base de données de recueil des DPE qui doit être remplie par les diagnostiqueurs depuis 2013 a été mise en place par l'ADEME et constitue une nouvelle source particulièrement intéressante pour le secteur du logement. Le DPE comporte en effet un grand nombre de données sur le logement, définies par l'arrêté du 8 février 2012 modifiant l'arrêté du 15 septembre 2006 relatif au Diagnostic de performance énergétique pour les bâtiments existants proposés à la vente en France métropolitaine. Le site www.observatoire-dpe.fr permet d'obtenir des graphiques sur les DPE sur la base de requêtes, mais sans mettre pour l'instant à disposition des données sous un format réutilisable.

Il faut cependant souligner que, parfois, il apparaîtrait naturel aux acteurs que certaines données soient renseignées dans une base, à un niveau donné, alors qu'il n'en est rien. Les échanges du groupe de travail ont bien mis en évidence cette dichotomie entre la perception des acteurs et la réalité des données existantes, brutes ou agrégées.

b) Des données statistiques trop limitées

À l'insuffisance de collecte des données, s'ajoute la faiblesse des données statistiques disponibles soulignée dans le cadre du groupe de travail, à la suite de nombreux rapports sur ce sujet (cf. rapport CGEDD/IG-INSEE précité).

Il s'agit, dans ce cas, non pas de données enregistrées dans le cadre d'un dispositif ou d'une politique, mais de statistiques élaborées soit à partir de données de gestion existantes, soit à partir de celles collectées dans le cadre d'enquêtes.

Parmi les manques, il convient de souligner que la majeure partie des données statistiques disponibles ne le sont pas à un niveau local très fin, alors même que les collectivités locales sont de plus en plus impliquées dans les politiques de logement.

Les données agrégées à l'échelle nationale offrent un aperçu des politiques nationales mais ne rendent pas compte des réalités locales qui sont pourtant très diverses selon les territoires. Les données disponibles peuvent être agrégées au niveau régional ou départemental également. Mais souvent, les acteurs souhaiteraient des données au niveau des EPCI, voire à un niveau plus fin. Cette question de la granularité se pose plus largement, comme cela sera abordé ultérieurement.

Il a été noté que certaines bases de données se sont même appauvries au fil du temps (par exemple, Sit@del, du fait de la décentralisation de la procédure de demande de permis de construire).

2) En termes de format et de granularité de mise à disposition de données

a) *Formats des données*

La question du format de la mise à disposition est apparue comme centrale dans les échanges du groupe de travail.

Grâce au développement de l'informatique, il est possible aujourd'hui d'avoir accès à un large choix de types de données. Il y a bien évidemment le texte, mais aussi le son et l'image pour ne citer que les plus utilisés. Toutefois, aussi diverses soient-elles, les données sont toutes représentées en définitive par des 0 et des 1 (représentation binaire) pour qu'elles puissent être interprétables par l'ordinateur²⁴.

Une fois représentées grâce à un format, les données peuvent être stockées (enregistrées) sur support informatique via un fichier ayant lui aussi un format spécifique.

Pour réaliser le fichier qui va permettre le stockage de ces données, il faut en définir le format.

Ainsi, chaque logiciel qui crée un fichier (document texte, tableau...) utilise un format d'enregistrement pour le créer qui peut être spécifique au logiciel (propriétaire) ou standard et donc utilisé et compris par d'autres logiciels. Un format de données est dit ouvert si son mode de représentation a été rendu public par son auteur et qu'aucune entrave légale ne s'oppose à sa libre utilisation (droit d'auteur, brevet, copyright).

Les formats ouverts sont généralement créés dans un but d'interopérabilité. Un document enregistré dans un format ouvert sera indépendant du logiciel utilisé pour le créer, le modifier, le lire et l'imprimer. L'interopérabilité laisse le choix du logiciel pour utiliser le document libre.

Depuis le milieu des années 2000, certains États ou administrations (par exemple, l'administration fédérale suisse, la gendarmerie nationale française ou le gouvernement britannique) se sont montrés très attentifs à l'utilisation de formats ouverts. Au-delà de l'intérêt direct pour ces acteurs (qui tient essentiellement à l'interopérabilité), deux raisons expliquent ce soutien :

- d'une part, le secteur public ne peut adopter un format fermé pour ses échanges avec les administrés ou les usagers sans imposer l'utilisation d'un logiciel particulier, ce qui revient à donner à son éditeur une position dominante sur le marché, qui pourrait alors en abuser et rendre ce secteur dépendant de cet éditeur pour l'accès à ses propres données et outils ;
- d'autre part, manipulant des données sensibles (secret défense, informations fiscales ou médicales par exemple) et archivant ses documents, la puissance publique doit avoir l'assurance

²⁴ [Sciences des données : de la logique du premier ordre à la Toile](http://books.openedition.org/cdf/529) - Leçon inaugurale prononcée le jeudi 8 mars 2012 – Serge Abiteboul - <http://books.openedition.org/cdf/529>

que celles-ci restent consultables à l'avenir, même si le logiciel qui les a générées n'est plus disponible.

Plusieurs sujets relatifs au format ont été abordés dans le cadre des travaux du groupe.

Ainsi, lorsque des données sont rendues publiques sous format PDF, elles sont de fait difficilement exploitables, à moins de les ressaisir. Pourtant, les données issues de systèmes d'information ne sont généralement pas restituées sous ce format. Un retraitement est effectué de la part de l'entité productrice pour transformer les données au format PDF pour en protéger le contenu et garantir sa pérennité. Chaque donnée mise à disposition au format PDF est obligatoirement sous-tendue par un format source plus ou moins facilement lisible par une machine.

Par ailleurs, des acteurs ont souligné les difficultés rencontrées pour réutiliser les données proposées au format « Beyond 20/20 »²⁵. En effet, il faut connaître le nom des colonnes afin de procéder à des travaux de réutilisation. Il apparaîtrait nécessaire de disposer des bases sources afin de permettre le développement d'usages innovants, même si le format « Beyond » a été choisi a priori pour faciliter l'utilisation.

b) Granularité des données

Il existe aujourd'hui de nombreuses données produites par l'appareil statistique public (enquête logement de l'INSEE, statistiques sur les permis de construire, les permis d'aménager ainsi que les déclarations préalables, etc.), par des administrations ou des organismes chargés d'une mission de service public, comme la Société de gestion du fonds de garantie de l'accession sociale à la propriété (SGFGAS), la Banque de France, l'Anah, etc.

La mise à disposition de données agrégées au niveau départemental, voire intercommunal ou communal, ne semble pas toujours suffisante compte tenu des besoins identifiés par les professionnels du logement ainsi que par les organisations citoyennes qui demandent davantage de précision géographique dans les données transmises par l'administration, à des fins d'études notamment, ou pour développer des services géolocalisés. Les membres du groupe de travail insistent sur l'importance de descendre à une granularité fine tout en veillant à respecter le secret statistique et la vie privée.

En effet, mettre à disposition des prix d'achat ou de loyer au mètre carré à l'échelle d'une commune comme Paris ne représente pas d'intérêt pour des citoyens souhaitant avoir davantage de visibilité sur les prix en vue de l'acquisition d'un bien immobilier compte tenu de leur hétérogénéité. L'enjeu de la granularité des données a ainsi été souligné concernant les données sur les loyers qui devraient être produites dans le cadre du réseau des observatoires locaux des loyers mis en place par l'État, avec les collectivités locales.

²⁵ Beyond 20/20 est un visualisateur (browser) de données qui permet de constituer des tableaux multidimensionnels personnalisés à partir d'un fichier de données individuelles, de consulter et de manipuler des tableaux de données prédéfinis.

Les producteurs ont cependant souligné que la diffusion des données à une échelle plus fine dans le respect de la vie privée nécessitait des traitements, qui, certes, peuvent désormais être relativement automatisés, mais ne font pas partie de la mission de service public initiale.

Sur le sujet spécifique de la cartographie, enjeu important dans le champ du logement, pour pouvoir "attacher" de l'information à des territoires et les représenter sur des cartes, les réutilisateurs soulignent la nécessité de disposer de leurs formes géométriques (polygones : données vectorielles), géolocalisées précisément et accompagnées d'identifiants stables. Ils insistent également sur le format de la projection et demandent une projection globale (type WGS84, EPSG4326).

La priorité est de définir une granularité garantissant à la fois une confidentialité et un niveau de précision suffisants de chacun des jeux de données à ouvrir, pour autoriser une représentation cartographique pertinente.

L'IRIS²⁶ a été identifié par le groupe de travail comme étant, de manière générale, la maille géographique la plus pertinente en matière de données sur le logement. Les contours IRIS (IGN/INSEE) présentent l'avantage d'être étudiés pour satisfaire les règles du secret statistique (si une donnée agrégée concerne trop peu de ménages dans un IRIS, on l'agrège avec celle un ou de plusieurs autres IRIS). Grâce aux identifiants IRIS, il est possible de reconstituer les quartiers, les communes, etc. Cependant, aujourd'hui, ces contours ne sont pas ouverts et des attentes importantes ont été exprimées sur ce sujet dans le cadre du groupe de travail.

c) Profondeur temporelle des données et actualisation

Les réutilisateurs ont également souligné l'importance de pouvoir disposer à la fois de données anciennes et actuelles.

La profondeur temporelle présente notamment un intérêt important pour les travaux d'études.

Concernant les données historiques, leur mise à disposition peut se heurter à des questions de format. La mise à disposition des plus récentes d'entre elles se heurte quant à elle parfois aux temps de traitement longs et aux moyens nécessaires à leur mise à disposition.

Le mode de collecte peut également être à l'origine des délais avant la mise à disposition. Ainsi, par exemple, concernant la base de données Sit@del²⁷ produite par le SOeS, ce sont les centres instructeurs des permis de construire qui font remonter les données. Celles-ci sont ensuite constituées par date de

²⁶Afin de préparer la diffusion du recensement de la population de 1999, l'INSEE avait développé un découpage du territoire en mailles de taille homogène appelées IRIS2000 (Un sigle qui signifiait : « Ilots Regroupés pour l'Information Statistique » et qui faisait référence à la taille visée de 2 000 habitants par maille élémentaire). Depuis, l'IRIS (appellation qui se substitue désormais à IRIS2000) constitue la brique de base en matière de diffusion de données infra-communales. Il doit respecter des critères géographiques et démographiques et avoir des contours identifiables sans ambiguïté et stables dans le temps. Les communes d'au moins 10 000 habitants et une forte proportion des communes de 5 000 à 10 000 habitants sont découpées en IRIS. Ce découpage constitue une partition de leur territoire. La France compte environ 16 100 IRIS dont 650 dans les DOM.

²⁷Les données de Sit@del2 proviennent des formulaires de permis de construire, des formulaires de déclarations d'ouverture de chantier (DOC) et des déclarations attestant l'achèvement et la conformité des travaux (DAACT). Ces informations sont transmises mensuellement par les centres instructeurs.

prise en compte, puis, dans un deuxième temps, par date réelle des événements. Les données en date de prise en compte sont publiées avec un délai d'un mois. En revanche, le délai de publication en date réelle est plus long, entre 6 mois et 2 ans en fonction de leur typologie.

Compte tenu de la décentralisation de l'instruction des permis de construire, l'État a une maîtrise plus limitée des remontées de données et les collectivités, qui mettent en œuvre cette compétence depuis peu, peuvent prioriser l'instruction des permis de construire plutôt que la collecte des données publiques.

Concernant les données statistiques en matière de logement, la question de la fréquence de l'enquête logement réalisée par l'INSEE a été évoquée (la précédente édition date de 2006 et la prochaine est en cours pour l'édition 2013/2014). Elle constitue une source de qualité particulièrement appréciée des acteurs, qui regrettent cependant qu'elle ne puisse être élaborée plus fréquemment. Il convient de rappeler en effet que cette enquête repose sur une collecte de données très importantes, qui nécessite donc une mobilisation de moyens à la hauteur du volume de la collecte.

3) En termes d'ouverture de nouvelles données publiques

Les attentes des réutilisateurs portent enfin sur des données publiques existantes mais qui ne sont aujourd'hui pas mises à disposition.

Ainsi, par exemple, les données sur les mutations d'immeubles à titre onéreux, sur le financement des opérations (accession en particulier), sur la réalisation des objectifs en termes de développement de l'offre de logement social par commune (loi SRU), sur les DPE, sur les logements vacants (soumis à la taxe afférente), sur le foncier en général, avec une précision géographique la plus fine possible, ont été mentionnées.

Il s'agit de données qui existent, mais qui ne sont pas aujourd'hui mises à disposition, alors qu'elles sont identifiées comme pouvant contribuer à une meilleure connaissance et transparence de l'action publique.

À noter que se développe depuis quelques années une ouverture des données pour les collectivités locales, qui sont au cœur de la politique de logement et ont besoin de données produites par l'administration centrale pour mener au mieux leurs missions.

Ainsi, dans ce cadre, depuis juillet 2011, les collectivités et les EPCI ont accès à une nouvelle donnée, appelée « Demandes de valeurs foncières » (DVF)²⁸. Il s'agit d'une base de données contenant l'intégralité des mutations foncières et immobilières mise à disposition gratuitement par l'administration fiscale sous format électronique, par le biais d'un portail sécurisé et sous certaines conditions (ces conditions étant les mêmes que pour les EPA, EPF et services de l'État, pour lesquels la mise en œuvre est en cours).

Grâce à cette ouverture, les collectivités locales disposent désormais d'une donnée gratuite, exhaustive, dans un domaine particulièrement important : les marchés fonciers, de l'immobilier et de

²⁸http://www.legifrance.gouv.fr/affichCodeArticle.do;jsessionid=5E9A0B36CC20BE00C74F188862EFD555.tp_djo09v_3?idArticle=LEGIARTI000006315265&cidTexte=LEGITEXT000006069583&dateTexte=20080121 et <http://www11.minefi.gouv.fr/boi/boi2007/13repub/textes/13k1007/13k1007.pdf>

l'aménagement. La donnée est néanmoins diffusée sous forme brute et ne peut être exploitée sans précautions préalables. Elle exige que les utilisateurs puissent bien en comprendre les contenus pour l'exploiter au mieux.

Dans le cadre du groupe de travail, il a été suggéré que les données DVF transmises au groupe DVF pourraient être mises à disposition d'un tiers de confiance chargé de leur anonymisation avant une mise en Open Data.

De même, la base Filocom qui constitue une source particulièrement utile au secteur du logement, peut être mise à disposition des collectivités locales ou de cabinets d'études pour le compte de la collectivité dans un certain cadre. Les données sur la vacance détenues par l'administration fiscale peuvent également faire l'objet d'une mise à disposition auprès des collectivités locales.

Comme expliqué précédemment, la question des redevances a fait l'objet d'échanges au sein du groupe de travail. Si les réutilisateurs souhaitent de manière générale une mise à disposition gratuite, conformément à la logique de l'Open Data et telle que réaffirmée dans le cadre des travaux du CIMAP, les producteurs ont aussi fait part des surcoûts liés à la mise à disposition des données, notamment concernant leur collecte et leur anonymisation. La question du coût de la mise à disposition est notamment évoquée pour les bases notariales²⁹.

4) En termes d'accessibilité et de visibilité des sources de données mises à disposition

Les travaux de ce groupe ont été l'occasion d'échanges riches entre producteurs et utilisateurs de données, permettant de partager les contraintes des uns et les attentes des autres. Ils ont également permis de souligner le déficit de connaissance de l'existant, ou les difficultés à le trouver.

Ainsi, concernant le site de l'INSEE qui constitue une source d'informations particulièrement riche, plusieurs utilisateurs ont noté des difficultés à se repérer parfois, ce qui engendre une sous-utilisation des possibilités offertes par l'INSEE.

Concernant les données mises à disposition par le SOeS (service statistique au sein du Commissariat général du Développement Durable), qui se sont développées au cours des dernières années, elles ne sont pas non plus toujours connues, d'autant que leur « rattachement » au champ du logement ne se fait pas forcément de manière évidente. Le SOeS alimente en effet le site dédié www.statistiques.developpement-durable.gouv.fr en présentant les informations qu'il produit sur les thèmes du logement, ce site portant également sur les données relatives à la construction en général, aux transports, à l'énergie et au climat, à l'environnement et au développement durable. Il faut noter que les données mises à disposition par le SOeS sont également présentées sous forme de cartographies dynamiques (Géoidd) permettant d'interpréter les données et de valoriser le travail d'analyse réalisé par le service statistique du ministère.

²⁹ Les bases de données notariales, qui font partie des attentes exprimées par les réutilisateurs de données ont été écartées du périmètre de réflexion du groupe de travail, du fait des discussions en cours sur les textes d'application de la loi du 28 mars 2011.

Concernant les données produites par le ministère en charge du logement, il convient de souligner qu'un nouveau site a été mis en place récemment pour centraliser les informations sur le financement du logement social : <http://www.financement-logement-social.territoires.gouv.fr/>. Il rassemble les informations réglementaires et fait le lien vers les systèmes d'information sur ce sujet.

De plus, le site relatif au Système national d'enregistrement des demandes de logement social comporte des tableaux de bord sous format Excel : <http://sne.info.application.territoires.gouv.fr/>.

Ces sources de données ne sont pas référencées sur www.data.gouv.fr, ni accessibles facilement depuis le site du ministère en charge du logement.

Enfin, il convient de souligner l'avancée en termes de connaissances des sources de données que représente le travail effectué dans le cadre de la mission CGEDD/INSEE, évoqué précédemment. L'ensemble des fiches est aujourd'hui disponible en annexes du rapport, mis à disposition sur le site du CGEDD. Le présent rapport comporte en annexe³⁰ une synthèse de la plupart de ces fiches concernant son périmètre, qui constitue une cartographie des données en matière de logement. Pour chaque source, sont précisées les conditions de mise à disposition des données et les perspectives, le cas échéant.

³⁰ Annexe 4

PARTIE III – Propositions de pistes de travail

Au-delà du principe de la mise à disposition des données publiques dans le domaine du logement qui doit être poursuivi, dans le respect du cadre légal, de la vie personnelle et du secret statistique, les réflexions du groupe permettent de dégager des principes qui pourraient contribuer à renforcer l'accès aux données dans le champ du logement.

A) Améliorer le format de mise à disposition (formats ouverts, interopérables)

L'objectif de l'ouverture des données publiques est de favoriser et de faciliter les réutilisations et les réinterprétations, de la manière la plus automatisée et la plus standardisée possible. Les données brutes – telles que produites ou utilisées par les administrations à des fins de service public – sont en ce sens extrêmement intéressantes. Il est préférable de les diffuser dans des formats structurés, sans avoir recours à des options de présentation (couleurs, cellules fusionnées, fichiers à plusieurs onglets..), ni à des fonctions de présentations (macros, liens croisés dynamiques...).

Pour en permettre une réutilisation simple par le plus grand nombre, le groupe de travail suggère de présenter ces données dans des formats ouverts (Exemple : CSV, JSON, XML, RDF...) qui permettent la réutilisation sans restriction d'accès ni de mise en œuvre, par opposition à un format fermé ou propriétaire. La circulaire du Premier ministre du 19 septembre 2012³¹, sur l'usage du logiciel libre dans l'administration, encourage l'usage de ces formats réutilisables et ouverts.

Dans la mesure du possible, l'ouverture des données publiques requiert la diffusion des données brutes et complètes dans des formats normalisés qui permettent une réutilisation simplifiée dans des applications. En complément, les données peuvent également être diffusées sous forme de flux accessibles à travers des interfaces de programmation (API).

Il apparaît souhaitable que les données diffusées soient les plus exhaustives et les plus précises possible, avec une granularité fine dans le respect des différents secrets, et qu'elles s'appuient sur des référentiels partagés et des nomenclatures décrites et publiées.

Les auditions ont permis de repérer les freins à la réutilisation liés au formalisme des données publiques en matière de logement et de trouver des solutions simples lorsque cela était possible.

Par exemple, comme mentionné précédemment, le référencement de service de visualisation de données tel que Beyond 20/20 constitue déjà une réutilisation des données et complexifie de manière significative les potentielles nouvelles réutilisations. Il est apparu envisageable au SOeS de travailler à une mise à disposition des bases de données sources dans des formats facilitant leurs réutilisations potentielles (par exemple sur la base Sit@del).

³¹ http://circulaire.legifrance.gouv.fr/pdf/2012/09/cir_35837.pdf

De même, les informations rendues disponibles dans les comptes du logement pourraient l'être dans un format réutilisable. « L'état du logement » est mis à disposition sous format Excel, il s'agit d'un exemple pertinent.

Dans un autre registre, des associations ont émis le souhait que les fichiers statistiques sur le suivi du dispositif DALO, mis à disposition sous format PDF de manière régulière, soient mis également à disposition en format CSVcsv. Il a d'ailleurs été souligné à cette occasion l'exemplarité du dispositif de suivi mis en œuvre au moment de son développement, à la fois avec des outils et dans la transparence au travers de la mise à disposition des fichiers statistiques sur le site Internet du comité de suivi DALO.

B) Accroître la granularité des données mises à disposition :

Comme expliqué précédemment, les membres du groupe de travail insistent sur l'importance de descendre à une granularité fine sans pour autant porter atteinte au secret statistique.

Les producteurs soulignent quant à eux, la nécessité de faire des traitements en matière d'anonymisation, qui, certes, peuvent désormais être relativement automatisés, mais qui ne font pas partie de la mission de service public initiale et représentent un coût.

Les fichiers liés aux prêts à taux zéro sont un exemple significatif des attentes des acteurs et des difficultés du producteur. Aujourd'hui mis à disposition à une échelle départementale, il est cependant envisageable d'aller plus loin dans la granularité, mais cette volonté se confrontera à un moment donné à la question du secret statistique qui ne peut être résolue qu'avec des traitements qui présentent un coût supplémentaire.

L'administrateur général des données, qui a pour mission de veiller à la production ou l'acquisition de données essentielles pourrait accompagner les producteurs concernés dans le processus d'anonymisation des contenus afin de permettre la mise à disposition de ces informations publiques en Open Data.

Enfin, comme précisé précédemment dans ce rapport, l'IRIS a été identifié par le groupe de travail comme étant, de manière générale, la maille géographique la plus pertinente en matière de données logement. Il convient cependant de pouvoir disposer de ces contours. L'INSEE et l'IGN, propriétaires du découpage en IRIS, ont confirmé au groupe de travail que des actions sont actuellement en cours pour ouvrir plus largement les contours données dans des conditions plus favorables pour permettre leur réutilisation, au niveau des îlots tels que définis par l'INSEE.

C) Prévoir le recensement des données dans le cadre des délégations de compétences ou transfert de compétences

Si la question de la granularité des données est un enjeu majeur en matière de logement, il n'est possible d'aller vers une plus grande finesse que si les données existent. Or, il s'avère que, du fait de la

décentralisation de certains dispositifs ou de certaines politiques, le recueil de données s'est finalement appauvri. Ont été cités notamment les exemples du Fonds de Solidarité Logement et des Permis de Construire. Dans le cas du Fonds de Solidarité Logement, les données sur les aides délivrées existent bien dans chaque département, mais il est difficile d'en avoir une vue d'ensemble, sauf à organiser une enquête de remontée de données, ce qui est coûteux en temps, à la fois pour le niveau national, mais aussi pour les collectivités locales, qui ne disposent pas de données homogènes et doivent donc éventuellement faire des retraitements pour répondre à une telle enquête.

En conséquence, les acteurs invitent à réfléchir le dispositif de remontée de données au niveau national, dans un format standardisé, au moment du transfert ou de la délégation de compétences (à noter que dans ce second cas de figure, la nécessité d'évaluation permet en général d'éviter cet écueil).

D) Mettre à disposition de nouveaux jeux de données en matière de logement

Le formalisme et la granularité des données produites par les administrations sont des éléments essentiels permettant de faciliter les usages.

Toutefois, comme évoqué précédemment, les attentes des réutilisateurs de données de logement portent également sur l'ouverture et la mise à disposition en Open Data de nouvelles données publiques.

Dans le cadre des auditions du groupe de travail, chacune des sources a fait l'objet d'échanges avec les producteurs, ce qui a permis d'identifier des données publiques à fort potentiel de transparence et d'innovation. Le travail de cartographie élaboré par la mission commune IG-INSEE/ CGEDD dans le cadre de son rapport sur l'organisation de la production de statistiques sur le logement a constitué une aide importante dans l'identification des sources.

À partir de ces travaux, il est proposé en annexe du présent rapport³², une cartographie de synthèse qui mentionne, le cas échéant, les avancées prévues en termes de mise à disposition de données, sur la base des échanges qui ont eu lieu avec les producteurs dans le cadre de ce débat (en dernière colonne, Remarques et perspectives).

Il en ressort cependant que, en dehors des données qui peuvent être produites à partir des bases fiscales³³, peu de sources ne sont pas du tout mises à disposition et les marges de progrès sont surtout dans l'approfondissement ou dans la construction des bases. Des possibilités existent cependant, comme sur les données sur la réalisation des objectifs de construction de logements sociaux dans le cadre de la SRU par exemple, sur les aides publiques délivrées, à la fois pour le logement social et pour les ménages, ou sur la demande de logement social, dont le système d'information est cependant en cours de construction. Les producteurs ont pu faire part de leur travail sur le sujet.

³² Annexe 4

³³ En particulier les données sur les valeurs foncières, sur les mutations ou celles sous-jacentes à la base Filocom, sur lesquelles des démarches sont en cours.

Ainsi, la DHUP (Direction de l'Habitat, de l'Urbanisme et des Paysages) qui gère les systèmes d'information du ministère en charge du logement, a déjà initié, dans le cadre du développement de ces systèmes, la mise à disposition de données sur la demande de logement social comme sur le financement du logement social (sur les sites précités : <http://sne.info.application.territoires.gouv.fr/> et <http://www.financement-logement-social.territoires.gouv.fr/>). Ce travail se poursuivra, dans la limite des moyens disponibles, car la mise à disposition de fichiers sur un site Internet prend toujours nécessairement du temps, même s'il est plus réduit lorsque l'outil produit les fichiers de manière automatisée.

Concernant l'Anah par exemple, elle souhaite travailler sur la mise à disposition de données relatives aux aides qu'elle délivre.

E) Penser la mise à disposition des données lors de la mise en place du système d'information

Il convient de souligner qu'il est ressorti des échanges l'importance de penser cette mise à disposition le plus en amont possible lorsque s'élabore un système d'information, bien que, justement, la complexité des projets SI en fait souvent une priorité de second ordre.

Cette volonté a cependant été posée très clairement dans le cadre de la mise en place des observatoires locaux des loyers en cours de construction. Et elle répond à des attentes fortes d'acteurs très divers sur les données de loyers du parc privé, et ce à une échelle fine. Ainsi, l'arrêté du 10 novembre 2014 relatif aux observatoires locaux des loyers et aux modalités de communication et de diffusion de leurs données prévoit celles qui devront obligatoirement être mises à disposition par les observatoires locaux des loyers. Il prévoit également que l'ensemble des données soient disponible sur un site internet national.

Aujourd'hui, la DHUP confirme intégrer en amont de la construction des systèmes d'information la question de la mise à disposition des données. C'est ce qui a permis d'avancer sur les données relatives la demande de logement social comme exposé précédemment. Cette question est bien prévue également dans la construction du système d'information sur les expulsions locatives (Exploc). À l'inverse, il est difficile de progresser en termes de mise à disposition de données lorsque le système d'information n'a pas été conçu en le prévoyant.

Pour l'avenir, a été mentionné le futur registre national des copropriétés qui pourrait constituer une source de données particulièrement riche et pour lequel il serait donc pertinent de penser, le plus en amont possible, à la mise à disposition des données.

De même, à titre d'exemple, il serait utile que les réflexions sur le carnet de santé numérique pour le logement, prévu dans le projet de loi relatif à la transition énergétique, intègrent dès le départ la question de l'ouverture possible des données qui seront recensées grâce à cette carte.

F) S'appuyer sur la mission Etalab et l'administrateur général des données pour mettre en œuvre le processus de mise à disposition des données logement

Dans le cadre des travaux du groupe de travail logement, de nombreuses sources de données à fort potentiel de transparence et d'innovation ont été identifiées. Toutefois le caractère « sensible » de ces dernières, du fait de leur précision géographique et/ou de leur protection par un secret légal, ne permet pas aujourd'hui d'en exploiter tout le potentiel.

Or, faciliter la circulation des données au sein de l'administration comme à destination des chercheurs, entreprises et citoyens et permettre la mise à disposition de tous des données clés comptes parmi les principaux chantiers de l'administrateur général des données.

Le logement fait partie des secteurs stratégiques permettant de renforcer l'efficacité des politiques publiques et de contribuer à l'amélioration de la qualité des services rendus aux usagers.

Il est proposé que le suivi de la mise en œuvre des recommandations du présent rapport puisse s'appuyer sur la mission Etalab. En particulier, concernant les données dont la mise à disposition en Open Data ne peut être envisagée en l'état, il est suggéré aux administrations productrices concernées de travailler conjointement avec l'administrateur général des données. Celui-ci pourrait les aider dans la définition de techniques et d'outils permettant de faciliter l'anonymisation.

Il paraît en effet souhaitable qu'il puisse aider les administrations productrices afin de faciliter la mise à disposition des données stratégiques en matière de logement.

G) Rendre visible les données existantes et projets de réutilisation des producteurs à travers data.gouv.fr

Enfin, comme les travaux du groupe l'ont montré, favoriser l'accès aux données publiques passe aussi par le fait de rendre plus visible celles mises à disposition.

Comme évoqué précédemment, les sources de données sur le logement, notamment du SOeS, du ministère en charge du logement et de l'INSEE, pourraient être rendues plus visibles et accessibles.

Il semblerait opportun de centraliser au maximum les données disponibles et de multiplier les entrées pour les faire mieux connaître. Les solutions possibles sont diverses. À ce sujet, le rapport CGEDD/IG-INSEE préconise d'ailleurs un site Internet unique.

En tout état de cause, il serait utile que le site du ministère en charge du logement puisse constituer un point d'accès, qui renvoie vers ce lieu unique (de type portail).

De plus, la plate-forme data.gouv.fr doit aussi servir d'entrée et d'appui à la diffusion des données. Outil au service du fonctionnement démocratique et de l'efficacité de l'action publique, la plate-forme permet

aux producteurs de données de mettre à disposition simplement et rapidement les données produites ou reçues dans le cadre de leur mission de service public dans un format réutilisable. Elle a vocation à offrir de la visibilité pour l'ensemble des acteurs publics et privés qui s'inscrivent dans une démarche d'ouverture et de mise à disposition des données. Chacun des producteurs, en plus de la mise à disposition sur sa propre plateforme ou sur une plateforme dédiée au logement, a tout intérêt à mettre ses données en Open Data sur la plateforme data.gouv.fr qui, par un simple référencement, lui donne de la visibilité et augmente la fréquentation de son site.

Par ailleurs, une fois les données mises à disposition, il est intéressant de constater la diversité ainsi que la qualité des usages qui sont fait par les réutilisateurs et par les producteurs.

Il s'agit donc d'un outil qui permet aux producteurs de données de suivre et valoriser les réutilisations qui ont été faites des différents jeux de données. Cartographies, datavisualisations, graphiques, applications web ou mobiles, tous les usages développés sur la base des données sont associés aux contenus qui ont permis leur développement.

L'ouverture des données a d'autant plus d'impact qu'elle s'inscrit dans le dialogue et la concertation. Échanger avec les utilisateurs de ses données, répondre à leur questions, recueillir leurs critiques ou leurs suggestions, c'est l'un des changements essentiels que produit cette démarche. C'est l'une des dimensions du portail data.gouv.fr qui autorise les utilisateurs, administrations, citoyens, chercheurs...et créateurs à dialoguer avec le producteur de la donnée, à partager des points de vue différents, à améliorer la donnée, à la croiser avec d'autres, voire à disposer des données qui ne sont pas celles de l'État.

Conclusion

Le groupe de travail dont le présent rapport relate les échanges a permis de confirmer, dans le secteur du logement, les enjeux à la fois démocratiques, économiques et d'amélioration de l'action publique de la mise à disposition des données publiques.

Les réflexions ont aussi montré la sensibilité et la complexité de la question. L'Open Data, si elle est de plus en plus connue, reste une logique et une culture encore nouvelles pour de nombreux acteurs. Il est donc important d'apporter des précisions sur son sens afin d'éviter les incompréhensions ou de lever les inquiétudes. Il convient également de montrer comment elle s'articule avec d'autres enjeux liés aux données, en particulier en termes de données effectivement collectées ou produites.

De plus, se pose également une question de moyens : concilier la logique d'ouverture des données publiques avec le respect de la vie privée empêche souvent de mettre à disposition directement celles produites par les administrations, les organismes rattachés ou par d'autres organismes dans le cadre d'une mission de service public. Des traitements préalables sont alors nécessaires.

Cette question se pose d'autant plus que la granularité de la mise à disposition est fine. Or, elle est au cœur du sujet dans le domaine du logement. Les réutilisateurs souhaitent un degré de détail le plus fin possible qui pose des difficultés en termes de recueil de données, de respect de la vie privée et de coûts de traitement d'anonymisation. Pour autant, là aussi, des avancées semblent possibles, à plus ou moins brèves échéances selon les bases.

Aussi, les réflexions du groupe ont permis de faire émerger des propositions sur les modalités de mise à disposition des données (format, visibilité,...). De plus, à l'heure où les systèmes d'information se développent fortement dans le cadre de l'action publique en matière de logement, il faut également signaler, d'un point de vue méthodologique, l'importance de penser la mise à disposition des données en même temps que la conception du système d'information.

Le groupe a par ailleurs montré la nécessité de l'articulation de la question de la mise à disposition des données avec le rôle des collectivités territoriales. En effet, ces dernières jouent un rôle de plus en plus important dans le champ du logement, que ce soit dans le cadre de compétences décentralisées, déléguées ou volontaires. D'une part, il apparaît important de penser au recueil et à la mise à disposition des données afférentes au moment de la délégation ou de la décentralisation de compétences. D'autre part, les collectivités locales sont des acteurs particulièrement intéressés par la mise à disposition de données publiques nationales afin de les aider à mener leurs propres politiques.

Avant de conclure ce rapport, il convient de souligner que ce groupe a constitué un lieu d'échanges riches entre acteurs très divers. Il a ainsi permis à des jeunes entrepreneurs de l'économie numérique d'échanger directement avec les producteurs de données au sein de l'administration, comme avec des statisticiens pour qui celles-ci sont la matière première pour différents usages. Ces échanges ont certainement été en eux-mêmes féconds, car ils ont permis une meilleure connaissance des enjeux et contraintes de chacun et le croisement de cultures. Il faut aussi souligner qu'ils ont contribué à diffuser auprès de différents relais ce que représente le mouvement de l'Open Data et dans quel contexte

international s'inscrit la politique portée sur ce sujet dans le cadre de la modernisation de l'action publique.

Enfin, les propositions faites par le groupe et synthétisées dans ce rapport constituent un point d'étape, qui dresse l'état des réflexions et apporte de premières pistes de travail, reflétant le plus souvent une volonté réelle des producteurs d'avancer en la matière. Ces derniers soulignent cependant les obstacles, par exemple sur l'anonymisation des données : l'administrateur des données qui vient d'être nommé pourrait contribuer à les aider à les lever. Les échanges ont vocation à se poursuivre, pour continuer d'avancer en matière d'ouverture des données publiques, et pour faire en sorte que cette ouverture alimente les réflexions et les pratiques dans le secteur du logement.

ANNEXES

ANNEXE 1

LETTRE DU PRESIDENT DU CONSEIL NATIONAL DE L'HABITAT

ANNEXE 2

Liste des participants au groupe de travail

ANNEXE 3

Liste des auditions et entretiens

ANNEXE 4

Cartographie des données

Annexe 1 – Lettre du Président du Conseil National de l'Habitat



MINISTÈRE DE L'ÉGALITÉ DES TERRITOIRES ET DU LOGEMENT

Conseil national de l'habitat

Paris, le 30 SEP. 2013

cnh@developpement-durable.gouv.fr

Le Président du Conseil national de l'habitat

à

Mesdames et Messieurs
les membres du Conseil national de l'habitat

Madame, Monsieur,

Le gouvernement a souhaité donner un élan fort au mouvement d'ouverture des données publiques (« open data »), au travers de l'effort de modernisation de l'action publique (MAP). L'animation de cette politique est confiée à la Mission « Etalab », au sein du Secrétariat Général de la Modernisation de l'Action Publique (services du Premier ministre).

Le Comité Interministériel pour la Modernisation de l'Action Publique (CIMAP), présidée par le Premier ministre, a réaffirmé le principe de gratuité de la réutilisation des données publiques et décidé de la mise en œuvre de débats thématiques ouverts afin de favoriser l'ouverture de données stratégiques à fort impact sociétal ou à fort potentiel d'innovation sociale et économique. L'une des thématiques retenues est le logement.

Il m'a été proposé que ce débat soit organisé au sein du Conseil national de l'habitat en se fondant notamment sur sa représentativité de la pluralité du secteur du logement. J'ai accepté cette proposition avec enthousiasme.

Ce débat, qui bénéficiera d'un secrétariat assuré par la mission Etalab, donnera lieu au premier trimestre 2014 à un rapport, remis au Premier ministre et à la ministre chargée du logement, qui formulera des recommandations en termes d'ouverture cohérente de données relatives au logement, de modalités et de conditions de réutilisation associées. Ce rapport précisera en outre les bénéfices attendus de la démarche.

Une présentation de la démarche, des modalités du débat et de ses objectifs sera faite par la mission Etalab, à l'occasion de la prochaine réunion du CNH, prévue le 28 octobre prochain.

.../...

Je vous propose qu'un groupe de travail amené à faire vivre le débat et ainsi faire émerger des propositions concrètes soit mis en place dès le mois de novembre, sous la présidence de Mme Géraldine CHALENCON, directrice générale de l'agence nationale pour l'information sur le logement. Je vous remercie de m'informer dès à présent, et au plus tard d'ici la prochaine réunion du CNH, de votre souhait de participer à ce groupe de travail. Vous pourrez à cet effet adresser votre réponse par courriel au secrétariat du CNH : cnh@developpement-durable.gouv.fr.

Je vous remercie pour votre participation à ce groupe de travail et vous prie de croire, Madame, Monsieur, à l'assurance de mes sentiments les meilleurs.



Le Président,



Michel PIRON

Annexe 2 – Liste des participants au groupe de travail

Membres du Conseil National de l'Habitat

Agence Nationale de l'Habitat (ANAH) :

- Philippe Cassagne/Pierre Sitkho

Association Nationale de Défense des Consommateurs et usagers (CLCV) :

- Edward Watteevw/Yasmina Heligon

Banque de France

- Marc Chazelas

Caisse des Dépôts et Consignations (CDC) :

- Michèle Pfimmer/ Josselin Kalifa

Confédération de l'Artisanat et des Petites Entreprises du Bâtiment (CAPEB) :

- Sophie Gourvenec, Sébastien Sauvaget

Confédération Française des Travailleurs Chrétiens

- Jean-Charles Jimenez

Conseil National de l'Ordre des Architectes :

- Sylvie Soulas-Perrot

Conseil Supérieur du Notariat :

- Christine Rey du Boissieu / Olivier Compère

Crédit Agricole SA :

- Barbara Delafenestre

Direction Générale de la Cohésion Sociale (DGCS) :

- Pierre Haenel, Bruno Lixi

Familles Rurales

- Jean-François Lemaître

Fédération Nationale de l'Immobilier (FNAIM) :

- Nathalie Ezerzer, Lionel Peynet

Fédération Française du Bâtiment (FFB) :

- Eric Quinton

Fédération Nationale des Coopératives HLM

- Isabelle Roudil

Habitat et Développement

- Nicolas Rousseau

Société de Gestion du Fonds de Garantie à l'Accession Sociale (SGFGAS) :

- François de Ricolfis

Union Nationale des Associations Familiales (UNAF) :

- Marilia Mendes

Union Sociale pour l'Habitat

- Martin de Bettignies

Autres participants :

- Pascal Berteaud, Stéphanie Carvalheiro, Claude Penicand, IGN
- Jean-Claude Driant, chercheur
- Jean-Louis Fournier, DVF
- Caroline Gerber, consultante, ancienne directrice de l'association Etudes Foncières
- Eduardo Lorrain, Kelquartier
- François Massot, Nelson Herbin, Regards Citoyens
- Adrien Paville, Site Home'n Go,

- Brice Person (blogueur, start-up sur le logement)
- Rayna Stamboliyska, Open Knowledge Foundation France (OKFN),

Pour le Ministère en charge du Logement :

Direction de l'habitat, de l'urbanisme et des paysages :

- Thierry Chatelain
- Laure Dexcidieux
- Inès Bouchikhi
- Carole Dahan-Chocron, CNH
- Yann Ludmann

Service de l'Observation et des Statistiques (SOeS) :

- Guillaume Houriez

Secrétariat Général :

- Charles-Guillaume Blanchon
- Marc Leobet
- Grégoire Philippon

Conseil Général de l'Environnement et du Développement Durable (CGEDD) :

- Jean-Paul Le Divenah, en tant que rapporteur du CNH
- Patrick Laporte et Jacques Friggit, pour la mission d'Inspection CGEDD/INSEE

Animation du groupe de travail :

- Géraldine Chalencon, Agence Nationale pour l'information sur le logement
- Romain Tales, Etalab

AUDITIONS

Le 15 janvier 2014 :

- Agence Nationale de l’Habitat (ANAH)
- Société de gestion du Fonds de garantie de l’accession sociale à la propriété (SGFGAS)
- Open Knowledge Foundation France
- Kelquartier

Le 13 mars 2014 :

- SOeS
- DHUP
- CGDD

Le 27 mars 2014 :

- DGFIP

Le 17 avril 2014 :

- INSEE

Le 30 juin 2014 :

- OLAP
- ANIL

ENTRETIENS

Le 27 janvier 2014

- Groupe DVF

Le 25 février 2014

- Fédération Nationale des Agences d’Urbanisme

Le 12 mars 2014

- Fondation Abbé Pierre

Annexe 4 – Cartographie des données

Cette cartographie a été établie à partir des fiches contenues à l'annexe 12 du rapport CGEDD-IG-INSEE³⁴. Elle en reprend la plupart des fiches, ciblées sur le champ du rapport (logement).

La dernière colonne présente, en plus de remarques de différents ordres, les évolutions en termes d'ouverture de données par les producteurs, telles qu'elles ont été présentées au cours des échanges avec ces derniers dans le cadre du travail mené.

³⁴ Inspection générale de l'Insee N°1.7.25 - Conseil général de l'Environnement et du Développement Durable N°009075-02 Rapport sur l'organisation du service statistique dans le domaine du logement - précité