



autorité de régulation
des communications électroniques,
des postes et de la distribution de la presse

RÉPUBLIQUE FRANÇAISE

MA CONNEXION INTERNET

Documentation de prise en main technique

03 décembre 2020

A decorative graphic in the bottom right corner consisting of a dense, overlapping pattern of thin, light grey lines that form a fan-like shape pointing towards the right.

ISSN n°2258-3106

1 Introduction

L'Arcep a mis à disposition des utilisateurs une version bêta de son site cartographique « [Ma Connexion Internet](#) », pour partager l'avancée de ses travaux et lancer une phase de co-construction avec l'ensemble des acteurs.

Pour cette version bêta, les données présentées sur le site ainsi que l'open data associé sont donc à prendre en l'état. Elles comportent un certain nombre de limitations, sur lesquelles devront encore porter des améliorations avant qu'une première version complète du site puisse voir le jour.

Cette documentation est un document technique qui vise à décrire précisément l'ensemble des données utilisées, des traitements réalisés et des éléments produits dans ce projet pour faciliter la prise en main de Ma connexion internet. Elle vise donc à alimenter avec transparence la phase bêta en donnant la matière nécessaire aux acteurs intéressés pour comprendre l'avancement des travaux de l'Arcep.

Pour participer à l'amélioration du site l'ARCEP vous invite à :

- vous approprier les documents publiés et répertoriés sur la page <https://www.arcep.fr/cartes-et-donnees/nos-cartes/documentation-technique-ma-connexion-internet.html>. Une présentation générale des fonctionnalités de Ma connexion internet et du projet est disponible sur cette même page ;
- remonter vos contributions à l'adresse mail contributionmci@arcep.fr ;
- vous approprier les algorithmes qui ont été mis à votre disposition dans le [dépôt GitHub "algo_maconnexioninternet"](#) ;
- reVISIONNER les 3 webinars qui vous ont été proposés en courant du mois de mai (disponibles sur la page documentation technique ci-dessus).

Par ailleurs, une foire aux questions est accessible à cette adresse : <https://www.arcep.fr/nos-sujets/ma-connexion-internet.html>.

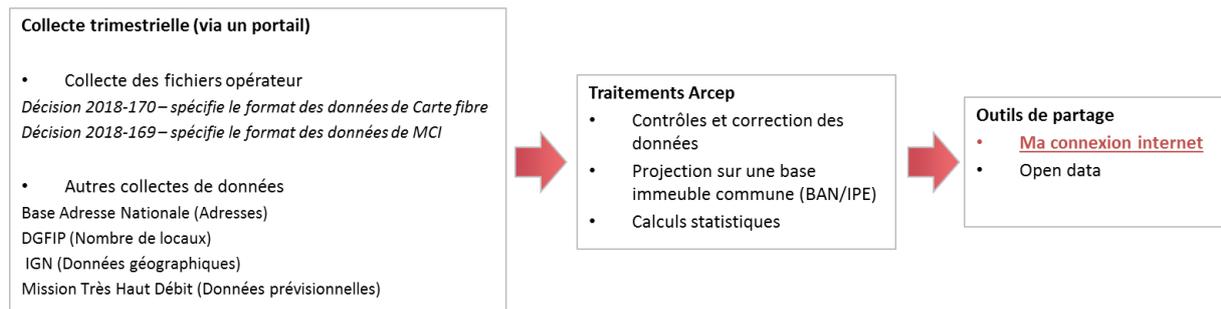
2 Plan du document

1	Introduction.....	2
2	Plan du document.....	3
3	Organisation générale du traitement des données pour la production des cartes de Ma Connexion Internet	4
4	Collecte des données	5
4.1	Cadre Juridique.....	5
4.2	Données collectées.....	5
5	Traitements réalisés sur les données	7
5.1	Constitution des référentiels d'immeubles et de locaux	7
5.2	Géocodage des données opérateurs.....	17
5.3	Calcul des éligibilités.....	18
5.4	Calcul des statistiques	29
6	Restitutions	35
6.1	Génération des cartes pour les informations affichées à l'immeuble et enjeux de fonctionnalités et de performances.....	35
6.2	Utilisation des cartes	36
6.3	Utilisation de l'open data	40
7	Annexe 1 – Données collectées.....	41
8	Annexe 2 – Données collectées.....	43
9	Annexe 3 – Technologies utilisées et quelques chiffres.....	45
9.1	Les technologies utilisées pour les différentes étapes :.....	45
9.2	Quelques chiffres.....	46
10	Annexe 4 – Rappel des limitations et améliorations prévues.....	47

3 Organisation générale du traitement des données pour la production des cartes de Ma Connexion Internet

Pour réaliser ces cartes, l'ARCEP a projeté les résultats des calculs d'éligibilité, obtenus à partir des données des opérateurs, sur un référentiel d'immeubles (constitué à partir de la Base Adresse Nationale et de l'IPE¹).

La création des cartes de Ma connexion internet se fait en plusieurs étapes, décrites en détail dans ce document:



- Collecte des données d'infrastructure (opérateurs d'infrastructures) et des données commerciales (opérateurs commerciaux) par mail (cf. partie 4 Collecte des données) ;
- Import des données dans la base de données Arcep et traitements sur les données lors de cet import en base de données (cf. partie 5 Traitements réalisés sur les données) ;
- Constitution d'un référentiel d'adresses, d'immeubles à partir de données sur les adresses et des fichiers des réseaux en fibre optique¹. Ces fichiers, produits par les opérateurs, sont régulièrement mis à jour et permettent de compléter le référentiel immeuble (cf. partie 5.1 Constitution des référentiels d'immeubles et de locaux);
- Calcul de l'éligibilité des immeubles technologie par technologie, issue de la projection des données opérateur d'éligibilité sur le référentiel immeuble qui sert de référentiel commun (cf. partie Calcul des éligibilités);
- Calculs statistiques de l'éligibilité des locaux, à partir des éligibilités des immeubles (cf. partie Calcul des statistiques);
- Copie dans une base de données cartographique des données d'éligibilité du référentiel immeuble et des données statistiques, pour affichage sur le site « Ma connexion internet » (cf. partie Restitutions) ;
- Génération des fichiers de l'open data, pour publication sur le site <https://www.data.gouv.fr/fr/datasets/ma-connexion-internet-beta/> (cf. partie Utilisation de l'open data).

¹ Fichiers « IPE » : Informations Préalablement Enrichies, description de l'ensemble du parc d'adresses FTTH ciblées, signées, en cours de déploiement ou déployées. Tout changement est notifié à J+1 par les opérateurs

4 Collecte des données

4.1 Cadre Juridique

Pour collecter les données nécessaires auprès des opérateurs, l'ARCEP se repose sur la décision n° 2018-0169 – relative aux contenus et aux modalités de publication de cartes de couvertures des réseaux et des services d'accès à internet en situation fixe, et aux modalités de transmission des informations sous-jacentes.

Ces décisions permettent à l'Autorité de :

- De collecter des **informations auprès des opérateurs** fournissant au public des services d'accès à internet en situation fixe et qui, directement ou à travers les groupes qui en détiennent le contrôle, disposent sur le marché de détail grand public d'au moins 1 000 abonnements actifs ;
- De collecter des **informations sur les technologies** filaires DSL et fibre optique (bout en bout et à terminaison en câble coaxial) ainsi que sur les technologies hertziennes satellitaires et hertziennes terrestres (THD Radio, HD Radio et 4G fixe) ;
- De collecter auprès des opérateurs commerciaux, pour chaque type de technologies (sauf le DSL) et d'infrastructures d'accès, des données relatives aux **débits descendants et aux débits montants**.
- De définir le contenu des informations à collecter (formats des fichiers, champs) ;
- De définir les modalités de transmission des informations ;
- De définir le calendrier des transmissions (trimestrielles).

4.2 Données collectées

Sources des données collectées

Pour produire les cartes de Ma connexion internet, l'Arcep utilise 3 grands ensembles de données :

- les données des opérateurs ;
- les données sur les adresses et les locaux ;
- les données de référence géographiques administratives (mailles administratives).

Les données collectées auprès des opérateurs proviennent :

- des opérateurs d'infrastructures : informations sur les réseaux, les zones de couverture ;
- des opérateurs commerciaux : informations sur leur présence sur les réseaux et les débits qu'ils proposent.

Les données sur les adresses et les locaux (autres que celles provenant des opérateurs) proviennent des 2 sources suivantes :

- Base Adresse Nationale, de licence ODBL : 24,6 millions d'adresses au T1 2020 ;
- Fichier des Propriétés Bâties (FPB), de la Direction générale des Finances publiques (DGFIP) : 23,8 millions d'adresses au T1 2020 soit 40,4 millions de locaux.

Les données de référence géographiques administratives proviennent des 2 sources suivantes :

- ADMIN EXPRESS de l'IGN : décrit le découpage administratif français ;
- COG (Code Officiel Géographique) de l'INSEE : recense les changements de codes INSEE liés notamment à des fusions de communes

Outil de collecte

Les données sont actuellement collectées par mail. Les agents de l'ARCEP déposent ensuite les fichiers des opérateurs par un portail dédié avec une vérification et une validation des données. Une fois que le portail de collecte sera ouvert sur le web, les opérateurs devront y déposer directement leurs fichiers.

Suite à ces contrôles, le portail de collecte renvoie des logs d'erreurs aux agents de l'ARCEP qui les transfèrent aux opérateurs pour qu'ils corrigent les fichiers erronés.

Une liste des fichiers collectés auprès des opérateurs est mise en annexe du présent document (annexe 1).

5 Traitements réalisés sur les données

5.1 Constitution des référentiels d'immeubles et de locaux

Le site cartographique « Ma connexion internet » se compose de deux cartes.

Sur la carte des débits, l'utilisateur peut visualiser des données d'éligibilité au niveau de chaque immeuble (points noirs aux niveaux les plus zoomés). Pour produire cette carte, un **référentiel d'immeubles** a donc été constitué.

Sur la carte des statistiques, l'utilisateur peut visualiser des taux de locaux éligibles pour différentes mailles administratives. Pour produire ces statistiques, il a été aussi nécessaire de constituer une **base de locaux associant un nombre de locaux à chaque immeuble** du référentiel immeuble.

Les sources premières d'informations initialement prévues par l'Arcep étaient :

- la BAN pour les adresses ;
- le Fichier des Propriétés Baties (FPB), provenant de la Direction Générale des Finances Publiques (DGFIP) pour le nombre de locaux pour les immeubles ;
- et l'utilisation des adresses, et plus particulièrement du géocodeur de la BAN pour effectuer les appariements entre ces deux jeux de données, ainsi qu'avec les données réseau des opérateurs.

Néanmoins, au regard du caractère incomplet des données de la BAN dans certaines zones du territoire, ainsi que des taux d'appariement insuffisants constatés entre les différentes données sur la base des adresses, ce schéma a été amené à évoluer, en particulier :

- en donnant un poids plus important aux informations issues des fichiers « IPE » des déploiements FttH lorsqu'elles existent ;
- en donnant un poids plus important aux informations issues du fichier « FPB » (données cadastrales) en dehors des zones très denses ;
- en effectuant une partie des appariements sur une base de proximité géographique.

Néanmoins, la BAN reste la base utilisée comme pivot pour associer les données entre elles sur la base des adresses, à partir de son géocodeur. La BAN est donc aussi la base privilégiée pour la recherche d'adresses sur le site.

5.1.1 Structure du référentiel immeuble utilisé par l'Arcep

Le référentiel immeuble de l'Arcep est constitué de deux tables : une table des adresses et une table des immeubles, plusieurs immeubles pouvant être reliés à la même adresse.

Les immeubles sont les entités affichées sur la carte des débits.

Deux critères sont primordiaux lors de la constitution du référentiel immeuble. Le référentiel doit être le plus complet possible (en agrégeant plusieurs bases existantes si nécessaire) et éviter au maximum les doublons (présenter plusieurs fois le même immeuble avec des points distincts).

Pour cela, et au regard des données effectivement disponibles et de leurs caractéristiques, l'Arcep a traité de façon distincte les 3 zones suivantes :

- les zones très denses (ZTD) ;

- les Zones IPE : zones géographiques à moins de 40 mètres d'un immeuble IPE², hors des ZTD ;
- les Zones hors IPE : zones géographiques à plus de 40 mètres d'un immeuble IPE, hors des ZTD.

En zones très denses

Dans les zones très denses, l'IPE n'est pas obligatoirement complet et peut ne pas recenser tous les immeubles. La base adresse est donc constituée en ajoutant à l'IPE toutes les adresses de la BAN qui n'ont pas été retrouvées dans l'IPE.

Base adresse (données de la table adresse) :

Elle est constituée de l'ensemble des adresses provenant de la BAN auxquelles sont ajoutées les adresses de l'IPE qui sont considérées absentes de la BAN car aucun appariement³ n'a été possible.

Il y a donc trois cas de figure dans la création de la base adresse :

Cas	Source de l'adresse
1/ Adresse BAN. Au moins un immeuble IPE a été apparié à cette adresse	BAN
2/ Adresse BAN – sans immeuble IPE apparié (22% des adresses)	BAN
3/ Adresse issue des immeubles IPE non appariées à la BAN (3% des adresses)	IPE

Référentiel immeuble (données de la table immeuble) :

Il est constitué de l'ensemble des immeubles provenant de l'IPE ainsi que de la BAN (un immeuble BAN est créé par adresse BAN) lorsque l'adresse ne présente pas d'immeubles IPE apparié (cas 2/ dans le tableau précédent).

Il y a donc deux cas de figures dans la création de la table immeuble :

Cas	Source de l'immeuble
Immeuble issu de l'IPE	IPE
Immeuble créés à partir des adresse BAN ⁴ – sans immeuble IPE appariés (21 % des immeubles)	BAN

Suite à une analyse détaillée, il apparaît que cette approche génère beaucoup de faux positifs, liés à des difficultés dans l'association d'adresses à des immeubles et à des différences d'adressage. Il est donc envisagé dans une prochaine version de ne s'appuyer plus que sur les IPE, dans la mesure où ceux-ci sont *a priori* désormais plus complets.

² Dans cette partie un immeuble IPE est un immeuble provenant du fichier IPE dont le statut n'est pas 'abandonné'

³ Un immeuble IPE est dit apparié à la BAN lorsque son adresse a été trouvée dans le référentiel d'adresse BAN. Cette appariement est fait par l'opérateur d'infrastructure ou lors d'un processus de géocodage réalisé par l'Arcep et décrit en partie 5.2 de ce document

⁴ Dans cette partie un immeuble BAN peut correspondre à un ou plusieurs bâtiments

En dehors des zones très denses (ou « zones moins denses ») et dans les zones référencées par l'IPE (zone située à moins de 40 mètres d'un immeuble IPE)

Pour les immeubles : primat à l'IPE

En dehors des zones très denses, l'obligation qui pèse sur l'opérateur d'infrastructure de recenser tous les immeubles à rendre raccordables dans son IPE conduit à ce que l'IPE recense, sauf omission, l'ensemble des immeubles de la zone. Au regard des défauts d'appariement qui peuvent résulter du croisement de plusieurs bases, il apparaît alors préférable de s'appuyer sur l'IPE pour constituer le référentiel d'immeubles.

Ce choix peut être fait dans les zones où l'IPE est disponible. Ces zones sont a priori délimitées par les ZAPM (Zone Arrière du Point de Mutualisation). L'IPE pourrait ainsi être retenu comme référentiel d'adresses et d'immeubles à l'intérieur des ZAPM existantes, et le cadastre en dehors de ces zones. Néanmoins, leur import n'ayant pas encore été prévu dans l'outil Ma Connexion Internet, les ZAPM n'y sont pas à date disponibles. Pour délimiter les zones de choix entre IPE ou Cadastre, une approximation a ainsi été effectuée, consistant à ne retenir des points du cadastre que lorsqu'ils sont suffisamment distants de tout point des IPE. En pratique, le seuil de [40m] a été retenu.

Le référentiel immeuble est constitué de l'ensemble des immeubles provenant de l'IPE. Aucun immeuble provenant de la BAN ou du cadastre n'est présent.

Pour les adresses : primat à la BAN

La base adresse (données de la table adresse) est constituée de l'ensemble des adresses provenant de la BAN, du cadastre et de l'IPE. Les adresses qui ont pu être appariées d'une part entre la BAN et l'IPE ou d'autre part entre la BAN et le cadastre, ne sont présentes qu'une seule fois - leurs informations de géolocalisation (adresse et emplacement géographiques) proviennent de la BAN. Les adresses IPE n'ayant pas pu être appariées aux adresses BAN, sont présentes dans la table adresse. Les adresses du cadastre n'ayant pas pu être appariées aux adresses BAN, sont aussi présentes dans la table adresse.

En dehors des zones très denses (ou « zones moins denses ») et dans les zones non référencées par l'IPE (zone située à plus de 40 mètres d'un immeuble IPE)

Pour les immeubles : primat au Cadastre

Le référentiel immeuble est constitué uniquement d'immeubles provenant du cadastre.

Pour les adresses : primat à la BAN

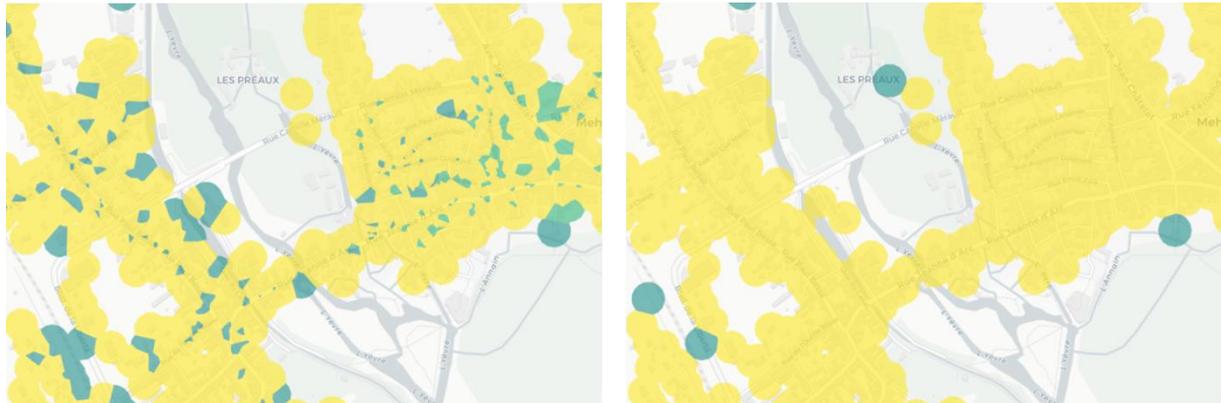
Les adresses et leurs informations de géolocalisation (adresse et emplacement géographiques) proviennent de la BAN. Les adresses du cadastre n'ayant pas pu être appariées aux adresses BAN sont aussi présentes dans la table adresse.

Précisions sur l'algorithme utilisé

En pratique, la constitution du référentiel immeuble revient à intégrer l'ensemble des immeubles de l'IPE (sauf immeubles 'abandonnés') et de n'ajouter que les immeubles BAN qui sont situées dans des communes de la zone très dense ou, hors de ces communes, de n'ajouter que les immeubles du cadastre qui se trouvent à plus de 40 mètres de tout immeuble IPE.

Résultats

Référentiel immeuble qui aurait été constitué à partir de tous les immeubles IPE et BAN Référentiel immeuble constitué suivant le principe décrit ci-avant



Exemple de Mehun-sur-Yèvre, commune intégralement fibrée : Les immeubles en jaune sont issus des IPE. Les immeubles en vert proviennent des adresses de la BAN du FPB pour lesquelles aucun immeuble IPE n'est apparié. Les immeubles verts ont été supprimés lorsqu'ils sont en zone fibrée, car ils sont probablement en doublons avec les immeubles IPE, ce qui crée des « faux négatifs » en termes d'éligibilité à la fibre.

5.1.2 Structure du référentiel de locaux (logements et locaux à usage professionnels) utilisé par l'ARCEP

Pour constituer ce référentiel de locaux, chaque immeuble du référentiel immeuble s'est vu associer un nombre de locaux.

Le Fichier des Propriétés Baties (FPB), provenant de la Direction Générale des Finances Publiques (DGFIP), est constitué des adresses et des coordonnées des immeubles du cadastre ainsi que du nombre de locaux associés.

Les immeubles issus de l'IPE se sont vu associer le nombre de locaux qui leur est associé dans le fichier FPB lorsqu'un appariement direct⁵ a été possible. En pratique, on apparie les immeubles IPE aux adresses BAN par un géocodage défini au 5.2 du présent document, puis on apparie les immeubles du cadastre aux adresses de la BAN, par un appariement défini dans la suite de cette partie.

⁵ 3 tours d'appariement, voir ci-dessous.

Les immeubles issus de l'IPE, qui n'ont pas pu être appariés à un immeuble du cadastre, se sont vu associer le nombre de locaux qui leur est associé dans le fichier IPE.

La BAN ne donnant pas d'indication sur le nombre de locaux, les immeubles issus de la BAN (zone très dense) se sont vus associer les données sur les locaux du FPB. En pratique, on apparie les adresses du cadastre aux adresses de la BAN par un appariement large⁶ défini dans la suite du document.

Or il n'a pas été possible d'apparier toutes les adresses du cadastre à la BAN (le taux d'appariement est de 78% des adresses du cadastre). Pour chaque adresse du cadastre qui n'a pas été appariée, on réalloue le nombre de locaux indiqué dans le FPB, à l'immeuble le plus proche provenant de notre référentiel situé dans la même commune et à moins de 200m.

Suite à cette réallocation, environ 5% des adresses du cadastre ne sont toujours pas appariées à la BAN.

Appariement direct des adresses en 3 tours

Le processus d'appariement suivant a pour objectif d'associer les adresses du FPB aux adresses équivalentes de la BAN. Pour maximiser le nombre d'adresses appariées, nous réalisons 3 tours de d'appariement sur les données du FPB:

Tour	Action
1 ^{er} tour – égalité des quadruplets	Egalité sémantique de la concaténation des 4 champs suivants : code insee, id fantoir, numéro de voie et répétition.
2 ^e tour – appariement double (géographique et sémantique)	Recherche d'une adresse BAN en renforçant les critères de proximité des coordonnées géographiques et en relâchant légèrement les critères de champs sémantiques d'adresses. L'égalité du quadruplet est remplacée par un critère sur la distance de levenshtein : <ul style="list-style-type: none">- Pour les géométries BAN se trouvant à moins de 70 mètres d'une adresse FPB, la distance de levenshtein doit être inférieure ou égale à 7 ;- Pour les géométries BAN se trouvant entre 70 mètres et 500 mètres d'une adresse FPB, la distance de levenshtein doit être inférieure ou égale à 3.
3 ^e tour – amélioration sémantique puis appariement double (géographique et sémantique)	Applications d'une correction sémantique sur le champ répétition (tous les termes sont remplacés par « bis »), puis application du même appariement qu'au 2 ^e tour

Réallocation

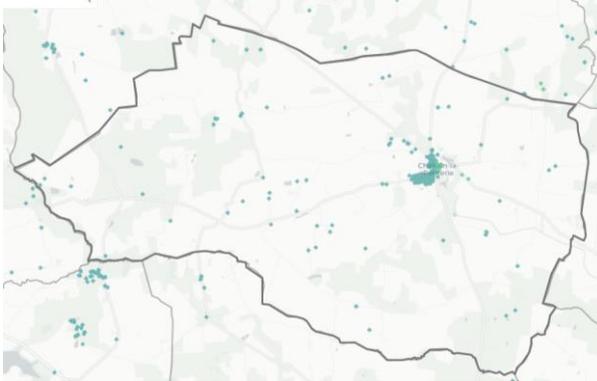
⁶ 4 tours d'appariement : l'appariement directe et la réallocation, voir ci-dessous.

Pour chaque adresse du cadastre qui n'a pas été appariée, on réalloue le nombre de locaux provenant du FPB à l'immeuble le plus proche provenant de notre référentiel situé dans la même commune et à moins de 200m.

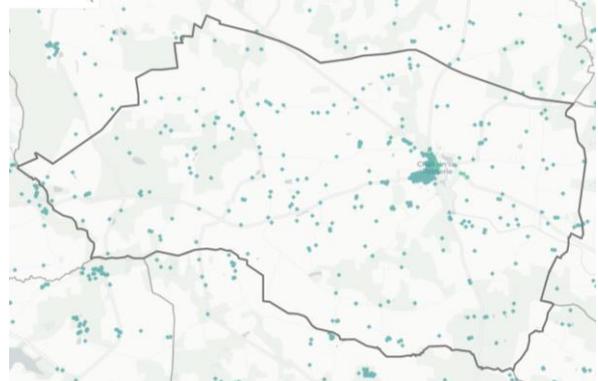
Evolution effectuées

Le référentiel immeuble est dorénavant plus complet, particulièrement en zones rurales. Ci-dessous l'exemple de la commune de Challain-la-Potherie montre qu'une grande quantité d'immeubles ont été ajoutés avec les données du FPB (cadastre). Il s'agit principalement des immeubles ayant une « pseudo-adresse », c'est-à-dire qui ne présentent pas de nom de rue, non présents dans la BAN

Base immeuble du T3 2019



Base immeuble du T1 2020



Exemple de Challain-la-Potherie : Les points sur la carte de gauche représentent les immeubles présents sur les cartes du site cartographique lors de la publication des données du T3 2019 (issus de la BAN) et les points de la carte de droite représentent les immeubles présents sur les cartes du site cartographique lors de la publication des données du T1 2020 (issus du cadastre)

Une modification non prévue de nomenclature de la DGFIP pour catégoriser les locaux du FPB, n'avait pas permis d'intégrer les locaux professionnels du fichier FPB dans la première version de la carte et de l'open data publiés en avril 2020. 2,9 millions de locaux professionnels ont donc ainsi pu être récupérés dans la publication du T1 2020. Le liste des catégories DGFIP retenues pour constituer le référentiel de locaux est mise à disposition en annexe.

Un nettoyage des fichiers IPE a aussi permis de supprimer 2,1 millions de locaux dans des immeubles en doublons ou fantômes, qui avaient un impact sur les statistiques et sur la représentation des débits :

- soit en dédoublant toutes les adresses avec des informations différentes (exemple des cas de doublons Orange-SFR en zone AMII),
- soit en perturbant le calcul des Voronoï lorsque les points sont strictement superposés (cas des changements d'identifiants).

Limitations et améliorations prévues sur la carte

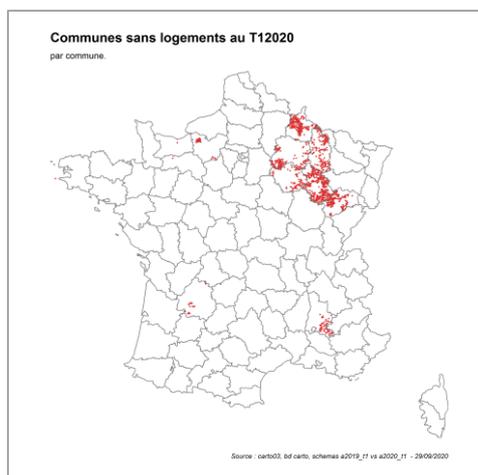
Communes sans immeuble

Le cadastre n'étant pas entièrement numérisé, certaines adresses n'ont pas de coordonnées géographiques (plus de 200 000). Bien que présentes dans le référentiel de bâtiments de Ma connexion internet, ces adresses n'apparaissent pas sur la carte et ne sont pas comptabilisées dans les calculs statistiques.

1 039 communes (voir localisation sur la carte ci-dessous), essentiellement situées dans la région du Grand Est, ne possèdent que des adresses non numérisées et sont donc totalement vides.

500 communes supplémentaires ont récemment été numérisées dans le cadastre et pourront être directement intégrées dans une prochaine mise à jour.

D'ici là et pour les communes qui ne sont toujours pas numérisées, cette limitation pourra être levée en privilégiant les adresses de la BAN dans les 1 039 communes.



Localisation des communes dont le cadastre n'est pas encore numérisé

Augmentation du nombre de locaux

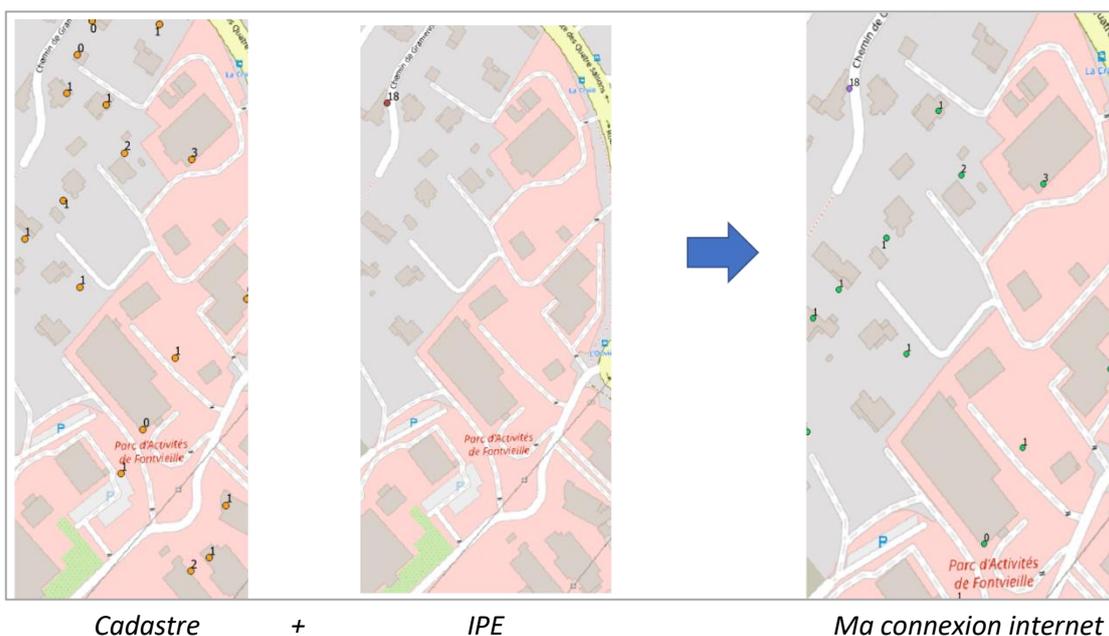
Le référentiel de locaux estimé à peu plus 43 millions au T1 2020 est surévalué du fait de doubles comptages d'immeubles et de locaux. Ces doubles comptages peuvent être réduits en améliorant les algorithmes, mais un nombre significatif de doubles comptages sont issus d'incohérences entre les bases utilisées et ne pourront être corrigés qu'avec une meilleure cohérence entre les adresses de l'IPE, de la BAN et du cadastre.

	INSEE	FPB	MCI T1 2020 (FPB & IPE)
Nombre de locaux (en millions)	37,2	40,4	43,1

Double comptage d'immeubles

Les adresses du cadastre et de l'IPE sont associées (géocodage) aux adresses BAN. La BAN sert donc de pivot pour relier un immeuble IPE et une adresse cadastrale. Or seules 78% des adresses du cadastre s'associent à une adresse BAN.

Pour éviter des doubles comptes d'immeubles, il a été décidé de récupérer les immeubles du cadastre non associés à la BAN uniquement lorsqu'ils sont situés à plus de 40m de l'immeuble IPE. Néanmoins dans certaines zones, et comme le montre l'exemple ci-dessous, cette distance n'est pas toujours suffisante pour éviter de comptabiliser à tort des doublons.



Dans l'exemple ci-dessus, l'immeuble IPE en haut à gauche, possède l'ensemble des locaux de la zone industrielle. Le cadastre est plus précis et donne la position de chaque immeuble de la zone industrielle ainsi que leur nombre de locaux. Mais ces immeubles n'ont pas les mêmes adresses dans l'IPE et dans le cadastre, l'association n'a donc pas réussi et les immeubles sont comptabilisés deux fois.

Double comptage de locaux

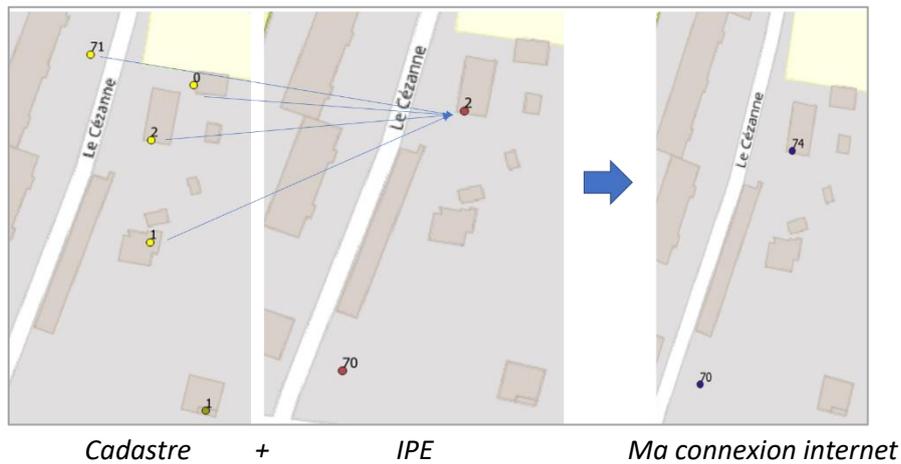
En raison de discussions en cours avec certains opérateurs sur la capacité à placer en open data les données de décompte de locaux issues de leur IPE, l'Arcep effectue à ce stade une substitution, autant que possible, des locaux issus de l'IPE par des données provenant de la DGFIP.

Néanmoins, ce remplacement reste tributaire de la capacité à associer des bâtiments à une adresse BAN. Certaines associations se basant sur des critères de distance entraînent des doubles comptages, comme le décrit l'exemple ci-dessous.

Cette limitation pourrait être levée :

- en ne remplaçant que les immeubles qui présentent un nombre de locaux proche entre le référentiel IPE et le référentiel FPB ;
- ou si les discussions en cours avec certains opérateurs sur la capacité à placer en open data

les données de décompte de locaux issues de leur IPE aboutissent.



Dans l'exemple ci-dessus, nous pouvons voir que l'adresse du FPB ayant 71 locaux (en haut à gauche) est associée à l'immeuble IPE le plus proche (immeuble avec 2 locaux dans l'IPE). Or l'immeuble avec le nombre de locaux correspondant est situé plus bas dans l'IPE. Mais les adresses étant très différentes, aucune association sur l'adresse n'a pu aboutir et l'association a donc été effectuée sur le seul critère de distance (au plus proche). Il y a donc un double compte de ces locaux.

Incohérences entre les bases sur le nombre de locaux d'un immeuble

L'IPE et le FPB nous fournissent un nombre de locaux à l'immeuble. Il s'avère que les deux bases peuvent associer un nombre de locaux très différents pour les mêmes immeubles (parfois l'IPE associe un nombre de locaux plus important pour un immeuble donné, parfois c'est l'inverse).



Dans l'exemple ci-dessus, nous pouvons voir que les adresses du FPB associent plus de locaux que l'IPE pour ces immeubles.

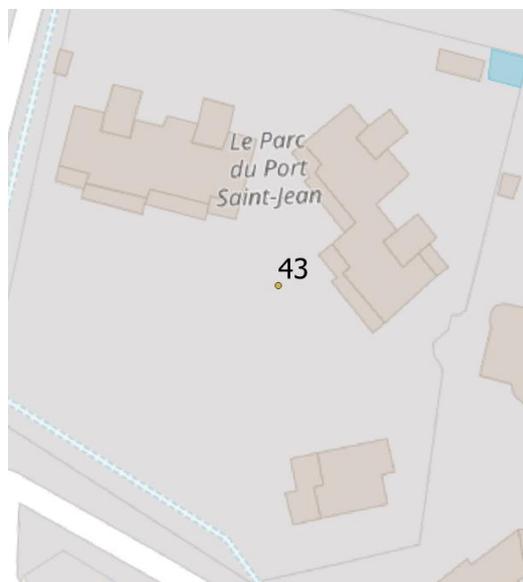
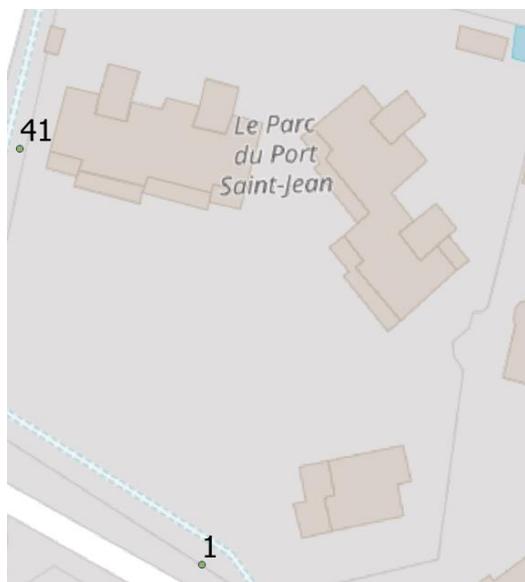
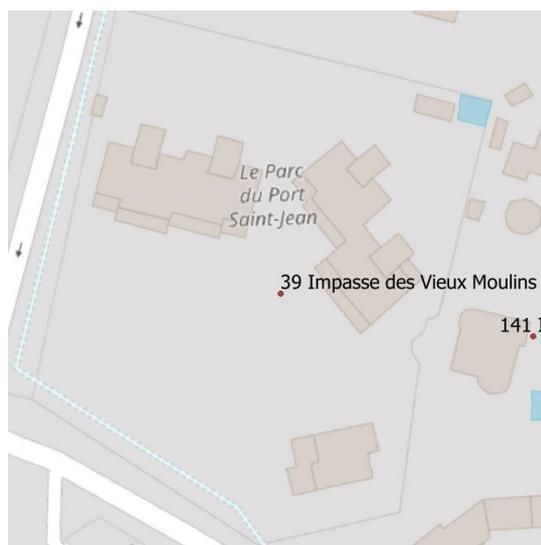
Incohérences entre les bases sur le nombre de locaux d'une adresse

L'IPE et le FPB nous fournissent un nombre de locaux à l'immeuble. Cet immeuble est relié à une adresse. Il apparaît que plusieurs adresses peuvent être associées aux mêmes immeubles ou groupe d'immeubles. Les locaux sont alors associés à des adresses différentes dans les deux bases.

IPE



Cadastre (avec locaux du FPB)



Dans l'exemple ci-dessus, nous pouvons voir que l'IPE associe 41 locaux à l'adresse « 2 rue andré bellon » et un seul local à l'adresse 31 chemin des vieux moulins, qui semblent représenter le même groupe d'immeubles, alors que le FPB ne repertorie pas l'adresse « 2 rue andré bellon » et associe 43 locaux à l'adresse « 39 impasse des vieux moulins ».

5.2 Géocodage des données opérateurs

5.2.1 Principe

Les données transmises par les opérateurs comportent des informations sur l'emplacement des adresses (pour la fibre et le câble) ou sur l'emplacement des Point de Concentration (pour le cuivre et le câble).

Il est nécessaire d'apparier ces informations réseaux sur un référentiel unique d'adresses (commun entre toutes les technologies). Cette étape est appelée ici « géocodage ».

La base adresse qui a été retenue dans Ma Connexion Internet pour réaliser cet appariement est la Base Adresse Nationale (BAN). La BAN sert ainsi de référentiel pivot pour associer les différentes données réseaux entre elles, ainsi qu'aux différents référentiels utilisés (BAN, FPB). Le géocodage est donc le procédé qui apparie ici une information réseau donnée par l'opérateur (à l'adresse ou l'immeuble) et l'adresse correspondante dans la Base Adresse Nationale.

Les lignes cuivre, les adresses coax et les immeubles fibres sont géocodés via une instance locale de l'API de la BAN en utilisant les paramètres suivants :

- Numéro de voie
- Répétition (Bis, Ter, a, b, c, ...)
- Type et nom de rue
- Code INSEE de la commune
- Coordonnées de la donnée réseau (x,y)

5.2.2 Critères appliqués

L'API de la BAN renvoie une liste d'adresses de sa base susceptibles de correspondre à l'adresse recherchée. Ces adresses sont notées selon des critères sémantiques appliqués aux champs textuels des adresses, ainsi que des critères de proximité géographiques des coordonnées. L'adresse BAN sélectionnée doit remplir les critères suivants :

- Avoir un score supérieur à 0,6 (l'adresse avec le meilleur score est sélectionnée en cas de candidates multiples)
- Le type de l'adresse BAN doit être 'house number' (car les autres types ne correspondent pas à des adresses)
- La distance entre les coordonnées de l'adresse BAN et les coordonnées fournies par l'opérateur (adresses pour les technologies fibre et coax, point de concentration (PC) pour la technologie cuivre) doit être inférieure à une distance maximum définie par l'Arcep qui dépend de la technologie (voir tableau ci-dessous).

Technologie	Type de coordonnées	Distance maximum	Explication distance
Cuivre	Au PC	1 5000 m	Les PC peuvent être très

			éloignés de l'adresse en zone rural.
Fibre	À l'immeuble	150 m	Les immeubles peuvent être éloignés de leur adresse de quelques centaines de mètres
Coax	À l'adresse	150 m	Valeur sélectionnée par similarité avec ce qui se fait pour la fibre

Distance maximum entre la coordonnée fournie par l'opérateur et la coordonnée BAN associée

5.2.3 Géocodage en 3 tours

Pour maximiser le nombre d'adresses géocodées, nous réalisons 3 tours de géocodage sur les données des opérateurs non géocodées :

Tour	Action
1 ^{er} tour – géocodage simple	Géocodage du champ adresse tel que fourni par l'opérateur
2 ^e tour – amélioration sémantique	<ul style="list-style-type: none"> - Applications de corrections sémantiques sur le champ adresse principalement pour reconstituer des abréviations (ex : « r » est remplacé par « rue ») - Puis, nouveau géocodage du champ adresse
3 ^e tour – géocodage « inverse »	Uniquement pour la fibre et le coax. Recherche d'une adresse BAN en renforçant les critères de proximité des coordonnées géographiques et en relâchant légèrement les critères de champs sémantiques d'adresses.

5.3 Calcul des éligibilités

5.3.1 Principe général

Les calculs d'éligibilité visent à déterminer l'éligibilité d'un immeuble à une technologie⁷ grâce aux données transmises par les opérateurs d'infrastructures et les opérateurs commerciaux (données ponctuelles pour les technologies filaires et données surfaciques pour les technologies hertziennes).

Ils se font en deux temps :

- tout d'abord au travers du traitement des données des opérateurs d'infrastructure, pour venir associer les extrémités de réseau à des immeubles ;

⁷ Immeuble : dans le sens de la base adresse et immeuble créée par l'Arcep – voir la partie dédiée pour plus d'informations

- ensuite au travers du traitement des données des opérateurs commerciaux, pour venir déterminer quel opérateur commercial utilise quel réseau de quel opérateur d'infrastructure, et en proposant quel service dessus.

Ces traitements permettent *in fine* de remplir la table d'éligibilité qui est le « plus grand dénominateur commun » entre toutes les technologies - avec les paramètres suivants :

- Identifiant de l'immeuble (issu de notre base immeuble)
- Technologie
- Opérateur
- Débit montant
- Débit descendant
- Limitation de données (si l'opérateur commercial inclut dans son offre une limitation du volume des données échangées)
- Saturation (pour les technologies sans fil, traduit une interruption temporaire de commercialisation de l'offre car une zone est saturée en raison d'un nombre déjà trop important d'utilisateurs pour continuer à proposer une offre commerciale)

5.3.2 Technologies filaires

Les calculs de l'éligibilité des technologies filaires sont réalisés en 2 étapes :

- Dans un premier temps, il est nécessaire de rattacher les données sur les réseaux transmises par les opérateurs d'infrastructures (données ponctuelles), aux immeubles du référentiel immeuble ; il s'agit d'associations point à point, puisque les données transmises par les opérateurs d'infrastructures filaires sont ponctuelles ;
- Dans un second temps il est nécessaire d'associer à ces immeubles les débits proposés par les opérateurs commerciaux.

5.3.2.1 Cuivre

Les fichiers de données sur le réseau cuivre contiennent des informations par ligne d'abonné. Pour chaque ligne d'abonné, les informations contiennent :

- l'adresse de la ligne (information textuelle, mais sans coordonnées)
- le PC de rattachement
- les coordonnées de ce PC
- l'atténuation au niveau de ce PC

A contrario, ces informations ne contiennent ni les coordonnées du local relié par cette ligne, ni l'atténuation de la ligne au niveau de ce local.

Etape n°1. Association des données réseaux aux immeubles du référentiel

Cette étape s'effectue en 2 temps :

1. Suite au géocodage des données cuivre, un code BAN est associé à la plupart des adresses des lignes cuivre des fichiers opérateurs. Ce code BAN permet d'apparier les données d'infrastructure aux immeubles du référentiel immeuble et ainsi de calculer un affaiblissement au niveau de chaque immeuble suivant la méthode développée ci-après.

2. Pour les immeubles qui n'ont pas été associés à une ligne cuivre par le géocodage, mais qui ont néanmoins un nombre de locaux non nul et qui ne se situent pas en Zone Locale Immeuble Neuf⁸, ils sont traités comme suit :
- On cherche un immeuble proche, en supposant qu'ils sont rattachés au même PC => On associe à l'immeuble l'affaiblissement du PC⁹ de l'immeuble le plus proche (géocodé ou non), situé dans un rayon maximal de 500 mètres ;
 - Si aucun immeuble n'a été trouvé, on cherche un PC proche, mais dans un rayon plus grand => on associe à l'immeuble l'affaiblissement du PC le plus proche, situé dans un rayon maximal de 2 500 mètres ;
 - Pour les autres immeubles, l'ARCEP associe un affaiblissement null. Ce mécanisme pour les immeubles restants pourrait être revu dans la version post beta.
- On calcule ensuite l'affaiblissement à l'immeuble avec la méthode développée ci-après.

Etape n°2. Calcul des affaiblissements à l'immeuble.

La donnée d'affaiblissement transmise par l'opérateur d'infrastructure (Orange) est fournie, pour chaque ligne, au niveau du Point de concentration (PC).

L'ARCEP associe à chaque immeuble un affaiblissement au PC, suivant la méthode indiquée à l'étape 1 de la présente section.

Pour calculer l'affaiblissement à l'immeuble, l'ARCEP récupère l'affaiblissement au PC de la ligne, et y ajoute une estimation de l'affaiblissement supplémentaire apporté par le tronçon de la ligne situé entre le PC et l'immeuble, calculé suivant la formule suivante :

$$\text{affaiblissement_immeuble} = \text{affaiblissement_PC} - 3.0 + ((\text{longueur_pc_immeuble} * \sqrt{2}) + \text{longueur_technique}) * 15/1000$$

Avec :

- affaiblissement_PC : Valeur d'affaiblissement au PC, déterminé suivant la méthode expliquée à l'étape1 de la présente section.
- 3.0 dB : affaiblissement ajouté par Orange par défaut à la valeur mesurée au PC pour tenir compte de l'affaiblissement supplémentaire en aval du PC. Nous enlevons donc cette valeur dans le calcul de l'affaiblissement
- longueur_pc_immeuble: distance à vol d'oiseau, en mètres, entre le PC et l'immeuble (géocodé) de la ligne cuivre
- longueur_technique : longueur dans le NRA et en partie privative, nous la fixons par défaut à 100m
- 15/1000 : atténuation par mètre de la ligne cuivre(en dB/m)

⁸ Zone Locale immeuble Neuf : ensemble d'immeubles de construction récente identifiés par Orange dont les logements sont fibrés mais pas cuivrés.

⁹ L'affaiblissement du PC est calculé en faisant la moyenne des affaiblissements au PC des lignes actives qui lui sont rattachées. Dans le cas où le PC ne comporte que des lignes inactives, l'affaiblissement du PC est calculé en prenant l'affaiblissement au PC maximal de toutes les lignes inactives qui lui sont rattachées.

Etape n°3. Détermination de la présence des OC et des technologies DSL proposées

Les opérateurs commerciaux transmettent à l'Arcep leur présence au niveau de NRA ainsi que la technologie utilisée (ADSL ou VDSL).

Cette présence permet de déterminer les immeubles pour lesquels ces opérateurs commerciaux proposent une offre (l'ensemble des lignes reliées au NRA sur lesquels l'OC est présent sauf les lignes sans service et les lignes reliées à un PC sans service).

Le débit disponible pour l'utilisateur est calculé à partir de ces informations et de l'atténuation de la ligne ; il n'est pas fourni par l'opérateur. La méthodologie de calcul est ainsi la même pour tous les opérateurs. Le calcul est effectué suivant la méthode développée ci-dessous (la fonction utilisée dans le calcul est différente entre les technologies DSL).

Etape n°4. Calcul des débits à l'immeuble

Après détermination de l'affaiblissement à l'immeuble, un second calcul permet de calculer le débit théorique proposé à l'immeuble.

La formule de calcul des débits cuivre dépendant de la technologie utilisée au niveau du NRA (ADSL ou VDSL2). L'Arcep croise les déclarations de l'opérateur d'infrastructure (Orange), sur les technologies proposées sur ses NRA, avec les déclarations des opérateurs commerciaux présents sur ces NRA (technologie sur laquelle l'OC propose des offres).

Suivant les technologies déclarées par les opérateurs (OI et OC) les classes de débits cuivre sont alors calculées suivant les formules des tableaux ci-dessous¹⁰.

Débits descendants :

<u>Seuils VDSL</u>	<u>Seuils ADSL</u>	<u>Débits correspondants</u>
[0-11 dB[100 Mbits/s – 20 Mbits/s
[11-34 dB[[0-34 dB[25 Mbits/s – 6 Mbits/s
[34-49 dB[[34-49 dB[10 Mbits/s – 2 Mbits/s
[49-78 dB[[49-78 dB[4 Mbits/s – 0,512 Mbits/s
[78 dB-...]	[78 dB-...]	Inéligible HD

La classe de débit « Inéligible HD » est associée aux immeubles pour lesquels l'affaiblissement des lignes est null.

Débits montants :

<u>Seuils VDSL</u>	<u>Seuils ADSL</u>	<u>Débits correspondants</u>
[0-14dB[10 Mbits/s – 2 Mbits/s
[14-78 dB[[0-78 dB[4 Mbits/s – 0,512 Mbits/s
[78 dB-...]	[78 dB-...]	Inéligible HD

¹⁰ On associe la classe de débit « inéligible HD » aux immeubles ayant un affaiblissement null.

La classe de débit « Inéligible HD » est associée aux immeubles pour lesquels l'affaiblissement des lignes est null.

Etape n°5. Cas particuliers des lignes avec un affaiblissement au PC inférieur ou égale à 78dB

En cohérence avec les critères d'éligibilité utilisés par les opérateurs, sont affichées éligibles HD, c'est-à-dire que le client peut souscrire à un abonnement en DSL, toutes les lignes pour lesquelles la valeur de l'affaiblissement au PC (point de concentration) est inférieur ou égale à 78dB, quand bien même l'atténuation au niveau de l'immeuble est supérieure. Le débit associé est le débit plancher ([0,512-4[en débit montant et descendant).

Etape n°6. Calcul des débits pour les statistiques sur les locaux

Après détermination des classes de débits à l'immeuble, une correspondance de classes permet de déterminer les classes de débit pour les statistiques effectuées sur les locaux de ces immeubles. Les correspondances entre les classes de débits (montants et descendants) est données dans le tableau ci-dessous.

Classe(s) de débit potentielle(s) pour l'affichage sur la carte	1 Gb/s et +	100 Mb/s à 1 Gb/s	20 Mb/s à 100 Mb /s	6 MB/s à 25 Mb/s	2 Mb/s à 10 Mb/s	512 kb/s à 4 Mb/s	Inéligible HD
(classes de débits pour les statistiques)	idem	idem	30 à 100	8 à 30	3 à 8	512 à 3	idem
n° de la classe	7	6	5	4	3	2	1

Tableau de comparaison des classes de débit des deux cartes du site cartographique « Ma connexion internet »

Impacts du calcul de l'éligibilité cuivre sur les classes de débits des cartes du site cartographique « Ma connexion internet »

Les deux cartes du site cartographique « Ma connexion internet » n'affichent pas les mêmes classes de débits (cf tableau précédent).

Les classes de débits de la carte des débits inférieures à 100mbits/s se recouvrent (exemple la classe 2 à 10 mbit/s recouvre la classe 6 à 25 mbits/s entre 6 et 10 mbits/s).

Le choix de ces recouvrements provient des incertitudes de calcul du débit cuivre. En effet, le calcul de l'atténuation théorique de la ligne constitue en pratique un prédicteur très imparfait du débit disponible, car de nombreux perturbateurs peuvent s'ajouter. Pour une atténuation donnée, le débit disponible présente une grande dispersion (cf.graphique ci-dessous).

Le mode « débit » présente une information individuelle, pour un utilisateur. Les classes du mode débit permettent d'assurer avec une bonne fiabilité la disponibilité réelle des débits présentés, y compris aux limites.

Le problème ne se pose pas dans la carte des statistiques, car l'agrégation vient en grande partie limiter les effets de la dispersion observée dans la relation entre affaiblissement et débit.

Paramètres

Les données suivantes permettent donc de déterminer l'éligibilité du cuivre :

Éléments de réseau	Paramètres
NRA_OI (Nœud de Raccordement Abonné – Opérateur d'Infrastructure)	<ul style="list-style-type: none"> - Ouverture au VDSL - Ouverture à l'ADSL2+
LD (Ligne Principale cuivre)	<ul style="list-style-type: none"> - PC de rattachement de la ligne - Code BAN de l'adresse rattachée à la ligne - Coordonnées de l'adresse BAN
PC	<ul style="list-style-type: none"> - Code PC - Affaiblissement du PC¹¹
NRA_OC (Nœud de Raccordement Abonné – Opérateur Commercial)	<ul style="list-style-type: none"> - Présence commerciale de l'OC - Type de présence (dégrouper/bitstream) - Ouverture au VDSL (si présence en dégroupage)
PC sans service (Point de Concentration)	<ul style="list-style-type: none"> - Code PC
LD sans service (Ligne Principale cuivre)	<ul style="list-style-type: none"> - Numéro de la ligne

Evolutions effectuées

Le calcul d'atténuation réalisé par l'Arcep prend pour hypothèse d'ajouter un affaiblissement correspondant à la distance entre le PC et la terminaison de la ligne, à celui observé au PC afin d'être au plus proche de la réalité terrain concernant les débits.

Néanmoins, le critère d'affaiblissement inférieur à 78 dB pour déterminer l'éligibilité HD est dorénavant apprécié au PC, comme habituellement par Orange, au lieu de l'être en bout de ligne comme précédemment. Cette évolution a entraîné une éligibilité supplémentaire au DSL pour 0,5 millions de locaux jusqu'ici inéligibles.

¹¹ Cette valeur est calculé par l'Arcep. Il s'agit de l'affaiblissement moyen au PC des lignes actives reliées à ce PC. Si le PC n'a que des lignes inactives, la valeur retenue est l'affaiblissement maximum des lignes inactives reliées à ce PC

5.3.2.2 Fibre

Le calcul de l'éligibilité fibre présente moins de complexité que celui du cuivre, car les données des immeubles ont été directement constituées à partir des IPE. Par ailleurs le niveau de service fourni est suffisamment indépendant des caractéristiques de la ligne pour qu'il n'en soit pas tenu compte.

Etape n°1. Détermination des immeubles raccordables ou raccordables sur demande et ouverts aux services commerciaux

Les immeubles IPE du référentiel immeuble proviennent directement des données opérateurs d'infrastructure.

Seuls les immeubles au statut « déployés », pour lesquels le PM est déployé et qui sont ouverts aux services commerciaux (date de mise en service commerciale dépassée à la fin du trimestre, ou non renseignée) peuvent être considérés comme éligibles.

Etape n°2. Détermination de la présence des OC et des débits proposés.

Les opérateurs commerciaux déclarent leur présence au niveau des points de mutualisation (PM) et indique les débits commerciaux qu'ils proposent sur la zone arrière.

Ainsi, tout immeuble satisfaisant les critères ci-dessus et pour lequel au moins un OC est présent au PM de rattachement de l'immeuble est éligible à la fibre.

Face aux limitations actuelle de la collecte de données, qui n'englobe pas tous les FAI (OC), un opérateur commercial fictif a été créé par l'Arcep sur tous les PM déployés sur lesquels aucun opérateur commercial n'était présent, dans la mesure où les lignes fibre sont quasiment toujours éligibles à au moins une offre commerciale.

Ces informations permettent à l'Arcep de déterminer la présence des OC et les débits proposés, pour chaque immeuble.

Paramètres

Les données suivantes permettent de déterminer l'éligibilité de la fibre:

Éléments de réseau	Paramètres
PM_OI (Point de Mutualisation - Opérateur d'Infrastructure)	<ul style="list-style-type: none">- État du PM- Date de mise en service commerciale
Immeuble	<ul style="list-style-type: none">- État de l'immeuble
PM_OC (Point de Mutualisation – Opérateur Commercial)	<ul style="list-style-type: none">- Présence commerciale de l'OC au PM- Débit montant et descendant
Immeubles_sans_service_OC	<ul style="list-style-type: none">- Code de l'immeuble- Reference du PM

Evolutions effectuées

L'Arcep récupère dorénavant des données de la part de 16 opérateurs commerciaux ce qui permet d'être plus complet sur les offres proposées par les opérateurs. Néanmoins, environ 600 000 locaux

ne sont actuellement couverts par aucun opérateur commercial.

5.3.2.3 Câble

Etape n°1. Association des données réseaux aux immeubles du référentiel

Les données transmises par les opérateurs d'infrastructures du câble coaxial ne présentent pas toutes le même niveau de précision. Elles peuvent être fournies:

- Au niveau de l'adresse : on connaît le nombre de prises pour chaque adresse ;
- Au niveau du point de concentration câble (PCC) : on connaît le nombre de prises pour chaque PCC.

Suite au géocodage des données câble, un code BAN est associé à la plupart des adresses des fichiers opérateurs. Ce code BAN permet d'apparier les données d'infrastructure aux immeubles du référentiel immeuble.

Tous les immeubles n'ayant pas été associés à une adresse des fichiers opérateurs, par le géocodage et se trouvant à moins de 50 mètre d'un PCC, sont associés au PCC le plus proche et donc éligibles au câble.

Etape n°2. Détermination de la présence des OC et des débits proposés.

Les opérateurs commerciaux déclarent leur présence commerciale à la TDR, ou au PCC lorsqu'ils sont aussi opérateur d'infrastructure. Ils déclarent également le débit fourni à l'utilisateur. Ces informations permettent à l'ARCEP de déterminer la présence des OC et d'y associer les débits proposés (déclarés), pour chaque immeuble.

Paramètres

Les données suivantes permettent de déterminer l'éligibilité du coax :

Éléments de réseau	Paramètres
PCC (Point de Concentration Câble)	<ul style="list-style-type: none">- Coordonnées du Point de Concentration Câble- Nombre de prises- Débit montant- Débit descendant
Adresse à laquelle est situé un logement raccordé en câble coaxial	<ul style="list-style-type: none">- Code BAN de l'adresse
PCC_OC (Tête de réseau – Opérateur Commercial)	<ul style="list-style-type: none">- Débit montant et descendant
TDR_OC (Tête de réseau – Opérateur Commercial)	<ul style="list-style-type: none">- Débit montant et descendant

5.3.3 Hertzien

Les calculs de l'éligibilité des technologies hertziennes sont réalisés en 2 étapes :

- Dans un premier temps, il est nécessaire de rattacher les données sur les réseaux transmises par les opérateurs d'infrastructures (données surfaciques), aux immeubles du référentiel immeuble ; il s'agit de calculs d'intersection entre des points et des zones, puisque les données transmises par les opérateurs d'infrastructures hertziennes sont surfaciques ;
- Dans un second temps il est nécessaire d'associer à ces immeubles les débits proposés par les opérateurs commerciaux.

5.3.3.1 4G fixe

Association des données réseaux aux immeubles du référentiel et détermination de la présence des OC et des débits proposés.

Les opérateurs d'infrastructure fournissent des fichiers représentant la surface du territoire couvert par leurs réseaux. Les opérateurs commerciaux transmettent l'identifiant des réseaux sur lesquels ils proposent une offre commerciale ainsi que les débits associés.

Par recoupement géographique (coordonnées de l'immeuble se trouvant dans la zone de couverture du réseau), ces informations permettent à l'ARCEP de déterminer la présence des OC et l'éligibilité pour chaque immeuble.

Les débits commerciaux proposés sont transmis par les opérateurs commerciaux. Néanmoins, il s'agit de débits théoriques maximaux, les débits réels étant susceptibles de varier selon la configuration de réception radio et selon la charge du réseau. L'Arcep a ainsi fait le choix, à ce stade, d'associer les débits descendants de tous les opérateurs, au maximum à la classe de débits n°4 (cf. tableau section 5.3.3.5) et les débits montants de tous les opérateurs, au maximum, à la classe de débits n°3 (cf. tableau section 5.3.3.5).

Paramètres

Les données suivantes permettent de déterminer l'éligibilité de la 4G fixe:

Éléments de réseau	Paramètres
Réseau_OI (Réseau mobile – Opérateur d'Infrastructure)	<ul style="list-style-type: none">- Zone de couverture du réseau mobile- Ouverture commerciale
Réseau_OC (Réseau mobile – Opérateur d'Infrastructure)	<ul style="list-style-type: none">- Présence commerciale de l'OC sur le réseau- Limitations mensuelles des données (en Go)

Evolutions effectuées

Des cartes de couverture plus précises ont été obtenues de la part des opérateurs (les cartes de couvertures ne suivent plus les contours communaux et pour la plupart, ne sont plus directement reprises à partir des cartes du réseau 4G mobile).

Des données d'éligibilité à la 4G fixe ont été ajoutées dans les territoires d'outre-mer.

5.3.3.2 Satellite

Association des données réseaux aux immeubles du référentiel et détermination de la présence des OC et des débits proposés.

Les opérateurs d'infrastructure fournissent des fichiers représentant la surface du territoire couvert par leurs réseaux. Les opérateurs commerciaux transmettent l'identifiant des réseaux sur lesquels ils proposent une offre commerciale ainsi que les bandes de fréquence.

Par recoupement géographique (coordonnées de l'immeuble se trouvant dans la zone de couverture du réseau), ces informations permettent à l'ARCEP de déterminer la présence des OC pour chaque immeuble.

Les débits commerciaux proposés sont transmis par les opérateurs commerciaux. Néanmoins, il s'agit de débits théoriques maximaux, les débits réels étant susceptibles de varier selon la configuration de réception radio et selon la charge du réseau. L'Arcep a ainsi fait le choix, à ce stade, d'associer les débits descendants et montants de tous les opérateurs au maximum à la classe de débits n°4 (cf. tableau section 5.3.3.5).

Paramètres

Les données suivantes permettent de déterminer l'éligibilité du satellite :

Éléments de réseau	Paramètres
Réseau_OI (Réseau mobile – Opérateur d'Infrastructure)	<ul style="list-style-type: none">- Zone de couverture du réseau satellite- Ouverture commerciale
Réseau_OC (Réseau mobile – Opérateur d'Infrastructure)	<ul style="list-style-type: none">- Présence commerciale de l'OC sur le réseau- Débit montant et descendant- Limitations mensuelles des données (en Go)

Evolutions effectuées

Les données d'opérateurs satellites proposant des offres dans les territoires d'outre-mer sont dorénavant intégrées.

5.3.3.3 Autres technologies radio

Association des données réseaux aux immeubles du référentiel et détermination de la présence des OC et des débits proposés.

Les opérateurs d'infrastructure fournissent des fichiers représentant la surface du territoire couvert par leurs réseaux en indiquant leurs bandes de fréquence. Les opérateurs commerciaux transmettent

l'identifiant des réseaux sur lesquels ils proposent une offre commerciale ainsi que les débits associés.

Par recoupement géographique (coordonnées de l'immeuble se trouvant dans la zone de couverture du réseau), ces informations permettent à l'ARCEP de déterminer la présence des OC pour chaque immeuble.

Pour les technologies hertziennes à très haut débit (« THD radio »), qui disposent de fréquences dédiées et d'obligations associées, l'Arcep a fait le choix, à ce stade, d'associer les débits descendants au maximum à la classe de débits n°5 (cf. tableau section 5.3.3.5) et les débits montants, au maximum, à la classe de débits n°3 (cf. tableau section 5.3.3.5).

Pour les technologies hertziennes à haut débit (wifi, wifimax, wmax), l'Arcep a fait le choix, à ce stade, d'associer les débits descendants à la classe de débits n°4 (cf. tableau section 5.3.3.5) et les débits montants à la classe de débits n°2 (cf. tableau section 5.3.3.5).

Paramètres

Les données suivantes permettent de déterminer l'éligibilité des technologies haut et très haut débit radio (ex : wifi, wifimax, wimax, ...):

Éléments de réseau	Paramètres
Site_OI (Site radio – Opérateur d'Infrastructure)	<ul style="list-style-type: none"> - Zone de couverture du réseau hertzien - Ouverture commerciale
Site_OC (Site radio – Opérateur commercial)	<ul style="list-style-type: none"> - Présence commerciale de l'OC sur le site - Débit montant et descendant - Limitations mensuelles des données (en Go)

Evolutions effectuées

Les données de 8 opérateurs commerciaux proposant des offres haut et très haut débit radio sont dorénavant collectées par l'Arcep.

5.3.3.4 Récapitulatif

Les calculs expliqués dans les sections précédentes permettent d'associer chaque technologie à des classes de débits (cf. tableau ci-dessous).

Technologie	Classe(s) de débit potentielle(s) pour l'affichage sur la carte	1 Gb/s et +	100 Mb/s à 1 Gb/s	20 Mb/s à 100 Mb /s	6 MB/s à 25 Mb/s	2 Mb/s à 10 Mb/s	512 kb/s à 4 Mb/s	Inéligible HD

	(classes de débits pour les statistiques)	idem	idem	30 à 100	8 à 30	3 à 8	512 à 3	idem
	n° de la classe	7	6	5	4	3	2	1
FttH	De 100 Mbit/s à plus de 1 Gbit/s (selon déclaration opérateur)	X	X					
Câble	De 30 Mbit/s à plus de 1 Gbit/s (selon déclaration opérateur)	X	X	X				
DSL	D'inéligible à 100 Mbit/s (selon affaiblissement de la ligne)			X	X	X	X	X
THD Radio	[30 Mbit/s – 100 Mbit/s]			X				
4G fixe	[8 Mbit/s – 30 Mbit/s] (pourra évoluer selon la maturité de la technologie)				X			
HD Radio	[8 Mbit/s – 30 Mbit/s]				X			
Satellite	[8 Mbit/s – 30 Mbit/s]				X			

5.4 Calcul des statistiques

5.4.1 Principe général

Les données d'éligibilité sont utilisées pour calculer les taux de couverture (% de locaux éligibles) aux différentes mailles administratives.

Ces taux de couverture sont calculés par classe de débits et par technologie (présence d'un réseau et d'une offre commerciale sur ce réseau) et sont établis à partir des données relatives à l'immeuble qu'on retrouve dans le référentiel immeuble (commune, département ou région d'appartenance) ou dans les données d'éligibilité (technologie et débits disponibles).

5.4.2 Evolutions effectuées

Evolution effectuées

Le nouveau référentiel de locaux (logements et locaux professionnels) décrit en 5.1.2 a entraîné une augmentation de la taille des zones. En effet, le référentiel de locaux estimé à 43 millions au T1 2020 est surévalué du fait de doubles comptages d'immeubles et de locaux. Ces doubles comptages

peuvent être réduits en améliorant les algorithmes, mais un nombre significatif de doubles comptages sont issus d'incohérences entre les bases utilisées et ne pourront être corrigés qu'avec une meilleure cohérence entre les adresses de l'IPE, de la BAN et du cadastre.

	INSEE	FPB	MCI T1 2020 (FPB & IPE)
Nombre de locaux (en millions)	37,2	40,4	43,1

La correction des algorithmes utilisés pour calculer l'éligibilité au FttH a permis de mieux estimer le taux de couverture de cette technologie.

En effet, les OC transmettent un fichier qui liste les immeubles des PM sur lesquels ils sont présents et pour lesquels ils ne proposent pas d'offre. Or, le script développé par l'équipe projet Arcep rendait sans service les immeubles de ces lignes, pour tous les opérateurs et pas uniquement pour les opérateurs indiquant ces lignes sans service.

La correction de cette erreur a entraîné une augmentation importante du nombre de locaux éligibles au FttH de l'ordre de 0,3 million .

5.4.3 Résultats statistiques complémentaires à ceux affichés sur la carte des statistiques du site « Ma connexion internet »

Nombre d'adresses/immeubles des fichiers source

	FPB	BAN
Nombre d'adresses/immeubles (en millions)	23,8	24,6

Nombre d'immeubles recensés

Statistiques en millions	France entière		en ZTD		en ZIPRI		en ZIPU	
	T3 2019	T1 2020	T3 2019	T1 2020	T3 2019	T1 2020	T3 2019	T1 2020
Taille des zones en immeubles	25,1	24,1	1,4	1,4	9,3	8,7	14,4	14,4
Dont Nombre d'immeubles IPE	14,0	14,6	1,1	1,1	8,9	8,2	4,0	5,3
Dont Nombre d'immeubles BAN	11,0	2,9	0,3	2,9	0,4		10,3	

Dont Nombre d'immeubles du Cadastre	-	9,7	-	-	0,5	-	9,2
--	---	-----	---	---	-----	---	-----

Nombre de locaux recensés

Statistiques <i>en millions</i>	France entière			en ZTD		
	INSEE	MCI		INSEE	MCI	
		T3 2019	T1 2020		T3 2019	T1 2020
Taille des zones en locaux	37,2	42,3	43,1	6,5	7,8	8,1
Dont Nombre de locaux FPB ou IPE d'immeubles IPE	-	30,1	30,7	-	7,1	7,2
Dont Nombre de locaux FPB associés à des immeubles du cadastre ou BAN	-	12,2	12,5	-	0,8	0,9

Statistiques <i>en millions</i>	en ZIPRI			en ZIPU		
	INSEE	MCI		INSEE	MCI	
		T3 2019	T1 2020		T3 2019	T1 2020
Taille des zones en locaux	14,2	18,2	17,2	16,5	16,2	17,8
Dont Nombre de locaux FPB ou IPE d'immeubles IPE	-	17,7	16,5	-	5,3	6,9
Dont Nombre de locaux FPB associés à des immeubles du cadastre ou BAN	-	0,5	0,7	-	10,9	10,9

Statistiques de couverture des technologies filaires

Statistiques <i>en millions</i>	Couverture France entière			Couverture en ZTD		
	OBS	MCI		OBS	MCI	
	T3 2019	T3 2019	T1 2020	T3 2019	T3 2019	T1 2020
Taille des zones	37,2	42,3	43,1	6,5	7,8	8,1

FttH	16,7 45,5%	16,2 38%	19,4 46%	5,6 87%	5,6 72%	5,9 73%
Câble	26% (9,5)	22% (9,4)	23% (9,9)	79% (5,2)	68% (5,3)	69% (5,6)
DSL	99,7%	98% (41,4)	100% (42,5)	-	97% (7,6)	97% (7,9)

Statistiques <i>en millions</i>	Couverture en ZIPRI			Couverture en ZIPU		
	OBS	MCI		OBS	MCI	
	T3 2019	T3 2019	T1 2020	T3 2019	T3 2019	T1 2020
Taille des zones	14,2	18,2	17,2	16,5	16,2	17,9

FttH	8,2 60%	8,2 45%	9,9 58%	2,9 17%	2,4 15%	3,6 20%
Câble	29% (4,0)	21% (3,8)	24% (4,1)	2% (0,3)	1% (0,2)	1% (0,2)
DSL	-	98% (17,7)	99% (17,0)	-	98% (16,0)	99% (17,9)

Statistiques spécifiques à la technologie cuivre

Statistiques DSL <i>en % des locaux (volume en millions)</i>	Couverture France entière
---	---------------------------------

	MCI
Inéligibles	2% (0,7)
[0,512 Mbit/s – 3 Mbit/s]	14% (5,8)
[3 Mbit/s – 8 Mbit/s]	22% (9,4)
[8 Mbit/s – 30 Mbit/s]	49% (20,9)
Très haut débit	15% (6,4)
Total en millions de lignes	42,1

Statistiques par niveaux de débits

Nombre de locaux couverts <i>en millions</i>	Couverture France entière			Couverture en ZTD			Couverture en ZIPRI			Couverture en ZIPU		
	OBS	MCI		OBS	MCI		OBS	MCI		OBS	MCI	
	T1 2020	T1 2020	T3 2019	T1 2020	T1 2020	T3 2019	T1 2020	T1 2020	T3 2019	T1 2020	T1 2020	T3 2019
FttH	19,5 49%	19,4 45%	16,2 38%	5,9 82%	5,9 73%	5,6 72%	9,8 62%	9,9 58%	8,3 46%	3,8 23%	3,6 20%	2,3 14%
BHD	N/A	43,1 100%	41,7 99%	N/A	8,1 100%	7,8 100%	N/A	17,2 100%	17,8 99%	N/A	17,8 100%	16,1 98%
BHD Sans satellite	N/A	41,6 96%	40,1 95%	N/A	7,8 96%	7,5 96%	N/A	16,4 95%	16,7 93%	N/A	17,4 97%	15,9 96%
BHD Filaire (Ni Satellite, ni 4G Fixe)	N/A	36,0 83%	34,1 81%	N/A	7,8 96%	7,5 96%	N/A	15,0 87%	16,7 93%	N/A	13,2 74%	15,9 96%
THD	25,2	26,8	23,5	6,7	7,3	7,0	11,8	12,1	11,0	6,8	7,4	5,5

	63%	62%	56%	93%	90%	89%	74%	70%	61%	40%	41%	34%
THD Filaire <i>Sans THD</i> <i>Radio</i>	N/A	26,3 61%	N/A	N/A	7,3 90%	N/A	N/A	12,0 70%	N/A	N/A	6,9 39%	N/A

6 Restitutions

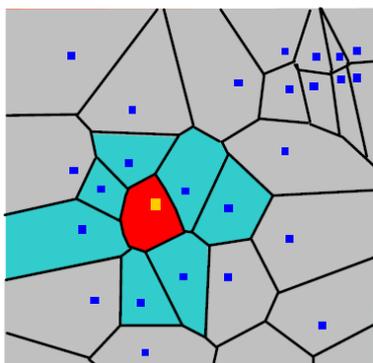
6.1 Génération des cartes pour les informations affichées à l'immeuble et enjeux de fonctionnalités et de performances

Affichage des données

Pour que les informations d'éligibilité à l'immeuble soit lisibles à l'échelle d'une carte France entière, l'ARCEP a décidé d'utiliser des tampons construits sur des polygones de Voronoï.

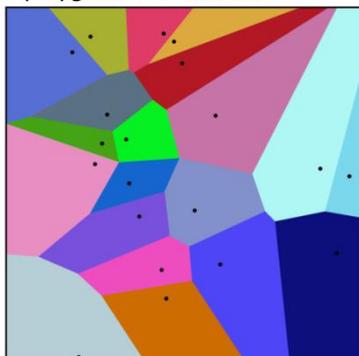
La carte des débits est donc constituée d'une série de polygones (Polygones de Voronoï) formés autour de l'emplacement de chaque point représentant les immeubles. Les polygones de Voronoï sont créés de sorte que chaque emplacement dans un polygone est plus proche du point immeuble présent dans ce polygone que de tout autre point immeuble.

Par exemple, dans cette figure, le point jaune est entouré d'un polygone, affiché en rouge. Chaque emplacement dans le polygone rouge est plus proche du point jaune que de tout autre point (points bleu foncé).

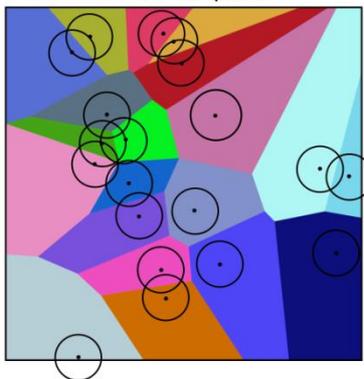


Des tampons de 50 mètres sont ensuite réalisés autour des points, tout en prenant en compte le contour de la commune de chaque point. La carte affiche alors le résultat de la triple superposition, des tampons, des polygones de Voronoï et des contours de leurs communes comme le décrit la figure ci-dessous.

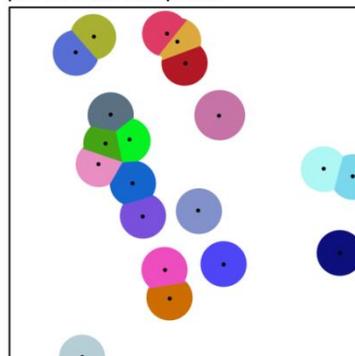
1 polygones de voronoï



2 définition de tampons



3 découpe des polygones à partir des tampons



Généralisation et simplification des zones tampons

La génération et l'affichage des multiples zones tampons étant couteux en ressources, des simplifications et agrégations sont ensuite effectuées sur ces polygones, pour alléger les affichages sur de grandes zones, à différents niveaux de zoom :

- Niveau de zoom 14 à 16 : réalise une agrégation¹² par commune des polygones sans aucune simplification;
- Niveau de zoom 10 à 13 : simplification¹³ des polygones agrégés inférieurs à 7 500 m² et réagrégation des polygones supérieurs à 7 500 m²;
- Niveau de zoom 6 à 9 : simplification des polygones agrégés inférieurs à 15 000 m² et réagrégation des polygones supérieurs à 15 000 m².

6.2 Utilisation des cartes

Le site cartographique « Ma connexion internet », mis à disposition des utilisateurs, comprend 2 cartes : la carte des débits et la carte des statistiques.

6.2.1 Carte des débits

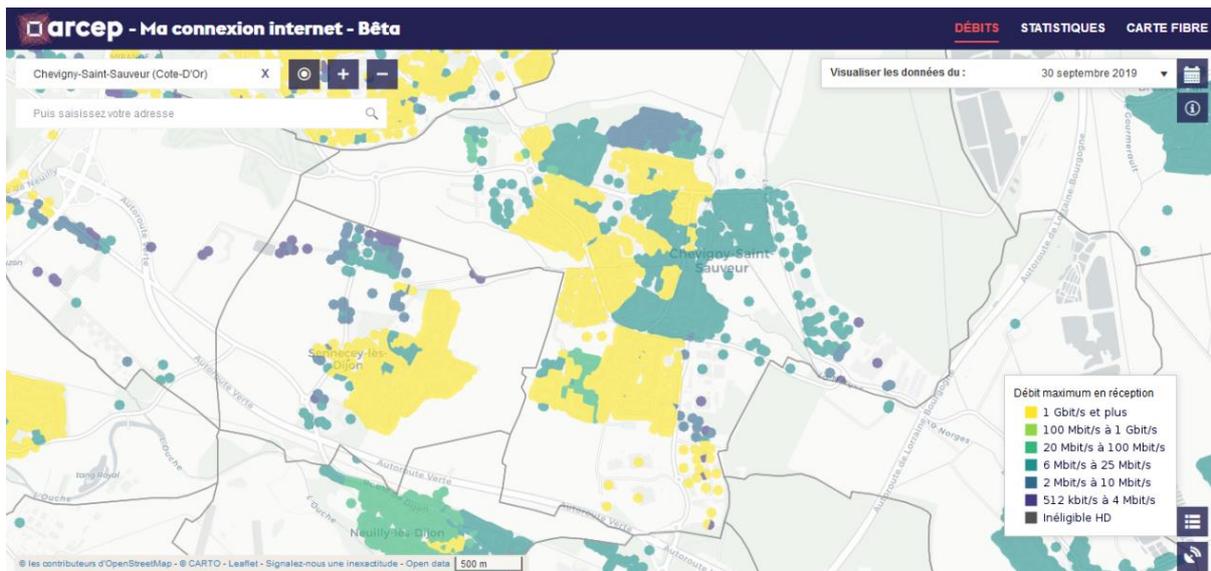
La carte des débits affiche les éligibilités disponibles pour chaque immeuble. Cette éligibilité est affichée par opérateur commercial, par technologie et par classe de débits.

Description de la carte

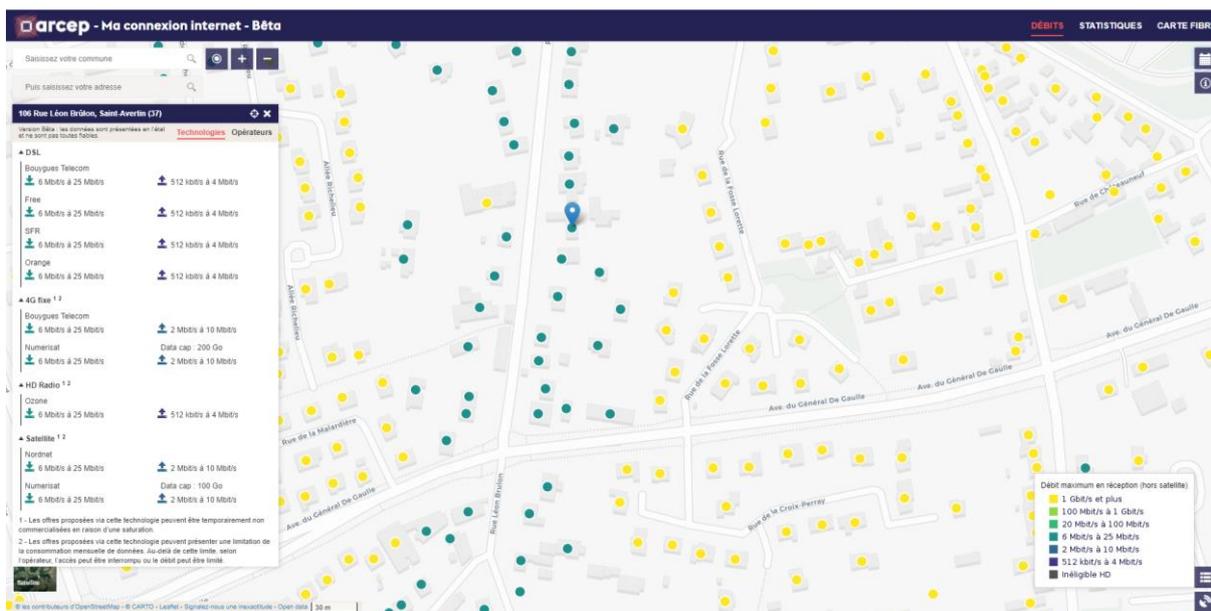
L'éligibilité de chaque immeuble est représentée par des surfaces d'une couleur associée à une classe de débit (cf légende de l'illustration ci-dessous).

¹² Agrégation : chaque polygone est à la base unique car lié à un immeuble (chaque point de la carte débit a un polygone associé). L'agrégation consiste à unir ces polygones par classe de débit et par maille administrative.

¹³ Simplification : la simplification consiste à réduire le nombre d'arêtes ou de sommets pour avoir des formes géométriques plus simples et à supprimer les trous des multipolygones.



Aux deux niveaux de zoom maximum, chaque immeuble est représenté par un point noir et un clic sur l'immeuble permet, par ouverture d'un panneau latéral, d'obtenir des informations plus détaillées sur l'éligibilité de cet immeuble : les opérateurs et technologies disponibles ainsi que les débits proposés par les opérateurs (cf. illustration ci-dessous).



(Capture d'écran de la carte des débits dans sa version actuelle)

Description de la légende de la carte

Chaque technologie est associée à une ou plusieurs classes de débits (cf. illustration ci-dessous) suivant les méthodes indiquées à la section 5.3.

Technologie	Classe(s) de débit potentielle(s) pour l'affichage sur la carte	1 Gb/s et +	100 Mb/s à 1 Gb/s	20 Mb/s à 100 Mb/s	6 Mb/s à 25 Mb/s	2 Mb/s à 10 Mb/s	512 kb/s à 4 Mb/s	Inéligible HD
	<i>(classes de débits pour les statistiques)</i>	idem	idem	30 à 100	8 à 30	3 à 8	512 à 3	idem
FttH	De 100 Mbit/s à plus de 1 Gbit/s (selon déclaration opérateur)							
Câble	De 30 Mbit/s à plus de 1 Gbit/s (selon déclaration opérateur)							
DSL	D'inéligible à 100 Mbit/s (selon affaiblissement de la ligne)							
THD Radio	[30 Mbit/s – 100 Mbit/s]							
4G fixe	[8 Mbit/s – 30 Mbit/s] (pourra évoluer selon la maturité de la technologie)							

Description des niveaux de zoom possibles sur les cartes

Les cartes permettent de zoomer entre le niveau 6 (France métropolitaine visible entièrement) et le niveau 18 (précision au quartier). Le tableau ci-dessous liste les différents niveaux de zoom accessibles sur les cartes du site « Ma connexion internet » et leurs échelles associées.

Niveau de zoom	Echelle
6	100 km
7	50 km
8	30 km
9	20 km
10	10 km
11	5 km
12	2 km
13	1 km
14	500 m
15	300 m
16	100 m
17	50 m
18	30 m

Limitations et améliorations prévues sur la carte

La carte des débits en version bêta ne permet pas de filtrer les informations affichées par technologie ou par opérateur ni de présenter des polygones de voronoï de couleur aux niveaux les plus zoomés (niveaux de zoom n°17 et n°18). Ces deux fonctionnalités, bien que prévues initialement, nécessitent un affichage dynamique, très lourd à tuiler en cache, pour tous les niveaux de zoom. La solution actuelle ne possède pas un niveau de performance suffisant pour permettre cet affichage dynamique.

L'équipe projet réfléchit actuellement aux possibilités de réintégrer les filtres.

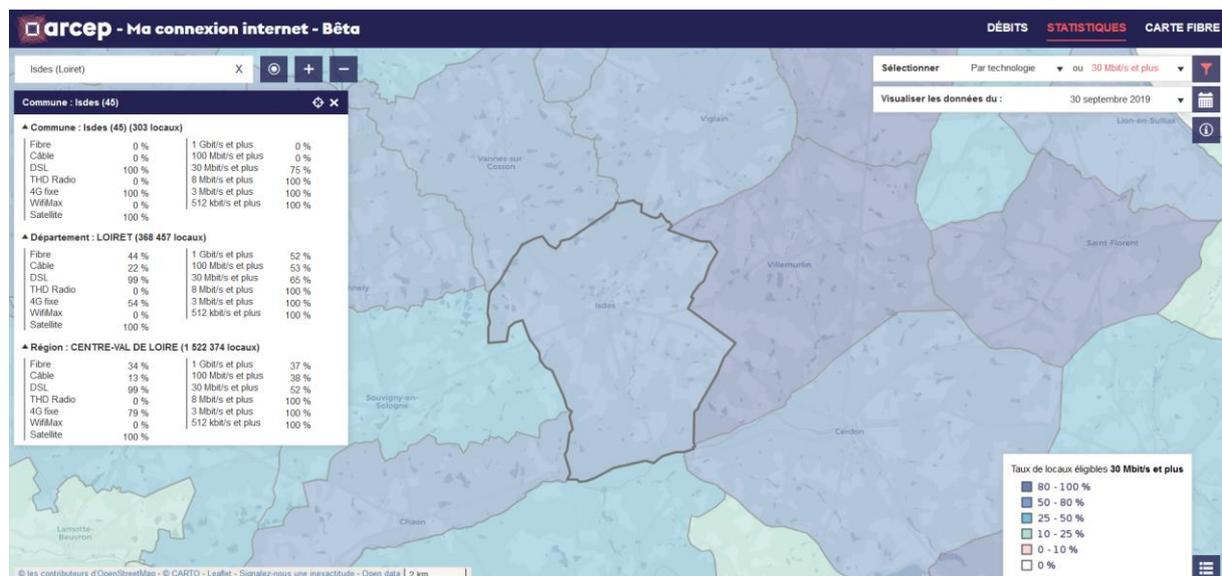
La version bêta de ma connexion internet renvoie l'utilisateur, au clic sur l'onglet « Carte Fibre », vers le site internet cartefibre.arcep.fr. Il est prévu à terme d'intégrer le contenu de [cartefibre](http://cartefibre.arcep.fr) directement dans ma connexion internet. En attendant, [cartefibre](http://cartefibre.arcep.fr) demeure la référence pour les données d'infrastructure FttH.

6.2.2 Carte des statistiques

Sur la carte des statistiques, l'utilisateur peut visualiser des taux de locaux éligibles (la « couverture ») pour différentes mailles administratives (commune, département et région).

Un filtre (ou le panneau latéral) permet de visualiser :

- les taux de couverture pour l'ensemble des technologies d'accès fixe (Fibre, Câble, Cuivre, THD Radio, 4G fixe, HD Radio) ;
- les taux de couverture en bon haut débit, en très haut débit ou en Gbit/s.



Capture d'écran de la carte des statistiques dans sa version actuelle

6.3 Utilisation de l'open data

L'Arcep met à disposition un open data, constitué de 11 fichiers :

- Le référentiel immeuble défini dans ce document ;
- Les données d'éligibilité de chaque immeuble par technologie et par classe de débit ;
- Les statistiques du nombre de locaux éligibles par classe de débit pour chaque maille administrative (commune, département, région) ;
- Les statistiques du nombre de locaux éligibles par technologie pour chaque maille administrative (commune, département, région) ;
- Les statistiques du nombre de locaux éligibles pour l'ensemble des technologies filaires et terrestres pour chaque maille administrative (commune, département, région).

Ces données permettront aux utilisateurs de reconstituer les 2 cartes du site « Ma connexion Internet » (à l'exception des informations sur les opérateurs commerciaux du panneau latéral de la carte des débits).

Ces données permettront aussi aux utilisateurs de reconstituer les statistiques pour d'autres mailles administratives (EPCI par exemple) et de réutiliser le référentiel immeuble pour d'autres cartes.

Limitations et améliorations

En raison d'un besoin de fiabilisation des résultats (version bêta), il a été décidé de ne pas publier les informations nominatives concernant les opérateurs.

L'Arcep prévoit néanmoins de publier ces informations dans des versions futures.

7 Annexe 1 – Données collectées

Le tableau ci-dessous liste des opérateurs transmettant leurs données pour affichage sur le site « Ma connexion internet », en fonction des technologies et de leurs statuts (Opérateurs d’Infrastructures et/ou Opérateurs Commerciaux).

Les cases bleues représentent les données récupérées pour la dernière mise à jour.

	Cuivre		Fibre		Câble		Radio		4G fixe		Satellite	
	OI	OC	OI	OC	OI	OC	OI	OC	OI	OC	OI	OC
Orange												
SFR												
Bouygues Telecom												
Free												
OI FttH ¹⁴												
WSG												
MTVC												
Orne THD												
UEM Metz												
OMT												
SRR												
Canal plus telecom												
ZEOP												
K-Net												
Wibox												
Tubeo												
Videofutur												
Syndicat des communes du pays de Bitche												
Ozone												
SRR												
OWI TEL												
Guyacom												
Vialis												
Axione												
LGNUM												
LGTEL												
Nomotech												
Apinet												
Xilan												
Dauphin												
FAIMaison												

¹⁴ Sont intégrées les données de tous les opérateurs d’infrastructure fibre transmettant leurs données IPE à l’Autorité au titre de la Décision n° 2018-170 – relative à la collecte d’informations concernant le déploiement et la commercialisation de réseaux fixes à haut et très haut débit

8 Annexe 2 – Données collectées

Le table ci-dessous présente le classement effectué par l'Arcep en logement ou local professionnel des catégories du FPB à partir des désignations (« destinations ») fournies par Etalab.

type de local (source : destination etalab)	logement/local pro (destinations ignorées = adresse non générée par etalab si seulement cette destination)
appartement	logement
atelier-artisanal	local pro
atelier-industriel	local pro
autre-etablissement	local pro
autre-hebergement	local pro
bureaux	local pro
carriere	ignorée
centre-de-loisirs	local pro
centre-medico-social-creche	local pro
centre-thermal-reeducation	local pro
chantier	ignorée
chenil-vivier	local pro
commerce	local pro
dependance-batie-isolee	ignorée
dependance-commerciale	ignorée
depot	ignorée
divers	local pro
ecole-privee	local pro
etablissement-detente-bien-etre	local pro
gare	ignorée
hopital	local pro
hotel	local pro
installations-techniques	ignorée
local-commun	ignorée
maison	logement
maison-de-retraite	local pro
marche	ignorée

parc-de-stationnement	ignorée
port-de-plaisance	local pro
residence-hoteliere	local pro
salle-de-loisir	local pro
salle-de-spectacle	local pro
site-industriel	local pro
sol-de-maison	ignorée
station-service	ignorée
stockage-specifique	ignorée
terrain-de-camping	local pro
terrain-de-sport	local pro

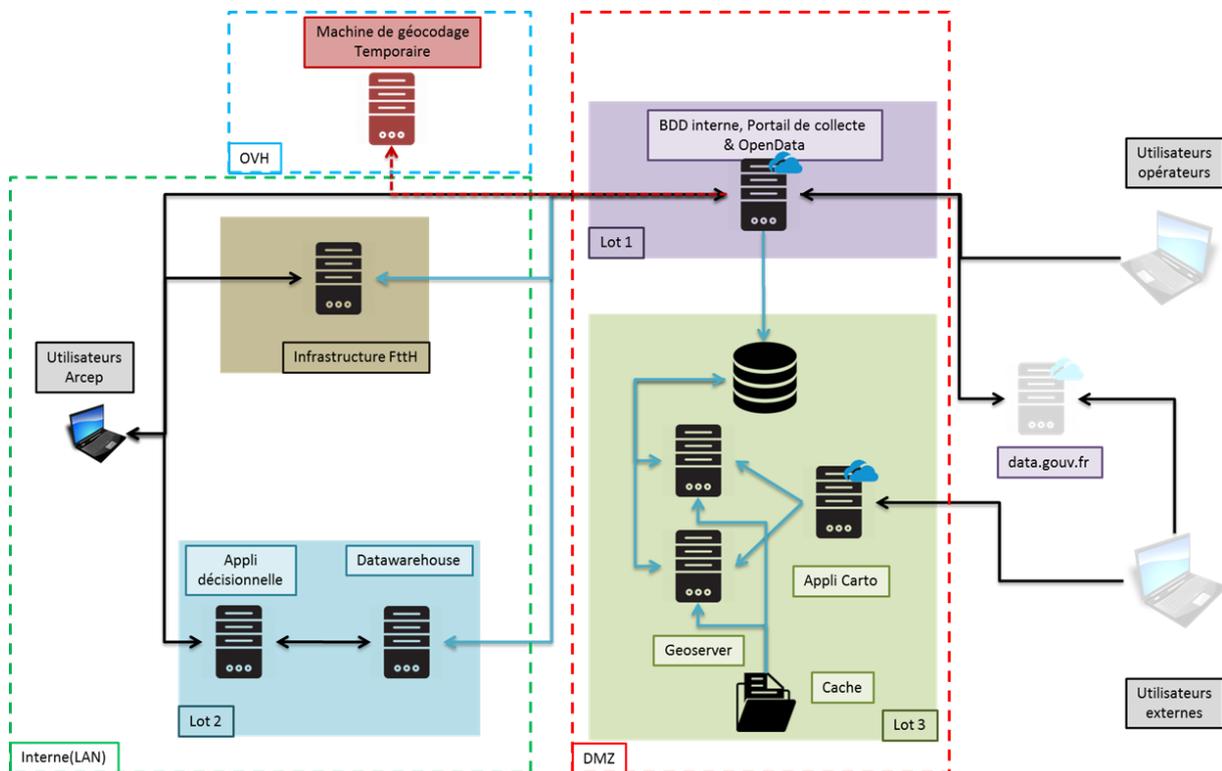
9 Annexe 3 – Technologies utilisées et quelques chiffres

9.1 Les technologies utilisées pour les différentes étapes :

L'architecture logicielle actuelle se base sur un système d'exploitation Linux Debian 9.4, 64-bits (Stretch). Les composants installés sont :

- Système d'exploitation : Debian 9;
- Base de données : PostgreSQL- PostGIS;
- Cartographie : librairie Leaflet ;
- Développement web : JavaScript, HTML, CSS, PHP;
- Serveur de tuilage : GeoServer ;
- Fonds de carte : OpenStreetMap ;
- Serveur WEB : Apache.

L'infrastructure globale du projet est décrite sur le schéma suivant :



L'application cartographique affiche des couches raster publiées via geoserver.

Afin de répondre à la charge des consultations, ont été déployées deux instances Tomcat qui hébergent chacune deux instances geoserver. L'équilibrage de charge est réalisé par Apache.

Certaines couches sont mises en cache via GeoWebCache afin de décharger la base de données à petite échelle.

Les instances Geoserver se connectent à une seule instance de base de données qui est installée sur une machine virtuelle distincte.

9.2 Quelques chiffres

Ces chiffres sont présentés pour les données d'un seul trimestre, à savoir le T3 2019 de la version bêta. L'ouverture de la collecte à un plus grand nombre d'opérateurs pour les prochaines versions ainsi que l'ajout de trimestres successifs entrainera d'importantes augmentations sur ces chiffres :

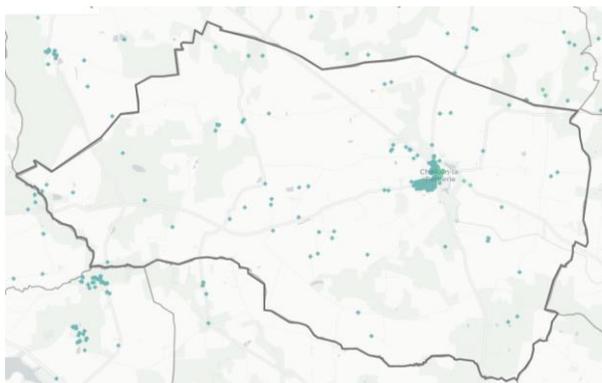
- Nombre de fichiers collectés : plus de 150 fichiers (sans compter les fichiers OI FttH) ;
- Volumes de données collectées : 7 Go ;
- Volume de données de la base : 14 Go ;
- Nombre de traitements : plus de 15 traitements sur chacun des fichiers ;
- Temps de traitement entre la fin de la collecte et la mise à jour de la carte : environ 3 semaines ;
- Taille de l'open data : 6,5 Go (4,8 Go compressé).

10 Annexe 4 – Rappel des limitations et améliorations prévues

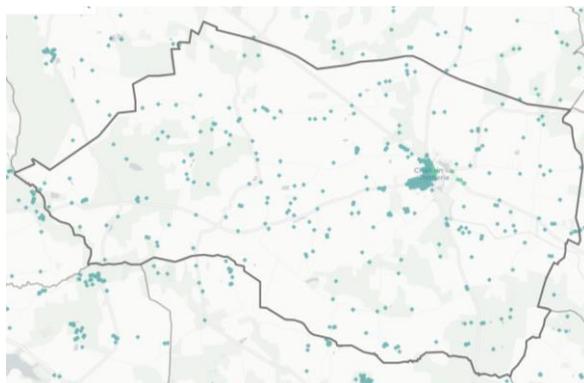
Evolutions effectuées sur les référentiels immeubles et locaux

Le référentiel immeuble est dorénavant plus complet, particulièrement en zones rurales. Ci-dessous l'exemple de la commune de Challain-la-Potherie montre qu'une grande quantité d'immeubles ont été ajoutés avec les données du FPB (cadastre). Il s'agit principalement des immeubles ayant une « pseudo-adresse », c'est-à-dire qui ne présentent pas de nom de rue, non présents dans la BAN

Base immeuble du T3 2019



Base immeuble du T1 2020



Exemple de Challain-la-Potherie : Les points sur la carte de gauche représentent les immeubles présents sur les cartes du site cartographique lors de la publication des données du T3 2019 (issus de la BAN) et les points de la carte de droite représentent les immeubles présents sur les cartes du site cartographique lors de la publication des données du T1 2020 (issus du cadastre)

Une modification non prévue de nomenclature de la DGFIP pour catégoriser les locaux du FPB, n'avait pas permis d'intégrer les locaux professionnels du fichier FPB dans la première version de la carte et de l'open data publiés en avril 2020. 2,9 millions de locaux professionnels ont donc ainsi pu être récupérés dans la publication du T1 2020. Le liste des catégories DGFIP retenues pour constituer le référentiel de locaux est mise à disposition en annexe.

Un nettoyage des fichiers IPE a aussi permis de supprimer 2,1 millions de locaux dans des immeubles en doublons ou fantômes, qui avaient un impact sur les statistiques et sur la représentation des débits :

- soit en dédoublant toutes les adresses avec des informations différentes (exemple des cas de doublons Orange-SFRen zone AMII),
- soit en perturbant le calcul des Voronoï lorsque les points sont strictement superposés (cas des changements d'identifiants).

Evolutions effectuées sur l'éligibilité cuivre

Le calcul d'atténuation réalisé par l'Arcep prend pour hypothèse d'ajouter un affaiblissement

correspondant à la distance entre le PC et la terminaison de la ligne, à celui observé au PC afin d'être au plus proche de la réalité terrain concernant les débits.

Néanmoins, le critère d'affaiblissement inférieur à 78 dB pour déterminer l'éligibilité HD est dorénavant apprécié au PC, comme habituellement par Orange, au lieu de l'être en bout de ligne comme précédemment. Cette évolution a entraîné une éligibilité supplémentaire au DSL pour 0,5 millions de locaux jusqu'ici inéligibles.

Evolutions effectuées sur l'éligibilité à la fibre

L'Arcep récupère dorénavant des données de la part de 16 opérateurs commerciaux ce qui permet d'être plus complet sur les offres proposées par les opérateurs. Néanmoins, environ 600 000 locaux ne sont actuellement couverts par aucun opérateur commercial.

Evolutions effectuées sur l'éligibilité à la 4GF

Des cartes de couverture plus précises ont été obtenues de la part des opérateurs (les cartes de couvertures ne suivent plus les contours communaux et pour la plupart, ne sont plus directement reprises à partir des cartes du réseau 4G mobile).

Des données d'éligibilité à la 4G fixe ont été ajoutées dans les territoires d'outre-mer.

Evolutions effectuées sur l'éligibilité au satellite

Les données d'opérateurs satellites proposant des offres dans les territoires d'outre-mer sont dorénavant intégrées.

Evolutions effectuées sur l'éligibilité aux autres technologies radio

Les données de 8 opérateurs commerciaux proposant des offres haut et très haut débit radio sont dorénavant collectées par l'Arcep.

Evolutions effectuées sur les calculs des statistiques

Le nouveau référentiel de locaux (logements et locaux professionnels) décrit en 5.1.2 a entraîné une augmentation de la taille des zones. En effet, le référentiel de locaux estimé à 43 millions au T1 2020 est surévalué du fait de doubles comptages d'immeubles et de locaux. Ces doubles comptages peuvent être réduits en améliorant les algorithmes, mais un nombre significatif de doubles comptages sont issus d'incohérences entre les bases utilisées et ne pourront être corrigés qu'avec une meilleure cohérence entre les adresses de l'IPE, de la BAN et du cadastre.

	INSEE	FPB	MCI T1 2020
--	--------------	------------	------------------------

			(FPB & IPE)
Nombre de locaux (en millions)	37,2	40,4	43,1

La corrections des algorithmes utilisés pour calculer l'éligibilité au FttH a permis de mieux estimer le taux de couverture de cette technologie.

En effet, les OC transmettent un fichier qui liste les immeubles des PM sur lesquels ils sont présents et pour lesquels ils ne proposent pas d'offre. Or, le script développé par l'équipe projet Arcep rendait sans service les immeubles de ces lignes, pour tous les opérateurs et pas uniquement pour les opérateurs indiquant ces lignes sans service.

La correction de cette erreur a entraîné une augmentation importante du nombre de locaux éligibles au FttH de l'ordre de 0,3 million .

Limitations sur la carte

La carte des débits en version bêta ne permet pas de filtrer les informations affichées par technologie ou par opérateur ni de présenter des polygones de voronoï de couleur aux niveaux les plus zoomés (niveaux de zoom n°17 et n°18). Ces deux fonctionnalités, bien que prévues initialement, nécessitent un affichage dynamique, très lourd à tuiler en cache, pour tous les niveaux de zoom. La solution actuelle ne possède pas un niveau de performance suffisant pour permettre cet affichage dynamique.

L'équipe projet réfléchit actuellement aux possibilités de réintégrer les filtres.

La version bêta de ma connexion internet renvoie l'utilisateur, au clic sur l'onglet « Carte Fibre », vers le site internet cartefibre.arcep.fr. Il est prévu à terme d'intégrer le contenu de cartefibre directement dans ma connexion internet. En attendant, cartefibre demeure la référence pour les données d'infrastructure FttH.

Limitations sur l'open data

En raison d'un besoin de fiabilisation des résultats (version bêta), il a été décidé de ne pas publier les informations nominatives concernant les opérateurs.

L'Arcep prévoit néanmoins de publier ces information dans des versions futures.

Limitations sur les référentiels immeubles et locaux

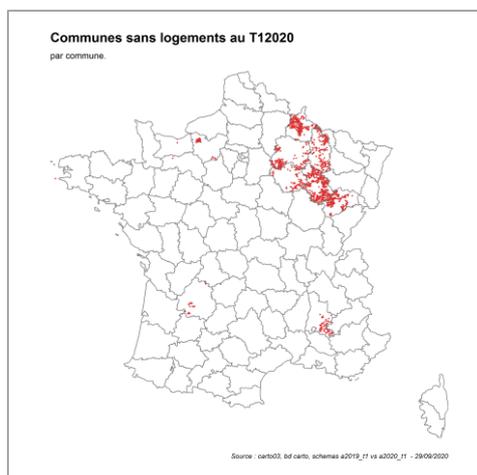
Communes sans immeuble

Le cadastre n'étant pas entièrement numérisé, certaines adresses n'ont pas de coordonnées géographiques (plus de 200 000). Bien que présentes dans le référentiel de bâtiments de Ma connexion internet, ces adresses n'apparaissent pas sur la carte et ne sont pas comptabilisées dans les calculs statistiques.

1 039 communes (voir localisation sur la carte ci-dessous), essentiellement situées dans la région du Grand Est, ne possèdent que des adresses non numérisées et sont donc totalement vides.

500 communes supplémentaires ont récemment été numérisées dans le cadastre et pourront être directement intégrées dans une prochaine mise à jour.

D'ici là et pour les communes qui ne sont toujours pas numérisées, cette limitation pourra être levée en privilégiant les adresses de la BAN dans les 1 039 communes.



Localisation des communes dont le cadastre n'est pas encore numérisé

Augmentation du nombre de locaux

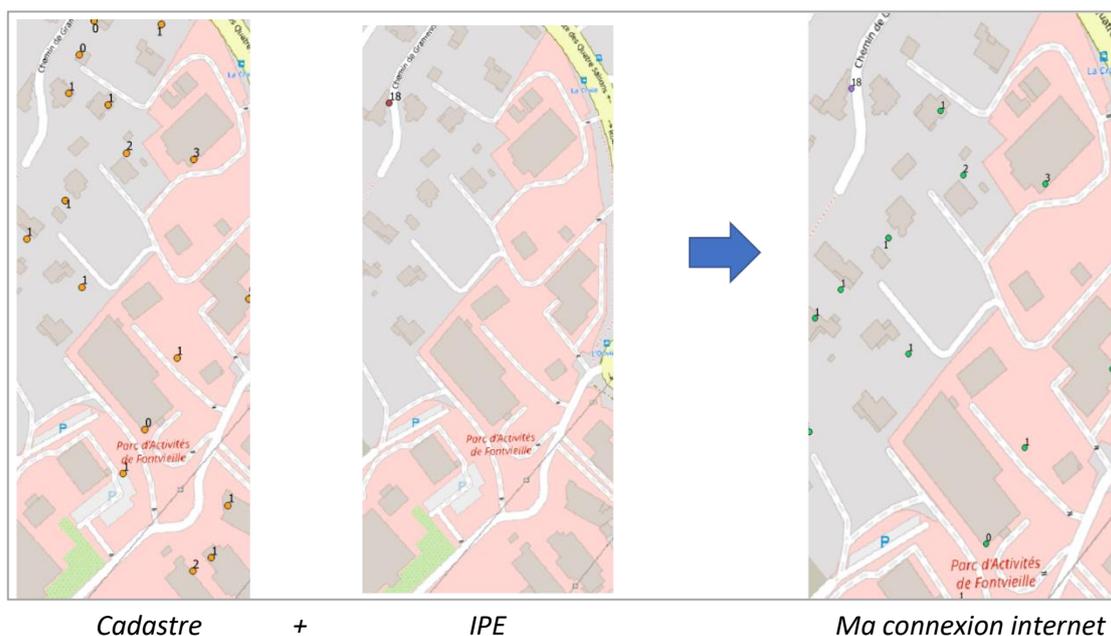
Le référentiel de locaux estimé à peu plus 43 millions au T1 2020 est surévalué du fait de doubles comptages d'immeubles et de locaux. Ces doubles comptages peuvent être réduits en améliorant les algorithmes, mais un nombre significatif de doubles comptages sont issus d'incohérences entre les bases utilisées et ne pourront être corrigés qu'avec une meilleure cohérence entre les adresses de l'IPE, de la BAN et du cadastre.

	INSEE	FPB	MCI T1 2020 (FPB & IPE)
Nombre de locaux (en millions)	37,2	40,4	43,1

Double comptage d'immeubles

Les adresses du cadastre et de l'IPE sont associées (géocodage) aux adresses BAN. La BAN sert donc de pivot pour relier un immeuble IPE et une adresse cadastrale. Or seules 78% des adresses du cadastre s'associent à une adresse BAN.

Pour éviter des doubles comptes d'immeubles, il a été décidé de récupérer les immeubles du cadastre non associés à la BAN uniquement lorsqu'ils sont situés à plus de 40m de l'immeuble IPE. Néanmoins dans certaines zones, et comme le montre l'exemple ci-dessous, cette distance n'est pas toujours suffisante pour éviter de comptabiliser à tort des doublons.



Dans l'exemple ci-dessus, l'immeuble IPE en haut à gauche, possède l'ensemble des locaux de la zone industrielle. Le cadastre est plus précis et donne la position de chaque immeuble de la zone industrielle ainsi que leur nombre de locaux. Mais ces immeubles n'ont pas les mêmes adresses dans l'IPE et dans le cadastre, l'association n'a donc pas réussi et les immeubles sont comptabilisés deux fois.

Double comptage de locaux

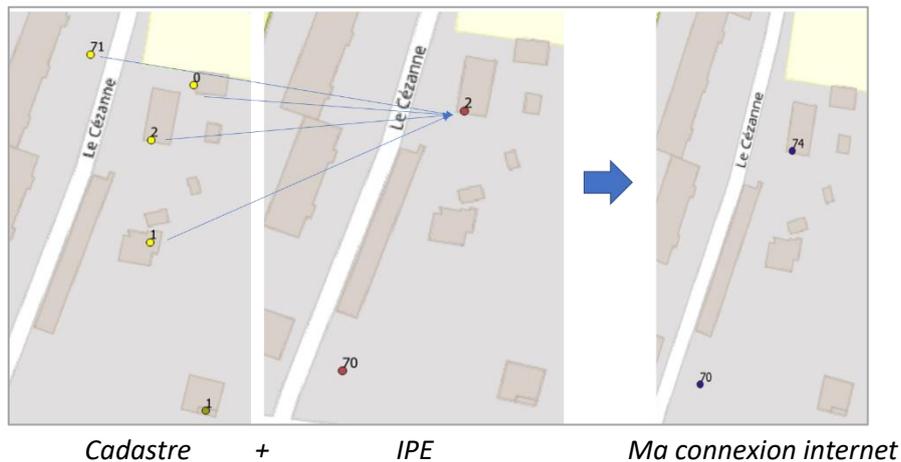
En raison de discussions en cours avec certains opérateurs sur la capacité à placer en open data les données de décompte de locaux issues de leur IPE, l'Arcep effectue à ce stade une substitution, autant que possible, des locaux issus de l'IPE par des données provenant de la DGFIP.

Néanmoins, ce remplacement reste tributaire de la capacité à associer des bâtiments à une adresse BAN. Certaines associations se basant sur des critères de distance entraînent des doubles comptages, comme le décrit l'exemple ci-dessous.

Cette limitation pourra être levée :

- en ne remplaçant que les immeubles qui présentent un nombre de locaux proche entre le référentiel IPE et le référentiel FPB ;
- suite aux discussions en cours avec certains opérateurs sur la capacité à placer en open data

les données de décompte de locaux issues de leur IPE.



Dans l'exemple ci-dessus, nous pouvons voir que l'adresse du FPB ayant 71 locaux (en haut à gauche) est associée à l'immeuble IPE le plus proche (immeuble avec 2 locaux dans l'IPE). Or l'immeuble avec le nombre de locaux correspondant est situé plus bas dans l'IPE. Mais les adresses étant très différentes, aucune association sur l'adresse n'a pu aboutir et l'association a donc été effectuée sur le seul critère de distance (au plus proche). Il y a donc un double compte de ces locaux.

Incohérences entre les bases sur le nombre de locaux d'un immeuble

L'IPE et le FPB nous fournissent un nombre de locaux à l'immeuble. Il s'avère que les deux bases peuvent associer un nombre de locaux très différents pour les mêmes immeubles (parfois l'IPE associe un nombre de locaux plus important pour un immeuble donné, parfois c'est l'inverse).



Dans l'exemple ci-dessus, nous pouvons voir que les adresses du FPB associent plus de locaux que l'IPE pour ces immeubles.

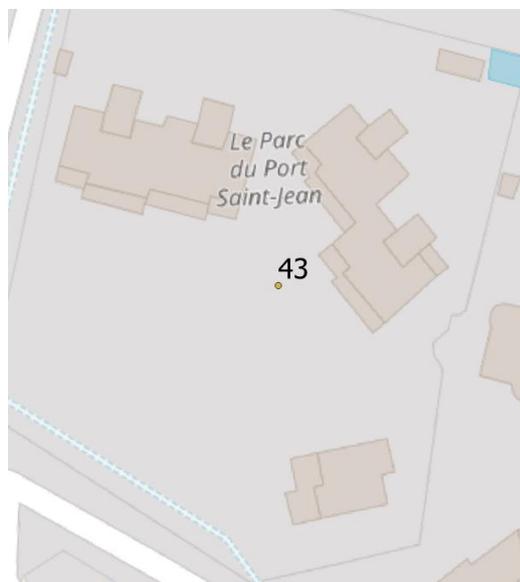
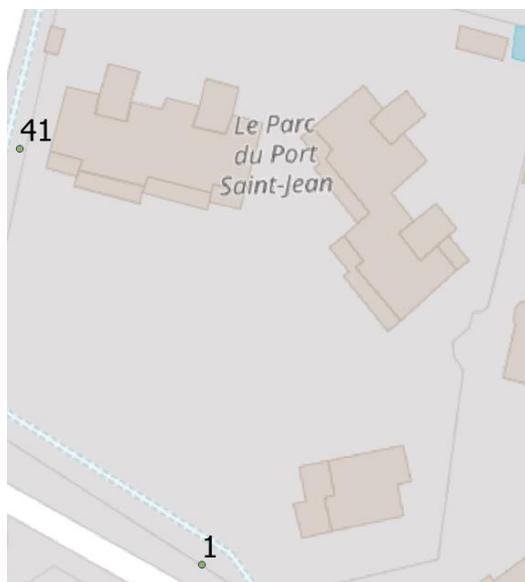
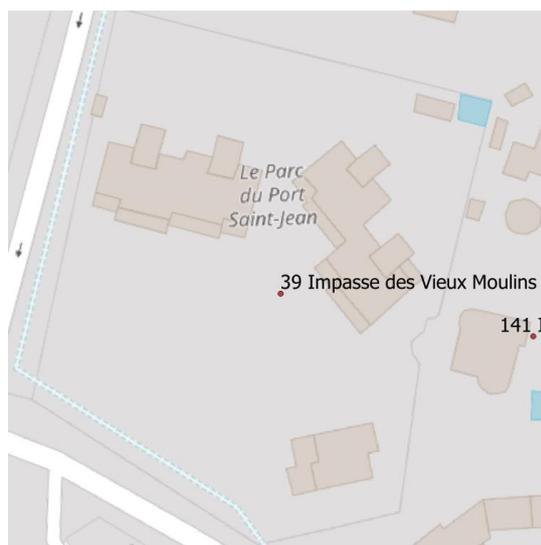
Incohérences entre les bases sur le nombre de locaux d'une adresse

L'IPE et le FPB nous fournissent un nombre de locaux à l'immeuble. Cet immeuble est relié à une adresse. Il apparaît que plusieurs adresses peuvent être associées aux mêmes immeubles ou groupe d'immeubles. Les locaux sont alors associés à des adresses différentes dans les deux bases.

IPE



Cadastre (avec locaux du FPB)



Dans l'exemple ci-dessus, nous pouvons voir que l'IPE associe 41 locaux à l'adresse « 2 rue andré bellon » et un seul local à l'adresse 31 chemin des vieux moulins, qui semblent représenter le même groupe d'immeubles, alors que le FPB ne repertorie pas l'adresse « 2 rue andré bellon » et associe 43 locaux à l'adresse « 39 impasse des vieux moulins ».