

Introduction à la production de données

**Mon adresse, mon débit
avec **Ma connexion internet****

29 mai 2020

Maconnexioninternet.arcep.fr

Ma connexion internet est un **outil cartographique** de référence à destination des consommateurs et des élus, permettant d'afficher, pour un immeuble donné, les opérateurs qui y proposent les services et les débits qui y sont disponibles.

Collecte (trimestrielle)

Décision 2018-169 – spécifie le format des données de MCI

Source de données complémentaires :

Opérateurs (dont IPE)
BAN (Adresses)
DGFIP (Nombre de locaux)
IGN (Données géographiques)

Traitements Arcep

- Contrôles et corrections
- Création d'une référentiel commun d'immeuble (BAN/IPE)
- Projection d'informations d'éligibilité sur ce référentiel
- Calculs statistiques

Partage et exploitation

Ma connexion internet
Open data
Page avec documentation technique

Collecte trimestrielle basée sur la **Décision 2018-169** (faisant référence à l'article L. 36-7 du CPCE) **qui :**

- définit les contenus et les modalités de **mise à disposition du public**, par les opérateurs de **cartes** relatives à la **disponibilité, à la qualité** et à la couverture des réseaux et des services d'accès à internet en situation fixe
- définit les contenus et les modalités de **transmission à l'Arcep d'informations relatives à la couverture des réseaux et des services d'accès à internet** en situation fixe, en vue de leur mise à disposition du public par l'Autorité, notamment sous forme de **cartes numériques**.
 - => ne sont concernés que les opérateurs qui disposent sur le marché de détail grand public **d'au moins 1 000 abonnements actifs**
 - => type de fichier (CSV, shapefile..)
 - => noms des entêtes des champs
 - => données des champs et format à respecter (entier, VarChar(10), booléen...)

Maconnexioninternet.arcep.fr

Ma connexion internet est un **outil cartographique** de référence à destination des consommateurs et des élus, permettant d'afficher, pour un immeuble donné, les opérateurs qui y proposent les services et les débits qui y sont disponibles.

Collecte (trimestrielle)

Décision 2018-169 – spécifie le format des données de MCI

Source de données complémentaires :

Opérateurs (dont IPE)
BAN (Adresses)
DGFIP (Nombre de locaux)
IGN (Données géographiques)

Traitements Arcep

- Contrôles et corrections
- Création d'une référentiel commun d'immeuble (BAN/IPE)
- Projection d'informations d'éligibilité sur ce référentiel
- Calculs statistiques

Partage et exploitation

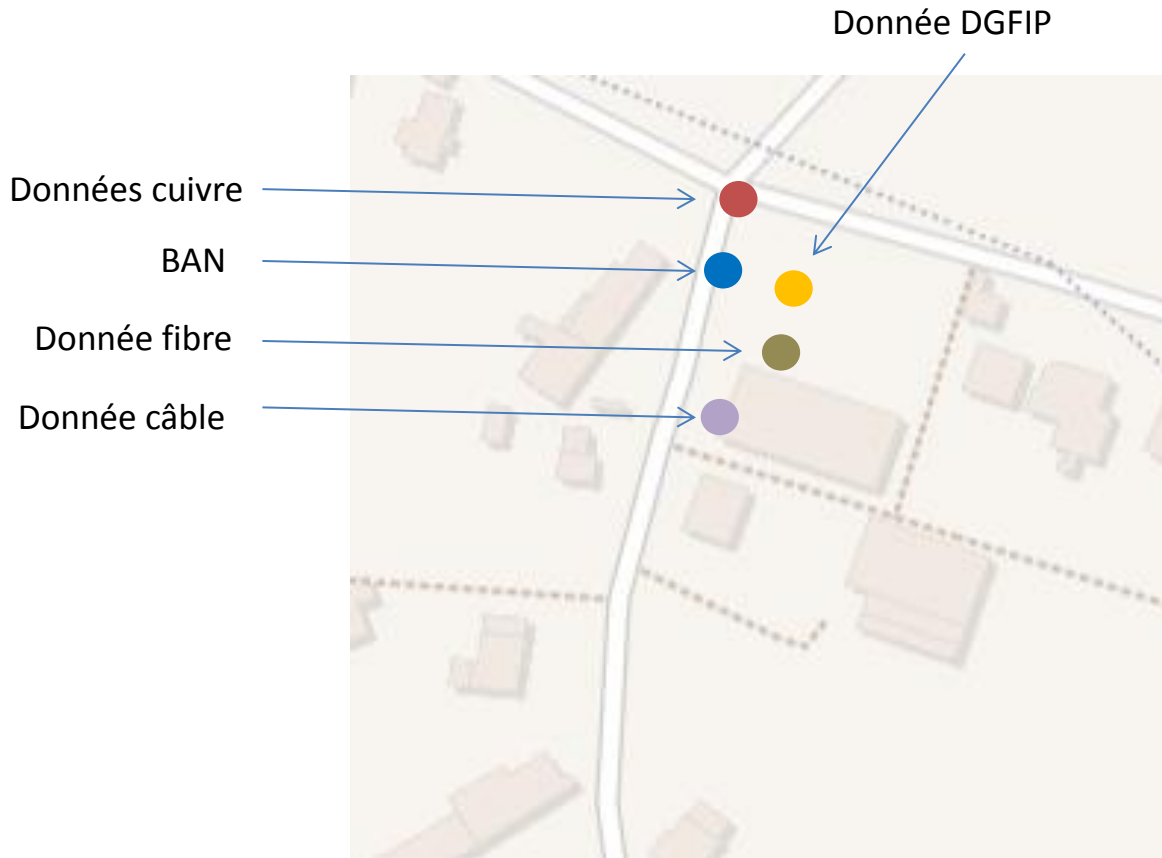
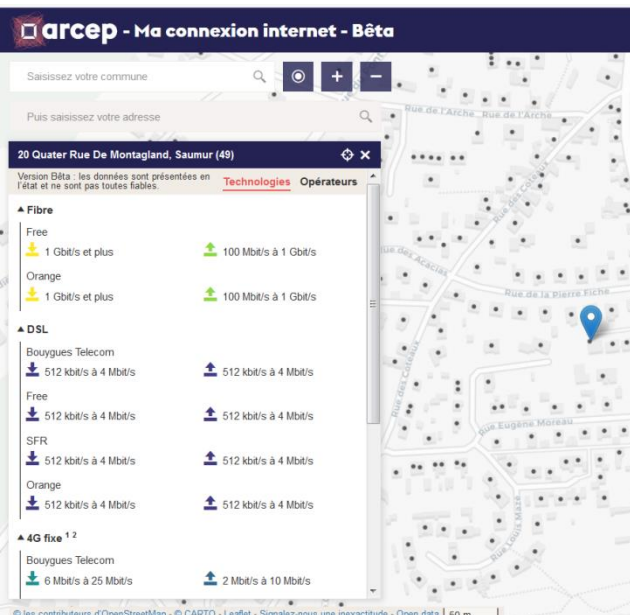
Ma connexion internet
Open data
Page avec documentation technique

Objectifs

- 1. Présentation sur les traitements effectués par l'ARCEP**
 - 1. Référentiel d'immeubles et de locaux**
 - 2. Les calculs d'éligibilité**
- 2. Cas pratiques – le géocodage à la BAN**
 - 1. Exemple de géocodage avec l'outil de géocodage en ligne d'ETALAB**
 - 2. Détail du fonctionnement du script de géocodage des données DGFIP**

Pourquoi un référentiel commun?

Nécessité d'avoir l'éligibilité à la maille de l'immeuble or multitude de sources d'adresses en entrée



Pourquoi un référentiel commun?

Géocoder toutes les bases à une base commune permet de rattacher à ce référentiel, les informations des autres référentiels

La BAN sert de base de « pivot » pour tous les référentiels des données d'entrée

arcep - Ma connexion internet - Bêta

Saisissez votre commune

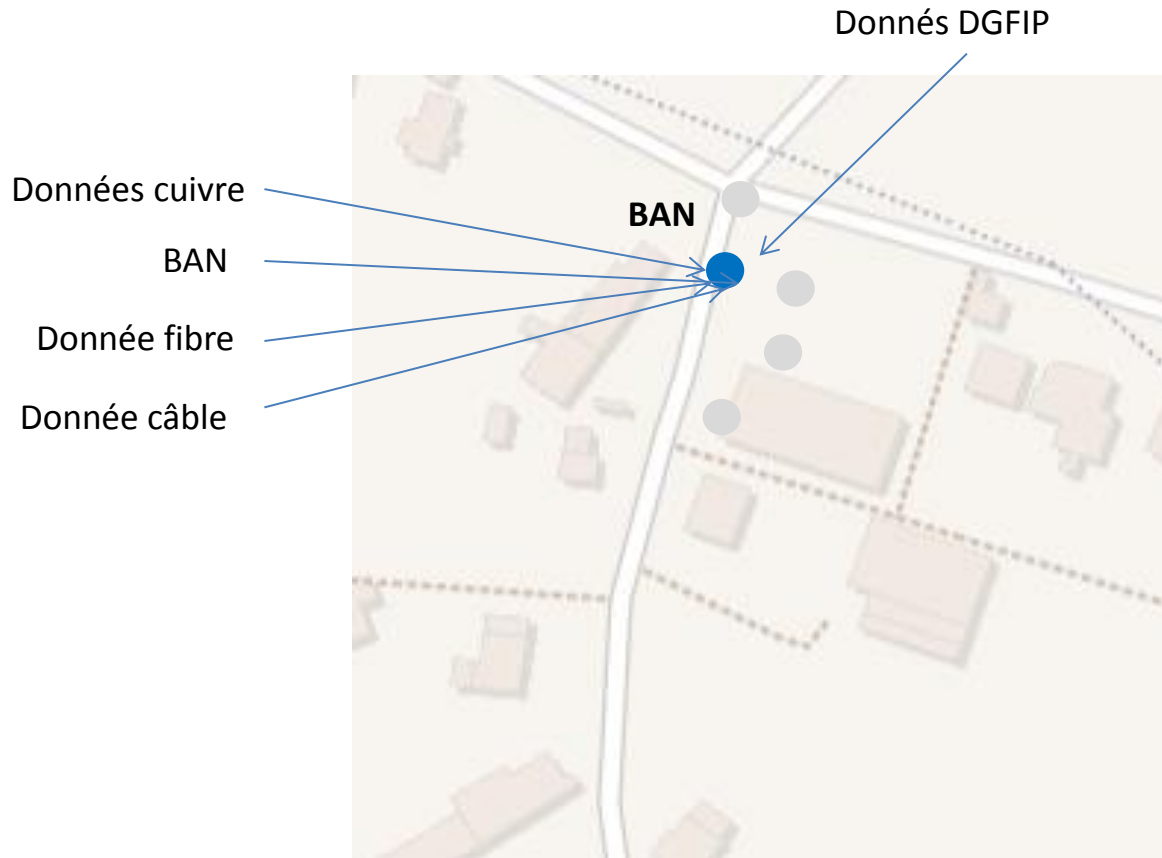
Puis saisissez votre adresse

20 Quater Rue De Montagland, Saumur (49)

Version Bêta : les données sont présentées en l'état et ne sont pas toutes fiables. Technologies Opérateurs

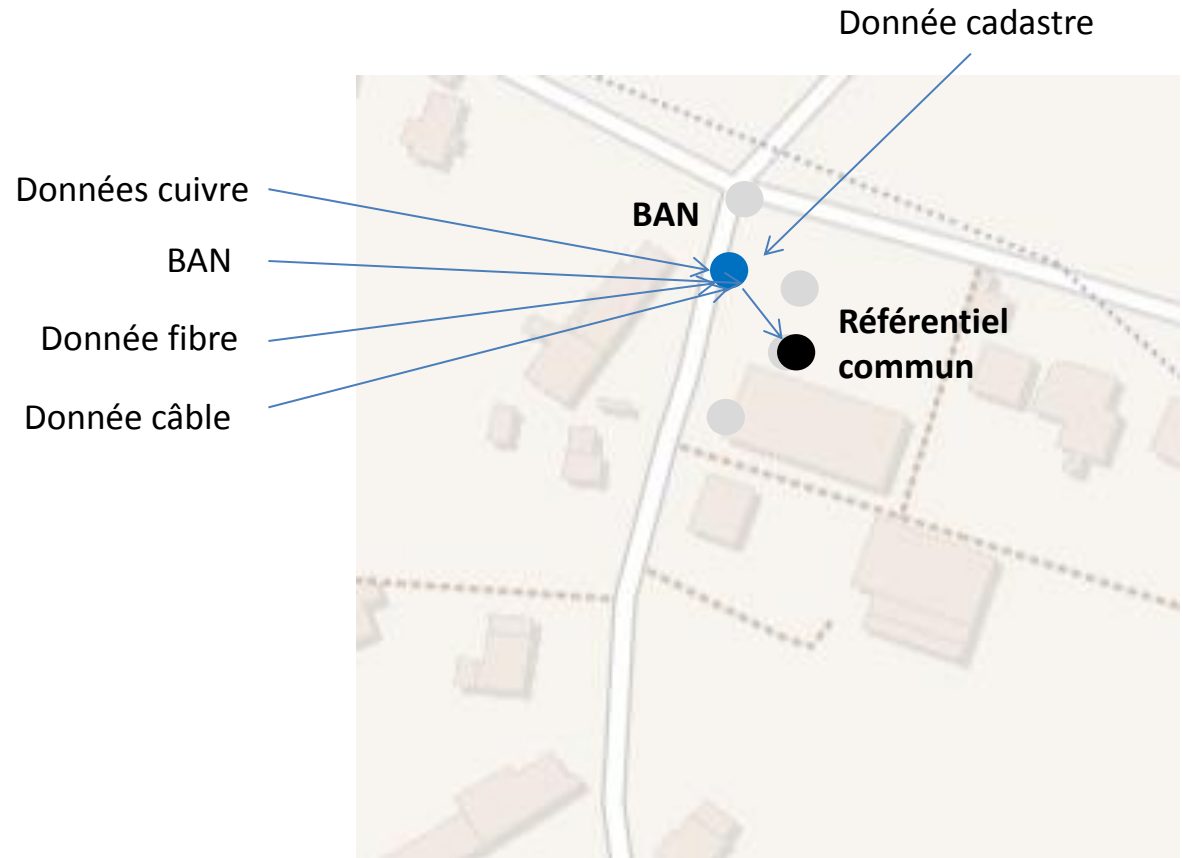
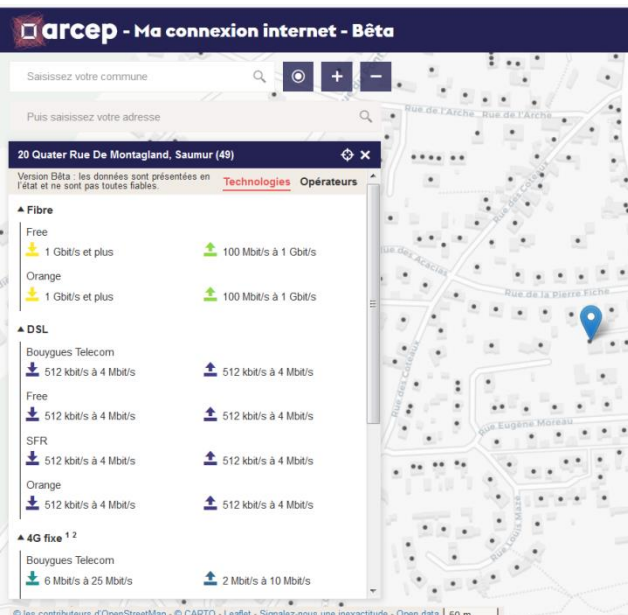
- Fibre**
 - Free
 - 1 Gbit/s et plus 100 Mbit/s à 1 Gbit/s
 - Orange
 - 1 Gbit/s et plus 100 Mbit/s à 1 Gbit/s
- DSL**
 - Bouygues Telecom
 - 512 kbit/s à 4 Mbit/s
 - Free
 - 512 kbit/s à 4 Mbit/s
 - SFR
 - 512 kbit/s à 4 Mbit/s
 - Orange
 - 512 kbit/s à 4 Mbit/s
- 4G fixe**
 - Bouygues Telecom
 - 6 Mbit/s à 25 Mbit/s

© les contributeurs d'OpenStreetMap - © CARTO - Leaflet - Signalez-nous une inexactitude - Open data | 50 m



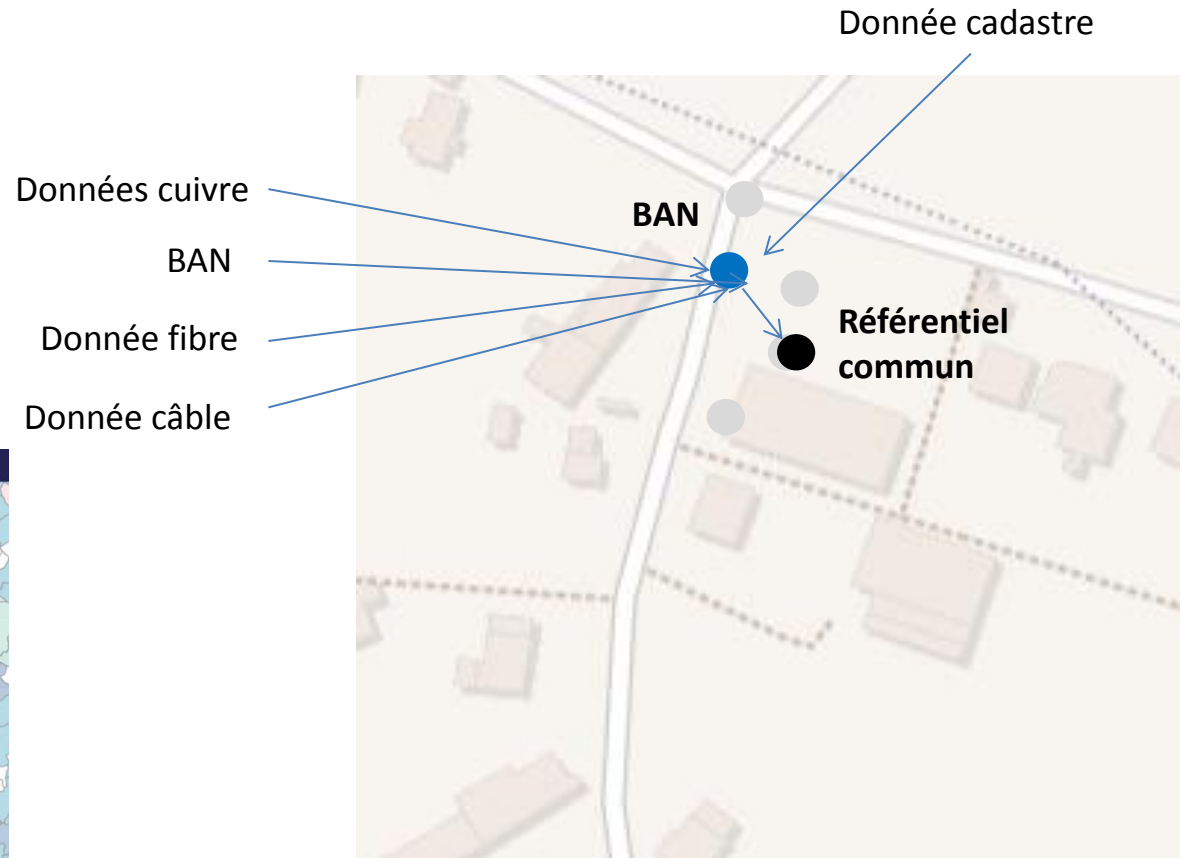
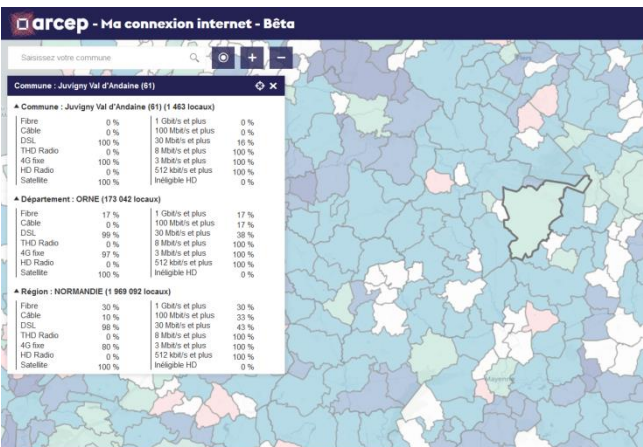
Pourquoi un référentiel commun?

Géocoder toutes les bases à une base commune permet de rattacher à ce référentiel, les informations des autres référentiels



Pourquoi un référentiel commun?

Géocoder toutes les bases à une base commune permet de rattacher à ce référentiel, les informations des autres référentiels



Référentiel d'immeubles et de locaux

Le référentiel d'immeubles

Faire primer l'IPE



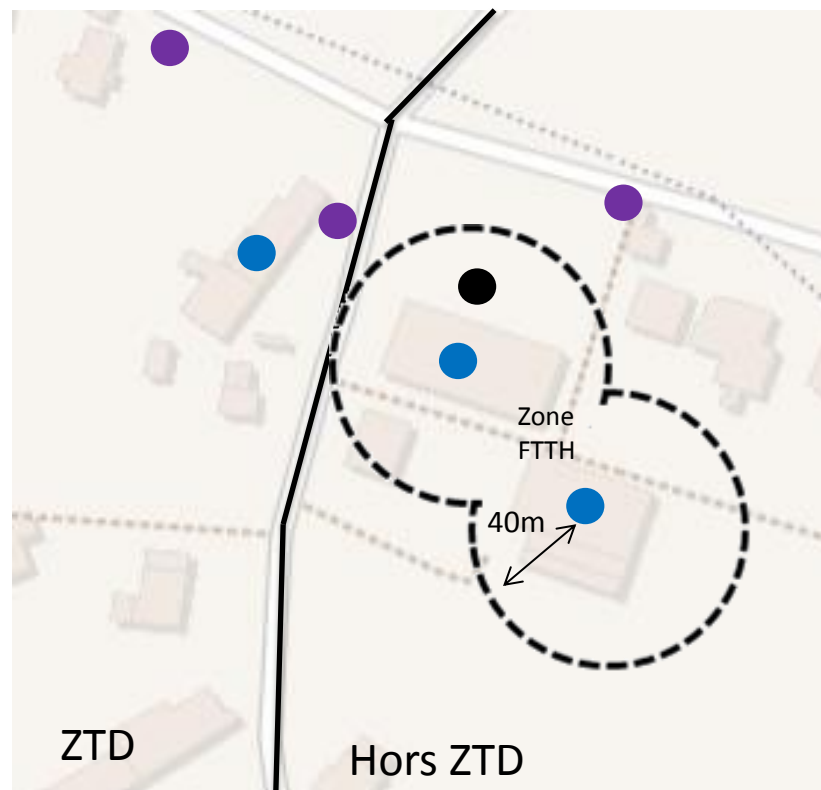
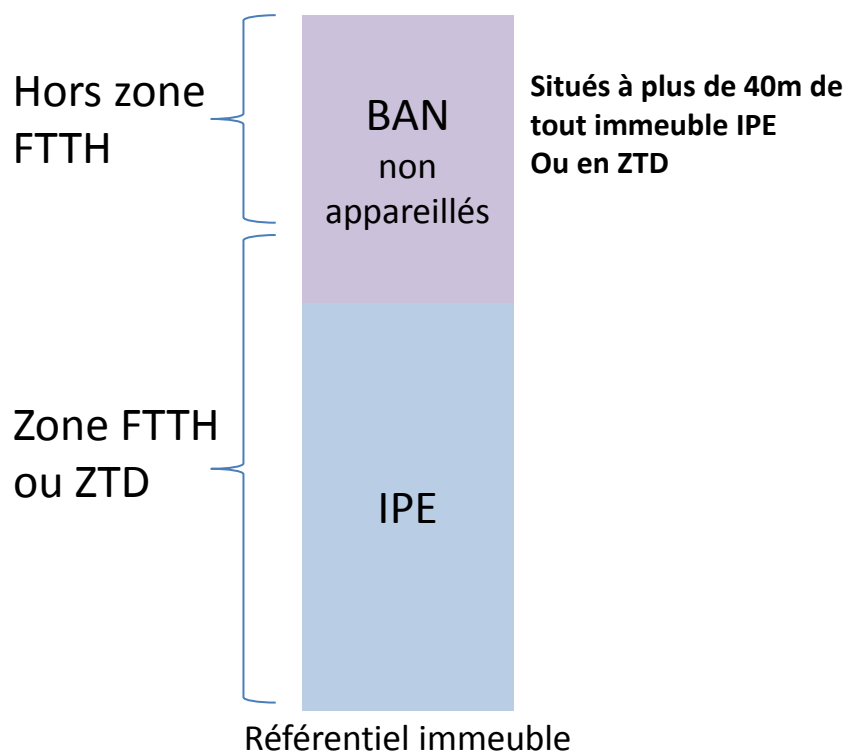
Référentiel immeuble



Le référentiel d'immeubles

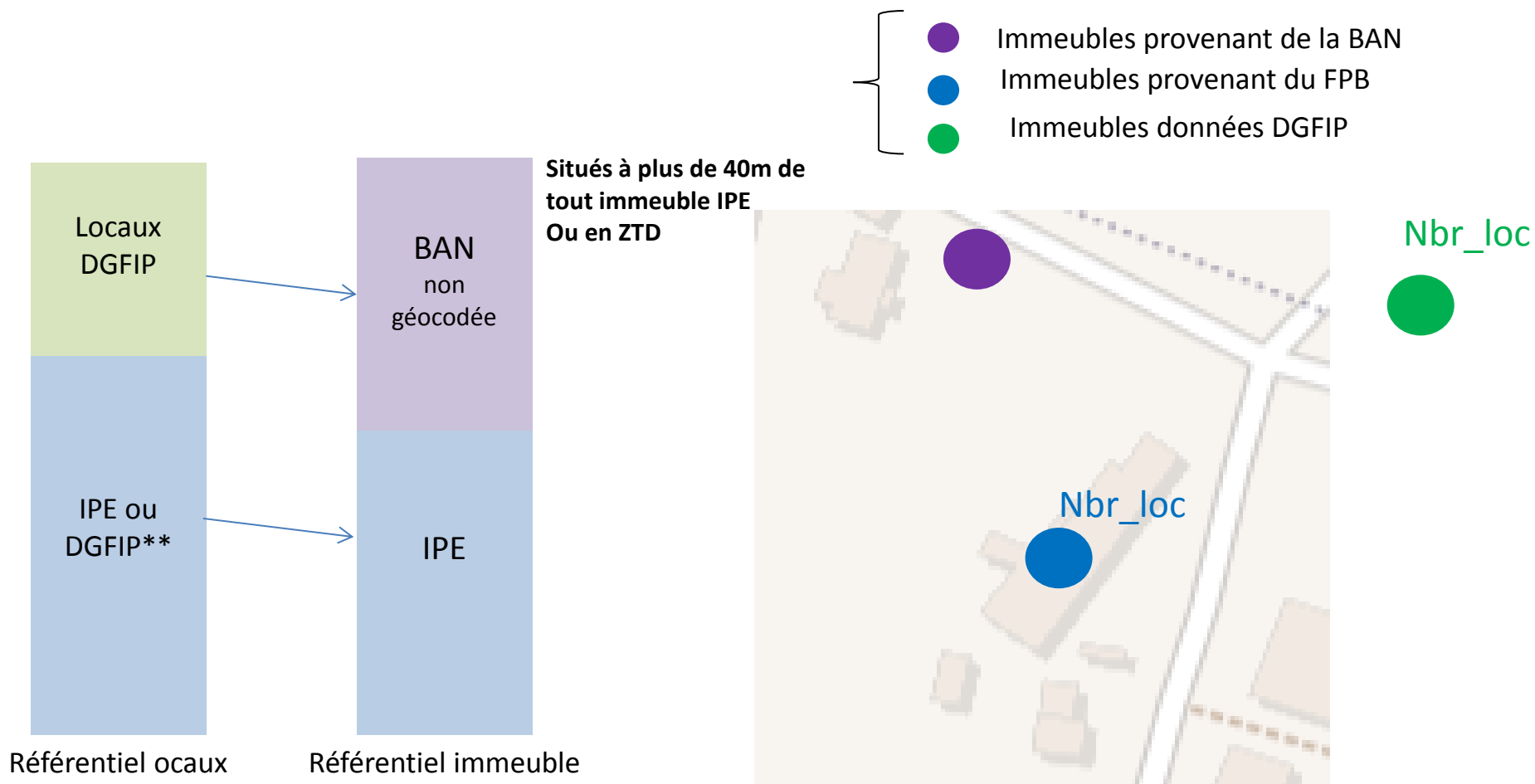
Le complément BAN

- Immeubles provenant de la BAN
- Immeubles provenant du FPB
- Immeubles BAN supprimé



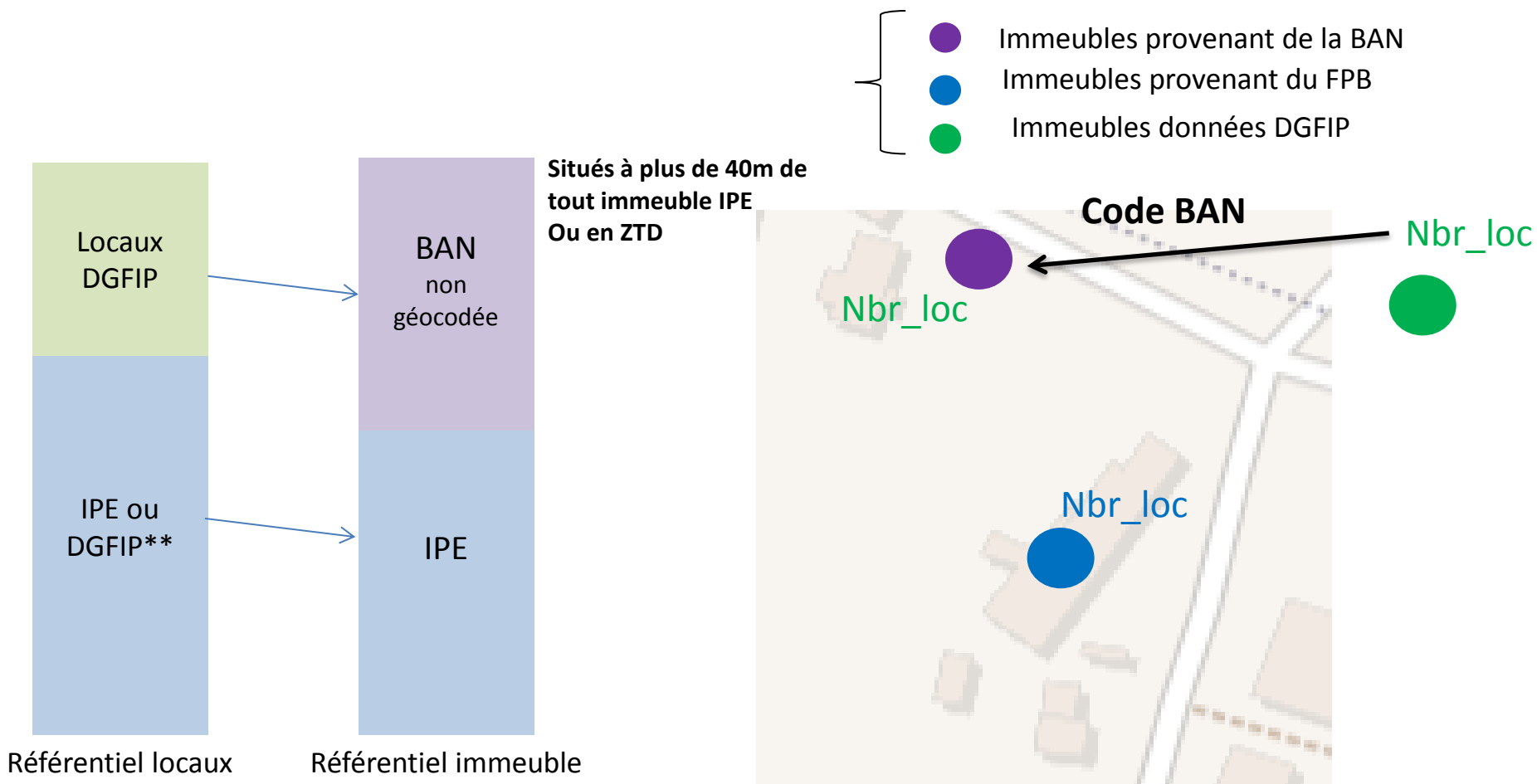
Le référentiel de locaux

Etape n°2 : le rattachement des locaux de l'IPE et du FPB



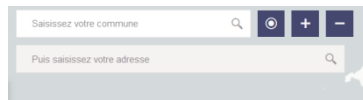
Le référentiel de locaux

Etape n°2 : le rattachement des locaux de l'IPE et du FPB grace au géocodage des données DGFIP à la BAN



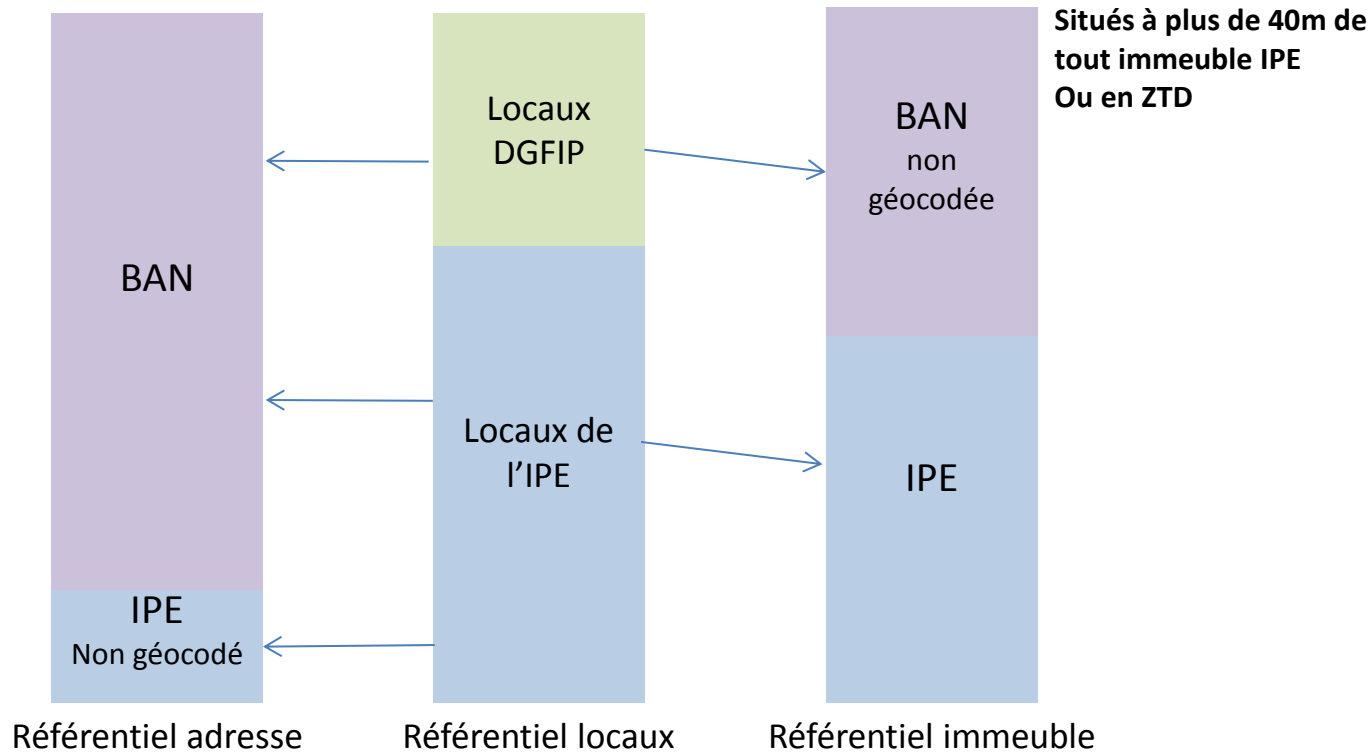
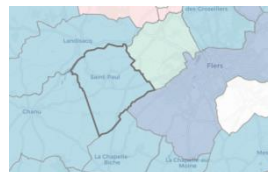
Le référentiel d'adresses

Faire primer la BAN



Saisissez votre commune

Puis saisissez votre adresse

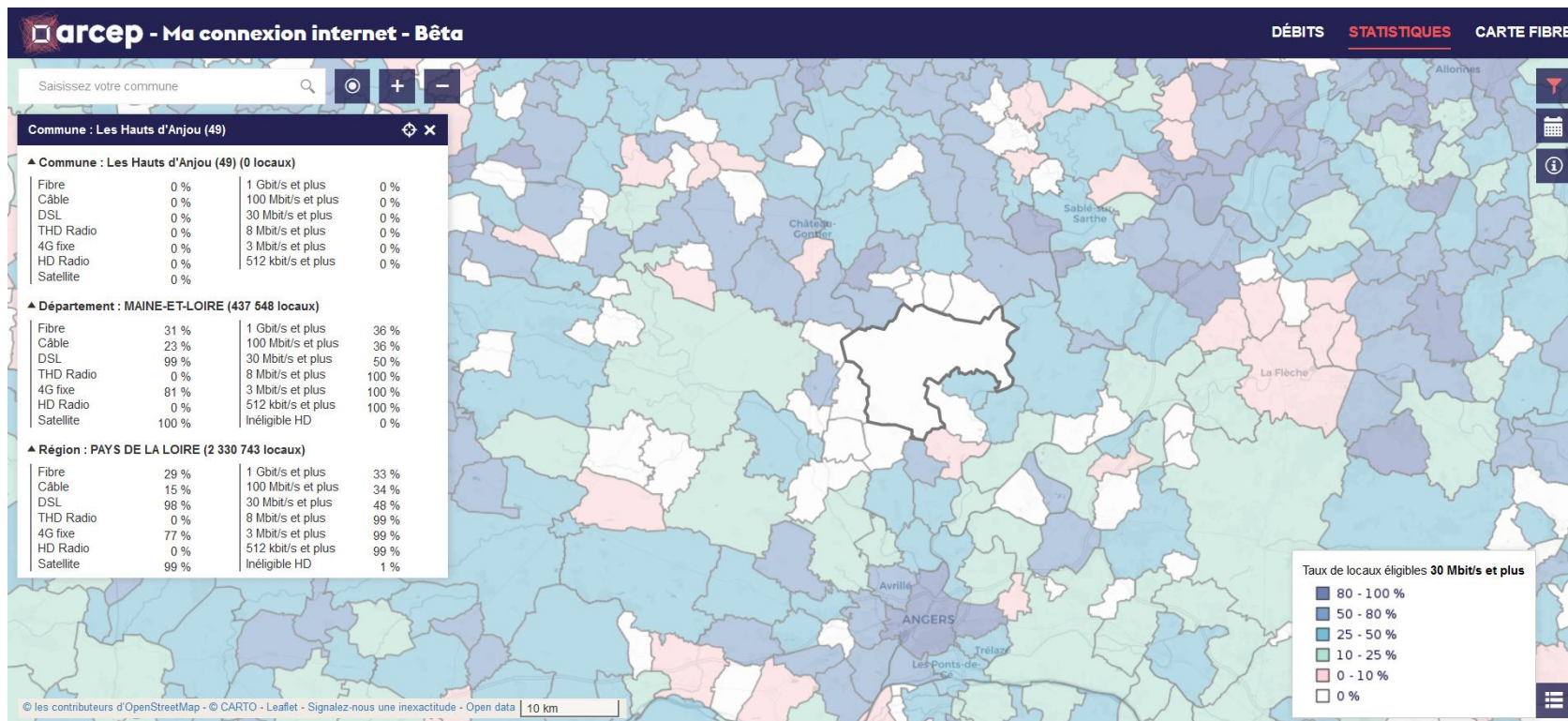


Les évolutions en cours des référentiels

306 communes vides en raison d'écarts entre le référentiel de communes BAN et celui du COG (fusions, communes nouvelles).

⇒ Objectif : Meilleure prise en compte des modifications de communes

Exemple de la commune Les Hauts-d'Anjou qui est une commune nouvelle créée de la fusion des communes Brissarthe, Contigné, Cherré, Champigné, Marigné, Sœurdres et Querré.

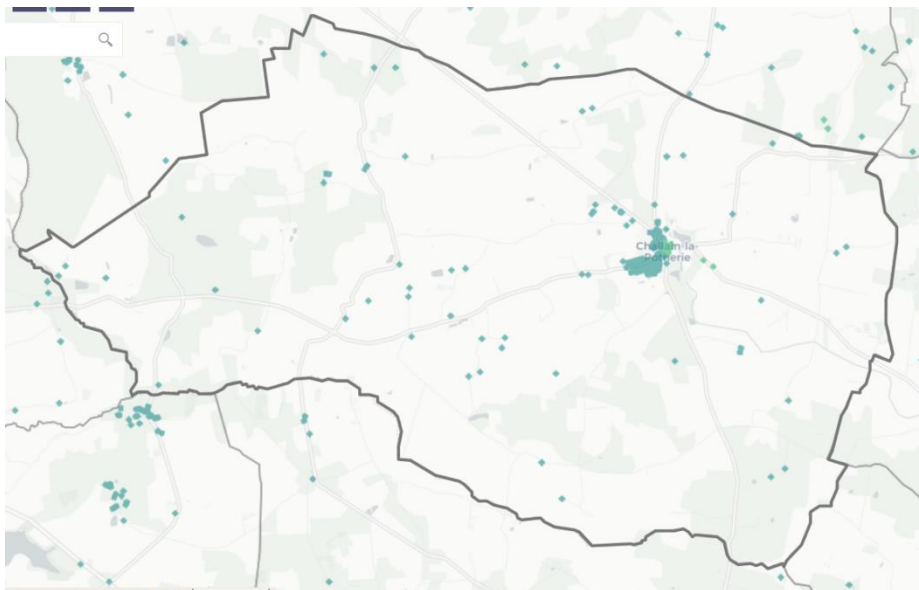


Les évolutions en cours des référentiels

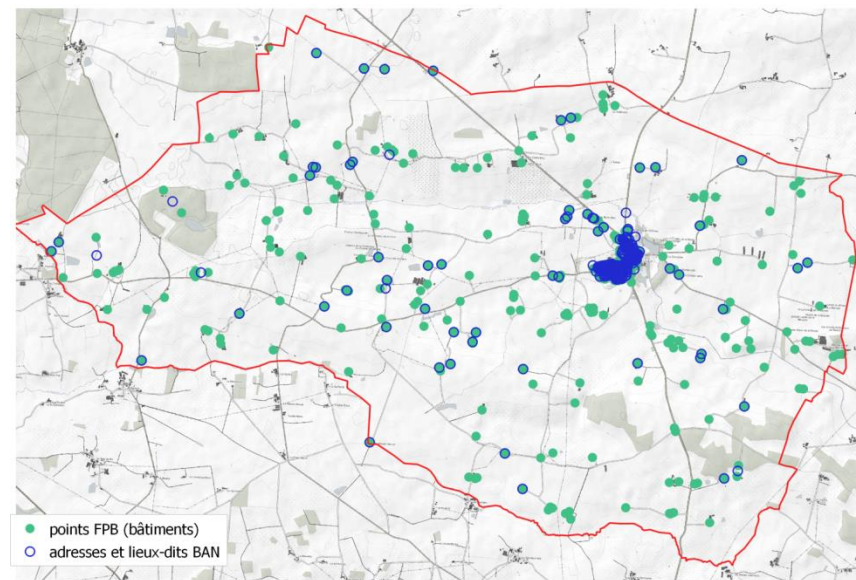
Meilleure prise en compte des immeubles dans les zones peu denses (lieux-dits).

Pour cela il est prévu d'utiliser la géométrie du cadastre (données DGFiP) en complément de la BAN.

Commune de Challain-la-potherie



actuellement



évolution prévue

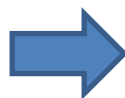
Les calculs d'éligibilité

L'éligibilité cuivre

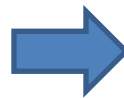
Immeuble est appareillé à
une ligne (géocodage)



affaiblissement de la
ligne au PC
[donnée d'entrée]



affaiblissement
du PC



affaiblissement à
l'immeuble



Le débit proposé à
l'immeuble



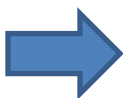
Immeuble n'est pas
appareillé à une ligne

L'éligibilité cuivre

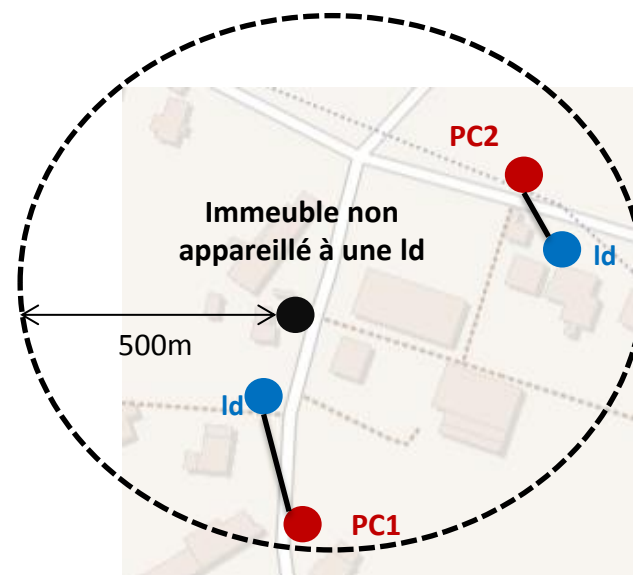
Cas des immeubles dont les code BAN ne se retrouvent pas dans les lignes géocodées

Hypothèses de départ :

- L'immeuble n'est pas situé en Zone Locale Immeuble Neuf



On associe à l'immeuble l'affaiblissement du PC de la ligne la plus proche (géocodé ou non), situé à moins de 500m

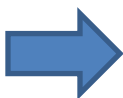


L'éligibilité cuivre

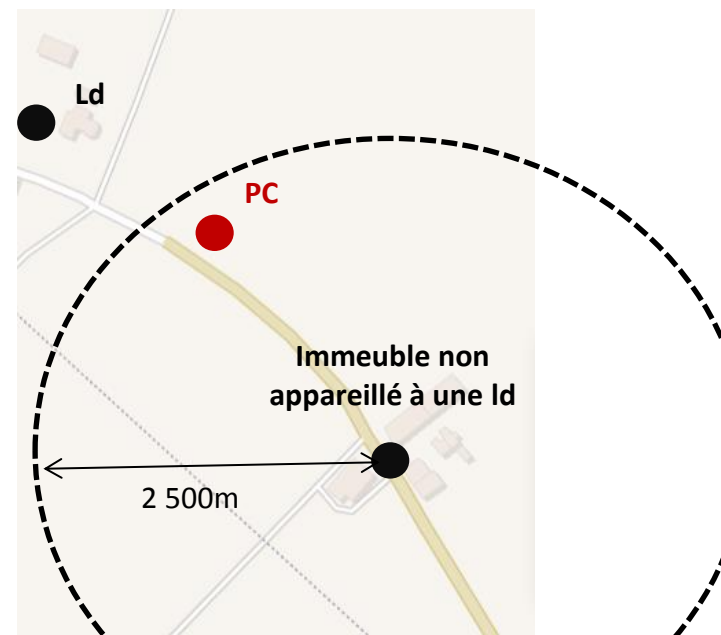
Cas des immeubles dont les code BAN ne se retrouvent pas dans les lignes géocodées

Hypothèses de départ :

- L'immeuble n'est pas situé en Zone Locale Immeuble Neuf
- Pas de ligne située à moins de 500m



on associe à l'immeuble l'affaiblissement du PC le plus proche, situé à moins de 2 500m



L'éligibilité cuivre

Calculer l'affaiblissement à l'immeuble à partir de l'affaiblissement au PC

affaiblissement_immeuble

=

affaiblissement_PC - 3.0 + ((longueur_pc_immeuble * | / 2) + longueur_technique) * 15 / 1000

⇒ longueur_technique = 100m

⇒ affaiblissement_PC :

- moyenne des affaiblissements au PC des lignes actives qui lui sont rattachées
- si le PC ne comporte **que des lignes inactives**, affaiblissement au PC maximal de toutes les lignes inactives qui lui sont rattachées

=> longueur_pc_immeuble : distance à vol d'oiseau entre les coordonnées du PC et celles de l'immeuble

L'éligibilité cuivre

Calculer le débit à l'immeuble à partir de l'affaiblissement à l'immeuble (tableaux de correspondance)

Débits descendants

| Seuils VDSL | Seuils ADSL | Débits correspondants |
|-------------|-------------|---------------------------|
| [0-11 dB[| | 100 Mbits/s – 30 Mbits/s |
| [11-34 dB[| [0-34 dB[| 30 Mbits/s – 8 Mbits/s |
| [34-49 dB[| [34-49 dB[| 8 Mbits/s – 3 Mbits/s |
| [49-78 dB[| [49-78 dB[| 3 Mbits/s – 0,512 Mbits/s |
| [78 dB-...] | [78 dB-...] | Inéligible |

Débits montants

| Seuils VDSL | Seuils ADSL | Débits correspondants |
|-------------|-------------|---------------------------|
| [0-10 dB[| | 8 Mbits/s – 3 Mbits/s |
| [10-14 dB[| | 8 Mbits/s – 3 Mbits/s |
| [14-78 dB[| [0-78 dB[| 3 Mbits/s – 0,512 Mbits/s |
| [78 dB-...] | [78 dB-...] | Inéligible |

technologie du NRA (VDSL/ADSL)



OI

technologie des offres commerciales
(VDSL/ADSL)



OC

L'éligibilité cuivre

Calculer le débit à l'immeuble à partir de l'affaiblissement à l'immeuble

Débits descendants

| Seuils ADSL | Débits correspondants |
|-------------|---------------------------|
| | 100 Mbits/s – 30 Mbits/s |
| [0-34 dB[| 30 Mbits/s – 8 Mbits/s |
| [34-49 dB[| 8 Mbits/s – 3 Mbits/s |
| [49-78 dB[| 3 Mbits/s – 0,512 Mbits/s |
| [78 dB-...] | Inéligible |

Exemple d'un immeuble présentant un affaiblissement de 34 dB en ADSL se situe dans la classe de débit 3 à 8 Mbits/s.

Aux extrémités des classes, la **probabilité d'être dans telle ou telle classe est incertaine**.

Dans notre exemple, l'immeuble peut certainement avoir un débit proche mais supérieur à 8Mbits/s et se situer dans la mauvaise classe.

Face à la grande **dispersion** des débits cuivre, pour **fiabiliser l'affichage des données** ponctuelles il a été décidé de présenter des **classes légèrement superposées**.

Débit maximum en réception (hors satellite)

- 1 Gbit/s et plus
- 100 Mbit/s à 1 Gbit/s
- 20 Mbit/s à 100 Mbit/s
- 6 Mbit/s à 25 Mbit/s
- 2 Mbit/s à 10 Mbit/s
- 512 kbit/s à 4 Mbit/s
- Inéligible HD

L'éligibilité cuivre

Améliorations prévues

Si l'affaiblissement de l'immeuble est supérieur à 78 dB => inéligible

Or pour Orange c'est le seuil d'affaiblissement au PC qui permet de le déterminer.

Cela entraîne une augmentation trop importante du nombre de locaux inéligibles au DSL.

=> Améliorer la précision des calculs et la définition de l'éligibilité HD dans MCI.

L'éligibilité fibre

Hypothèses pour le calcul de l'éligibilité fibre

- état immeuble : raccordable à la demande ou déployé
- état du PM : déployé
- date de mis en service commerciale : antérieur à aujourd'hui ou null
- id de l'immeuble n'est pas dans les immeubles sans service (par opérateur commercial)



Immeubles éligibles



Débit indiqué par l'OC

L'éligibilité fibre

Améliorations prévues

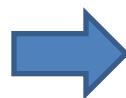
=> récupérer des données de la part d'un **plus grand nombre d'opérateurs commerciaux**

=> mettre en place un **nettoyage industrialisé des fichiers IPE** (environ 2 millions de locaux dans des immeubles en doublons ou fantômes). Ces immeubles et leurs locaux associés entraînent de fausses inéligibilités à la fibre.

L'éligibilité câble

Hypothèses pour le calcul de l'éligibilité câble

- immeubles appareillés à une adresse du fichier d'entrée
- Immeubles situés à moins de 50 mètres d'un PCC
- Présence d'un OC sur le PCC
- Adresse n'est pas dans les adresse sans service



Débit indiqué par l'OC



Immeubles éligibles

L'éligibilité des technologies hertziennes (4G fixe, radio, satellite)

Immeubles situés dans la couverture

et

sont reliés à un réseau/site sur lequel se
trouve un OC



Immeubles éligibles

| | Limitation débit montant | Limitation débit descendant |
|-----------------------------|-------------------------------------|--|
| 4G fixe | 8mbit/s | 30mbits/s |
| HD radio (wimax, wifimax..) | 3mbit/s | 30mbit/s |
| THD radio | 8mbit/s | 100mbit/s |
| Satellite | 30mbit/s | 30mbit/s |

L'éligibilité des technologies hertziennes (4G fixe, radio, satellite)

Amélioration prévue pour la 4G fixe

Des discussions sont en cours pour obtenir des données de couverture plus précises de la part des opérateurs => surévaluation de la couverture en 4G fixe

Amélioration prévue pour les technologies radio

Récupérer des données à partir d'un plus grand nombre d'opérateurs

Le géocodage

Le géocodage à l'API de la BAN

3 critères pour associer une adresse BAN à nos adresses des fichiers d'entrée :

- **2 critères sur la sémantiques** : un score supérieur à 0,6 et un type 'house number' ou 'locality'
- **1 critère géographique** : les coordonnées de l'adresse BAN et les coordonnées fournies par l'opérateur sont inférieure à

| Technologies | Type de coordonnées | Distance max |
|--------------|---------------------|--------------|
| Cuivre | Au PC | 1 500m |
| Fibre | A l'immeuble | 150m |
| Coax | A l'adresse | 150m |

Le géocodage au référentiel adresse – géocodage des données DFIP

Premier tour est un géocodage par égalité d'un quadruplet

| | Code BAN | Numéro de voie | Nom de la voie | Id_fantoir | Nom de la commune | Code_insee | répétition |
|-----------|----------------------|----------------|-------------------|------------|---------------------------|------------|------------|
| Table BAN | 30271_005 0_00272 | 272 | Chemin de l'Ecole | 0050 | Saint-Julien-de-Cassagnas | 30271 | b |

Est-ce que la concaténation du quadruplet est égale?

| | Code BAN | Code | Numéro de voie | Nom de la voie | Code_insee | répétition |
|--------------------|----------|------------------|----------------|-------------------|------------|------------|
| Table données DGFI | | 30271-0050-272-b | 272 | Chemin des Ecoles | 30271 | b |

Le géocodage au référentiel adresse – géocodage des données DFIP

Premier tour est un géocodage par égalité d'un quadruplet

| | Code BAN | Numéro de voie | Nom de la voie | Id_fantoir | Nom de la commune | Code_insee | répétition |
|-----------|------------------|----------------|-------------------|------------|---------------------------|------------|------------|
| Table BAN | 30271_0050_00272 | 272 | Chemin de l'Ecole | 0050 | Saint-Julien-de-Cassagnas | 30271 | b |

30271_0050_00272

**Est-ce que la concaténation du quadruplet est égale?
=> oui on associe le code ban à l'adresse FPB**

| | Code BAN | Code | Numéro de voie | Nom de la voie | Code_insee | répétition |
|----------------------|----------|------------------|----------------|-------------------|------------|------------|
| Table données DGFIIP | | 30271-0050-272-b | 272 | Chemin des Ecoles | 30271 | b |

Le géocodage au référentiel adresse – géocodage du FPB

Premier tour est un géocodage par égalité d'un quadruplet

| | Code BAN | Numéro de voie | Nom de la voie | Id_fantoir | Nom de la commune | Code_insee | répétition |
|-----------|----------------------|----------------|-------------------|------------|---------------------------|------------|------------|
| Table BAN | 30271_005 0_00272 | 272 | Chemin de l'Ecole | 0050 | Saint-Julien-de-Cassagnas | 30271 | b |

Est-ce que la concaténation du quadruplet est égale?

=> Non on ne peut pas associer de code ban à cette adresse FPB

| | Code BAN | Code | Numéro de voie | Nom de la voie | Code_insee | répétition |
|---------------------|----------|------------------|----------------|-------------------|------------|------------|
| Table données DGFIP | | 30271-0051-272-b | 272 | Chemin des Ecoles | 30271 | b |

Le géocodage au référentiel adresse – géocodage du FPB

Deuxième tour est un géocodage inverse

| | Code BAN | Numéro de voie | Nom de la voie | Id_fantoir | Nom de la commune | Code_insee | répétition |
|-----------|------------------|----------------|-------------------|------------|---------------------------|------------|------------|
| Table BAN | 30271_0050_00272 | 272 | Chemin de l'Ecole | 0050 | Saint-Julien-de-Cassagnas | 30271 | b |

Est-ce que la les trois champs entourés sont égaux?

=> Si oui on compare aussi le nom de la voie en calculant la distance de levenshtein

| | Code BAN | Code | Numéro de voie | Nom de la voie | Code_insee | répétition |
|---------------------|------------------|------------------|----------------|-------------------|------------|------------|
| Table données DGFIP | 30271_0050_00272 | 30271-0051-272-A | 272 | Chemin des Ecoles | 30271 | b |

Le géocodage au référentiel adresse – géocodage des données DFIP

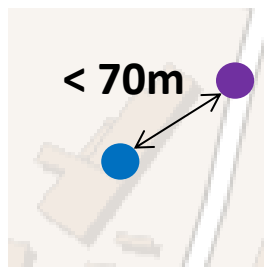
La distance de Levenshtein est égale au nombre minimal de caractères qu'il faut supprimer, insérer ou remplacer pour passer d'une chaîne de caractères à l'autre.

| | |
|-------------|-------------------|
| données BAN | Nom de la voie |
| | Chemin de l'Ecole |

| | |
|---------------|-------------------|
| données DGFIP | Nom de la voie |
| | Chemin des Ecoles |

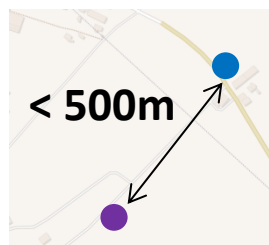
=> distance de Levenshtein =4

Si distance de Levenshtein ≤ 7 et



- ⇒ Même adresse
- ⇒ On associe le code BAN à l'adresse FPB

Si distance de Levenshtein ≤ 3 et



Le géocodage au référentiel adresse – géocodage du FPB

Amélioration sémantique puis relance du 2eme tour de géocodage

| | Code BAN | Numéro de voie | Nom de la voie | Id_fantoir | Nom de la commune | Code_insee | répétition |
|-----------|----------------------|----------------|-------------------|------------|---------------------------|------------|------------|
| Table BAN | 30271_005 0_00272 | 272 | Chemin de l'Ecole | | Saint-Julien-de-Cassagnas | 30271 | b |

**Si la répétition dans la BAN est « b »
on ne regarde l'égalité que du duo « numéro et nom de voie »
=> Puis on recommence le géocodage du 2eme tour**

| Table données DGFI | Code BAN | Code | Numéro de voie | Nom de la voie | Code_insee | répétition |
|--------------------|------------------|------------------|----------------|-------------------|------------|------------|
| | 30271_0050_00272 | 30271-0050-272-A | 272 | Chemin des Ecoles | 30271 | bis |

Le géocodage au référentiel adresse – géocodage des données DFIP

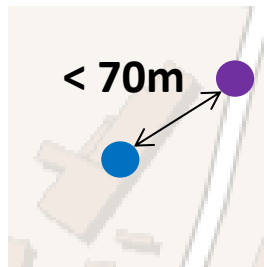
La distance de Levenshtein est égale au nombre minimal de caractères qu'il faut supprimer, insérer ou remplacer pour passer d'une chaîne de caractères à l'autre.

| | |
|-------------|-------------------|
| données BAN | Nom de la voie |
| | Chemin de l'Ecole |

| | |
|---------------|-------------------|
| données DGFIP | Nom de la voie |
| | Chemin des Ecoles |

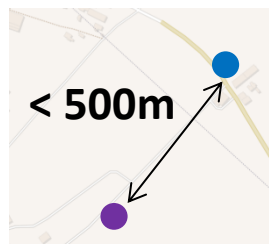
=> distance de Levenshtein =4

Si distance de Levenshtein ≤ 7 et



- ⇒ Même adresse
- ⇒ On associe le code BAN à l'adresse FPB

Si distance de Levenshtein ≤ 3 et

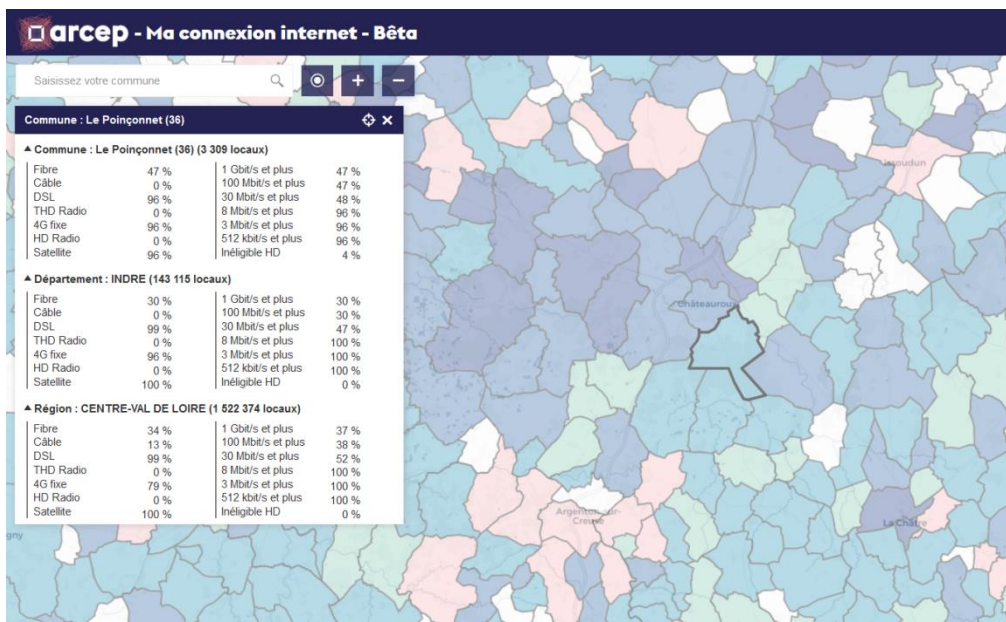
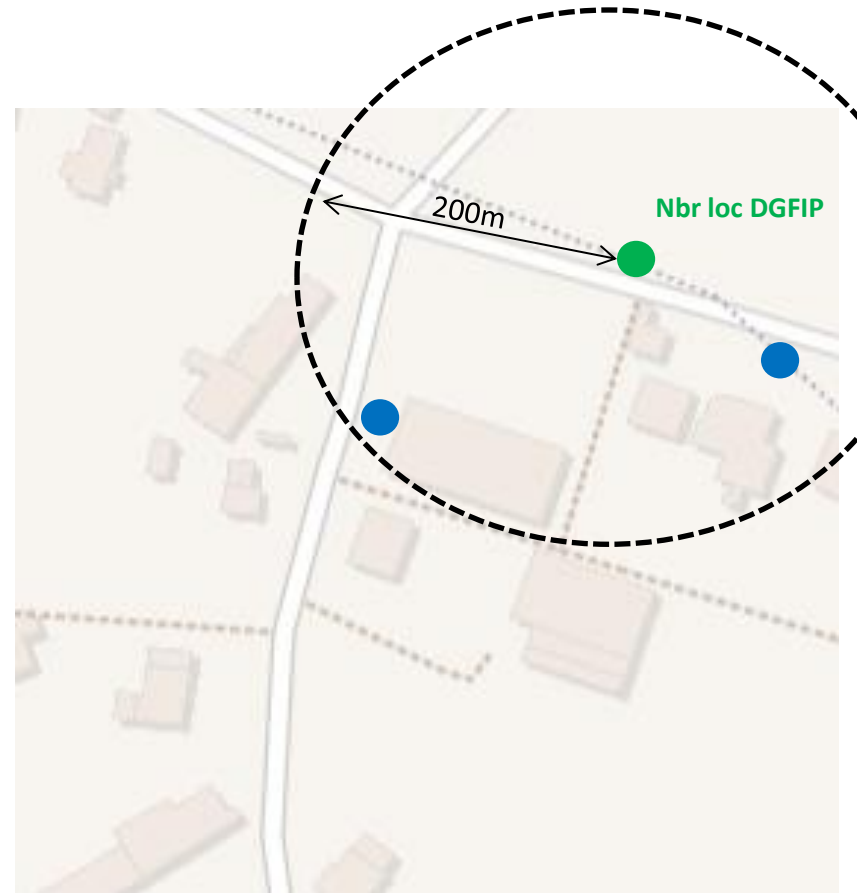


Le géocodage au référentiel adresse – géocodage des données DFIP

3^{ème} tour de géocodage basé que sur la distance

On associe les adresses DGFIIP non géocodées à l'adresse BAN la plus proche et situées dans la même commune

=> Le plus petit niveau administratif de nos cartes statistiques est la maille de la commune





Suite de la phase d'échanges :

Groupe de travail expert 1 : Réutiliser les données de Ma connexion internet