



**Cerema**

Centre d'études et d'expertise sur les risques,  
l'environnement, la mobilité et l'aménagement

# La Base des Arrêts de Transport Collectif - BATO -

Silvio Rousic – Cerema

# Motivation

L'intérêt d'un référentiel des arrêts de transport collectif est de disposer d'un **identifiant unique** utilisable par différentes applications, notamment pour l'information voyageur:

- calcul d'itinéraires et affichage des prochains passages temps réel
- billettique
- maintenance des équipements
- Etc...

En France (en 2016), seule l'Ile-de-France (le STIF) avait créé un référentiel d'arrêts régional (appelé Reflex) avec des Id uniques et pérennes (profil Netex). Projet « Pass-Pass » (du SMIRT) pour le Nord-pas-de-Calais.

Rien en vue pour les autres régions, mais outre le fait que cela puisse prendre plusieurs années, cela engendrerait autant de référentiels que de régions → Quid des identifiants uniques ?



# Genèse de la proposition

Pour différentes raisons, un **projet national** n'a pas encore émergé, alors qu'un profil (modèle, format d'échange, définition des identifiants) pour les données d'arrêt, partagé au niveau national et basé sur la norme européenne Netex, a été élaboré en 2013 à l'initiative de l'**AFIMB\***.

Le Cerema disposait de la **base ATC** (commanditée par le SOeS), établie à partir de données Open Data ou sous convention et des arrêts d'OSM. La base ATC utilise les champs définis dans le modèle d'arrêts partagé de l'AFIMB.

En prenant exemple sur le modèle de la Base Adresses Ouverte **BANO** et suite à des premiers échanges avec la communauté OSM France, le Cerema a proposé de créer une Base Arrêts de Transport Ouverte **BATO**, s'appuyant sur la base ATC, qui permettrait d'initier un référentiel national des arrêts de transport collectif.

\* *AFIMB = agence française pour l'information multimodale et la billettique*

# Périmètre

Le périmètre de la BATO s'est limité aux arrêts « physiques » (par opposition aux regroupements en arrêts commerciaux ou pôles d'échange).

La BATO est **gérée indépendamment de la base OSM et regroupe des données de plusieurs sources** :

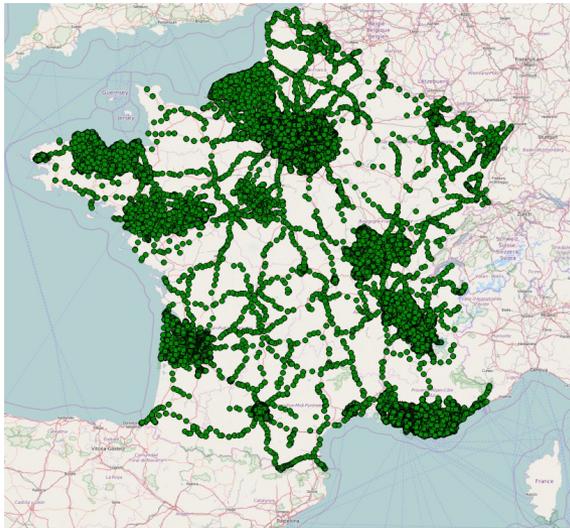
- OSM bien sûr
- Open data – données structurées (GTFS, Neptune) ou SIG (CSV, SHP)
- Référentiel officiel d'arrêts locaux (base Reflex du STIF)

Les données sont réutilisables (sous licence ODBL).

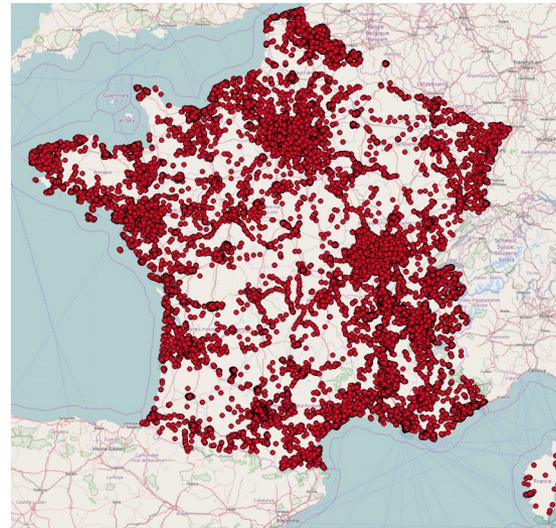
# Constitution initiale

L'initialisation de la BATO a consisté à:

- extraire certains arrêts de la base ATC (réseaux open data uniquement) et fusionner en une seule couche
- intégrer les arrêts OSM



Base ATC = 112 817 arrêts



Base OSM = 98 933 arrêts

# Traitements spécifiques

Il a fallu effectuer quelques **corrections** sur les données collectées :

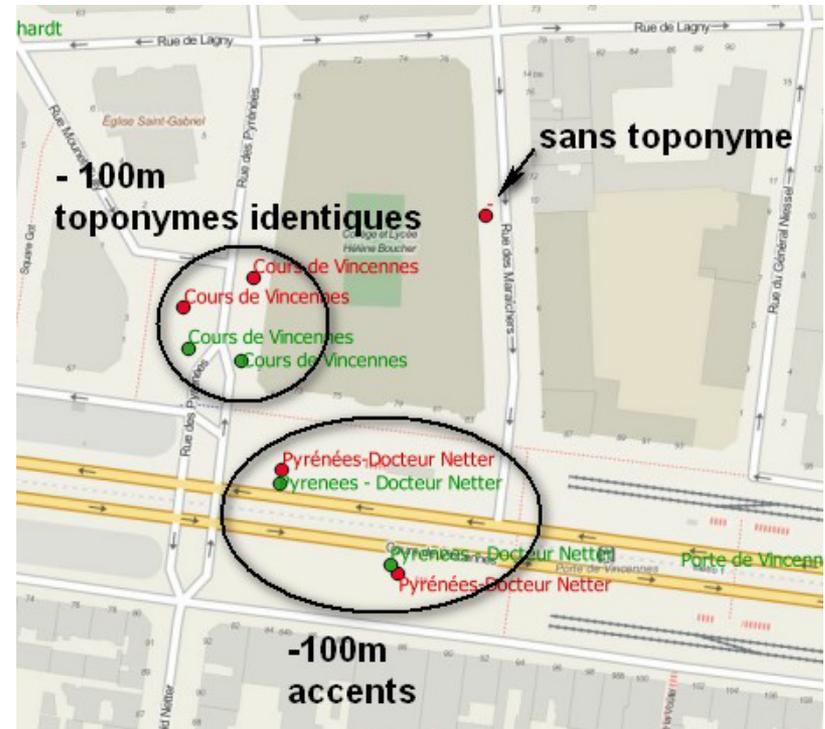
- suppression d'arrêts OSM **sans toponymes**
- suppression de « **doublons** » base ATC/OSM établis selon certains critères :

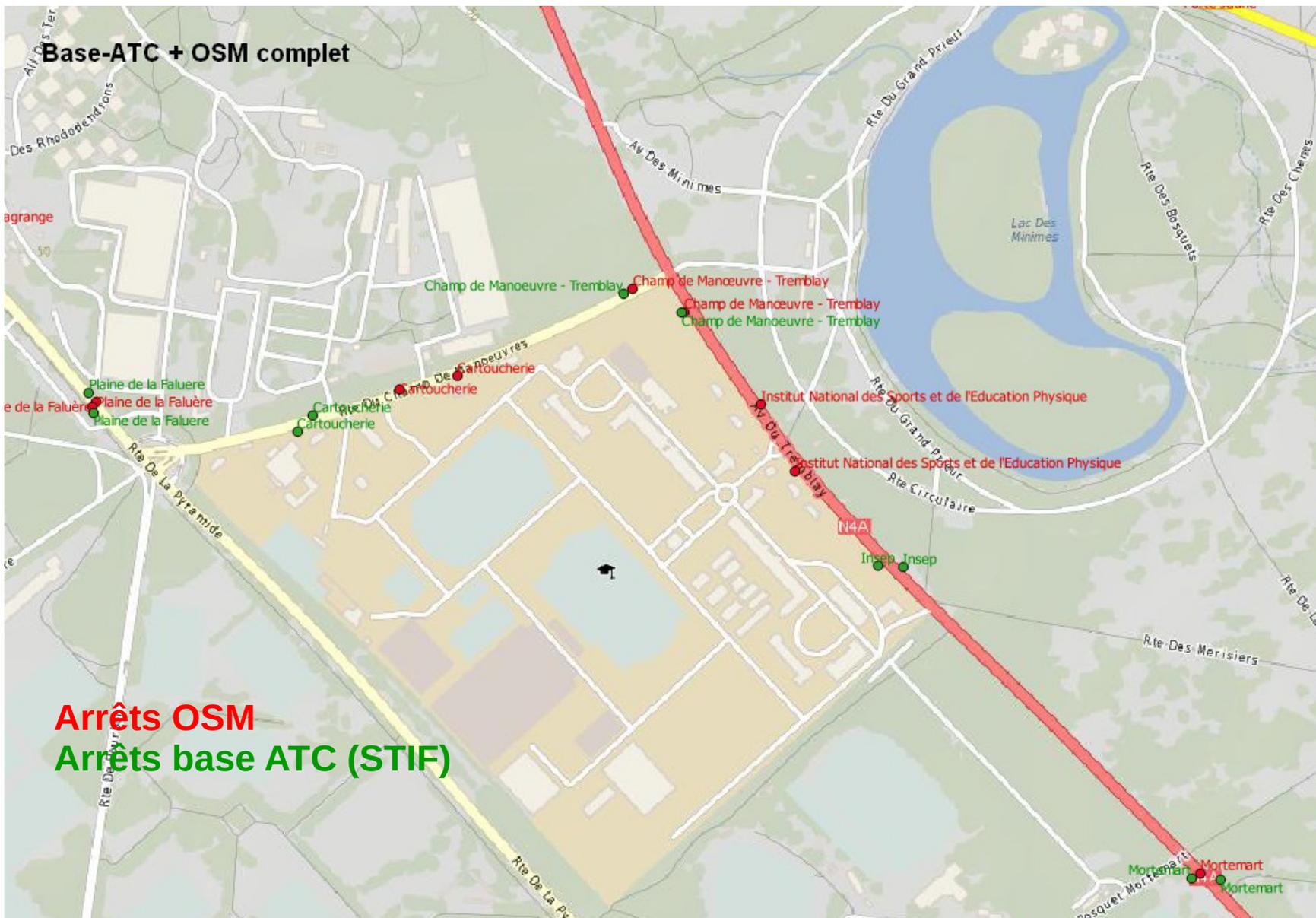
- Distance < 100m
- toponymes identiques

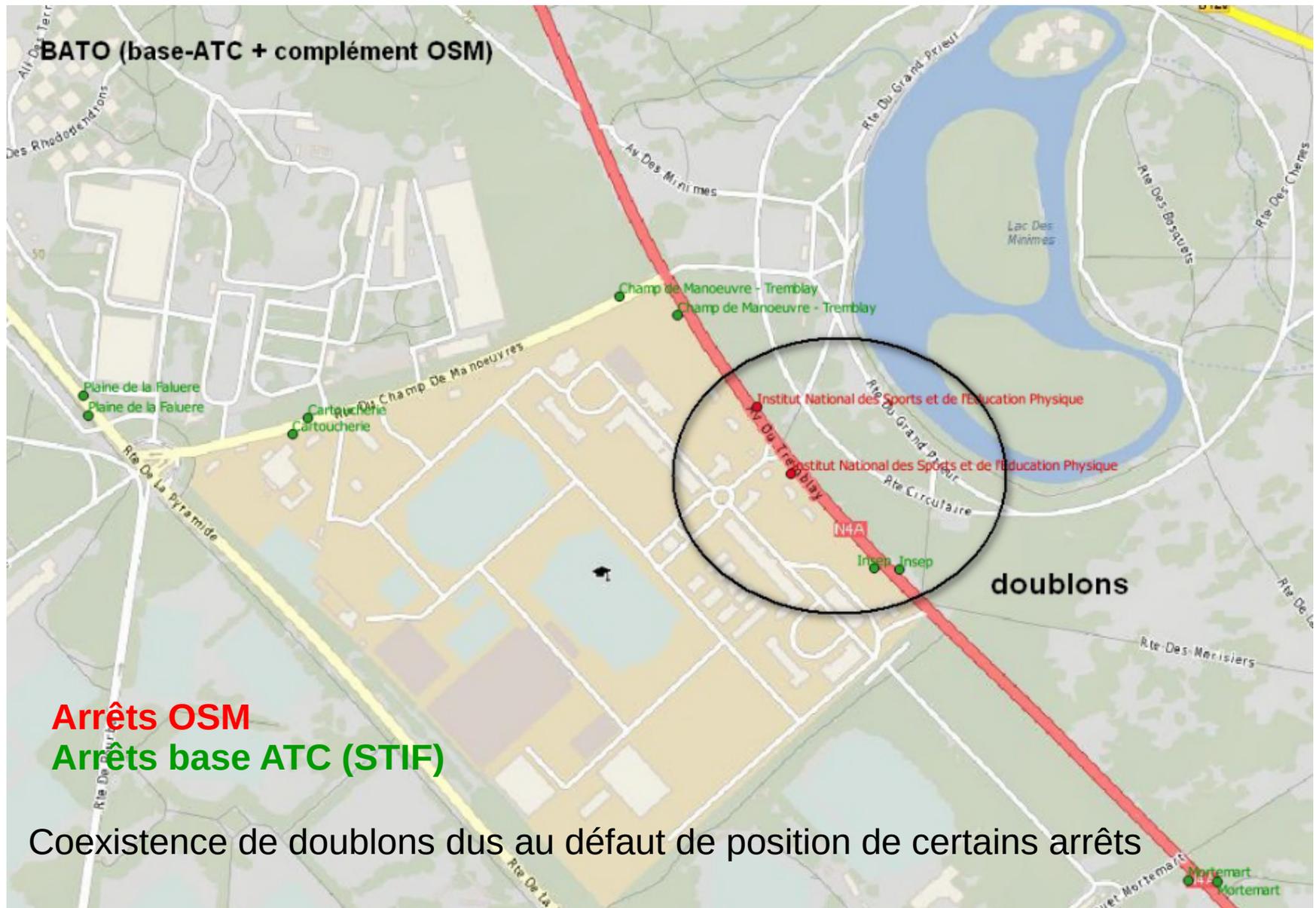
**Normalisation** des toponymes :

- les accentuations
- les abréviations
- la casse

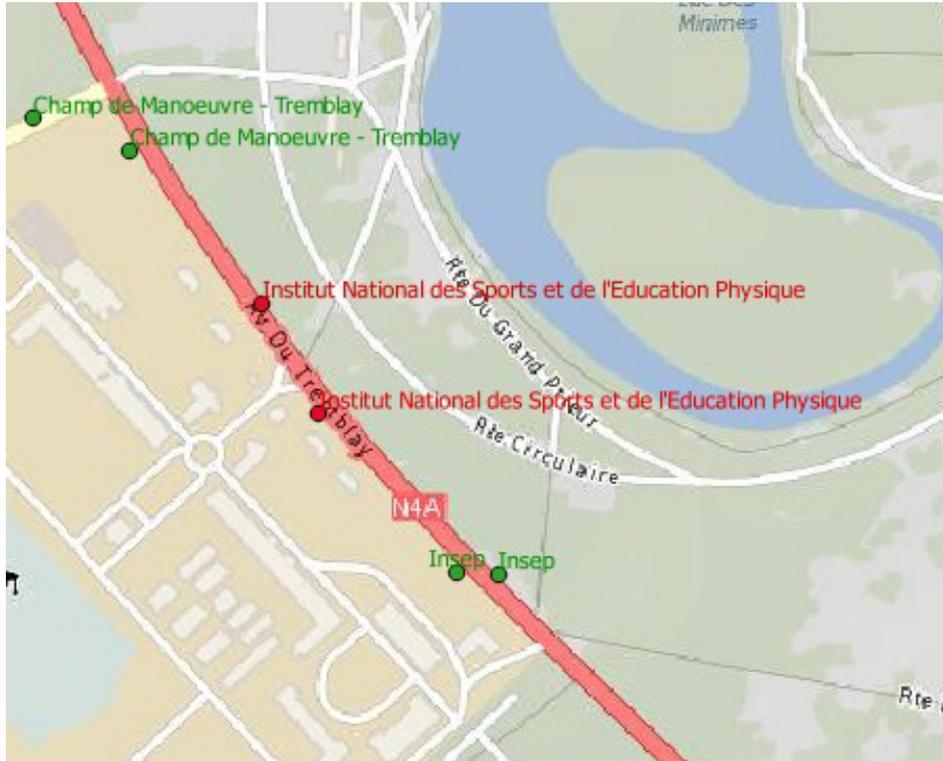
Mais quelques doublons subsistent...







# Doublons : qui a raison ?



## Arrêts OSM:

- toponyme en entier (mais erroné\*)
- position correcte

## Arrêts base ATC (STIF):

- abréviation correcte
- position fausse

\* *INSEP* = *Institut national du sport, de l'expertise et de la performance*

En présence de doublons, le choix de l'arrêt à conserver n'est pas toujours simple... ni facilement automatisable.

# Résultat final

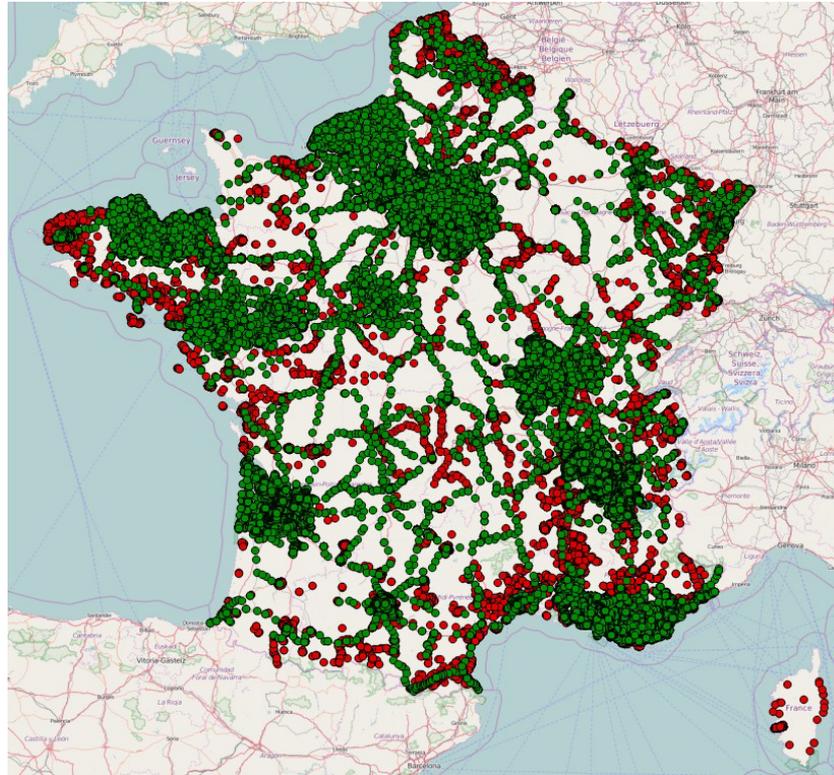
A l'issue de ces traitements, la **1ère version de la base BATO** contient :

112 817 arrêts ATC  
57 976 arrêts OSM

Détection des doublons =  
- 40 957 arrêts OSM

→ arrêts open data et du  
STIF privilégiés

+ Identifiant Netex !



Soit **170 793 arrêts pour la base BATO** (nb d'arrêts des SIM : 313 521)



# Un constat

Le projet BATO a dressé un bilan sur la difficulté d'avoir accès à des données ouvertes dans la thématique des transport collectifs, mais également un besoin de complétude des données collaboratives.

Fort de ce constat, le Cerema a présenté en 2016 la BATO lors de la 4ème édition du *State of the Map*, organisée par l'association OpenStreetMap France (à Clermont-Ferrand).

L'objectif était de solliciter la communauté OSM sur la préfiguration d'un référentiel afin d'avancer en l'absence de projet national.

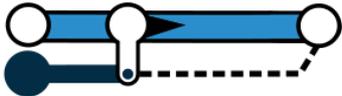
On peut aujourd'hui dire que ce vœu est en partie exaucé avec le projet « **Jungle Bus** », qui vise à faciliter la contribution à OpenStreetMap pour les données liées au transport public.

Des progrès restent encore à faire sur l'ouverture des données.

# Merci de votre attention

Silvio Rousic – Cerema

**BATO**



<https://github.com/bato-fr>

<https://twitter.com/ProjetBATO>