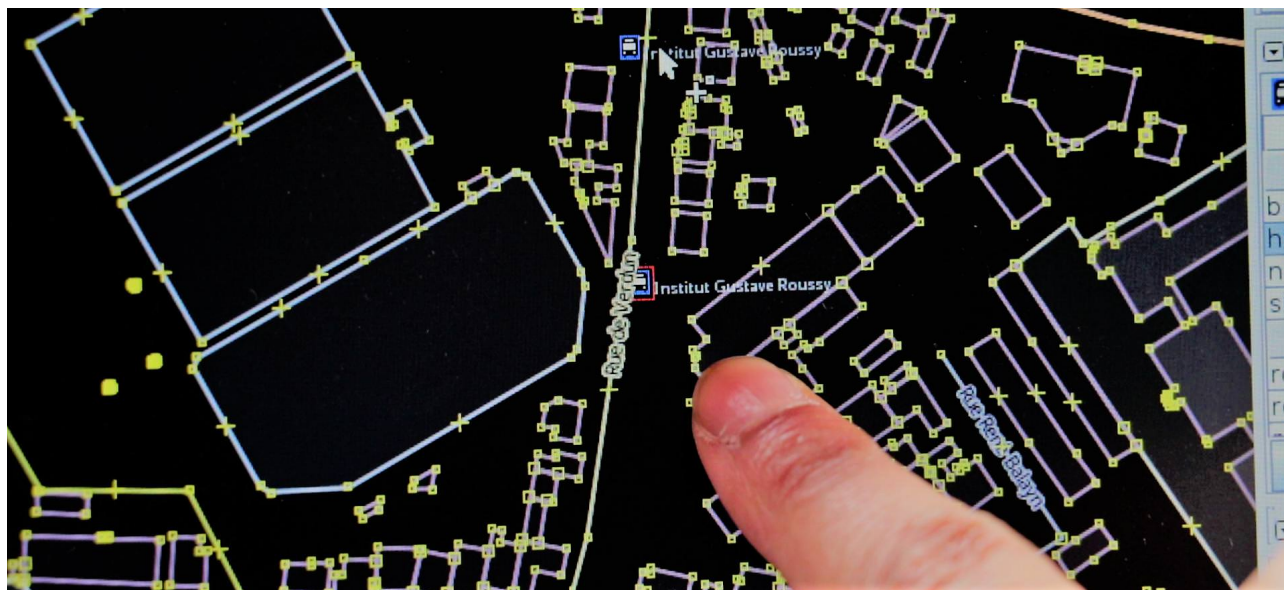


Audit comparatif des données Open Data des arrêts de bus de la région Île-de-France - mise à jour de février 2019

Publié le 28 février 2019 par Jungle Bus (<https://junglebus.io>)



Ce document est une analyse qui approfondit et complète l'audit publié en avril 2018 (http://junglebus.io/iledefrance/audit_2018_04/) puis ses mises à jour de juillet 2018 (http://junglebus.io/iledefrance/audit_2018_07/) et novembre 2018 (http://junglebus.io/iledefrance/audit_2018_11/).

Introduction

Île-de-France Mobilités publie en Open Data les données des lignes et des arrêts de transport d'Île-de-France.

L'association Jungle Bus s'est donnée pour mission de cartographier les réseaux de bus et de soutenir les contributeurs OpenStreetMap qui poursuivent ce but, et ce partout dans le Monde. Elle porte en outre une attention toute particulière aux réseaux

de transport d'Île-de-France.

Une première étude a été réalisée par Jungle Bus en avril 2018. Son objectif était de comparer les données d'Île-de-France Mobilités et celles d'OpenStreetMap afin d'évaluer dans quelle mesure on pouvait utiliser l'une pour enrichir l'autre. Une méthodologie ainsi que des indicateurs y ont été introduits ; le présent document propose d'étudier les évolutions de ces indicateurs suite aux mises à jour de données effectuées depuis. À noter que deux précédentes mises à jour ont déjà été publiées dans ce but.

L'audit mené en avril a révélé des disparités de modélisation des objets de transport et propose en conséquence un modèle pour comparer le plus petit dénominateur commun entre les deux sources. C'est le routepoint, défini comme un arrêt desservi par une ligne vers une destination.

Sur cette base, l'étude a mis en lumière l'hétérogénéité des contributions OpenStreetMap ainsi que leur non-exhaustivité, notamment concernant les lignes de bus.

Malgré cette incomplétude, sur une sélection restreinte permettant de garantir une comparaison des mêmes objets entre les deux référentiels, on constate qu'OpenStreetMap possède des données d'excellente qualité qui permettent de proposer des enrichissements aux données officielles d'Île-de-France Mobilités sur un panel de sujets utiles à l'information voyageur.

Vous pouvez consulter l'audit initial dans son intégralité (http://junglebus.io/iledefrance/audit_2018_04/), ainsi que ses mises à jour de juillet 2018 (http://junglebus.io/iledefrance/audit_2018_07/) et novembre 2018 (http://junglebus.io/iledefrance/audit_2018_11/) pour obtenir la méthodologie de calcul de chaque indicateur proposé ainsi que des conclusions plus détaillées.

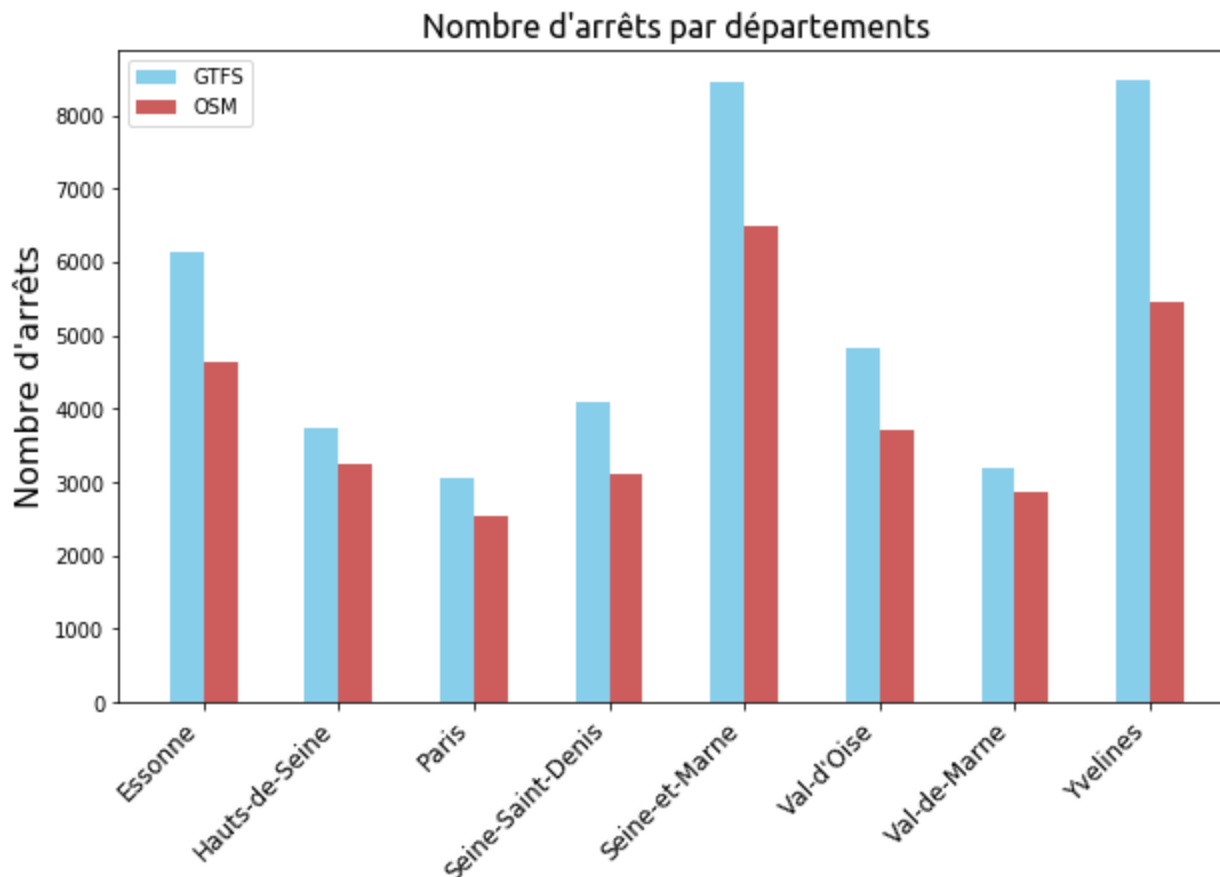
Voyons à présent de quelle manière les chiffres ont évolué.

Analyses quantitatives

Nombre d'arrêts

Voici la répartition du nombre relatif d'arrêts dans les deux référentiels, synthétisée

par département.



Le pourcentage global de couverture d'OpenStreetMap en arrêts sur la région Île-de-France est de **76 %**.

Comme lors de nos précédentes mises à jour, on constate une légère progression de ce chiffre. Le département de l'Essonne est celui dans lequel on constate la plus grande amélioration de couverture.

Nombre de lignes

Le pourcentage global de couverture en lignes de bus sur la région Île-de-France est de **69 %** (contre un peu moins de 50 % en avril dernier).

Le nombre de lignes cartographiées par les contributeurs OpenStreetMap continue donc de croître régulièrement.

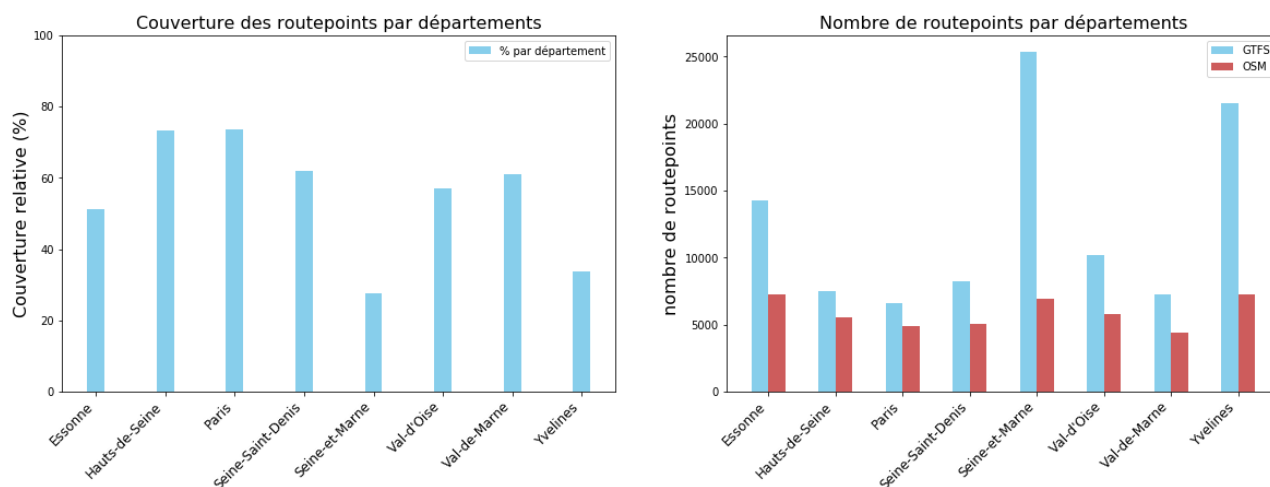
Nombre de routepoints par département

Pour mitiger les divergences de modélisation entre les deux sources, depuis avril 2018 notre étude s'articule largement autour du **route point**, c'est-à-dire un arrêt logique représentant un arrêt de bus desservi par une ligne dans une direction donnée.

Voici la répartition du nombre relatif de route points par département.

Le graphique de gauche représente la couverture en route points de chaque département de la Région.

Le second graphique présente le nombre de route points de chaque département (en bleu dans l'open data d'Île-de-France Mobilités, et en rouge dans OpenStreetMap).



Le pourcentage global de couverture en route points sur la région Île-de-France est de **47 %** contre **25 %** en avril dernier.

Nombre de tracés de lignes

On constate que **43%** de lignes disposent d'un tracé dans OpenStreetMap (cela correspond à 829 lignes, soit 62% des lignes existantes dans OpenStreetMap). Pour rappel, Île-de-France Mobilités ne publie pas de tracés de lignes en open data que l'on pourrait utiliser pour comparaison.

Ces tracés pourraient cependant venir enrichir les données publiées en open data, sous la forme de tracés de circulation dans le GTFS ou comme base pour réaliser des supports cartographiques.

Analyses qualitatives

L'objectif des analyses qualitatives est d'évaluer l'écart entre les deux sources sur un ensemble de critères, afin de déterminer dans quelle mesure il est possible d'enrichir les données Open Data officielles à l'aide des données OpenStreetMap.

Rappels méthodologiques : Nous ne conserverons que les routepoints (arrêt logique représentant un arrêt de bus desservi par une ligne dans une direction donnée) des deux sources que nous avons pu faire correspondre, en utilisant les référentiels d'Île-de-France Mobilités (REFLEX pour les arrêts, CODIFLIGNE pour les lignes) et une comparaison de chaînes de caractères pour les directions.

Seuls 27% des routepoints sont ainsi retenus pour la suite de cette étude (contre 17% lors du précédent audit).

Cette méthode est volontairement restrictive afin de limiter les faux positifs.

Les données mises en correspondances des deux sources sont disponibles au téléchargement à cette adresse (https://raw.githubusercontent.com/Jungle-Bus/ref-fr-STIF/master/audit_routepoints/audits/2019_02/images/mapping_des_routepoints.csv).

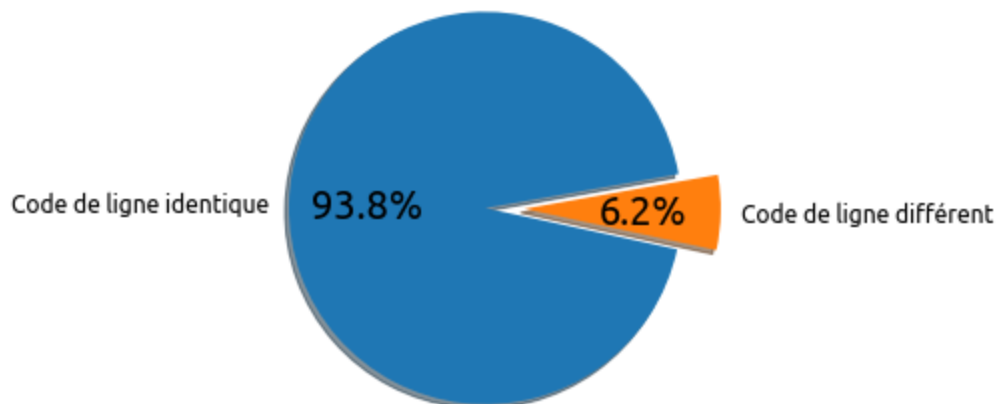
Qualité des numéro de lignes

Lors de l'audit initial, nous avons constaté

- qu'environ 95 % des lignes avaient un code de ligne identiques entre les deux sources
- qu'OpenStreetMap propose une dénomination plus proche de ce que constatera un voyageur sur le terrain dans les cas restants

Nous constatons en février que l'écart s'est creusé puisqu'à présent plus de 6% des lignes ont un code différent :

Analyse de concordance des codes de lignes sur les routepoints



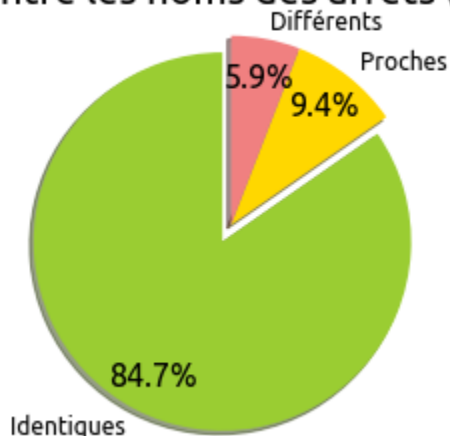
Qualité des noms

Lors de l'audit initial, nous avons constaté que les données du référentiel officiel sur le nommage des arrêts sont très hétérogènes et manquent d'uniformisation en comparaison avec celles d'OpenStreetMap.

Cependant, après un traitement de remise en cohérence sur les données Open Data, environ 95 % des arrêts avaient un nom identique ou proche dans les deux sources.

Sur ce sujet, nous constatons que ces constats restent d'actualité.

Similitude entre les noms des arrêts (sans casse)



Qualité des positions

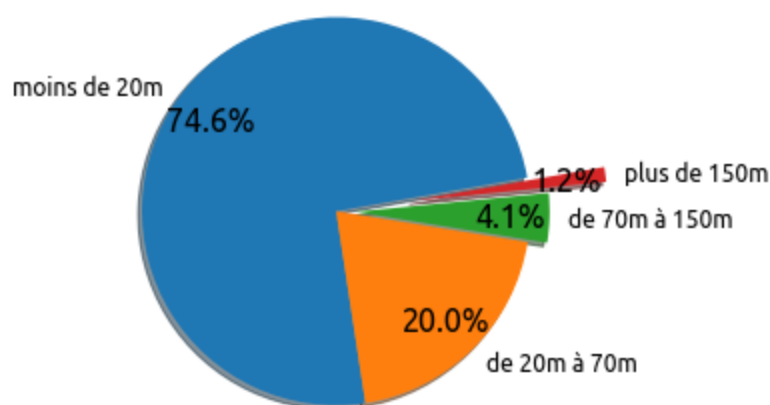
Lors de l'audit initial, nous avons constaté que l'essentiel des arrêts officiels se trouve

à une distance raisonnable de leurs homologues d'OpenStreetMap.

Cependant, quelques centaines d'arrêts étaient situés dans le GTFS à plus de 70 mètres de leur version crowdsourcée par la communauté.

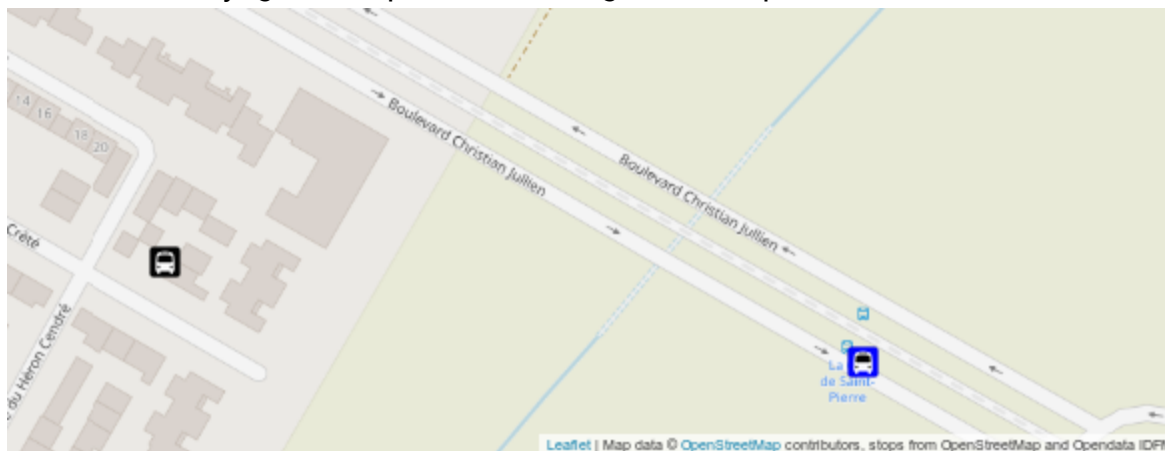
De plus, au cours de nos mises à jour de l'audit, nous avons pu constater une augmentation du nombre de routepoints mal géolocalisée plus rapide que l'augmentation du nombre de routepoints pris en compte dans cette étude. Ces constats restent d'actualité et nous comptons désormais plus de 5% des routepoints à plus de 70 mètres.

Distance entre les arrêts officiels et OSM



Certains de ces arrêts sont placés à plusieurs kilomètres de leur position réelle, à l'image de l'arrêt "Haras Campus Saint-Thérèse" qui devrait être dans la commune d'Ozoir-la-Ferrière mais se retrouve, d'après l'Open Data officiel dans celle de Chaumes-en-Brie.

D'autres encore, situés dans des bâtiments, ne permettent pas de proposer une information voyageur de qualité aux usagers susceptibles de l'utiliser :



Arrêt *La Clef de Saint-Pierre*, Tzen 1 en direction de *Lieusaint Moissy Gare RER*, en noir d'après les données officielles d'Île-de-France Mobilités, en bleu d'après OpenStreetMap

Enfin, afin de rendre possible l'amélioration des données officielles à partir d'OpenStreetMap, le détail des écarts de positions est disponible en téléchargement à cette adresse (https://raw.githubusercontent.com/Jungle-Bus/ref-fr-STIF/master/audit_routepoints/audits/2019_02/images/%C3%A9cart_positions_routepoints.csv).

Conclusion

La quantité de données disponibles dans OpenStreetMap sur la thématique des bus d'Île-de-France continue de progresser lentement vers la complétude totale.

Les constats précédemment établis lors de notre étude initiale de comparaison entre les données d'Île-de-France Mobilités et celles d'OpenStreetMap se confirment dans le temps avec l'augmentation de la couverture en données d'OpenStreetMap.

Malgré leur non-exhaustivité, les données crowdsourcées sont de très bonne qualité et leur utilisation permettrait d'améliorer l'information voyageur sur un certain nombre de sujets, dont la position des arrêts.

Crédits

Cet audit a été réalisé par Jungle Bus (<https://junglebus.io>), grâce au soutien de Cityway (<https://www.cityway.fr/>), dans le cadre du projet m2i (<http://www.mob2i.fr>).

Contactez-nous à l'adresse contact-arobase-junglebus.io (<http://contact-arobase-junglebus.io>) ou via notre compte Twitter BusJungle (<https://twitter.com/BusJungle>).



(<https://junglebus.io>)

Les données utilisées pour cette étude sont :

- les données OpenStreetMap, sous licence ODbL (© les contributeurs d'OpenStreetMap), datées du 23 février 2019
- les données GTFS mises à disposition en OdbL par Île-de-France Mobilités sur le portail opendata.stif.info (<http://opendata.stif.info>), datées du 21 février 2019

Le code source utilisé pour préparer ces données et calculer les différents indicateurs présentés est consultable sur l'organisation github de Jungle Bus (<https://github.com/Jungle-Bus/ref-fr-STIF>).

Les résultats de cet audit (texte et graphiques) sont disponibles ici sous licence CC-BY-ND (<https://creativecommons.org/licenses/by-nd/4.0>).