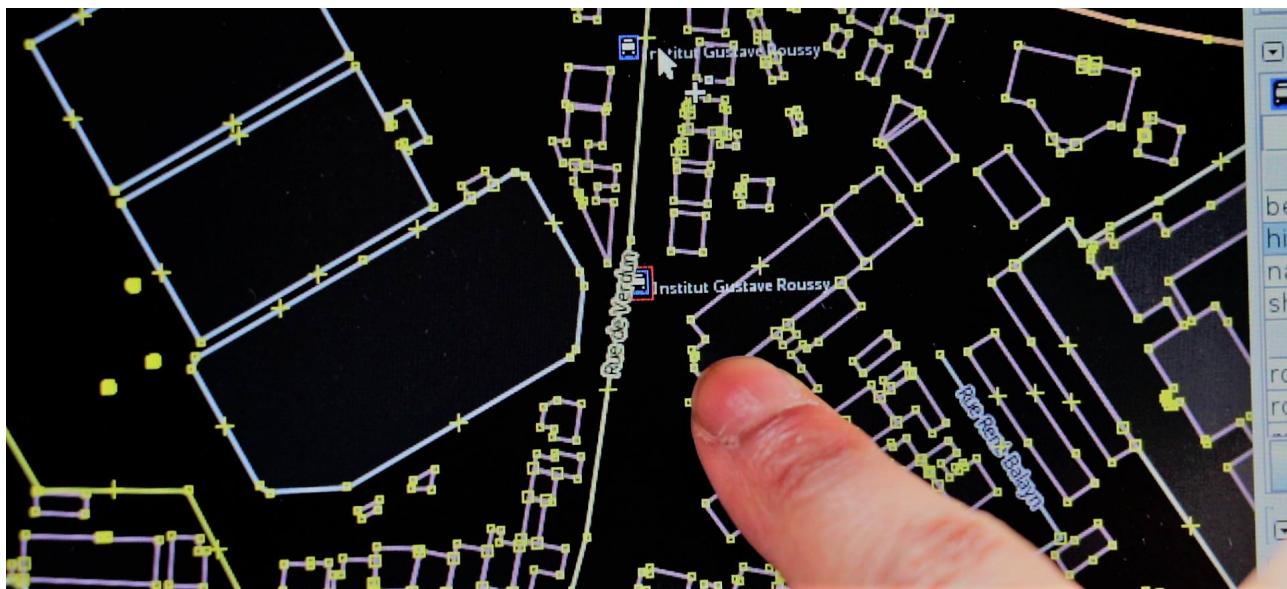


# Audit comparatif des données Open Data des arrêts de bus de la région Île-de-France - mise à jour de novembre

Publié le 26 novembre 2018 par l'association Jungle Bus (<https://junglebus.io>)



Ce document est une analyse qui approfondit et complète les précédents audits publiés en avril 2018 ([http://junglebus.io/iledefrance/audit\\_2018\\_04/](http://junglebus.io/iledefrance/audit_2018_04/)) puis en juillet 2018 ([http://junglebus.io/iledefrance/audit\\_2018\\_07/](http://junglebus.io/iledefrance/audit_2018_07/)).

## 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 six mois. À noter qu'un deuxième document avait déjà été produit en juillet afin de mesurer les évolutions à trois mois.

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 très restrictive permettant de garantir une comparaison des mêmes objets entre les deux référentiels, on constate qu'OpenStreetMap possède des données d'excellente qualité. Elles permettent de proposer des enrichissements aux données officielles d'Île-de-France Mobilités sur un panel de sujets utiles à l'information voyageurs.

Vous pouvez consulter le précédent audit dans son intégralité ([http://junglebus.io/iledefrance/audit\\_2018\\_04/](http://junglebus.io/iledefrance/audit_2018_04/)), ainsi que sa première mise à jour ([http://junglebus.io/iledefrance/audit\\_2018\\_07/](http://junglebus.io/iledefrance/audit_2018_07/)), pour obtenir les indicateurs et conclusions plus détaillés.

Voyons à présent de quelle manière les chiffres ont évolué au cours des six derniers mois.

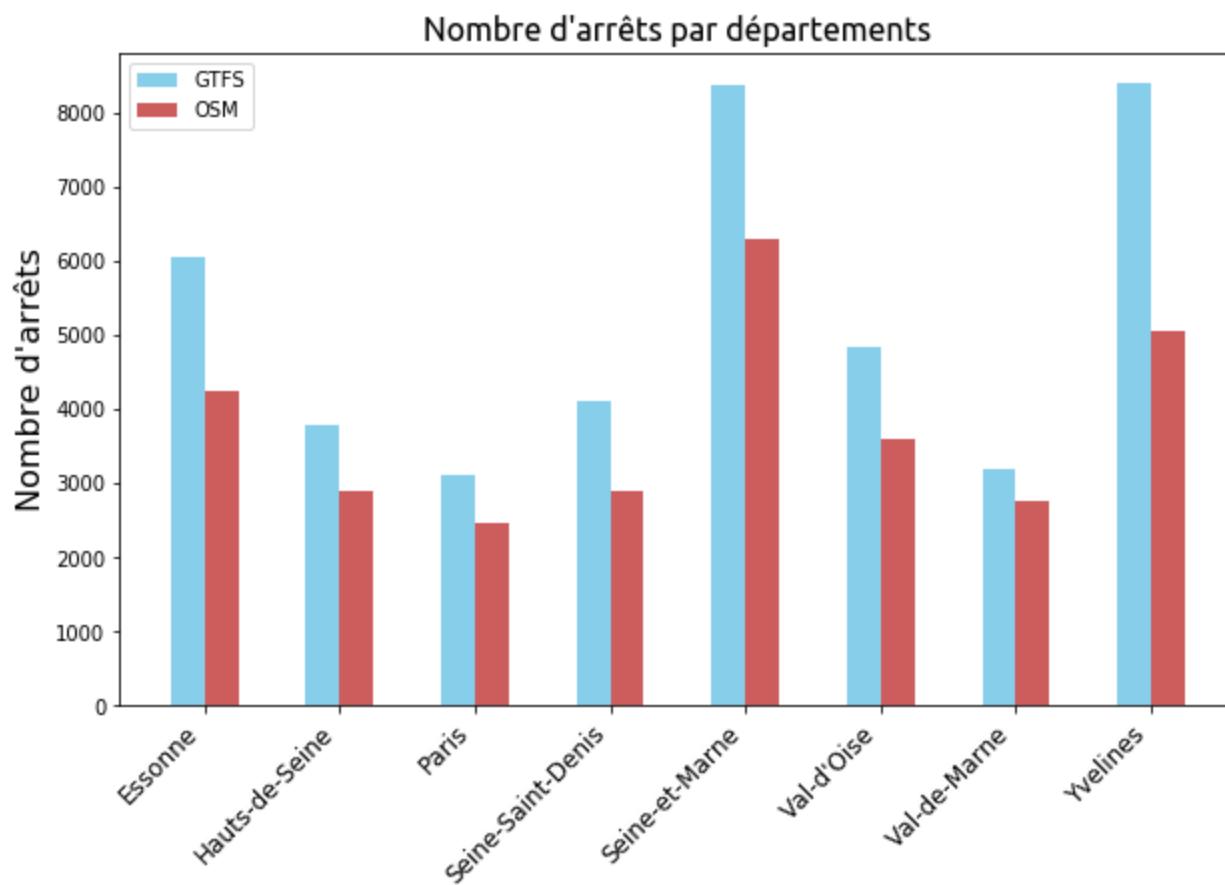
## Analyses quantitatives

---

### Nombre d'arrêts

---

Voici la répartition du nombre relatif d'arrêts synthétisée par département.



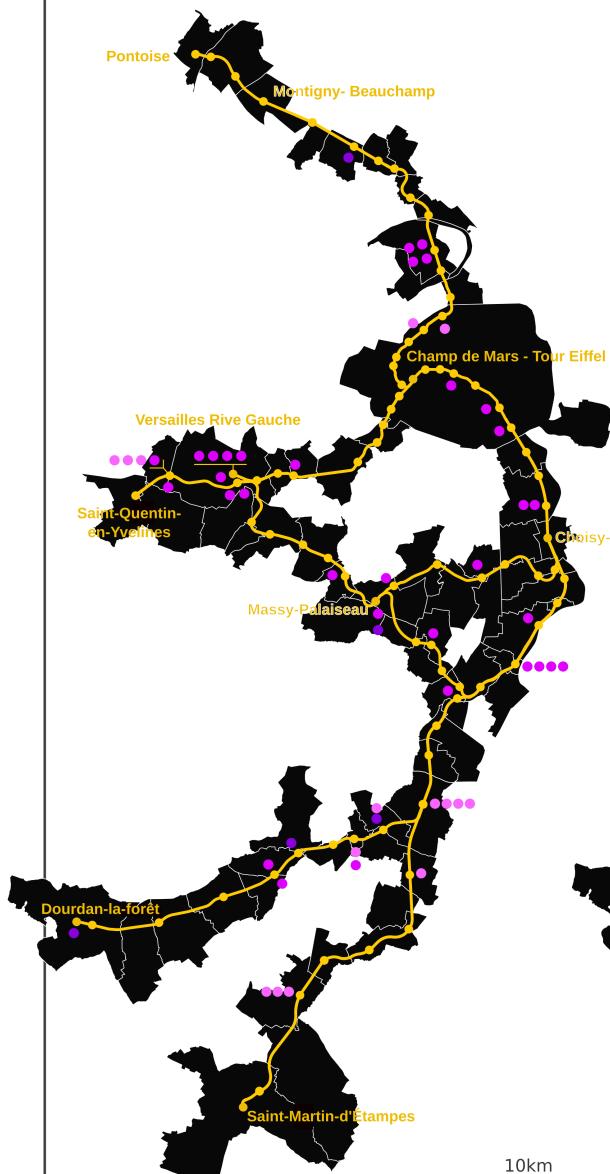
Le pourcentage global de couverture en arrêts sur la région Île-de-France est de 71,5 %.

Ce chiffre a progressé puisque le pourcentage de couverture en arrêts n'était que de 61,8 % en avril dernier. On constate que les écarts se sont resserrés dans les trois départements où il manquait le plus de données : l'Essonne, la Seine-et-Marne et les Yvelines.

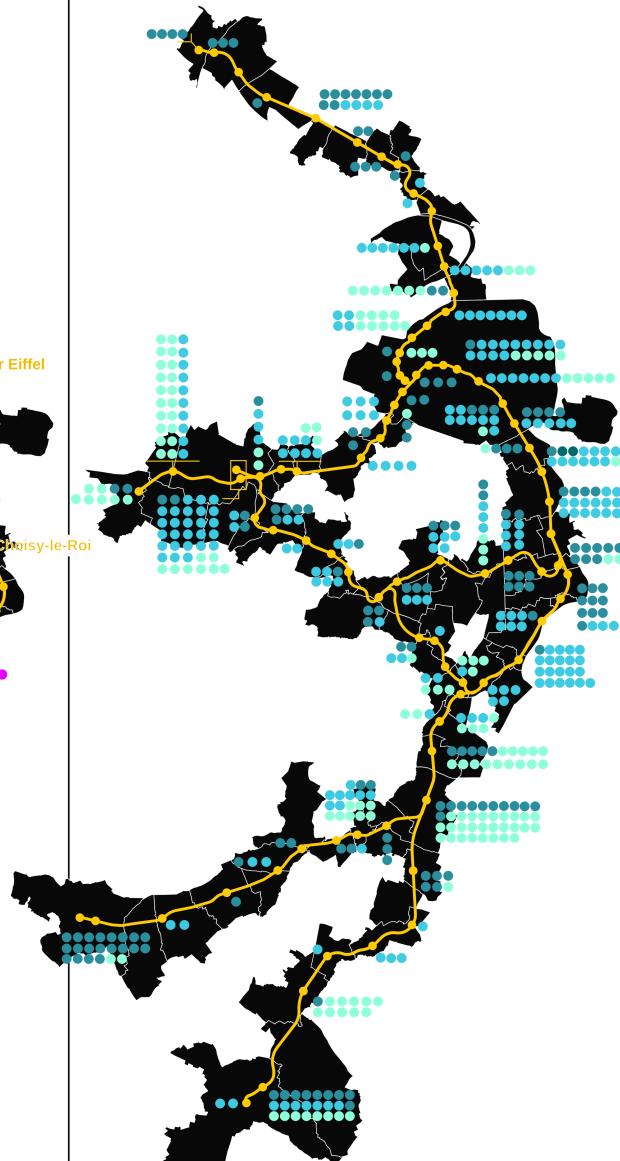
Parmi les initiatives menées par la communauté, on peut citer par exemple la campagne de contribution qui s'est concentrée sur les arrêts de bus le long du RER C entre septembre et novembre. Au cours de cette campagne, 591 arrêts de bus ont été soit créés soit mis à jour. L'Essonne est le département qui a le plus bénéficié de cette campagne de contribution avec 84 éditions en septembre, 81 en octobre et 74 en novembre. Les Yvelines, Paris et le Val-de-Marne ont également bénéficié de ces éditions avec en moyenne 45 arrêts édités par mois.

## Accroissement du renseignement des arrêts de bus le long du RER C entre septembre et novembre 2018

De nouveaux arrêts de bus sur l'ensemble du réseau



Mise à jour progressive et sectorielle :  
Saint-Cyr et Brétigny puis,  
Versailles et le tronçon Choisy-le-Roi - Champs de Mars  
et enfin Dourdan, Saint-Martin d'Étampes et Pontoise



### Éléments composant le RER C

- terminus de tronçon
- arrêt
- tracé

### Nouveaux arrêts de bus

- novembre
- octobre
- septembre

### Arrêts de bus mis à jour

- novembre
- octobre
- septembre

## Nombre de lignes

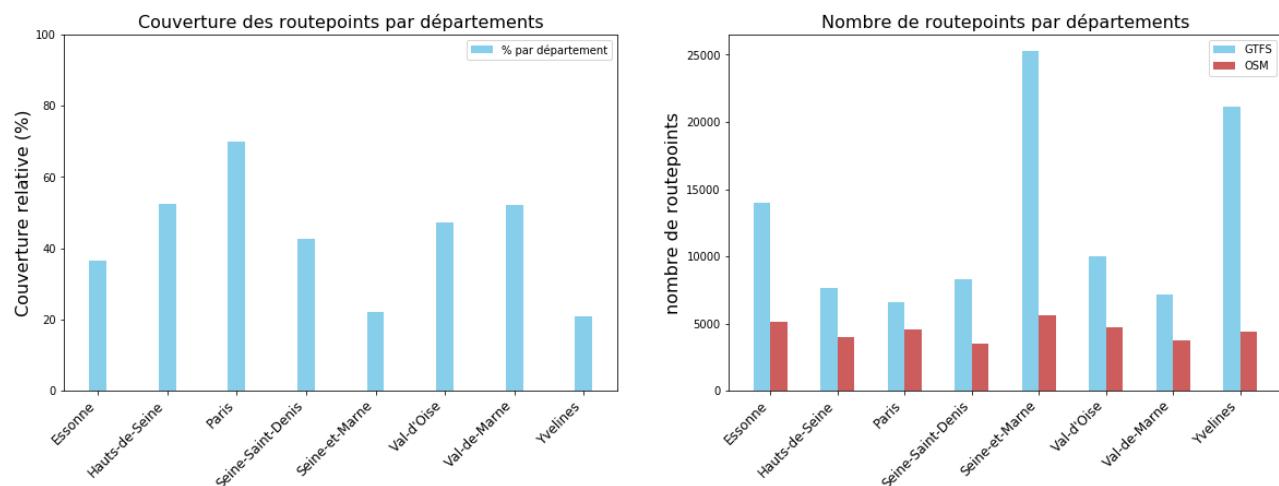
Le pourcentage global de couverture en lignes de bus sur la région Île-de-France est de 59 % (contre un peu moins de 50 % en avril dernier).

## Nombre de routepoints par département

Pour mitiger les divergences de modélisation entre les deux sources, depuis avril notre étude s'articule largement autour du routepoint, 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 routepoints par département.

Le graphique de gauche montre la couverture en routepoints de chaque département de la Région. Le second graphique présente le nombre de routepoints 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 routepoints sur la région Île-de-France est de 35,5 % contre 25 % en avril dernier.

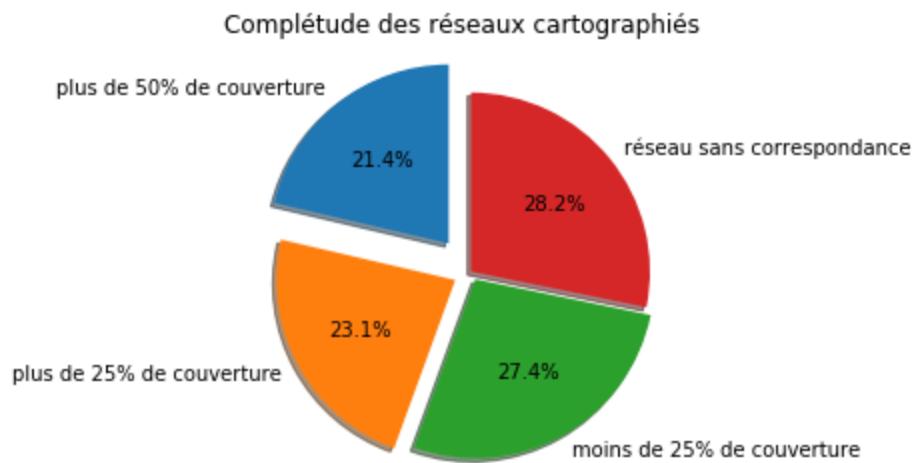
Là encore, les efforts menés par la communauté en Essonne et dans les Yvelines sont remarqués puisqu'il s'agit des deux départements où l'augmentation est la plus significative.

## Nombre de routepoints par réseau

En avril dernier, l'analyse révélait que moins de 10% des réseaux tels que définis par Île-de-France Mobilités détenaient plus de 50 % de couverture en routepoints dans OpenStreetMap.

La plupart des réseaux OPTILE n'étaient d'ailleurs pas représentés dans OpenStreetMap.

La situation a évolué très positivement en six mois puisque la part des réseaux sans correspondance a drastiquement diminuée tandis que les réseaux couverts à plus de 50% a au contraire augmentée :



## Nombre de tracés de lignes

---

En novembre, on constate que 41% de lignes disposant d'un tracé d'après OpenStreetMap

(cela correspond à 787 lignes, soit 70% des lignes présentes 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.

À noter que les indicateurs présentés ici ne garantissent pas que le tracé est continu ou exploitable - nous vous invitons à consulter l'audit de juillet où les détails méthodologiques sont précisés pour en savoir plus.

# 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 17% des routepoints sont ainsi retenus pour la suite de cette étude (contre 6% en avril dernier).

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/2018\\_11/images/mapping\\_des\\_routepoints.csv](https://raw.githubusercontent.com/Jungle-Bus/ref-fr-STIF/master/audit_routepoints/audits/2018_11/images/mapping_des_routepoints.csv)).

## Qualité des numéros de lignes

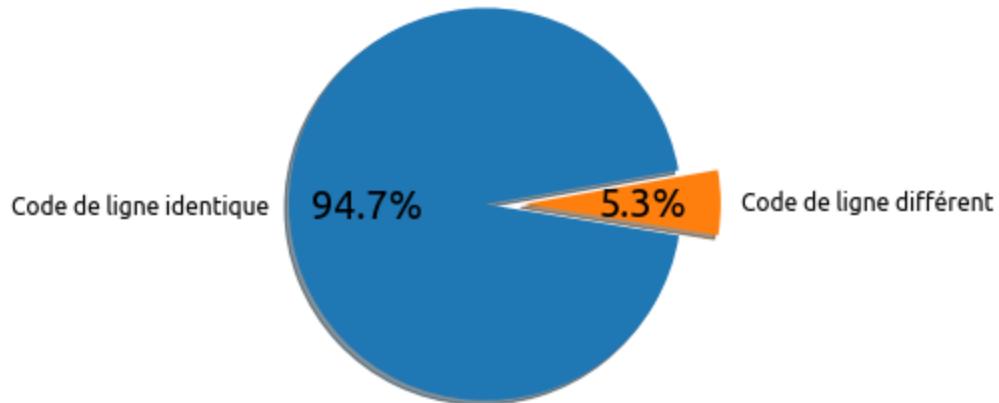
---

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

Ces constats restent d'actualité en novembre :

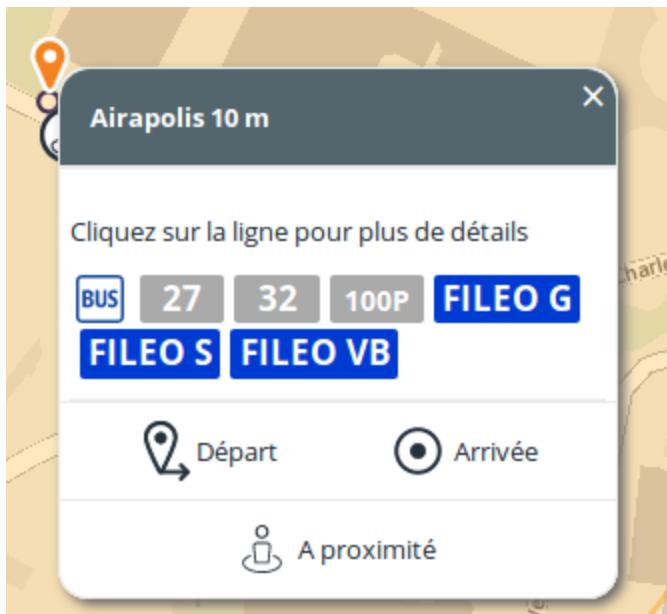
## Analyse de concordance des codes de lignes sur les routepoints



Voici par exemple le détail des codes de lignes du réseau Filéo :

OpenStreetMap	Île-de-France Mobilités
Goussainville	FILEO G
Othis	FILEO O
Roissy Sud	FILEO V
Roissy Sud	FILEO VE
Roissy Sud	FILEO T
Roissy Sud	FILEO SL
Saint-Pathus	FILEO SP
Sarcelles	FILEO S
Survilliers	FILEO FO
Villiers-le-Bel	FILEO VB

Cela se matérialise par exemple de la sorte dans l'information voyageur :



capture d'écran des lignes passant à un arrêt, sur Vianavigo

Pourtant, les intitulés "FILEO G" ou encore "FILEO VB" ne correspondent à rien d'observable par un voyageur en cherchant son arrêt.



signalétique affichée en gare de Roissypôle

Les dénominations proposées par OpenStreetMap sont cette fois encore plus proches de la réalité du terrain et plus susceptibles d'être utiles au voyageur.

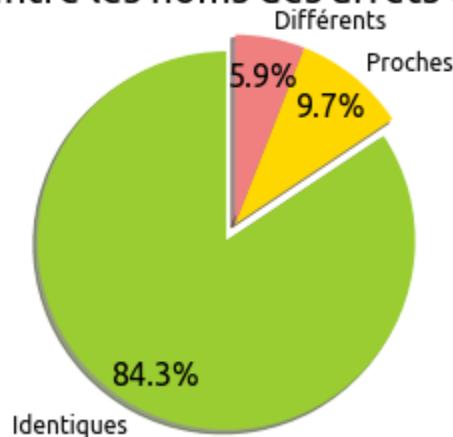
## Qualité des noms

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.

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 crowdsourcés.

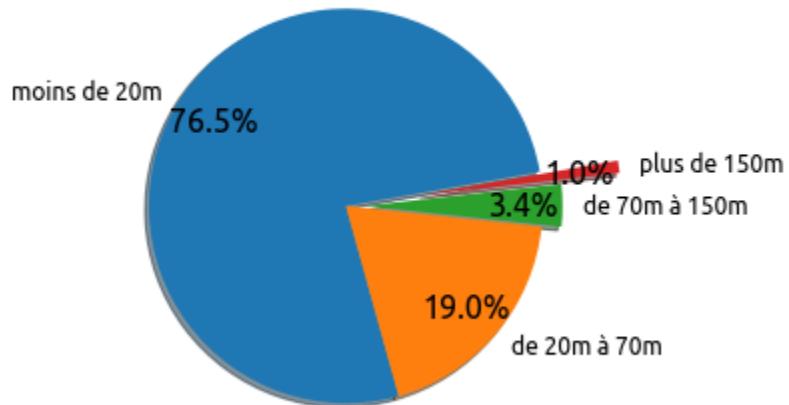
Cependant, quelques centaines d'arrêts étaient situés dans le GTFS à plus de 70 mètres de leur version crowdsourcée, avec certains à plus de 10 minutes de marche à pied.



écart de positions entre les deux sources pour les arrêts de la gare routière de Torcy

Ce constat se vérifie toujours dans les données du mois de novembre, avec **782 routepoints** (correspondants à **562 arrêts**) situés à plus de 70m.

### Distance entre les arrêts officiels et OSM



Voici un tableau récapitulatif du nombre de routepoints utilisés dans les différents audits successifs :

Audit	% de routepoints retenus	Augmentation par rapport à avril
Avril	6%	-
Juillet	14,9%	+148%
Novembre	17%	+183%

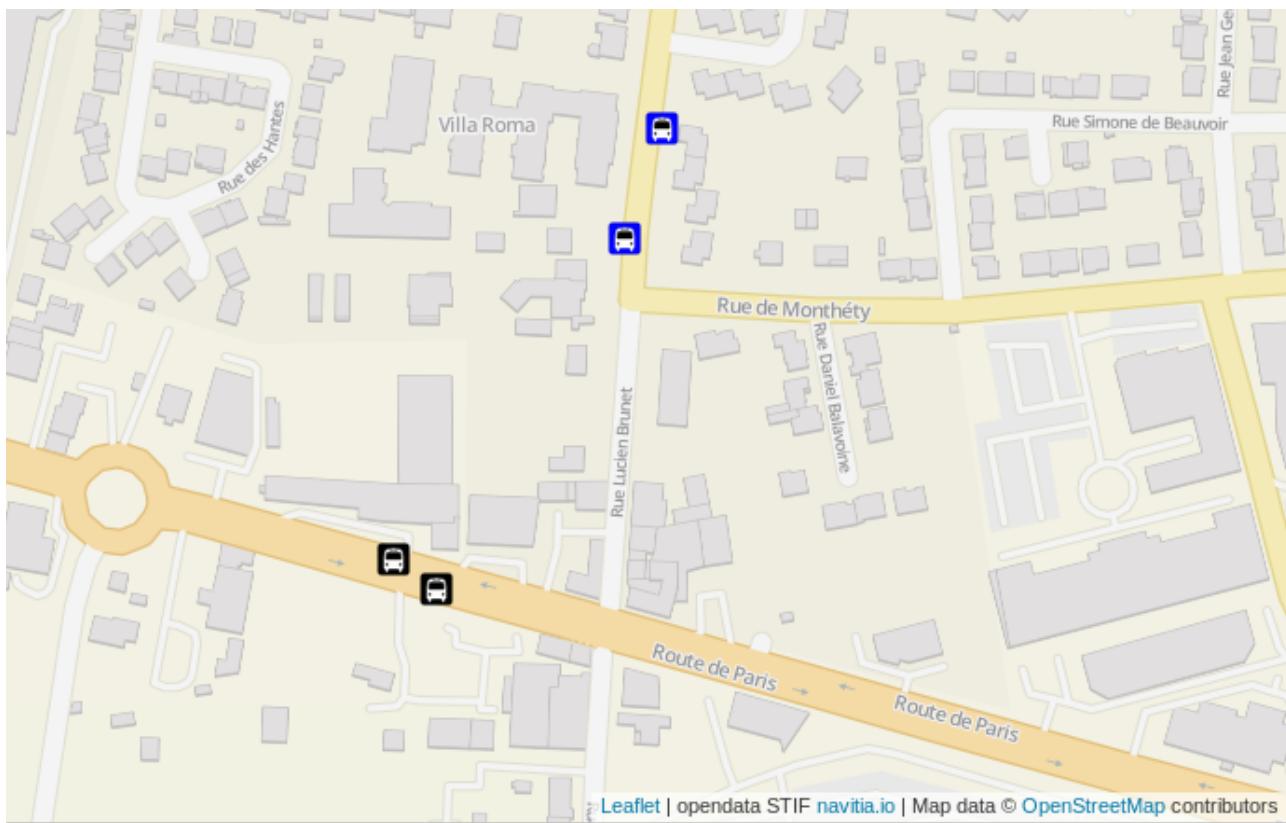
Et voici un tableau récapitulatif du nombre d'arrêts mal géolocalisés dans les audits successifs :

Audit	nombre d'arrêts à plus de 70m	Augmentation par rapport à avril
Avril	174	-
Juillet	450	+158%
Novembre	562	+223%

On constate donc que le nombre de routepoints et d'arrêts mal géolocalisés a augmenté beaucoup plus que le nombre de routepoints retenus pour l'audit !

Voici un exemple de problème de géolocalisation pouvant être très perturbant pour un usager : les arrêts ne sont ici pas situés sur le trajet effectif de la ligne.

Par ailleurs, la position des arrêts telle que proposée dans OpenStreetMap est confirmée par le plan de ligne du transporteur (<https://www.transdev-idf.com/ligne-A/ozoir-la-ferriere-pontault-combault/003-SIT>).



Arrêts "Pavé" de la ligne A, en noir d'après l'open data 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/2018\\_11/images/%C3%A9cart\\_positions\\_routepoints.csv](https://raw.githubusercontent.com/Jungle-Bus/ref-fr-STIF/master/audit_routepoints/audits/2018_11/images/%C3%A9cart_positions_routepoints.csv)).

## Conclusion

L'enrichissement des données de transport en Île-de-France par la communauté OpenStreetMap se poursuit. Comme lors du précédent audit, cette amélioration est notable et permet non seulement d'augmenter le nombre d'arrêts rapprochés entre les données crowdsourcées et les données officielles, mais également de confirmer les différentes mises en qualité évoquées dans les deux précédents audits (correction de la position des arrêts, amélioration des libellés des arrêts ou utilisation des tracés des lignes).

Le dynamisme de la communauté se confirme également dans la durée. La

participation à l'amélioration des outils de la communauté ou la mobilisation de celle-ci sur des campagnes de contribution locales restent des leviers très forts pour disposer de données encore plus complètes qu'actuellement.

## Crédits

Cet audit a été réalisé par l'association 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](mailto: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 16 novembre 2018
- 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 16 novembre 2018

Le code source utilisé pour préparer ces données et calculer les différents indicateurs présentés est consultable sur l'organisation github de l'association (<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>).