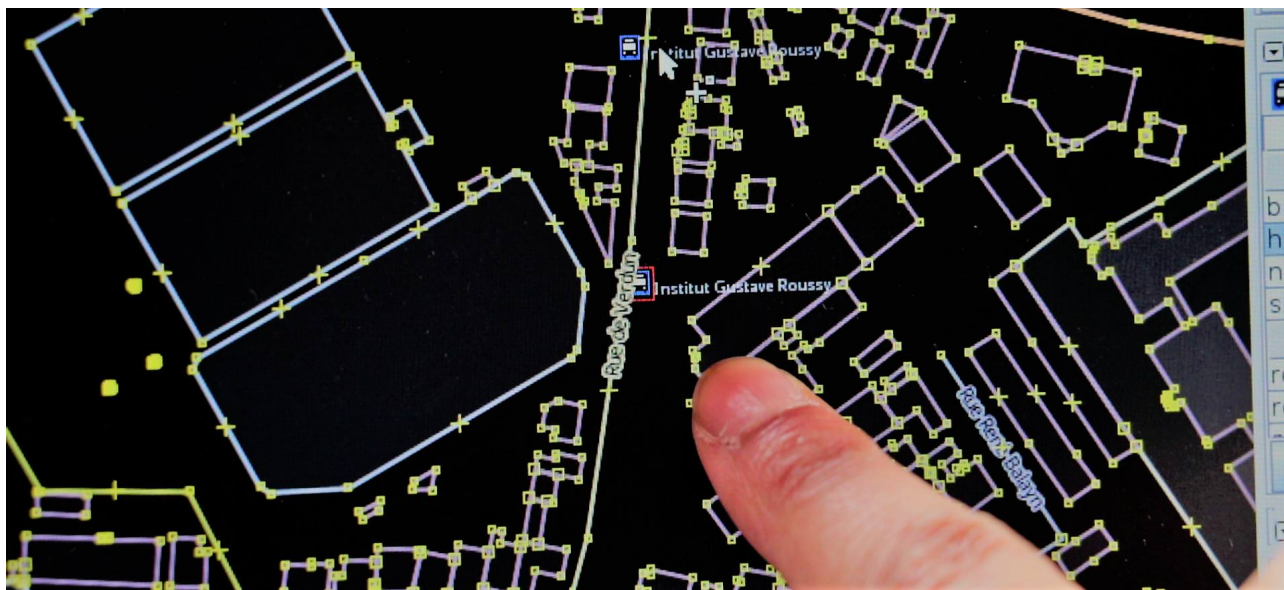


Audit comparatif des données Open Data des arrêts de bus de la région Île-de-France - mise à jour de juillet

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



Ce document est une analyse qui approfondit et complète le précédent audit publié le 26 avril 2018 (http://junglebus.io/iledefrance/audit_2018_04/).

Introduction

Île-de-France Mobilités publie en Open Data les données des lignes et les 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 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/) pour obtenir les indicateurs et conclusions plus détaillés.

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

Analyses quantitatives

L'étude d'avril dernier (http://junglebus.io/iledefrance/audit_2018_04/) présentait des estimations de la complétude des données OpenStreetMap en les comparant aux données officielles relatives à l'offre de transport publiées par Île-de-France Mobilités.

Les données officielles publiées en juillet présentent d'importantes disparités par rapport à celles d'avril : on y retrouve environ 350 lignes de moins, ainsi que de nombreux changements.

Certains de ces changements correspondent à l'évolution naturelle de l'organisation des transports : restructuration de réseaux (par exemple *Apolo 7*) ou augmentation des lignes de substitution pour travaux pendant l'été.

En revanche, il faut également noter que les lignes scolaires, qui ne circulent pas en ce moment, ont complètement disparu des données d'offre officielles (GTFS). Cela correspond d'après le Guide Transports Scolaires (https://www.iledefrance-mobilites.fr/wp-content/uploads/2017/12/Depliant_TS_SEPT-2017_MEL.pdf) à environ 750 circuits spéciaux.

Ces données restent présentes dans OpenStreetMap, puisque seule la composante géographique des données transport y est représentée.

En conséquence, nous avons choisi de ne pas présenter les mises à jour des indicateurs quantitatifs réalisés au précédent audit (nombre d'arrêts, nombres de lignes, nombre de routepoints) : les chiffres seraient largement surestimés du fait de l'exclusion des données scolaires d'une des sources.

Nous pourrions calculer à nouveau ces indicateurs lors de notre prochain audit, en octobre 2018.

Nombre de tracés de lignes

Bien que des grilles horaires présentant les trajets sur une carte soient disponibles pour la quasi-totalité des lignes, et qu'Île-de-France Mobilités réalise des plans détaillés présentant les lignes de bus d'une zone, il n'existe aujourd'hui pas de source officielle en Open Data qui serait réutilisable par d'autres acteurs pour disposer du tracé géographique des lignes.

Il pourrait être envisagé de les mettre à disposition dans un format cartographique standard via le portail Open Data existant, ou de les ajouter directement, sous forme de tracé de course au sein du GTFS existant.

Par ailleurs il est possible, lorsqu'on cartographie une ligne de transport, d'en indiquer le tracé dans OpenStreetMap. Nous pouvons donc évaluer le nombre de lignes qui disposent d'un tracé.

À noter qu'il est possible pour les contributeurs de n'indiquer qu'une portion du trajet (par exemple la portion de la lignes qu'ils ont emprunté au moment de leur relevé sur le terrain).

Afin de ne pas comptabiliser à tort des lignes faiblement cartographiées, nous avons choisi de limiter notre étude aux lignes mesurant plus de 3 kilomètres (en moyenne entre ses différents parcours).

Cela écarte de fait les plus petites lignes du réseau (telles que par exemple le bus 34 des Courriers d'Île-de-France (https://www.keolis-cif.com/fileadmin/Sites/cif/documents/Lignes/plans/Plan_ligne_34.pdf) ou encore le bus Monastère (<https://www.ratp.fr/sites/default/files/plans-lignes>

/Reseaux/Bus-Noctilien/Bus/572.pdf)) mais offre l'avantage déterminant de limiter les faux positifs.

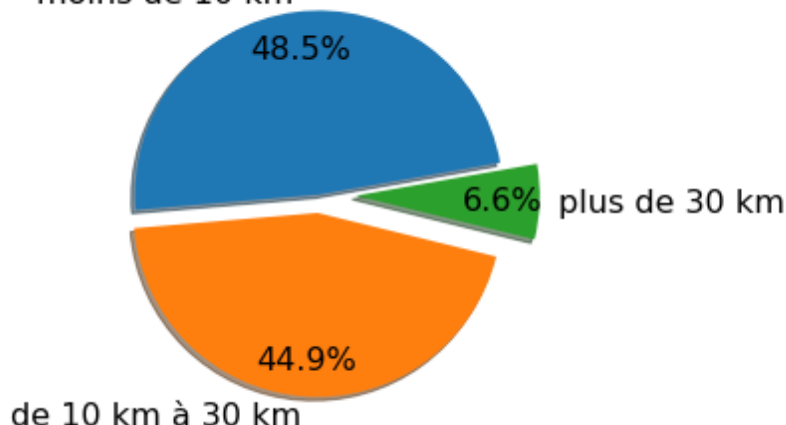
Avec ce mode de calcul, nous comptons **71 %** de lignes présentes dans OpenStreetMap qui disposent d'un tracé.

Nous ne rapporterons pas ce pourcentage au nombre de lignes officielles d'Île-de-France, en raison de l'indisponibilité des lignes scolaires à cette date dans les données officielles, mais cela représente exactement 756 lignes.

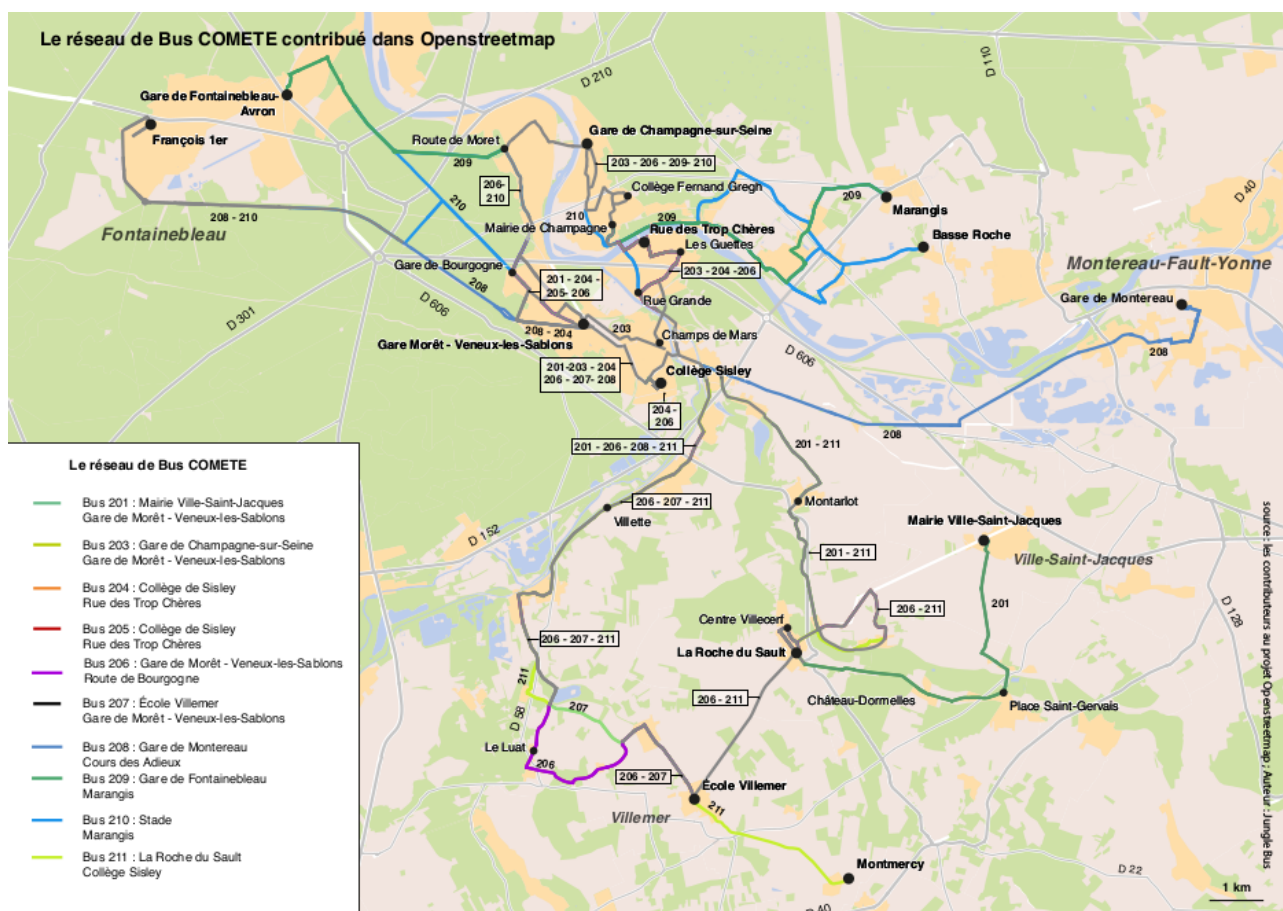
Dans l'optique d'envisager un enrichissement des données officielles, une étude plus approfondie devrait être réalisée afin d'évaluer la continuité des trajets et la complétude des lignes.

En l'absence de source officielle, il est difficile d'estimer la qualité de ce résultat. Une analyse par longueur moyenne de ligne permet cependant d'identifier des tendances encourageantes.

Longueur moyenne des lignes de bus d'après OSM



On retrouve en effet, parmi les lignes de moins de 10 km, les lignes des petits réseaux urbains (*Orgebus*, *Probus*, *Chavilbus*, etc). Et à l'inverse, parmi les lignes de plus de 30 km, on reconnaît les lignes suivies par les cars de type express (*Express A14*, *Seine-et-Marne Express*) ou les réseaux plus longue distance de la région (*Réseau Interurbain de Rambouillet*, *Le Bus Direct*).



Voici une carte du réseau Comète réalisée grâce aux tracés présents dans OpenStreetMap.

Du fait de la non-disponibilité de données officielles publiées par Île-de-France Mobilités, il est aujourd'hui impossible de réaliser une carte de ce type ou un plan des lignes qui desservent un lieu touristique ou une ville autrement qu'en se basant sur des données issues d'OpenStreetMap.

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 : Pour mitiger les divergences de modélisation entre les deux sources, notre étude portera sur des routepoints, c'est-à-dire un arrêt logique représentant un arrêt de bus desservi par une ligne dans une direction donnée. Par ailleurs, nous ne conserverons que les routepoints 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. 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_07/images/mapping_des_routepoints.csv).

Après la mise en correspondance des deux sources, **14,9 %** des routepoints officiels (soit un total de 15 137) sont retenus pour poursuivre l'analyse.

En comparaison, lors de l'audit d'avril 2018 (http://junglebus.io/iledefrance/audit_2018_04/), seuls 6 % des routepoints officiels (soit environ 7000) étaient retenus.

On notera que l'augmentation en pourcentage est surévaluée, du fait de la disparition dans les données officielles des nombreux routepoints correspondant à des circuits scolaires.

Néanmoins on constate tout de même que le nombre brut de routepoints a doublé, et il convient de saluer les efforts qui ont été menés par la communauté OpenStreetMap sur la mise en qualité des données existantes ainsi que sur l'intégration des référentiels d'Île-de-France Mobilités.

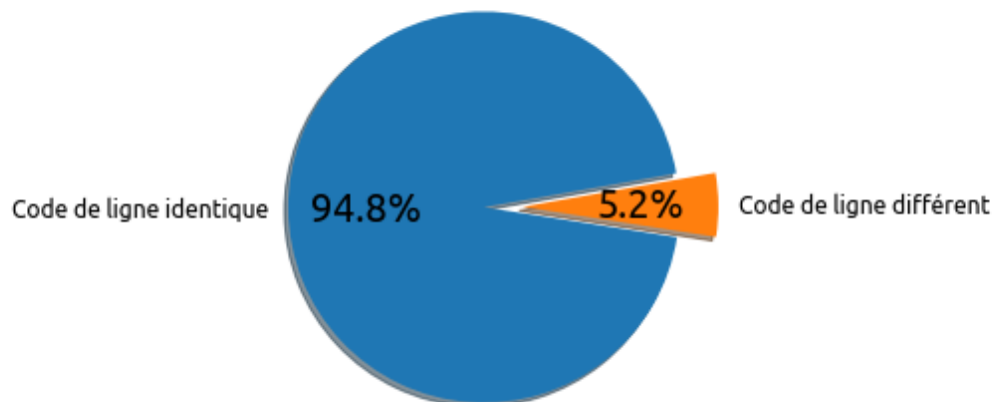
Qualité des numéro de lignes

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

Ces constats restent d'actualité en juillet.

Analyse de concordance des codes de lignes sur les routepoints



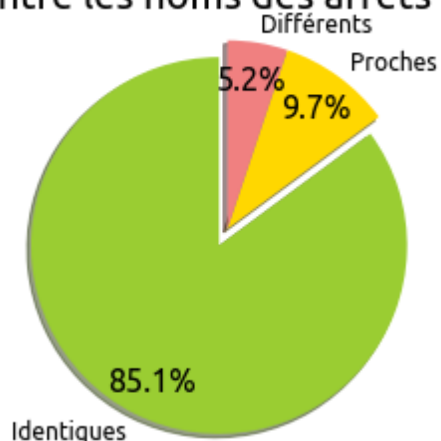
Qualité des noms

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.

Ces constats restent d'actualité.

Similitude entre les noms des arrêts (sans casse)



Qualité des positions

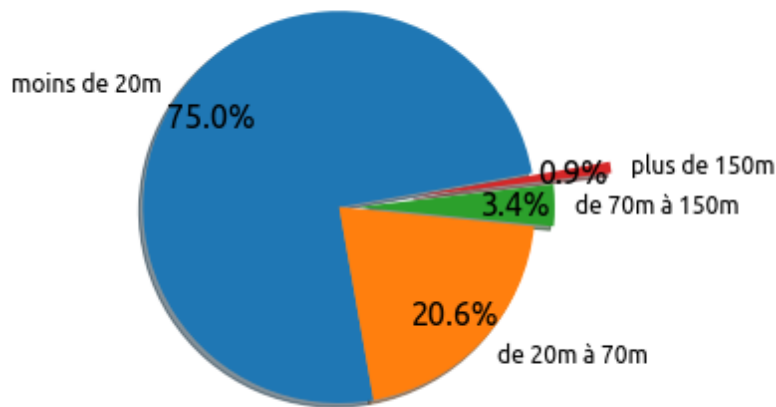
Lors de l'audit initial, nous avons constaté que l'essentiel des arrêts officiels se trouve

à une distance raisonnable de leurs homologues crowdsourcés.

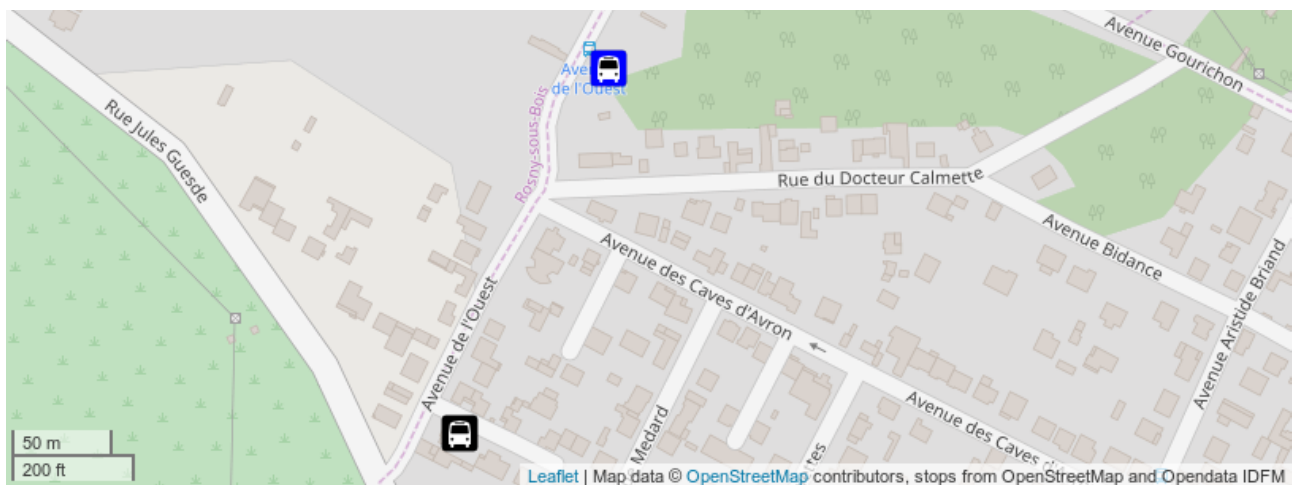
Cependant, plus de 170 arrêts étaient situés dans le GTFS à plus de 70 mètres de leur version crowdsourcée, avec certains à plus de 10 minutes de marche à pied.

Ce chiffre a malheureusement augmenté : nous avons à présent 450 arrêts (soit 634 routepoints) situés à plus de 70 mètres. Cette augmentation ne concerne pas un réseau de transport en particulier. En volume, les gros réseaux tels que ceux de la RATP et de Noctilien sont naturellement plus représentés.

Distance entre les arrêts officiels et OSM



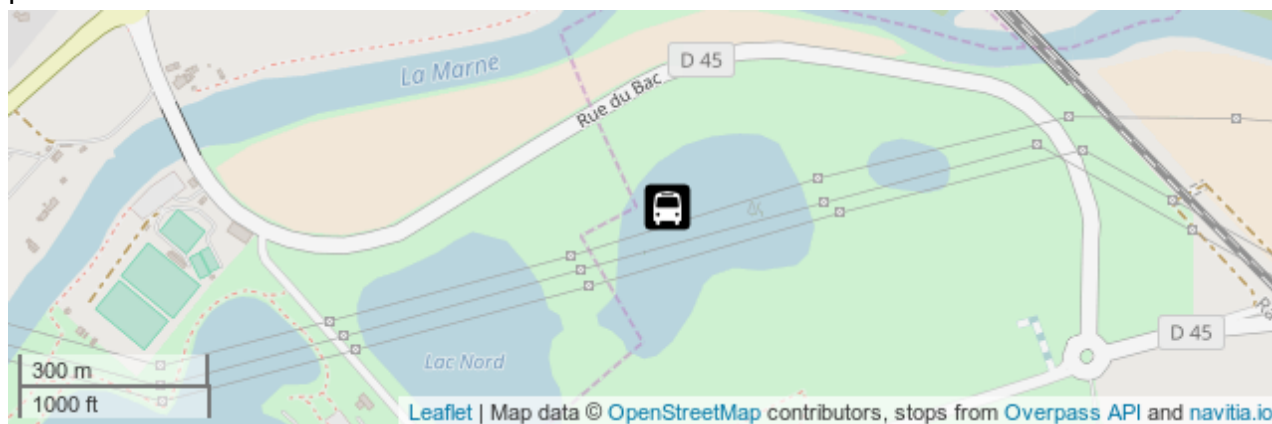
Le bon positionnement des arrêts est pourtant un prérequis pour produire de l'information voyageur de qualité. En effet le guidage efficace d'un voyageur vers un arrêt de bus est remis en cause s'il est dans les données positionné à tort dans la mauvaise rue ou à près de 200 mètres de sa position crowdsourcée.



Arrêt Avenue de l'Ouest, ligne Titus 3 vers la Gare de Rosny sous Bois. Le marqueur bleu indique la position d'après OpenStreetMap, le noir celle d'après le référentiel

officiel d'Île-de-France Mobilités

Certains arrêts, comme celui-ci, positionné dans un étang, sont manifestement mal positionnés.



Arrêt Base de Loisir, ligne Pep's 14

On constate donc à nouveau qu'OpenStreetMap permet, sur un ensemble d'arrêts contrôlé, d'apporter des positions plus fiables que celles du référentiel officiel.

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_07/images/%C3%A9cart_positions_routepoints.csv) (https://raw.githubusercontent.com/Jungle-Bus/ref-fr-STIF/master/audit_routepoints/audits/2018_07/images/%C3%A9cart_positions_routepoints.csv) (analysée réalisée mi-juillet).

Conclusion

Il nous semble que cette analyse mise à jour propose des résultats particulièrement encourageants. Même si la disparition des nombreuses lignes scolaires du référentiel officiel d'Île-de-France Mobilités ne nous a pas permis de mettre à jour les indicateurs quantitatifs mis en place lors de l'audit précédent, on constate une augmentation de la quantité de données OpenStreetMap disponibles pour la comparaison. Le nombre de routepoints strictement comparables entre les deux sources a en effet doublé. Cela dénote la vitalité de la communauté et son engagement concret sur les problématiques liées au transport.

De plus, les tendances proposées en avril dernier se renforcent ; tous les indicateurs qualitatifs confirment leurs conclusions. On constate même une augmentation du nombre d'arrêts dont la position d'après le référentiel officiel mériterait d'être vérifiée.

Enfin, les tracés des lignes de bus dans les données OpenStreetMap sont nombreux et proposent une réponse pertinente à l'absence de publication des tracés officiels en Open Data par Île-de-France Mobilités.

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@junglebus.io (<http://contact@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 juillet 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 12 juillet 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/>).