

Audit comparatif des données Open Data des arrêts de bus de la région Île-de-France

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



Contexte

Île-de-France Mobilités (anciennement STIF) s'est investi depuis plusieurs années dans une démarche de publication systématique des données concernant les lignes et les arrêts de transport, collectées et agrégées auprès des opérateurs de transport de la région Île-de-France.

Par ailleurs, 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.

En effet la communauté OpenStreetMap (OSM) sur laquelle repose Jungle Bus cartographie le Monde entier, rue après rue, arrêt de bus après arrêt de bus. Ces données crowdsourcées sont également librement accessibles et ouvertes à la réutilisation.

Le but de ce document est de produire une étude portant sur la comparaison entre ces deux sources de données.

Île-de-France Mobilités collecte les données auprès des opérateurs de transport de la région. Son référentiel de données peut donc prétendre à l'exhaustivité sur son périmètre. Néanmoins, d'autres réseaux coexistent sur ce territoire et sont exclus de fait de ce référentiel. On peut notamment citer certaines navettes municipales ou encore des lignes privées telles que les navettes privées IKEA ou encore des navettes de musée.

Dans les données OpenStreetMap, la collecte est assurée par une communauté de passionnés, principalement bénévoles et amateurs. La couverture est donc moins exhaustive. En revanche, une fois créée, elle est réputée plus détaillée et plus précise que les données disponibles officiellement en Open Data.

Aucune source de données n'est exempte d'erreurs qu'il s'agisse d'imprécisions ou d'anomalies. Pourtant, afin de guider un voyageur, la fiabilité et la cohérence des informations sont cruciales, aussi bien concernant les horaires prévus de passage des bus que sur les repères utiles pour la compréhension des voyageurs (code de ligne, couleur associée, etc.)

En effet, un voyageur ne comprend pas et pourrait même s'inquiéter si l'arrêt indiqué dans son application n'a pas le même nom que celui affiché, ou s'il se situe en réalité dans une rue voisine !

Les objectifs de cette étude sont multiples :

- évaluer la complétude des données OpenStreetMap
- mesurer les différences entre ces deux sources sur un échantillon comparable
- évaluer le potentiel des données crowdsourcées d'OpenStreetMap pour enrichir les données d'Île-de-France Mobilités, en particulier sur la localisation des arrêts
- identifier les moyens d'actions pour améliorer la couverture OpenStreetMap dans le but de maximiser le potentiel d'enrichissement des données d'Île-de-France Mobilités.

Nous réaliserons une étude quantitative puis qualitative des données, puis analyserons les résultats et proposerons enfin notre plan d'action futur.

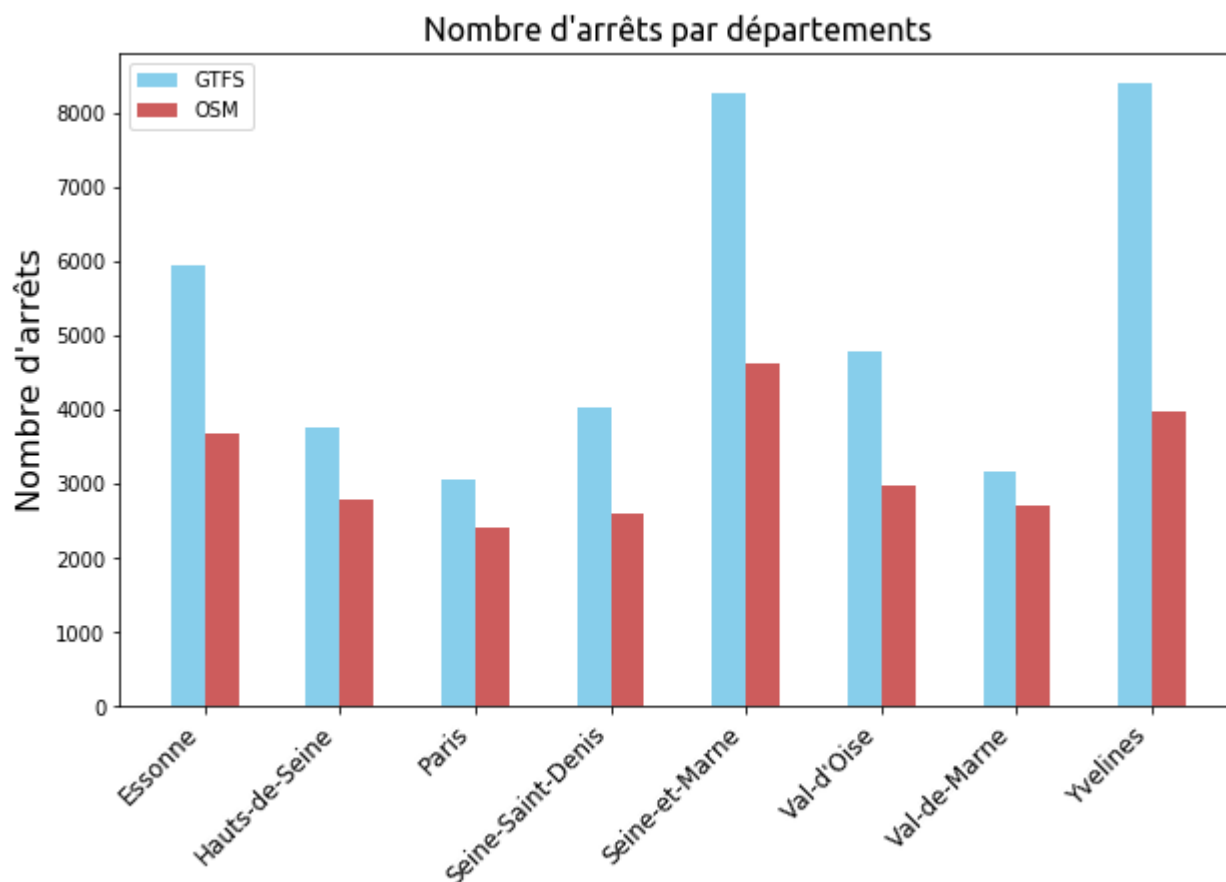
Analyses quantitatives

Nombre d'arrêts

Commençons par évaluer la quantité d'arrêts de bus présents dans OpenStreetMap. Afin de rendre cet indicateur plus parlant, nous effectuons une mesure relative au nombre d'arrêts répertoriés par Île-de-France Mobilités dans le GTFS qui est le jeu

de données décrivant l'offre de transport.

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



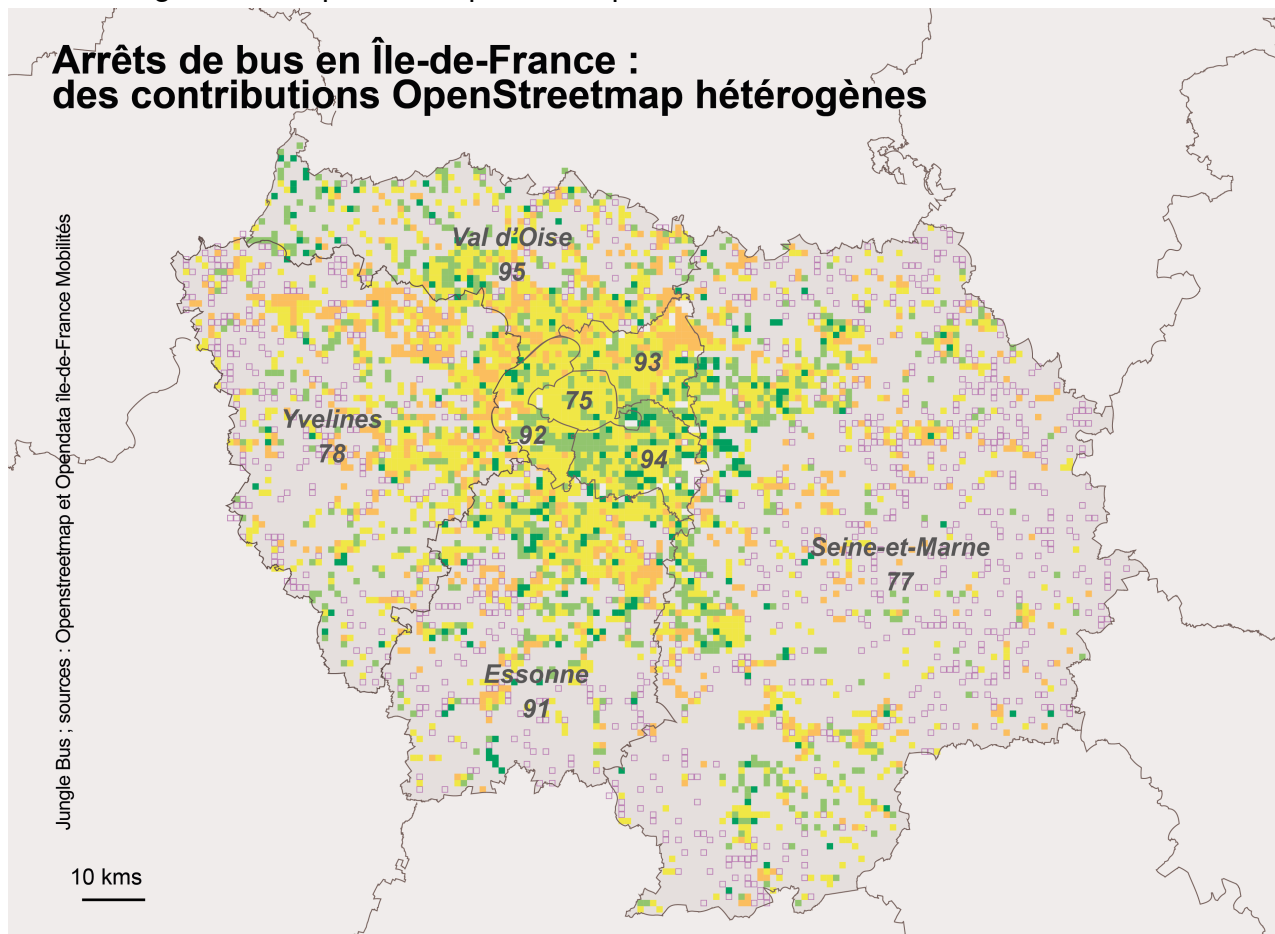
nom	nombre d'arrêts GTFS	nombre d'arrêts OSM	couverture OSM
Essonne	5934	3685	62,1%
Hauts-de-Seine	3751	2781	74,1%
Paris	3056	2413	79,0%
Seine-Saint-Denis	4028	2601	64,6%
Seine-et-Marne	8262	4609	55,8%
Val-d'Oise	4769	2981	62,5%
Val-de-Marne	3161	2716	85,9%
Yvelines	8385	3976	47,4%

NB : les objets qui représentent les arrêts sont typés par mode dans OpenStreetMap. Ce n'est pas le cas pour les arrêts dans le GTFS. Ainsi, les chiffres proposés ici sont légèrement sous évalués car les arrêts OpenStreetMap considérés sont

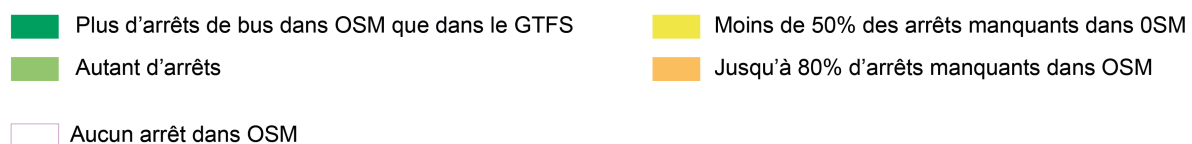
exclusivement des arrêts de bus, tandis que les arrêts GTFS comptent également des arrêts d'autres modes (tram, métro, RER, etc)

On constate une bonne couverture sur Paris et le Val-de-Marne, tandis que les Yvelines et la Seine-et-Marne semblent avoir moins de la moitié des arrêts présents.

Une analyse géographique par grille nous permet d'observer la répartition territoriale avec une granularité plus fine que les départements.



Rapport entre le nombre d'arrêts GTFS et le nombre d'arrêts de bus dans OSM
par cellule de 1000 mètres de côté



L'hétérogénéité spatiale de la contribution à OSM est mise en évidence par cette carte. Paris et surtout les communes limitrophes Est et Sud-Est présentent un niveau de complétude élevé. On note ensuite des zones irrégulièrement renseignées. Certaines mailles qui présentent une complétude totale correspondent probablement à des espaces globalement très riches en données. Ce sont les zones éloignées du centre de la région qui sont, de manière générale, moins complètes dans OpenStreetMap.

On souligne notamment la majorité du département Seine-et-Marne qui ne présente

aucun arrêt dès qu'on s'éloigne du tracé des lignes de transport en commun ferré. Les Yvelines et l'Est de l'Essonne sont également des territoires plus pauvres en données. Des campagnes de contribution pourraient efficacement re-stimuler la contribution localement.

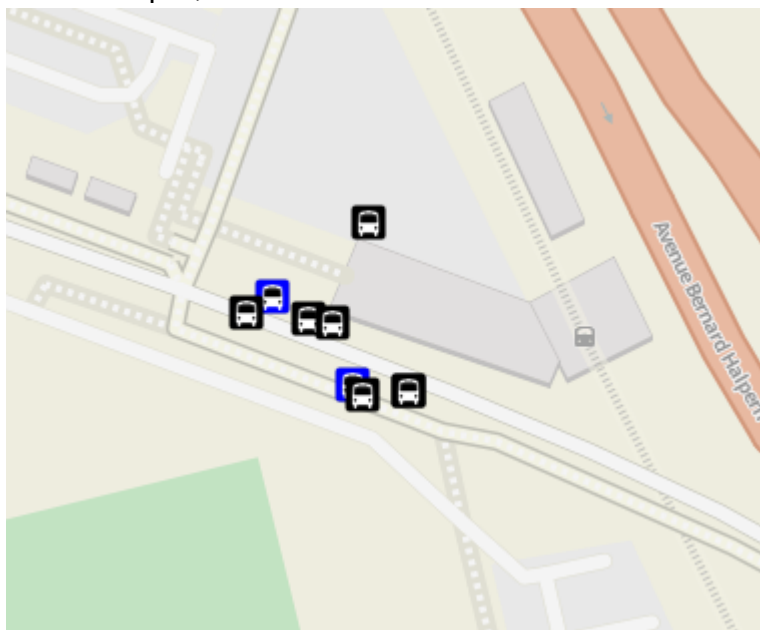
Le pourcentage global de couverture en arrêts sur la région Île-de-France est de **61,8 %**.

Nombre de routepoints

Le nombre d'arrêts est un indicateur intéressant pour mesurer la capacité de crowdsourcing de la communauté OpenStreetMap, mais n'est pas très fin, du fait des différences de modélisation des données.

En effet, un arrêt de bus dans OpenStreetMap représentera un lieu physique où un voyageur attend un véhicule, tandis que les arrêts du GTFS sont souvent dupliqués par opérateur.

Par exemple, voici la situation à la station de Pointe du Lac, dans le Val-de-Marne.



Sur le terrain et dans OpenStreetMap, il existe deux arrêts : un de chaque côté de la rue (pictogrammes bleus sur la carte).

Dans les données officielles du GTFS, il y a 6 arrêts : deux pour le réseau "RATP", deux pour le réseau "STRAV" et deux pour le réseau "Plateau de Brie" (pictogrammes noirs sur la carte).

Enfin, dans le référentiel REFLEX (qui est un autre jeu de données proposé en Open Data par Île-de-France Mobilités et qui agrège les arrêts par proximité), il y a 3 codes d'arrêts.

Dans le cas présent, le comptage naïf du nombre d'arrêts dans OpenStreetMap et dans le GTFS présenté plus tôt laisse supposer, à tort, que 4 arrêts seraient manquants dans OSM.

Afin de mitiger ces difficultés de comparaison, nous proposons donc de prendre en compte non seulement les informations de l'arrêt, mais aussi celui des lignes qui le desservent. Ainsi, on ne parlera plus de l'arrêt "Pointe du Lac", mais de l'arrêt logique "Pointe du Lac de la ligne 393 vers la gare de Sucy-Bonneuil". Cet arrêt logique sera dénommé "routepoint" par la suite dans ce document.

Afin de produire cet indicateur, il est nécessaire, en complément des arrêts, d'analyser les lignes de bus, ainsi que les terminus de chaque parcours de ligne.

Nombre de lignes

La communauté OpenStreetMap cartographie également les lignes : l'enchaînement des arrêts, mais aussi les différents trajets de chaque ligne de bus sont répertoriés. Ces tracés ne sont pas proposés par Île-de-France Mobilités sur sa plateforme Open Data. Il s'agit pourtant d'une information à forte valeur ajoutée pour l'information voyageur ; nous pourrions réaliser une étude sur ce sujet dans des audits ultérieurs.

Le pourcentage global de couverture en lignes de bus sur la région Île-de-France est de **49,7 %**.

Les lignes sont donc bien moins cartographiées par la communauté OpenStreetMap que les arrêts. Cela s'explique en partie par la relative complexité de cette tâche qui est de facto inaccessible aux contributeurs débutants.

Nombre de terminus de ligne

Les lignes d'OpenStreetMap comportent en moyenne 2,07 terminus tandis que les lignes du GTFS comptent en moyenne 3,17 terminus.

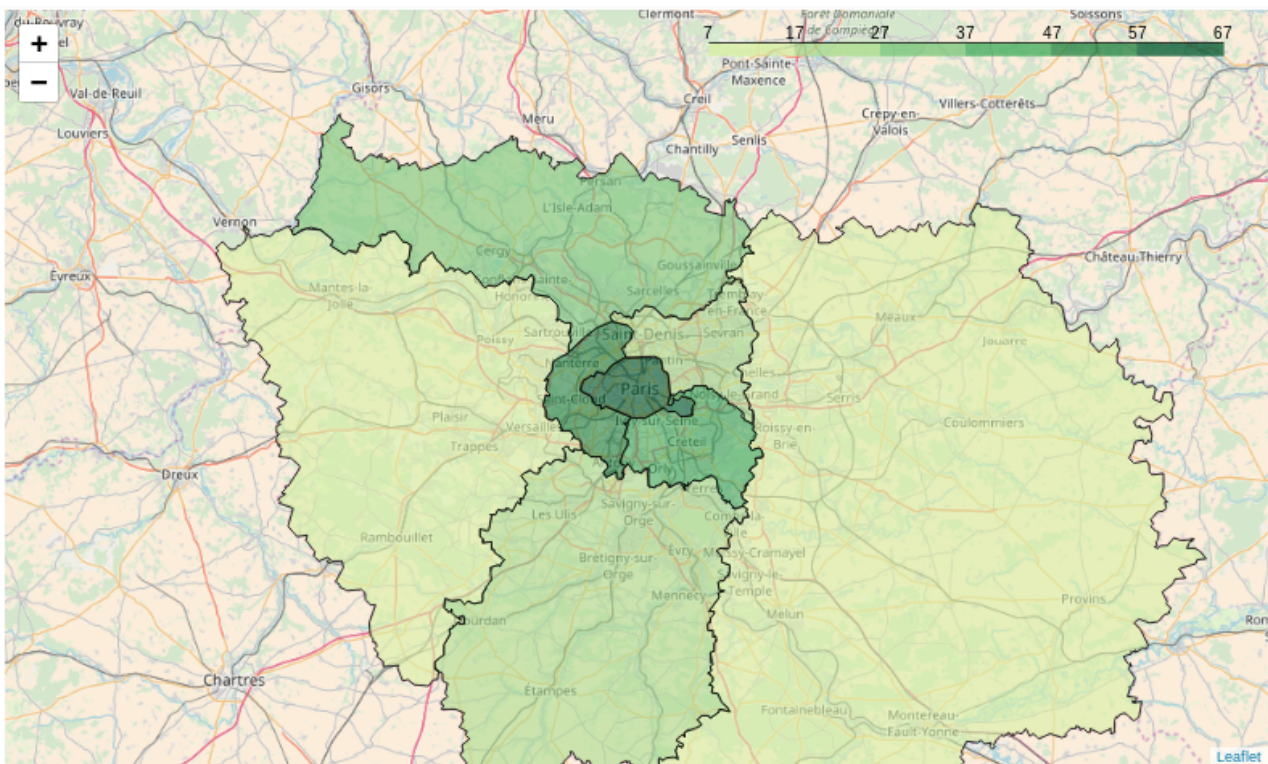
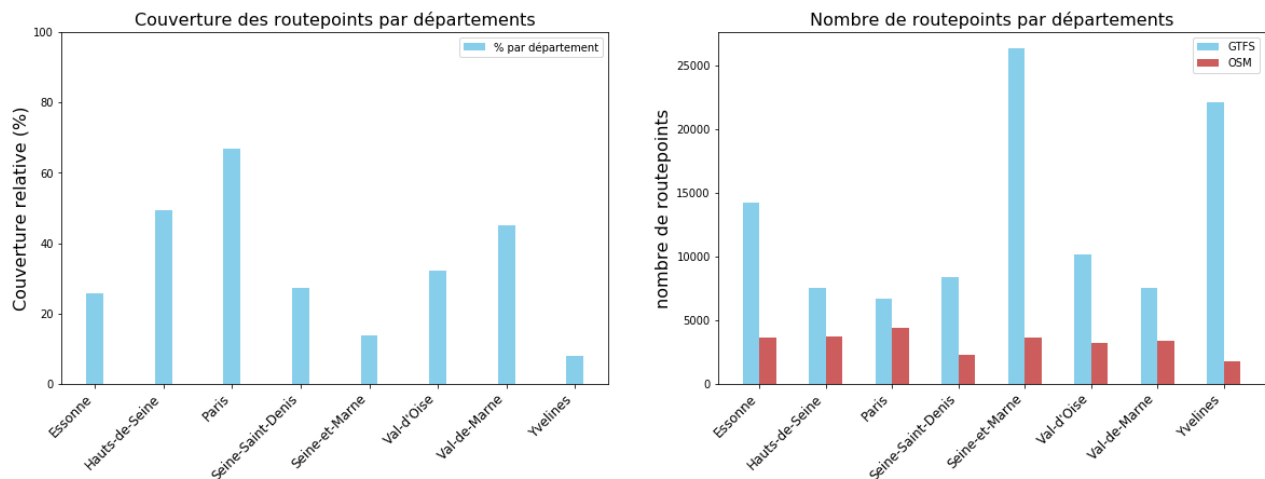
Les lignes cartographiées par la communauté OpenStreetMap ne couvrent donc pas tous les trajets alternatifs de chaque ligne.

Là encore, plus que la quantité de données crowdsourcée, c'est la différence de modélisation qui est avant tout déterminante. En effet, les terminus partiels de parcours qui n'existent qu'à certaines heures n'ont par exemple pas vocation à être modélisés dans OpenStreetMap, qui est une base purement géographique, alors qu'on les retrouve dans le GTFS.

Ainsi, la ligne de bus RATP 57 aura pour terminus uniquement "Gare de Laplace" et "Porte de Bagnolet" dans OpenStreetMap, alors que le GTFS comporte un terminus partiel supplémentaire à "Gare de Lyon", au milieu de la ligne.

Nombre de routepoints par département

Voici la répartition du nombre relatif de routepoints par département.



On constate à nouveau une bonne couverture sur Paris et le Val-de-Marne, tandis que les départements moins cartographiés des Yvelines et de la Seine-et-Marne, qui sont également ceux qui disposent du plus grand nombre de routepoints officiels, ressortent encore plus.

Le pourcentage global de couverture en routepoints sur la région Île-de-France est de **25,4 %**.

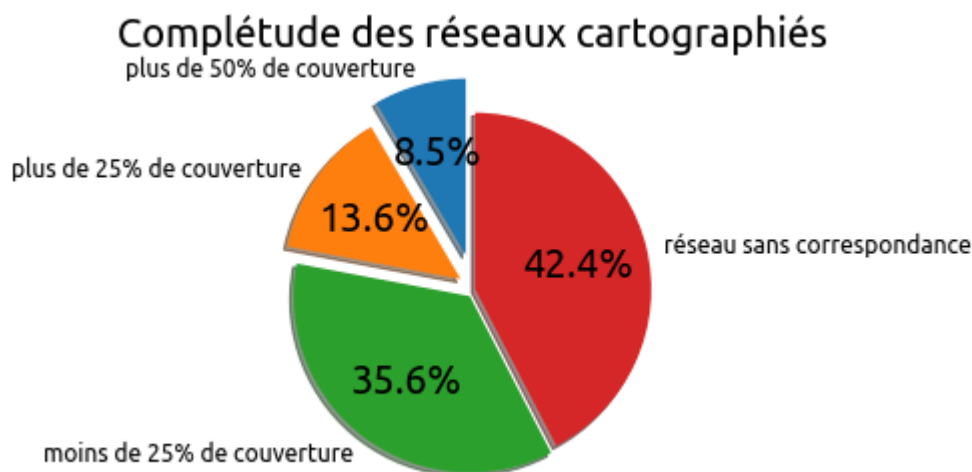
Nombre de routepoints par réseau

L'analyse par département permet de visualiser les écarts géographiques, mais ne

reflète pas très bien l'organisation des transports ; une analyse par réseau permet de voir la couverture d'OpenStreetMap avec une granularité plus fine.

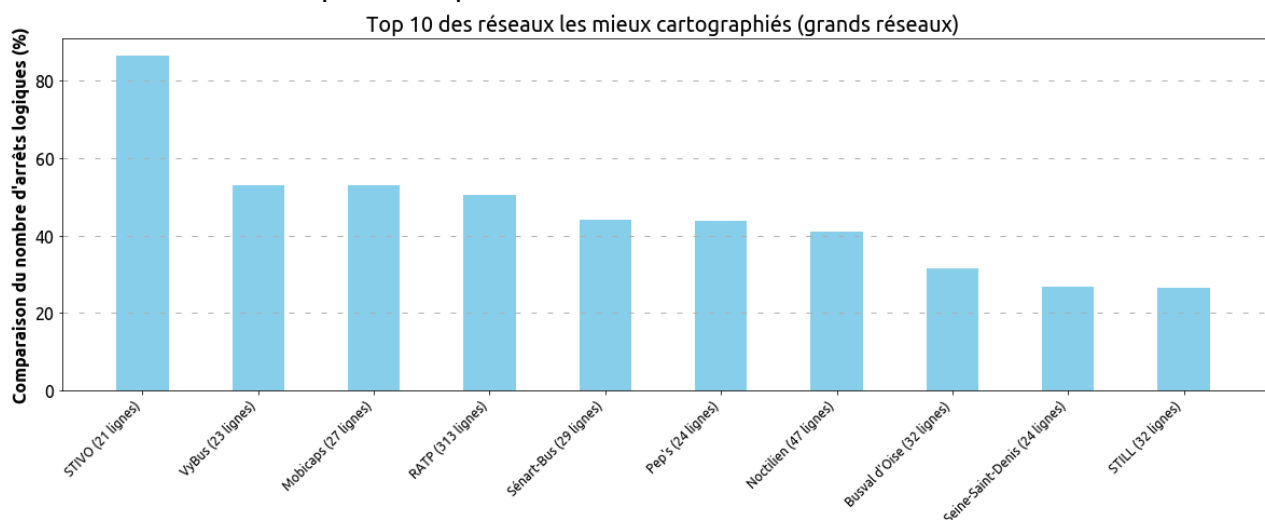
Nous avons pour cela tenté de faire correspondre chaque "agency" proposée par le GTFS avec les notions de réseau (network) et transporteur (operator) proposées dans OpenStreetMap afin de calculer la complétude des données crowdsourcées par réseau. Là encore, des différences de modélisation sont à signaler.

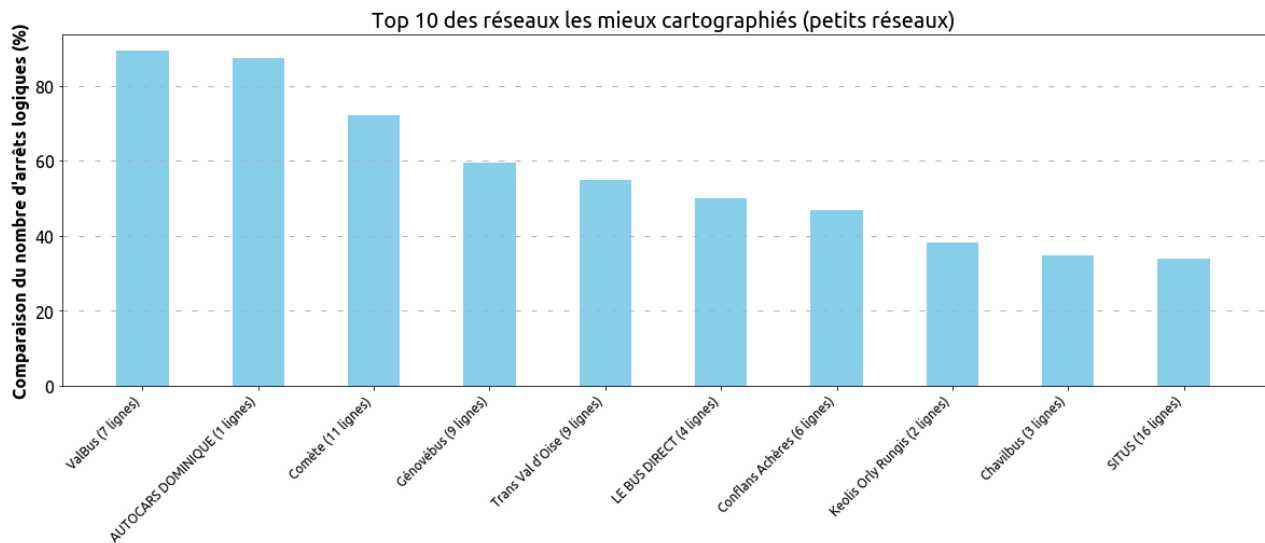
Plus de 40% des "agency" du GTFS ne semblent pas trouvables dans OpenStreetMap.



On constate ainsi que moins de 10% de réseaux peuvent prétendre à la complétude.

Voici les réseaux les plus complets :





Ces premiers indicateurs quantitatifs nous permettent de mesurer la quantité de données OpenStreetMap disponible à la comparaison avec les données officielles.

Analyses qualitatives

Voyons à présent s'il est possible d'utiliser ces données crowdsourcées pour apporter des améliorations aux données officielles, ou a minima, détecter des cas qui nécessitent une attention particulière.

Nous limiterons l'analyse aux données strictement comparables : nous allons donc faire correspondre autant que possible chaque routepoint d'OpenStreetMap avec un routepoint du référentiel officiel avant de mesurer les différences entre les deux sources.

Nous l'avons vu précédemment, un routepoint se décrit de telle sorte : "Arrêt Pointe du Lac, ligne 23 en direction de Créteil".

Il y a donc 3 composantes :

- arrêt
- ligne
- destination

Retrouver chaque élément dans les deux référentiels n'est pas une opération triviale : si une ligne est facilement identifiable par son code de ligne, il est plus difficile de reconnaître automatiquement un même nom d'arrêt ou de terminus dans des jeux de données.

Afin de limiter ces problèmes, nous utiliserons les codes officiels des lignes et des arrêts mis à disposition en Open Data par Île-de-France Mobilités ; la communauté a

en effet commencé à les intégrer dans OpenStreetMap afin de favoriser l'interopérabilité des données.

Les codes sur les arrêts (REFLEX, ZDEr_ID_REF_A) et les codes sur les lignes (CODIFLIGNE) nous permettent de faire correspondre les deux premières composantes.

Pour le nom de la destination, nous utiliserons une comparaison de chaîne de caractères.

Cette méthode est bien sûr loin d'être idéale.

Elle permet par exemple de faire correspondre :

- "Sucy-Bonneuil RER" avec "Gare de Sucy-Bonneuil"
- "A. Lebrun" avec "Auguste Lebrun"
- "Centre commercial" avec "Noisy-le-Roi - Centre Commercial le Cèdre"

mais ne va pas faire correspondre :

- "Massy" avec "Gare RER de Massy Palaiseau"
- "Saint-Maur Créteil" avec "SAINT-MAUR - CRETEIL RER"
- "Cimetière" avec "Cimetière de Saint-Maur-des-Fossés"

Après la mise en correspondance des deux sources, seuls 6% des routepoints officiels sont retenus pour poursuivre l'analyse.

Les critères utilisés pour faire la correspondance sont volontairement stricts afin de garantir que les données sont réellement comparables et portent sur le même objet logique, quitte à réduire la quantité d'objets considérés.

Pour réaliser une correspondance plus efficace entre les deux sources, une des deux solutions suivantes devrait être privilégiée :

- disposer d'un référentiel de parcours, maintenu et par Île-de-France Mobilités et intégré par la communauté dans OSM
- disposer d'un tableau de correspondance entre les 5639 destinations possibles de parcours officiels et leurs équivalents dans OSM

Par ailleurs, des actions devront être menées afin de terminer l'intégration des référentiels d'arrêts et de lignes d'OpenStreetMap.

Nous proposons de nous focaliser pour l'instant sur les indicateurs suivants :

- écart sur le numéro de la ligne
- écart sur le nom de l'arrêt
- écart de distance pour l'arrêt entre les deux sources

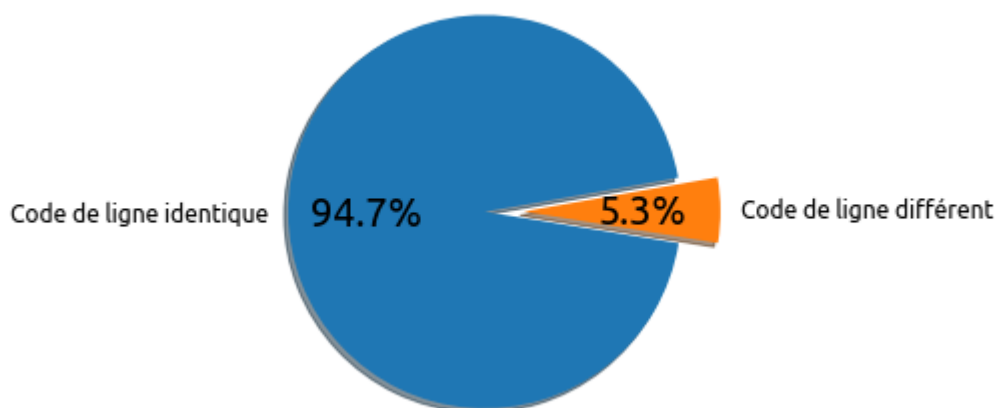
Nous pourrions ajouter d'autres mesures ultérieurement telle que l'accessibilité des arrêts aux fauteuils roulant ainsi que le détail des indicateurs sus-cités par zone géographique ou par réseau.

Qualité des numéro de lignes

Le numéro de ligne est bien souvent pour un voyageur le moyen d'identifier sa ligne, il est donc particulièrement important que celui proposé dans les différents supports d'information voyageur soit cohérent.

Nous mesurons ici les écarts de code de lignes entre nos deux sources.

Analyse de concordance des codes de lignes sur les routepoints



L'essentiel des codes de ligne est donc identique entre les deux sources.

Si on regarde plus en détail les cas où il y a des écarts, deux typologies de différences se dessinent.

Le premier cas correspond au cas où les données OpenStreetMap ont un code numérique et pas les données officielles. C'est le cas par exemple de

- la traverse Ney Flandre
 - code de ligne dans les données officielles : NEY-FLA
 - code de ligne dans les données OpenStreetMap : 519
- la traverse de Charonne
 - code de ligne dans les données officielles : CHARONN
 - code de ligne dans les données OpenStreetMap : 501

Il s'agit d'une ligne ayant un nom couramment utilisé pour la désigner en complément de son numéro.

Il existe un champ pour indiquer un nom complet de ligne aussi bien dans les données GTFS (route_long_name, fichier routes.txt) que dans les données

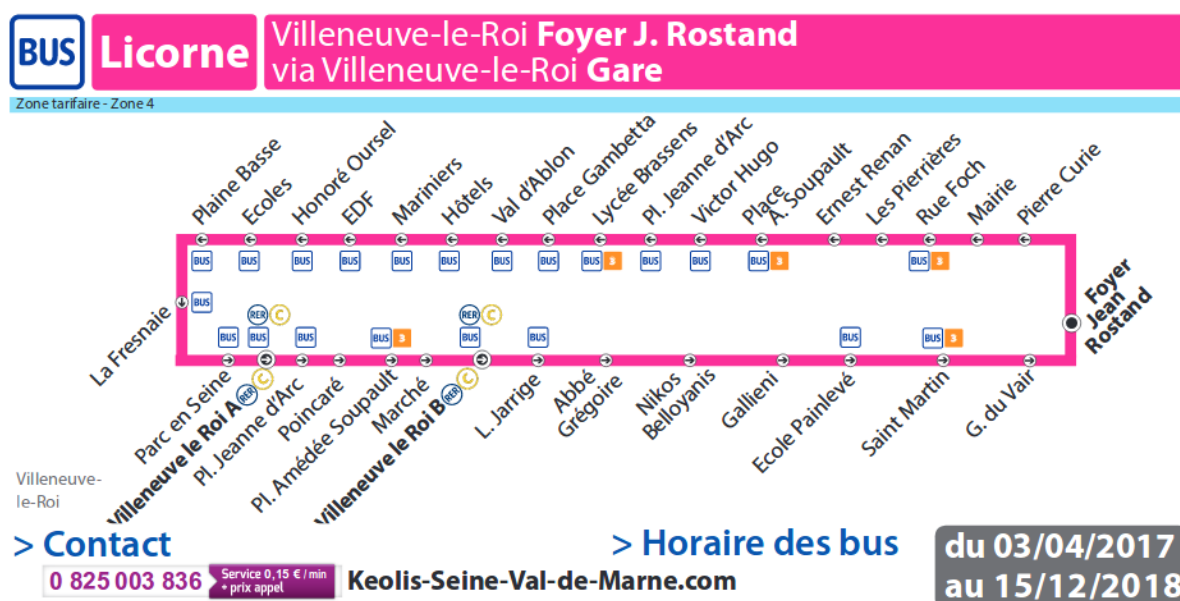
OpenStreetMap (tag name). C'est pour ces lignes plutôt celui-ci qui devrait être utilisé.

L'autre typologie d'écart concerne les codes tronqués.

C'est le cas par exemple de :

- la ligne Citalien
 - code de ligne dans les données officielles : CITA
 - code de ligne dans les données OpenStreetMap : Citalien
- la ligne Licorne
 - code de ligne dans les données officielles : LIC
 - code de ligne dans les données OpenStreetMap : Licorne

D'après le plan officiel ci-dessous, "Licorne" est le libellé effectivement proposé au voyageur :



OpenStreetMap propose une dénomination plus exacte de ce que constatera un voyageur sur le terrain.

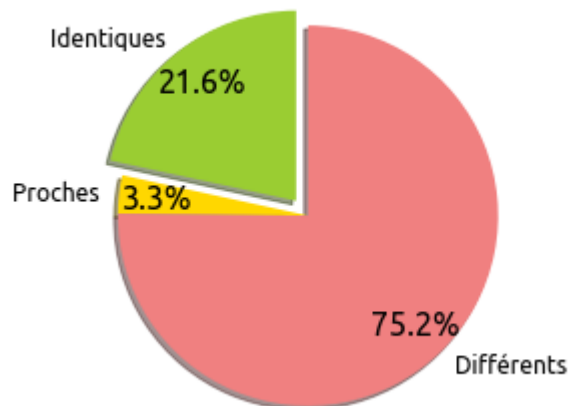
Qualité des noms

Le nom de l'arrêt est un élément de base de l'information voyageur : il permet de rassurer l'utilisateur et de lui indiquer où descendre lorsqu'il prend le bus dans un endroit qui ne lui est pas familier.

L'analyse suivante mesure les écarts entre les noms des arrêts de nos deux sources. La méthode de comparaison est la même que celle utilisée pour faire correspondre les destinations de lignes.

Voici la répartition observée :

Similitude entre les noms des arrêts



Les noms semblent donc assez peu concordants entre les deux sources de données !

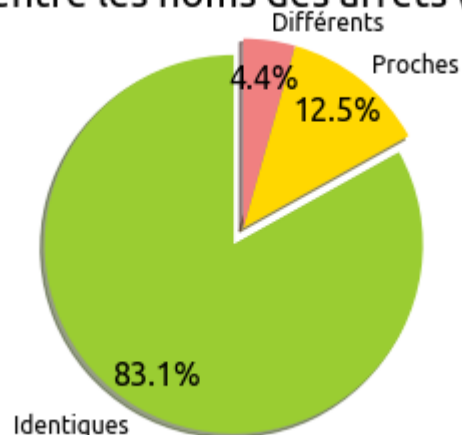
Une analyse plus détaillée permet de constater que de nombreux noms d'arrêts sont écrits intégralement en lettres capitales dans les données officielles.

Le fait d'avoir un nom capitalisé empêchera rarement à un voyageur de retrouver son arrêt sur le terrain, mais ce mode d'expression est peu recommandé dans le cadre d'information voyageur numérique : en effet, il est communément admis qu'écrire tout en majuscules est désagréable pour le lecteur, et traduit les cris ou la colère.

Le fait que certains arrêts soient capitalisés mais pas tous donne par ailleurs un manque de cohérence assez déstabilisant.

Refaisons maintenant la même analyse en uniformisant au préalable la casse des deux jeux de données :

Similitude entre les noms des arrêts (sans casse)



Si on ignore la casse, les noms sont donc globalement identiques dans les deux jeux de données.

Voici quelques typologies de différences constatées :

Les variations sémantiques : il s'agit de cas où les noms désignent sans ambiguïté le même lieu, mais en utilisant des termes proches, par exemple

- “OZOIR RER” contre “Gare d'Ozoir-la-Ferrière”
- “Gare de Torcy” contre “Torcy RER”

Les variations d'orthographe : les bonnes pratiques d'OpenStreetMap recommandent en effet d'éviter les abréviations et de respecter les règles d'orthographe usuelles, tandis qu'aucune règle ne semble en vigueur dans les données officielles, ce qui entraîne un certain manque de cohérence et d'uniformité des données officielles :

- “CC Bonneuil” contre “Centre Commercial Bonneuil”
- “A. Baron” contre “Antoine Baron”
- “EGLISE DE PANTIN-METRO” contre “Église de Pantin - Métro”

En conclusion, les données du référentiel officiel sur le nommage des arrêts sont très hétérogènes et manquent d'uniformisation en comparaison avec celles d'OpenStreetMap.

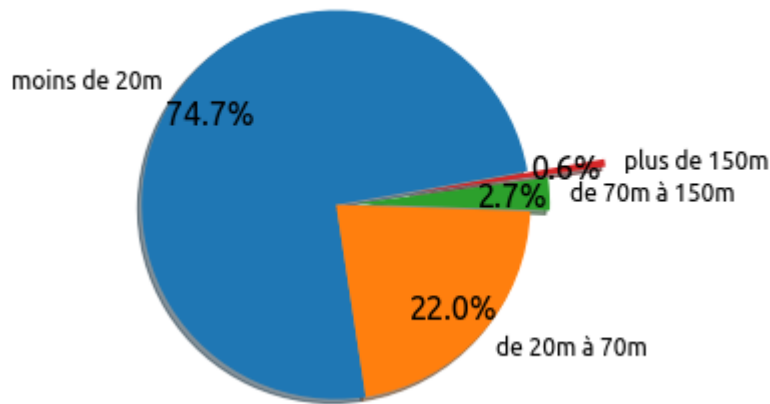
Qualité des positions

Afficher au voyageur des positions d'arrêts correctes est une nécessité pour offrir un service de qualité, et permettre de guider efficacement l'utilisateur vers les transports en commun. La précision de cette position est une condition nécessaire pour construire des services voyageurs modernes tels que le guidage piéton ou en fauteuil roulant lors des correspondances, au début ou à la fin de son voyage.

Mais c'est aussi une nécessité pour faciliter l'usage des transports en commun afin par exemple de proposer des temps de correspondances crédibles.

Voici la répartition des distances géographiques entre les deux sources :

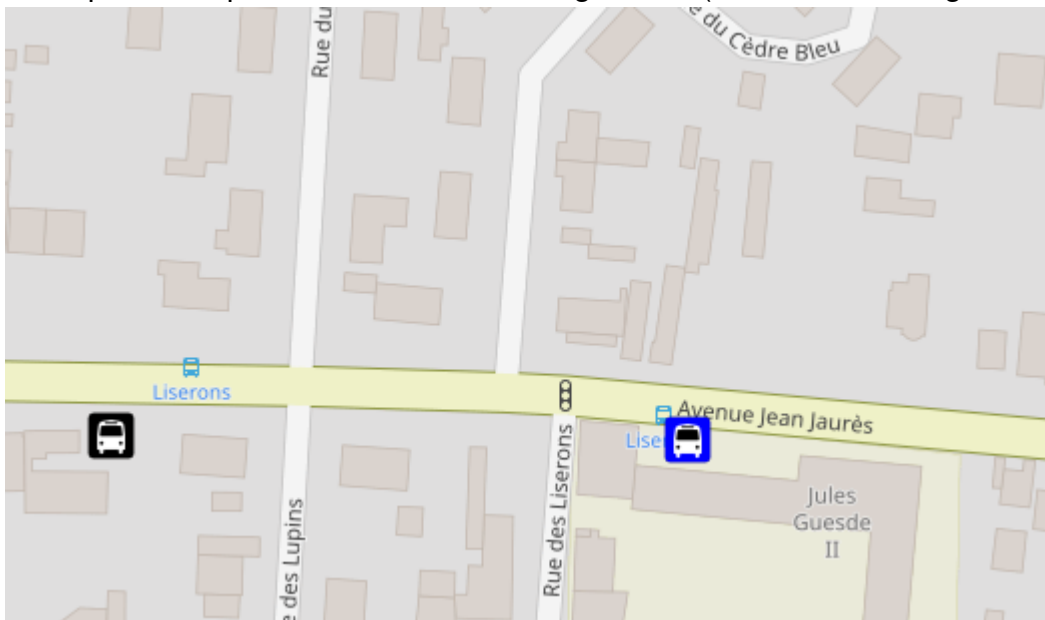
Distance entre les arrêts officiels et OSM



On constate que l'essentiel des arrêts officiels se trouve à moins de 70 mètres de leurs homologues crowdsourcés, même si l'on observe quelques arrêts situés déraisonnablement loin, jusqu'à 778 mètres soit plus de 10 minutes de marche à pied.

La comparaison de nos deux sources nous permet d'identifier 241 routepoints (qui correspondent à 174 arrêts) qui sont dans le GTFS à plus de 70 mètres de leur version crowdsourcée ; des actions de vérification voire de correction peuvent être envisagées sur ces arrêts.

Voici par exemple l'arrêt Liserons à Argenteuil (réseau R'bus, lignes 3 et 9)



Le pictogramme bleu localise l'arrêt dans OpenStreetMap, et cette position est bonne (vérifiable à l'aide de l'imagerie aérienne de l'IGN). Le pictogramme noir correspond à la position dans les données officielles : il s'agit donc de l'endroit où un utilisateur le cherchera en utilisant l'information voyageur à sa disposition, deux croisements plus loin de l'endroit où il se situe en réalité.

Il est important de noter qu'un petit écart de distance peut suffire à déstabiliser le voyageur, en le guidant à tort du mauvais côté de la rue, ou dans une autre rue à la recherche d'un arrêt de bus qui n'est pas situé à cet endroit.

L'exemple suivant comporte un écart de distance de moins de 40 mètres, mais l'arrêt n'étant pas situé dans la bonne rue, le voyageur a de grands risques de ne jamais le trouver s'il n'a à sa disposition que l'information voyageur officielle.



Carnot (réseau RATP, ligne 318)

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

Constats et moyens d'action

Voici un rappel des objectifs de cette étude :

- évaluer la complétude des données OpenStreetMap
- mesurer les différences entre les deux sources de données sur un échantillon comparable
- évaluer le potentiel des données crowdsourcées d'OpenStreetMap pour enrichir les données d'Île-de-France Mobilités, en particulier sur la localisation des arrêts
- identifier les moyens d'actions pour augmenter la couverture OpenStreetMap dans le but de maximiser le potentiel d'enrichissement des données d'Île-de-France Mobilités.

Notre analyse nous a permis de faire plusieurs constats concernant les données.

Tout d'abord les contributions de la communauté OpenStreetMap sont

géographiquement hétérogènes. Les zones les plus proches du centre de la région sont bien mieux cartographiées que les zones lointaines. La densité de population, et donc de contributeurs OpenStreetMap, est un facteur déterminant dans la quantité de données contribuées.

Ensuite, nous constatons que les deux sources de données ne sont pas comparables en l'état, du fait de grandes différences de modélisation. Nous avons pour cela défini un modèle pour comparer le plus petit dénominateur commun entre les deux sources : le routepoint, qui représente un arrêt desservi par une ligne vers une destination.

Bien qu'OpenStreetMap disposent d'environ 60% des arrêts en comparaison avec les données officielles on constate que seul un quart environ des routepoints existent dans OpenStreetMap.

En effet, dans OpenStreetMap, les lignes (et donc le fait qu'une ligne desserve un arrêt) sont bien moins cartographiés que les objets représentant les arrêts.

Afin de pouvoir mesurer plus finement les différences qualitatives entre les deux sources, nous avons par la suite construit un échantillon de données comparables, notamment en utilisant les codes REFLEX et CODIFLIGNE d'interopérabilité d'Île-de-France Mobilités que la communauté OpenStreetMap a commencé à intégrer à ses données.

Les hypothèses que nous avons choisies sont volontairement restrictives afin de garantir une comparaison fiable des données.

Sur cet échantillon contrôlé, la précision de la position, les noms d'arrêts et les numéros de ligne sont proches entre nos deux sources.

Nous constatons toutefois des limitations dans les données officielles. Il existe une grande disparité dans les formats de nommage des arrêts, qui ne reflète bien souvent pas la réalité du terrain. L'analyse précise de quelques cas particuliers a également montré la véracité et la qualité des métadonnées renseignées par les contributeurs OpenStreetMap.

Enfin, sur plus de 150 arrêts, OpenStreetMap et les données officielles divergent en terme de position d'arrêts : il s'agit d'autant de cas où les données crowdsourcées permettent dès aujourd'hui d'apporter des corrections de grande valeur pour les voyageurs franciliens.

Les moyens d'actions présentés ci-dessous sont autant de leviers permettant d'augmenter la proportion de données où un rapprochement fiable entre les deux sources est réalisable. Ils permettront donc d'identifier de nombreuses autres sources d'amélioration des données officielles proposées par Île-de-France Mobilités.

Tout d'abord, nous avons pu identifier plusieurs cas où les données crowdsourcées

permettent déjà d'apporter des améliorations dans les données officielles, notamment sur le positionnement des arrêts : il nous semble donc important de transmettre ces signalements à Île-de-France Mobilités dès maintenant.

Par ailleurs, nous avons filtré une grande quantité de données OpenStreetMap dans notre étude comparative, du fait de la difficulté pour les rapprocher de manière fiable avec les données officielles. Plusieurs actions peuvent être prises pour augmenter l'interopérabilité entre nos deux sources de données. Cela concerne aussi bien l'amélioration des outils de contrôle qualité de la communauté que l'intégration des codes de référence stables mis à disposition par Île-de-France Mobilités.

Enfin, il semble nécessaire d'augmenter la couverture des données d'OpenStreetMap pour en exploiter son plein potentiel en dynamisant les efforts de la communauté sur les sujets du transport en commun.

Cela inclut par exemple :

- d'améliorer les outils de collecte des contributeurs pour faciliter la saisie des informations liées aux lignes qui passent aux arrêts.
- de mobiliser la communauté lors de campagnes de contributions ciblées. Les zones périphériques de la région ainsi que les réseaux les plus éloignés de Paris sont de bons candidats, puisqu'ils sont aujourd'hui très peu pourvus en données.

Crédits

Cet audit a été réalisé par l'association Jungle Bus (<https://junglebus.io>), grâce au soutien de Cityway (<https://www.cityway.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 6 avril 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 5 avril 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>).