

M2 SIGAT - 2023/2024

LES DONNÉES DE FLUX AU SERVICE DES PROBLÉMATIQUES DU DÉPARTEMENT D'ILLE-ET-VILAINE

Rapport de projet



Paul Bourcier - Alexandre Blin - Jonas Constant - Hugo Frerou - Saïd Khatir
Antoine Le Doeuff - Tristan Leray - Killian Mouton - Félicia Noël - Angéline Pinilo



Grâce au réseau d'antennes de téléphonie mobile, des événements comme l'envoi de SMS, les appels ou même l'allumage des téléphones sont enregistrés. Ces événements donnent lieu à la production de données de localisation, à raison d'une toutes les 7 minutes en moyenne. Cela correspond à 205 événements par jour et par téléphone, soit 2 à 4 milliards de données de localisation par jour et par opérateur (Cerema, 2023). Chacun de ces événements étant localisé dans l'espace, la somme de ces derniers constitue une base de données conséquente. Depuis quelques années, ces données sont commercialisées, permettant notamment à des collectivités de les mobiliser pour décrire des flux de population. De par leur caractère massif, ces données pourraient offrir de nouvelles opportunités d'analyse qu'il convient d'explorer.

C'est ce genre de données émergentes que l'on nous a demandé d'étudier, à travers cet atelier professionnel. Notre commanditaire est le service ingénierie applicative du département de l'Ille-et-Vilaine. Ce service est rattaché à la direction des systèmes numériques. Plus particulièrement, la commande s'inscrit dans la mission innovation et développement de ce service. Le but de cette mission est de mener "un travail de veille technologique et méthodologique afin de répondre aux nouveaux besoins numériques de la collectivité".

Le travail qu'il nous a été demandé de mener s'inscrit donc pleinement dans ces objectifs. Il se divise en trois axes :

- La réalisation d'un état de l'art des offres de données existantes
- L'analyse de la pertinence de telles offres, du point de vue des problématiques rencontrées par les différents services métiers utilisant des données de flux au département d'Ille-et-Vilaine.
- L'évaluation des capacités techniques du SIG (Système d'information géographique) du département à intégrer ces données.

Le type de données de flux étudié dans le cadre de cet atelier a été précisé avec le commanditaire : la commande porte sur des données dites massives et qui ne sont pas issues de comptages par réseau de capteurs. Les données étudiées ici sont produites à partir de deux technologies différentes : d'une part la téléphonie mobile, et d'autre part les réseaux de satellites.

À travers cette commande, l'objectif de nos commanditaires est d'anticiper de potentiels besoins des différents services métiers du département en matière de données de flux. L'idée est d'être prêt, si de tels besoins émergent, à fournir des outils de manipulation de ces données en interne, et éviter l'usage de plateformes tierces. L'usage de ce genre de plateformes va en fait à l'encontre de l'agenda de la direction des services numériques, qui tend vers le regroupement des outils en un seul géoportail interne, et à l'encontre de sa politique de gouvernance des données, axée sur la souveraineté et la maîtrise de ces dernières.

Table des matières

I/ Éléments de définition.....	4
1. Mobilité et données massives.....	4
2. Les données de téléphonie.....	6
Structure des réseaux de téléphonie.....	7
Les types de données de téléphonies mobiles.....	9
Résolution spatiale et temporelle.....	10
Les limites des données de téléphonie.....	10
3. Les données satellitaires.....	11
Structure et fonctionnement des données satellitaires.....	11
Les limites des données de GPS.....	12
Les réglementations des données GPS.....	14
II/ L'état de l'art.....	15
1. Les solutions écartées.....	15
2. Tableau comparatif des solutions retenues.....	15
3. Quelques cas d'usages.....	17
III/ Cas d'usages potentiels pour le département et mise en application dans son SIG..	18
1. Données massives de flux : quels usages potentiels pour les services métiers du département ?.....	18
2. Illustration d'un cas d'intégration de données externes dans le SIG du département de l'Ille-et-Vilaine.....	21
3. Estimation du poids de la donnée.....	25
Conclusion.....	26
1. Rétrospective de l'atelier par rapport à la commande.....	26
2. Retour sur la gestion de projet.....	28
Références.....	30
Annexes - Fiches fournisseurs de données.....	32
Geovelo.....	32
Outdoor Vision	35
Waze for cities.....	37
Google Maps Platform	39
Flux Vision.....	40
Strava	42

I/ Éléments de définition

1. Mobilité et données massives

Dans cette étude, le terme de mobilité est entendu au sens large comme le mouvement des êtres humains (individus et groupes) dans l'espace et dans le temps. La mobilité est une caractéristique intrinsèque du comportement humain, qui, de tout temps, a dû se déplacer pour subsister. Dans les sociétés modernes occidentales globalisées, où la voiture individuelle s'est normalisée, elle occupe une place centrale dans la structuration des territoires et dans la vie des individus. Ainsi, en 2020, les ménages français consacrent en moyenne 12,5 % de leur revenu disponible aux transports (DATALAB, 2023). Quotidiennement, les Français passent en moyenne 1h à se déplacer en 2019 (Service des données et études statistiques, 2021). La résolution temporelle et spatiale des trajets des individus est souvent régulière et périodique donnant lieu à des schémas de mobilité propres à chaque territoire. La compréhension de ces schémas, tant dans leur structure que dans leur évolution, constitue un enjeu majeur pour les acteurs publics et notamment les collectivités territoriales.

Table 1. Typologie des données de mobilité. D'après (Cerema, 2022)

	Collecte active	Collecte passive
Bonne maîtrise du niveau d'échantillonnage	1 Échantillons représentatifs d'une population, obtenus par sondage aléatoire	3 Observations tendant vers l'exhaustivité du flux en un point du réseau
Faible maîtrise du niveau d'échantillonnage	2 Recherche d'une variabilité de l'échantillon, sans garantie de représentativité	4 Observations qui peuvent être nombreuses, mais dont on ne maîtrise pas qualitativement l'échantillon

Le *Big Data* est né dans les années 2010 suite à la généralisation de l'utilisation des téléphones mobiles — le taux de pénétration de la carte SIM en France atteint 100 % en 2010¹ (ARCEP, 2012) — et à une demande toujours plus importante des utilisateurs pour un accès rapide à des données de qualité, des techniques efficaces de compression, des applications de visualisation ou encore des garanties de sécurité (Yaqoob et al., 2016). Il est généralement défini en trois axes (les trois V) : le volume (la taille de la donnée), la vitesse (la vitesse d'entrée et de sortie de la donnée) et la variété (la source et le type de la donnée). Un quatrième axe est ajouté par IBM et Microsoft pour la véracité, ce terme faisant référence au "désordre" et à la fiabilité des données. McKinsey & Co. a ajouté la

¹ En juin 2023, le nombre de cartes SIM en activité était de 82,7 millions (ARCEP, 2023)

valeur comme quatrième V, celle-ci faisant référence à la valeur des connaissances cachées dans les données (Yaqoob et al., 2016). D'une manière plus commune, le *Big Data* est défini comme une collection de grandes quantités de données complexes qui rendent difficile leur gestion efficace par les technologies de traitement des données les plus modernes (Philip Chen & Zhang, 2014). Les données associées aux *Big Data* sont : les données issues de l'Internet Of Things (IoT), les données des réseaux sociaux, les données de téléphonie mobile, les données web ou encore les données de capteur (Yaqoob et al., 2016). L'ensemble de ces données est potentiellement géocodée auquel cas, elles tombent dans le type 4 de la typologie du Cerema.

En fonction de la législation des pays, la collecte et la distribution de données concernant les individus peuvent être soumises à des restrictions. En France, les données propres à la détermination de la position individuelle, étant considérées comme des données à caractères personnelles², elles doivent de respecter les limites définies par la Commission Nationale Informatique et Libertés (CNIL) et par le Règlement Général de Protection des Données (RGPD). Ainsi, les producteurs doivent respecter les règles suivantes du secret statistiques (INSEE, 2023):

- Pour les fichiers agrégés : il ne doit pas être possible d'identifier une personne (article L.312-1-2 du code des relations entre le public et l'administration). Le risque de réidentification par inférence dans les données agrégées doit également être pris en compte. Ce risque correspond aux cas où de nouvelles informations sur un individu peuvent déduites à partir des données ;
- Pour les fichiers de données individuelles : « *il n'est pas possible de fournir des données individuelles sur les personnes morales en respectant strictement l'anonymat, sauf à supprimer du fichier des variables considérées comme essentielles pour l'analyse de tels fichiers : activité économique, localisation, indicateur de la taille* » (INSEE, 2023).

Ces restrictions ont un impact notable sur la résolution spatiale des jeux de données disponibles sur le marché dans la mesure où les données doivent être agrégées pour respecter la législation. Les opérateurs sont également contraints de supprimer les données après un an de stockage légal (Cerema, 2022, 2023). Les offres de services identifiées sur le marché français concernent uniquement des données de téléphonie mobile et des données GPS. Il convient d'indiquer dès à présent que l'ensemble de ces solutions sont des données prétraitées et que, de fait, le caractère de *Big Data* ne peut leur être raisonnablement associé dû au caractère structuré de la donnée fournit ainsi que sa masse (les données étant agrégées). En d'autres termes, la

² Selon l'article 4 du RGPD, les « données à caractère personnel » correspondent à toute information se rapportant à une personne physique identifiée ou identifiable.

donnée brute collectée par le producteur peut être du *Big Data*, mais la donnée distribuée ne l'est plus à proprement parler.

2. Les données de téléphonie

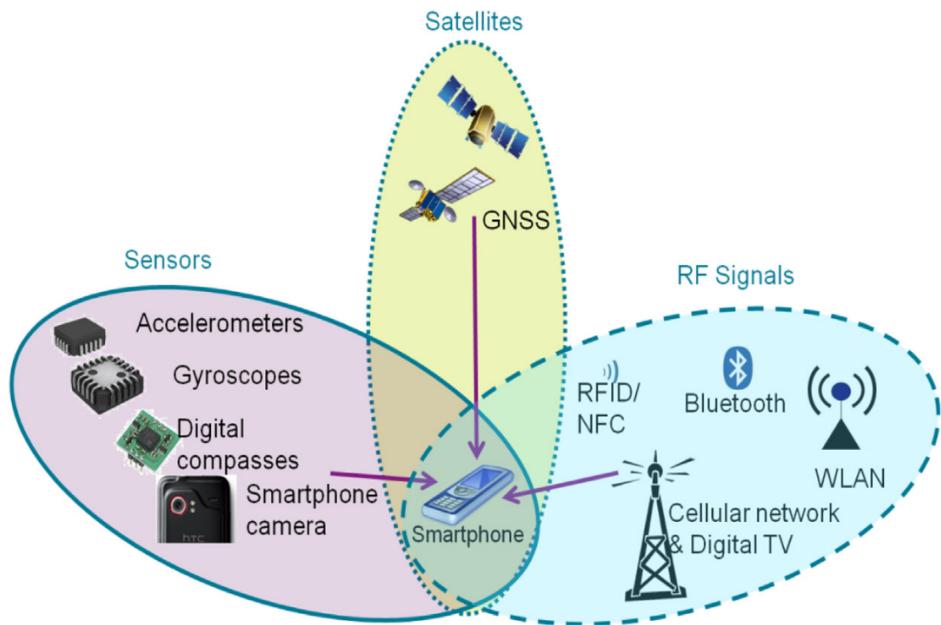


Figure 1. Systèmes et techniques de collecte de données pour les téléphones mobiles. Figure reproduite depuis Pei et al., 2013

Les téléphones portables permettent de collecter un nombre important de données variées, et ce, au moyen de technologies embarquées diverses. Wang et al., 2018 identifie trois systèmes d'acquisition de données par les portables modernes : les radios fréquences (par ex. bluetooth, réseau de téléphonie), les satellites (Global Navigation Satellite Systems (GNSS)) et les capteurs (par ex. accéléromètres, gyroscopes) (fig. 1). La position du téléphone peut être déterminée, plus ou moins précisément, à partir des deux premiers types. Dans cette section, nous nous intéresserons aux données de téléphonie mobile.

Dans le cadre de l'analyse des mobilités, les données de téléphonie mobile ont marqué un tournant, permettant l'acquisition massive de données géolocalisées en temps réel (Barbosa et al., 2018). Ainsi, à partir des années 2010, conjointement au développement du *Big Data*, les données relatives aux téléphones mobiles sont devenues un sujet à part entière dans la littérature scientifique (Blondel et al., 2015; Calabrese et al., 2014; Okmi et al., 2023).

Structure des réseaux de téléphonie

Un réseau téléphonique typique³ est composé de stations de transmission de base (*Base Transceiver Station* (BTS)) couramment appelée antenne relais. Ces BTS sont équipées d'antennes et de systèmes voués à faciliter la communication entre le réseau et les téléphones qui lui sont associés. Chaque BTS couvre un espace défini, nommé cellule, lui-même composé de trois sous-cellules ; il s'agit des unités de base du réseau téléphonique. Un contrôleur de station de base (*Base Station Controller* (BSC)) est associé à un ensemble de BTS afin de gérer le réseau radio, l'établissement des appels et des transferts entre les BTS pendant les appels ou les connexions de données. On appelle aire de localisation (LA) l'aire couverte par les cellules des BTS sous la responsabilité d'un BSC. De manière générale, des dizaines, voire des centaines de cellules partagent un seul LA. Les BSCs sont eux-mêmes gérés par un *Mobile Switching Center* (MSC) chargé de la gestion des connexions de bout en bout (par exemple, entre utilisateurs mobiles, ou des utilisateurs mobiles vers d'autres réseaux (des réseaux téléphoniques, d'autres réseaux mobiles ou des réseaux internet, etc.)). Ces MSCs gèrent également les exigences de mobilité et de transfert pendant l'appel ou la connexion de données, et s'occupent de la facturation. Un MSC contient souvent un registre de localisation des visiteurs (*visitor location register*, (VLR)), qui est une base de données enregistrant la localisation de tous les abonnés mobiles actuellement présents dans la zone de service du MSC (fig. 2) (Fekih et al., 2021; Huang et al., 2019).

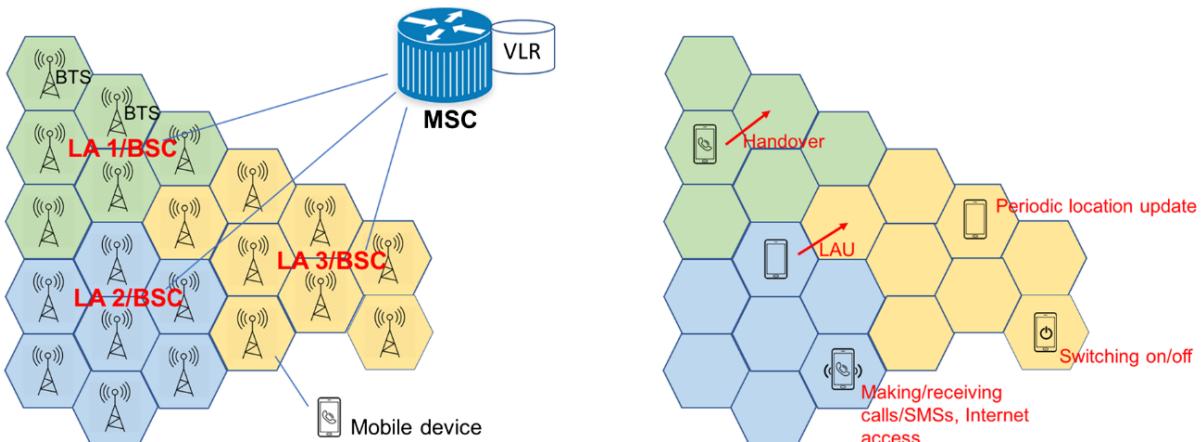


Figure 2. Structure d'un réseau téléphonique (gauche) et évènements réseaux enregistrés dans les données orientées réseau (droite). Figure reproduite depuis Huang et al., 2019

Dans un souci de facturation, les réseaux téléphoniques déterminent constamment et fréquemment la localisation des téléphones, et ce, même s'ils sont en veille. La localisation du téléphone est généralement calculée à partir du BTS à laquelle il se

³ Il existe de nombreux types de réseau, nous prenons l'exemple d'un réseau de type Global System for Mobile Communications (GSM), 2G. La structure spatiale des autres réseaux (UMTS, LTE, etc.) est globalement calquée sur le GSM.

connecte. Ainsi, la précision spatiale de la localisation varie en fonction de la distribution des BTS sur le territoire.

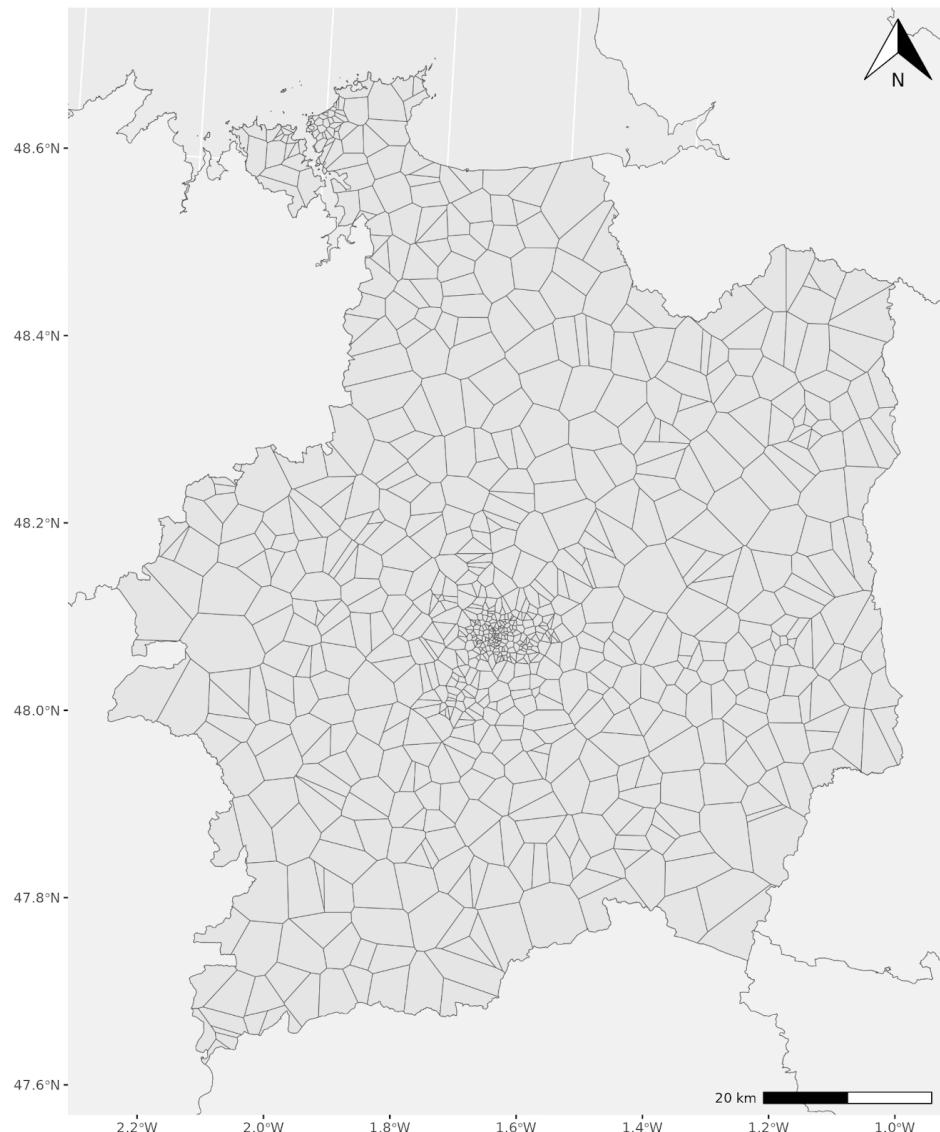


Figure 3. Cellules du réseau téléphonique en Ille et Vilaine modélisé par la méthode des polygones de Voronoï. Données : ANFR, 2023

En France, il y a trois principaux opérateurs qui se répartissent le marché de la téléphonie mobile (entre 15 et 35 % du marché respectivement) : Orange, Bouygues et SFR (Cerema, 2023). Le département d'Ille-et-Vilaine compte 854 BTS en service au 24 novembre 2023. 30 % des installations appartiennent à Orange, 25 % à SFR, 25 % à Bouygues Telecom et 20 % à Free, sachant que plusieurs installations d'opérateurs différents peuvent coexister sur un même BTS. Environ 20 % des BTS en service sont concentrées dans la ville de Rennes pour une densité 3,26 BTS par km². Dans les autres communes du département, la densité médiane de BTS par km² est de 0,08 (ANFR, 2023). Les polygones de Voronoï sont régulièrement employés pour modéliser les cellules du réseau téléphonique (Bachir et al., 2019; Doyle et al., 2011; Fekih et al., 2021; García

et al., 2016; Nabavi Larijani et al., 2014). La figure 3 présente le réseau cellulaire du département de l'Ille-et-Vilaine modélisé par la méthode des polygones de Voronoï.

Les types de données de téléphonies mobiles

Deux groupes de données de localisation sont enregistrés dans les réseaux mobiles. Le premier est axé sur les événements (*event-driven*) et le second sur le réseau (*network-driven*) (Bachir et al., 2019; Blondel et al., 2015; Calabrese et al., 2014).

Les données mobiles basées sur les événements (*event-driven*) sont enregistrées dès que le téléphone utilise activement un service (par ex. envoyer un SMS (*Short Message Service*), passer un appel ou utiliser internet). Ces données sont collectées par l'opérateur dans un objectif de facturation. Deux groupes peuvent être distingués au sein de cette catégorie :

- Enregistrement des détails de l'appel (*Call Detail Records*, CDRs) : un CDR contient le détail d'un appel ou d'un message. Il n'existe pas de standard quant au contenu des CDRs. Ils contiennent généralement les numéros de téléphone encryptés (origine/destination), une empreinte temporelle, une durée, un type de communication, l'ID de l'antenne téléphonique d'origine associée, et, quelques fois, l'ID de l'antenne associée au téléphone ciblé.
- Enregistrement des détails du protocole internet (*Internet Protocol Detail Records*, IPDRs) : il s'agit des logs de l'accès à internet, (c.-à-d. les détails de l'usage d'internet).

Les données mobiles axées sur le réseau (*network-driven*) enregistrent la mise à jour de la localisation des données suite aux types suivants d'action sur le réseau :

- Éteindre ou allumer le téléphone.
- Mise à jour de l'aire de localisation (*Location Area Update*, LAU) : dès qu'un téléphone en veille passe d'une cellule du réseau à une autre.
- Transfert : quand un utilisateur passe d'une cellule à une autre durant un appel ou s'il est en train d'utiliser internet.
- Passer un appel, envoyer un SMS ou accéder à un service internet. À l'inverse des CRDs, ces événements n'enregistrent que la localisation et l'empreinte temporelle du téléphone d'origine, sans information concernant le téléphone de destination.

- Localisation périodique : si le téléphone est resté fixe durant une période donnée, généralement quelques heures, une demande de localisation périodique est effectuée.

Il ressort que les données axées sur le réseau sont plus nombreuses que celles basées sur les événements, mais ne permettent pas d'obtenir d'informations concernant l'utilisateur autres que sa localisation.

Résolution spatiale et temporelle

La fréquence des données dépend à la fois de leur type (axés sur l'événement ou sur le réseau) et de l'utilisateur lui-même. Par exemple, chaque appel vocal génère un CDR. Cependant, le même appel peut donner lieu à plusieurs enregistrements orientés réseau si le téléphone se déplace le long de plusieurs cellules pendant l'appel. Ainsi, comme indiqué plus haut, les données de CDR sont moins massives que celles orientées réseau. À titre d'exemple, Jiaqi, 2018 a évalué que l'intervalle de temps médian entre deux points consécutifs étaient de 90 secondes dans une donnée orientée réseau. Le Cerema indique quant-à-lui, pour chaque opérateur français, un événement toutes les 7 minutes en moyenne par téléphone, soit 205 événements par jour et téléphone (c.-à-d. 2 à 4 milliards de données de localisation par jour et par opérateur au total) (Cerema, 2023).

La résolution spatiale des données de téléphonie est sujette à une forte variabilité en fonction du nombre de BTS disponible dans la zone où l'utilisateur se connecte au réseau. Ainsi, la précision peut être de l'ordre de quelques mètres dans des centres urbains disposant d'une forte densité de BTS à plusieurs kilomètres en zone rurale (Huang et al., 2019). Des techniques existent afin de préciser la localisation des terminaux mobiles, mais, en raison des lois sur l'anonymisation des données, elles peuvent être prohibées, comme en France où la triangulation des données de position est interdite (Bachir et al., 2019; Cerema, 2022). La résolution spatiale des données peut être affinée par un croisement avec des données GPS — celles-ci étant plus précises que les données de téléphonie mobiles. À notre connaissance, en France, seul Orange propose une telle solution, permettant, dans des zones où la densité de données GPS et la densité de BTS est suffisamment importante, de descendre à un maillage de 50 par 50 mètres.

Les limites des données de téléphonie

De nombreuses sources d'incertitudes sont associées aux données de téléphonies mobiles. On peut citer : le fonctionnement complexe et évolutif du réseau en lui-même ; la précision spatiale et temporelle comme évoqué plus haut ; le double équipement ou les téléphones éteints qui rendent difficile l'association d'un mobile à une personne ; la représentativité des données. Concernant cette dernière source d'incertitudes, les données collectées par un opérateur ne concernent que la partie des utilisateurs utilisant une carte SIM de leur réseau. Dans ce cadre, si l'opérateur souhaite analyser ses

données au regard de l'ensemble de la population, il lui faudra les redresser. Ainsi, la donnée finale inclura nécessairement des incertitudes qu'il faudra ajouter à celles liées à la collecte des données brutes. De nouveau, le détail du redressement de l'échantillon n'est pas révélé par les opérateurs.

3. Les données satellitaires

Structure et fonctionnement des données satellitaires

Prenons comme exemple le système GPS (*Global Positioning System*). Ce système repose sur une constellation de satellites, dépendante des USA, en orbite autour de la Terre, conçus pour fournir des informations de localisation et de temps précises à quiconque possède un récepteur GPS. D'autres systèmes satellitaires équivalents au GPS existent, comme le GLONASS (équivalent Russe), Beidou (équivalent Chinois) ou encore Galiléo (équivalent européen). Les données de position issues de satellites s'appuient en général sur plusieurs constellations à la fois. C'est donc souvent un abus de langage de parler de "données GPS", les récepteurs de signal satellite utilisés aujourd'hui combinant les systèmes.

À des milliers de kilomètres au-dessus de la Terre, plus de 30 satellites GPS émettent constamment des signaux radio portant des données cruciales, notamment leur identité, leur position orbitale et l'heure exacte à laquelle le signal a été émis. Ces signaux se propagent à travers l'espace en direction de la Terre.

Lorsqu'un récepteur GPS, qu'il soit intégré à un téléphone, à une montre, à un système de navigation automobile ou à un appareil dédié, capte ces signaux de quatre satellites ou plus, le processus de triangulation entre en jeu. En mesurant le temps que met chaque signal pour atteindre le récepteur, celui-ci peut calculer la distance entre lui-même et chaque satellite.

Les distances mesurées à partir de plusieurs satellites sont ensuite utilisées par le récepteur GPS pour effectuer des calculs complexes, déterminant ainsi la position exacte en termes de latitude, longitude et altitude. La précision de cette localisation dépend du nombre de satellites auxquels le récepteur a accès.

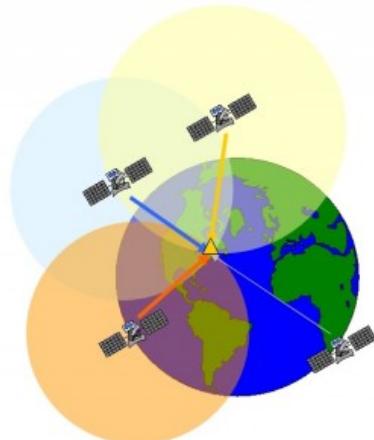


Figure 4. Triangulation du système GPS.

Données : 3grt-solutions.com,
2020

Pour illustrer nos propos, nous pouvons citer les applications de navigation automobile (Waze, google maps, ...), les applications de fitness (Strava, Garmin, ...), les gestions de flottes automobiles pour les entreprises, l'agriculture de précision, la localisation des appels d'urgence, et de nombreux autres utilisations qui accompagnent notre quotidien.

Ces données de flux GPS nous offrent un aperçu des mouvements dans l'espace. En traquant les changements de position au fil du temps, on peut déduire des détails intéressants sur la vitesse, la direction et les habitudes de déplacement des objets ou des personnes équipés de récepteurs GPS.

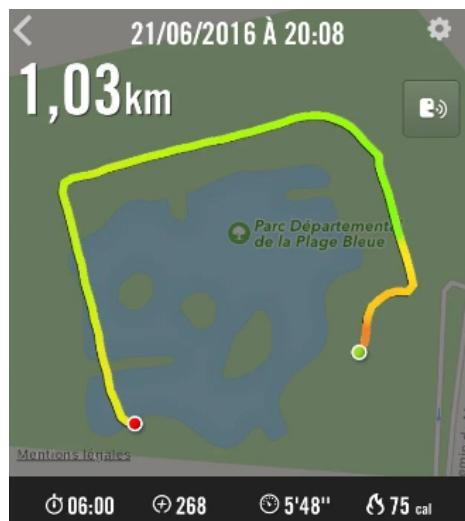


Figure 5. Vitesse et altitude captée par GPS.

Données : 01net.com, 2016

La figure 4 ci-dessus illustre les capacités des données de flux collectées par les GPS. En effet, sur cette représentation d'application de running, il est possible de voir le trajet précis de l'individu, son allure à chaque instant, ainsi que l'altitude entre le point de départ et d'arrivée.

Les limites des données de GPS

Bien sûr, ces données ne sont pas sans leurs questions éthiques. La précision avec laquelle on peut suivre les mouvements des individus soulève des préoccupations légitimes en matière de vie privée. Les gardiens de ces données, qu'il s'agisse d'entreprises ou d'organisations, doivent jongler avec les avantages pratiques et les implications de confidentialité.

L'une des limitations les plus notables se manifeste en milieu urbain, où les signaux GPS peuvent être réfléchis par des bâtiments et des obstacles, entraînant une moindre précision, surtout dans des ruelles étroites. Cet "effet canyon" impacte également le nombre de satellites visibles par le récepteur (une ligne de vue directe avec au moins quatre satellites étant nécessaire pour localiser le récepteur). Dans des environnements où la visibilité du ciel est limitée, comme dans des canyons profonds ou des forêts denses, la précision de la localisation peut être compromise.

Les obstacles naturels, comme les montagnes ou des conditions météorologiques défavorables, peuvent également affecter la réception du signal GPS. Les signaux peuvent être bloqués ou atténusés, entraînant une perte momentanée de précision, surtout en cas de mauvais temps ou de couverture nuageuse.

Par ailleurs, le nombre de constellations de satellites pris en compte par le récepteur influence la qualité du positionnement. L'image suivante illustre la problématique des traces GPS en milieux urbains avec des montres de running de différentes gammes.

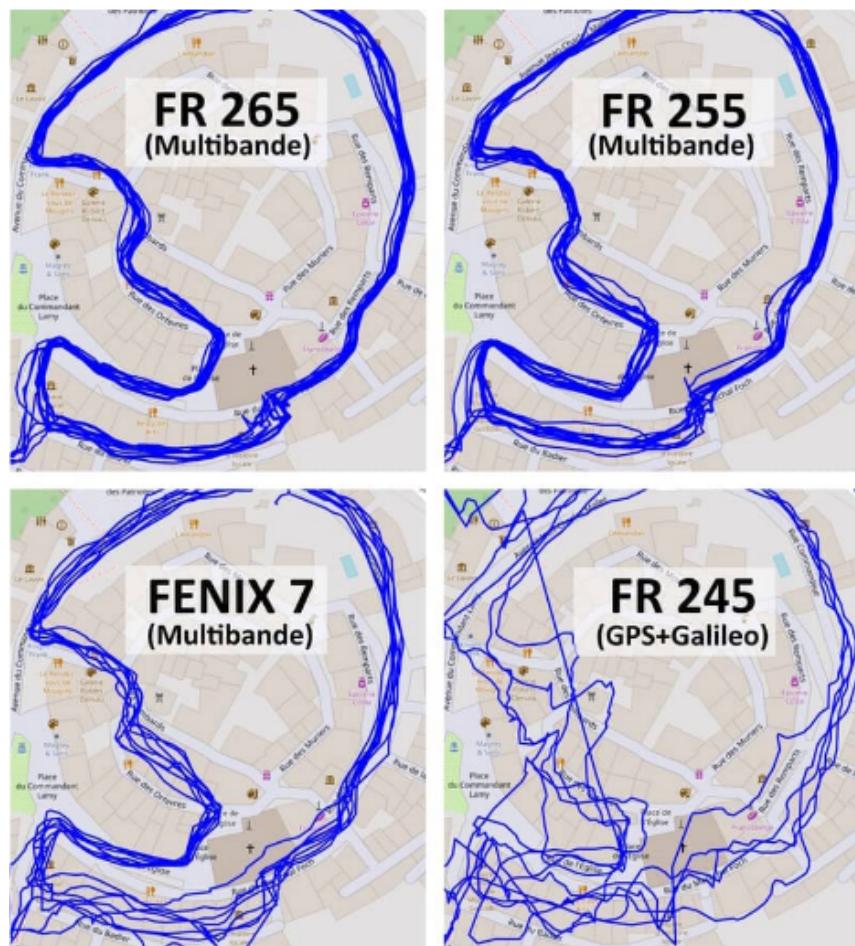


Figure 6. Précision GPS en milieu urbain.

Données : sport-passion.com

L'utilisation constante du GPS peut également poser des problèmes de vie de la batterie, en particulier sur les appareils mobiles. Cette consommation d'énergie accrue peut réduire considérablement la durée de vie de la batterie, obligeant les utilisateurs à gérer soigneusement leur utilisation du GPS.

Les réglementations des données GPS

Les lois entourant les données GPS jouent un rôle crucial dans la protection de la vie privée, la sécurité et la régulation de l'utilisation de ces informations sensibles. Ces réglementations varient d'un pays à l'autre, mais certaines tendances communes émergent pour encadrer cette technologie omniprésente.

En matière de protection de la vie privée, les lois exigent souvent des entreprises qu'elles soient transparentes quant à la manière dont elles collectent, stockent et utilisent les données GPS. Le consentement éclairé est un principe clé, obligeant les entreprises à obtenir une approbation explicite des individus avant de collecter leurs données de localisation.



Figure 7. Précision GPS en milieu urbain.

Données : support-apple.com, 2023

La durée de conservation des données est un autre aspect régulé. De nombreuses juridictions imposent des règles spécifiques sur le temps pendant lequel les données GPS peuvent être stockées. Les entreprises doivent définir des politiques claires sur la conservation des données et supprimer les informations sensibles une fois qu'elles ne sont plus nécessaires. En pratique, la durée de conservation des données GPS en France

peut dépendre de divers facteurs, tels que le contexte de la collecte, le consentement des individus, les obligations légales et réglementaires, ainsi que les politiques de conservation de l'entité responsable du traitement des données. Les entreprises ou organisations collectant ces données doivent définir des politiques de conservation spécifiques, et ces politiques doivent être conformes aux exigences du RGPD.

C'est sur ces technologies que s'appuient les données proposées par les différentes offres analysées dans le cadre de ce travail. Nous proposons ainsi un état de l'art des solutions existantes.

III/ L'état de l'art

Nous avons décidé de réaliser un état de l'art sur les offres de données de flux, afin de les recenser et de pouvoir les comparer les unes aux autres.

1. Les solutions écartées

Mais avant cela, nous avons écarté certaines offres qui ne répondraient pas à la demande émise par le département de l'Ille-et-Vilaine. Tout d'abord, il y a les services proposant un accès aux données uniquement en consultation, sous forme de datavisualisation. En effet, ce type de services implique l'usage d'une plateforme tierce (généralement un portail web), ce qui va à l'encontre de la politique de la direction des systèmes numériques en matière de données. L'objectif étant, au contraire, de récupérer des données "brutes" pour les centraliser dans une seule plateforme à destination des services métier, et dont le département a la maîtrise. Ensuite, il y a les applications qui ressemblent à des offres de services déjà choisies comme MyTraffic dont le fonctionnement est très similaire à celui de Waze, mais qui comptabilise moins d'utilisateurs. Et enfin, nous avons écarté toutes les données utilisant la méthode du comptage par capteur, le département ayant déjà plusieurs capteurs installés à divers endroits, et notre commanditaire souhaitait que nous orientions nos recherches sur des données issus d'autres solutions.

2. Tableau comparatif des solutions retenues

Voici les offres de services qui correspondent aux attentes précisées par les commanditaires, on en recense 6. A noter que nous les détaillons chacune en annexe.

Solutions	Producteur	Méthode de production	Echelle spatiale	Accessibilité	Granularité
Strava Metro	Strava	Collecte des données utilisateurs	Traces GPS	API	Temps réel
Waze for Citizens	Waze	Collecte des données utilisateurs	Traces GPS	API	Temps réel
Flux vision	Orange Business	Utilisation des données mobiles	IRIS	CSV sur le cloud d'Orange, Power BI	toutes les 30 minutes
Strava Metro	Strava	Collecte des données utilisateurs	Traces GPS	API	Temps réel
Cycling Insights	Geovelo / Eco-compteur	Collecte des données utilisateurs	IRIS	Données brutes (nombreux formats disponibles) à télécharger manuellement	4 ou 5 flux par jour
Outdoor Vision	PRNSN ⁴	Aggrégation de traces GPS de plusieurs producteurs	Traces GPS	Plateforme de visualisation (téléchargement .geojson possible)	3 fois/an
Google Map Platform	Google	Collecte des données utilisateurs	Traces GPS	API ou SDK	Temps réel

4 Pôle ressource national des sports de nature

3. Quelques cas d'usages

Afin de mieux comprendre comment les différentes données de flux sont utilisées par les collectivités territoriales, nous vous présentons un exemple de cas d'usages par offre de service.

La donnée Flux Vision (offre tourisme) est utilisée dans la majorité des Agences de Développement Touristique (ADT) pour estimer la fréquentation touristique tant qualitativement que quantitativement. La RATP utilise Flux Vision pour analyser les flux de déplacement et ainsi adapter leur réseau à la demande. Nous pouvons également citer la ville de Mulhouse, qui a mobilisé la donnée Flux Vision afin de mieux comprendre la fréquentation de son centre ville. Cette analyse permettrait de le redynamiser en pilotant mieux les horaires des commerces, la gestion du stationnement et les transports.

En 2016, l'agglomération de Versailles Grand Parc a décidé de s'associer avec Waze afin de mieux pouvoir appréhender le trafic. En effet, chaque année depuis 2010 se déroule la course à pied Paris-Versailles qui impacte plusieurs communes et des riverains se plaignaient de ne plus savoir comment se déplacer. La collaboration est donc née pour contrer ce problème. Elle est gratuite et fonctionne sur la base de l'échange. Waze partage des données sur le trafic dans l'agglomération tandis que cette dernière partage des informations sur les grands évènements, les travaux de voirie et autres incidents routiers sur l'application, de manière à en informer les automobilistes.

Quant à la solution Outdoor Vision, 48 départements ont pu être intégrés au projet, afin de tester les fonctionnalités de la plateforme. Parmi eux, le département du Gers a pu mobiliser cette plateforme pour analyser des données de fréquentation de sites naturels, lui permettant d'aboutir à des réalisations telles que :

- Sécurisation d'un gouffre sur un sentier karstique dans les bois de Broustes
- Pose de jalons regroupant différents balisages
- Aménagement des abords du lac Bousquetara avec des bancs et tables de pique-niques

Toutefois, il ne s'agit que d'un usage de consultation de la plateforme, et non une récupération des données de celle-ci. À ce jour, seuls 5 départements possèdent le statut "d'expert", soit la version payante d'Outdoor Vision permettant d'exporter les traces GPS depuis la plateforme.

III/ Cas d'usages potentiels pour le département et mise en application dans son SIG

1. Données massives de flux : quels usages potentiels pour les services métiers du département ?

Outre l'identification des offres de services de données massives de flux via la réalisation d'un état de l'art, la commande qui nous a été faite consiste également en l'exploration des cas d'usages possibles de ces données pour les différents services métiers, présents au département.

Notre approche, pour réaliser cet aspect de la commande, a sensiblement évolué au cours de l'atelier. En effet, au début de celui-ci, l'approche consistait plutôt à chercher ce que des collectivités en dehors du département de l'Ille-et-Vilaine ont pu faire à partir des offres identifiées. L'idée était alors de considérer ces exemples de réalisation comme des pistes de cas d'usages pour les services métiers en Ille-et-Vilaine. Nous avons d'ailleurs évoqué certains de ces cas d'usages en partie 2 de ce rapport. Notre approche vis-à-vis de cette partie de nos missions a changé durant le temps de l'atelier, puisque nous avons eu la possibilité de contacter directement les services métiers de l'Ille-et-Vilaine, afin de comprendre leurs potentiels besoins en matière de données de flux, et donc d'imaginer avec eux des cas d'usages. Cette prise de contact s'est faite via l'organisation d'une réunion dans les locaux du département. Cette réunion a consisté en la présentation de notre travail d'une part, et en l'animation d'un "world café" d'autre part, soit un atelier de débat mouvant ayant permis aux services métiers présents de discuter de leurs problématiques quotidiennes en matière d'usage (ou non-usage) de données. Dans une moindre mesure, certains services métiers ont pu être contactés via l'organisation d'entretiens téléphoniques/visio. Nous présentons ici les pistes de cas d'usages tels qu'évoqués à partir de cette nouvelle approche.

Avant d'avoir pu interroger une partie des services métiers du département, nous avions identifié avec nos commanditaires les domaines d'applications potentiels des données massives de flux, en lien avec les compétences du département. Quatre grandes thématiques ont alors été identifiées :

- la gestion des routes à la charge du département
- la gestion et le développement des pistes cyclables
- la gestion des espaces naturels sensibles (E.N.S)
- l'action sociale

Après la prise de contact avec certains services métiers, il s'est avéré que ces thématiques correspondent assez bien à de potentiels cas d'usages des données de flux. Toutefois, nous n'avons pas reçu de réponse de la part du service en charge du

développement des pistes cyclables et ne pouvons donc pas nous exprimer sur leurs usages des données de mobilité. Quelques autres cas d'usages ne relevant pas de ces thématiques nous sont également parvenus.

Pour la thématique de la gestion des E.N.S, qui relève de la mission espaces naturels et paysages du service patrimoine naturel, les données massives de flux pourraient être utilisées pour améliorer la finesse des études de fréquentation. Ces études sont menées au moins une fois par an pour mieux connaître la fréquentation des espaces naturels dont le département a la charge. Il s'agit d'estimations du nombre de visiteurs dans ces espaces, ainsi que des itinéraires qu'ils empruntent. Ces études permettent de comprendre les pratiques des usagers de ces espaces, afin de remplir au mieux les missions du département, à savoir l'accueil du public et la préservation écologique des milieux. Cette politique d'accueil et de préservation est pilotée par le plan de gestion, qui évolue périodiquement. Jusqu'alors, elles sont réalisées grâce à des données issues de capteurs (une vingtaine de sites en sont équipés en Ille-et-Vilaine) et des données d'enquêtes réalisées sur le terrain auprès des visiteurs. Croiser ces informations à des données massives de flux permettrait d'obtenir des estimations plus fiables, en s'affranchissant de certains biais, notamment la faible couverture temporelle des enquêtes (souvent menées le week-end uniquement), et en améliorant la précision spatiale des itinéraires étudiés. Dans ce cas, les données massives de flux peuvent avoir une plus-value si elles ont une bonne granularité temporelle et spatiale. Cette amélioration de la finesse des études de fréquentation pourrait se traduire par différents exemples de cas d'usages :

- mesure de l'usage des parkings mis à disposition des visiteurs
- établissement d'une signalétique éloignant les visiteurs des zones à préserver et les guidant vers les zones visitables

Nous avons également relevé qu'une dimension attributaire des données de flux pourrait être utilisée par ce service métier pour mieux détecter les différentes pratiques de mobilité au sein de ces espaces, notamment pour prévenir des conflits d'usage entre vélo et marche à pied.

Quant à la thématique de l'action sociale, les services métiers concernés au département ont principalement mis en avant une potentielle utilité des données massives de flux pour l'évaluation de la fréquentation des lieux d'accueil, comme les CDAS (centre départemental d'action sociale) ou les centres PMI (protection maternelle et infantile). Là encore, il s'agit de compléter des informations récoltées jusqu'alors par comptage. Un premier cas d'usage évoqué est qu'une meilleure précision dans l'estimation de la fréquentation des lieux d'accueil permettrait d'adapter les horaires d'ouvertures et les effectifs en fonction des pics de fréquentation. Cette amélioration pourrait aussi conduire à mieux évaluer le non-recours au droit, en comparant les individus fréquentant les lieux d'accueil et le total des ayant droit.

Ensuite, la compréhension de la zone de provenance des visiteurs de ces lieux d'accueil, à travers une dimension attributaire des données massives de flux, permettrait d'améliorer ces évaluations. Un cas d'usage a par exemple émergé autour de cette idée, à propos de l'aide aux personnes handicapées. En Ille-et-Vilaine, il existe une MDPH (Maison Départementale pour les Personnes Handicapées) qui accueille ces personnes, et réalise notamment des évaluations de handicap afin de mettre en place des aides spécifiques. En parallèle de cette MDPH, il existe des annexes dans le territoire, appelées CLIC (Centres locaux d'information et de coordination). Ces annexes ne proposent pas d'évaluation de handicap, mais permettent toutefois d'avoir des renseignements. Une dimension attributaire de la donnée de fréquentation permettrait par exemple de connaître la provenance des personnes fréquentant la MDPH uniquement pour obtenir des renseignements, et les réorienter vers une CLIC s'il en existe une plus proche de chez eux que de la MDPH.

Outre l'amélioration de ces études de fréquentation des lieux d'accueil, des données massives de flux permettant de connaître le trafic routier, notamment en temps réel ou proche temps réel, pourraient être mobilisées pour certains usages. Nous avons pu récolter des cas d'usages potentiels concernant l'analyse des déplacements du personnel de santé se rendant à domicile, ainsi que des familles d'accueil en charge d'enfants dans le cadre de l'ASE (Aide sociale à l'enfance). Une meilleure connaissance du trafic routier pourrait permettre une optimisation des trajets de ces acteurs.

Concernant la thématique des routes, plusieurs besoins impliquant l'usage de données massives de flux ont émergé. D'abord, de la même manière que pour les thématiques évoquées jusqu'à présent, ces nouvelles sources de données sont pressenties comme pouvant enrichir les données déjà à disposition. Dans le cas des routes, les services manipulent déjà, outre les données sur l'infrastructure du réseau (limitations de vitesse, poids maximal autorisé sur un ouvrage...), des données sur l'intensité du trafic et les types de véhicule en circulation. Ces données étant issues d'un réseau de capteurs, la finesse des flux récoltés est nécessairement plus limitée qu'avec des données massives de flux (le nombre de capteurs installés étant forcément plus faible que le nombre de ping d'un navigateur GPS par exemple). En outre, une amélioration de la précision spatiale et temporelle de la connaissance du trafic routier pourrait permettre de mieux calibrer l'installation de déviations routières.

Ensuite, la possibilité d'obtenir des informations en temps réel sur le trafic (ou quasi temps réel) a pu être évoquée. C'est notamment la connaissance d'événements ayant lieu sur la route qui pourrait intéresser le service des routes au département. L'enjeu relevé est la détection plus rapide d'accidents, permettant d'améliorer les temps d'interventions.

Enfin, la présence d'une dimension attributaire à ces données massives de flux pourrait être mobilisée également. En particulier, la connaissance du poids du véhicule. En effet, le département est en charge d'un certain nombre d'ouvrages comme des ponts, dont certains présentent un poids maximal autorisé (le tonnage sur ouvrage). Une dimension attributaire des données massives de flux comportant le poids des véhicules pourrait

permettre d'estimer leur poids cumulé sur un ouvrage (s'il n'est pas équipé de capteurs) et donc de prendre des dispositions en cas de surcharge.

Des cas d'usages des données massives de flux plus "transversaux" et ne relevant pas des trois thématiques évoquées jusqu'à présent nous sont également parvenues à travers notre prise de contact. Nous pouvons citer :

- aménagement des horaires d'embauche du personnel en fonction du trafic routier, afin d'éviter les embouteillages
- amélioration des études d'accessibilité des personnes aux services de base
- meilleure compréhension des pratiques de mobilités en dehors des déplacements domicile - travail, notamment des collégiens
- amélioration de la compréhension de l'occupation des locaux du département dans le temps

Cette première prise de contact nous a permis d'identifier un certain nombre de cas d'usages concernant les données massives de flux. Nous répondons donc par la positive à la question "Les données de flux sont-elles intéressantes en vue d'applications métiers telles que pressenties par les commanditaires ou manifestées par les services métiers ?". Toutefois, notre analyse ne permet pas de savoir quelle offre de données (ou quelles offres) doit être choisie spécifiquement, ni pour quels cas d'usage, parmi ceux identifiés. D'une part, car cette phase de prise de contact fut relativement sommaire et ne constitue pas une formalisation des besoins complète (pour chaque thématique, seuls quelques membres du service métier concerné ont pu être contactés). D'autre part, ces cas d'usages identifiés n'ont pas réellement été mis en perspective vis-à-vis de ce que les offres de données identifiées peuvent apporter. En d'autres termes, nous n'avons pas demandé aux membres du département auditionnés ce qu'ils pensaient de telle offre de données ou telle autre.

Reste la question de la capacité du SIG à intégrer ces nouvelles offres de données. Nous explorons désormais cette problématique à travers un exemple d'intégration, puis par le prisme du volume de la donnée.

2. Illustration d'un cas d'intégration de données externes dans le SIG du département de l'Ille-et-Vilaine

Afin de mieux se représenter le parcours de données issues d'une source externe, son intégration dans un SIG et son utilisation au sein des métiers du département de l'Ille-et-Vilaine, nous proposons ci-après un exemple illustré avec un cas d'usage précis.

Prenons pour illustration un employé du département Ille-et-Vilaine travaillant au service des routes et souhaitant analyser les risques d'accidents.

Le responsable dispose déjà de données internes fournies par le département : le tracé des routes, les panneaux de circulations (dont ceux indiquant les limitations de vitesses, par exemple), l'emplacement des radars, etc.

Il aurait besoin en supplément de données de fréquentation des axes routiers pour une zone d'étude (que nous allons appeler pour l'exemple zone A). C'est ici que vont intervenir les données externes de mobilité. Ici nous prendrons l'exemple du fournisseur de données Waze.

1. Intégration des données Waze dans le Système de Gestion de Base de Données (SGBD) du département de l'Ille-et-Vilaine via une API

Avant que le responsable puisse faire une demande via un outil de cartographie, voyons d'abord comment la donnée est intégrée dans le SIG du département de l'Ille et Vilaine.

Après avoir passé un accord entre Waze et l'Ille-et-Vilaine, Waze fournit une API ainsi qu'une clé ou bien un jeton d'authentification pour permettre au département d'accéder à leurs données. L'accès sera dans tous les cas restreint selon les termes définis dans l'accord passé entre les deux participants. Dans le cas du département de l'Ille-et-Vilaine, les données accessibles concerteront sans doute uniquement le périmètre du département. Une surcouche de filtres pourra être ensuite appliquée par les différents responsables métiers demandant accès à cette donnée pour ne charger que la partie qui les intéressent. Il peut s'agir de filtres de plages de dates, de zones géographiques particulières ou d'autres paramètres spécifiques selon le niveau de détail requis et des attributs disponibles de la donnée.

Aparté

L'API n'est pas la seule manière d'intégrer des données externes. Tous les fournisseurs de données identifiés dans l'état de l'art ne proposent pas cette possibilité. L'utilisateur doit parfois lui-même télécharger les données (en .csv ou shapefile selon le format de fichier proposé) et les intégrer lui-même dans son SIG. Pour assurer un service continu (comme la possibilité de consulter les données à jour et régulièrement), la DSN du département de l'Ille-et-Vilaine devra s'assurer de réaliser régulièrement ces téléchargements de données depuis les différents sites des fournisseurs de données avec lesquels ils auront contractualisé, et les retravailler dans un système d'extraction, transformation et chargement de la donnée (un ETL).

2. Edition, transformation et chargement des données dans le SGBD grâce à un ETL

Une fois les données chargées dans le SGBD de l'Ille-et-Vilaine, c'est-à-dire les lignes (entités) et les colonnes (variables), il est possible que des traitements soient mis en place sur celles-ci avant qu'elles ne parviennent aux utilisateurs finaux à travers un logiciel de cartographie.

Il n'est pas rare que les formats de données ne soient pas compatibles avec les paramètres et réglages d'un logiciel : comme le format de la date (jour/mois/année), de l'heure, le symbole qui sert de séparateur de virgule, certains caractères spéciaux, la présence de séparateurs de milliers, ou encore le format du fichier (.geojson, .kml, .shp), etc.

Dans notre exemple, préparer la donnée à l'avance permet d'éviter le maintien de la donnée brute telle que Waze la fournit et de devoir la modifier à chaque utilisation pour qu'elle s'adapte au logiciel de cartographie utilisé. La direction des services numériques du département mettra alors en place différentes règles via un ETL afin de nettoyer les jeux de données, calculer des indicateurs ou encore créer de nouvelles variables à partir d'une ou plusieurs autres variables (en extrayant des chaînes de caractères ou en les concaténant, par exemple). Grâce à ce système, la préparation des données sera complètement automatisée.

Toutes ces modifications de données sont transformables via un logiciel de type ETL (sigle anglais pour *Extract, Transform and Load*, soit Extraire, Transformer et Charger). Au département de l'Ille-et-Vilaine, c'est le logiciel FME qui endosse ce rôle. De très nombreux formats de données sont pris en charge par cet outil qui est directement relié aux bases de données du système d'information géographique du département. FME va chercher en entrée les données des fournisseurs, les nettoie, les transforme, les filtre, puis les réinjecte en sortie dans une nouvelle base de données. Ce sont ces données qui seront envoyées vers les utilisateurs finaux lorsque ces derniers les requêteront.

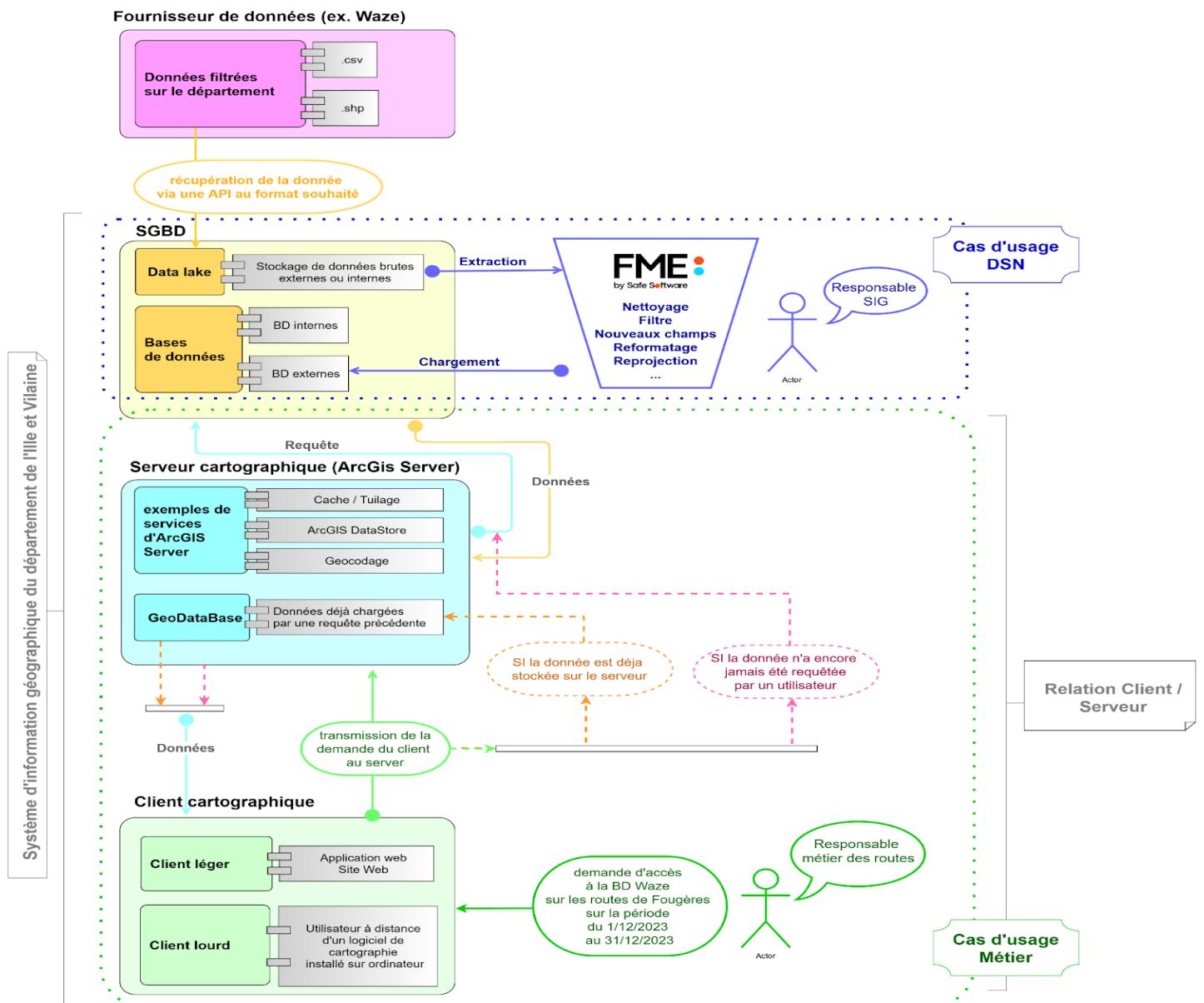
3. Exemple d'intégration de données externe dans une architecture SIG et d'un parcours client

Pour éviter les redondances d'explications et afin de permettre une bonne compréhension de l'architecture SIG ainsi que des interactions entre chaque bloc, nous recourons ici à un diagramme.

Il faut toutefois lire ce diagramme avec du recul. Il faut noter qu'un système d'information géographique peut être doté de plusieurs serveurs cartographiques et de plusieurs clients cartographiques. Nous proposons ici une version simplifiée pour une meilleure compréhension avec un seul serveur et un seul cas d'usage client (bien qu'il puisse prendre la forme "lourde" ou "légère").

Notons aussi que la communication entre un SGBD, un serveur et un client n'est pas aussi directe que le diagramme peut laisser croire. Chaque bloc utilise son langage : par exemple un SGBD utilise le SQL, un serveur peut utiliser java, php ou d'autres langages et un client peut utiliser du html par exemple s'il s'agit d'un client web. Ainsi, le schéma ci-dessous ne détaille ni les langages dans lesquels communiquent les différents blocs, ni des outils complémentaires qui pourraient exister dans cette architecture permettant par exemple la traduction de requêtes avec un protocole HTTP. Ces informations sont volontairement omises. D'abord parce que cela rendrait le schéma illisible sans nécessairement apporter une compréhension plus claire d'un cas d'usage métier (qui est notre objectif ici). Mais aussi parce que les possibilités sont très nombreuses.

Exemple d'une demande de consultation de données ici de la base de données Waze



3. Estimation du poids de la donnée

Comme expliqué précédemment, le SIG du département est en capacité de récupérer des données via une API ou bien par téléchargement direct de fichiers de données. Les offres de services sélectionnés dans le cadre de l'état de l'art proposant l'une ou l'autre de ces méthodes quant à l'accès à leurs données, ces dernières sont donc toutes théoriquement intégrables dans le SIG du département. Si la question du format n'est pas un frein à la capacité du SIG à intégrer ces données selon notre expertise, la question du volume se pose.

Nous ne proposons ici, non pas une analyse de la "solidité" du système informatique du département quant à l'intégration de données parfois issues de Big Data (cela sort de notre champ de compétences), mais bien une estimation des volumes de données que représentent les offres identifiées. Malgré nos sollicitations, très peu de fournisseurs de données ont pu nous partager des extraits de données sur un IRIS et nous permettre ainsi de réaliser des estimations.

Cette estimation peut donner une première idée :

- de l'espace de stockage à allouer sur les bases de données du SIG
- de la consommation de bande passante que pourrait représenter l'intégration de telles données

Nous présentons, dans le tableau ci-après, l'ensemble des estimations de volume que nous avons pu réaliser :

Solutions	Producteur	Format	Échelle d'un IRIS	Estimation Ille-et-Vilaine
Flux vision	Orange Business	.csv	environ 1,125 Mo pour une journée ⁵	328,7 Go (ensemble de l'année) ⁶
Outdoor Vision	PRNSN	.geojson	environ 288 Ko ⁷ (janvier 2018 à août 2023)	1,2 Go (janvier 2018 à août 2023)
Waze for Cities	Waze	.csv	40 Ko pour un an de données	20 Mo pour un an de données

⁵ Moyenne de la taille des fichiers pour 4 Iris en zone rural (Roz-sur-Couesnon, Cancale, Saint-Juste et Mezieres-sur-Couesnon, du 2 au 5 octobre 2023).

⁶ Le volume présenté ici ne correspond pas à la multiplication de la valeur adjacente par le nombre de jour dans une année. En raison de la variabilité du nombre de données entre zone rurale et zone urbaine, et sous conseil de notre interlocuteur chez Orange, nous avons augmenté le multiplicateur : 1,125 Mo par 800.

⁷ poids moyen par iris en considérant que le nombre de traces est similaire dans chaque Iris

Le choix d'une offre ne doit cependant pas uniquement se baser sur cette estimation. En effet, la quantité de données réellement nécessaire dépend avant tout des usages prévus. Ainsi, il est par exemple probable que la donnée flux vision soit utile pour des études ponctuels dans le temps et l'espace, et qu'il ne soit donc pas nécessaire de télécharger l'ensemble des 328,7 Go de donnée, qui correspondent au poids d'une année pour toute l'emprise de l'Ille-et-Vilaine avec une granularité temporelle de 30 minutes.

Conclusion

1. Rétrospective de l'atelier par rapport à la commande

L'accès à l'information

Notre étude a permis de répondre à la problématique qui est de participer au travail de veille technologique et méthodologique pour répondre aux besoins numériques du département, menée par la mission innovation et développement du service ingénierie applicative. La commande initiale s'est déclinée en trois étapes, la première aborde la réalisation d'un état de l'art de données existantes.

Au cours de cette étape, différents niveaux de difficulté se sont manifestés. La première porte sur la prise de contact avec les différents fournisseurs de données. Une fois avoir écrémé les différentes offres de services disponibles sur le marché des données massives de flux, nous avons répertorié 10 fournisseurs à contacter. Seulement 5 (Waze, Flux Vision, Outdoor Vision, GeoVelo, Strava) ont répondu à nos questions sur des délais de réponse très variés. A ce jour, seul Flux Vision a pu nous fournir un échantillon de données pour 4 IRIS de l'Ille et Vilaine. Ensuite, un travail de veille et de recherche a été réparti pour chaque étudiant afin de renseigner et documenter sous forme de fiche les informations sur un producteur de données. Ce travail mené, a été dépendant de la disponibilité et l'accès à la documentation d'une source de données (métadonnées, détail de l'offre, etc). Certains fournisseurs partagent peu d'informations ou le moins possible sur les caractéristiques techniques de leurs produits. Par exemple, Orange Business Service n'offre pas la possibilité de visualiser un échantillon de données pour comprendre la structure attributaire et spatiale du produit. Mais également, aucune métadonnée n'y est divulguée. Donc il est difficile de se projeter sur un potentiel usage sans avoir plus d'informations.

changement de méthode de recherche

La deuxième étape visait à trouver des cas d'usages de données de flux de personnes initiés par d'autres collectivités qui répondent à des problématiques territoriales. L'idée est de pouvoir s'inspirer de ce qui a déjà été fait pour pouvoir se projeter sur de potentiels cas d'usages internes des services métiers du département. C'est pourquoi nous avons commencé par orienter nos recherches sur des cas d'usages d'autres départements

français utilisant des données de flux pour analyser la fréquentation humaine d'un espace. A ce stade, nous nous sommes rendu compte que la cible de recherche était trop restreinte. De ce fait, nous nous sommes étendus aux collectivités territoriales en général, de là, nous avons agrandi le champ d'opportunité par d'autres cas d'usages telles que les métropoles/communauté de commune et la société Waze pour une meilleure gestion du trafic routier.

Ce travail a été initié pour identifier les opportunités que l'usage de certains produits peut apporter à la connaissance, l'étude et l'aide à la décision des thématiques inhérentes aux compétences du département (gestion de route, gestion des E.N.S, accueil et service sociaux etc). Mais le bilan tiré de ce travail est que nous ne pouvons pas nous appuyer uniquement sur ces quelques cas d'utilisations (mentionnées dans la partie 2) de par leur faible nombre, mais surtout sans connaître les réels besoins des services métiers interne du département.

L'ouverture des perspectives : se rapprocher des services métiers

C'est pourquoi il est devenu essentiel d'initier la collecte des besoins spécifiques des services métiers internes. Cette démarche ciblée vise à aligner précisément les cas d'usage identifiés avec les sources de données répertoriées, offrant ainsi une vision exhaustive et adaptée à notre mission. L'organisation de réunions et d'ateliers "world café" a été particulièrement enrichissante, facilitant la discussion autour des besoins en données de flux et contribuant à l'identification des cas d'usages potentiels. Cette démarche a abouti à la mise en lumière de thématiques majeures telles que la gestion des routes départementales, la préservation des espaces naturels sensibles et l'action sociale. À cela s'est ajoutée la découverte d'une thématique additionnelle, à savoir la gestion de l'accueil des usagers dans les établissements départementaux, élargissant ainsi la palette des opportunités à explorer malgré le constat d'un marché relativement restreint des données massives en France.

Cependant faire correspondre les offres de données disponibles sur le marché avec les cas d'usage identifiés n'est pas une mince à faire. D'abord car notre changement de méthode d'identification de cas d'usage pour le département à retarder l'éclosion de potentielles opportunités. Parce que la récolte des besoins métier s'est produite tardivement (à la fin de l'atelier). Mais aussi par le peu d'échantillons de données disponibles à ce jour (seulement un accès à un échantillon de 4 IRIS sur le département). Il est donc difficile aujourd'hui d'apporter une solution pour répondre à ces cas d'usages avec les offres de service que nous avons répertoriées.

Il reste encore des questions en suspens pour mettre en adéquation la donnée et la problématique métier. Comme la granularité spatiale qui dépend souvent du mode d'acquisition de la donnée (réseaux mobiles ou GPS), l'un modélisé en semis de point ou

réseaux linéaire, l'autre en découpage spatial (ex : IRIS pour Orange). Mais aussi de la résolution temporelle, la fréquence de diffusion de la donnée et de son volume. Ainsi ces multitudes de possibilités soulignent l'importance d'adapter nos sources de données en fonction des besoins précis de chaque cas d'usage pour garantir une analyse optimale et pertinente. Et donc de procéder une étude encore plus approfondie des besoins métiers pour déterminer quelles offres serait la plus adéquate.

Le SIG du département étant, d'après nos critères, en capacité d'intégrer les données identifiées dans l'état de l'art, c'est bien sur l'analyse de la correspondance entre les besoins au département et les caractéristiques des données identifiées qu'il faut se focaliser à l'avenir.

2. Retour sur la gestion de projet

Dans le cadre de l'atelier, nous avons mis en œuvre une méthodologie qui a été amenée à évoluer en fonction de nos recherches. Les différents points représentent les éléments qui ont été nécessaires pour mener à bien le projet et répondre aux attentes fixés par les commanditaires.

Appropriation du sujet et recherche bibliographique

Le projet a débuté par la phase d'appropriation du sujet, au cours de laquelle nous avons pu délimiter le champ d'étude. Cette étape a permis une compréhension des enjeux, des problématiques centrales, et a été renforcée par des recherches bibliographiques. Cela nous a permis d'étoffer nos connaissances sur le sujet dans le but de distinguer plusieurs grandes thématiques à étudier.

Répartition des groupes sur différentes thématiques

Nous nous sommes donc répartis le travail autour de ces thématiques. Cette démarche a permis de maximiser l'efficacité de nos recherches. Chaque membre du groupe s'est concentré sur différents aspects permettant une analyse approfondie du sujet. Tels que la production de données, les offres de services, le recueil d'informations sur le département et une recherche sur les serveurs et logiciels utilisés par le département.

Élaboration de la note de cadrage

Le but de ces recherches était de les mettre en commun en vue de la rédaction de la note de cadrage. Nous avons donc croisé nos différentes connaissances et intégré les attentes du commanditaire pour la réaliser. Ce document a permis de fixer les missions à effectuer pour le temps restant.

Concentration sur le contenu des rendus de l'atelier

Une fois la note de cadrage validée, notre attention s'est portée sur le contenu de l'atelier. La répartition du travail s'est effectuée par sous-groupes de 2 ou 3 personnes, chaque groupe a approfondi les thématiques sur lesquelles il était investi.

Analyse approfondie des offres de services

Lorsque les informations nécessaires étaient collectées, l'ensemble du groupe s'est attaché à des recherches autour des offres de services. Cette tâche a constitué une partie importante du temps fixé à l'atelier. Les recherches, les entretiens avec des personnes en lien avec certaines données et les demandes d'accès à des échantillons de données ont été des éléments clés dans l'enrichissement de cette thématique. Les modalités de ces recherches ont pu varier en fonction de la donnée notamment dans la prise de contact (mail, formulaire préétablie à renseigner ou encore contact en lien avec des personnes travaillant sur la donnée ou spécialiste de la donnée en question).

Communication régulière avec le commanditaire

Afin de mener au mieux notre atelier, il a été nécessaire d'entretenir un échange régulier (généralement hebdomadaires) avec le commanditaire en dehors de nos réunions au sein du groupe. Ces réunions ont permis de répondre aux attentes du commanditaire, de résoudre de potentiels points de blocage et de parfois ajuster notre démarche, cela a assuré une progression cohérente durant l'atelier. Parmi les points de blocage on retrouve notamment les prises de contact auprès de personnes ou d'organismes en lien avec la donnée. Les modalités de contact afin d'obtenir des échantillons ou des informations plus précises ont parfois été des freins à l'avancement de notre travail, par exemple Strava ne dispose pas d'adresse mail pour échanger autour de leur donnée.

Utilisation d'outils de gestion

Pour une organisation fluide, deux outils ont été privilégiés : Zotero, pour la conservation et la mise en forme de la bibliographie consultée, et le diagramme de Gantt, permettant de planifier et de visualiser le temps attribué à chaque tâche prédefinie. Ces outils ont facilité la coordination au sein du groupe et assuré une bonne gestion du temps.

Dans le cadre de cet atelier nous avons réalisé trois rendus comprenant une première présentation orale devant le Conseil Départemental d'Ille-et-Vilaine. Cet échange visait à présenter notre travail aux différents métiers concernés et à susciter de potentiels idées de cas d'usage. Ainsi qu'une seconde présentation orale public, accompagnée d'un rapport écrit.

En résumé, cette méthodologie, de l'appropriation du sujet à la présentation finale, a été nécessaire pour la gestion de notre atelier.

Cependant certaines limites se sont présentées à nous. Parmi elles le nombre de personnes au sein du groupe (10) a parfois compromis les échanges, l'évolution des attentes du commanditaire face aux résultats présentés...

Références

- ANFR (2023) 'Données sur les réseaux mobiles'. Available at: https://data.anfr.fr/visualisation/information/?id=observatoire_2g_3g_4g&location=11.48.13448,-1.50101 (Accessed: 15 December 2023).
- ARCEP (2012) *Les services de communications électroniques en france. Année 2010 – Résultats définitifs*.
- ARCEP (2023) *Les services de communications électroniques en france. 2eme trimestre 2023*.
- Bachir, D. et al. (2019) 'Inferring dynamic origin-destination flows by transport mode using mobile phone data', *Transportation Research Part C: Emerging Technologies*, 101, pp. 254–275. Available at: <https://doi.org/10.1016/j.trc.2019.02.013>.
- Barbosa, H. et al. (2018) 'Human mobility: Models and applications', *Physics Reports*, 734, pp. 1–74. Available at: <https://doi.org/10.1016/j.physrep.2018.01.001>.
- Blondel, V.D., Decuyper, A. and Krings, G. (2015) 'A survey of results on mobile phone datasets analysis', *EPJ Data Science*, 4(1), pp. 1–55. Available at: <https://doi.org/10.1140/epjds/s13688-015-0046-0>.
- Calabrese, F., Ferrari, L. and Blondel, V. (2014) 'Urban Sensing Using Mobile Phone Network Data: A Survey of Research', *ACM Computing Surveys*, 47, pp. 1–20. Available at: <https://doi.org/10.1145/2655691>.
- Cerema (2022) 'Collecte et utilisation de données de mobilité pour la modélisation des déplacements : Des enquêtes ménages-déplacements aux données massives'. Cerema (Données de mobilité pour la modélisation des déplacements). Available at: <https://doc.cerema.fr/Default/doc/SYRACUSE/593053/collecte-et-utilisation-de-donnees-de-mobilite-pour-la-modelisation-des-deplacements-des-enquetes-me> (Accessed: 8 November 2023).

Cerema, l'environnement (2023) ‘Données issues des antennes de la téléphonie mobile : Données de mobilité pour la modélisation des déplacements’. Cerema (Données de mobilité pour la modélisation des déplacements). Available at: <https://doc.cerema.fr/Default/doc/SYRACUSE/594129/donnees-issues-des-antennes-de-la-telephonie-mobile-donnees-de-mobilite-pour-la-modelisation-des-dep> (Accessed: 8 November 2023).

DATALAB (2023) *Chiffres clés des transports - Édition 2023*. Ministère de la transition écologique et de la cohésion des territoires.

Doyle, J. et al. (2011) ‘Utilising Mobile Phone Billing Records for Travel Made Discovery’, in. Available at: <https://www.semanticscholar.org/paper/Utilising-Mobile-Phone-Billing-Records-for-Travel-Doyle-Hung/f7f092c1944da437b3934341c9d5bae57c47be58> (Accessed: 16 December 2023).

Fekih, M. et al. (2021) ‘A data-driven approach of origin-destination matrix construction from cellular network signalling data: a case study of Lyon region (France)’, *Transportation*, 48. Available at: <https://doi.org/10.1007/s11116-020-10108-w>.

Flux Vision : l'alliance unique des données GPS et opérateur (2022). Available at: <https://www.youtube.com/watch?v=Y0OooJtNYsl> (Accessed: 20 November 2023).

García, P. et al. (2016) ‘Big Data Analytics for a Passenger-Centric Air Traffic Management System A Case Study of Door-to-Door Intermodal Passenger Journey Inferred from Mobile Phone Data’, in. Available at: <https://www.semanticscholar.org/paper/Big-Data-Analytics-for-a-Passenger-Centric-Air-A-of-Garc%C3%ADa-Herranz/d84ba04382f0d6dfbba068b8113242e47c82916b> (Accessed: 16 December 2023).

Huang, H., Cheng, Y. and Weibel, R. (2019) ‘Transport mode detection based on mobile phone network data: A systematic review’, *Transportation Research Part C: Emerging Technologies*, 101, pp. 297–312. Available at: <https://doi.org/10.1016/j.trc.2019.02.008>.

INSEE (2023) *Guide du secret statistique*.

Jiaqi, K.C. (2018) *Transport Mode Detection Using Cellular Signaling Data (Case Study of Graz and Vienna, Austria)*. Geographisches Institut der Universität Zürich.

Nabavi Larijani, A. et al. (2014) ‘Investigating the Mobile Phone Data to Estimate the Origin Destination Flow and Analysis; Case Study: Paris Region’, in. *Transportation Research Procedia*. Available at: <https://doi.org/10.1016/j.trpro.2015.03.006>.

Okmi, M. et al. (2023) 'Mobile Phone Data: A Survey of Techniques, Features, and Applications', *Sensors*, 23(2), p. 908. Available at: <https://doi.org/10.3390/s23020908>.

Pei, L. et al. (2013) 'Human Behavior Cognition Using Smartphone Sensors', *Sensors*, 13(2), pp. 1402–1424. Available at: <https://doi.org/10.3390/s130201402>.

Philip Chen, C.L. and Zhang, C.-Y. (2014) 'Data-intensive applications, challenges, techniques and technologies: A survey on Big Data', *Information Sciences*, 275, pp. 314–347. Available at: <https://doi.org/10.1016/j.ins.2014.01.015>.

Service des données et études statistiques (2021) *Enquête Mobilité des Personnes 2019*. Ministère de la transition écologique et de la cohésion des territoires.

Torre-Bastida, A.I. et al. (2018) 'Big Data for transportation and mobility: recent advances, trends and challenges', *IET Intelligent Transport Systems*, 12(8), pp. 742–755. Available at: <https://doi.org/10.1049/iet-its.2018.5188>.

Wang, Z., He, S.Y. and Leung, Y. (2018) 'Applying mobile phone data to travel behaviour research: A literature review', *Travel Behaviour and Society*, 11, pp. 141–155. Available at: <https://doi.org/10.1016/j.tbs.2017.02.005>.

Yaqoob, I. et al. (2016) 'Big data: From beginning to future', *International Journal of Information Management*, 36(6, Part B), pp. 1231–1247. Available at: <https://doi.org/10.1016/j.ijinfomgt.2016.07.009>.

Annexes - Fiches fournisseurs de données

Geovelo



Nom de la donnée : Geovelo

Geovelo est une application mobile proposant aux utilisateurs des itinéraires sécurisés pour les vélos selon différents critères (touristique, sécurisé, le plus direct et “équilibré”, c'est-à-dire un mélange des 3 premiers critères).

Geovelo est un très bon calculateur d'itinéraire vélo, puisqu'il contribue largement à cartographier les aménagements pour les collectivités sur OpenStreetMap et dispose donc d'une très bonne connaissance réseau. Ainsi, certains itinéraires proposés ne sont pas forcément retrouvés sur d'autres applications plus connues comme Google maps.

Les utilisateurs sont libres de partager leur itinéraire avec géovélo. De leur côté, ils peuvent consulter leurs statistiques (distance parcourue, vitesse, etc) : “L'application récolte et anonymise les données des cyclistes lors de leur parcours afin d'aider les villes à développer et améliorer leurs réseaux cyclables”.

Méthodes de production de la donnée :

3 méthodes :

- les parcours enregistrés avec le mode navigation, avec la position GPS
- les parcours enregistrés manuellement (la vitesse n'est pas enregistrée), dans ce cas-ci on ne peut pas savoir si l'itinéraire est exactement le même que celui proposé par l'application ou bien s'il y a eu des détours.
- les itinéraires enregistrés grâce à la détection automatique d'activités, si l'utilisateur a bien activé ce réglage dans son application.

Les services proposés par Géovélo auprès des décideurs publics :

Les données partagées par les utilisateurs de géovélo et proposées aux collectivités territoriales par Géovélo permettent d'analyser la pratique du vélo sur un territoire.

Voici quelques exemple d'informations et d'indicateurs que peuvent apporter les données

- segments les plus fréquentés
- segments les moins fréquentés
- zones à potentiel accidentogène
- meilleure compréhension de l'état du réseau cyclable
- indicateur temporel (pic et creux de la journée) : la donnée n'est plus communiquée à l'heure mais bien par tranches horaires de pic et de creux de fréquentation
- jour de la semaine
- l'accidentalité (zones de “points noirs” où les accidents sont fréquents et où les aménagements sont jugés peu sécurisants)
- analyse de discontinuité
- analyse de zone de désir de stationnement

Transformation en besoin métier : aide à la décision relative au développement, à la planification du réseau cyclable

En partenariat avec Eco-Compteur, Geovelo propose aux professionnels et notamment aux territoires l'accès à une plateforme baptisée **Cycling Insights** qui met en avant divers indicateurs aux travers de tableaux de bords et cartes, basés sur la donnée produite par les utilisateurs de l'application.

Aujourd'hui leur **API** ne sert que pour intégrer leurs calculateurs d'itinéraires dans d'autres applications ou d'autres plateformes. C'est ce qu'a fait Ile-de-France Mobilités par exemple.

Geovelo propose également le partage de **données brutes**, mais l'échelle la plus fine dans ce cas-là est l'IRIS. Les formats disponibles sont très nombreux : table csv, shapefile, geojson...

À noter qu'il faut être client de la solution Cycling Insights. Le téléchargement de données brutes passe directement par cette plateforme.

Compte-tenu d'une demande croissante, Geovelo travaille actuellement sur le développement d'un web service de type **WFS** pour permettre à ses clients de charger la donnée dans leur SIG respectif et de croiser la donnée Géovélo avec leurs bases de données clients. Cette solution devrait voir le jour à la fin de l'année 2024.

Estimation du poids de la données:

N'ayant pas eu accès à la plateforme, l'estimation du poids de la donnée ne peut s'effectuer que sur des données opendata. Les tracés des pistes cyclables diffusées via OSM peuvent donner une idée du poids des données mais sans dimensions temporelles. Mais ce genre de données est plutôt léger (quelques centaines de kilooctets) et n'est pas du tout représentatif de jeux de données produites par des utilisateurs quotidiens d'une application mobile. Il est donc plus difficile d'estimer le poids de la donnée produite par les utilisateurs de Geovelo sur une semaine avec une fréquence quotidienne par exemple.

Limites

Il ne sera pas possible de retracer des itinéraires complets avec le point de départ et le point d'arrivée. Cela permet de renforcer l'anonymisation des données et d'éviter toute possibilité de retracer par qui a été emprunté tel ou tel itinéraire. Ceci est mis en place dans un souci de protection de la vie privée des utilisateurs et de respect du règlement général de la protection des données (RGPD).

Sources :

<https://geoveloo.app/fr/blog/comment-les-donnees-geoveloo-aident-les-territoires/>
<https://geoveloo.app/fr/territories/>
<https://geoveloo.app/fr/api/>
<https://www.francemobilites.fr/solutions/geoveloo-territoires>

Outdoor Vision



Nom de la donnée : Outdoor Vision

Nom du fournisseur : Pôle ressources national sports de nature, mission d'appui du ministère chargé des Sports

URL : <https://outdoorvision.fr/>

Catégorie du fournisseur : Agrégateur de données issus de divers producteurs (principalement : Garmin, Suunto, Polar, IGNrando, VisoRando)

et Mode de partage : Plateforme de visualisation en ligne permettant, entre autre, de :

- Visualiser les flux de pratiquants selon des périodes
- Compter les passages de pratiquants dans une zone géographique ou le long d'itinéraires
- Identifier les principaux points de départs, d'arrivées et zones de pauses
- Visualiser la répartition des activités
- Comparer la fréquentation entre deux périodes

Modalité de livraison des données : plateforme de visualisation; export geojson des traces agrégées depuis la plateforme.

Métriques (granularité et temporalité des données) :

Les données sont des traces GPS agrégées, issues d'au moins 5 traces de minimum 3 contributeurs différents. Un accès direct à des données non agrégées pourrait être envisagé à l'avenir pour des cas spécifiques à discuter avec les clients et avec l'approbation des fournisseurs de données partenaires. Il ne s'agit toutefois pas de données brutes telles que produites par le fournisseurs, puisque ces données ont à ce stade été nettoyées (suppression des valeurs aberrantes,...) et pseudonymisées (de façon à ce qu'il ne soit pas possible pour Outdoor vision de remonter jusqu'à l'utilisateur). Concernant la temporalité, de nouvelles données sont mises sur la plateforme 3 fois par

an environ. L'objectif est, à l'avenir, d'améliorer cette fréquence afin d'avoir une mise à jour saisonale des données.

Estimation du poids de la donnée :

En Ille-et-Vilaine : 149 756 traces et 20 953 pratiquants du 1er janvier 2018 (date des plus anciennes données disponibles) au 20 août 2023, pour un poids d'environ 1,2 Go.

Cas d'usages ciblés :

IGN → enrichissement des référentiels cartographiques

Fédérations sportives → aide à l'organisation de compétitions (adéquation itinéraires empruntés et parcours prévus), détection des zones de pratiques et donc des adhérents potentiels

Collectivités → meilleure compréhension des itinéraires empruntés afin d'améliorer leur accessibilité, leur mise en valeur, protéger la biodiversité, créer des pistes cyclables...

(Bureaux d'étude, gestionnaires de sites naturels ou d'espace naturels protégés sont aussi identifiés comme des utilisateurs cibles)

Limites :

Données nécessairement incomplètes car basées sur les dons de traces GPS par les pratiquants. Il n'y a pas encore de redressement statistique des données, bien que ce soit envisagé à l'avenir. De plus, le profil des contributeurs influence la forme des traces visibles sur la plateforme : environ 60% de pratiquants occasionnels et 40% de praticiens intensifs

Coût :

- Formule découverte gratuite : 5 accès à la plateforme pour de la visualisation uniquement, accès aux webinars/formations du club utilisateurs (plusieurs dates dans l'année) et autre ressources type tutoriels d'utilisation
- Formule expert payante : une vingtaine d'accès, la possibilité d'exporter des données en geojson, la possibilité d'importer des référentiels vecteurs afin de réaliser certains traitements (comme compter le nombre de départs dans une zone). Le tarif est variable selon la quantité de traces disponibles (3 seuils allant jusqu'à 5000 euros par an). Les utilisateurs expert sont "tête de réseau" et s'engagent à promouvoir l'outil auprès des collectivités locales.

Sources :

- site officiel d'Outdoor Vision (<https://outdoorvision.fr/>)
- brochure de présentation (https://outdoorvision.fr/medias/Brochure_presentation_Outdoorvision.pdf)
- entretien avec Christophe Martinez, coordinateur du projet Outdoorvision

Waze for cities



Waze est une application mobile d'assistant d'aide à la conduite et d'assistance de navigation. L'itinéraire calculé peut se mettre à jour en temps réel grâce à certaines informations liées à l'état du trafic. Chaque utilisateur peut mettre à jour les tracés et caractéristiques des routes, noms de lieux et adresses ainsi que signaler des routes fermées (événements, travaux,...) via une application web distincte de l'application mobile.

Méthodes de production de la donnée :

- les utilisateurs partagent leurs données de localisation et signalent des conditions de circulation en temps réel
- Waze collecte des données sur la vitesse du véhicule, les arrêts, les itinéraires empruntés et les signalements d'incidents (accidents, dangers sur la route ou conditions météorologiques)
- les données sont collectées et traitées en temps réel par les serveurs de Waze qui utilisent des algorithmes pour agréger et analyser les informations provenant de plusieurs utilisateurs
- Les informations traitées sont ensuite utilisées pour mettre à jour les cartes en temps réel
- Waze encourage les utilisateurs à fournir des commentaires sur la qualité des informations fournies. Les retours des utilisateurs contribuent à améliorer la précision et la fiabilité des données.

Les services proposés par Waze auprès de décideurs publics :

- gestion du trafic : accès à des données en temps réel sur le trafic routier
- communication en temps réel avec les utilisateurs de l'application
- accès à des tableaux de bord et des outils analytiques spécifiques

Transformation en besoin métier :

L'aide à la décision pour améliorer la mobilité urbaine est cruciale pour des déplacements urbains efficaces. Elle intègre l'analyse de données, l'utilisation de technologies émergentes, et la consultation des parties prenantes. En combinant ces éléments, elle permet de prendre des décisions informées, favorisant la mise en place d'initiatives telles que des systèmes de transport plus efficaces, des solutions de mobilité durable, et des infrastructures adaptées. Cette approche contribue à créer des environnements urbains plus durables et agréables.

Mode de partage :

Les données en temps réel partagées par Waze avec les départements et les autorités locales sont généralement fournies via une API. Les données sont souvent fournies au format JSON, un langage de balisage léger utilisé pour structurer et échanger des informations entre les applications. Cette approche basée sur une architecture d'interface de programmation permet une communication fluide et instantanée entre Waze et les entités départementale, régionale et gouvernementale. Facilitant ainsi la mise à jour en temps réel des informations liées à la circulation, aux conditions routières et aux événements susceptibles d'impact sur la mobilité. L'utilisation du format JSON offre une flexibilité accrue dans le traitement et l'interprétation des données, permettant aux autorités locales de tirer pleinement partie des informations transmises par Waze.

Estimation du poids de la donnée :

Le poids des données de Waze varie généralement de quelques kilooctets à quelques dizaines de kilooctets. Cette variation dépend de facteurs tels que la densité du trafic et le nombre d'utilisateurs actifs. Les données, généralement transmises au format JSON, sont légères, favorisant une communication rapide entre Waze et les autorités locales. Cette légèreté facilite la mise à jour en temps réel des informations sur la circulation et contribue à une gestion efficace du trafic urbain.

Limites :

La qualité et la précision des données fournies par Waze dépendent du nombre d'utilisateurs actifs dans une région donnée. Dans les zones moins fréquentées comme les ENS, les données peuvent être moins fiables. De plus, les utilisateurs de Waze ne représentent pas nécessairement toute la population, ce qui peut biaiser la représentation du trafic.

Coût : gratuit en échange de données voiries

Sources :

- [Waze for Cities — Wazeopedia](#)
- [En Mayenne. L'application de navigation routière Waze s'associe avec le Département \(ouest-france.fr\)](#)

Google Maps Platform



Google Maps Platform

Nom du fournisseur : Google

URL : <https://mapsplatform.google.com/>

Catégorie du fournisseur : Producteur, agrégateur et revendeur

Mode de partage : Données brutes et par plateforme

Modalité de livraison de la donnée et des solutions :

API

SDK (Kit de développement logiciel). C'est un ensemble d'outils fourni avec une plateforme matérielle (généralement), un système d'exploitation ou un langage de programmation. Essentiellement de la mise en place de solutions mobiles.

Cas d'usage potentiel :

Évaluer les heures d'affluence, le degré de fréquentation d'un lieu grâce aux données de géolocalisation de Google.

- Conservation du patrimoine
- Evaluation de la fréquentation des ENS

Google Maps Geolocation API : Ce service permet de déterminer la position géographique d'un appareil en fonction des signaux des antennes de téléphonie mobile, des points d'accès Wi-Fi et des capteurs d'un appareil. Cela peut être utile pour obtenir la position d'un utilisateur même s'il ne dispose pas de la fonctionnalité GPS activée.

Google Maps JavaScript API : Cette API permet d'obtenir la position géographique de l'utilisateur à partir du navigateur web. Cela nécessite que l'utilisateur donne son consentement pour partager sa position.

Estimation du poids de la donnée : Pas d'informations

Temporalité de la donnée : Temps réel

Coût : Adapté aux besoins, selon le nombre de requêtes et l'emprise du territoire.

Limites :

La grande quantité de données peut être une problématique. Effectivement, la complexité des données ne permet pas toute de suite de cibler un cas d'usage. Nous n'avons également pas d'informations sur la production de certaines données.

Source :

API Geolocation

[https://developers.google.com/maps/documentation/geolocation/overview?
_gl=1*c97rdz*_ga*MTc2Mjg2NDY4OS4xNjk0NTI3OTk4*_ga_NRWSTWS78N*MTcwMDU3Mzc3NC4yLjEuMTcwMDU3NzlwNC4wLjAuMA..&hl=fr](https://developers.google.com/maps/documentation/geolocation/overview?_gl=1*c97rdz*_ga*MTc2Mjg2NDY4OS4xNjk0NTI3OTk4*_ga_NRWSTWS78N*MTcwMDU3Mzc3NC4yLjEuMTcwMDU3NzlwNC4wLjAuMA..&hl=fr)

API Maps JS

[https://developers.google.com/maps/documentation/javascript?
_gl=1*193cd0x*_ga*MTc2Mjg2NDY4OS4xNjk0NTI3OTk4*_ga_NRWSTWS78N*MTcwMDU3Mzc3NC4yLjEuMTcwMDU3NzU1MS4wLjAuMA..&hl=fr](https://developers.google.com/maps/documentation/javascript?_gl=1*193cd0x*_ga*MTc2Mjg2NDY4OS4xNjk0NTI3OTk4*_ga_NRWSTWS78N*MTcwMDU3Mzc3NC4yLjEuMTcwMDU3NzU1MS4wLjAuMA..&hl=fr)

Flux Vision

Nom du fournisseur : Orange

URL : <https://www.orange-business.com/fr/solutions/data-intelligence-iot/flux-vision>

Catégorie du fournisseur : producteur et revendeur

Mode de partage : Données brutes et visualisation

Modalité de livraison des données :

- Données brute : CSV
- Visualisation : PBI
- Les deux données sont mises à disposition sur le cloud d'orange

Métriques :

- Résolution spatiale :
 - IRIS
 - Les nouvelles innovations permettent d'affiner la résolution dans les zones où le réseau est suffisamment dense (via intégration de données GPS)
- Résolution temporelle :
 - Au minimum, toutes les 10 minutes

Dans la mesure où les offres sont "à la carte", l'ensemble des métriques peuvent évoluer en fonction de la zone étudiée (rural / urbain).

Estimation du poids de la donnée :

- cf. tableau partie 3

Cas d'usages effectifs (usages publics) :

- Tourisme
 - Les observatoires des départements étudient la fréquentation touristique grâce aux données flux vision ciblé sur le profil touristique des individus.

Exemple de cas d'usages ciblés :

- Social :
 - Panneau d'information pour influer sur le non-recours au droit (offre retail)
- Voirie :
 - Mieux appréhender les flux de population pour gérer les réseaux de route départemental (offre transport)

Cas d'usages envisagés pour notre atelier :

- croisement du flux vision avec les data interne (capteur, etc)

Limites :

- Données issues uniquement des utilisateurs possédant une carte SIM Orange (soit entre 30 à 40 % des utilisateurs de cartes SIM). Les données sont ensuite redressées à la population totale.
- La résolution spatiale (IRIS), ne permet pas une analyse fine des mobilités en zone rurale.

Coût :

- Le prix varie en fonction de l'offre choisie,

Sources :

Chambreuil, P., Jeon, J.Y. and Barba, T. (2022) 'The value of network data confirmed by the Covid-19 epidemic and its expanded usages', *Data & Policy*, 4, p. e4. Available at: <https://doi.org/10.1017/dap.2021.31>.

Orange, (2022) *Flux Vision : l'alliance unique des données GPS et opérateur* (2022).

Available at: <https://www.youtube.com/watch?v=Y0OooJtNYsI> (Accessed: 20 November 2023).

Orange (2023) *Flux Vision* | Orange Business. Available at: <https://www.orange-business.com/fr/solutions/data-intelligence-iot/flux-vision> (Accessed: 20 November 2023).

Schut, P.-O. and Beaudouin, S. (2021) ‘L’usage des données issues des réseaux de téléphonie mobile. Étude de cas à Roland-Garros’, *Réseaux*, 230(6), pp. 249–272.

Available at: <https://doi.org/10.3917/res.230.0249>.

Strava



Nom du fournisseur : Strava

URL : <https://metro.strava.com/>
<https://developers.strava.com/>
<https://www.strava.com/heatmap#8.97/-1.82843/48.16797/hot/all>

Catégorie du fournisseur :

Producteur de la donnée :

- La production de données est issue des activités diverses des utilisateurs de l’application (course à pied, vélo, marche, etc).
- La donnée est ensuite partagée par Strava et peut être mise à disposition.

Agrégateur :

- Dans le cas d’une plateforme comme Strava, l’agrégateur est l’entreprise en elle-même. Strava collecte les données des utilisateurs en accord avec les règles de confidentialité.
- Certains indicateurs présent dans la donnée sont créée par Strava (classement, carte de chaleur, utilisation d’un segment, etc)
- Lors de l’agrégation et du partage par Strava les données sont anonymisé

Mode de partage :

- Il existe plusieurs sources de partage autour des données Strava :
- Strava Métro : fournit des données agrégées pour le traitement de problématiques liées au transport ou à la fréquentation de certains lieux. Possiblement ouvert

gratuitement pour les projets de recherche et pour les organisme public (demande en cours en attente de réponse)

- Strava Global Heatmap : carte interactive permettant de visualiser, les itinéraires fréquemment utilisés par les utilisateurs. L'information sur le type d'activité pratiqué est également partagée. <https://www.strava.com/heatmap#8.97-1.82843/48.16797/hot/all>
- API : possibilité d'accès aux données sous forme d'abonnement <https://developers.strava.com/playground/#/Segments/getSegmentById> <https://www.strava.com/legal/api>

Modalité de livraison des données :

Strava Métro :

- Nécessite de faire une demande sur leur site pour avoir accès aux données de manière gratuite.
- Le format de données fournies varient en fonction de la demande formulée (probablement GeoJson, CSV, KML) .

API Strava :

- Nécessite de s'inscrire en tant développeurs sur le site web, puis de créer une application sur la plateforme développeurs de Strava pour obtenir une clé API
- Concernant l'authentification, le protocole OAuth est utilisé. Cela signifie que lorsqu'un utilisateur autorise une application à accéder à ces données Stava. L'utilisateur est redirigé vers le site WEB de Strava pour autoriser l'accès et ainsi recevoir un jeton permettant de réaliser les requêtes
- Récupération des données en faisant des requêtes https aux endpoints
- Les données renvoyés par l'API sont au format JSON

Granularité et temporalité des données

API Strava :

- La granularité des données disponible à partir de l'API Strava dépend des endpoints utilisés. Par exemple pour les données d'activités, la granularité sera très fine car on peut obtenir des points précisément placés. De la manière pour les données utilisateurs, les détails de données varient en fonction des autorisations fixées.
- La temporalité des données à récupérer dépend des endpoints utilisés et des requêtes formulées. Pour les données d'activités, il est possible de récupérer la date et l'heure du début et de la fin de l'activité.

Strava Métro :

- La granularité des données dépendra également du partenariat passé avec Strava. Pour la problématique du département, il pourrait être intéressant d'obtenir une granularité à l'échelle la plus fine possible (itinéraire, route, etc)

- La temporalité des données dépend aussi du partenariat. Il est possible de récupérer les données sur une période souhaitée. Les données fournies peuvent également être mises à jour de manière régulière (mensuelle, trimestrielle, annuelle).

Estimation du poids de la donnée :

- Cela dépend de la zone à étudier, plus elle est importante plus les données le seront aussi. Selon mes recherches concernant les données de Strava on peut de données de "masse". Variation également dans le type de format utilisés (dépend du partenariat avec Strava)

Cas d'usages ciblés (usagers, métiers) :

- Département des transports de l'Oregon aux Etats-Unis
- Ville de Portland : mieux connaître les besoins et pratique des cyclistes
- Mairie de Paris
- Métropole de Lille, Lyon et Marseille
- Etude de caractérisation des services écosystémiques de la tourbières des saisis : utilisation de la donnée Strava Suunto (montre connecté) pour la fréquentation des pistes de ski (hiver) et course à pied/trail l'été
- Étude quantitative de la forêt de Haguenau, ONF : dans le contrat de projet la donnée de Strava est citée mais pas dans la synthèse de l'étude. Par contre, utilisation de la donnée flux vision d'orange

Limites :

API Strava :

- Prise en main de l'API et des requêtes SQL
- Disponibilité des données en fonction de l'autorisation fixé par les usagers
- Outil gratuit mais possible de devenir payant si Strava estima cela nécessaire

Strava Métro :

- Possible restriction d'accès à certains indicateur, type de données ou zone géographiques en fonction du partenariat établie
- Dépend de la volonté de Strava de fournies les données ou non

Les données portent uniquement sur les utilisateurs de l'application Strava

Coût :

API Strava :

- L'outil est actuellement gratuit. Mais Strava peut facturer l'utilisation de l'API notamment pour les professionnels. Il n'existe pas de prix spécifique sur les prix cela dépend de l'utilisation de la personne ou de l'organisme.

Strava Métro :

- L'accès aux données peut être ouvert gratuitement en réalisant une demande sur leur site web

Sources :

<https://metro.strava.com/>

<https://developers.strava.com/>

<https://www.strava.com/heatmap#8.97/-1.82843/48.16797/hot/all>

<https://www.strava.com/legal/api>

Webinar: How Strava Metro can Support your Active Transportation Decision Making :

<https://stravametro.zendesk.com/hc/en-us/articles/4402463916183--Webinar-How-Strava-Metro-can-Support-your-Active-Transportation-Decision-Making>

Glossaire et dictionnaire de données :

<https://stravametro.zendesk.com/hc/en-us/articles/1500001573281-Glossary-Data-Dictionary>

Cas d'usage

<https://veillecarto2-0.fr/2018/09/24/strava-reseau-social-et-big-data/>

<https://cartonumerique.blogspot.com/2018/03/big-data-et-choix-damenagement-urbain.html>

<https://cartonumerique.blogspot.com/2018/02/strava-et-le-monde-des-big-data.html>

Grégory Vermersch

<https://www.linkedin.com/pulse/strava-m%C3%A9tro-publie-des-nouvelles-donn%C3%A9es-tr%C3%A8s-pour-de-vermersch/?originalSubdomain=fr>

<https://dumas.ccsd.cnrs.fr/dumas-04033961/document>

<https://www.strava.com/settings/api>