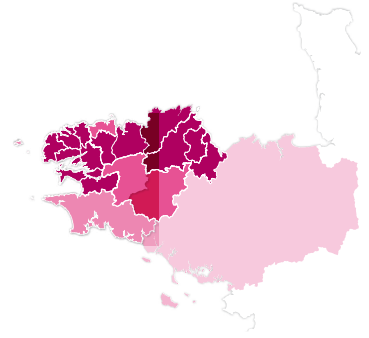


MÉGADONNÉES ET SOURCES ÉMERGENTES :

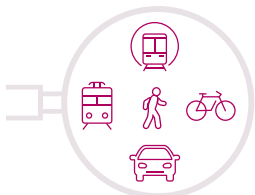
QUELS ENJEUX POUR L'OBSERVATION DES MOBILITÉS ?



#12 | Octobre 2024 OBSERVATOIRE | Note d'analyse



Crédit : ultramansk - Adobe stock



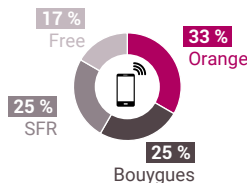
EN FRANCE
Les enquêtes mobilité certifiées Cerema couvrent **3/4** de la population et **41 %** des communes (Cerema, 2022)

4 737 377 737

c'est le nombre de valeurs produites en une année (2022) par les stations de mesure du trafic du réseau routier national non concédé (Cerema, Directions interdépartementales des routes).

Ces mégadonnées nécessitent donc le recours aux outils de l'intelligence artificielle pour être rendues intelligibles et opérationnelles.

EN FRANCE
Le marché de la téléphonie mobile est dominé par Orange, Bouygues, SFR et Free. (Arcep, Idate, 2022)



Les données de mobilité produites par les opérateurs doivent donc être redressées et généralisées à l'ensemble de la population.

La connaissance de la mobilité constitue un enjeu majeur de l'action publique car elle s'avère indispensable pour préparer, suivre et évaluer les politiques locales en matière de transport mais aussi dans d'autres champs (urbanisme, planification, logement, développement économique, etc.). Les collectivités expriment ainsi un besoin grandissant d'informations spécifiques pour réaliser efficacement leurs missions de service public. En parallèle, elles produisent aussi de manière croissante des données dans différents domaines, en particulier celui de la mobilité, sommées de répondre à la nécessaire ouverture des données publiques.

La loi d'orientation des mobilités (2019) poursuivait l'objectif de faire disparaître les « zones blanches de la mobilité » en couvrant l'ensemble du pays par des autorités organisatrices, conduisant de nombreuses communautés de communes à se saisir de la compétence *mobilité*. Cinq ans après la promulgation de la loi, force est de constater que toutes les collectivités ne disposent pas du même niveau d'information en la matière. Certaines sont couvertes par une enquête ménages-déplacements, d'autres sont accompagnées par un prestataire ou ont fait l'acquisition de données ad hoc, d'autres encore comptent une ingénierie dédiée au sein de leurs services. Mais nombre d'entre elles peinent encore à décrypter les pratiques de déplacements de leurs habitant·es et à proposer des solutions en conséquence.

Alors qu'émergent de nouvelles sources de données, notamment celles produites par l'usage massif des appareils connectés, il est légitime de s'interroger sur la pertinence de ces informations émanant désormais des acteurs privés. Comment ces données sont-elles élaborées et quels usages peut-on en faire ? Quelles complémentarités trouver avec les données publiques utilisées jusque-là ?

L'Adeupa a souhaité répondre à la problématique qui lui a été posée lors du comité de pilotage de l'observatoire de la mobilité de l'Ouest breton d'octobre 2023.

Ce rapport offre une approche synoptique et critique des données de mobilité actuellement à disposition et régulièrement utilisées dans la sphère publique. L'émergence des données massives ne rebat pas complètement les cartes mais implique une indispensable adaptation méthodologique et épistémologique pour que leur intégration dans le système de connaissance global soit réussie.

Introduction

La **gestion des données territoriales** s'est récemment imposée comme un **enjeu majeur de l'action publique** sous l'impact des évolutions législatives, comme la loi pour une république numérique (2016) et l'entrée en vigueur du règlement général sur la protection des données (RGPD, 2018), mais aussi en raison des **nouveaux usages de données numériques de plus en plus massives**.

Les données constituent désormais une **ressource incontournable dans la connaissance territoriale** et dans le **suivi des politiques publiques**, venant bouleverser les champs d'action et les manières de faire. Par exemple, en facilitant un usage mieux coordonné des modes de transport, le numérique s'impose parmi les solutions permettant de diminuer la congestion routière ou la consommation d'énergie.

L'émergence de **nouveaux acteurs privés** dans le domaine des techniques numériques a transformé l'écosystème de la donnée territoriale jusque-là dominé par quelques institutions publiques. Localement, les **collectivités s'emparent du sujet** en élaborant des stratégies territoriales autour de la donnée (cf. Service public de la donnée, SPLD), tandis que les citoyennes et citoyens produisent également de l'information de manière collaborative, à l'instar des projets de cartographie participative.

Si les **collectivités produisent de la donnée**, en particulier des données géographiques, dont une partie émane du secteur de la mobilité (AOM¹, opérateurs de transport, etc.), elles expriment un **besoin d'informations numériques spécifiques** pour prévoir la demande, comprendre les besoins des

habitants, évaluer leurs actions, planifier et adapter leur offre de transport.

La réponse à ce besoin peut se trouver dans l'exploitation des données issues de la téléphonie mobile ou des systèmes de navigation embarqués, dans le comptage automatisé par analyse vidéo et traitement par l'IA, voire dans des sources plus indirectes. **Les espoirs soulevés par ces données émergentes questionnent légitimement les acteurs publics de la mobilité** : ces nouvelles sources peuvent-elles se substituer en tout ou partie aux données classiques ? Quelles complémentarités trouver entre ces différentes sources ? Quels enjeux en matière de financement, de partenariat public-privé et de gouvernance ?

Lors du comité de pilotage de l'observatoire de la mobilité de l'Ouest breton qui s'est déroulé à Morlaix en octobre 2023, les participants – élus et techniciens – ont affirmé l'intérêt de **produire un éclairage sur ces enjeux en matière de connaissance des pratiques de déplacement des personnes et le suivi des politiques publiques**. Au carrefour de ces enjeux, l'Adeupa a produit ce rapport à partir d'une analyse de la littérature et d'entretiens avec des experts de la donnée et des techniciens des services des collectivités. Il s'agit de permettre de mieux comprendre les intérêts et les limites de ces données émergentes produites par des acteurs privés, en particulier celles issues de la téléphonie mobile. Avant cela, l'étude rappelle quels sont les enjeux liés à ces données de mobilité, comment elles sont utilisées par les agences et les collectivités et quelles sont les sources classiques toujours mobilisées dans les différents travaux.

Un service public local de la donnée pour...

- Maîtriser la donnée
- S'assurer du bon recueil de la donnée,
- Protéger les données
- Éviter la confiscation des données par des tiers

Des finalités : améliorer les réponses aux besoins des populations et des acteurs du territoire pour les missions de service public ; pour des actions publiques ou privées au service de l'intérêt général ; pour aider au développement d'actions privées bénéfiques au territoire.

Dans ce cadre, l'Adeupa a été missionnée pour jouer un rôle de tiers de confiance, garant de ces principes.

1. Autorité organisatrice de la mobilité.

La donnée : quels enjeux pour les territoires et les citoyens ?



Le secteur de la mobilité vit une véritable transformation numérique

Durant la dernière décennie, de **nouvelles sources de données** sont progressivement apparues pour suivre et analyser les déplacements, complétant un ensemble de sources existantes. Elles émanent d'une part de la **diffusion des téléphones mobiles** et des objets connectés et géolocalisés utilisés au quotidien, notamment ceux nous aidant dans notre mobilité (système de navigation), et d'autre part directement des **exploitants, AOM et autres acteurs de la mobilité** qui produisent de l'information numérique en continu.

Jusqu'à-là relativement restreint à quelques opérateurs et à des institutions publiques spécialisées dans la production d'information territoriale (Insee, Cerema²), le panorama des producteurs et des utilisateurs de données de mobilité s'est donc considérablement élargi avec la **révolution numérique**. Si les géants du web (notamment les Gafam – Google, Apple, Facebook, Amazon et Microsoft) se sont largement imposés en la matière, d'autres entreprises demeurent des acteurs de premier ordre dans le domaine, à l'instar des opérateurs de covoiturage ou de partage de véhicules, des fournisseurs de calculateurs d'itinéraires ou de solutions de billettique. **Utilisateurs continus des données publiques de mobilité**, ces entreprises sont aussi devenues des **producteurs majeurs d'information numérique** grâce à la masse d'utilisateurs de leurs solutions.



Du point de vue de la recherche, big data ne veut pas nécessairement dire rich data. »

Ollion et Boelart, 2015

2. Créé en 1946, l'Insee (Institut national de la statistique et des études économiques) est une direction générale du ministère de l'Économie et des Finances qui collecte, produit, analyse et diffuse des informations sur l'économie et la société françaises. Il est notamment chargé d'organiser et de contrôler les enquêtes de recensement, d'en exploiter les données et d'en diffuser les résultats.

Le Cerema (Centre d'études et d'expertise sur les risques, l'environnement, la mobilité et l'aménagement) est un établissement public relevant du ministère de la Transition écologique et de la Cohésion des territoires qui accompagne l'État et les collectivités territoriales pour l'élaboration, le déploiement et l'évaluation de politiques publiques d'aménagement et de transport.

Des sources émergentes qui interrogent nos manières de faire

La mobilité constitue l'un des domaines où l'émergence des mégadonnées provoque en effet d'importants changements méthodologiques et porte des promesses majeures.

Ces données sont produites par des opérateurs privés de téléphonie ou des services commerciaux numériques, échappant de fait à l'**ouverture³ attendue des données**. Mais, par l'importance de la diffusion d'appareils connectés dans la société, ces données sont particulièrement **massives** : couvrant de manière quasi exhaustive le territoire et s'inscrivant dans le temps, elles nécessitent des techniques de recueil et de traitement particulières. Plus récemment encore, l'**intelligence artificielle** ouvre de nouveaux horizons en matière d'analyse de ces données dans l'observation voire l'anticipation des pratiques de déplacement.

La **collecte régulière et automatique** de données grâce à des capteurs RFID⁴, des puces embarquées dans des téléphones ou des cartes de transport, fournit un tableau jusque-là hors de portée des analystes. En particulier, les données issues de la **téléphonie mobile** permettent d'obtenir des informations précises et fiables sur les trajets d'une vaste population – l'automatisation des enregistrements et

3. Les termes et les concepts soulignés sont définis dans les encadrés.

4. Les puces RFID permettent d'identifier et de localiser des objets ou des personnes (*radio frequency identification*).

l'accélération des traitements permettant d'étudier des échantillons considérables – avec une actualisation régulière.

De nombreux acteurs territoriaux, élus comme techniciens, se questionnent sur la pertinence de ces nouvelles données pour **répondre à leurs besoins en matière d'action publique**, en particulier au regard du coût et de la faible récurrence de mise à jour des enquêtes ménages-déplacements et constatant les besoins grandissants de **connaissance des pratiques de déplacement** des habitants, qui demeurent parfois dans un angle mort de l'observation territoriale.

La collecte de ces données issues des appareils mobiles est **passive** et donc beaucoup **moins contraignante** que les enquêtes par questionnaires ou entretiens. Les informations recueillies permettent notamment de **tracer l'itinéraire des usagers**. Elles représentent une opportunité pour répondre à certaines questions pour lesquelles les sources traditionnelles ne sont pas adaptées ou sont difficiles à mettre en œuvre (temporalité courte, population spécifique, échelle régionale ou nationale). En revanche, elles sont caractérisées par une **plus faible richesse en matière de caractérisation des déplacements** (modes, motifs) **et des individus** (genre, âge, classe sociale, etc.). Se pose également la **question d'ordre éthique** de la provenance de ces données émergentes, produites par des acteurs privés, marchandisées et, donc, non ouvertes alors qu'elles constituent une source d'information d'intérêt public.

Les mégadonnées ou données massives (en anglais big data) désignent ces ensembles d'informations numériques (c'est-à-dire encodées sous forme de chiffres) dont la production et le traitement ont été rendus possibles par une série de changements techniques (Ollion et Boelart, 2015) : croissance de l'usage d'internet, généralisation des téléphones intelligents, multiplication des capteurs de toute sorte (notamment les puces de géolocalisation par satellites, les accéléromètres, etc.), augmentation de la puissance de calcul des ordinateurs. Elles sont généralement caractérisées par la formule des « 3 V » identifiant ses propriétés : un volume considérable, une forte variété et une vélocité de la captation (permanente et immédiate).

L'ouverture des données (open data) a pour objet la diffusion libre, gratuite et universelle des données produites et collectées par les services publics (établissement, collectivité) et privés. Elles sont diffusées selon une méthode de licence ouverte garantissant leur libre accès et leur réutilisation par tous. En France, ce mouvement a été accompagné par la mise en ligne de la plateforme « data.gouv.fr », où plusieurs milliers de jeux de données sont désormais disponibles, et la mise en place de nouvelles structures administratives comme la direction interministérielle du numérique et du système d'information et de communication de l'État (Dinsic) ou la mission Etalab placée sous l'autorité du Premier ministre. Cependant, la marchandisation croissante des données issues des entreprises percutent frontalement cette ouverture nécessaire à l'intérêt public.

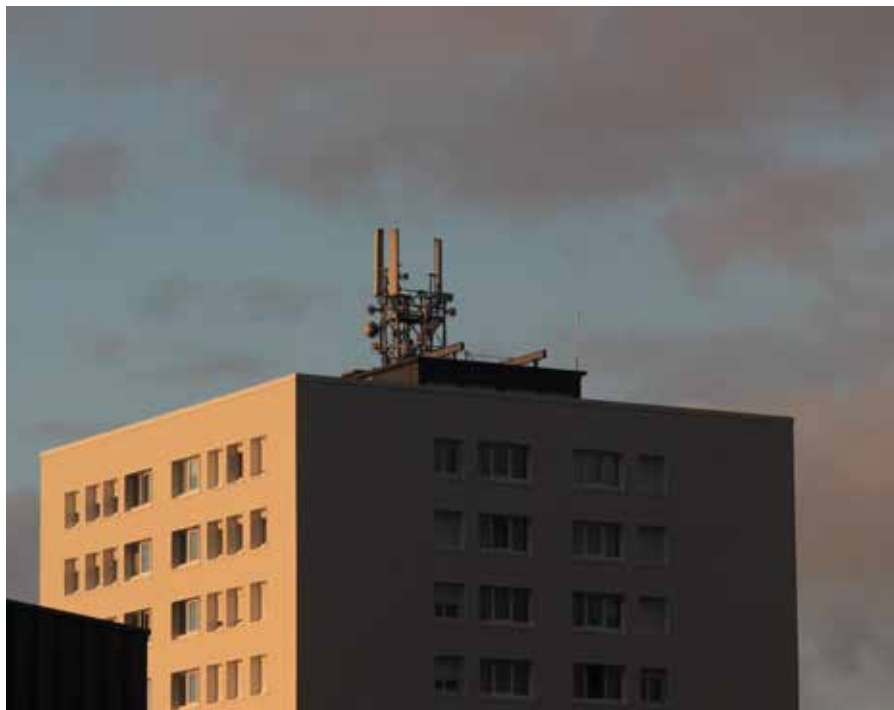
Les obligations légales

La loi d'orientation des mobilités a précisé et complété un règlement de la Commission européenne de 2017 encadrant la mise à disposition, dans l'ensemble de l'Union, de services d'informations sur les déplacements multimodaux. Les AOM, les opérateurs de transport et les fournisseurs de services de mobilité ainsi que les gestionnaires d'infrastructures sont donc tenus de mettre à disposition, par l'intermédiaire du point d'accès national (PAN), **les données de leurs services et réseaux**. C'est à l'Autorité de régulation des transports qu'a été confiée la mission de contrôler la conformité des publications de données et les modalités de réutilisation par les différents acteurs présents sur le marché. Cela concerne à la fois **les données statiques et les données dynamiques** (en temps réel).

Les AOM régionales sont les contributeurs majoritaires de l'offre de données des transports en commun. Les AOM, les collectivités locales, les opérateurs de mobilité et les fournisseurs de services sont les autres contributeurs au PAN.

Les **systèmes d'information multimodale (SIM)** fournissent des informations sur les déplacements et la circulation, couvrant plusieurs modes de transport et permettant la comparaison entre ces modes. Leur fonctionnement repose sur le traitement de données statiques ou dynamiques de mobilité. La publication des données de mobilité requise par le règlement délégué 2017/1926 est nécessaire au déploiement de SIM. La directive européenne relative aux systèmes de transports intelligents (STI) pose l'innovation dans les transports, notamment le développement des SIM, comme une solution pour limiter la congestion routière et la consommation d'énergie.

Le problème central posé par l'organisation des déplacements est celui des choix à faire aujourd'hui en fonction de l'idée que l'on se fait de l'avenir.



Antenne de téléphonie mobile à Brest - Crédit : Adeupa

À quoi servent les données de mobilité ?

Au sein des collectivités, les besoins et les usages de données de mobilité varient selon plusieurs facteurs : antériorité de l'exercice de la compétence mobilité, types de services proposés aux habitants, présence ou non d'un opérateur de transport collectif local, etc. Cependant, le problème central posé par l'organisation des déplacements est celui des choix à faire aujourd'hui en fonction de l'idée que l'on se fait de l'avenir. La réponse à cette question centrale est en partie à trouver du côté de **l'estimation des volumes de déplacements** et de leur **traduction en termes de trafic et de flux**.

La modélisation des déplacements : un outil indispensable pour l'action publique

La **modélisation des déplacements** permet de mettre en relation des variables décrivant l'espace urbain (données sociodémographiques, demande de transport, générateurs de déplacements) d'une part et son système de transport (offre de transport, capacités routières, réseaux de voirie, lignes de transports collectifs, aménagements cyclables) d'autre part. Les modèles de déplacements sont des outils d'aide à la décision qui permettent de connaître, de comprendre et de prévoir les mobilités. De manière concrète, la modélisation s'appuie sur des procédures de calculs dans lesquelles des variables sont introduites (situation de référence) et conduisent à l'estimation d'autres variables (scénarios). Cela permet de mettre en lumière des liens de causalité entre des variables décrivant les individus,

les ménages et le territoire et des variables décrivant les comportements de mobilité. **La situation de référence** est constituée à partir d'un ensemble de données, dont une partie est recueillie par enquêtes (caractéristiques socioéconomiques, graphe de l'offre de transport, matrice origine-destination caractérisant la demande de transport). A partir du modèle calé, les **scénarios** permettent, quant à eux, d'estimer et de simuler une nouvelle demande de transport et donc une éventuelle modification de l'offre, en lien avec des demandes d'études ponctuelles (réalisation d'un échangeur, construction d'un centre commercial, adaptation d'une ligne d'autobus, etc.) ou des études de plus grande ampleur (schéma directeur de voirie, projet de ligne de transport en site propre, contournement autoroutier, etc.).

Les matrices de déplacements : une photographie de la demande de mobilité

La modélisation nécessite l'usage de **matrices de déplacements**, communément appelées matrices origine-destination (O-D). Ce sont des tableaux synoptiques qui compilent le nombre de déplacements entre des zones d'un territoire d'analyse. Elles permettent de dresser un portrait de la demande de transport en fonction de différentes variables de segmentation (selon le mode de transport, le motif de déplacement, la période de la journée ou de la semaine, etc.). Ces informations sont généralement produites grâce à des enquêtes réalisées auprès des usagers des transports durant leurs déplacements ou au bord des routes auprès des conducteurs. Les personnes sont interrogées par un enquêteur sur leur déplacement en cours (origine, destination,

motif, etc.). Aujourd'hui, c'est pour élaborer ces matrices que les données de suivi des véhicules sont de plus en plus utilisées (voir p. 15).

Les données de mobilité servent ainsi à **l'élaboration, au suivi et à l'évaluation des politiques publiques** et des projets d'aménagement. Les usages sont très variés : positionnement de bornes de recharge des véhicules électriques, évaluation de la pertinence de nouveaux aménagements cyclables et de leur impact prévisible sur le partage modal, impact des flux touristiques sur le fonctionnement d'un réseau en période d'affluence, etc.

L'information-voyageur, l'autre grand domaine nécessitant des données sur la mobilité

Afin d'améliorer l'attractivité et l'efficacité des services de mobilité, il est indispensable **d'informer les usagers sur l'offre disponible**. L'ouverture des données de mobilité a été initiée en 2017 sur le fondement d'une directive européenne. Il s'agit en particulier de diffuser plusieurs informations : horaires, localisation des arrêts ou des stations, plan des lignes, itinéraires, tarification, etc. L'information peut provenir des exploitants ou prestataires de services, des autorités organisatrices de la mobilité, mais aussi de tiers, par exemple des fournisseurs d'application qui agrègent ces informations.

L'information est de deux types :

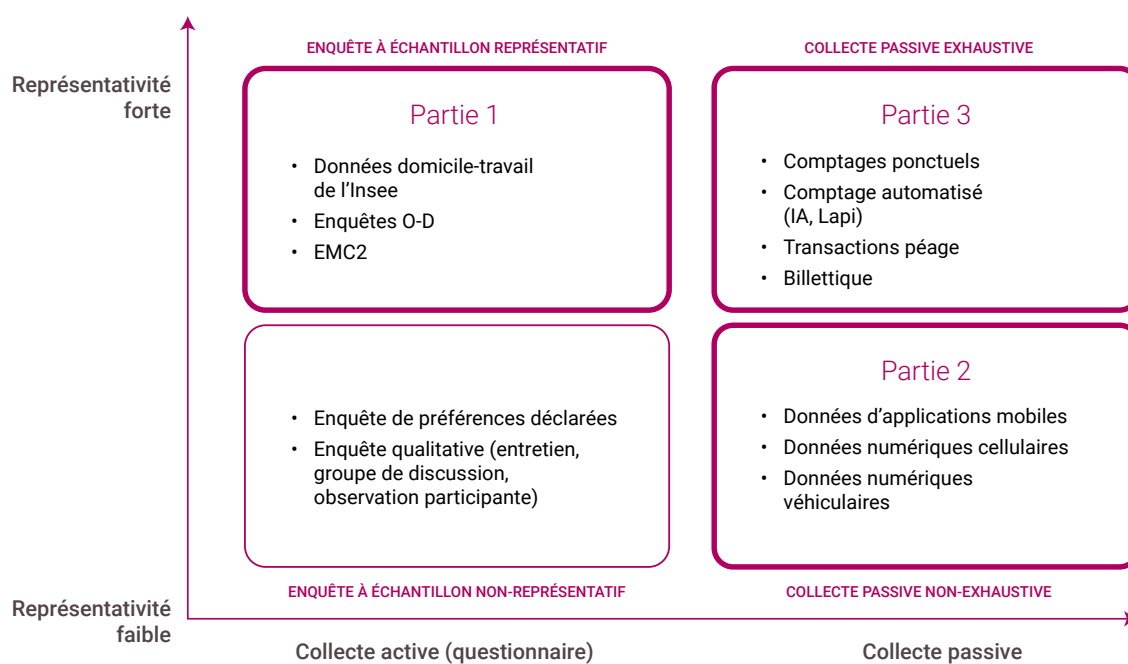
- statique (ou théorique) : temps de parcours théorique, horaires de passage, etc. ;
- dynamique (ou en temps réel) : temps de parcours effectif, horaires de passages en direct, etc.

Dans chaque pays de l'Union européenne, un **point d'accès national** unique doit rassembler, en ligne, l'ensemble des jeux de données relatifs à la mobilité : en Allemagne, c'est la Mobilithek, en Belgique, c'est le Nap ITS⁵, aux Pays-Bas, le NDOV Loker. En France, c'est la plateforme transport.data.gouv.fr qui donne accès à l'ensemble de ces données.

5. National Access Point for Intelligent Transport Systems.

Dans les parties suivantes, nous présenterons la plupart des sources de données sur la mobilité régulièrement utilisées par les collectivités, les agences d'urbanisme et d'autres structures publiques ou parapubliques. L'ensemble de ces données peut être classé sur deux axes : d'une part, celui de la représentativité des données et, d'autre part, celui du type de recueil, actif ou passif (Bousquet et Tricoche, Cerema, 2022). La typologie qui en découle permet d'isoler quatre grands blocs. Trois d'entre eux seront analysés dans cette publication : les enquêtes à échantillons représentatifs (partie I.), les données issues d'une collecte passive non-exhaustive (partie II.) et les données issues d'une collecte passive exhaustive (partie III.). Les approches plus qualitatives, telles que l'enquête de préférences déclarées (qui permet de connaître le comportement des usagers face à des situations hypothétiques) ou les enquêtes par entretiens, nécessitant une méthodologie spécifique, ne seront pas traitées ici. Enfin, les données portant sur l'offre de mobilité seront présentées dans la partie IV.

Typologie des données de mobilité : classification selon le type de collecte et la représentativité



D'après le Cerema, 2022

Partie 1

Les données à représentativité forte et faisant l'objet d'une collecte active : les enquêtes par sondage aléatoire

Ces données sont les plus couramment utilisées pour produire des modèles de déplacement. Elles s'appuient sur des enquêtes par sondage aléatoire, qui s'appuient sur un échantillonnage maîtrisé et un processus d'enquête robuste utilisant généralement un questionnaire. Les informations sont riches et comptent de nombreuses variables individuelles.

Les données domicile-travail de l'Insee

Des données stables et en accès libre mais représentant partiellement les déplacements des individus

Comment sont-elles produites ?

Durant le recensement de la population, l'Insee récolte des informations sur trois variables que sont le logement, le ménage et la personne. Depuis le millésime 2006 (et la mise en place du recensement en continu à partir de 2004), l'Insee estime la population, chaque année, à partir d'un échantillon composé de logements tirés au sort dans des communes de 10 000 habitants et plus, de résidences collectives (cités universitaires, maisons de retraite,

casernes, etc.) et de logements tirés au sort dans des communes de moins de 10 000 habitants. L'échantillonnage est étalé sur cinq années et c'est l'année médiane qui fait office de référence : ainsi, par exemple, les données diffusées en 2014 sont issues des enquêtes organisées entre 2009 et 2013 et décrivent la situation de l'année 2011. Les questionnaires sont auto-administrés sur papier ou en ligne.

Le cas particulier des migrations alternantes

Sur cette base, l'Insee construit d'autres unités observées, comme les familles, les migrations résidentielles et les navettes domicile-travail et domicile-études. Ce sont ces deux unités qui nous intéressent, rassemblées sous le concept de « migrations alternantes ». Il est important de noter que l'unité mesurée est bien une **navette**, c'est-à-dire un aller-retour d'un lieu à un autre, et non pas un **déplacement**. Ni la fréquence quotidienne, ni la fréquence hebdomadaire des déplacements ne sont observées. Surtout, **ces navettes ne sont pas des déplacements effectivement réalisés**.

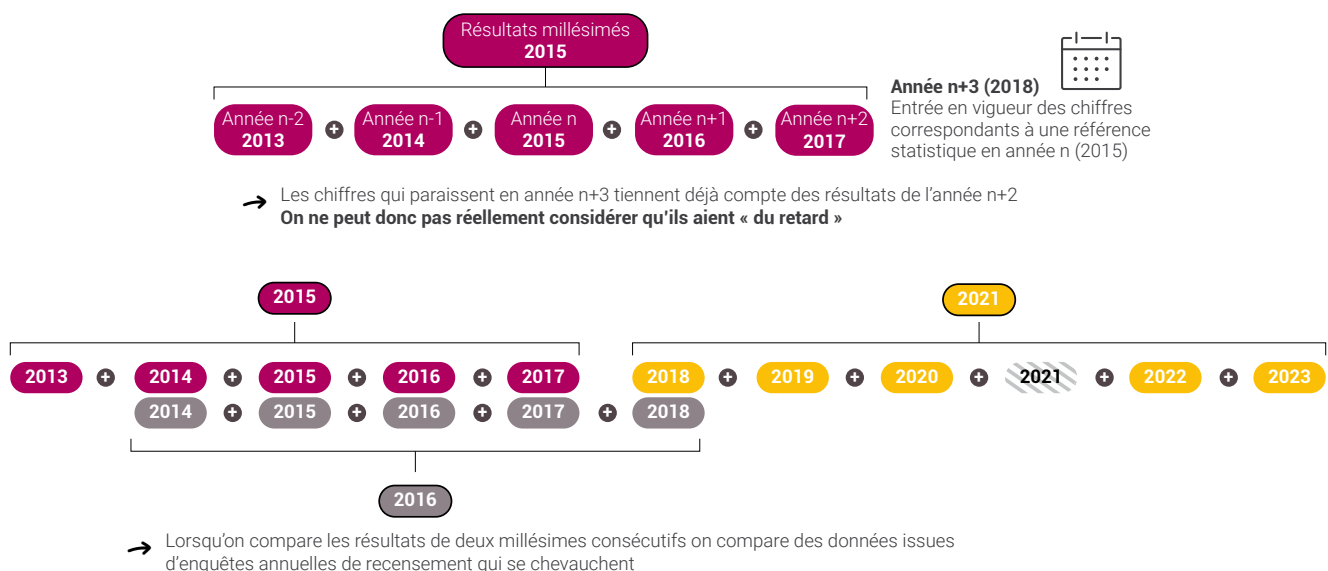
Les navettes **domicile-travail** concernent les personnes résidant habituellement en

France, âgées de quinze ans au moins et occupant un emploi (salarié ou non, à temps plein ou non, notamment en stage ou en apprentissage).

Les navettes **domicile-études** concernent les personnes résidant habituellement en France, âgées de deux ans ou plus et inscrites dans un établissement scolaire ou universitaire l'année en cours (le terme « études » renvoie donc à l'ensemble du parcours scolaire et universitaire, dont l'apprentissage).

Lors du recensement, les personnes déclarent leur mode **principal habituel** utilisé pour se déplacer vers leur lieu de travail (ce mode n'est pas disponible pour les navettes domicile-études). **Principal** signifie que le mode retenu est celui qui, dans le cadre d'un déplacement multimodal, est réalisé sur la plus longue distance (un déplacement composé d'un premier trajet à vélo vers une gare puis d'un trajet en train et, enfin, d'un dernier trajet à pied sera donc enregistré sous le mode « transport en commun »). **Habituel** renvoie au mode le plus utilisé, dans la semaine ou durant l'année (ainsi, un actif se déplaçant à vélo durant les beaux jours mais utilisant sa voiture le reste du temps se verra attribuer le mode « voiture »).

Pourquoi est-ce qu'on ne peut pas comparer deux années successives du recensement ?



Quels usages ?

Les données sont mises à disposition sous deux formes :

- La matrice des navettes de commune de résidence à commune de travail ou d'études (unité : la navette) ; permet, par des agrégations, d'obtenir une matrice à des échelles supérieures, par exemple les EPCI. En revanche, il n'existe pas de données infracommunales ;
- Les fichiers-détails qui permettent de produire une matrice à façon, en caractérisant les individus qui effectuent les navettes (unité : l'individu). Ces caractéristiques individuelles sont le sexe, la tranche d'âge, la motorisation, le type de ménage, la CSP et le type d'emploi occupé.

L'utilisation immédiate des migrations alternantes pour construire des **matrices origine-destination** pour le motif travail n'est pas possible. Elles peuvent cependant être utilisées pour de la **modélisation** simple, de manière croisée avec une enquête ménages-déplacements représentative du territoire modélisé. D'autres motifs peuvent ainsi être extrapolés. La fiabilité de la matrice obtenue passe obligatoirement par une confrontation à des données obtenues sur le terrain (comptages par exemple).

Il est à noter que, depuis l'enquête annuelle de recensement de 2016, les modalités de réponse à la question portant sur le mode ont évolué pour mieux rendre compte des évolutions des pratiques. Ainsi, la modalité « deux-roues » a été éclatée entre « vélo (y compris à assistance électrique) » et « deux-roues motorisé ». L'un des usages courants de cette information est de calculer la part modale des déplacements domicile-travail d'un territoire.

Ce jeu de donnée permet également de produire des indicateurs qui permettent de comprendre le fonctionnement des territoires, en calculant, par exemple, la part des navettes domicile-travail internes à un territoire ou le solde entre navetteurs entrant et sortant : ces indicateurs permettent d'illustrer le degré d'autonomie – ou de dépendance – d'un territoire⁶.

6. Il convient cependant de garder en mémoire que les déplacements pour le travail ne constituent qu'une partie des flux réalisés au quotidien par les habitants (voir ci-contre). Également, l'approche de la part modale par l'unité du déplacement implique un biais par rapport à l'unité du voyageur-kilomètre qui mesure, quant à lui, les distances parcourues. L'unité du déplacement a comme désavantage le fait de mettre sur le même plan un trajet à pied de quelques centaines de mètres avec un trajet en voiture ou en train de plusieurs dizaines de kilomètres. Or, les implications en matière de coût du déplacement ou encore d'émission de GES nécessitent de s'intéresser plutôt aux distances parcourues.

Limites et biais connus

Comparaison dans le temps

Le recensement permet de calculer des évolutions en tenant compte du fait que la construction de l'échantillon est étalée sur cinq années et donc que les résultats sont construits par glissement. Seul un cinquième de l'échantillon est mis à jour chaque année ce qui interdit d'analyser des évolutions annuelles. Les comparaisons doivent ainsi être réalisées à partir de l'année N-5 (le millésime 2017 peut être comparé au millésime 2012).

Représentativité des motifs

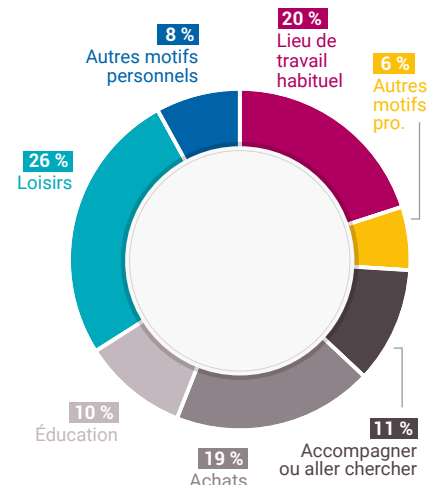
Le recensement ne mesure que deux types de mobilité quotidienne : celle liée au travail et celle liée aux études. Or, ce ne sont pas les seules raisons pour lesquelles nous nous déplaçons au quotidien. En effet, l'enquête nationale mobilité des personnes de 2019 (voir ci-après) montre que la mobilité liée au travail représente, en moyenne, un déplacement sur cinq réalisés chaque jour. À cela il faut ajouter un déplacement sur dix réalisés pour les études. Ainsi, si elle ne représente qu'un déplacement quotidien sur cinq, la mobilité liée au travail représente près de 30 % des kilomètres parcourus quotidiennement en France. Elles sont donc particulièrement structurantes dans notre quotidien et demeurent la cible privilégiée de la plupart des politiques de transport.

Autres biais

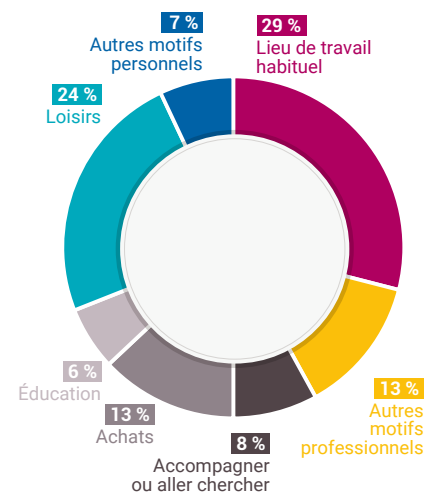
Le recensement ayant lieu en début d'année (janvier-février), cela induit des effets saisonniers au niveau local (par exemple sur le personnel saisonnier hivernal).

L'Insee déconseille de mener des analyses pour des zones de moins de 2 000 habitants. Un effectif de 500 navettes minimum est nécessaire pour considérer des ordres de grandeur.

Répartition du nombre de déplacements en France selon le motif



Répartition du nombre de voyageurs-km en France selon le motif



Source : SDES, Insee
Enquête Mobilité des Personnes 2018-2019

En résumé - Synthèse des atouts et limites des données de navettes de l'Insee

+	-
Coût : données en accès libre et gratuit	Mesure des flux et non des déplacements effectivement réalisés
Excellente représentativité de la population	Ne mesure que les navettes pour le travail et les études, pas les autres motifs
Possibilité de comparaison à l'échelle nationale	Le questionnaire auto-administré engendre des biais de déclaration (temps de trajet, modes utilisés)
Actualisation annuelle mais comparaison temporelle possible tous les cinq ans	Pas de données infracommunales

Les enquêtes déplacements et les enquêtes origine-destination

Une source pensée et élaborée pour observer la mobilité quotidienne d'un territoire

Les enquêtes déplacements ou enquêtes mobilités sont réalisées auprès d'un échantillon de ménages et offrent une « photographie » des déplacements réalisés un jour donné, la veille de l'enquête des personnes sélectionnées (ou le samedi précédent lorsque les questions portent sur les déplacements du week-end). Il existe deux principaux producteurs : les collectivités accompagnées du Cerema pour les enquêtes locales (comme pour l'**EMD du pays de Brest**) ; le Service des données et études statistiques (SDES) pour l'enquête nationale mobilité des personnes. Plus restreintes dans le temps, les enquêtes origine-destination (O-D) sont produites pour les besoins de la modélisation auprès des usagers des transports au cours de leur voyage ou auprès des conducteurs au bord des routes.

EMC2 : un dispositif historique dont le Cerema est le garant

Il s'agit de la principale source de données ad hoc fournissant **des informations robustes sur les déplacements des habitants** d'un territoire. À l'origine conçues pour les besoins de la modélisation des déplacements dans les années 1960, leur usage s'est étendu pour répondre aux besoins de connaissance des comportements. Aujourd'hui, l'ensemble des protocoles est regroupé sous l'appellation Enquête de mobilité certifiées Cerema (EMC2). C'est donc le Cerema qui suit l'élaboration des enquêtes et en assure la validité. Mais l'initiative revient aux collectivités territoriales qui en financent la majeure partie, en faisant notamment appel à des bureaux d'études spécialisés pour la passation des entretiens.

C'est à ce jour **la source la plus riche pour analyser le fonctionnement d'un territoire** en matière de mobilité. La rigueur méthodologique attachée à leur réalisation permet à l'ensemble des enquêtes d'être comparées dans le temps et l'espace, bien que tout le territoire national ne soit pas couvert. Il convient ainsi de conserver les mêmes périmètres d'une enquête à l'autre, ou tout du moins d'intégrer le périmètre ancien en totalité dans le périmètre élargi.

Le dispositif actuel est composé d'une enquête cœur mais aussi d'enquêtes complémentaires optionnelles (recueillant par exemple des données supplémentaires afin de disposer de plus d'informations sur les comportements de déplacements) et d'enquêtes satellites (ciblage d'une sous-population comme les non-résidents et les visiteurs afin de comprendre la mobilité des touristes).

L'Enquête ménages déplacements (EMD) du pays de Brest, 2019

Cette enquête visait à mieux connaître les pratiques de mobilité dans le pays de Brest, bassin de déplacements du quotidien de la grande majorité des habitants et correspondant au périmètre du Scot. Le PLU Facteur 4 de Brest métropole (2014) avait également inscrit la réalisation d'une telle enquête.

Répondant au standard national, elle peut ainsi être comparée avec d'autres territoires et avec les enquêtes antérieures. L'EMD demeure toujours la source principale de connaissance de la mobilité des habitants dans ce territoire. La précédente enquête de ce type portant sur le pays de Brest a été réalisée en 2003-2004. Depuis, de nombreux paramètres ont évolué (structure et répartition de la population, modes de vie et aspirations des habitants, etc.). La mise en service de la ligne A du tramway a été réalisée en 2012 et a été suivie par le téléphérique en 2016.

Cette connaissance des pratiques représente donc un enjeu majeur pour Brest métropole, maître d'ouvrage de l'EMD, et ses partenaires financiers (État, Région, Conseil départemental du Finistère, pôle métropolitain du pays de Brest) afin d'alimenter la construction des politiques de déplacements et d'aménagement. D'un coût global d'environ 500 000 euros, la passation a été réalisée par la société Alyce entre le 9 octobre 2017 et le 5 février 2018 ; elle a enquêté 5 510 ménages regroupant 7 588 habitants de 5 ans et plus. Ce coût ne compte pas les traitements et analyses.

Le Cerema a assuré la préparation, le suivi de l'enquête, le contrôle de la conformité au standard, l'apurement des fichiers, le redressement des données ainsi que l'exploitation standard. L'Adeupa a, quant à elle, été chargée de l'analyse et de la valorisation des résultats, travail toujours en cours, dans le cadre de l'Observatoire de la mobilité ou à travers d'autres missions confiées à l'agence (planification, études).

Les résultats de l'EMD sont mis à disposition [en ligne par Brest métropole](#).

Une procédure lourde mais récemment allégée

Un plan de sondage permet d'élaborer un échantillon représentatif de la population à partir d'une sélection de ménages résidant dans le périmètre de l'enquête dans lesquelles certaines personnes, ou toutes les personnes de 5 ans et plus habitant au moins trois nuits en semaine, sont interrogées sur les déplacements qu'elles ont réalisés durant les dernières 24 heures. L'unité de l'échantillonnage est donc le logement utilisé comme résidence principale, à partir d'un tirage réalisé dans le fichier des propriétés bâties de la Direction générale des finances publiques.

La méthode standard préconise de stratifier le territoire d'enquête en sous-ensembles géographiques de taille comparable en matière de population et présentant une relative homogénéité des comportements de mobilités. Parmi chacun de ces sous-ensembles, 70 ménages et 160 personnes au minimum sont enquêtés.

La collecte est réalisée entre septembre et avril, hors vacances scolaires, du mardi au samedi, et ce durant huit semaines au minimum pour lisser les résultats et réduire les biais liés à des phénomènes ponctuels. Une enquête complémentaire peut être réalisée pour connaître la mobilité durant le week-end.

La passation est réalisée en face-à-face au domicile et au téléphone. Les ménages tirés au sort sont auparavant informés par courrier. C'est l'enquêteur qui remplit le questionnaire, l'entretien durant en général une vingtaine de minutes.

Le Cerema estime que les enquêtes ont une validité d'environ dix ans. Ainsi, il conviendrait de les réaliser de manière régulière selon ce pas de temps.

La réalisation d'une EMC2 nécessite des moyens techniques et financiers importants, ce qui peut rebuter des territoires. Pour simplifier, on considère que le coût varie entre 1,5 et 2 € par habitant. Ce ratio permet d'évaluer le coût total pour le périmètre retenu. Il ne prend pas en compte le coût de traitement et d'analyse des données. Également, une enquête de ce type nécessite un processus long réunissant de nombreux acteurs. En moyenne, il faut compter deux ans à deux ans et demi entre la prise de décision et la livraison des résultats.

Quelles sont les informations fournies ?

Les résultats permettent de reconstituer plusieurs variables de mobilité :

- Connaissance de la mobilité des personnes
- Budget-temps et budget-distance
- Équipement des ménages
- Comportements de mobilité et habitudes de déplacements
- Flux de voyageurs par mode

En particulier, des variables sont recueillies sur les déplacements :

- Motifs à l'origine et à la destination
- Lieux d'origine et de destination (à la zone)
- Modes utilisés par trajet
- Horaires de déplacement
- Distance modélisée à partir de la distance à vol d'oiseau

Enfin, de nombreuses informations sur le ménage sont recueillies : taille du ménage, type de logement, nombre de véhicules à disposition, type de stationnement, âge, genre, profession et CSP, permis et abonnement de transport, zone fine du lieu de travail ou d'études, etc.

Principaux biais connus

L'absence des résidences collectives dans le fichier des propriétés bâties induit une sous-représentation de certaines populations comme les étudiants. Un échantillon spécifique peut être ajouté afin d'enquêter les résidences étudiantes. C'est particulièrement important dans les villes étudiantes.

Également, les déplacements réalisés par les personnes ne résidant pas dans le périmètre ne sont pas pris en compte. C'est une limite qui est à considérer en particulier dans les villes attirant des actifs qui résident loin du centre et qui vont utiliser les réseaux de transport urbains (par exemple, dans l'EMD du pays de Brest, les habitants de Morlaix qui travaillent à Brest ne sont pas comptabilisés).

En résumé - Synthèse des atouts et limites des données des enquêtes ménages-déplacements

+	-
Données représentatives de l'ensemble des déplacements	Un coût important et une mise en œuvre technique qui peut être lourde
Comparaisons possibles dans l'espace et dans le temps	Un processus long d'environ deux ans de la préparation à la diffusion des résultats
Une durée de vie d'environ dix ans	Les déplacements observés ne concernent que les habitants du territoire
Alimente les modèles de déplacements	Une nécessaire valorisation des résultats à anticiper

L'enquête mobilité des personnes

Réalisées régulièrement depuis les années 1960 par les services ministériels (c'est désormais le Service de la donnée et des études statistiques du ministère de la Transition écologique qui en a la charge), les enquêtes mobilité des personnes (anciennement enquête nationale transports et déplacements ou simplement enquête Transports) offrent une source d'information essentielle pour évaluer la mobilité des Français à l'échelle de l'ensemble du pays, en particulier dans les territoires peu denses non couverts par une EMC2. Produites en 1966-1967, 1973-1974, 1981-1982, 1993-1994, 2007-2008 et 2018-2019, ces enquêtes permettent de suivre dans le temps long l'évolution des pratiques et du parc de véhicules. Elles autorisent aussi des comparaisons locales avec les enquêtes du Cerema, plusieurs indicateurs étant communs.

La passation pouvant être jugée longue et fastidieuse par les personnes enquêtées, cela peut provoquer une sous-déclaration de mobilité liée au manque d'implication du répondant.

Le redressement réalisé après l'enquête permet en partie de corriger ces biais.

Les enquêtes origine-destination (O-D)

Ces enquêtes sont produites par les exploitants de réseaux de transports ou de réseaux routiers, par les AOM et les maîtres d'ouvrage de projets de transport. La passation, généralement réalisée par un prestataire spécialisé, dépend fortement du type de déplacement observé : les déplacements routiers d'une part, ceux en transports collectifs d'autre part.

Elles permettent la modélisation des déplacements en estimant les flux d'échange et de transit, et en élaborant les matrices O-D. Elles sont utilisées seules ou en combinaison avec d'autres dispositifs (comptage automatisé, données de billettique, etc.). Si leur méthode n'évolue pas entre les différentes campagnes d'enquêtes, elles peuvent enfin servir à l'observation d'évolutions temporelles. En général, elles sont actualisées tous les cinq à dix ans.

Les enquêtes O-D en bord de route

Ces enquêtes ont pour objet d'interroger les conducteurs de véhicules légers et de poids-lourds sur le déplacement qu'ils sont en train d'effectuer. Leur heure de passage, le pays d'immatriculation, le nombre d'occupants du véhicule, le type de véhicules, l'origine et la destination ainsi que les motifs sont les variables les plus souvent collectées. Un comptage est généralement effectué en parallèle.

La réalisation de ces enquêtes nécessite de prendre des dispositions de sécurité et doit le plus souvent être encadrée par les forces de l'ordre qui sont habilitées à dévier les véhicules vers le point d'enquête. Le dispositif est ainsi difficile à mettre en œuvre sur les axes très fréquentés ou dans des espaces où la vitesse est élevée. Plusieurs méthodes de collecte existent selon les objectifs (voir schéma ci-après). La plus connue est l'enquête dite « cordon » qui permet d'estimer l'ensemble des flux qui entrent et sortent d'une zone.

L'échantillon est composé aléatoirement en bord de route, en arrêtant régulièrement des groupes de véhicules. Le biais de non-réponse est négligeable car les refus sont peu courants. Pour certaines catégories, des échantillons minimums sont définis. Comme pour d'autres enquêtes, ce sont les mardis et les jeudis qui sont généralement sélectionnés, hors vacances scolaires et, parfois, seulement aux heures de pointe.

La majeure partie des variables observées concernent le véhicule (heure de passage, type de véhicule, nombre d'occupants, nationalité, origine, destination, etc.) et son conducteur (motif, fréquence de réalisation du trajet). Véhicules légers, utilitaires et poids-lourds sont enquêtés.

Les données brutes sont ensuite redressées à partir des comptages manuels puis des comptages automatiques.

Les enquêtes O-D dans les transports collectifs

Ces enquêtes se déroulent dans les véhicules de transport pendant les trajets (ou parfois aux arrêts) et consistent à interroger les passagers sur leur déplacement en cours (arrêts de montée/descente, origine, destination, motif, correspondance, etc.). Si ce sont généralement des enquêteurs qui déroulent le questionnaire, celui-ci peut parfois être auto-administré. Les enquêtes embarquées permettent d'administrer un questionnaire relativement élaboré, ce qui autorise le recueil d'informations détaillées sur le passager.

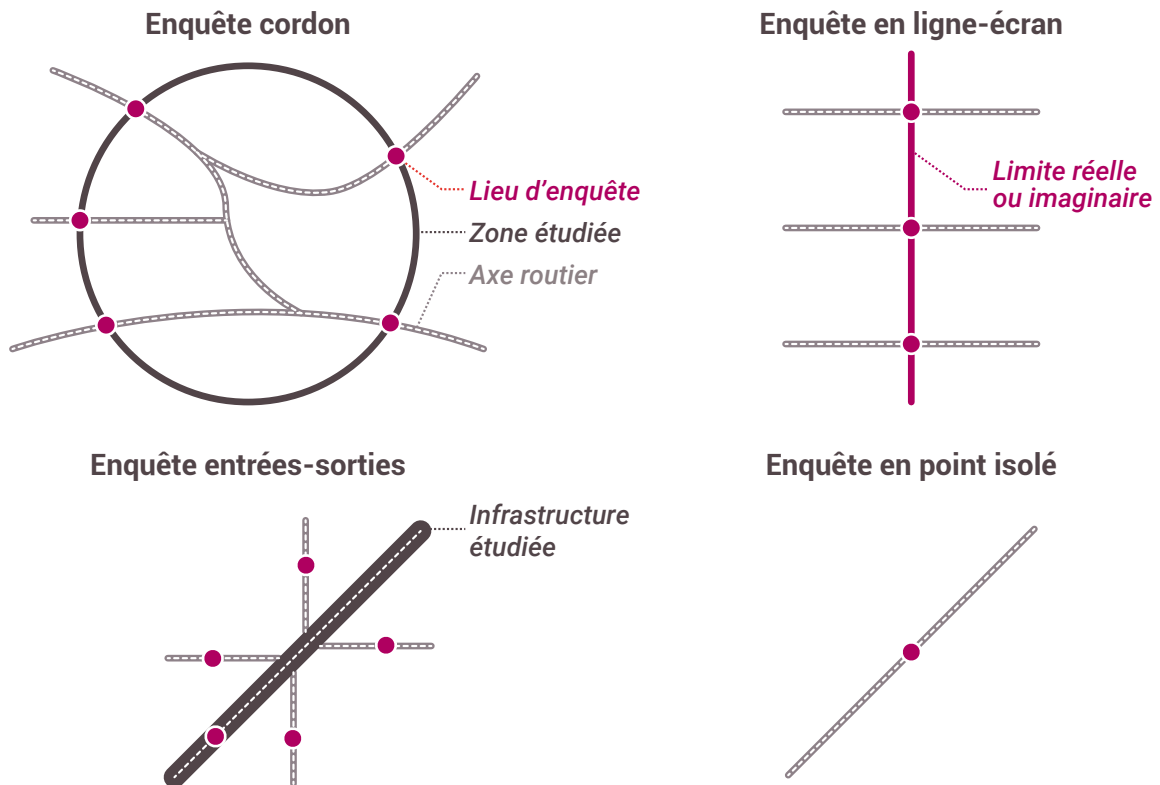
L'enquête a lieu durant toute une journée ou, a minima, durant les heures de pointe, essentiellement les mardis ou jeudis hors vacances scolaires. Il faut généralement répéter cela plusieurs jours durant un à deux mois consécutifs, afin d'obtenir des résultats représentant huit à 10 jours effectifs. Des « photographies » peuvent être réalisées d'autres jours pour des besoins spécifiques (départ en week-end pour le transport ferroviaire, le samedi et dimanche pour estimer les flux liés aux touristes).

Les résultats sont ensuite redressés grâce à des comptages automatiques ou manuels. In fine, cela permet de reconstituer un jour moyen élaboré sur la base d'une dizaine de jours d'enquête.

Si la comparaison temporelle entre deux enquêtes est rendue difficile en raison des évolutions du réseau de transport (ce qui n'est pas le cas pour des réseaux de transport lourd, notamment ferroviaire), les matrices origine-destination peuvent en revanche être comparées.

L'enquête déplacements reste à ce jour la source la plus riche pour analyser le fonctionnement d'un territoire en matière de mobilité.

Enquête OD routière : les différentes méthodes de collecte



D'après Bousquet, Cerema, 2022

Partie 2

Les données à représentativité faible et faisant l'objet d'une collecte passive : les mégadonnées

Ces données sont caractérisées par le très grand nombre d'unités observées, ce qui nous amène à parler de données massives ou mégadonnées. Elles sont issues des terminaux connectés que nous utilisons quotidiennement ou, tout du moins, régulièrement. Il s'agit en premier lieu des **téléphones mobiles** mais aussi des **systèmes d'aide à la navigation** présents dans nos véhicules. Les opérateurs de téléphonie, les fournisseurs de service d'aide à la navigation et les développeurs d'applications collectent ainsi des **masses considérables de données géolocalisées**. En particulier, la téléphonie mobile présente des atouts inédits en matière de pénétration et d'utilisation par la population. La représentativité spatiale et démographique de ces données est relativement fiable.

Une importante hétérogénéité en matière de production et d'acteurs

Le recueil de ces traces est assez hétérogène. Elles ont deux points en commun. D'une part, il s'agit de **sous-produits issus de l'usage massif des outils connectés** ; les données brutes ne deviennent des informations sur les déplacements qu'après avoir été collectées, purgées, agrégées et enrichies par les divers opérateurs et entreprises spécialisées. D'autre part, ces informations sont **produites de manière passive par les individus**.

Parmi les producteurs, on trouve notamment des opérateurs de téléphonie mobile qui commercialisent ces données, soit directement à des collectivités ou des laboratoires de recherche par exemple, soit à des entreprises intermédiaires qui vont se charger de traiter plus finement et d'enrichir ces données à différentes fins : comprendre les déplacements de la population, réaliser des études de marché, faciliter la circulation routière, etc. L'essentiel de ces sociétés étant privées, **elles protègent leurs sources d'informations et les traitements réalisés par le secret industriel**. Ainsi, **ces données présentent des biais relativement similaires**. Enfin, le paysage des commercialisateurs est devenu, ces dernières années, relativement dense et il est parfois difficile d'identifier le bon interlocuteur en fonction de ses besoins.

Trois types de données selon leur mode de production

Actuellement, trois grands types de données de déplacement sont commercialisés. Les deux premiers sont directement issus de l'usage des téléphones mobiles. Le troisième provient, quant à lui, de l'usage des applications mobiles et des systèmes d'aide à la navigation utilisés en voiture.

Parmi les deux premiers types, il convient de différencier deux sources issues de la téléphonie mobile qui se distinguent par la méthode de recueil. Ils feront l'objet des deux prochaines parties.

- La première provient du **réseau d'antennes** disséminées dans le territoire et auquel les téléphones mobiles se connectent régulièrement (on parle généralement de **Floating Mobile Data, FMD**). La localisation du propriétaire du téléphone peut ainsi être réalisée directement à partir de l'antenne à laquelle il s'est connecté lors d'un événement (appel, message, etc.) ou de manière automatique (envoi régulier d'un signal de localisation). Actuellement, seuls Orange et SFR proposent une offre commerciale à partir des données de leur réseau. Comme environ 95 % de la population dispose d'un téléphone portable⁷ et que chacun des opérateurs représente 15 à 35 % du marché, ces données ont un fort potentiel de représentativité.
- La seconde provient, quant à elle, du **système de positionnement géographique embarqué du téléphone** (GNSS⁸) dont les **traces** sont recueillies par des applications tierces (on parle de données **SDK pour Software Development Kit**, kit de développement logiciel, c'est-à-dire un extrait de code fourni aux éditeurs d'application mobiles afin d'être intégrées dans les applications pour transmettre les données de géolocalisation).

7. Insee, Enquêtes sur les TIC auprès des ménages, 2021

8. Le sigle GNSS (géolocalisation et navigation par un système de satellites) désigne le système de localisation et de navigation, associant plusieurs systèmes à couverture mondiale, notamment le système GPS (américain), le système Glonass (russe) et le système Galileo (européen), pour répondre aux besoins des utilisateurs des services terrestres, maritimes et aéronautiques.

Trace numérique

Ce sont des données informationnelles créées à chaque fois que l'on se sert d'un outil numérique connecté, comme un ordinateur, un téléphone, un système de navigation embarqué, etc. L'ensemble de ces traces sont enregistrées sur des serveurs de bases de données et peuvent être traitées par des algorithmes à des fins diverses. Elles sont dites géo-numériques si des données (ou métadonnées) géographiques permettant leur spatialisation leur sont associées. Il peut s'agir de coordonnées géographiques, d'un toponyme, d'un numéro de borne Wi-Fi, d'une cellule de réseau de téléphonie mobile ou encore d'un nom de lieu (Merickskay, 2019).

Trois grands types de données de déplacement sont commercialisés. Les deux premiers sont directement issus de l'usage des téléphones mobiles. Le troisième provient, quant à lui, de l'usage des applications mobiles et des systèmes d'aide à la navigation utilisés en voiture.

Les données numériques cellulaires issues des antennes téléphoniques (FMD)

Comment sont-elles produites et quels usages peut-on en faire ?

Les téléphones mobiles produisent régulièrement des informations de localisation (les traces) qui sont associées à un même appareil grâce à un alias téléphonique affilié à la carte SIM. D'une part, afin de repérer rapidement les téléphones en cas de communication, ceux-ci émettent régulièrement un signal de localisation (plusieurs fois par jour) : ce sont les **données de signalisation**. D'autre part, à chaque fois qu'un terminal mobile est utilisé lors d'un événement (effectuer un appel, envoyer un message ou transférer des données), il se connecte à une antenne : ce sont les **données de facturation** ou de transaction (Bonnell et al., 2017). Les premières permettent de connaître en permanence la localisation d'un téléphone mais à l'échelle de zones de localisation (location area) qui comptent de nombreuses antennes. Les secondes sont plus précises car elles identifient l'antenne à laquelle le téléphone a borné, mais uniquement lors d'un événement (schéma p. 13). Actuellement, ce sont les données de signalisation, à l'échelle des zones de localisation, qui sont utilisées (Bonnell et al., 2017).

En France, à l'heure actuelle, Orange, avec Flux Vision, et SFR, avec Geostatistics, proposent une offre commerciale avec des traitements à façon de ces données.

Des algorithmes de traitement produisent des journaux de déplacement pour chaque alias téléphonique à partir de ces données brutes. Celles-ci sont massives : on estime entre deux et quatre milliards le nombre de signalements par jour et par opérateur

pour la France (Charpe et Harache, 2023). À partir de l'ensemble de ces journaux, les **déplacements sont reconstitués** et les origines et destinations obtenues sont affectées à des zones géographiques, notamment les Iris, après des traitements statistiques (utilisation de la propagation du signal tenant compte de la topologie des antennes ou encore de la topographie). Les opérateurs conservent les données durant douze mois pour des raisons légales. Ainsi, passé ce délai, elles sont détruites. Il faut alors anticiper si une période en particulier doit être observée.

Ces données peuvent servir à la réalisation de **matrices origine-destination** et d'enquêtes ad hoc de mobilité sur un territoire en particulier. Elles apportent des éléments de connaissance sur la fréquentation d'un site ou d'un territoire, par exemple en lien avec l'événementiel. Parce qu'elles distinguent les déplacements d'**échange** et de **transit** des déplacements internes à une zone donnée, ces données sont également utilisées en complément d'autres sources pour construire des modèles de déplacement, en particulier en l'absence d'enquête déplacements (Charpe et Harache, 2023).

Quelles variables sont observées et quels sont les principaux biais connus ?

Les unités observées sont les alias téléphoniques associés aux cartes SIM. Il est ainsi possible de reconstituer les volumes de déplacements de personnes entre les zones d'un territoire par pas de temps (par heure, par type de jour, etc.), en distinguant les déplacements d'échange et de transit ainsi que les résidents et les non-résidents. Ainsi, les personnes étrangères sont incluses ; cependant, si les résidents européens bénéficient de la prise en charge courante des frais d'itinérance par les opérateurs locaux, ce n'est pas le cas des touristes extra-européens dont les usages sont encore très hétérogènes.

En France, à l'heure actuelle, Orange, avec Flux Vision, et SFR, avec Geostatistics, proposent une offre commerciale avec des traitements à façon de ces données.

La précision de la localisation dépend fortement de la localisation des antennes.

En milieu urbain, la taille d'une cellule associée à une antenne est de quelques centaines de mètres de diamètre ; en milieu moins dense, le diamètre d'une cellule peut s'établir à plusieurs kilomètres, en particulier si le maillage d'antennes est diffus. Une cellule peut ainsi couvrir plusieurs communes.

Le maillage ne correspondant pas aux périmètres administratifs desquels relèvent les données sociodémographiques, **les opérateurs doivent effectuer des traitements statistiques** mobilisant la propagation du signal **ce qui induit un biais de localisation**. En ville, l'importante densité de cellules, dont la propagation n'est pas immuable (leur couverture varie en fonction d'éléments extérieurs comme le relief ou la présence de plans d'eau), rend complexe l'affectation d'un événement de localisation à une zone en particulier.

Les traces sont dépourvues d'informations sur les personnes. Il est donc indispensable de les enrichir par des données sociodémographiques. Pour cela, les

Les différents types de déplacement

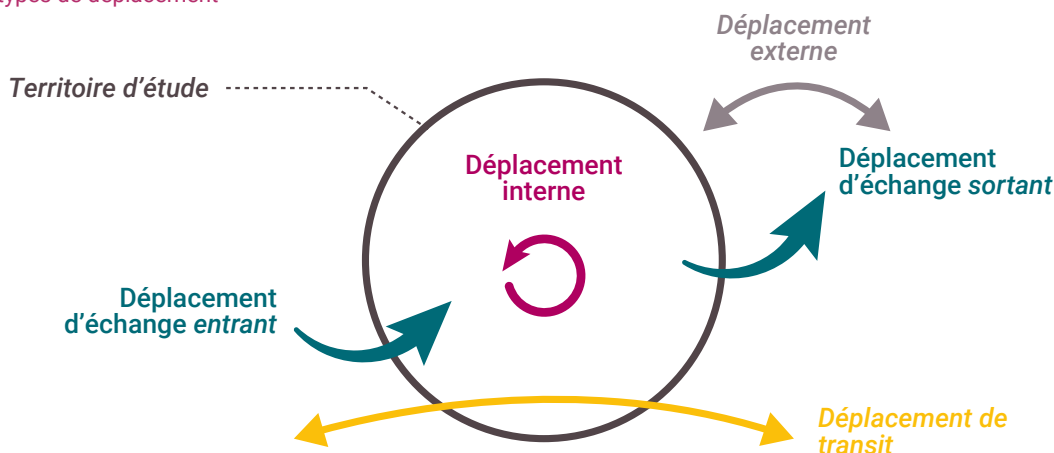
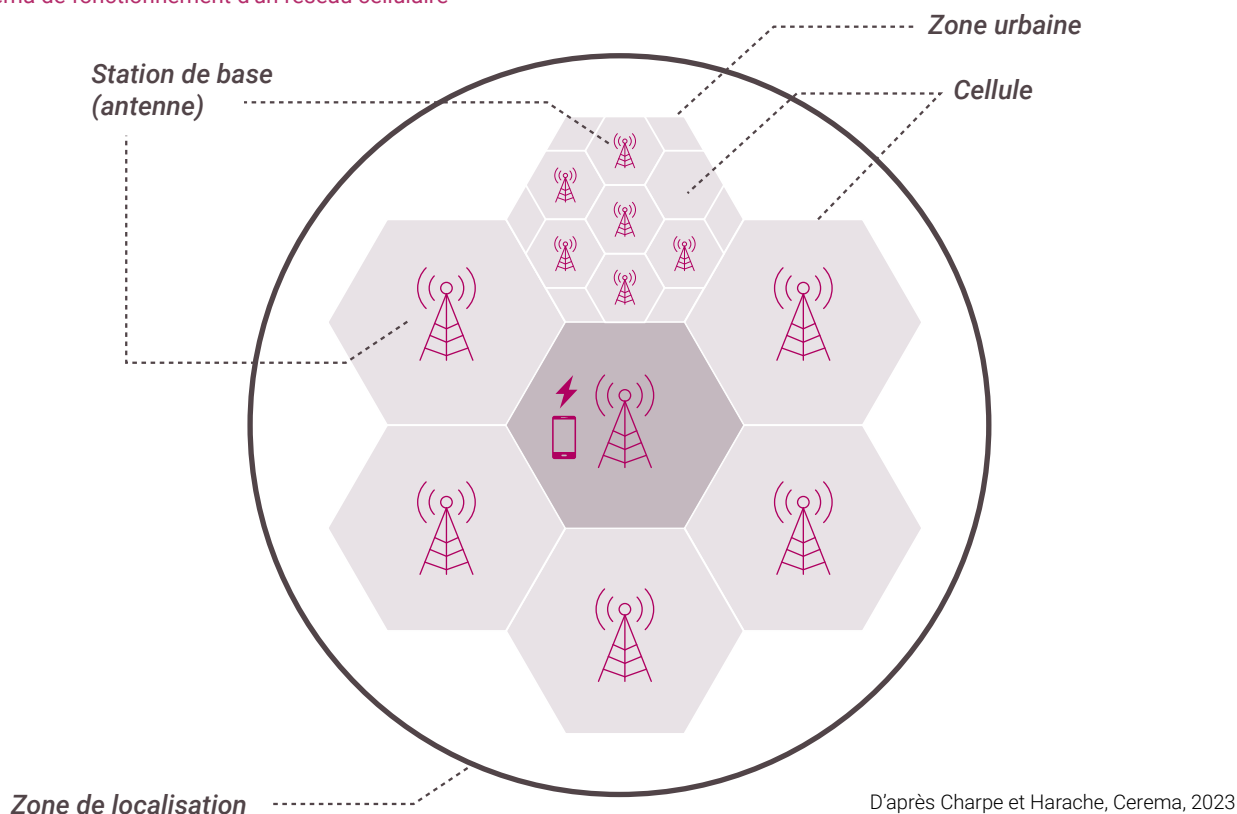


Schéma de fonctionnement d'un réseau cellulaire



D'après Charpe et Harache, Cerema, 2023

opérateurs utilisent en partie les données de facturation. Cependant, leur fiabilité interroge (suivi des changements d'adresse, cas des lignes contractées pour des tiers, etc.). Un algorithme attribue un lieu de domiciliation en se basant par exemple sur la récurrence de localisation pendant la nuit. **Ces traitements sont propres à chacun des opérateurs, sans uniformisation.**

La détermination des modes de transport à partir de ces données est très difficile compte tenu de leur faible précision géographique. Cela ne peut se faire que dans quelques cas : lorsque les infrastructures sont bien dissociées spatialement (autoroute et chemin de fer par exemple). Mais, la plupart du temps, cette distinction ne peut être établie. La distinction des modes selon la vitesse est également très difficile : en ville, la moyenne de déplacement d'un cycliste, d'un automobiliste ou d'un autobus est sensiblement la même (15 à 25 km/h). Concernant les motifs, seuls le travail et les études peuvent actuellement être déterminés a posteriori en s'appuyant sur la base de la récurrence de localisation d'un alias à une période donnée (en journée notamment).

Les téléphones éteints et le multi-équipement induisent un premier biais dans la donnée, même si les opérateurs assurent traiter cette problématique. Aucune donnée fournie ne permet d'identifier les personnes

car elles sont anonymisées et agrégées. Le taux de pénétration des opérateurs chez les personnes âgées, moins équipées en téléphonie mobile, questionne notamment la représentativité de ces données.

D'autres biais sont liés à des problématiques techniques : fonctionnement complexe et évolutif du réseau téléphonique, définition

de zonage, échantillonnage temporel irrégulier et hétérogène (les téléphones se connectent plus souvent en journée, lorsqu'ils sont utilisés, produisant davantage de données brutes à ce moment).

En résumé - Synthèse des atouts et limites des données FMD

+	-
Données disponibles à l'échelle nationale avec une précision temporelle et spatiale satisfaisante	La précision de la localisation dépend du maillage des antennes
Données disponibles rapidement et à une fréquence importante	Méthodes de redressement et de traitements statistiques opaques et propres à chaque opérateur et à leur part de marché qui varie d'un territoire à l'autre
Bonne représentativité de la population ; les étrangers sont notamment inclus mais de manière variable selon leur origine	La qualification des données (détermination des motifs, des modes, des caractéristiques sociodémographiques) est limitée
Possibilité de réaliser des comparaisons saisonnières durant une même année	Complexité de suivi dans le temps, en particulier le temps long, et nécessité de travailler auprès du même opérateur pour limiter les biais méthodologiques
Données à coût limité si le périmètre demeure contenu	Données fermées et payantes

Les données d'applications mobiles (SDK)

Comment sont-elles produites ?

Les téléphones mobiles contiennent de nombreux dispositifs qui permettent de connaître sa localisation et ses déplacements : antenne, puce GNSS, accéléromètre, gyroscope, etc.

Certaines applications installées sur les téléphones recueillent automatiquement ces informations géolocalisées (données GNSS). **Leurs bases comptent ainsi plusieurs millions d'utilisateurs actifs.** Les applications sont grand public et pas seulement dédiées à la mobilité : information régionale, météorologie, jeux, vie quotidienne, etc. Ces données sont apparues plus récemment et sont donc caractérisées par des questionnements importants quant à leur fiabilité. Ainsi, les bases de ces opérateurs comptent trois à quatre millions d'utilisateurs actifs en France

Ces données brutes correspondent à une série de coordonnées GNSS captées tout au long de la journée et associées à un identifiant de terminal. Par le biais de **traitements algorithmiques**, elles sont ensuite converties en **trajets** permettant de reconstituer des **déplacements** porte-à-porte en origine-destination (voir ci-dessous).

Certaines applications installées sur les téléphones recueillent automatiquement ces informations géolocalisées (données GNSS).

Traitements algorithmiques

La Cnil définit un algorithme comme « une suite finie et non ambiguë d'instructions permettant d'aboutir à un résultat à partir de données fournies en entrée. » Ils sont notamment utilisés par des administrations pour aider à la décision (classement de dossiers prioritaires, ciblage de candidatures, etc.) voire pour automatiser la prise de décision (calcul de l'impôt de l'ensemble des foyers fiscaux). Ils sont aussi utilisés par des entreprises du numérique (moteurs de recherche, réseaux sociaux, tarification dynamique en fonction de la demande, etc.).

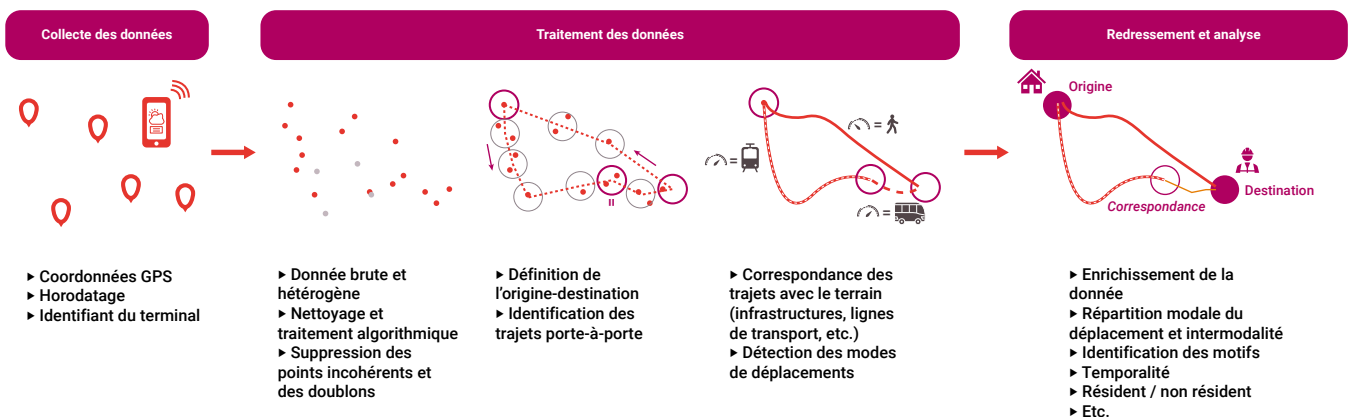
L'usage d'algorithmes par le secteur public soulève de nombreux enjeux car ils sont censés opérer au service de l'intérêt général. En ce sens, les algorithmes publics sont des formes de l'action publique et sont, à ce titre, soumis à la même forme d'exigence de redevabilité. Les administrations qui les mettent en œuvre doivent ainsi pouvoir rendre des comptes sur leur usage. La loi pour une République numérique et plus récemment le RGPD (règlement général sur la protection des données) ont introduit de nouvelles dispositions les concernant. Elles visent à introduire une plus grande transparence de l'administration dans l'usage de ces systèmes, en particulier quand ils sont utilisés pour prendre des décisions.

Par exemple, dans le cas des données de téléphonie, les entreprises utilisent des algorithmes de manière automatisée pour traiter des données brutes géolocalisées et les transformer en données de mobilité (voir Figure 2).

Si les algorithmes sont élaborés par des humains, de nouvelles techniques ont été développées depuis les années 2000 avec l'émergence des données massives et les progrès de l'intelligence artificielle. L'apprentissage automatique (machine learning) donne aux ordinateurs la capacité d'améliorer leurs performances dans la résolution de tâches à partir des données qu'ils traitent et par l'intermédiaire de modèles mathématiques. L'apprentissage profond (deep learning) est une méthode apparentée qui met en œuvre des algorithmes en utilisant des réseaux de neurones artificiels pour résoudre des problèmes complexes (définitions de la Cnil).

L'opacité des algorithmes varie donc en fonction de leur complexité mais aussi des choix des entreprises qui les développent, dans le but de garantir le secret industriel. Certains sont caractérisés par une grande transparence dans l'interprétation des résultats (boîte blanche), tandis que d'autres ne permettent qu'une transparence limitée (boîte grise). Si elle augmente leur performance, l'intelligence artificielle opacifie aussi les ressorts fonctionnels qui motivent la décision prise par l'algorithme (boîte noire), tendant à reléguer l'expert au rôle de facilitateur technique (Ermans et al., 2020).

D'une donnée brute à une donnée enrichie, les étapes de reconstitution des déplacements



Quelles différences avec les données FMD ?

Les usages, les intérêts, les limites et les biais sont sensiblement les mêmes que ceux des données issues des antennes (FMD). Cependant, les données SDK sont plus précises que celles issues des antennes (5 à 10 mètres). Elles permettent plus facilement d'identifier les infrastructures sur lesquelles les individus se déplacent et, moyennant des analyses fines, les pratiques de rabattement et d'intermodalité.

Les traces issues des applications sont donc caractérisées par une **sous-représentativité des personnes les plus jeunes et les plus âgées** (moins de 12 ans et plus de 70 ans) qui possèdent et utilisent moins le téléphone et ces applications. L'activation de la localisation étant indispensable, il existe un biais très peu documenté en lien avec l'acceptabilité et les types d'usagers qui activent ou non cette option intrusive. Il existe aussi un risque plus grand de perte de données liées à l'autonomie de la batterie ou à la désactivation automatique de la localisation. Enfin, le bon fonctionnement du dispositif nécessite que les applications soient mises à jour régulièrement.

Les données numériques véhiculaires (FCD)

La connaissance des conditions de trafic est indispensable à la gestion des réseaux routiers. Les dispositifs fixes traditionnels de connaissance du trafic (boucles et stations de recueil de données) sont complétés par d'autres dispositifs comme la surveillance vidéo ou les systèmes de détection automatique d'incidents. Depuis plusieurs années, **un nouveau marché de données de trafic se dessine** à travers les informations fournies par les FCD.

Les Floating Car Data⁹ (« données flottantes automobiles » ou « données de véhicules traceurs ») regroupent l'ensemble des **informations numériques issues des véhicules en circulation** et provenant des appareils connectés et embarqués : ordinateur de bord, terminaux de navigation, applications mobiles, etc. Ces objets connectés laissent derrière eux une somme importante de sous-produits numériques (identifiant du véhicule, horodatage, localisation, vitesse, cap, etc.), des données qui, en premier lieu, servent à fournir une **information en temps réel aux véhicules en circulation** (temps de parcours, incidents, etc.). Également, une fois agrégées et enrichies par les différentes sociétés (TomTom, Coyote, Waze, Autoroutes Trafic, etc.), ces données fournissent des informations générales sur les **conditions**

En résumé - Synthèse des atouts et limites des données SDK

+	-
Données disponibles à l'échelle nationale avec une précision temporelle et spatiale satisfaisante	Méthodes de redressement et de traitements statistiques opaques et propres à chaque opérateur et à leur part de marché qui varie d'un territoire à l'autre
Données disponibles rapidement et à une fréquence importante	La qualification des données (détermination des motifs, des modes, des caractéristiques sociodémographiques) est limitée
Bonne représentativité de la population ; les étrangers sont notamment inclus mais de manière variable selon leur origine	Complexité de suivi dans le temps, en particulier le temps long, et nécessité de travailler auprès du même opérateur pour limiter les biais méthodologiques
Possibilité de réaliser des comparaisons saisonnières durant une même année	Données fermées et payantes
Données à coût limité si le périmètre demeure contenu	Sous-représentation des personnes les plus jeunes et les plus âgées

de trafic et les origines-destinations. À l'instar des données issues de la téléphonie mobile, les informations recueillies nécessitent d'être purgées et assemblées par des **traitements algorithmiques propres à chaque opérateur et protégés par le secret commercial** (Ermans et al., 2018).

Des sources et des méthodes de production hétérogènes

Il existe plusieurs méthodes de collecte des FCD, selon le type de dispositif embarqué :

- le **positionnement GNSS** : localisation à intervalles réguliers des systèmes embarqués (dont les téléphones) par les satellites ou les antennes-relais (cf. Figure 2) ; il s'agit de la méthode la plus courante ;
- la **capture d'adresses Bluetooth** : des stations ad hoc réparties à intervalles réguliers le long des axes routiers et équipées de capteurs détectent les adresses Bluetooth des véhicules en circulation ; les informations recueillies permettent de calculer la vitesse moyenne des véhicules, leur temps de parcours et leur nombre ;
- les **infrastructures de péage** (radio-identification) : elles sont équipées de bornes capables de communiquer avec un appareil embarqué (télépéage) et donc de faire remonter de l'information ;
- le **positionnement Wi-Fi** : les bornes Wi-Fi publiques, notamment installées en milieu urbain, détectent les terminaux qui s'y connectent (téléphones, ordinateurs) ; il est alors possible de retracer leur parcours.

Pour constituer leur base de données, des entreprises agrègent des informations issues de ces différentes méthodes. C'est le cas de TomTom qui associe les données des terminaux de navigation classiques, des terminaux connectés, de ceux embarqués et des applications mobiles. La société commercialise ainsi deux types de bases de données (en temps réel et en temps différé).

Au-delà des biais déjà présentés, la principale limite de cette source est son caractère monomodal.

Des usages ciblés pour connaître et anticiper le trafic routier

Par rapport aux outils de comptage classiques ponctuels utilisés pour évaluer le trafic, les données FCD présentent plusieurs intérêts : recueil continu, historicité du recueil, couverture spatiale.

Les données en continu permettent de **superviser le trafic en temps réel** afin de connaître l'état de fluidité sur l'ensemble des sections d'un réseau routier grâce à la vitesse moyenne de déplacement des véhicules. Ces données en continu sont également utilisées pour alerter en cas d'incident et intervenir rapidement. Certaines collectivités diffusent cette information au grand public et mettent en place des alertes automatisées en cas de perturbation du trafic.

Les données historicisées servent, quant à elles, à la réalisation d'**études de trafic** : congestion, temps de parcours moyen, origines-destinations, comptage TMJO (taux moyen jours ouvrés), diagrammes de flux de circulation.

9. On parle aussi de Floating Vehicle Data

Les données de Waze

Waze est une application mobile d'aide à la navigation et d'assistance à la conduite reposant sur une production participative de ses utilisateurs. L'entreprise qui développe cet outil appartient à Google depuis 2013. La particularité de l'application est que les utilisateurs peuvent y déclarer les événements de circulation (congestion, accidents, travaux, contrôles de police). Cette application est régulièrement critiquée car certaines de ses recommandations incitent les conducteurs à circuler sur les axes secondaires afin d'éviter les embouteillages, causant des nuisances dans les communes et sur les routes qui ne sont pas calibrées pour accueillir un nombre important d'automobiles.

À l'instar des autres applications et systèmes de navigation par satellites, Waze recueille une masse considérable de données de circulation. Plusieurs agences ont expérimenté l'usage de ces données afin de mesurer la congestion routière et les irrégularités de trafic, à l'instar de l'Agorah, l'agence d'urbanisme de la Réunion. A travers son service Waze for Cities, la société américaine peut fournir des données de trafic aux collectivités, aux services de l'État ou aux établissements universitaires. En échange, Waze récupère des éléments à propos des fermetures de routes (dans le cadre de travaux), des incidents de trafic ou des événements programmés.

Les données peuvent notamment permettre d'identifier en temps réel les incidents de circulation, mais aussi de qualifier les fluctuations temporelles de circulation (entre les périodes de vacances et hors vacances par exemple), les impacts d'événements ponctuels, etc. La précision de l'information permet d'observer quotidiennement et rue par rue la longueur et la durée moyennes des embouteillages sur le réseau routier. Cependant, cette information est produite par l'algorithme interne de Waze dont le fonctionnement est opaque. Ainsi, comme pour toutes les données issues de la téléphonie mobile, les résultats sont à interpréter avec précaution.

<https://www.agorah.com/index.php/2023/08/25/la-data-waze-outil-de-mesure-de-la-congestion-routiere/>

Conclusion sur les mégadonnées

Ces données émergentes présentent de nombreux biais. Avec l'usage, certains sont progressivement levés ou atténués. Pour beaucoup d'entre eux, ils sont intrinsèques à la donnée elle-même. En effet, **ces traces ne sont pas produites pour réaliser des analyses sur la mobilité**, contrairement aux données issues du recensement ou des enquêtes déplacements. Elles sont engendrées automatiquement par notre usage des appareils connectés. Elles nécessitent de nombreux traitements algorithmiques et de redressements. Il existe donc un biais lié au choix de l'opérateur. Par ailleurs, il n'existe à l'heure actuelle pas de jeux de données croisées provenant de plusieurs opérateurs. Également, **ces données sont produites par des entreprises privées qui protègent ces méthodes de redressement** par le secret industriel. Ainsi, **la donnée livrée est opaque.** En particulier, la généralisation de la donnée à l'ensemble de la population demeure encore peu

documentée. D'autant que les techniques qui permettent la captation de ces traces ne cessent d'évoluer et de se diversifier, de même que les populations qui les utilisent. Cela implique une certaine instabilité du processus de production des indicateurs qui questionnent sur la possibilité de réaliser des séries temporelles.

Cependant, **de nombreux travaux de recherche ont démontré l'intérêt de ces données dans la compréhension de la mobilité**, en particulier les données de la téléphonie mobile qui présentent l'avantage d'être produites de manière massive.

Plusieurs experts se sont notamment prêtés au jeu de comparaison entre les données de téléphonie mobile et celles produites à travers les méthodes classiques (enquêtes déplacements, comptages routiers, etc.). La plupart des rapports apporte la conclusion que ces données émergentes présentent un **potentiel réel pour comprendre la structure des déplacements** sur les plans spatial, temporel, de durée, de distance. Une comparaison des données de signalisation

commercialisées par Orange (FMD) avec l'enquête globale transport de l'Île-de-France (2010) a permis de mesurer un écart de 9 % en volume du nombre de déplacements (Bonnell et al., 2017). Cependant, ces écarts peuvent atteindre 70 à 80 % sur certains couples origine-destination. Plus récemment, une étude de fiabilité des données GPS a été réalisée pour la métropole d'Aix-Marseille par le Laboratoire aménagement économie transports (Laet) en s'appuyant sur les résultats de l'EMC2 de 2019. Les conclusions pointent un écart de volume de 9 % avec, localement, des écarts pouvant être plus conséquents. Si la répartition modale apparaît correcte, la marche est sous-estimée tandis que les parts de la voiture et des transports collectifs sont surestimés (Meyer, 2022).

Cet exercice de validation des données par une méthode classique semble encore aujourd'hui indispensable, ce qui permet de **tempérer le caractère abordable des solutions** proposées généralement « clef en main » par les entreprises spécialisées.



Tenter de faire parler toutes ces traces numériques hétérogènes, c'est se confronter aux techniques multiples de transformation des données en informations et en savoirs, en cherchant à ouvrir les boîtes noires algorithmiques dont elles sont issues. »

B. Merickskay, 2019

Les appellations des données massives de mobilité

Type de données massives	Autre appellation	Méthode de recueil	Exemples de fournisseurs
Donnée numériques cellulaires	FMD	Passive (antennes)	Flux Vision (Orange), Geostatistics (SFR)
Données numériques véhiculaire	FCD	Passive (boîtier embarqué)	Coyote, TomTom
Données d'application mobile		Active (application dédiée)	Geco Air, Modum
Données d'application mobile	SDK	Passive (autres applications)	Single Spot, Hove (Keolis)

Source : Les Rendez-Vous Mobilités, Cerema, mai 2024

Partie 3

Les données à représentativité forte et faisant l'objet d'une collecte passive

Ces données sont obtenues par une collecte passive et exhaustive ou quasi exhaustive et permettent de dénombrer des flux, des véhicules ou des personnes en un point précis d'un réseau (routier ou de transport collectif). Les données produites sont pauvres en informations individuelles, à moins de coupler et d'enrichir la collecte grâce à une autre méthode. Ce sont des données généralement utilisées pour caler les modèles de déplacement.

Les données de billettique des opérateurs de mobilité

Les opérateurs de mobilité produisent une quantité importante de données qui remontent de l'usage des services de transport : autobus, tramway mais aussi vélos en libre-service, etc. Contrairement aux informations statiques ou dynamiques sur l'offre, les données liées à l'usage (fréquentation, statistiques, finances) sont plus rarement diffusées car elles dépendent de la volonté de transparence des AOM et des opérateurs de transport. Pourtant ces informations contribuent à mieux connaître les usagers ; en effet, elles permettent l'analyse des profils dans l'espace et dans le temps mais aussi de reconstituer une partie significative des trajets ou de faire des prévisions.

Le trafic routier : comptage et comptage automatisé

Le comptage routier répond à deux principaux objectifs : la mise en œuvre et le suivi des politiques routières (conception, entretien, sécurité, exploitation) et la fourniture de données statistiques qui rendent compte de la demande en matière de déplacement des personnes et des marchandises. Les données de trafic routier ont ainsi plusieurs usages : dimensionnement des chaussées, des ouvrages, du stationnement, évaluation de la pollution, évaluation des reports de trafic, programmation de la maintenance du réseau, etc. Il est ainsi nécessaire de produire régulièrement des données fiables et précises qui permettent aux gestionnaires routiers d'assurer leur mission d'exploitation.

Les compteurs de trafic permanents ou ponctuels sur les routes et autoroutes s'appuient sur différentes techniques : boucles électromagnétiques, capteurs magnétiques, tuyaux pneumatiques, caméras, etc. Ils vont notamment permettre d'établir le trafic moyen journalier annuel (TMJA), de connaître la répartition entre véhicules légers et poids-lourds, d'estimer le débit par heure et le taux d'occupation de la chaussée, etc. Une fois agrégées en séquences horaires ou journalières, l'utilisation la plus courante des données est réalisée en temps différé à des fins statistiques (observatoire, tableau de bord, etc.). Les données sont également utilisées au fil de l'eau par les gestionnaires afin de connaître en temps réel ou légèrement

différé le débit, la vitesse, le taux d'occupation de la chaussée. Cela permet également d'afficher de manière dynamique les temps de parcours ou les problèmes ponctuels (congestion).

La lecture automatisée des plaques d'immatriculation (Lapi)

La lecture automatisée des plaques d'immatriculation a été initialement développée pour la verbalisation des véhicules. À l'aide de caméras, ce dispositif détecte les véhicules et les visualise afin de lire leur plaque grâce à une technique de reconnaissance optique des caractères. Avec le temps, de nouveaux usages sont apparus et cette méthode constitue désormais une source de données qui peut être utilisée afin de modéliser les déplacements, d'analyser le trafic, etc.

La détection d'une plaque ne signifie pas nécessairement que l'information est exploitable (mauvaise lecture). Néanmoins, la détection sert d'abord à connaître le débit apparent. Les unités observées sont les véhicules disposant d'une plaque à l'avant : véhicules légers, véhicules utilitaires légers, poids-lourds, autobus et autocars. L'horaire de passage, le type de véhicules et la plaque d'immatriculation sont les variables produites.

Cela permet de connaître le nombre de véhicules par catégorie et, lorsqu'un dispositif ad hoc a été mis en place, les temps de parcours et les origines-destinations (points d'entrée et de sortie du réseau).



Station gare Ti Vélo à Landerneau - Crédit : Romainbehar

En post-traitement, des logiciels peuvent exploiter la donnée et fournir des informations supplémentaires en croisant avec les informations du système d'immatriculation des véhicules : marque, modèle, puissance, date de mise en circulation, classe environnementale, etc. Cela nécessite un relevé intégral des plaques et la mise en place d'une procédure spécifique avec le ministère de l'Intérieur (garantissant l'anonymat des données).

Les principaux biais tiennent dans le défaut de lecture des plaques : en cas de fort trafic, certains véhicules seront masqués ; de même, un véhicule dont la plaque n'aurait pas été lue en entrée ou en sortie sera ainsi considéré comme en transit. Les caméras peuvent aussi enregistrer deux fois de suite le même véhicule ou bien détecter d'autres signes alphanumériques sur un véhicule et le comptabiliser une autre fois. L'apurement de la base brute permet d'éliminer une partie de ces éléments avant l'application de l'algorithme de reconnaissance des véhicules. Enfin, un travail humain peut être nécessaire pour contrôler cette étape algorithmique (vérification visuelle des images).

Le comptage des usagers des modes actifs : cyclistes et piétons

À l'instar du comptage des véhicules motorisés, le comptage des cyclistes et piétons est indispensable pour préparer et évaluer un projet urbain ou d'aménagement piéton et cyclable, pour connaître la pratique et assurer la sécurité des usagers. Le comptage est effectué aussi bien dans les espaces denses des centres urbains que sur les axes périurbains et ruraux, en particulier ceux à destination des déplacements de loisir.

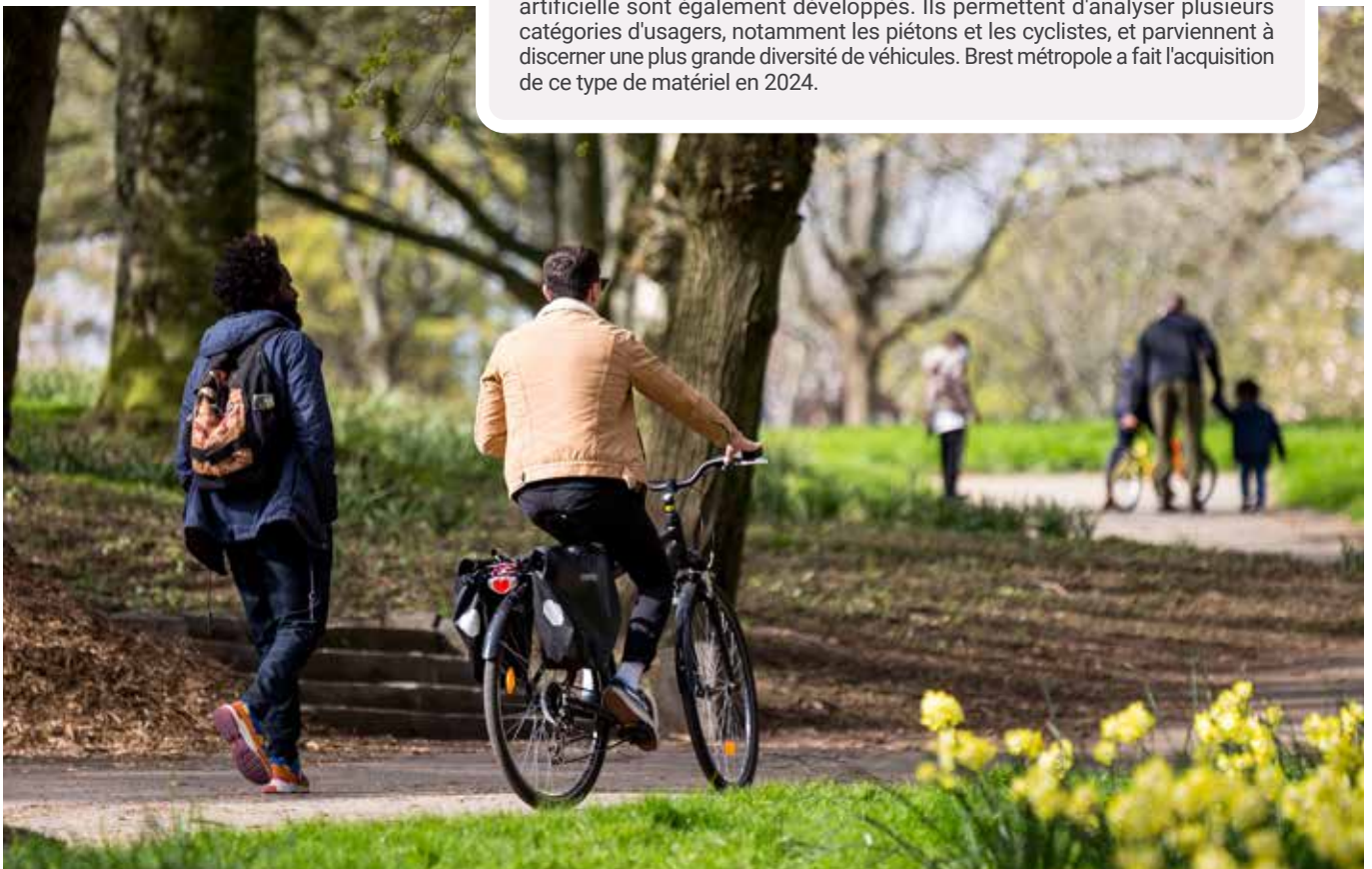
Le comptage des cyclistes a en particulier connu un essor important en 2020 et durant les années suivantes, alors que la pandémie de coronavirus avait accéléré le développement de l'usage du vélo en France. À Paris notamment, le comptage des cyclistes sur certains axes récemment aménagés, à l'instar de la rue de Rivoli, revêt une dimension politique qui dépasse le seul cadre de la production de données.

Comme pour la mesure du trafic routier, il existe plusieurs techniques de comptage des cyclistes, que l'on peut classer selon leur caractère intrusif ou non : les capteurs intrusifs dans la chaussée sont permanents (boucles électromagnétiques, capteurs piézoélectriques, magnétomètres, fibres optiques), tandis que les capteurs dits non-intrusifs sont permanents ou temporaires (tubes pneumatiques, radars, capteurs infrarouge, analyse vidéo, etc.).

Le comptage des piétons constitue également un enjeu d'aménagement de l'espace, d'autant que ce mode de déplacement est parfois le grand oublié des politiques de mobilité. Difficile à mesurer, la marche est pourtant un mode majeur, en ville comme dans les espaces ruraux : l'évaluer permet ainsi de comprendre, par exemple, les ressorts de l'attractivité commerciale d'un centre-ville.

L'association Vélo et territoires met à disposition des données de comptage des cyclistes sur la plateforme nationale des fréquentations, qui rassemble plus de 1 700 compteurs en France. Afin d'harmoniser la production, la diffusion et l'usage des données de comptage, un standard de données a été élaboré en lien avec la Direction interministérielle du numérique.

Enfin, des dispositifs de comptage par analyse vidéo mobilisant l'intelligence artificielle sont également développés. Ils permettent d'analyser plusieurs catégories d'usagers, notamment les piétons et les cyclistes, et parviennent à discerner une plus grande diversité de véhicules. Brest métropole a fait l'acquisition de ce type de matériel en 2024.



Crédit : Nacer Hammoui - Brest métropole

Partie 4

Les informations sur l'offre de mobilité : des jeux de données très hétérogènes

Le domaine de la mobilité et le secteur des transports brassent d'autres informations numériques dont il est difficile d'établir un portrait exhaustif. Dans cette dernière partie, nous présentons donc une sélection de données parmi les plus utilisées – et peut-être les plus utiles – pour l'élaboration de projets et le suivi de politiques publiques.

Représenter l'offre de transports collectifs

Les données relatives aux réseaux de transport (description géographique et qualification de l'offre de service) sont

indispensables afin de transmettre aux usagers les informations nécessaires pour réaliser leurs déplacements.

Trois catégories d'informations statiques sont de première nécessité :

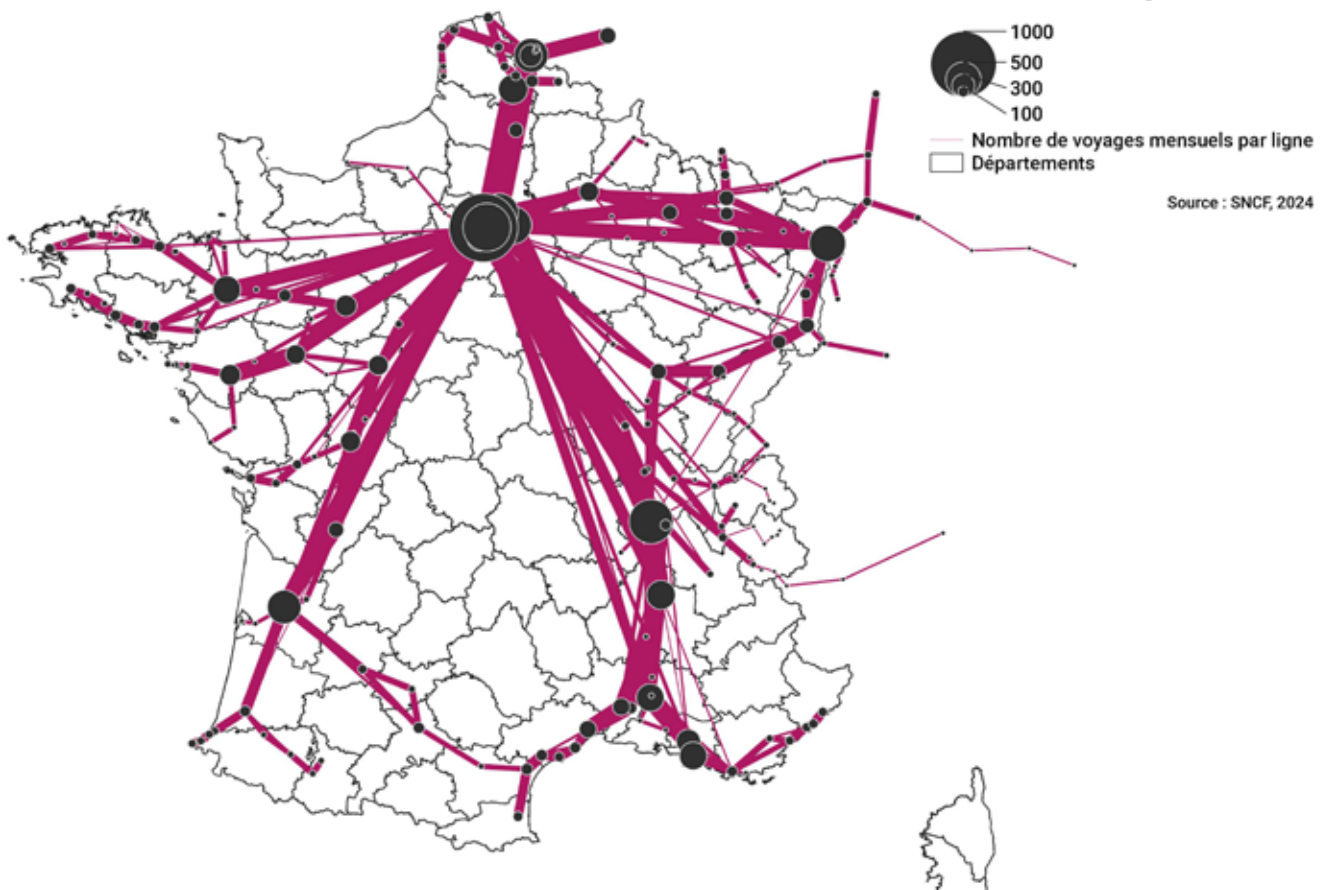
- les arrêts : localisation des points d'arrêts des lignes avec leurs attributs principaux (libellé, sens, nom des lignes le desservant), identification des correspondances éventuelles, qualification de l'accessibilité pour les personnes à mobilité réduite ;
- les lignes : géographie des lignes de transport en commun (libellé, sens), distinction des modes de transport

(tramway, bus) ;

- l'offre de service : qualification de l'offre (horaires journaliers, précision sur les services de semaine, du week-end), informations tarifaires (prix du tickets, types d'abonnements, etc.), information commerciale (nom de la compagnie qui opère le réseau).

En parallèle, des informations dynamiques (en temps réel) peuvent également être diffusées. Il s'agit notamment des horaires effectifs aux arrêts (prenant en compte les retards liés à la circulation par exemple) et des perturbations en cours.

Offre théorique de voyages en TGV



Les formats GTFS et Netex

Afin de faciliter la diffusion et la consultation ergonomique de ces informations, et pour permettre aux développeurs de créer des outils d'information multimodale et des calculateurs d'itinéraires, des formats informatiques standardisés ont été progressivement élaborés.

Deux d'entre eux sont largement diffusés pour publier les données statiques relatives aux transports en commun : le format **GTFS** (General Transit Feed Specification, spécification générale pour les flux relatifs aux transports en commun) et le format **Netex** (Network Timetable Exchange). Ces deux formats sont dits statiques ; ils coexistent et sont complémentaires.

Le format GTFS a été initialement conçu au sein des services de l'autorité organisatrice des transports de Portland au milieu des années 2000 en collaboration avec Google. Devenu libre en 2009, il s'est généralisé depuis lors et est utilisé par des dizaines de milliers d'agences de transport à travers le monde. Aux États-Unis, il est considéré comme le standard de base. En Europe, s'il est massivement utilisé par des AOM, il n'a pas d'existence officielle.

Ce standard se compose de plusieurs flux, c'est-à-dire des fichiers textes dont chacun modélise et définit un aspect spécifique des informations de transports en commun. La plupart des flux intégrés sont les suivants :

- stops : liste et localise tous les points d'arrêt ;
- stops_times : présente les horaires des courses aux points d'arrêt ;
- agency : regroupe les informations sur le réseau de transport (compagnie, nom) ;
- calendar et calendar_dates : contiennent le calendrier de circulation ;
- trips : détaille les courses sous la forme d'une table de liaison ;
- shapes : permet le tracé d'une route sur une carte ;
- transfers : présente les correspondance entre plusieurs points d'arrêts.

Le format Netex est défini au niveau européen par le Comité européen de normalisation. Il permet l'échange informatique d'horaires de transports publics et des informations associées. Plus riche que le standard GTFS, il met

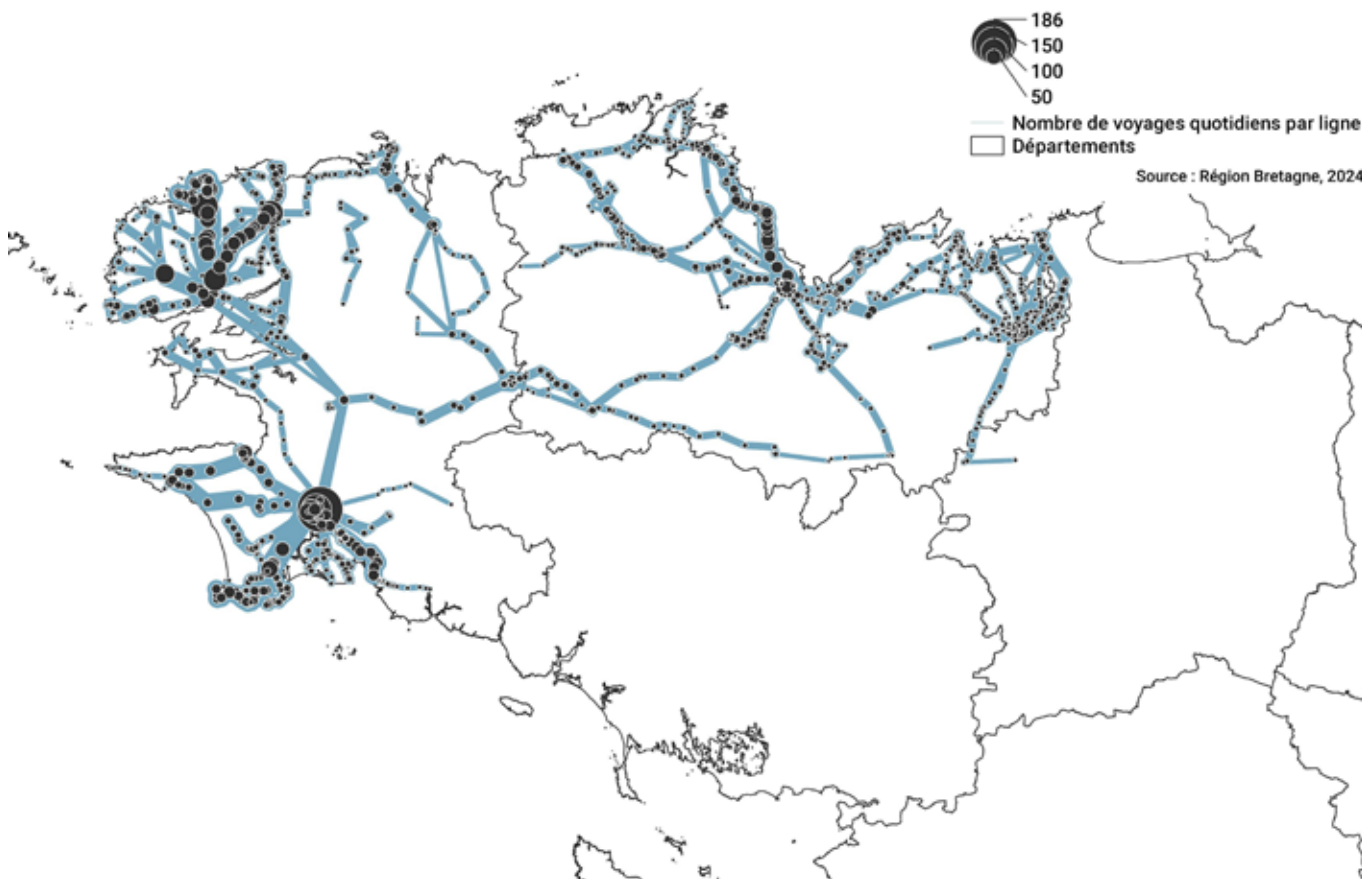
en place des règles communes pour assurer la diffusion de l'ensemble de ces données entre les opérateurs de transports, les collectivités, les développeurs d'applications. Il permet également l'interopérabilité entre les différents services. Il a vocation à devenir, à terme, le format de référence pour l'échange de données d'offre théorique de transport collectif.

Les données respectant le format Netex sont contenues dans un seul fichier .xml appelé profil qui définit l'ensemble d'un réseau : arrêts, lignes, horaires, tarifs, accessibilité, etc.

En France, c'est le format Neptune (Norme d'échange profil transport collectif utilisant la normalisation européenne), émanation de Netex, qui fait référence pour l'échange de données d'offre théorique de transport collectif.

En résumé, la complémentarité des deux formats tient dans leur destination et leur position dans le processus de production et de diffusion de la donnée : en amont, le format Netex est utilisé pour produire, affiner et intégrer la donnée dans un seul

Offre théorique de voyages en autocars BreizhGo (Finistère et Côtes-d'Armor)





Crédit : Franck Bétermin - Brest métropole

et même fichier ; le format GTFS, plus simple, couvre, quant à lui, seulement les informations suffisantes pour répondre aux requêtes de base de la planification de trajets au sein de plusieurs fichiers dans un dossier compressé. Enfin, à partir de Netex, il est possible de produire un ensemble de données GTFS (l'inverse n'étant pas possible).

Outre l'information des voyageurs, l'intérêt de telles données est par exemple de pouvoir cartographier l'offre effective de transport dans un territoire et d'observer les éventuels déséquilibres spatiaux.

Les données de fréquentation des gares de la SNCF

Le portail de ressources SNCF Open Data met à disposition plus de 200 jeux de données, comme la liste des gares ou des passages à niveaux, les horaires mais aussi des informations en lien avec leurs offres tarifaires. Une autre information est particulièrement utilisée dans le domaine de l'observation territoriale : la fréquentation en gares. Ce jeu de données représente la fréquentation annuelle de l'ensemble des 3 000 gares voyageurs depuis 2015 en France. Ces informations sont notamment issues des données de billetterie pour le

trafic national et régional hors Île-de-France. La méthodologie de calcul se précise année après année avec une reprise de l'historique pouvant expliquer des variations dans les données publiées d'une année sur l'autre.

Cette information permet de qualifier l'usage des gares et d'envisager des solutions afin de renforcer leur attractivité (stationnement, intermodalité, adaptation du réseau de transport local).

Les données portant sur le vélo : la nécessaire standardisation de la donnée

Depuis plusieurs années, le déploiement des aménagements cyclables (pistes mais aussi stationnement) s'accélère et la pratique du vélo progresse. C'est pourquoi l'ouverture des données sur les aménagements cyclables nécessaires à l'information voyageur est une obligation européenne. Car, dans le même temps, des applications de calcul d'itinéraires cyclables sont produites pour aider les cyclistes dans leurs déplacements du quotidien ou de loisir. Enfin, un autre domaine en pleine expansion est celui des vélos en libre-service, dont l'offre est opérée par une multitude d'entreprises et nécessite la diffusion de données en temps réel.

Cependant, l'hétérogénéité des aménagements, tant dans leur forme que leur qualité, demeure un obstacle important. C'est pourquoi une base nationale des aménagements cyclables (BNAC), produit de la collaboration entre le point d'accès national et l'association Vélo et territoires, encadre l'harmonisation des données de l'ensemble du territoire national. La base fournit un standard minimal de complétude des données, avec les informations indispensables à fournir.

Les informations géographiques sont notamment consultables sur le projet collaboratif de cartographie OpenStreetMap, sur l'application GeoVelo ou encore sur le système d'information géographique en ligne de l'association Vélo et territoires (Velodatamap).

Pour en savoir +



Politiques cyclables de l'Ouest breton : un état des lieux

Observatoire de la mobilité n° 8, Adeupa, mai 2021

En guise de conclusion

Vers une indispensable révolution méthodologique et épistémologique

L'apparition et l'usage de plus en plus généralisé de ces mégadonnées constituent un **enjeu nouveau pour les sciences sociales** et les **organismes d'observation des politiques publiques** à l'instar des agences d'urbanisme. En effet, l'engouement positif et l'émulation dont jouissent les mégadonnées et, plus généralement, les solutions issues d'innovations techniques – ce qui caractérise notamment l'enthousiasme autour des villes intelligentes (smart cities) – doivent nous interroger, et ce à plusieurs titres (Ermans, 2018).

Premièrement, **ces données massives nécessitent une maîtrise technique importante** et un traitement préalable à l'analyse qui ne sont pas négligeables. Cela implique d'incorporer, au sein des collectivités et des organismes parapublics, de nouvelles compétences méthodologiques. D'autant que, très souvent, ces données sont fragmentées (par opérateurs, par AOM, par questionnaires de réseaux, etc.) : pour en tirer profit de manière efficace, il est nécessaire de construire des services complets capables d'apporter une vision globale des mobilités.

Aussi, leur relative nouveauté, leur émanation d'outils techniques et leur caractère ubiquiste peuvent **produire un sentiment d'objectivité naturelle** : on parle de biais d'automatisation. Or, il convient alors d'être prudent dans le processus de passage du savoir à l'action. Car l'augmentation exponentielle de données numériques à disposition ne peut améliorer en soi la connaissance : les données étant construites et devant être interprétées, il demeure nécessaire de disposer d'une connaissance du domaine étudié (mobilité, logement, tourisme, etc.).

Enfin, la limite principale de ces mégadonnées tient dans leur **pauvreté en matière d'information sociodémographique**. Il est pourtant nécessaire de garder à l'esprit que la connaissance des enjeux de mobilité ne peut se réduire au prisme rationnel des flux de déplacement, à l'heure où la **nécessité d'adaptation des politiques publiques aux besoins de toutes et tous** paraît plus que jamais essentielle, en lien avec les enjeux sociaux que l'on connaît : vieillissement de la population, questions de genre, identité culturelle, handicap, etc.

Ainsi, plus largement, ces évolutions font émerger un **nouveau paradigme en matière de production de la connaissance** qui soulève également des enjeux épistémologiques.

En effet, en matière de déplacements, certains indicateurs sont trop souvent considérés par les chercheurs et les experts comme une information qui parle d'elle-même, comme une donnée qui permettrait de se détacher de toute idéologie. Or, nous nous déplaçons pour plusieurs raisons, et cela dépasse le simple fait de se rendre d'un point A à un point B. Au-delà du mouvement dans l'espace – qui est la première dimension de la mobilité – chacun de nos déplacements a une portée sociale, une signification qui lui est associée (marcher pour le plaisir, faire du vélo pour protéger l'environnement, covoiturer afin d'aider un proche, etc.) et pour laquelle on produit des récits.

Enfin, la mobilité est porteuse d'une troisième dimension : c'est la pratique, ou l'expérience du mouvement (conduire une voiture peut être vécu comme un moment crispant, prendre l'autobus comme un moment pour soi, etc.).

Et ces **trois dimensions sont imbriquées** ; nos pratiques interagissent avec les récits, s'y conforment ou s'y opposent (« je suis conscient des impacts de mes déplacements mais je n'ai pas d'autre choix que de prendre ma voiture »), le mouvement a un impact sur les pratiques (« je me rends chaque jour au travail donc je cherche à rendre ce déplacement agréable »), etc. La mobilité ne peut donc pas être réduite à une simple ligne tracée sur une carte.

Il convient alors de maintenir les **démarches d'analyse croisées**, hybridant les sources et les méthodes, mêlant les approches de plusieurs disciplines, tout en **laissant une place raisonnable à l'expérimentation et aux solutions émergentes**. Ces nouvelles données, une fois replacées dans la panoplie d'outils méthodologiques, constituent ainsi une **opportunité pour approcher une certaine exhaustivité** et bénéficier d'une **capacité de mobilisation** hautement plus importante que l'élaboration d'une enquête déplacements.

L'analyse des déplacements des personnes peut alors partir de la **compréhension de leurs pratiques**, ce qui permet de faire de l'individu la référence. La dimension sociale est alors considérée non plus seulement comme une simple caractéristique de l'individu mais comme une condition de production de son rapport à l'espace.

L'apparition et l'usage de plus en plus généralisé de ces mégadonnées constituent un enjeu nouveau pour les sciences sociales et les organismes d'observation des politiques publiques à l'instar des agences d'urbanisme.

Bibliographie

- ART, *État des lieux de l'ouverture et de l'utilisation des données de mobilité*, juin 2022.
- Patrick Bonnel, et al. « Apports et limites des données passives de la téléphonie mobile pour la construction de matrices origine-destination », *Revue d'Économie Régionale & Urbaine*, vol. n°4, 2017, pp. 647-672.
- Aurélie Bousquet, « Enquêtes origine-destination », *Données de mobilité pour la modélisation des déplacements*, Cerema, fiche n°1, novembre 2022.
- Aurélie Bousquet et Julie Tricoche, « Collecte et utilisation de données de mobilité pour la modélisation des déplacements : des enquêtes ménages-déplacements aux données massives », *Données de mobilité pour la modélisation des déplacements*, Cerema, fiche chapeau, novembre 2022.
- Alice Charpe et Julien Harache, « Données issues des antennes de la téléphonie mobile », *Données de mobilité pour la modélisation des déplacements*, Cerema, fiche n°3, février 2023
- Thomas Ermans, Céline Brandeleer et Michel Hubert, « L'apport des big data pour l'étude de la mobilité en Région de Bruxelles-Capitale : enjeux, opportunités et défis », *Note technique de l'Observatoire de la mobilité de la Région de Bruxelles-Capitale*, septembre 2018.
- Boris Mericskay, « Potentiels et limites des traces (géo)numériques dans l'analyse des mobilités : l'exemple des données de la plateforme de covoiturage BlaBlaCar », *Cybergeogeo*, document 889, avril 2019.
- Vincent Meyer, « Utilisation des données GPS de téléphonie mobile : un complément aux EMC2 ? », *Journée d'échanges sur la mobilité urbaine*, Cerema, octobre 2022 [URL].
- Christophe Nunes, Sylvie Deyris, Sébastien Pabion, Maria Tebar, « Connaître la mobilité grâce aux lectures de plaques d'immatriculation », *Données de mobilité pour la modélisation des déplacements*, Cerema, fiche n°5, mai 2024.
- Étienne Ollion et Julien Boelaert. « Au-delà des big data. Les sciences sociales et la multiplication des données numériques », *Sociologie*, vol. 6, n° 3, 2015, pp. 295-310.
- Wilfried Raballand, Maxime Le Corre, *Pertinence des données de téléphonie mobile pour les matrices origine-destination*, 45^e Congrès ATEC ITS France – Les Rencontres de la Mobilité intelligente, janvier 2018.

L'Adeupa tient à remercier les interlocuteurs et interlocutrices interrogées dans le cadre de la réalisation de cette étude : Victor Antonio et Claire Chusseau-Hamard (Brest métropole), Barbara Christian (Cerema), Anselme Le Brozec (RD Brest), Boris Mericskay (Université Rennes 2), Vincent Meyer (Métropole Aix-Marseille-Provence).

Pour aller plus loin

Avec l'Adeupa



[Impact économique et environnemental du Tour de France 2021](#)
Rapport d'étude - Mai 2022



[Les résultats principaux de l'EMD](#)
Observatoire de la mobilité n°4 - Mai 2019



[Décarbonation des mobilités](#)
Observatoire de la mobilité n°10 - Septembre 2023



[Les résultats de l'EMD : les déplacements du Samedi](#)
Observatoire de la mobilité n°5 - Décembre 2019



[Le cadre de vi\(II\)e en 2023](#)
Observatoire socio-économie de Mon réseau grandit - Décembre 2023

Et ailleurs

- [Circulation routière et congestion - Premiers résultats de l'exploitation des data Waze](#)
Agorah - Juillet 2023

LES OBSERVATOIRES | MOBILITÉ

Direction de la publication

Yves Cléach

Réalisation

Julien Gingembre

François Rivoal

Maquette et mise en page

Timothée Douy

Relecture

Magali Can

Tirage

150 exemplaires

Contact

contact@adeupa-brest.fr

Dépôt légal

4^e trimestre 2024

ISSN

2267-4411

Référence

24-159



AGENCE D'URBANISME DE BREST • BRETAGNE
18 rue Jean Jaurès - 29200 BREST
Tél. 02 98 33 51 71

www.adeupa-brest.fr



LICENCE OUVERTE
OPEN LICENCE