

Les données de téléphonie mobile

Une source de connaissance sur la population et ses déplacements



Marie-Pierre Joubert*

Les données de téléphonie mobile offrent des perspectives uniques pour la statistique publique, notamment pour étudier la présence de la population sur le territoire. Ainsi, grâce à des partenariats avec des opérateurs (Orange, Bouygues, SFR), l'Insee a pu analyser très rapidement les déplacements de population pendant la crise de la Covid-19, fournissant des informations cruciales pour calibrer les services publics. Ces données, collectées via la signalisation du téléphone sur le réseau, permettent de suivre les mobilités avec une précision temporelle fine. Cependant, elles posent des défis méthodologiques, au sujet en particulier de la couverture géographique et de la représentativité. Les collaborations avec les opérateurs, sous des cadres législatifs stricts aux niveaux national et européen, permettent d'accéder à des informations anonymisées et agrégées. Ces dernières complètent les sources traditionnelles, offrant une vision plus fine des dynamiques de population, importantes pour certaines politiques publiques.

 Mobile network data provides exceptional opportunities for official statistics, especially for studying the population's presence in a given area. Thanks to partnerships with operators (Orange, Bouygues, SFR), INSEE was able to quickly analyse population mobility during the Covid-19 crisis, providing essential information for adjusting public services. These data, collected via network signaling, enable mobility to be tracked with a high degree of temporal precision. However, they raise methodological challenges, particularly in terms of geographical scope and representativeness. Collaboration with operators, under strict national and European legislative frameworks, provides access to anonymised and aggregated information. This complements traditional sources, providing a more detailed view of population dynamics, which are important for various public policies.

* À la date de rédaction de l'article, responsable adjointe du SSP Lab, Insee.
marie-pierre.joubert@finances.gouv.fr

Dans quelle mesure la population s'est-elle confinée ailleurs que dans sa résidence principale pendant la crise de la Covid-19, par exemple chez de la famille, des amis ou dans une résidence secondaire ? La réponse à cette question était cruciale en cette période pour calibrer les services publics, notamment sanitaires. Or, les données habituellement utilisées par la statistique publique pour localiser la population (recensement de la population, sources fiscales) renseignent sur l'adresse des résidences principales et secondaires, mais pas sur leur occupation en temps réel. Grâce à un partenariat avec trois opérateurs de téléphonie mobile (Orange, Bouygues, SFR), l'Insee a pu mobiliser les données collectées par ces derniers grâce aux signaux envoyés par les téléphones sur leurs réseaux, pour analyser les déplacements de population pendant la crise sanitaire (Sémécurbe et al., 2020). Il a pu ainsi fournir aux préfetures un précieux éclairage sur le nombre de personnes présentes sur leur territoire.

Cet exemple marquant d'usage des données de téléphonie mobile s'inscrit dans le contexte de multiples collaborations entre l'Insee et les opérateurs : partenariats de recherche, projets européens grâce auxquels les instituts statistiques de plusieurs pays collaborent pour élaborer des méthodologies de combinaison des données, projets en lien avec le milieu académique financés par l'Agence nationale de la recherche, etc. Le présent article décrit les enjeux méthodologiques liés à l'utilisation de ces données, ainsi que les diverses modalités permettant au statisticien public d'y accéder, avec leurs avantages et inconvénients respectifs. Il conclut sur les perspectives offertes, notamment au niveau européen.

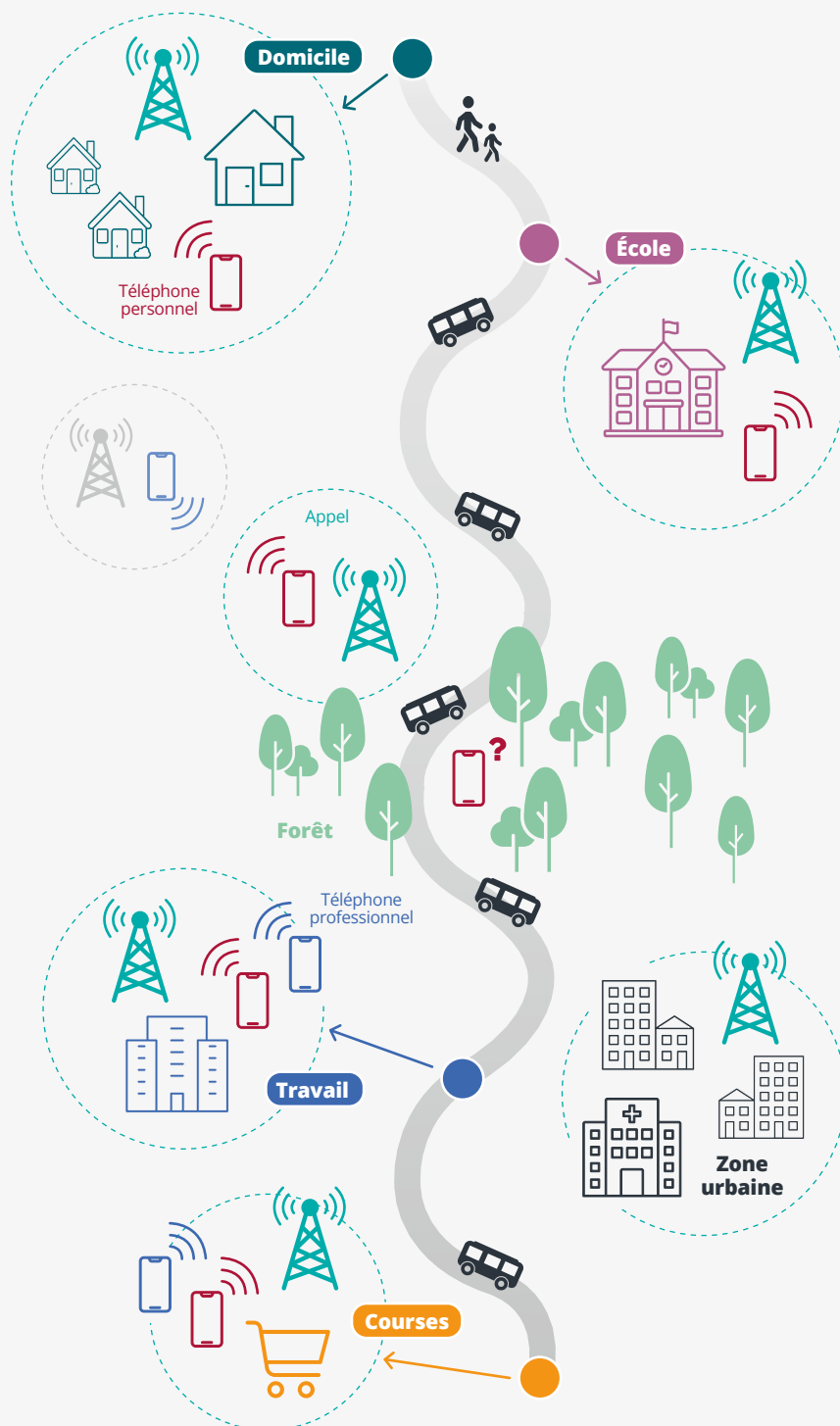
► Du téléphone mobile à la base de données : que collectent les opérateurs ?

Ce matin à 7h, madame Zaoui allume son téléphone en se réveillant. À 8h, elle dépose ses enfants à l'école, puis prend le bus vers son travail. En chemin, elle appelle son frère. La communication s'interrompt au moment où elle traverse la forêt, mais reprend dès qu'elle rejoint la zone urbaine. Elle passe ensuite la journée au bureau. Elle profite de la pause méridienne pour aller faire quelques courses au centre commercial le plus proche ; elle emporte alors avec elle son téléphone professionnel, qu'elle a allumé en arrivant au bureau (*figure 1*). Quelles informations sont collectées par l'opérateur téléphonique et consignées dans ses bases de données ?

Les comptes rendus d'appels

Les comptes rendus d'appel de facturation ou *Billing Call Detail Records* (CDR) sont enregistrés à chaque fois qu'un usager utilise le réseau mobile pour envoyer ou recevoir un appel ou un SMS, ou que les applications embarquées dans le mobile transfèrent des données. Autrement dit, il s'agit de tout évènement actif (à la « demande » du mobile) qui laisse une trace de facturation. Ici, un CDR est donc enregistré quand madame Zaoui appelle son frère. Il contient l'heure de l'appel, sa durée et la localisation des antennes les plus proches des téléphones de madame Zaoui et de son frère durant leur communication. Les données sont anonymisées, en conformité avec le règlement général sur la protection des données (RGPD). Par ailleurs, **le contenu de la**

► **Figure 1 - Suivi des connexions téléphoniques de madame Zaoui**



conversation ne fait pas partie des données enregistrées : seules les métadonnées¹ de l'appel sont conservées.

Dans le cadre des CDR, aucune information n'est enregistrée lorsque l'utilisateur se déplace en portant simplement son téléphone sans interagir avec le réseau. L'information est donc parcellaire d'un point de vue temporel. Sakarovitch et al. (2019) soulignent l'impact de cette irrégularité temporelle sur la production de statistiques sur la population.

Les données de signal

Les données de signalisation complètes (données de signal ou *signaling data*) comprennent également des informations liées à l'itinérance du téléphone mobile, qui visent à garantir la qualité de connexion au réseau. En effet, en dehors de tout évènement actif, si le téléphone se déplace significativement, il se signale automatiquement afin d'être joignable sur sa nouvelle position.

Les données de signal contiennent l'identifiant de l'antenne la plus proche du téléphone. Leur grand avantage par rapport aux CDR est leur fréquence temporelle plus élevée. Ainsi, selon Bonnetain et al. (2021), le temps médian entre deux évènements de signalisation est inférieur à 1 minute pour 90 % des utilisateurs, tandis qu'il est de plus de 30 minutes dans les CDR².

Contrairement aux CDR, cependant, les données de signal ne contiennent pas d'information sur les contacts entre usagers. Par ailleurs, les personnes comme madame Zaoui qui possèdent deux téléphones, un personnel et un professionnel, risquent d'être comptées deux fois. Ce phénomène est difficile à prendre en compte et fait partie des nombreuses difficultés méthodologiques à résoudre pour produire des statistiques publiques à partir de ces données.

► Quelle que soit la maille d'agrégation des données, de nombreuses difficultés méthodologiques à résoudre —

Les principales incertitudes inhérentes aux données de téléphonie mobile sont l'incertitude temporelle, l'incertitude sur la couverture de la population et l'incertitude spatiale (Ricciato et al., 2020). L'incertitude temporelle est surtout problématique lors du traitement des CDR. En effet, la fréquence temporelle à laquelle ces dernières fournissent une information dépend de l'usage du réseau téléphonique par l'utilisateur. Dans le cas des données de signal, il est plus facile de compléter les informations manquantes entre deux remontées de données puisque celles-ci sont plus nombreuses. Les deux autres types d'incertitudes concernent à la fois les CDR et les données de signal. D'autres questions méthodologiques s'ajoutent par ailleurs, que doit résoudre le statisticien public pour envisager d'utiliser ces données.

¹ Une métadonnée est une donnée qui fournit de l'information sur une autre donnée.

² Avec le développement des applications mobiles qui échangent des données, la différence entre CDR et données de signal a tendance à s'atténuer. En effet, les CDR intègrent aussi ces échanges de données, dont la fréquence est souvent élevée.

Gérer l'incertitude sur les caractéristiques de la population des abonnés de l'opérateur...

Travailler avec un unique opérateur peut conduire à des imprécisions ou biais dans les décomptes de population.

La couverture du territoire n'est pas identique suivant les opérateurs (ARCEP, 2024). Par ailleurs, les caractéristiques sociodémographiques de leurs abonnés sont en général différentes. Travailler avec un unique opérateur peut ainsi conduire à des imprécisions ou biais dans les décomptes de population pour certaines zones géographiques où la couverture réseau serait plus faible ou la population moins abonnée à cet opérateur spécifique.

Lorsque le redressement des données n'est pas effectué par l'opérateur, le statisticien public a besoin d'obtenir des informations sociodémographiques sur les clients de façon à pouvoir redresser les données. Or, pour ne pas divulguer le secret des affaires, l'opérateur ne diffuse pas toujours cette information à une échelle géographique fine, ce qui diminue la qualité des redressements effectués.

... et identifier des zones de domicile

Un élément-clé pour le statisticien travaillant sur les données de téléphonie mobile est l'identification de la zone de domicile la plus probable.

Un élément-clé pour le statisticien travaillant sur les données de téléphonie mobile est l'identification de la zone de domicile la plus probable. Cette information est cruciale pour faire le lien avec les sources de la statistique publique, comme le recensement de la population, et estimer ainsi la représentativité des données, voire recalculer ces dernières sur les chiffres du recensement. Le domicile est également un prérequis à de nombreuses analyses, par exemple celle des déplacements domicile-travail. Diverses méthodes

d'attribution de domicile existent et les résultats peuvent différer selon les méthodes, ce qui introduit là encore de l'incertitude.

Pour les données issues des CDR, Vanhoof et al. (2018) comparent cinq algorithmes classiques qui identifient respectivement le domicile comme le lieu :

- où la majorité des activités téléphoniques (appels ou SMS, émis ou reçus) ont été effectués ;
- où le nombre maximum de jours distincts avec des activités téléphoniques a été observé ;
- où la plupart des activités téléphoniques ont été enregistrées entre 19h00 et 9h00 ;
- où la plupart des activités téléphoniques ont été enregistrées en considérant un périmètre de 1 000 mètres autour d'une antenne-relais ;
- où la plupart des activités téléphoniques ont été enregistrées entre 19h00 et 9h00 en considérant un périmètre de 1 000 mètres autour d'une antenne-relais.

À un niveau agrégé, quel que soit l'algorithme, la différence de répartition de la population avec celle obtenue à partir du recensement n'est pas négligeable, ce qui illustre les difficultés méthodologiques posées par ces données. Vanhoof et al. (2018) insistent sur l'importance de créer un jeu de données ad hoc permettant de valider les résultats, par exemple en disposant, pour un échantillon d'utilisateurs, à la fois de leur domicile réel et des traces laissées par leurs communications téléphoniques. Connaître la répartition des parts de marché de l'opérateur est également une information très importante pour la qualité de l'attribution du domicile, de même qu'en savoir plus sur les habitudes d'usage du téléphone.

Pour les données de signal, Suarez Castillo et al. (2023) identifient la zone de domicile grâce à un algorithme qui s'appuie sur la fréquence et la localisation des signaux enregistrés. Ils agrègent ensuite ces données à un niveau spatial fin. Ils les comparent alors à celles de la population résidente estimée à ce même niveau à partir du dispositif **Filosofi**³, issu du rapprochement des données fiscales et des données sur les prestations sociales. Cette comparaison leur permet de déduire des poids pour redresser les données de téléphonie mobile. De premiers résultats prometteurs laissent augurer de futures avancées méthodologiques.

Gérer l'incertitude spatiale en estimant la couverture de l'antenne

Du point de vue spatial, disposer uniquement de l'information sur la localisation de l'antenne la plus proche apporte beaucoup d'incertitude, surtout dans les zones rurales peu couvertes en antennes (Sakarovitch et al., 2019). En effet, dans cette situation, on approxime la couverture géographique de l'antenne par un polygone selon la méthode inventée par le mathématicien russe Gueorgui Voronoï⁴.



Les opérateurs disposent des connaissances de radio-ingénierie leur permettant d'estimer de façon fiable la couverture de l'antenne. Suivant les modalités de collaboration avec eux, ils peuvent mettre à disposition ces informations.



Or la couverture réelle est souvent éloignée de cette représentation : elle dépend de la nature de l'antenne, des bâtiments alentours et même de la météo.

Des méthodes probabilistes permettent d'améliorer un peu la précision de l'estimation (Salgado et al, 2021 ; Ricciato et Coluccia, 2021 ; Gootzen et Tennekkes, 2022). Bonnetain et al. (2021) ont par ailleurs développé un algorithme permettant d'augmenter significativement la précision spatiale des déplacements reconstruits par la téléphonie mobile dans un environnement urbain.

³ <https://www.insee.fr/fr/metadonnees/source/serie/s1172>.

⁴ Voronoï a élaboré un algorithme mathématique qui permet, partant d'un ensemble discret de points, de partitionner l'espace en polygones autour de ces points avec la propriété suivante : pour un point p de l'ensemble discret de points P, tous les points contenus dans le polygone associé à p sont plus proches de p que d'aucun autre point de l'ensemble P.

Toutefois, les opérateurs disposent des connaissances de radio-ingénierie leur permettant d'estimer de façon fiable la couverture de l'antenne. Suivant les modalités de collaboration avec eux, ils peuvent mettre à disposition ces informations. Une incertitude demeure dans certains cas où la couverture d'une antenne intersecte celle de ses voisines, mais cela reste plus précis que l'estimation à partir des polygones de Voronoï.

Les questions méthodologiques selon le niveau d'agrégation des données

Sous réserve du respect de conditions strictes garantissant la non-divulgence de la vie personnelle des individus, ainsi que la sécurité des données, **certains partenariats permettent aux instituts de statistique publique d'accéder à des données individuelles pseudonymisées**. Ces données contiennent une ligne par identifiant de téléphone et minute de connexion au réseau, avec l'information sur la localisation de l'antenne la plus proche du téléphone. S'il s'agit de CDRs, les informations sur l'identifiant du téléphone contacté et la localisation de l'antenne la plus proche de ce téléphone sont également disponibles. Ces données sont de **taille massive**. Les traiter nécessite donc une infrastructure de stockage spécifique et des méthodes de science des données (ou *data science*) adaptées. La plupart du temps, les statisticiens se déplacent dans les locaux de l'opérateur pour travailler directement sur les serveurs sécurisés.

Les données individuelles ont l'avantage d'être exhaustives et de permettre à l'analyste de maîtriser de bout en bout les traitements effectués. La méthodologie est ainsi intégralement connue par l'analyste et des statistiques descriptives à façon peuvent être calculées. Ceci ne lève pas pour autant toutes les difficultés méthodologiques précitées, comme celles liées à l'évaluation des parts de marché de l'opérateur ou à la couverture spatiale de l'antenne.

Dans d'autres types de partenariats, l'opérateur fournit directement des données agrégées. L'opérateur a ainsi procédé lui-même au redressement des données pour les rendre représentatives de l'ensemble de la population résidant en France, ainsi qu'à leur projection géographique sur des zonages administratifs (commune, *Iris*⁵). Ces données sont en général d'une taille raisonnable, les traitements en sont donc grandement facilités. Toutefois, le plus souvent, les opérateurs ne divulguent pas l'intégralité des traitements méthodologiques effectués et tronquent les données, par exemple en ne diffusant pas d'information sur les déplacements de commune à commune représentant moins de 20 individus. Ceci restreint les usages possibles et rend la statistique publique dépendante d'un travail en amont qu'elle ne maîtrise pas.

La plupart du temps, les informations mises à disposition consistent en des décomptes de populations ayant été présentes dans une zone géographique donnée durant une période de temps donnée, par exemple une demi-heure. Il s'agit également de matrices origine-destination indiquant le nombre de personnes s'étant rendues d'une commune A à une commune B pendant un intervalle de temps donné. Ces informations permettent d'étudier la population présente et les déplacements de population. Elles nécessitent toutefois un cadrage méthodologique supplémentaire. Par exemple, faut-il définir une durée d'arrêt

⁵ L'Iris constitue la brique de base en matière de diffusion de données infracommunales.
<https://www.insee.fr/fr/metadonnees/definition/c1523>.



La plupart du temps, les informations mises à disposition consistent en des décomptes de populations ayant été présentes dans une zone géographique donnée durant une période de temps donnée, par exemple une demi-heure.



minimal dans une zone pour considérer qu'une personne y a été présente ? Ou encore, une personne doit-elle être comptée deux fois dans une zone si elle l'a traversée plusieurs fois ?

Les données sont souvent segmentées suivant la « zone de nuitée » qui peut être assimilée au domicile de l'individu et est directement calculée par l'opérateur. Aude et al. (2024) utilisent ces données agrégées en complément des données traditionnelles de l'Insee pour décrire le fonctionnement des territoires. Les quartiers de Lyon y sont d'abord caractérisés par une typologie basée sur leur population et leur parc de logements,

obtenues grâce aux sources de données traditionnelles de la statistique publique. Puis leur fréquentation en journée est analysée grâce aux données de téléphonie mobile.

Que ces données soient individuelles ou agrégées, y accéder nécessite d'établir un partenariat entre un ou plusieurs opérateur(s) de téléphonie mobile, un ou plusieurs institut(s) de statistique publique et parfois d'autres acteurs, par exemple issus du milieu académique.

► Des modalités d'accès qui concilient respect de la vie privée, qualité des informations et enjeux commerciaux des opérateurs

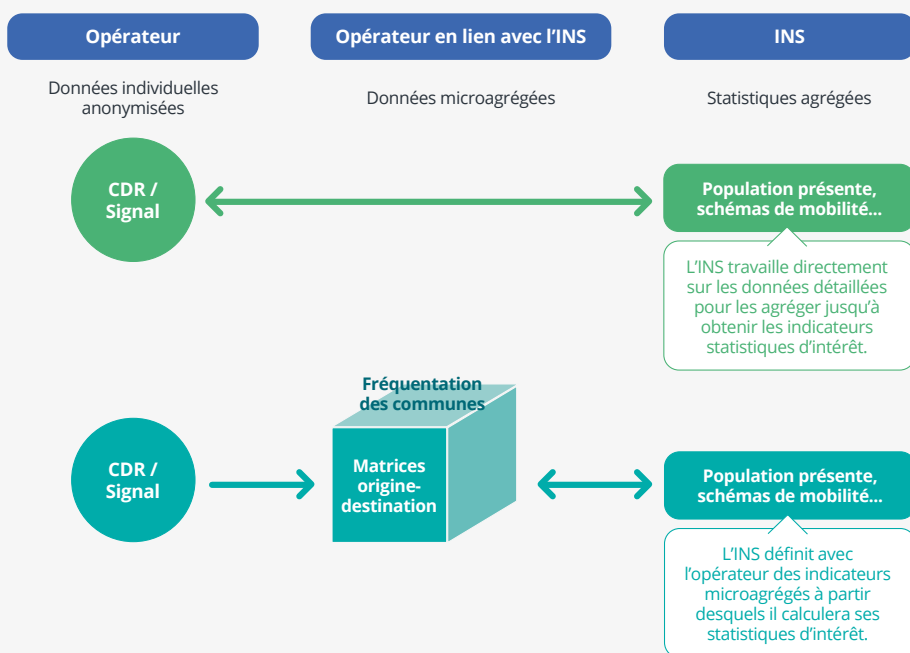
L'Insee a expérimenté plusieurs types de partenariats...

À ce jour, **deux principaux types de partenariat ont été expérimentés par l'Insee** : l'un permet d'accéder aux données détaillées, l'autre aux données agrégées (*figure 2*).

L'accès aux données détaillées a été obtenu pour la période 2016-2022 dans le cadre d'une convention tripartite entre Orange-Labs (le laboratoire de recherche en sociologie d'Orange), l'Insee et Eurostat⁶. Chaque partie permettait aux autres d'accéder à ses bases détaillées dans ses propres locaux. L'opérateur Orange apportait son expertise méthodologique et analytique de l'utilisation de la téléphonie mobile pour l'analyse sociale et économique et fournissait toutes les informations nécessaires pour utiliser les données. L'Insee assurait la disponibilité de sources susceptibles d'être utiles pour améliorer et évaluer la qualité des statistiques dérivées de la téléphonie mobile. Des accords de confidentialité s'assuraient de la non-diffusion d'informations sensibles issues des bases de l'un ou l'autre des partenaires. Plusieurs publications de recherche ont résulté de ce partenariat ; par exemple, l'une d'entre elles confronte le zonage en aires urbaines aux données de téléphonie mobile (Combes et al., 2017).

⁶ Eurostat est l'Office statistique de l'Union européenne.

► **Figure 2 - Modalités d'accès des instituts nationaux de statistique aux données de téléphonie mobile**



CDR : Comptes rendus d'appel ou *Call Detail Records*.
INS : Institut national de statistique.

Dans le contexte exceptionnel de la crise de la Covid-19, trois des quatre opérateurs majeurs ont répondu favorablement aux sollicitations de l'Insee et engagé des collaborations philanthropiques et limitées dans le temps.

Dans le contexte exceptionnel de la crise de la Covid-19, trois des quatre opérateurs majeurs ont répondu favorablement aux sollicitations de l'Insee et engagé des collaborations philanthropiques et limitées dans le temps. Des accords de confidentialité ou bons de commande ont été établis pour encadrer la livraison et l'utilisation des données agrégées. Disposer des données de trois opérateurs a permis d'améliorer la qualité des statistiques produites, qui ont été diffusées dans deux communiqués de presse et une publication *Insee Analyses* (Galiana et al., 2020). Toutefois,

de façon à ne pas révéler leurs parts de marché respectives à une échelle géographique fine, les opérateurs ont fourni des données déjà ajustées pour les rendre représentatives de la population. Cela a rendu difficile le retraitement par l'Insee, de même que la documentation des limites méthodologiques.

Un troisième type de partenariat est actuellement en cours. Il s'agit d'un partenariat de recherche, financé par l'Agence nationale de la recherche, entre l'université Gustave Eiffel, Orange et l'Insee. Le projet a pour objectif d'utiliser des données de téléphonie mobile,

combinées à d'autres sources de données, pour estimer en continu des indicateurs de présence et de mobilité des personnes, sur la zone du Grand Lyon. Le service commercial d'Orange met à disposition des partenaires des données agrégées qui correspondent au besoin du projet (présence et matrices origine-destination). Les équipes de recherche et développement d'Orange participent aux réunions et apportent l'ensemble des éléments méthodologiques nécessaires à la bonne compréhension des données. De plus, un post-doctorant travaille sur les bases détaillées dans les locaux d'Orange pour concevoir de nouveaux indicateurs expérimentaux. Enfin, une enquête ad hoc est menée sur un échantillon de volontaires, pour confronter leurs trajectoires réelles et les traces repérées par le réseau, ce qui est précieux pour mieux comprendre les données. Ce partenariat, bien que donnant principalement accès à des données agrégées, permet donc de continuer à améliorer la compréhension des données et à illustrer leur intérêt pour la statistique publique. Les récentes avancées législatives au niveau européen ouvrent toutefois la voie à un accès plus généralisé aux données détaillées.

... qui s'inscrivent dans le contexte juridique français et européen...

De manière générale, on entend par **base de données privée** toute base de données collectée ou produite par des organismes de droit privé dans le cadre de leurs activités. Ces données peuvent concerner des tiers, notamment des personnes physiques, et donc être à ce titre protégées par le RGPD, ce qui est le cas pour les données de téléphonie mobile. Elles peuvent être l'objet, de la part de leur détenteur, d'une valorisation à des fins commerciales ou d'une publication. Elles peuvent contenir des informations dont le secret est protégé par la loi et relever pour ce motif des obligations dues au secret professionnel. La circulation de ces données est régie par plusieurs lois imbriquées aux niveaux français et européen, qui protègent donc d'une part la propriété de la base de données (droit d'auteur, de propriété), d'autre part la diffusion du contenu de ces bases, notamment ce qui est lié au caractère de leur contenu (comme les données à caractère personnel).

En France, **la loi du 7 octobre 2016 pour une République numérique⁷** favorise l'ouverture de l'accès aux données publiques (open data) et encourage la circulation des données, la protection des individus dans la société du numérique et l'accès au numérique pour tous. Cette loi introduit un article 3 bis dans la loi du 7 juin 1951⁸, qui oblige les personnes morales de droit privé à « transmettre par voie électronique sécurisée au Service statistique public, à des fins exclusives d'établissement de statistiques, les informations présentes dans les bases de données qu'elles détiennent, lorsque ces informations sont recherchées pour les besoins d'enquêtes statistiques obligatoires. Ces données ne peuvent pas ensuite être transmises à un tiers. » Pour accéder aux données privées suivant ce cadre, il est nécessaire de passer par plusieurs étapes définies dans le décret d'application : concertation avec les détenteurs de données, étude de faisabilité et d'opportunité, avis du Conseil national de l'information statistique (Cnis), parution d'un arrêté et surtout substitution d'une enquête statistique. **Les données de téléphonie mobile n'ont pour le moment pas été mobilisées pour obtenir des informations qui, sinon, auraient été collectées par une enquête. Elles ne sont donc a priori pas concernées par ce cadre légal.**

⁷ Voir les références juridiques en fin d'article.

⁸ Voir les références juridiques en fin d'article.



Dans sa version révisée adoptée par le Parlement et le Conseil en 2024, le règlement européen n° 223/2009 introduit la possibilité légale d'utiliser les données détenues par le secteur privé pour le développement et la production de statistiques officielles européennes.



L'utilisation des données de téléphonie mobile est encadrée, en plus du RGPD, par **le règlement européen n° 223/2009⁹**, qui établit le cadre légal pour le développement, la production et la diffusion des statistiques européennes. Dans sa version révisée adoptée par le Parlement et le Conseil en 2024, il introduit la possibilité légale d'utiliser les données détenues par le secteur privé « pour le développement et la **production de statistiques officielles européennes**, sur une base durable et selon des règles

équitable, claires, prévisibles et proportionnées, conformément au cadre des droits fondamentaux de l'Union. L'accès aux données détenues par le secteur privé devrait être garanti conformément au principe de rentabilité et ne doit pas entraîner de charge excessive pour les opérateurs économiques. » La question de la maille d'agrégation des données et du montant de l'éventuelle compensation financière des opérateurs reste ouverte, mais il s'agit d'une perspective prometteuse pour l'utilisation de ces données, sous réserve de la production de statistiques européennes rentrant dans le cadre du programme annuel de travail du système statistique européen. Pour mettre en place des collaborations constructives, le champ exact des données fournies aux instituts nationaux de statistique devra prendre en compte également les enjeux propres aux opérateurs.

... et prennent en compte les enjeux propres aux opérateurs pour établir des partenariats bénéfiques pour tous

Les opérateurs de téléphonie mobile sont intéressés par l'expertise technique des chercheurs et des statisticiens publics. En travaillant sur les données de téléphonie, qu'elles soient détaillées ou agrégées, ceux-ci soulèvent certains problèmes non encore résolus et mettent à l'épreuve les méthodes d'agrégation utilisées par l'opérateur, ce qui conduit à une amélioration générale de la qualité. De plus, les statisticiens publics ont accès à des données de référence qui peuvent aider à mieux calibrer les données des opérateurs privés, ou du moins à mettre en avant les limites des méthodes de calibrage actuelles.

Sur le plan de l'image publique, le point de vue des opérateurs est nuancé. D'un côté, ces derniers souhaitent maîtriser l'utilisation faite de données sensibles pour leurs utilisateurs, alors que les enjeux d'acceptabilité sociale des travaux menés sur les données sont cruciaux. D'un autre côté, travailler avec le milieu académique et la statistique publique illustre la contribution de l'opérateur aux sujets d'intérêt général, ce qui a un rôle positif sur son image (Sémécurbe et al., 2020 ; Coudin et al., 2021).

Les opérateurs craignent aussi la divulgation à leurs concurrents d'informations sensibles sur leurs parts de marché. Travailler avec les données détaillées anonymisées de plusieurs opérateurs nécessite donc de se positionner comme un tiers de confiance.

⁹ Voir les références juridiques en fin d'article.



Une bonne coopération avec les fournisseurs de données est un prérequis pour bien comprendre leur méthodologie de construction et produire des statistiques les plus robustes et documentées possible.



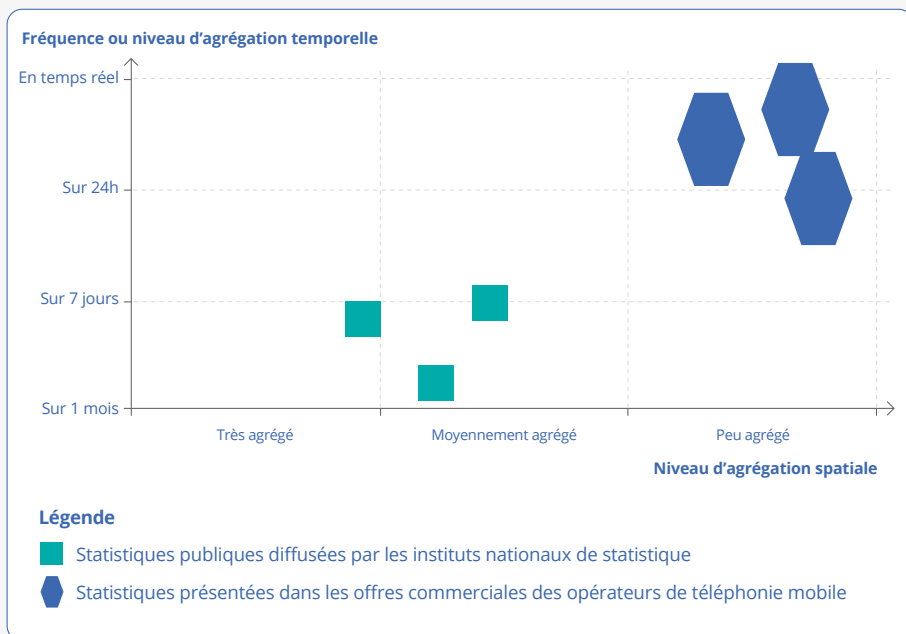
Ceci demande un investissement non négligeable dans des infrastructures techniques permettant de garantir le niveau de confidentialité requis.

Enfin, certains opérateurs commercialisent eux-mêmes des statistiques agrégées, notamment pour étudier la population présente et les mobilités. Quel que soit le cadre législatif, il est nécessaire d'échanger avec eux pour établir un partenariat qui soit bénéfique pour toutes les parties, d'autant que ces données ne sont pas directement exploitables par les statisticiens et doivent être retraitées par les opérateurs. Une bonne coopération avec les fournisseurs de

données est dans tous les cas un prérequis pour bien comprendre leur méthodologie de construction et produire des statistiques les plus robustes et documentées possible. En général, les indicateurs publiés par les instituts de statistique se situent à une maille géographique et temporelle beaucoup plus large que celle des indicateurs vendus par les opérateurs (*figure 3*).

Quels sont plus précisément les indicateurs pour lesquels l'utilisation des données de téléphonie mobile présente un intérêt pour les statistiques publiques ?

► **Figure 3 - Instituts nationaux de statistique et opérateurs de téléphonie : des niveaux d'analyse complémentaires***



* D'après une figure réalisée par Fabio Ricciato (Eurostat).

► Une source d'information riche et utile pour le système statistique public

Grâce à leur volume et à la richesse des informations qu'elles contiennent, les données de téléphonie mobile peuvent compléter celles du système statistique public pour étoffer les statistiques habituellement produites ou pour éclairer un nouvel angle d'analyse.

Étudier la population présente

Le taux d'incidence de la Covid-19 a-t-il augmenté en Île-de-France durant les Jeux Olympiques de 2024 ? La réponse varie suivant le dénominateur : population résidente au sens du recensement, ou population réellement présente à cette période estimée avec la téléphonie mobile. En revenant sur la situation de 2021, Tarantola et Hamidouche (2025), chercheurs à Santé Publique France, montrent qu'en prenant au dénominateur la population présente, le critère d'alerte maximal a été dépassé durant le mois d'août 2021, alors que ce n'est pas le cas si on met au dénominateur la population résidente (*figure 4*).

De nombreuses politiques publiques gagneraient en précision et en efficacité en complétant les données de population résidente par celles de la population présente.

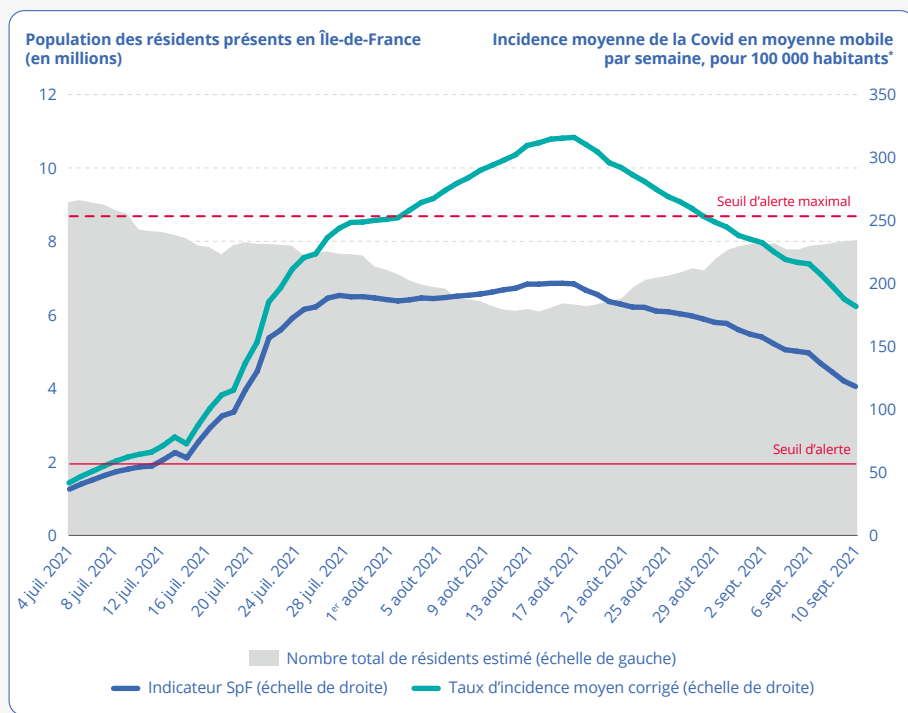
Au-delà des politiques sanitaires, de nombreuses politiques publiques gagneraient en précision et en efficacité en complétant les données de population résidente par celles de la population présente. Ce constat a été fait également par le Cnis. Ce dernier indique dans ses objectifs de moyen terme 2024-2028 que « la seule mesure de la population résidente ne suffit pas pour capter la dynamique et l'attractivité d'un territoire, la fréquentation de ses équipements et l'utilisation de ses ressources ». Il émet

le souhait que « l'ensemble des données publiques et privées soient mobilisées pour mesurer la population présente ». Insistons bien, malgré tout, sur le fait que la « population présente » est un autre concept que celui de « population résidente » et que seul le recensement de la population permet d'estimer cette dernière.

Pour estimer la population présente, l'étape de mise en cohérence (calage) des données de téléphonie mobile avec les données de référence est particulièrement importante. Pour ce faire, l'enjeu méthodologique décrit plus haut d'une bonne attribution du lieu de domicile est particulièrement important. Une fois les données calées, il est envisageable de les confronter à d'autres sources de données issues d'acteurs privés, de façon à aborder l'évolution de la population présente sous différents angles. Ainsi, la *figure 5* compare le nombre de personnes détectées avec la téléphonie mobile au nombre de personnes ayant réalisé une transaction par carte bancaire, demi-heure par demi-heure, sur la zone de Lyon en septembre 2022¹⁰. Les données de téléphonie mobile correspondent à la moyenne du nombre de personnes ayant borné dans un Iris (quartier) de la ville de Lyon. Ces données ont été mises à disposition par le service Flux Vision commercialisé par Orange Business, de façon agrégée. Il s'agit davantage d'un

10 Source : Gabrielle Gambuli et Chloé Breton (Insee, université Gustave Eiffel, Télécom Paris).

► **Figure 4 - Incidence de la Covid-19 pendant l'été 2021, par semaine glissante, calculée par Santé publique France (SpF) et corrigée**



* Au numérateur : nombre de tests positifs réalisés par les résidents d'Île-de-France en Île-de-France ; au dénominateur, nombre de résidents d'Île-de-France au sens de l'Insee pour l'indicateur SpF ou estimés avec les données de téléphonie pour l'indicateur corrigé.
Champ : Île-de-France.
Sources : Santé publique France ; Orange (données de signal) ; calculs Santé publique France.
D'après Tarantola et Hamidouche, à paraître.

indicateur de fréquentation que d'un indicateur de présence. En effet, une personne qui traverse plusieurs fois un Iris durant la demi-heure est comptabilisée plusieurs fois. Les données de transaction par carte bancaire correspondent au nombre de porteurs de cartes différents ayant effectué une transaction dans la zone durant la demi-heure. L'accès aux données individuelles accordé aux chercheurs dans le cadre de la chaire Finance digitale¹¹ a permis de construire un indicateur de fréquentation calculé avec une méthodologie identique à celle utilisée pour les données de téléphonie mobile.

En semaine, on observe avec les données de téléphonie mobile une augmentation de fréquentation de l'aire d'attraction de la ville de Lyon en journée, avec un premier pic lors des déplacements domicile-travail matinaux, un deuxième à la pause déjeuner et un troisième, plus important, au moment du retour à domicile. Les volumes de cartes bancaires ayant effectué une transaction, eux, augmentent nettement au moment de la pause déjeuner et en fin d'après-midi. Le samedi, les augmentations de fréquentation observées grâce

¹¹ Voir l'article de Boittelle et al. sur les données de transactions par carte bancaire CB dans ce même numéro.



Ces travaux exploratoires illustrent l'intérêt d'utiliser plusieurs sources de données complémentaires pour affiner l'analyse des comportements de la population.

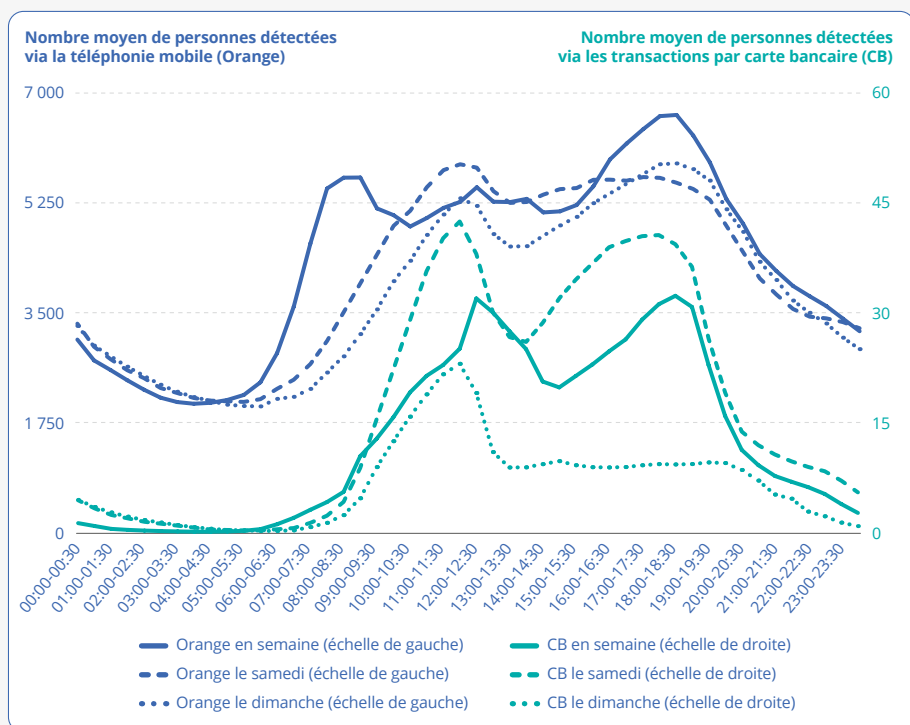


aux données de téléphonie mobile (en fin de matinée et dans l'après-midi) coïncident cette fois avec l'augmentation des transactions par carte bancaire. Enfin, le dimanche, les transactions par carte bancaire sont plus faibles, en particulier l'après-midi. Ces travaux exploratoires illustrent l'intérêt d'utiliser plusieurs sources de données complémentaires pour affiner l'analyse des comportements de la population. Combiner ces visions parcellaires de la situation réelle permet un vrai gain en qualité.

Aller au-delà des mobilités domicile-travail

L'Insee publie tous les dix ans différents zonages d'étude, destinés à mieux comprendre l'organisation du territoire et faciliter la production de statistiques territoriales. Ces zonages s'appuient souvent sur les déplacements domicile-travail mesurés par le

► **Figure 5 - Volume moyen de personnes détectées dans l'aire d'attraction de la ville de Lyon, avec la téléphonie mobile et les transactions par carte bancaire, en septembre 2022**



Sources : Flux Vision (données de signal) ; groupement des cartes bancaires CB ; calculs Insee.
D'après Gabrielle Gambuli et Chloé Breton (Insee, université Gustave Eiffel, Télécom Paris).

recensement. Cette source a l'avantage d'être exhaustive, fiable et bien documentée, mais les déplacements mesurés ne couvrent qu'une partie des mobilités : les déplacements des non-actifs (retraités, étudiants, etc.) façonnent aussi le territoire. Depuis la crise de la Covid-19, par ailleurs, le développement du télétravail a fait évoluer les mobilités quotidiennes des actifs. Par exemple, le zonage en **aires d'attraction des villes**¹² définit les pôles, densément peuplés et riches en emplois, puis leurs couronnes, communes dont 15 % des actifs occupés vont travailler dans le pôle. Les contours des aires d'attraction des villes évolueront sûrement si l'on considère l'ensemble des déplacements observés dans les données de téléphonie mobile et pas uniquement les déplacements domicile-travail.

Les mobilités sont en partie captées grâce à des enquêtes spécifiques (par exemple les enquêtes **EMC2**¹³ du Cerema¹⁴ ou l'enquête « **Mobilité des personnes** »¹⁵ du SDES¹⁶). Comme indiqué dans le **suivi d'avis du Cnis du 6 juin 2024**¹⁷, les données d'enquête sont riches d'informations concernant le profil des personnes présentes et leurs motifs de présence, mais elles ne permettent pas d'observer les variations hebdomadaires ou mensuelles et certaines ne couvrent que les espaces urbains. Ces enquêtes et les données de téléphonie mobile interviennent donc de façon complémentaire.

Analyser l'évolution de la ségrégation sociospatiale au fil de la journée

Les quartiers prioritaires de la politique de la ville concentrent de nombreuses difficultés sociales et économiques. Mieux comprendre dans quelle mesure ces quartiers sont isolés du reste des espaces urbains est un fort enjeu pour cibler les politiques publiques visant à diminuer cet isolement. Les indicateurs de mixité sociale pendant la nuit permettent d'identifier les zones où la ségrégation sociale sur le lieu de résidence est la plus forte. Toutefois, en fonction des mobilités en journée, cette ségrégation peut rester très marquée dans certains espaces et au contraire s'atténuer dans d'autres.

Étudier ce phénomène nécessite de combiner données de téléphonie mobile et données traditionnelles à l'échelle spatiale la plus fine possible. Bien sûr, l'appariement exact au niveau individuel n'est pas envisageable puisque toutes les données sont anonymisées. Il est toutefois possible d'attribuer à tous les porteurs de téléphone mobile résidant dans une zone géographique donnée (commune, **carreau**¹⁸, etc.) les caractéristiques sociodémographiques des habitants de cette zone. La mobilité moyenne observée en journée pour les habitants de cette zone permet ensuite d'étudier leur nouvelle répartition sur le territoire heure par heure et ainsi l'évolution de la ségrégation.

À 6 heures du matin, on considère que les personnes détectées dans les données de téléphonie mobile sont à leur lieu de domicile. À cette heure, en région parisienne, les zones où la part des personnes à bas revenus est supérieure à 20 % sont très concentrées dans le

¹² <https://www.insee.fr/fr/metadonnees/definition/c2173>.

¹³ <https://www.cerema.fr/fr/activites/mobilities/connaissance-modelisation-evaluation-mobilite/enquetes-mobilite-emc2>.

¹⁴ Le Cerema est le Centre d'études et d'expertise sur les risques, l'environnement, la mobilité et l'aménagement.

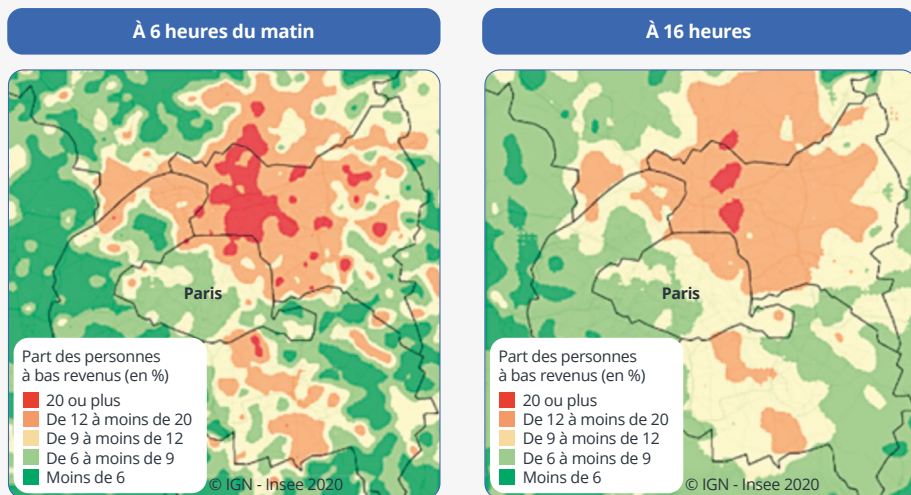
¹⁵ <https://www.statistiques.developpement-durable.gouv.fr/enquete-sur-la-mobilite-des-personnes-2018-2019?ist-enquete=true>.

¹⁶ Le service des données et études statistiques (SDES) assure les fonctions de service statistique des ministères chargés de l'environnement, de l'énergie, de la construction, du logement et des transports.

¹⁷ <https://www.cnis.fr/wp-content/uploads/2024/03/com-terr-2024-1-suivi-avis.pdf>.

¹⁸ L'Insee découpe le territoire en carreaux pour y diffuser de l'information statistique à un niveau faiblement agrégé. Selon les informations diffusées, il s'agit de carreaux de 1 km de côté ou, au plus fin, de 200 mètres de côté. Voir <https://www.insee.fr/fr/outil-interactif/7737357/documentation.html#carroyage>.

► Figure 6 - Évolution de la ségrégation sociospatiale en journée



Lecture : Les personnes à faibles revenus sont plus concentrées dans le nord-est durant la nuit, la ségrégation diminue au cours de la journée.

Champ : Unité urbaine de Paris.

Sources : Orange, CDR 2007 ; Insee, Filosofi 2014 ; calculs Insee. D'après les figures 1.b et 1.c publiées dans Galiana et al. (2020).

nord de Paris, notamment en Seine-Saint-Denis (*figure 6*). En revanche, à 16h, la répartition est beaucoup plus homogène. Les données de téléphonie mobile permettent ainsi d'éclairer l'évolution de la ségrégation sociospatiale en journée d'une façon que ne permettent pas du tout les sources traditionnelles des statistiques publiques (Galiana et al., 2020).

► Des perspectives européennes

L'enjeu de la prise en compte de l'itinérance

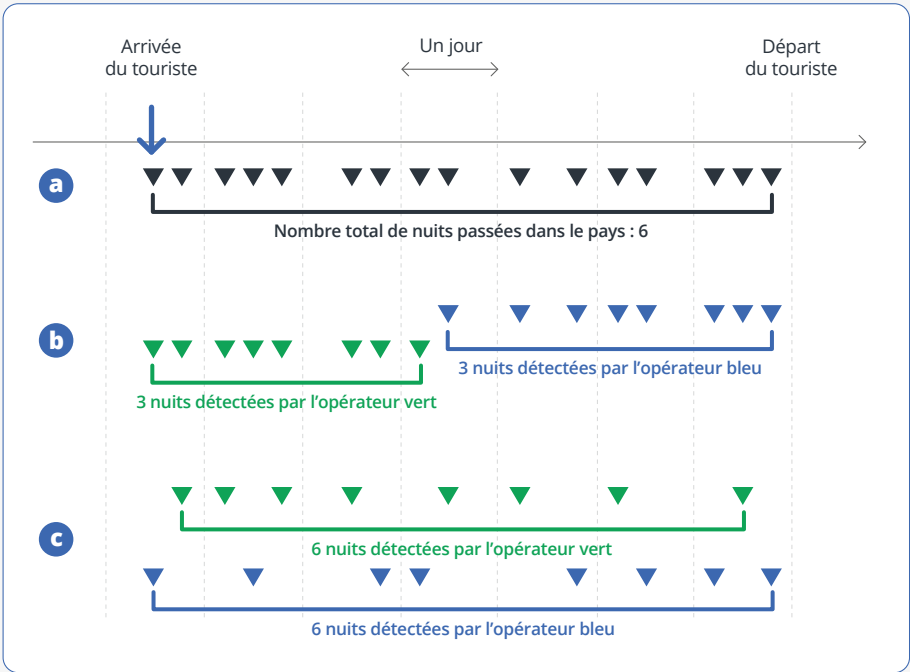
Au-delà de la France, l'usage des données de téléphonie mobile est un enjeu pour l'ensemble du système statistique public européen.

Au-delà de la France, l'usage des données de téléphonie mobile est un enjeu pour l'ensemble du système statistique public européen (*European Statistical System*, 2021). Certaines thématiques dépassent par définition les frontières nationales. Ainsi, étudier les mobilités touristiques nécessite de traiter au mieux le fait que l'abonné d'un opérateur peut en changer lorsqu'il visite un autre pays (phénomène d'itinérance ou *roaming*) et même plusieurs fois. Ces changements

peuvent advenir y compris lorsque l'abonné est dans son propre pays mais sur une zone frontalière. Le fait que les opérateurs ne collaborent pas au niveau des données individuelles peut mener à des doubles comptes. Ainsi, la *figure 7* présente le cas d'un

touriste passant six nuits en France. Dans le cas (b), l'opérateur vert détecte trois nuits, puis l'opérateur bleu détecte trois nuits. Il n'y a donc pas de doubles comptes, mais si l'institut de statistique a accès aux données d'un unique opérateur, le visiteur ne sera observé que partiellement. En revanche, dans le cas (c), les opérateurs vert et bleu détectent chacun six nuits. Si les deux opérateurs fournissent des décomptes agrégés, le visiteur sera compté en double. Il s'agit là des principaux cas, mais de nombreuses autres configurations peuvent exister. Une bonne collaboration entre opérateurs et entre instituts de statistique au-delà des frontières nationales est ainsi fondamentale pour estimer les flux avec une précision suffisante.

► **Figure 7 - Quelques conséquences possibles du phénomène d'itinérance**



Lecture : Un touriste passe 6 nuits en France. Dans le cas (b), chacun des deux opérateurs détecte trois nuits. Dans le cas (c), les opérateurs détectent chacun six nuits. Le triangle représente le signal émis par le téléphone.
Source : projet Multi-MNO : <https://cros.ec.europa.eu/landing-page/multi-mno-project>.

De nombreux projets en cours sous l'égide d'Eurostat

Eurostat finance depuis décembre 2023 et pour une durée de deux ans, le projet de recherche « *Trusted Smart Statistics: methodological developments based on new data sources* »¹⁹, appelé également projet MNO-MINDS²⁰. Ce projet regroupe dix pays européens et est coordonné par l'Italie. L'objectif est de développer un cadre méthodologique commun permettant de combiner données de téléphonie mobile et autres sources de données. Les données considérées appartiennent à deux grandes catégories : celles collectées par les instituts nationaux de statistique (INS) dans le



L'objectif est de se placer dans une perspective de production régulière de statistiques officielles et plus seulement de statistiques expérimentales.



but d'élaborer des statistiques (recensement, enquêtes) ; celles collectées en premier lieu pour des usages autres que statistiques et ensuite réutilisées pour les analyses socioéconomiques (données administratives, capteurs de trafic routier, données de billettique de transport en commun, etc.). Eurostat souhaite que ce panorama des sources prenne en compte l'arbitrage entre la qualité des données et leur coût (d'acquisition ou de traitement suivant la source). L'objectif est de se placer dans une perspective de production régulière de statistiques officielles et plus

seulement de statistiques expérimentales. De plus, il s'agit de considérer la disponibilité potentielle dans tous les pays européens. Des méthodes permettant de combiner les différentes sources de données seront également développées et diffusées et une enquête ad hoc servira à mieux comprendre les usages des téléphones : distinction entre téléphones personnel et professionnel, cas où un abonné principal gère les téléphones de toute sa famille, etc.

Par ailleurs, Eurostat finance un deuxième projet (Multi-MNO²¹), débuté en décembre 2023 pour une durée de deux ans, piloté par un cabinet de conseil (GOPA) et réunissant des INS de quatre pays (CBS pour les Pays-Bas, ISTAT pour l'Italie, GUS pour la Pologne et SURS pour la Slovénie), des entreprises spécialisées dans le traitement de données mobiles (Positium et Nommon) et cinq opérateurs de téléphonie mobile, issus de quatre pays différents (Orange Espagne, Vodafone Espagne, Vodafone Italie, A1 Slovénie et POST Luxembourg). L'objectif est de développer un traitement de données standardisé (ou *pipeline*) permettant d'agréger les données individuelles des opérateurs sur la base d'une méthode validée, de façon à produire des statistiques agrégées avec un bon niveau de qualité.

Le fait que deux opérateurs du même pays participent au projet permettra de démontrer, d'un point de vue technique et organisationnel, la possibilité de produire des statistiques à partir de données d'opérateurs concurrents au sein d'un même pays. Les opérateurs fourniront l'accès à leurs données pour tester, évaluer et améliorer le circuit de traitement développé durant le projet. Seules les données agrégées et anonymisées quitteront

¹⁹ Des statistiques de confiance à partir des objets connectés : développements méthodologiques basés sur les nouvelles sources de données. <https://www.insee.fr/fr/information/7681963>.

²⁰ <https://cros.ec.europa.eu/mno-minds>.

²¹ <https://cros.ec.europa.eu/landing-page/multi-mno-project>.



Une task force sur l'usage des données de téléphonie mobile pour la statistique publique a été lancée par Eurostat en 2021.



les serveurs des opérateurs de téléphonie mobile. L'objectif est que les traitements sur données individuelles développés dans le cadre du projet puissent être réalisés directement sur les serveurs des opérateurs. Certains éléments de la méthode d'agrégation resteront paramétrables, de façon à s'ajuster aux exigences nationales et notamment aux contraintes réglementaires, lesquelles peuvent différer suivant les pays. Une attention particulière sera portée au respect des informations

commerciales sensibles pour les opérateurs. Par exemple, les poids utilisés pour garantir la représentativité des statistiques ne seront pas publiés. Des experts du domaine juridique font partie du consortium.

Enfin, une task force sur l'usage des données de téléphonie mobile pour la statistique publique a été lancée par Eurostat en 2021. Son objectif est de coordonner et d'orienter les développements méthodologiques relatifs à l'utilisation de données de téléphonie mobile au sein du système statistique européen, de favoriser le partage des connaissances et les bonnes pratiques issues des expériences nationales, et de progresser vers la définition d'un cadre méthodologique commun pour l'ensemble du système statistique européen.

En septembre 2023, les 18 pays membres de la task force ont cosigné un article (*European Statistical System Task Force*, 2023) argumentant en faveur du développement d'une méthodologie commune à l'ensemble du système statistique européen pour traiter les données de téléphonie mobile. Une telle méthodologie doit être transparente et interprétable. Elle doit également permettre de comparer les statistiques obtenues dans les différents pays et de combiner différentes statistiques entre elles.

Remerciements

Pierre Bayard, Chloé Breton, Étienne Côme, Élise Coudin, Gabrielle Gambuli, Méлина Hillion, Sylvie Lagarde, Arnaud Legendre, Romain Lesur, Fanny Mikol, Latifa Oukhellou, Corinne Prost, Patrick Redor, Denis Renaud, Milena Suarez Castillo. Remerciement au projet MobiTIC (ANR-19-CE22-0010), financé par l'Agence nationale de la recherche en France.

► Fondements juridiques

- Règlement (CE) n° 223/2009 révisé du Parlement européen et du Conseil du 11 mars 2009 relatif aux statistiques européennes. In : *site de l'Union européenne*. [en ligne]. [Consulté le 27 décembre 2024]. Disponible à l'adresse : <https://eur-lex.europa.eu/legal-content/FR/TXT/?qid=1724682778358&uri=CELEX%3A32009R0223>.
- Règlement (UE) n° 2022/868 du Parlement européen et du Conseil du 30 mai 2022 portant sur la gouvernance européenne des données et modifiant le règlement (UE) n° 2018/1724. In : *site de l'Union européenne*. [en ligne]. [Consulté le 25 mars 2025]. Disponible à l'adresse : <https://eur-lex.europa.eu/eli/reg/2022/868/oj?locale=fr>.
- Loi n° 51-711 du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques. In : *site de Légifrance*. [en ligne]. Mise à jour le 25 mars 2019. [Consulté le 27 décembre 2024]. Disponible à l'adresse : <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000888573>.
- Loi n° 2016-1321 du 7 octobre 2016 pour une République numérique. In : *site de Légifrance*. [en ligne]. [Consulté le 25 mars 2025]. Disponible à l'adresse : <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000033202746>.

► Bibliographie

- ARCEP, 2024. *Mon réseau mobile*. [en ligne]. [Consulté le 27 décembre 2024]. Disponible à l'adresse : <https://monreseau-mobile.arcep.fr/>.
- AUDE, Johanne, DEBOUZY, Ivan, JOUBERT, Marie-Pierre, PRAMIL, Julien et GAMBULLI, Gabrielle, 2024. Cinq types de territoires diversement habités et inégalement fréquentés en journée – Aire d'attraction de la ville de Lyon. In : *Insee Analyses*. [en ligne]. 27 novembre 2024. Insee. N° 184. [Consulté le 27 décembre 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/8289238>.
- BONNETAIN, Loïc, FURNO, Angelo, EL FAOUZI, Nour-Eddin, FIORE, Marco, STANICA, Razvan, SMOREDA, Zbigniew et ZIEMLICKI, Cezary, 2021. TRANSIT: Fine-grained human mobility trajectory inference at scale with mobile network signaling data. In : *Transportation Research Part C: Emerging Technologies*. [en ligne]. Septembre 2021. Volume 130, 103257. [Consulté le 27 décembre 2024]. Disponible à l'adresse : <https://doi.org/10.1016/j.trc.2021.103257>.
- COMBES, Stéphanie, JOUBERT-DE BELLEFON, Marie-Pierre et VANHOOF, Maarten, 2017. Mining Mobile Phone Data to Detect Urban Areas. In : *Proceedings of the Conference of the Italian Statistical Society: SIS 2017, Statistics and Data Science: New challenges, new generations*. [en ligne]. Juin 2017. [Consulté le 27 décembre 2024]. Disponible à l'adresse : https://www.researchgate.net/publication/321278660_Mining_Mobile_Phone_Data_to_Detect_Urban_Areas.
- COUDIN, Élise, POULHES, Mathilde et SUAREZ CASTILLO, Milena, 2021. The French official statistics strategy: Combining signaling data from various mobile network operators for documenting COVID-19 crisis effects on population movements and economic outlook. In : *Cambridge University Press*. [en ligne]. 24 juin 2021. [Consulté le 27 décembre 2024]. Disponible à l'adresse : <https://doi.org/10.1017/dap.2021.1>.
- EUROPEAN STATISTICAL SYSTEM, 2021. European Statistical System (ESS) position paper on the future Data Act proposal – Access to privately held data is urgently needed for producing new, faster, more detailed official statistics. In : *site d'Eurostat*. [en ligne]. Juin 2021. [Consulté le 27 décembre 2024]. Disponible à l'adresse : <https://ec.europa.eu/eurostat/documents/13019146/13405116/main+ESS+position+paper+on+future+Data+Act+proposal.pdf/37f3b5c7-abfd-5a05-6be2-fdc4b87ee7d2?t=1631695372906>.
- EUROPEAN STATISTICAL SYSTEM TASK FORCE, 2023. Reusing mobile network operator data for official statistics: the case for a common methodological framework for the European Statistical System. In : *site de Eurostat*. [en ligne]. 12 septembre 2023. [Consulté le 27 décembre 2024]. Disponible à l'adresse : <https://ec.europa.eu/eurostat/web/products-statistical-reports/w/ks-ft-23-001>.
- GALIANA, Lino, SUAREZ CASTILLO, Milena, SÉMÉCURBE, François, COUDIN, Élise et JOUBERT-DE BELLEFON, Marie-Pierre, 2020. Retour partiel des mouvements de population avec le déconfinement. In : *Insee Analyses*. [en ligne]. 22 juillet 2020. Insee. N° 54. [Consulté le 27 décembre 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/4635407>.

- GALIANA, Lino, SAKAROVITCH, Benjamin, SÉMÉCURBE, François et SMOREDA, Zbigniew, 2020. La mixité sociale est plus forte en journée sur les lieux d'activité que pendant la nuit dans les quartiers de résidence. In : *Insee Analyses*. [en ligne]. 9 novembre 2020. Insee. N° 59. [Consulté le 27 décembre 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/4930403>.
- GOOTZEN, Yvonne et TENNEKES, Martijn, 2022. Bayesian location estimation of mobile devices using a signal strength model. In : *Journal of Spatial Information Science*. [en ligne]. Décembre 2022. [Consulté le 27 décembre 2024]. Disponible à l'adresse : <https://josis.org/index.php/josis/article/view/166>.
- RICCIATO, Fabio et COLUCCIA, Angelo, 2021. On the Estimation of Spatial Density from Mobile Network Operator Data. In : *IEEE Transactions on Mobile Computing*. [en ligne]. Décembre 2021. Pp. (99):1-1. [Consulté le 27 décembre 2024]. Disponible à l'adresse : https://www.researchgate.net/publication/357009792_On_the_estimation_of_spatial_density_from_mobile_network_operator_data.
- RICCIATO, Fabio, LANZIERI, Giampaolo, WIRTHMANN, Albrecht et SEYNAEVE, Gerdy, 2020. Towards a methodological framework for estimating present population density from mobile network operator data. In : *Pervasive and Mobile Computing*. [en ligne]. Octobre 2020. Volume 68. [Consulté le 27 décembre 2024]. Disponible à l'adresse : <https://doi.org/10.1016/j.pmcj.2020.101263>.
- SAKAROVITCH, Benjamin, JOUBERT-DE BELLEFON, Marie-Pierre, GIVORD, Pauline et VANHOOF, Maarten, 2019. Estimer la population résidente à partir de données de téléphonie mobile, une première exploration. In : *Économie et Statistique / Economics and Statistics*. [en ligne]. 11 avril 2019. N° 505-506. [Consulté le 27 décembre 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/3706213?sommaire=3706255>.
- SALGADO, David, SANGUIAO, Luis, OANCEA, Bogdan, BARRAGÁN, Sandra et NECULA, Marian, 2021. An end-to-end statistical process with mobile network data for official statistics. In : *EPJ Data Science*. [en ligne]. 29 avril 2021. [Consulté le 27 décembre 2024]. Disponible à l'adresse : <https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-021-00275-w>.
- SÉMÉCURBE, François, SUAREZ CASTILLO, Milena, GALIANA, Lino, COUDIN, Élise et POULHES, Mathilde, 2020. Que peut faire l'Insee à partir des données de téléphonie mobile ? – Mesure de population présente en temps de confinement et statistiques expérimentales. In : *Le blog de l'Insee*. [en ligne]. 15 avril 2020. [Consulté le 27 décembre 2024]. Disponible à l'adresse : <https://blog.insee.fr/que-peut-faire-linsee-a-partir-des-donnees-de-telephonie-mobile-mesure-de-population-presente-en-temps-de-confinement-et-statistiques-experimentales/>.
- SUAREZ CASTILLO, Milena, SÉMÉCURBE, François, ZIEMLIICKI, Cezary, TAO, Haixuan X. et SEIMANDI, Tom, 2023. Temporally Consistent Present Population from Mobile Network Signaling Data for Official Statistics. In : *Journal of Official Statistics*. [en ligne]. 10 décembre 2023. Vol. 39, N° 4, pp. 535–570. [Consulté le 27 décembre 2024]. Disponible à l'adresse : <https://sciendocom/article/10.2478/jos-2023-0025>.

- TARANTOLA, Arnaud et HAMIDOUCHE, Mohamed, 2025. Cell phone data to correct population estimates and Sars-Cov2 incidence in Ile-de-France during the summer of 2021. In : *Eurosurveillance*. À paraître.
- VANHOOF, Maarten, REIS, Fernando, PLOETZ, Thomas et SMOREDA, Zbigniew, 2018. Assessing the Quality of Home Detection from Mobile Phone Data for Official Statistics. In : *Journal of Official Statistics*. [en ligne]. Décembre 2018. Vol. 34, N° 4, pp. 935–960. [Consulté le 27 décembre 2024]. Disponible à l'adresse : https://www.researchgate.net/publication/329712197_Assessing_the_Quality_of_Home_Detection_from_Mobile_Phone_Data_for_Official_Statistics.