

Sources de données privées : panorama et perspectives



Romain Lesur*

Depuis la fin du XX^e siècle, le numérique a transformé l'économie, générant massivement des traces numériques exploitables par les entreprises et désormais par le système statistique public. Ce dernier s'intéresse à ces données pour compléter les enquêtes traditionnelles, recherchant une plus grande granularité spatiotemporelle, tout en tenant compte des enjeux technologiques, juridiques et méthodologiques qui leur sont associés.

Initialement appelées « big data », ces traces numériques sont aujourd'hui désignées comme des « données détenues par des opérateurs privés », marquant un déplacement d'enjeux vers les questions d'accès et de qualité. L'Insee a intégré les données de caisse des grandes surfaces alimentaires dans la production de l'indice des prix à la consommation et a étendu ses travaux exploratoires aux données de téléphonie mobile, de transactions par carte bancaire, de comptes bancaires ou encore de locations touristiques de courte durée.

Ces nouvelles sources enrichissent la connaissance économique et sociale, mais présentent des limites : couverture partielle, imprécisions géographiques ou temporelles, et absence de variables sociodémographiques. Leur exploitation nécessite des partenariats renforcés avec des opérateurs privés, des infrastructures sécurisées, et une méthodologie rigoureuse.

La stratégie européenne pour les données et la révision du règlement européen n° 223/2009 offrent désormais un cadre pour faciliter l'accès à ces sources. Malgré leur potentiel, ces dernières restent souvent complémentaires aux enquêtes, chacune ayant ses forces et faiblesses, rendant indispensable leur articulation méthodologique afin d'enrichir les statistiques publiques.

 Since the end of the 20th century, digital technology has transformed the economy, generating data at scale that can be exploited by businesses and now by the official statistical system. The latter is interested in this data to supplement traditional surveys, seeking greater spatiotemporal granularity, while taking into account the technological, legal and methodological issues associated with these new data sources.

Initially referred to as 'big data', these data are now referred to as 'privately held data', marking a shift in the focus towards data access and quality. INSEE has integrated scanner data from supermarkets into its consumer price index, and has extended its exploratory work to mobile network operators data, payment bankcards transactions, bank accounts and short-term tourist accommodation.

These new sources enrich our economic and social knowledge, but have their limitations: partial coverage, geographical or temporal inaccuracies, and the absence of socio-demographic variables. Their use requires stronger partnerships with private operators, secure infrastructures and a rigorous methodology.

The European data strategy and the revision of Regulation No 223/2009 now provide a framework to facilitate access to these sources. Despite their potential, these sources are still often complementary to surveys, each with its own strengths and weaknesses, making it essential to combine them methodologically in order to enrich official statistics.

* Chef de l'unité SSP Lab, Insee.
romain.lesur@insee.fr

Depuis la fin du XX^e siècle, l'économie s'est transformée sous l'effet du développement du numérique (Insee, 2019). Les entreprises ont modernisé leur fonctionnement, tant en interne, avec par exemple le recours fréquent à des systèmes intégrés de gestion¹, que dans leurs relations externes, avec notamment l'essor du commerce électronique. Conséquence de ces évolutions, **de très nombreux pans de l'activité des individus et des entreprises donnent désormais lieu à des traces numériques.**

Ces nouvelles données détenues par les entreprises représentent une opportunité pour celles-ci, car elles leur permettent d'optimiser leurs processus internes ou d'améliorer les services offerts à leurs clients (Vacher et Pradines, 2017). Depuis une quinzaine d'années, le système statistique public s'y intéresse également. Il y voit en effet un moyen de compléter ses sources d'information traditionnelles, à savoir les grandes enquêtes qu'il réalise, qui restent encore la fondation du système. Il cherche en particulier à bénéficier d'une plus grande granularité spatiotemporelle, tout en tenant compte des limites méthodologiques de ces données liées à des processus de gestion (Blanchet et Givord, 2017).

Où en est-on aujourd'hui de l'utilisation de ces traces numériques par le système statistique public ?

► Des « big data » aux « données détenues par des opérateurs privés »

Durant les années 2010, le terme le plus communément employé pour désigner ces sources était celui de **données massives** (ou **big data**). Le choix d'un tel vocabulaire soulignait en premier lieu le défi technologique que représentaient leur stockage et leur traitement. Les progrès réalisés depuis une quinzaine d'années ont désormais permis de résoudre ces enjeux, à tel point que le terme « big data » est de moins en moins usité (Tigani, 2023). La statistique publique maîtrise aujourd'hui ces technologies, qui constituent les soubassements des nouvelles plateformes de *data science* telles que le SSP Cloud (Comte et al., 2022). Ces infrastructures facilitent le traitement de données massives, notamment par le recours à de nouveaux formats de données (Dondon et Lamarche, 2023).

Désormais, ces traces numériques sont désignées sous le nom de **données détenues par des opérateurs privés**. Le glissement sémantique n'est pas anodin : parmi les multiples défis qu'elles posent, celui de la technologie aura été certainement le plus vite résolu. Ces nouvelles sources représentent donc le troisième type de données traité par la statistique publique, aux côtés des enquêtes et des données administratives. Elles partagent avec les données administratives le fait qu'elles n'ont pas été produites à des fins de statistique publique et qu'elles doivent donc d'abord être « qualifiées »², avant d'être éventuellement intégrées dans le système d'information statistique (Cotton et Haag, 2023).

¹ Systèmes permettant de centraliser et de rationaliser l'ensemble des données de gestion des entreprises (ressources humaines, comptabilité, activité commerciale, etc.).

² Selon Cotton et Haag (2023), « il est nécessaire d'échanger avec le producteur de la donnée afin de vérifier que la source est : exploitable (les données contenues peuvent être restructurées pour mesurer des concepts statistiques) ; complète (aucune sous-couverture évidente qui empêcherait son exploitation) ; disponible dans un délai raisonnable ; documentée (présence de métadonnées) ».



La première source de données détenues par des opérateurs privés exploitée par l'Insee dans le cadre de sa production statistique correspond aux données de caisse des enseignes de la grande distribution alimentaire.



La première source de données détenues par des opérateurs privés exploitée par l'Insee dans le cadre de sa production statistique correspond aux données de caisse des enseignes de la grande distribution alimentaire. L'Insee les exploite depuis 2020 pour élaborer l'indice des prix à la consommation (Leclair, 2019). À ce jour, il s'agit de la principale source de ce type pleinement intégrée dans un processus de production statistique de l'Insee.

La crise sanitaire de 2020 a renforcé des collaborations déjà établies et en a suscité de nouvelles, avec des détenteurs de données soucieux de leur responsabilité sociétale. Ces acteurs ont souhaité mettre leurs informations à disposition des pouvoirs publics pour les aider à agir durant cette situation d'urgence. Certaines collaborations se sont prolongées à l'issue de la crise sanitaire. L'Insee, et plus largement les systèmes statistiques français et européen, ont ainsi pu approfondir l'examen de ces nouvelles sources afin de qualifier leurs potentiels. Les travaux se sont démultipliés : ils explorent l'utilisation des données détenues par des banques, des plateformes de location immobilière de courte durée, des opérateurs de téléphonie mobile, mais aussi celles des programmes de fidélité (Galiana et Suarez Castillo, 2022), des compteurs communicants (Le Saout et al., 2024) ou encore des « schémas » de paiement par carte bancaire (voir ci-dessous, dans la partie consacrée aux données de transactions par carte bancaire)³.

Sur le long terme, et dans l'optique de leur intégration dans le système d'information statistique, les sources de données privées présentent **trois grands défis pour le service statistique public (figure 1)**. Le premier est celui de la base légale sur laquelle il pourra se fonder pour y accéder dans le cadre de ses missions. Le deuxième porte sur la mise en place d'un cadre partenarial soutenable et pérenne pour traiter ces informations sans rompre leur confidentialité. Enfin, le troisième défi est d'ordre méthodologique, ces données n'ayant pas été collectées pour la réalisation de statistiques publiques.

► Les nouvelles sources de données examinées par la statistique publique

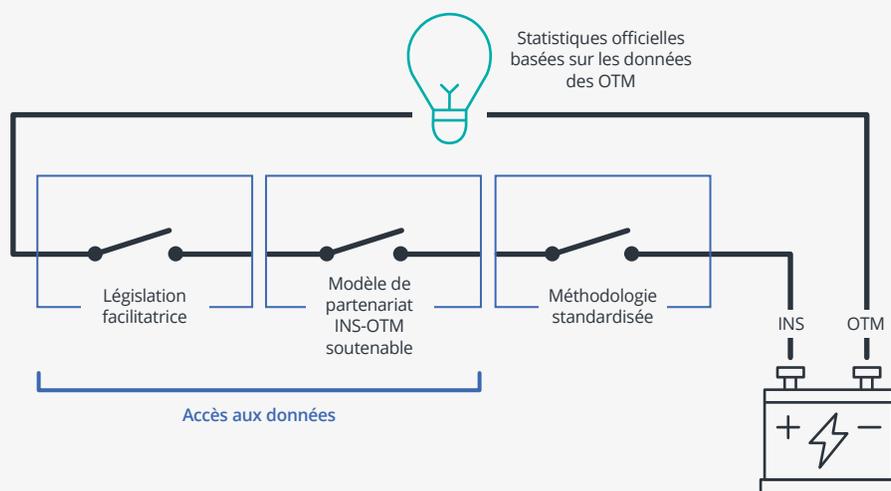
Après les données de caisse, l'Insee s'est donc intéressé à de nouvelles sources de données détenues par les opérateurs privés. Cet article en donne ici un panorama général (figure 2). De manière plus détaillée, les données de comptes bancaires ont fait l'objet d'un article dans le numéro précédent du *Courrier des statistiques* (Bonnet et Loisel, 2024). Les données de téléphonie mobile et de transactions par carte bancaire font, elles, l'objet des deux articles suivants de ce numéro⁴.

³ L'Insee a également exploré le potentiel des données recueillies par Google au travers du service Google Trends pour compléter les enseignements de la statistique publique. L'apport de ces données est cependant limité (Bortoli et Combes, 2015 ; Tavernier et Ourliac, 2020).

⁴ Voir les articles de Joubert sur les données de téléphonie mobile et de Boittelle et al. sur les données de transactions par carte bancaire CB dans ce même numéro.

► Figure 1 - « Fermer le circuit »*

« La production régulière de statistiques officielles basées sur les données des opérateurs de téléphonie mobile ne pourra être mise en place que lorsque toutes les questions en suspens auront été résolues. La définition d'une méthodologie appropriée est une question logiquement distincte de la définition d'un modèle durable d'accès aux données, qui implique l'établissement de modèles de partenariats soutenables entre l'institut national de statistique et les opérateurs de téléphonie mobile dans le cadre d'une législation facilitatrice. »*



INS : Institut national de statistique.

OTM : Opérateur de téléphonie mobile.

* D'après la figure 1 de l'exposé de position d'Eurostat sur l'utilisation à des fins de statistique publique des données de téléphonie mobile (Eurostat, 2023).

Les données de téléphonie mobile

Les données de téléphonie mobile offrent un potentiel désormais reconnu pour compléter la connaissance statistique sur la population et les territoires. Elles permettent d'analyser la présence et la mobilité des personnes dans une zone géographique. Ainsi, il est possible d'estimer la population présente à différents moments de la journée, de la semaine ou de l'année, de mesurer les déplacements quotidiens et même de créer des cartographies dynamiques de la population. Ces données fournissent également des éclairages sur les flux touristiques et la ségrégation résidentielle (Galiana et al., 2020).

L'exploitation de ces données présente **plusieurs avantages indéniables**. Leur richesse informationnelle, leur fréquence d'actualisation et leur granularité géographique complètent les sources traditionnelles de la statistique publique, en tout premier lieu le recensement de la population. Elles peuvent permettre d'analyser finement les déplacements domicile-travail, et par exemple d'appréhender les effets du télétravail, mais aussi de comparer la mobilité entre la semaine et le week-end. Elles peuvent également apporter un éclairage utile aux acteurs publics locaux sur les besoins de protection de la population (services de santé et d'urgence, sécurité et protection civile) ou aux entreprises du commerce de détail sur les besoins en points de vente.



Les données de téléphonie mobile permettent d'analyser finement les déplacements domicile-travail, d'appréhender les effets du télétravail, mais aussi de comparer la mobilité entre la semaine et le week-end.

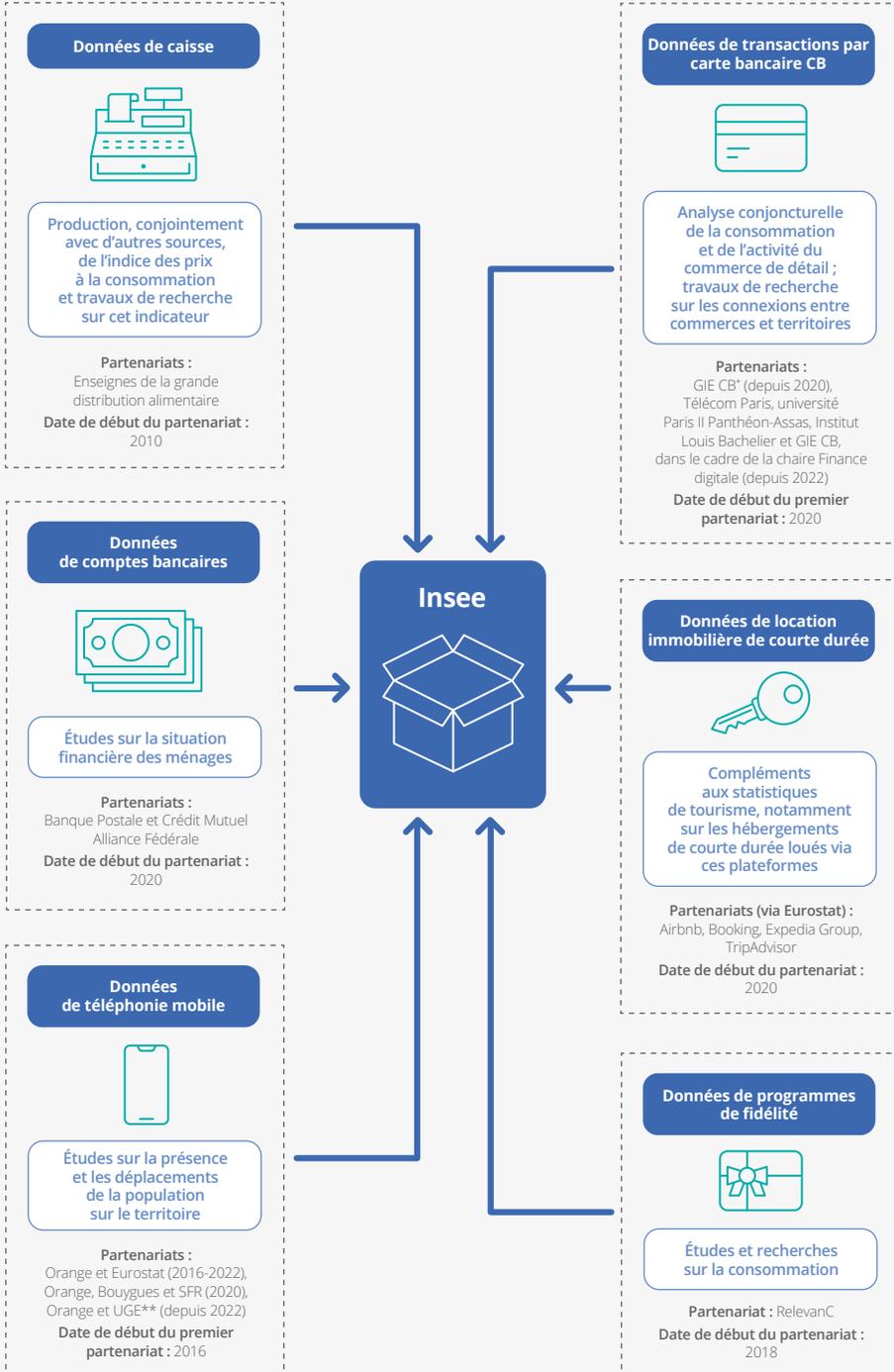


Cependant, leur utilisation soulève **de nombreux enjeux méthodologiques**. Les données de téléphonie mobile sont conçues et optimisées pour la gestion d'un réseau de télécommunications, et non pour la production de statistiques. Leur réutilisation à cette fin nécessite des traitements complexes et des méthodologies spécifiques (Suarez Castillo et al., 2023). Les principales difficultés sont les suivantes :

- L'enregistrement des interactions entre les téléphones mobiles et les antennes relais est irrégulier, en fonction par exemple de l'activité du téléphone, de la densité locale des antennes et du nombre d'utilisateurs qui y sont connectés. Cela engendre une incertitude temporelle pour l'estimation du nombre de personnes présentes dans une zone géographique, et ce tout particulièrement la nuit, lorsque des téléphones sont éteints ou en mode avion.
- La localisation des téléphones mobiles dépend de la couverture des antennes relais de chaque opérateur sur le territoire. Elle est approximative et plus ou moins précise selon les zones géographiques, en fonction de la densité de ces antennes. Cela crée une incertitude spatiale, qui peut aussi être plus élevée la nuit lorsque certaines antennes relais sont éteintes.
- La représentativité des données est partielle, si l'on se limite à un seul opérateur ; chacun ne couvre en effet qu'une partie de la population (Sakarovitch et al., 2019). L'estimation du nombre de personnes présentes à partir des données de téléphonie mobile se heurte aussi au fait que certaines personnes peuvent posséder plusieurs appareils ou abonnements, potentiellement chez différents opérateurs, ou inversement que le même appareil peut être utilisé par plusieurs personnes différentes.
- L'absence de variables sociodémographiques précises concernant les utilisateurs d'appareils mobiles limite les possibilités d'exploitation à des fins de statistique publique. L'hétérogénéité des formats de données et des informations chez les différents opérateurs requiert par ailleurs des traitements complexes.
- Il est indispensable également de définir clairement les concepts mesurés, par exemple celui de « touriste », qui est parmi les plus complexes à aborder à partir des données de téléphonie mobile. En effet, un client d'un opérateur étranger en itinérance peut être détecté sur plusieurs réseaux mobiles français, conduisant ainsi à de multiples comptages.
- Enfin, la stabilité des indicateurs peut être affectée par l'évolution rapide des usages et des technologies.

L'utilisation des données de téléphonie mobile soulève également **des enjeux juridiques importants**. Les questions de confidentialité, de protection de la vie privée et d'accès aux données sont centrales s'agissant de données aussi sensibles que celles de téléphonie mobile. La protection de la vie privée, le respect du règlement général sur la protection des données (RGPD) et la protection du secret des affaires doivent être garantis à chaque étape du processus de traitement de l'information. Ceci impose une organisation et des investissements spécifiques.

► **Figure 2 - Les données d'opérateurs privés déjà explorées par l'Insee**



* GIE CB : Groupement d'intérêt économique Cartes Bancaires CB.

** UGE : Université Gustave Eiffel.

L'exploitation des données de téléphonie mobile nécessite donc une compréhension approfondie de leurs caractéristiques et limites, l'adoption de méthodologies adaptées et la mise en place d'un cadre technique, organisationnel et juridique qui réponde aux exigences et obligations de l'ensemble des acteurs concernés (Coudin et al., 2021⁵). Pour ces raisons, Eurostat⁶ a lancé différents projets sur ces données, dont deux aboutiront en 2025 (**encadré 1**) :

- Le premier projet, intitulé « Multi-MNO », offrira une implémentation de référence d'un processus de traitement des données de téléphonie mobile garantissant leur confidentialité, depuis les opérateurs de téléphonie mobile vers les instituts nationaux de statistique.
- Le second projet, intitulé « MNO-MINDS », proposera un cadre méthodologique de référence pour combiner les données de téléphonie mobile avec d'autres sources, au sein d'un processus de production statistique cohérent.

► Encadré 1. Les projets européens sur les données de téléphonie mobile en 2025

En 2023, Eurostat a attribué des financements européens à deux projets portant sur les données de téléphonie mobile et dans lesquels l'Insee est impliqué : Multi-MNO et MNO-MINDS.

Le projet Multi-MNO* :

Ce projet, auquel l'Insee et Orange France sont associés dans le cadre de son comité consultatif, vise à définir, à des fins de production statistique, un ensemble de traitements standardisés de données ou *pipeline* (mis en œuvre chez les opérateurs de téléphonie mobile) et de flux de données (vers les instituts nationaux de statistique européens). Les traitements et flux de données ont été définis par le système statistique européen en partenariat avec les opérateurs. Afin à la fois de faciliter le déploiement du *pipeline* et la transparence des méthodologies appliquées chez les opérateurs, le projet propose une implémentation open source. Le *pipeline* a pour

caractéristique de protéger intégralement la confidentialité : seules des données agrégées seront transmises aux instituts de statistique, données qui auront donc été calculées suivant des méthodologies communes et transparentes. Ce projet vise également à démontrer la faisabilité de mise en œuvre du *pipeline* en l'appliquant à des données réelles.

Le projet MNO-MINDS** :

Ce projet est coordonné par l'institut national de statistique italien (Istat) et associe dix partenaires, dont l'Insee. L'objectif est de proposer des méthodes, ainsi que des librairies open source, dédiées à l'intégration des données de téléphonie mobile avec d'autres sources, à des fins de production régulière de statistiques publiques. Ce projet proposera également des formations à ces méthodes et outils.

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

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

Les données de transactions par carte bancaire

Les données issues des transactions par carte bancaire apportent des informations complémentaires à celles de la statistique publique, notamment pour le **suivi conjoncturel et la prévision des indicateurs économiques**. Une transaction par carte bancaire mobilise sept acteurs : l'acheteur, le commerçant, leurs banques respectives, le schéma de paiement (CB, Visa, Mastercard, etc.), le réseau interbancaire d'autorisation et le système interbancaire d'autorisation⁷. L'ensemble des transactions génère de

5 Voir également l'article de Joubert sur les données de téléphonie mobile dans ce même numéro.

6 Eurostat est l'institut statistique communautaire, direction générale de la Commission européenne.

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



Les données de transactions par carte bancaire peuvent améliorer le suivi conjoncturel et la prévision de certains indicateurs d'activité, comme l'indice de chiffre d'affaires du commerce de détail.



nombreuses données, dont les montants, dates et heures des opérations et les références bancaires des acheteurs et vendeurs. Une fois agrégées et enrichies, par exemple avec l'activité principale et la localisation du commerçant, ces données permettent d'analyser en profondeur les comportements de consommation et l'activité commerciale.

Face aux délais de collecte et de traitement de ses sources de données traditionnelles, l'Insee a cherché à mobiliser ces données, à haute fréquence et disponibles rapidement. Elles

peuvent en effet améliorer le suivi conjoncturel et la prévision de certains indicateurs d'activité, comme l'indice de chiffre d'affaires du commerce de détail. Les données du **groupement d'intérêt économique Cartes Bancaires CB (GIE CB)⁸**, qui est le schéma domestique français de paiement par carte et par mobile, couvrent un large spectre d'activités commerciales. Elles offrent un aperçu précis des dynamiques territoriales et sectorielles et ont notamment permis d'éclairer les conséquences économiques du confinement de 2020 (Insee, 2020).

Le partenariat entre l'Insee et le GIE CB – renforcé depuis la crise sanitaire et via la **chaire de recherche Finance digitale⁹** – repose sur une transmission sécurisée à l'Insee de données préagrégées. Ces dernières sont issues d'un travail de structuration et d'enrichissement des données individuelles anonymisées. Le dispositif permet l'exploitation d'informations détaillées sur chaque transaction tout en garantissant leur confidentialité. Des contrôles rigoureux, basés sur la comparaison entre données d'autorisation, de compensation et d'activité de paiement, permettent d'analyser la qualité et la fiabilité des agrégats exploités pour la production statistique.

Ces données présentent cependant **des limites intrinsèques**, qui nécessitent des corrections pour ne pas fausser les analyses :

- Elles ne couvrent qu'une partie des transactions réalisées en France. Elles ne concernent en effet que les paiements effectués par des personnes résidant en France auprès de commerçants affiliés à une banque française, et uniquement par carte bancaire CB. Elles n'intègrent donc pas les paiements en espèces ou par chèque, les virements ou les paiements en ligne via d'autres schémas de paiement que CB (par exemple les schémas internationaux Visa et Mastercard).
- Elles ne permettent pas de distinguer, au sein des paiements par carte, les dépenses professionnelles des dépenses personnelles.
- Elles sont sensibles à l'évolution de la couverture des dépenses par le schéma de paiement CB, liée notamment à la concurrence entre schémas et aux changements éventuels de comportements (par exemple l'essor du paiement par mobile).

⁸ <https://www.cartes-bancaires.com/cb/groupement/>.

⁹ <https://digital-finances.com/>. Voir à ce sujet l'article de Boittelle et al. sur les données de transactions par carte bancaire CB dans ce même numéro.

- Enfin, des imprécisions sur l'activité et la localisation des commerçants peuvent altérer l'analyse.

Durant la crise sanitaire, le caractère inframensuel de ces données et leur rapidité de mise à disposition ont permis d'obtenir des informations précieuses pour l'analyse conjoncturelle. Néanmoins, leur utilisation sur le moyen terme pour cette finalité reste encore une question ouverte, en raison d'une forte volatilité et d'une couverture fluctuante. En revanche, le chaînage des transactions pour une même carte offre la possibilité d'éclairer les **liens entre commerces et territoires** (c'est-à-dire où vont les habitants d'un territoire pour faire leurs achats). Il permet d'envisager l'utilisation de ces données pour analyser les effets de l'implantation ou de la disparition de certains points de vente sur l'activité commerciale infracommunale, ou bien l'évaluation de politiques publiques portant sur l'activité commerciale. Les comportements effectifs de consommation de la population observés à travers ces données peuvent également compléter les sources d'information déjà existantes sur les connexions entre territoires. En effet, les zonages actuels élaborés par l'Insee établissent ces connexions principalement à partir des déplacements domicile-travail (**aires d'attraction des villes¹⁰**) ou des distances des populations aux équipements les plus proches (**bassins de vie¹¹**).

Au total, ces données enrichissent la palette des indicateurs économiques en offrant une vision plus fine des comportements de consommation et des dynamiques territoriales, tout en posant des défis méthodologiques pour leur intégration dans les processus statistiques traditionnels.

Les données de comptes bancaires

Le numéro précédent du Courrier des statistiques présente les travaux menés par l'Insee à partir d'échantillons anonymisés de comptes bancaires de La Banque Postale et du Crédit Mutuel Alliance Fédérale (Bonnet et Loisel, 2024). L'institut s'est intéressé à ces données en tout premier lieu pour l'**analyse de la conjoncture économique**.



Les relevés mensuels de comptes bancaires permettent d'étudier les revenus et dépenses des ménages sur une base journalière.



En effet, leur fraîcheur, leur granularité fine, la grande taille des échantillons considérés et la variété des renseignements disponibles favorisent des analyses précises et quasiment en temps réel des comportements financiers des ménages. Ainsi, les relevés mensuels de comptes bancaires, délivrés dès la fin du mois suivant la période observée, voire au milieu du mois, permettent d'étudier les revenus et dépenses des ménages sur une base journalière. La diversité des informations recueillies autorise également une analyse fine de la **diversité des**

comportements des ménages et des **inégalités de consommation** (y compris sur des trajectoires longues, grâce à la possibilité de constituer des panels). Des populations spécifiques peuvent par ailleurs être étudiées avec davantage de précision, car les effectifs disponibles sont plus importants que dans les enquêtes traditionnelles.

¹⁰ <https://www.insee.fr/fr/metadonnees/definition/c2173>.

¹¹ <https://www.insee.fr/fr/metadonnees/definition/c2060>.

La possibilité d'étudier de manière détaillée l'impact de chocs économiques, tels que des variations brutales de prix ou de revenus, est un autre atout majeur de ces données. On peut ainsi estimer, par exemple, dans quelle mesure les ménages réagissent à une hausse de prix en puisant dans leur épargne ou en réduisant leur consommation. L'Insee a pu étudier grâce à ces données les effets budgétaires, redistributifs et environnementaux des remises sur le prix des carburants après les hausses liées à la guerre en Ukraine (Adam et al., 2023). Ce potentiel très riche d'analyse tient en grande partie à la nature même des données collectées, qui incluent les soldes de comptes, le détail et les dates des opérations, ainsi que certaines informations sociodémographiques. Les banques proposent de surcroît une typologie des dépenses, grâce à une catégorisation des paiements par carte selon le type d'établissement bénéficiaire, ce qui se révèle précieux pour analyser la composition de la consommation.

Ces nombreux avantages n'évident pas les limites des données, à commencer par un problème de représentativité :

- Les personnes non bancarisées ne sont pas prises en compte. Par ailleurs, l'observation des seuls clients d'une banque spécifique, même si cette clientèle couvre un large spectre de la population, est sensible aux effets de spécialisation de chaque établissement.
- La vision de la situation financière est partielle. En effet, les clients sont parfois multibancarisés et peuvent détenir d'autres revenus ou avoir recours à d'autres canaux de dépenses qui échappent à l'analyse. Les dimensions comme le patrimoine immobilier ne sont pas non plus accessibles.
- À ces biais s'ajoute la difficulté d'établir des correspondances exactes avec les concepts usuels de la statistique publique. La notion de ménage doit ainsi être approchée par celle de groupe familial, aboutissant à une mesure incomplète des revenus et des dépenses, si certains membres du ménage ont un compte dans une autre banque. De plus, certaines transactions ne traduisent pas de véritables opérations de consommation ou de perception de revenus, par exemple des transferts entre comptes d'un même individu entre des banques différentes.
- Ces indicateurs peuvent se révéler volatils et induire des conclusions hâtives démenties ensuite lorsque l'environnement économique est moins turbulent.

Plusieurs défis se posent pour maximiser le potentiel de ces données et améliorer la qualité des analyses. Le premier consiste à accroître la représentativité en développant de nouveaux partenariats avec davantage d'établissements afin, d'une part, de potentiellement mieux cerner les personnes multibancarisées et, d'autre part, d'englober un éventail plus large de ménages. Le second concerne la profondeur historique, puisqu'un historique plus ancien permet de mieux appréhender l'évolution des comportements financiers sur le long terme. Une fois ces conditions réunies, la désaisonnalisation¹² prendra tout son sens, car elle contribuera à atténuer une partie du bruit¹³ inhérent à la haute fréquence d'observation et facilitera la détection de tendances conjoncturelles.

¹² Désaisonnaliser consiste à appliquer un traitement statistique pour éliminer les effets dus aux phénomènes saisonniers.

¹³ Fluctuations non directement liées au phénomène que l'on cherche à analyser.

Les efforts d'amélioration de la catégorisation des flux financiers constitueront également un atout précieux pour rapprocher progressivement les transactions des définitions usuelles de la statistique publique. Enfin, la préservation de la confidentialité des données demeure centrale, afin de garantir l'anonymat des clients et de protéger leur vie privée. L'objectif à long terme est de parvenir à combiner ces données avec d'autres sources pour disposer d'une vision plus exhaustive du comportement économique des ménages, tout en maintenant des standards élevés de fiabilité et de confidentialité.

Les données de location immobilière de courte durée

Début 2020, Eurostat a conclu un accord d'échanges de données avec quatre plateformes de location de courte durée : *Airbnb*, *Booking*, *Expedia Group* et *TripAdvisor*. La collecte de données auprès des plateformes permet d'améliorer la qualité des statistiques européennes sur le tourisme. En effet, s'agissant du marché de l'hébergement, le segment des locations de vacances de courte durée n'est traditionnellement couvert que partiellement. Les acteurs de ce marché de location sont nombreux, ce qui rend la collecte de données plus difficile. Une partie importante des locations de vacances de la nomenclature d'activité européenne « vacances et autres hébergements de court séjour » (code NACE¹⁴ 55.2) n'est ainsi pas représentée.

Les données de location immobilière de courte durée viennent combler les lacunes des enquêtes traditionnelles sur le tourisme, qui ne couvrent pas systématiquement ce segment en plein essor.

Les données mises à disposition par ces plateformes couvrent les hébergements de courte durée (à l'exclusion des hôtels et campings) réservés par leur intermédiaire dans l'Union européenne (UE) et dans les pays de l'Association européenne de libre-échange (AELE¹⁵). Les plateformes fournissent trimestriellement à Eurostat des données comme les nombres de nuitées réservées et de voyageurs. Les statistiques obtenues ont un caractère expérimental, comprenant notamment des risques de double compte.

Ces données permettent d'éclairer l'impact économique et social des plateformes de location immobilière de courte durée dans différents pays européens (voir par exemple Ulrich (2021) pour la France). Elles viennent combler les lacunes des enquêtes traditionnelles sur le tourisme, qui ne couvrent pas systématiquement ce segment en plein essor. Leur utilisation est également envisagée pour mieux prendre en compte l'hébergement de courte durée via les plateformes dans le futur système de comptabilité nationale (Askenazy et Bourgeois, 2025). Ces travaux illustrent en même temps l'importance d'institutions telles qu'Eurostat dans la négociation d'accords avec des acteurs de l'économie numérique qui sont simultanément présents sur différents marchés nationaux. L'échelon européen est, dans ce contexte, mieux adapté pour négocier avec ces acteurs.

¹⁴ NACE : nomenclature statistique des activités économiques dans la Communauté européenne : <https://www.insee.fr/fr/metadonnees/definition/c2073> et <https://ec.europa.eu/eurostat/fr/web/nace>.

¹⁵ L'AELE compte actuellement quatre pays membres: l'Islande, le Liechtenstein, la Norvège et la Suisse.

► Évolutions institutionnelles et juridiques

Avec la publication de la **stratégie européenne pour les données**¹⁶ par la Commission européenne en 2020, l'échelon européen s'est emparé de la question de l'accès aux données détenues par le secteur privé à des fins d'intérêt général. Dans ce contexte, les institutions européennes ont également finalisé, fin 2024, **une révision du règlement n° 223/2009 relatif aux statistiques européennes**¹⁷, via l'adoption d'un règlement rectificatif : le règlement n° 2024/3018 modifiant le règlement n° 223/2009 relatif aux statistiques européennes¹⁸. Les nouvelles dispositions introduites par cette révision traitent notamment de l'accès aux sources de données privées par le système statistique européen. Ces différents textes précisent par ailleurs ce qu'ils entendent sous le terme de données (**encadré 2**).¹⁹

► Encadré 2. Les définitions des « données » par le droit européen

Dans le cadre de la stratégie européenne pour les données, l'article 2 du *Data Governance Act*²⁰ définit ce que le règlement entend comme étant une « donnée ». Il s'agit de « toute représentation numérique d'actes, de faits ou d'informations et toute compilation de ces actes, faits ou informations, notamment sous la forme d'enregistrements sonores, visuels ou audiovisuels ». Pour la première fois, un texte européen donne une définition juridique à la notion de « données ».

Pour le statisticien public, cette définition ne va pas nécessairement de soi, les données pouvant exister même si elles n'ont pas de représentation numérique (cas par exemple de tous les anciens

registres administratifs sur support papier). Il est d'ailleurs amusant de constater que deux ans et demi plus tard, le règlement n° 2024/3018 introduit dans le règlement n° 223/2009 relatif aux statistiques européennes une définition différente de la donnée : « toute représentation numérique ou non d'actes, de faits ou d'informations et toute compilation de tels actes, faits ou informations sur les unités observées ».

Le droit européen qui, jusqu'à récemment, ne définissait pas ce qu'était une donnée repose désormais sur deux définitions différentes de cette notion. Bien évidemment, la définition applicable à la statistique européenne est celle du règlement n° 223/2009.

* Voir les références juridiques en fin d'article.

La stratégie européenne pour les données

La stratégie européenne pour les données comporte deux textes majeurs : le *Data Governance Act* (règlement sur la gouvernance des données)²⁰ et le *Data Act* (règlement sur les données)²¹. Certains points intéressent en particulier la statistique publique.

Le *Data Governance Act* définit le cadre juridique des **espaces européens de données** (*data spaces*). La Commission européenne soutient leur développement pour faciliter la mise en commun et le partage des données dans des secteurs clés, notamment dans divers champs sectoriels de l'économie (agriculture, santé, énergie, transports, etc.). Si ces espaces se développent et se pérennisent, ils peuvent offrir des perspectives

¹⁶ https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_fr.

¹⁷ Voir les références juridiques en fin d'article.

¹⁸ Voir les références juridiques en fin d'article.

¹⁹ Au niveau de la France, par ailleurs, un rapport d'information de l'Assemblée nationale propose des évolutions du cadre juridique et des pratiques entourant l'accès aux sources de données privées afin de faciliter leur utilisation par l'Insee (Sala, 2023).

²⁰ Voir les références juridiques en fin d'article.

²¹ Voir les références juridiques en fin d'article.

intéressantes pour la statistique européenne. D'ores et déjà a été créé en mars 2025 l'**espace européen des données de santé**²². Datant d'avant le *Data Governance Act*, mais dans le même esprit, l'**espace de données Copernicus**²³ offre un accès aux images satellitaires ainsi qu'un écosystème d'outils et de données dérivées qui en facilitent l'usage²⁴.

Le *Data Act* porte principalement sur les données générées par l'Internet des objets²⁵ et sur la concurrence sur le marché du *cloud* (informatique en nuage). En outre, il définit un cadre juridique pour le partage de données des entreprises vers les pouvoirs publics (**encadré 3**). Ces règles concernent l'ensemble des pouvoirs publics, dont bien évidemment la statistique publique.

► **Encadré 3. Les nouveaux droits des pouvoirs publics introduits par le *Data Act* pour accéder à des données détenues par des entreprises en cas de besoin exceptionnel**

Le *Data Act* donne un cadre juridique au partage de données des entreprises vers les pouvoirs publics, communément appelé « B2G » pour *business to government*. Concrètement, les dispositions du chapitre V donnent plus de droits aux pouvoirs publics pour demander des données au secteur privé, mais seulement en cas de nécessité et pour un besoin exceptionnel, ce qui en limite la durée et la portée.

Si le besoin exceptionnel correspond à une situation d'urgence, les données sont mises à disposition gratuitement, et les pouvoirs publics accordent « une reconnaissance publique » aux entreprises qui le demandent.

Hors situation d'urgence, seul l'accès à des données à caractère non personnel est possible pour répondre à un besoin exceptionnel, et uniquement en tout dernier recours : après avoir épuisé tous les autres moyens dont disposent les pouvoirs publics, y compris l'adoption de nouvelles mesures législatives ou l'achat de ces données sur le marché. Lorsque ces conditions sont réunies, le détenteur de données a le droit de demander une compensation visant à couvrir les coûts techniques et organisationnels pour répondre à la demande de transmission de données, ainsi qu'une marge raisonnable.

La révision du règlement n° 223/2009 relatif aux statistiques européennes

La révision du règlement n° 223/2009 relatif aux statistiques européennes introduit d'abord de nouvelles définitions. Parmi celles-ci, un « détenteur de données » est « une personne physique ou morale ou toute autre entité qui a le droit [...] ou la capacité, de gérer et de mettre à disposition des données obtenues dans le cadre de son activité ».

Surtout, **cette révision dote Eurostat et les instituts nationaux de statistique européens de prérogatives nouvelles pour demander l'accès à des sources de données privées à des fins statistiques, en plus de celles prévues par le *Data Act* en cas de besoin exceptionnel**. Le texte s'applique à des données dont il convient de montrer qu'elles sont strictement « nécessaires pour le développement, la production

22 https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space-regulation-ehds_fr.

23 <https://www.copernicus.eu/fr/acces-aux-donnees>.

24 De nombreux instituts nationaux de statistique, dont l'Insee, explorent l'usage de l'imagerie spatiale et de la photographie aérienne pour la statistique publique. La plupart de ces données étant publiques, elles ne sont pas évoquées dans cet article.

25 Systèmes interconnectés permettant de superviser et de contrôler des objets à distance grâce à des échanges de données par Internet (Boudrot, 2021).

et la diffusion de statistiques européennes et que celles-ci ne peuvent être obtenues autrement ou que leur réutilisation entraînera une réduction considérable de la charge de réponse pesant sur les détenteurs de données et d'autres entreprises ». Il est important de noter que ces finalités statistiques comprennent les activités scientifiques et de recherche des instituts de statistique ainsi que l'établissement de bases de sondage.

S'agissant d'une potentielle compensation financière, le nouvel article 17 ter du règlement dispose que la mise à disposition des données et métadonnées est gratuite, mais aussi que « [l]orsque les données demandées [...] nécessitent un service de traitement spécifique, les États membres peuvent accorder une compensation au détenteur de données privé pour ce service, sauf lorsque le droit national proscrit l'indemnisation des détenteurs de données ». Par conséquent, la question de la gratuité de l'accès aux sources de données privées est renvoyée au droit de chaque État.



Le nouveau cadre européen invite à établir des relations partenariales avec les détenteurs de données privés.



Le nouveau cadre défini par cette révision invite aussi à établir des relations partenariales avec les détenteurs de données privés. En effet, après une demande de mise à disposition de données par Eurostat ou un institut national de statistique, un dialogue doit s'engager entre le détenteur de données privé et l'administration ayant formulé la demande « afin de discuter et de convenir des mesures requises pour la mise à disposition des données [...] en vue de conclure un accord ». La demande d'accès ne suffit donc pas,

il est nécessaire d'engager une discussion pour trouver un accord. Si aucun accord n'est conclu dans un délai de trois mois, une seconde demande peut être formulée. Elle deviendra alors opposable en droit, c'est-à-dire contraignante pour les détenteurs de données. L'accent mis sur le dialogue entre les acteurs du système statistique européen et les détenteurs de données privés est ainsi proche du mécanisme de concertation qu'a prévu le législateur aux travers des dispositions de l'article 3 bis de la loi n° 51-711 du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques²⁶.

Les nouvelles dispositions du règlement prévoient également qu'Eurostat puisse mettre en place une infrastructure sécurisée pour faciliter le partage de ces données au sein du système statistique européen. Ces dispositions sont en lien avec les investissements qu'Eurostat mène sur les infrastructures de calcul multipartite sécurisé (Ricciato, 2024) au travers notamment du projet JOCONDE²⁷.

► **Mettre en place un cadre partenarial soutenable**

L'orientation prise par le système statistique européen vise résolument à la mise en place de partenariats avec les détenteurs de données privés. Cette orientation était présente dès 2022 dans les conclusions du groupe européen de haut niveau sur l'utilisation des nouvelles sources de données pour la statistique publique (Eurostat, 2022).

²⁶ Voir les références juridiques en fin d'article.

²⁷ Joint On-demand Computation with No Data Exchange : <https://cros.ec.europa.eu/joconde>.

Cette voie correspond, dans les faits, à celle déjà entreprise dans différents pays européens ainsi qu'en France depuis les années 2010. La principale raison d'être de ces partenariats tient à la nature des données. Aucune source de données privée n'a été conçue à des fins statistiques. Ces sources ont toutes été conçues pour d'autres finalités : gérer un réseau de télécommunications, gérer un système de paiement par carte, offrir un service d'intermédiation et de mise en relation sur internet, gérer des comptes bancaires, etc.

L'investissement nécessaire pour construire des indicateurs qui répondent aux exigences de la statistique publique est conséquent et nécessite un réel travail partenarial. En effet, les statisticiens publics ne peuvent comprendre seuls la complexité que représente la gestion d'un réseau de télécommunications ou d'un système de paiement. Réciproquement, les détenteurs de données privés ne connaissent pas la statistique publique. Les données qu'utilise le service statistique public doivent notamment être alignées sur des référentiels, tels que les nomenclatures statistiques. À cet égard, les données diffusées en open data par la statistique publique participent à l'appropriation de ces référentiels par le secteur privé. Le répertoire [Sirene](#)²⁸ est ainsi très largement exploité par le secteur privé, diffusant par là-même l'utilisation de la [nomenclature d'activités française](#)²⁹ (NAF) au sein de cette sphère.

Par ailleurs, les obligations et exigences de protection de la vie privée auxquelles doivent répondre les détenteurs de données privés représentent autant de freins à la mobilité des données dont ils sont responsables (Desrochers, 2024). Suivant les circonstances, il peut être préférable qu'ils ne transfèrent pas leurs données brutes à l'extérieur de leur système d'information. Dès lors, il peut devenir nécessaire de concevoir et mettre en place des systèmes et des méthodologies qui, à la fois, répondent aux besoins de la statistique publique et prennent en compte ces contraintes. Dans certains cas, un co-investissement de la statistique publique et de ces opérateurs sera la meilleure voie pour trouver des solutions qui répondent aux besoins et contraintes de l'ensemble des partenaires. Dans cette optique, il pourrait être envisagé de recourir à des technologies améliorant la confidentialité ou « PETs » (*Privacy Enhancing Technologies*), telles que la confidentialité différentielle (*differential privacy*) (Tassi, 2019) ou le calcul multipartite sécurisé (Ricciato, 2024). Cependant, ces dernières restent encore particulièrement complexes et coûteuses à mettre en œuvre.



Les accords passés entre le système statistique public et des détenteurs de données privés restent néanmoins fragiles dans le temps.



Les accords passés entre le système statistique public et des détenteurs de données privés restent néanmoins fragiles dans le temps. La concurrence existant dans le secteur privé peut conduire à un changement de contrôle, voire à la disparition de certains d'entre eux. Les accords jusqu'ici trouvés avec des acteurs privés ont souvent reposé sur une prise de conscience de leur responsabilité sociétale. En ce sens, l'épisode de la crise sanitaire de 2020 a été un catalyseur très positif. Il est néanmoins impossible de présager de l'avenir, et de savoir si

²⁸ <https://www.insee.fr/fr/metadonnees/source/serie/s1020>.

²⁹ <https://www.insee.fr/fr/information/2406147>.

les partenariats avec les instituts de statistique continueront de trouver une place dans leurs stratégies. Les nouvelles prérogatives du système statistique européen – en ce qu’elles permettent de rendre obligatoire la transmission de données qu’ils détiennent – atténuent ces risques.

► Les principaux enjeux méthodologiques associés aux sources de données privées

Les sources de données privées proviennent d’acteurs agissant sur un marché concurrentiel. Ces détenteurs de données peuvent donc être nombreux, se différencier sur leurs offres et, partant, leurs clientèles. Dès lors, les données d’un seul acteur ne couvrent pas l’ensemble du champ et peuvent souffrir d’un problème de **représentativité**. Plusieurs stratégies sont alors possibles. On peut chercher à accroître le champ – en augmentant le nombre d’acteurs auprès desquels sont collectées les données – afin qu’il soit le plus large possible (Coudin et al., 2021). Dans certains cas, on peut aussi échantillonner les données ou les caler, si les sources de données comprennent des informations auxiliaires le permettant (Bonnet et Loisel, 2024).

Les sources de données privées souffrent très fréquemment de **différences de concept** au regard de ce que le statisticien cherche à mesurer. Par exemple, en ce qui concerne les données bancaires, les données d’une seule banque ne permettent pas d’avoir une mesure exhaustive de la situation financière des individus du fait du cas des personnes multibancairisées. De même, les paiements par carte bancaire ne mesurent pas les dépenses en espèces ou par chèque.

Le fait que les consommateurs aient recours aux services de plusieurs acteurs concurrents peut conduire à des **comptages multiples**. C’est le cas par exemple, de la téléphonie mobile pour les personnes disposant d’une ligne personnelle et d’une ligne professionnelle, ou bien des clients d’opérateurs étrangers en itinérance qui peuvent être détectés successivement chez différents opérateurs.

Dans certains cas, le **passage à l’unité statistique utilisée par le statisticien** est source de complexité. En effet, il n’y a pas toujours identité entre le souscripteur d’un service et ses utilisateurs : abonnement de téléphonie mobile au nom d’un parent utilisé par un enfant, carte de paiement utilisée par un autre membre de la famille, compte bancaire conjoint, etc.

Afin de corriger la plupart de ces problèmes, il est souvent procédé à des enquêtes pour en évaluer leurs ampleurs. Cela conduit à un paradoxe, car ces nouvelles sources de données ont souvent été présentées comme pouvant réduire le recours à des enquêtes.

► Quel avenir, alors, pour ces nouvelles sources ?

La documentation accumulée sur ces enjeux méthodologiques est le résultat des partenariats mis en œuvre dans le service statistique public avec des détenteurs de données privés. La connaissance de ces sources, de leur potentiel, mais aussi de leurs limites est précisément ce qui peut permettre, à terme, leur intégration aux côtés des autres sources mobilisées par le service statistique public.



Ces données privées offrent d'ores et déjà des connaissances inaccessibles à l'aide d'enquêtes. Elles ne permettent pas pour autant de s'y substituer, chaque source ayant ses forces et faiblesses.



Ces données privées offrent d'ores et déjà des connaissances inaccessibles à l'aide d'enquêtes. Elles ne permettent pas pour autant de s'y substituer, chaque source ayant ses forces et faiblesses. Les enquêtes restent ainsi le seul moyen de couvrir un champ de façon exhaustive. Par ailleurs, même si les prérogatives des instituts nationaux de statistique ont été renforcées par la révision du règlement n° 223/2009, les accès aux sources de données privées reposent sur des accords qui n'offrent pas la même garantie de pérennité.

Aujourd'hui, elles apparaissent donc bien comme des sources complémentaires, qui peuvent trouver leur place aux côtés des enquêtes et des sources administratives déjà exploitées par le service statistique public, et permettre ainsi d'enrichir la production statistique.

Comme Harford l'avait expliqué dès 2014, ces nouvelles traces numériques des entreprises privées ne sont pas la panacée, ne peuvent se substituer aux enquêtes déjà existantes, et ne peuvent faire l'économie d'une analyse méthodologique approfondie (Harford, 2014). Mais elles permettent d'offrir de nouveaux éclairages pour enrichir la connaissance de la société et de l'économie française.

► Fondements juridiques

- *Data Governance Act* : 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>.
- *Data Act* : règlement (UE) n° 2023/2854 du Parlement européen et du Conseil du 13 décembre 2023 concernant des règles harmonisées portant sur l'équité de l'accès aux données et de l'utilisation des données et modifiant le règlement (UE) n° 2017/2394 et la directive (UE) n° 2020/1828. In : *site de l'Union européenne*. [en ligne]. [Consulté le 25 mars 2025]. Disponible à l'adresse : <https://eur-lex.europa.eu/eli/reg/2023/2854/oj>.
- Règlement (UE) n° 2024/3018 du Parlement européen et du Conseil du 27 novembre 2024 modifiant le règlement (CE) n° 223/2009 relatif aux statistiques européennes. In : *site de l'Union européenne*. [en ligne]. [Consulté le 25 mars 2025]. Disponible à l'adresse : <https://eur-lex.europa.eu/eli/reg/2024/3018/oj>.
- Loi n° 51-711 du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques. In : *site de Légifrance*. Mise à jour le 25 mars 2019. [en ligne]. [Consulté le 25 mars 2025]. Disponible à l'adresse : <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000888573>.

► Bibliographie

- ADAM, Marine, BONNET, Odran, FIZE, Étienne, RAULT, Marion, LOISEL, Tristan, WILNER, 2023. L'ajustement de court terme de la consommation de carburant à des changements de prix – Des estimations menées à partir de données à haute fréquence. In : *Documents de travail*. [en ligne]. 7 juillet 2023. [Consulté le 25 mars 2025]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/7644834>.
- ASKENAZY, Philippe et BOURGEOIS, Alexandre, 2025. Vers une meilleure prise en compte de l'hébergement via des plates-formes en ligne au sein des comptes nationaux. In : *Documents de travail*. [en ligne]. 4 mars 2025. [Consulté le 25 mars 2025]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/8376381>.
- BLANCHET, Didier et GIVORD, Pauline, 2017. Données massives, statistique publique et mesure de l'économie. In : *L'économie française, coll. « Insee Références »*. [en ligne]. 11 juillet 2017. Pp. 59-77. [Consulté le 25 mars 2025]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/2894010?sommaire=2894036>.
- BONNET, Odran et LOISEL, Tristan, 2024. L'économie racontée par les données bancaires – Ce que nos relevés de comptes disent de nous. In : *Courrier des statistiques*. [en ligne]. 16 décembre 2024. Insee. N° N12, pp. 115-136. [Consulté le 25 mars 2025]. Disponible à l'adresse : <https://www.insee.fr/fr/information/8264558?sommaire=8264562>.
- BORTOLI, Clément et COMBES, Stéphanie, 2015. Apports de Google Trends pour prévoir la conjoncture française : des pistes limitées. In : *Note de conjoncture*. [en ligne]. 2 avril 2015. Insee. Pp. 43-56. [Consulté le 20 mai 2025]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/1408926?sommaire=1408931>.
- BOUDROT, Nicolas, 2021. Internet des objets, impression 3D, robotique : des technologies davantage utilisées par les grandes sociétés. In : *Insee Première*. [en ligne]. 21 avril 2021. Insee. N° 1854. [Consulté le 12 mai 2025]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/5356737>.
- COMTE, Frédéric, DEGORRE, Arnaud et LESUR, Romain, 2022. Le SSPCloud : une fabrique créative pour accompagner les expérimentations des statisticiens publics. In : *Courrier des statistiques*. [en ligne]. 20 janvier 2022. Insee. N° N7, pp. 68-87. [Consulté le 25 mars 2025]. Disponible à l'adresse : <https://www.insee.fr/fr/information/6035940?sommaire=6035950>.
- COTTON, Franck et HAAG, Olivier, 2023. L'intégration des données administratives dans un processus statistique – Industrialiser une phase essentielle. In : *Courrier des statistiques*. [en ligne]. 30 juin 2023. Insee. N° N9, pp. 104-125. [Consulté le 25 mars 2025]. Disponible à l'adresse : <https://www.insee.fr/fr/information/7635829?sommaire=7635842>.
- 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 : *Data & Policy*. [en ligne]. 24 juin 2021. Vol. 3, p. e10. [Consulté le 25 mars 2025]. Disponible à l'adresse : <https://doi.org/10.1017/dap.2021.1>.

- DESROCHERS, Pierre R., 2024. Access to Information and Privacy: Practical Approaches for Public Service Reform. In : *Canadian Public Administration*. [en ligne]. 19 décembre 2024. Volume 67, Issue 4. Pp. 562-572. [Consulté le 25 mars 2025]. Disponible à l'adresse : <https://doi.org/10.1111/capa.12582>.
- DONDON, Alexis et LAMARCHE, Pierre, 2023. Quels formats pour quelles données ? In : *Courrier des statistiques*. [en ligne]. 30 juin 2023. Insee. N° N9, pp. 86-103. [Consulté le 25 mars 2025]. Disponible à l'adresse : <https://www.insee.fr/fr/information/7635827?sommaire=7635842>.
- EUROSTAT, 2022. Empowering society by reusing privately held data for official statistics: final report prepared by the high level expert group on facilitating the use of new data sources for official statistics — A European approach — 2022 edition. In : *site de Eurostat*. [en ligne]. Publications Office of the european Union. [Consulté le 25 mars 2025]. Disponible à l'adresse : <https://ec.europa.eu/eurostat/web/products-statistical-reports/-/ks-ft-22-004>.
- EUROSTAT, 2023. Reusing Mobile Network Operator data for Official Statistics: the case for a common methodological framework for the European Statistical System, 2023 edition. In : *site de Eurostat*. [en ligne]. [Consulté le 16 mai 2025]. Publications Office of the european Union. Disponible à l'adresse : <https://ec.europa.eu/eurostat/web/products-statistical-reports/w/ks-ft-23-001>.
- GALIANA, Lino, SAKAROVITCH, Benjamin, SÉMÉCURBE, François et SMOREDA, Zbigniew, 2020. Évolution de la ségrégation pendant la journée et frictions spatiales : une analyse à partir de données de téléphonie mobile. In : *Documents de travail*. [en ligne]. 9 novembre 2020. Insee. N° G2020-12. [Consulté le 25 mars 2025]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/4925200>.
- GALIANA, Lino et SUAREZ CASTILLO, Milena, 2022. Fuzzy matching on big-data: an illustration with scanner and crowd-sourced nutritional datasets. In : *Proceedings of the 2022 ACM Conference on Information Technology for Social Good*. [en ligne] Pp. 331-337. [Consulté le 25 mars 2025]. Disponible à l'adresse : <https://doi.org/10.1145/3524458.3547244>.
- HARFORD, Tim, 2014. Big data: are we making a big mistake? In : *Financial Times*. [en ligne]. 28 mars 2014. [Consulté le 25 mars 2025]. Disponible à l'adresse : <https://www.ft.com/content/21a6e7d8-b479-11e3-a09a-00144feabdc0>.
- INSEE, 2019. L'économie et la société à l'ère du numérique. In : *Insee Références*. [en ligne]. 4 novembre 2019. pp 55-69. [Consulté le 25 mars 2025]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/4238635>.
- INSEE, 2020. Éclairage - Disparités territoriales de consommation : que disent les données de transaction par carte bancaire ? In : *Notes et points de conjoncture de l'année 2020*. [en ligne]. 15 décembre 2020. [Consulté le 29 janvier 2025]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/4488582?sommaire=4473296&q=cartes+bancaires>.

- LE SAOUT, Ronan, RIEDINGER, Nicolas et MESQUI, Bérengère, 2024. Les statistiques publiques de l'énergie – Enjeux passés, présents et futurs. In : *Courrier des statistiques*. [en ligne]. 8 juillet 2024. Insee. N° N11, pp. 51-71. [Consulté le 25 mars 2025]. Disponible à l'adresse : <https://www.insee.fr/fr/information/8203038>.
- LECLAIR, Marie, 2019. Utiliser les données de caisses pour le calcul de l'indice des prix à la consommation. In : *Courrier des statistiques*. [en ligne]. 19 décembre 2019. Insee. N° N3, pp. 61-75. [Consulté le 25 mars 2025]. Disponible à l'adresse : <https://www.insee.fr/fr/information/4254225?sommaire=4254170>.
- RICCIATO, Fabio, 2024. Steps Toward a Shared Infrastructure for Multi-Party Secure Private Computing in Official Statistics. In : *Journal of Official Statistics*. [en ligne]. 15 mars 2024. Volume 40, Issue 1. [Consulté le 25 mars 2025]. Disponible à l'adresse : <https://doi.org/10.1177/0282423X241235259>.
- SALA, Michel, 2023. Rapport d'information sur l'accès aux données privées : une nouvelle ressource pour l'Institut national de la statistique et des études économiques ? In : *Rapport d'information de l'Assemblée Nationale*. [en ligne]. N° 1312. 1^{er} juin 2023. [Consulté le 25 mars 2025]. Disponible à l'adresse : https://www.assemblee-nationale.fr/dyn/16/rapports/cion_fin/116b1312_rapport-information#_Toc256000048.
- SAKAROVITCH, Benjamin, 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 25 mars 2025]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/3706213?sommaire=3706255>.
- SUAREZ CASTILLO, Milena, SÉMÉCURBE, François, ZIEMLIKI, Cezary, TAO, Haixuan Xavier 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, 2023, pp. 535–570. [Consulté le 25 mars 2025]. Disponible à l'adresse : <https://doi.org/10.2478/jos-2023-0025>.
- TASSI, Philippe, 2019. Introduction – Les apports des Big Data. In : *Économie et Statistique / Economics and Statistics*. [en ligne]. 11 avril 2019. N° 505-506. [Consulté le 25 mars 2025]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/3705956?sommaire=3706255>.
- TAVERNIER, Jean-Luc et OURLIAC, Benoît, 2020. Google en sait-il plus que l'Insee sur les Français ? In : *Le blog de l'Insee*. [en ligne]. 18 décembre 2020. [Consulté le 19 mai 2025]. Disponible à l'adresse : <https://blog.insee.fr/google-en-sait-il-plus-que-linsee-sur-les-francais/>.
- TIGANI, Jordan, 2023. Big Data is Dead. In : *MotherDuck Blog*. [en ligne]. 7 février 2023. [Consulté le 25 mars 2025]. Disponible à l'adresse : <https://motherduck.com/blog/big-data-is-dead/>.
- ULRICH, Amandine, 2021. Hébergements proposés par des particuliers via des plateformes – En 2019, Paris et Nice dans le top 10 des villes les plus fréquentées de l'Union européenne. In : *Insee Première*. [en ligne]. 26 novembre 2021. Insee. N° 1879. [Consulté le 25 mars 2025]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/5871113>.

- VACHER, Thomas et PRADINES, Nadège, 2017. Cloud computing, big data : de nouvelles opportunités pour les sociétés. In : *Insee Première*. [en ligne]. 30 mars 2017. Insee. N° 1643. [Consulté le 25 mars 2025]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/2672067>.