

Les données de téléphonie mobile en tant que source pour les statistiques sur les déplacements entre le lieu de résidence et de travail en Belgique: deux applications statistiques

- Marc Debusschere, Pieter Dewitte,
Patrick Lusyne, Youri Baeyens -

n°10

ANALYSE

06.2020

Les données de téléphonie mobile en tant que source pour les statistiques sur les déplacements entre le lieu de résidence et de travail en Belgique: deux applications statistiques

Marc Debusschere, Pieter Dewitte, Patrick Lusyne, Youri Baeyens¹

¹ Statisticiens à Statbel (Direction générale Statistique - Statistics Belgium)

ABSTRACT

Depuis fin 2015, Statbel, l'office belge de statistique, a eu des contacts avec les trois opérateurs de réseau de téléphonie mobile actifs en Belgique afin d'utiliser les données de téléphonie mobile comme sources pour les statistiques publiques. Suite à cela, deux applications concrètes potentielles des données de téléphonie mobile pour la production de statistiques sur les déplacements entre le lieu de résidence et de travail ont pu être développées. Elles sont constituées d'une description détaillée des données nécessaires, des opérations à effectuer et des résultats statistiques attendus. Ces applications respectent bien entendu complètement la charte Statbel sur les big data et la vie privée².

La première application statistique possible vise à mesurer les profils de présence et d'absence sur le lieu de résidence et sur le lieu de travail pendant une année entière. La deuxième application complète cette approche statique. Elle consiste à relier de manière dynamique le lieu de résidence et le lieu de travail afin de créer une matrice « Lieu de résidence/Lieu de travail » beaucoup plus détaillée que l'actuelle qui est établie sur base de données administratives. Cette nouvelle approche serait également plus rapide que l'actuelle.

Enfin, la manière dont ces applications peuvent donner lieu à des statistiques expérimentales et, à terme, à de nouvelles productions statistiques sera examinée.

² Voir <https://statbel.fgov.be/fr/propos-de-statbel/vie-privee/statbel-big-data-et-respect-de-la-vie-privee> : Statbel, big data et vie privée.

TABLE DES MATIÈRES

<i>Les données de téléphonie mobile en tant que source pour les statistiques sur les déplacements entre le lieu de résidence et de travail en Belgique: deux applications statistiques</i>	1
<i>Abstract</i>	2
<i>Table des matières</i>	3
1. Les données de téléphonie mobile et les statistiques publiques	4
2. Les données de téléphonie mobile et Statbel	4
3. Applications statistiques	5
4. De meilleures statistiques démographiques selon le lieu de résidence et le lieu de travail	6
4.1. Contexte	6
4.2. Demande de données	7
5. Matrice « Lieu de résidence/Lieu de travail (origine et destination du déplacement) »	7
5.1. Contexte	7
5.2. Demande de données	8
5.2.1. Scénario 1	8
5.2.2. Scénario 2	8
6. Conclusion	9
7. Références	10

1. LES DONNEES DE TELEPHONIE MOBILE ET LES STATISTIQUES PUBLIQUES

Dès leur apparition au début du 19^e siècle, les statistiques publiques s'appuyaient sur des enquêtes menées auprès des citoyens et des entreprises. Afin de limiter les coûts et la charge des répondants, les fichiers administratifs ont été de plus en plus utilisés au cours des vingt dernières années. Entre-temps, la « troisième révolution de données » a commencé: l'exploitation de l'immense flot de « big data », en grande partie déstructuré, généré en continu dans notre société via les capteurs et caméras, les satellites, la communication « machine to machine », le commerce électronique, les paiements électroniques et les retraits, les activités en ligne en tout genre, les réseaux sociaux, etc. avec pour objectif de produire plus rapidement et plus efficacement la plupart des statistiques existantes, voire même de décrire des phénomènes inexplorés jusqu'à présent.

Le *Scheveningen Memorandum on Big Data and Official Statistics*³, réalisé par les DGINS⁴ et adopté par le CSSE⁵ le 27 septembre 2013, peut être considéré comme le début officiel de l'intégration des big data dans le Système Statistique Européen. Cela a été concrétisé par l'ESSnetBig Data I et II⁶, un vaste partenariat visant à étudier l'utilisation de divers types de big data, y compris les données de téléphonie mobile, et à les tester au moyen d'études pilotes.

Dans le contexte des big data, les données de téléphonie mobile constituent une source potentielle particulièrement prometteuse de statistiques, en particulier dans les domaines de la population, des migrations, du tourisme et de la mobilité. Leur utilisation devrait permettre d'obtenir plus rapidement, voire même immédiatement, des statistiques beaucoup plus détaillées, avec une couverture presque parfaite, sans biais de réponse, à moindre coût et sans devoir impliquer les citoyens et les entreprises. Par ailleurs, elles peuvent donner un accès direct à des phénomènes qui étaient jusqu'à présent impossibles à mesurer (tels que la population réelle présente par rapport à la population enregistrée ou des modèles de déplacements détaillés en fonction de la journée de travail, des conditions météorologiques,...). De nombreuses études pilotes en Europe et dans le monde ont démontré qu'elles constituaient une alternative potentielle aux sources de données plus traditionnelles utilisées par les instituts nationaux de statistique (De Meersman.e.a., 2016, bibliographie). Le *work package* du ESSnetBig Data sur les données de téléphonie mobile a amélioré la compréhension des données, de leurs possibilités et de leurs limites et a donné une impulsion décisive à la définition d'une méthodologie statistique pour les exploiter. Cependant, les progrès réels sont encore fortement entravés par l'accès extrêmement limité aux données de téléphonie mobile pour la production statistique et pour le test des différentes méthodes et approches.

2. LES DONNEES DE TELEPHONIE MOBILE ET STATBEL

Trois opérateurs de téléphonie mobile sont actifs en Belgique: Proximus (qui avait auparavant le monopole du marché), Telenet/Base et Orange Belgium, dont les parts de marché respectives en 2018 s'élevaient à environ 41 %, 30 % et 24,6 %⁷. Dès 2015, ces trois opérateurs ont été contactés par Statbel qui leur a proposé d'examiner conjointement leurs données et d'éventuellement les combiner à des données statistiques en vue d'une utilisation statistique par Statbel et d'une utilisation commerciale par les opérateurs.

Les discussions exploratoires avec Telenet/Base en 2015 et 2018 n'ont pas encore donné lieu à un projet concret, malgré l'intérêt manifesté à ces deux occasions.

Un premier contact avec Orange en 2016 est également resté sans suite immédiate, mais en 2018, une proposition plus concrète du MIT (Boston, USA) visant à combiner les données de signalisation du réseau de téléphonie mobile aux données de Statbel sur le revenu fiscal, afin d'étudier la ségrégation sociale à Bruxelles, a donné un motif de collaboration plus concret. Toutefois, en raison de mouvements au sein du personnel du MIT et d'Orange, ce projet a été mis en suspens et il n'y a aucune perspective de finalisation à court terme.

Un projet commun a débuté en décembre 2015 avec Proximus et Eurostat, l'Office statistique de l'Union européenne, afin d'analyser les données de signalisation agrégées du réseau Proximus pour leur contenu informatif et leurs possibilités

³ Voir <https://ec.europa.eu/eurostat/documents/42577/43315/Scheveningen-memorandum-27-09-13> (téléchargement PDF).

⁴ Conférence annuelle des directeurs généraux des instituts nationaux de statistique du Système statistique européen, voir <https://ec.europa.eu/eurostat/web/ess/about-us/ess-gov-bodies/dgins>.

⁵ Le Comité du système statistique européen. Voir <https://ec.europa.eu/eurostat/web/ess/about-us/ess-gov-bodies/essc>.

⁶ Voir <https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata/>.

⁷ Voir le site <https://www.internetproviders.be/overzicht-mobiel-internet-op-belgische-markt/>, consulté le 8 août 2019; le total n'est pas de 100 % en raison des parts de marché limitées d'opérateurs sans réseau propre.

d'utilisation éventuelles. Cela a donné lieu à 10 publications⁸ et à une conférence de presse conjointe en septembre 2016 présentant les résultats au grand public. En 2018, dans le contexte du projet MIT mentionné ci-dessus, Proximus a créé un fichier contenant les données de signalisation agrégées du réseau qui peuvent être reliées à un ensemble de données sur les revenus fiscaux, fournies de manière agrégée par Statbel, selon une méthode innovante et très détaillée. En effet, le lien n'étant pas effectué sur une base individuelle, tout problème de confidentialité ou de respect de la vie privée est alors exclu.

Le principal obstacle, non seulement en Belgique mais aussi dans les autres États membres de l'UE et même dans le monde entier, à l'utilisation de big data telles que les données de téléphonie mobile pour les statistiques publiques, est d'avoir accès aux données des entreprises privées. En effet, ces dernières voient les problèmes potentiels à les rendre accessibles mais pas d'avantages clairs. Les opérateurs de réseau mobile sont particulièrement réticents à accorder l'accès aux données, voire à les exploiter eux-mêmes, en raison des risques éventuels d'atteintes à la vie privée qui pourraient susciter l'indignation du grand public. En outre, les considérations de concurrence constituent un problème supplémentaire : malgré toutes les garanties offertes, les opérateurs sont très préoccupés par l'éventualité d'une fuite de données de signalisation stratégiques du réseau vers les concurrents. Cette réserve ne peut être levée que s'il peut être démontré que le risque est infime et que les applications commerciales ou statistiques peuvent créer suffisamment de valeur ajoutée.

Un document présenté en octobre 2018 lors de la conférence des DGINS à Bucarest (Debusschere, Waeyaert, Van Loon, 2018) cite quatre facteurs clés nécessaires à l'obtention de l'accès aux big data telles que les données de téléphone mobile par les statistiques publiques:

- 1) un projet clair et détaillé;
- 2) un engagement au haut niveau et un soutien actif;
- 3) l'instauration d'un climat de confiance en garantissant de manière absolue la confidentialité et le respect de la vie privée;
- 4) une législation spécifique.

La méthodologie proposée par Statbel pour le projet MIT est extrêmement pertinente pour le troisième point, les garanties strictes de sécurité des données, grâce à deux approches innovantes:

- ▶ Les données individuelles, tant les données statistiques que les données de téléphonie mobile, ne doivent pas quitter le data warehouse des propriétaires respectifs. Le respect de la vie privée et la confidentialité sont ainsi sous contrôle et les droits de propriétés sont assurés;
- ▶ Les données statistiques et de téléphonie mobile ne sont pas couplées individuellement, mais à un niveau géographique très détaillé, toutefois anonyme: les variables statistiques sont agrégées en fonction des zones couvertes par les antennes (via des fichiers de formes Voronoï fournis par l'opérateur de réseau mobile) afin qu'elles puissent être combinées de manière transparente avec les données agrégées de téléphonie mobile pour ces mêmes zones.

Il convient de noter que cette méthodologie convient également aux applications commerciales des opérateurs de réseau de téléphonie mobile, ce qui permet d'enrichir considérablement les données de téléphonie mobile avec des informations statistiques contextuelles appropriées. Cela pourrait inciter les opérateurs à coopérer avec les instituts nationaux de statistique pour leur avantage mutuel.

3. APPLICATIONS STATISTIQUES

L'article actuel met l'accent sur le premier des quatre facteurs critiques mentionnés dans le document des DGINS: le développement de « use cases » clairs et détaillés, plus particulièrement dans le domaine de la mobilité et des déplacements entre le lieu de résidence et de travail. Actuellement, Statbel calcule des statistiques sur les déplacements qui sont, soit limitées en ce qui concerne le niveau de détail géographique et/ou publiées à un intervalle de temps considérable. La matrice « Lieu de résidence/Lieu de travail » produite dans le contexte du Census décennal est basée sur des données administratives, au niveau du secteur statistique (ce qui correspond aux « localités » des communes), provenant du registre national et de la Banque Carrefour de la Sécurité Sociale. L'enquête trimestrielle sur les forces de travail (EFT) est une enquête par sondage

⁸ Voir [De Meersman e.a \(2016\)](#) pour un aperçu du projet.

qui contient également des questions sur les déplacements entre le lieu de résidence et de travail mais, en raison de la taille limitée de l'échantillon, aucun résultat n'est publié à un niveau plus détaillé que celui des trois régions.

Pendant les premières étapes de la recherche du potentiel des données de téléphonie mobile pour les statistiques publiques, d'innombrables études pilotes ont été réalisées en mettant l'accent sur les caractéristiques des données, leur validité statistique, leur fiabilité, leur qualité et les questions de méthodologie. Les résultats, du projet commun Statbel-Eurostat-Proximus, notamment, semblent prometteurs: il est fort probable que les données de téléphonie mobile puissent être utilisées pour remplacer, compléter, enrichir ou valider tout ou une partie des statistiques publiques dans des domaines tels que la population et la migration, la mobilité et les déplacements, les transports, le tourisme, etc. L'étape logique suivante consiste à développer des applications statistiques, en identifiant un produit statistique existant ou en développant un nouveau, en sélectionnant les données de téléphonie mobile nécessaires comme input et en spécifiant les opérations à effectuer sur ces données afin d'obtenir des statistiques publiques qui sont au moins partiellement basées sur les données de téléphonie mobile.

L'élément clé d'une application statistique est une demande concrète de données de téléphonie mobile qui soit suffisamment détaillée et réaliste en termes de complexité, de délai d'exécution de la requête, de taille de l'ensemble de données qui en résultent et dans la manière d'aborder les questions de confidentialité. Par ailleurs, afin d'être utilisable à plus long terme pour les statistiques publiques, l'application doit être durable, c'est-à-dire renouvelable à une fréquence convenue sans effort supplémentaire afin que la production régulière de statistiques devienne possible.

Les deux applications statistiques présentées ci-dessous décrivent en détail les données de téléphonie mobile, les agrégations et les calculs qui doivent servir de base pour établir les statistiques sur les déplacements entre le lieu de résidence et de travail. Elles peuvent naturellement remplir des objectifs secondaires, comme la détermination du mode de transport (par exemple en combinant ces résultats aux données sur l'utilisation du sol ou du registre des bâtiments), le calcul des temps de déplacement, l'estimation de l'impact environnemental, etc. Une autre utilisation possible est la « géographie sociale »: définir les zones résidentielles, de travail et de navette au sein d'un territoire, ou délimiter les « zones d'influence » des régions urbaines, ou l'expansion urbaine sous l'angle des déplacements.

4. DE MEILLEURES STATISTIQUES DEMOGRAPHIQUES SELON LE LIEU DE RESIDENCE ET LE LIEU DE TRAVAIL

4.1. Contexte

Les données de téléphonie mobile ont déjà été utilisées dans le cadre du projet Statbel-Eurostat-Proximus comme réplique et donc validation implicite des résultats du Censur dénombrant la population par lieu de résidence, basés sur le registre national belge (De Meersman e.a., 2016). Le résultat prometteur, et plus particulièrement la corrélation élevée de 0,85 entre les deux sources de données, permet d'identifier de manière plus précise les données de téléphonie mobile spécifiques qui peuvent rendre plus précises les statistiques démographiques relatives au lieu de résidence et au lieu de travail.

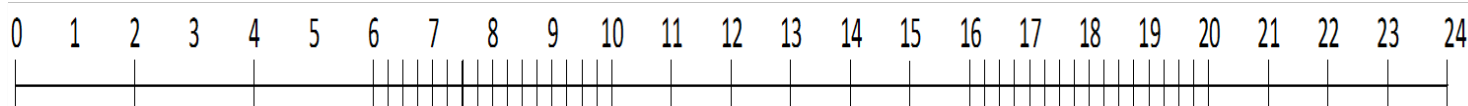
Le « domicile » inscrit au registre de la population est une bonne indication, bien qu'imparfaite, du lieu de résidence, lieu d'habitation réel. Les données de téléphonie mobile offrent une estimation alternative de cette dernière variable, bien que présentant des lacunes différentes. Par conséquent, les deux sources sont complémentaires, et leurs lacunes respectives peuvent être compensées par des informations issues de l'autre source. En combinant les deux et, selon le cas, en ajoutant d'autres ensembles de données (p.ex. CORINE Land cover⁹), le lieu de résidence peut probablement être mieux estimé et conduire à des résultats de recensement plus précis.

En ce qui concerne le lieu de travail, une approche similaire peut être développée, en combinant les données de la Banque Carrefour de la sécurité sociale (BCSS) et de la Banque Carrefour des Entreprises (BCE) avec les données de téléphonie mobile.

⁹ Voir <https://www.eea.europa.eu/publications/CORO-landcover> pour plus d'informations.

4.2. Demande de données

Le nombre d'appareils mobiles pour chaque cellule de Voronoï (N=11.000), extrapolé à partir de la part de marché de Proximus, mesuré à différents moments (N=45) sur une période de 24 heures, pour chaque jour de l'année civile 2016 (N=365) ou des 12 derniers mois disponibles, accompagné de la topologie (shapefile) des cellules Voronoï du premier jour, et dans la mesure où celles-ci évoluent (afin de prendre en compte tout changement survenant lors de leur conversion en d'autres cartes géographiques comme une grille km²).



- ▶ Cette période se compose des éléments suivants, avec des fréquences de mesure diverses:
- ▶ Entre 00:00 et 06:00 (période calme) toutes les 2 heures
- ▶ Entre 06:00 et 10:00 (heures de pointe du matin) toutes les 15 minutes
- ▶ Entre 10:00 et 16:00 (heures de bureau « normales ») toutes les heures
- ▶ Entre 16:00 et 20:00 (heures de pointe du soir) toutes les 15 minutes
- ▶ Entre 20:00 et 24:00 (repos du soir) toutes les heures

Outre l'analyse de la présence à un certain endroit sur la journée, l'utilisation de données sur une année complète permet également d'examiner les différences entre les jours ouvrables et les samedis, les dimanches, les autres jours ouvrables, les jours fériés, les périodes de vacances, les événements ponctuels (comme les attentats du 22 mars 2016, les grèves), les conditions climatiques, les influences saisonnières, etc.

L'ensemble de données contiendrait un peu plus de 180 millions d'enregistrements et la requête devrait pouvoir être exécutée annuellement (par exemple en janvier pour l'année civile précédente).

L'utilisation de 96 moments, c'est-à-dire toutes les 15 minutes, au lieu des 45 proposées doublerait la taille de l'ensemble de données, mais pourrait simplifier les paramètres de la requête.

5. MATRICE « LIEU DE RESIDENCE/LIEU DE TRAVAIL (ORIGINE ET DESTINATION DU DEPLACEMENT) »

5.1. Contexte

L'ensemble de données mentionnées ci-dessus fournit un bon aperçu global du lieu de résidence habituel, d'une part, et du lieu de travail de la population, d'autre part, mais pas de la relation entre les deux ni des déplacements de l'un à l'autre. Les totaux des cellules de Voronoï ne sont pas adéquats à cette fin, davantage de détails sont nécessaires. Une première possibilité consiste à tracer les appareils mobiles individuellement dans le temps et dans l'espace pour, ensuite, agréger les résultats. Il existe toutefois une autre solution qui éviterait tout problème lié à la protection de la vie privée: désigner le lieu de résidence et le lieu de travail les plus probables pour chaque appareil mobile à l'aide d'un algorithme qui vérifie où l'appareil est le plus souvent localisé à certains moments de la journée, et qui agrège ensuite ces données en un ensemble de données.

Cet ensemble de données de téléphonie mobile pourrait servir, dans le cadre du Censur, à valider la matrice « Lieu de résidence/Lieu de travail » développée actuellement par Statbel sur base des données administratives, mais aussi à améliorer sensiblement sa rapidité, son exactitude et son niveau de détail temporel et spatial.

La matrice « Lieu de résidence/Lieu de travail » est utilisée notamment par les partenaires sociaux (employeurs et syndicats) représentés au Conseil central de l'économie et par le Bureau fédéral du Plan pour leurs modèles économiques.

5.2. Demande de données

L'ensemble de données demandées devrait permettre de développer un algorithme pour déterminer le lieu de résidence et le lieu de travail. Selon les résultats, cette requête peut devoir être modulée pour d'autres combinaisons de moments distincts ou de périodes, sans toutefois affecter significativement le volume de données.

Concrètement, la demande porte sur les résultats d'un calcul effectué sur base des données du dernier mois d'octobre, qui est considéré comme le mois le plus « normal », « typique » sans jour férié et habituellement sans vacances scolaires. Deux algorithmes possibles sont présentés ci-dessous, avec un résultat similaire: un tableau croisé pour le nombre d'appareils mobiles d'environ 11.000 X 11.000 cellules de Voronoï. Cela correspond à un maximum théorique de 121.000.000 entrées, mais la plupart d'entre elles auront probablement une valeur nulle. Une entrée se compose d'un lieu de résidence, d'un lieu de travail, du nombre d'appareils mobiles et de la topologie (shapefile) des deux cellules de Voronoï.

5.2.1. Scénario 1

- ▶ Étape 1: déterminer, pour chaque téléphone mobile, la cellule de Voronoï qui peut être considérée comme le lieu de résidence le plus probable de la personne qui utilise l'appareil. Pour chaque appareil et pour chaque jour d'octobre, la cellule de Voronoï dans laquelle l'appareil est localisé à 04:00 est identifiée. Le lieu de résidence le plus probable est alors la cellule de Voronoï dans laquelle un appareil a été le plus souvent observé à 04:00 pendant le mois d'octobre 2016 (ou, en d'autres termes, le mode de répartition des cellules de Voronoï mesurées à 04:00 pour un appareil donné en octobre 2016).
- ▶ Étape 2: déterminer, pour chaque téléphone mobile, la cellule de Voronoï qui peut être considérée comme le lieu de travail le plus probable de la personne qui utilise l'appareil. Pour chaque appareil est déterminée la cellule de Voronoï dans laquelle il a été observé à 10h, 11h, 14h et 15h chaque jour ouvrable (du lundi au vendredi) d'octobre 2016. Le lieu de travail le plus probable est alors la cellule de Voronoï dans laquelle l'appareil se situe le plus fréquemment à ces moments (ou, en d'autres termes, le mode de répartition des cellules de Voronoï mesurées à 10h, 11h, 14h et 15h chaque jour ouvrable d'octobre 2016 pour un appareil donné).
- ▶ Étape 3: Après avoir défini le lieu de résidence et le lieu de travail les plus probables pour chaque appareil mobile, ceux-ci sont agrégés dans un tableau croisé lieu de résidence X lieu de travail d'environ 11.000 X 11.000 cellules, soit un total d'environ 121.000.000 valeurs (dont la grande majorité sera probablement nulle, étant donné qu'aucun appareil mobile « domicilié » dans la cellule x ne « travaille » dans la cellule y).

		Lieu de travail le plus probable en octobre 2016					
		Cellule 1	Cellule 2	Cellule 3	Cellule 3	Cellule 4	Cellule 5
Lieu de travail le plus probable en octobre 2016	Cellule 1						
	Cellule 2						
	Cellule 3						
	Cellule 4						
	Cellule 5						

5.2.2. Scénario 2

Dans le scénario 2, le lieu de résidence est déterminé de la même manière que dans le scénario 1, mais tandis que le lieu de travail le plus probable est toujours déterminé dans le scénario 1 pour un appareil mobile donné, le scénario 2 exclut les cas où l'appareil mobile est localisé trop rarement à un endroit.

Supposons que nous ayons n appareils mobiles et m cellules de Voronoï. Nous notons le nombre de fois (la *fréquence*) que le i -ème appareil se trouve dans la j -ème cellule de Voronoï (à 10h, 11h, 14h et 15h chaque jour de semaine d'octobre 2016)

comme f_{ij} . Nous notons l'indice de la cellule de Voronoï dans laquelle le i -ème appareil mobile est apparu le plus souvent comme Mod_i (le mode de répartition des cellules de Voronoï où le i -ème appareil mobile est apparu). Autrement dit,

$$f_{iMod_i} = \max_{j \in \{1, \dots, m\}} \{f_{ij}\}$$

Nous calculons la probabilité p que le i -ème appareil mobile soit situé dans la cellule de Voronoï Mod_i à un « moment de travail » aléatoire.

$$p_i = \frac{f_{iMod_i}}{\sum_{j=1}^m f_{ij}}$$

p_i donne une idée de la fréquence à laquelle une personne se trouve au même endroit pendant la journée de travail. Nous définissons le « lieu de travail le plus probable » comme la cellule de Voronoï Mod_i si p_i est plus grand ou égal à une certaine valeur minimum α . Si $p_i < \alpha$, le « lieu de travail le plus probable » est considéré comme non défini.

α peut être défini comme le décile inférieur de p . Ou, exprimé différemment:

$$\alpha = \inf \left\{ p \mid \frac{\#\{p_i \mid p_i \leq p\}}{n} \geq 0.1 \right\}$$

Dès lors que le « lieu de résidence le plus probable » et le « lieu de de travail le plus probable » ont été déterminés pour chaque appareil mobile, nous agrégeons le fichier. Nous voulons le nombre d'appareils mobiles par lieu de résidence le plus probable et lieu de travail le plus probable pour le mois d'octobre 2016. Nous obtenons ainsi un tableau croisé.

		Lieu de travail le plus probable en octobre 2016					
		Cellule 1	Cellule 2	Cellule 3	Cellule 3	Cellule 4	Cellule 5
Lieu de travail le plus probable en octobre 2016	Cellule 1						
	Cellule 2						
	Cellule 3						
	Cellule 4						
	Cellule 5						

6. CONCLUSION

Le but ultime de l'exploration du potentiel des données de téléphonie mobile et d'autres données, qui sont la propriété d'entreprises privées, comme source pour les statistiques publiques, est leur intégration dans des résultats statistiques réguliers calculés et publiés selon une méthodologie, une fréquence, une actualité, un niveau de détail et des normes de qualité définis au préalable. Le développement d'applications statistiques, use cases, est la deuxième étape logique de ce processus, après la phase exploratoire et avant le développement de statistiques expérimentales et, enfin, de statistiques publiques complètes.

Mais le développement de cas concrets est également crucial pour surmonter l'obstacle le plus important à la réalisation de cet objectif: le manque d'accès aux données. Les décideurs politiques et les autres utilisateurs potentiels, en particulier les opérateurs de réseau eux-mêmes, ne seront convaincus de cette nécessité que s'il peut leur être démontré que les coûts sont faibles, les risques parfaitement maîtrisables et les bénéfices potentiels importants. Les use cases sont une opérationnalisation concrète de ces résultats, car ils montrent quelles connaissances essentielles peuvent être obtenues à partir des données de téléphonie mobile.

Les deux applications présentées dans cet article donnent un aperçu concret et pratique de la manière dont il est possible d'obtenir des informations, très détaillées et à jour, sur les déplacements lieu de résidence-lieu de travail à partir de données de téléphonie mobile agrégées, sans aucune menace pour la vie privée ou la confidentialité des données et avec un coût marginal, une fois la requête initiale de la base de données établie. Des applications similaires peuvent être développées pour d'autres domaines présentant un intérêt élevé. Quelques exemples concrets:

- ▶ La population réellement présente, tant en nombre qu'en composition (répartition des habitants, touristes, travailleurs, visiteurs des hôpitaux ou écoles, vacanciers etc., en combinant avec des ensembles de données spatiales tels que le registre du bâtiment, l'utilisation du sol).
- ▶ La migration: migration de la main-d'œuvre et déplacements transfrontaliers via les données de roaming.
- ▶ Le tourisme: identification du nombre de visiteurs dans les localités et sites touristiques, événements avec l'origine, la durée du séjour, grâce aux données de localisation du réseau et des données de roaming.
- ▶ La mobilité et le transport de voyageurs: les flux de trafic avec leur origine et leur destination, le mode de transport (en combinant avec des ensembles de données spatiales tels que le registre des bâtiments ou de l'utilisation du sol, ou encore la mesure fine de l'intensité d'utilisation des routes, du chemin de fer (gares), des aéroports, etc.

Une étape suivante possible, qui devrait également présenter un grand intérêt pour les opérateurs de réseau, consisterait à combiner des données de téléphonie mobile agrégées avec un ensemble de données statistiques et spatio-temporelles agrégées de manière similaire, afin de créer des produits statistiques totalement nouveaux qui pourraient répondre à des questions statistiques ou commerciales qui n'étaient pas posées auparavant, puisqu'il était impossible d'y répondre. Par exemple : quel est l'impact des conditions météorologiques sur le trafic routier aux heures de pointe? d'où viennent les gens qui se rendent dans un lieu donné et quelles sont les caractéristiques démographiques, le revenu et le niveau d'instruction moyens de leur lieu d'origine?

7. RÉFÉRENCES

F. De Meersman, G. Seynaeve, M. Debusschere, P. Lusyne, P. Dewitte, Y. Baeyens, A. Wirthmann, C. Demunter, F. Reis, H.I. Reuter (2016): *Assessing the Quality of Mobile Phone Data as a Source of Statistics* ([mirror site](#)), Q2016 document de conférence, juin 2016 (téléchargement pdf)

M. Debusschere, N. Waeyaert, K. van Loon (2018): Key Factors for Obtaining Access to Big Data, document de conférence DGNIS , octobre 2018 (non publié, disponible sur demande auprès du premier auteur)

À PROPOS DE STATBEL

Statbel, l'office belge de statistique, collecte, produit et publie des chiffres fiables et pertinents sur l'économie, la société et le territoire belges.

Sur la base de sources de données administratives et d'enquêtes, Statbel produit des statistiques reposant sur des fondements scientifiques. Les résultats statistiques sont publiés de manière conviviale et sont disponibles pour tous en même temps.

Statbel utilise les données collectées uniquement à des fins statistiques. En tant qu'office de statistique, nous garantissons à tout moment la protection de la vie privée et des données confidentielles.

Visitez notre **site internet**

www.statbel.fgov.be

ou **contactez-nous**

e-mail: statbel@economie.fgov.be

Statbel (Direction générale Statistique - Statistics Belgium)
North Gate - Boulevard du Roi Albert II, 16, 1000 Bruxelles
E-mail: statbel@economie.fgov.be

Numéro d'entreprise
0314.595.348

Editeur responsable
Nicolas Waeyaert

North Gate
Boulevard du Roi Albert II, 16
1000 Bruxelles

