

Mobiele telefoondata als bron voor woon-werkstatistieken in België: Twee statistische toepassingen

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

nr.10

ANALYSE

06.2020

Mobiele telefoondata als bron voor woon-werkstatistieken in België: Twee statistische toepassingen

Marc Debusschere, Pieter Dewitte, Patrick Lusyne, Youri Baeyens¹

¹ Statistici bij Statbel (Algemene Directie Statistiek – Statistics Belgium)

ABSTRACT

Statbel, het Belgische statistiekbureau, heeft sinds eind 2015 contact gehad met de drie mobiele netwerkoperatoren actief in België in zijn pogingen om mobiele telefoondata als bron voor openbare statistieken te gebruiken. Een belangrijk resultaat daarvan is een reeks potentiële concrete toepassingen voor het opstellen van statistieken over woon-werkverkeer op basis van mobiele telefoondata. Deze toepassingen bestaan uit een gedetailleerde beschrijving van de vereiste data, de operaties die daarop uitgevoerd moeten worden en de te verwachten statistische resultaten. Die toepassingen respecteren uiteraard volledig het Statbel charter inzake big data en privacy².

De eerste statistische toepassing heeft als doel het meten van de patronen van aanwezigheid en afwezigheid op de woonplaats en de werkplaats gedurende een heel jaar. De tweede toepassing is complementair met deze statische benadering; ze wil de woonplaats en de werkplaats dynamisch koppelen om zo een « Matrix Woonplaats/Werkplaats » te kunnen opstellen die aanzienlijk gedetailleerder en tijdiger is dan de huidige op basis van administratieve data.

Tenslotte, en als conclusie, wordt besproken op welke manier deze toepassingen kunnen resulteren in experimentele statistieken en uiteindelijk in nieuwe statistische productielijnen.

² Zie <https://statbel.fgov.be/nl/over-statbel/privacy/statbel-big-data-en-privacy>: Statbel, big data en privacy

INHOUD

<i>Mobiele telefoondata als bron voor woon-werkstatistieken in België: Twee statistische toepassingen</i>	1
<i>Abstract</i>	2
<i>Inhoud</i>	3
1. <i>Mobiele telefoondata en openbare statistieken</i>	4
2. <i>Mobiele telefoondata en Statbel</i>	4
3. <i>Statistische toepassingen</i>	5
4. <i>Betere bevolkingsstatistieken volgens woonplaats en werkplaats</i>	6
4.1. <i>Achtergrond</i>	6
4.2. <i>Datavraag</i>	6
5. <i>Matrix Woonplaats/Werkplaats (herkomst en bestemming woon-werkverkeer)</i>	7
5.1. <i>Achtergrond</i>	7
5.2. <i>Datavraag</i>	7
5.2.1. <i>Scenario 1</i>	8
5.2.2. <i>Scenario 2</i>	8
6. <i>Besluit</i>	9
7. <i>Referenties</i>	10

1. MOBIELE TELEFOONDATA EN OPENBARE STATISTIEKEN

Openbare statistieken waren sinds hun begin in de vroege negentiende eeuw gebaseerd op enquêtes bij burgers en ondernemingen. Om de kosten en de belasting voor wie moet antwoorden te beperken, werd de laatste twintig jaar in toenemende mate een beroep gedaan op administratieve bestanden. Intussen is de ‘derde datarevolutie’ aangebroken: het exploiteren van de immense en grotendeels ongestructureerde vloedgolf van ‘big data’ die onze samenleving continu genereert via sensoren en camera’s, satellieten, *machine-to-machine* communicatie, e-business, elektronische betalingen en afhalingen, allerlei activiteit op het internet, sociale media, ... om de meeste bestaande statistieken sneller en beter te produceren en zelfs om fenomenen te beschrijven die tot nu toe volledig buiten beeld bleven.

Het *Scheveningen Memorandum on Big Data and Official Statistics*³, opgesteld door DGINS⁴ en aangenomen door het ESSC⁵ op 27 september 2013, kan beschouwd worden als de officiële start voor de integratie van big data binnen het Europees Statistisch Systeem. Dit werd concreet gemaakt door ESSnet Big Data I en II⁶, een uitgebreid samenwerkingsverband om het gebruik van diverse types van big data, waaronder mobiele telefoondata, te onderzoeken en via pilotstudies uit te testen.

Binnen het geheel van big data vormen mobiele telefoondata een bijzonder veelbelovende potentiële bron voor statistieken, vooral in de domeinen bevolking, migratie, toerisme en mobiliteit. Hun gebruik zou kunnen leiden tot snellere en zelfs onmiddellijke statistieken, in veel meer detail, met bijna volmaakte dekking, zonder antwoordbias, tegen een lagere kost en zonder noodzaak om burgers en ondernemingen lastig te vallen. Bovendien geven ze wellicht directe toegang tot fenomenen die tot nu onmeetbaar waren (zoals reëel aanwezige versus geregistreerde bevolking of gedetailleerde pendelpatronen naargelang de werkdag, weersomstandigheden, ...). Talloze pilotstudies in Europa en wereldwijd hebben aangetoond dat ze een mogelijk alternatief vormen voor de meer traditionele gegevensbronnen die de nationale statistische instituten gebruiken (De Meersman e.a., 2016, literatuurlijst). Het ESSnet Big Data *workpackage* rond mobiele telefoondata heeft het inzicht in de data en hun mogelijkheden en beperkingen ingrijpend verhoogd en een beslissende aanzet gegeven tot een statistische methodologie om ze te exploiteren, maar echte vooruitgang wordt nog steeds sterk gehinderd door de uiterst beperkte toegang tot mobiele telefoondata voor statistische productie en zelfs voor het testen van methodes en benaderingen.

2. MOBIELE TELEFOONDATA EN STATBEL

België heeft drie mobiele netwerkoperatoren: Proximus (de vroegere monopolist), Telenet/Base en Orange België, met respectieve marktaandeelen in 2018 van ongeveer 41%, 30% en 24.6%⁷. Allen werden ze vanaf 2015 gecontacteerd door Statbel met voorstellen om hun data gezamenlijk te exploreren en deze eventueel te combineren met statistische data, met het oog op statistische toepassing door Statbel en commerciële toepassing door de operatoren.

Verkennde gesprekken met Telenet/Base in 2015 en 2018 hebben nog niet tot een concreet project geleid ondanks de bij beide gelegenheden uitgedrukte interesse.

Een eerste contact met Orange in 2016 bleef ook zonder onmiddellijk gevolg, maar in 2018 bood een voorstel door MIT (Boston, US) om mobiel-netwerk-signaaldata te combineren met data over fiscale inkomens van Statbel voor een studie over sociale segregatie in Brussel een meer concrete aanleiding tot samenwerken. Door personeelwissels binnen MIT en Orange werd dit echter in wachtstand geplaatst en er zijn geen vooruitzichten op finalisering binnen een korte termijn.

Met Proximus en Eurostat, het Europees statistiekbureau, werd in december 2015 een gemeenschappelijk project opgestart om geaggregeerde netwerksignaaldata van Proximus te analyseren op hun informatie-inhoud en potentiële gebruiksmogelijkheden. Dit resulteerde in 10 publicaties⁸ en een gezamenlijke persconferentie in september 2016 om de resultaten publiek te presenteren. In de context van het hierboven vermelde MIT-project in 2018 heeft Proximus een bestand gecreëerd met geaggregeerde netwerksignaaldata die gelinkt kunnen worden aan een op maat gemaakte en eveneens

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

⁴ Jaarlijkse conferentie van de directeurs-generaal van de nationale statistische instituten van het Europees Statistisch Systeem, zie <https://ec.europa.eu/eurostat/web/ess/about-us/ess-gov-bodies/dgins>.

⁵ ESSC staat voor European Statistical System Committee. Zie <https://ec.europa.eu/eurostat/web/ess/about-us/ess-gov-bodies/essc>.

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

⁷ Zie <https://www.internetproviders.be/overzicht-mobiel-internet-op-belgische-markt/>, geconsulteerd op 8 augustus 2019; het totaal is niet 100% omwille van beperkte marktaandeelen van operatoren zonder eigen netwerk.

⁸ Zie [De Meersman e.a \(2016\)](#) voor een overzicht van het project.

geaggregeerde dataset van Statbel over fiscale inkomens, volgens een innovatieve en zeer gedetailleerde methode waarbij het linken evenwel niet op individuele basis gebeurde, zodat elk probleem met privacy of vertrouwelijkheid uitgesloten kon worden.

De grootste hinderpaal bij het gebruik van big data zoals mobiele telefoondata voor openbare statistieken, niet alleen in België maar ook in de andere EU-lidstaten en zelfs wereldwijd, is toegang krijgen tot data van privébedrijven als die wel mogelijke bezwaren maar geen duidelijk voordeel zien om deze toegankelijk te maken. Mobiele-netwerkoperatoren staan bijzonder weigerachtig tegenover het verlenen van toegang en zelfs tegenover het zelf exploiteren van de data omwille van de reële of gepercipieerde risico's op inbreuken op de privacy die zou kunnen leiden tot publieke verontwaardiging. Verder vormen concurrentieoverwegingen een bijkomend probleem: operatoren zijn ondanks alle geboden garanties zeer beducht voor het mogelijk lekken van strategische netwerksignaaldata naar de concurrenten. Deze terughoudendheid kan alleen gecounterd worden als aangetoond kan worden dat het risico infinitesimaal klein is en als commerciële of statistische toepassingen voldoende waarde kunnen creëren.

Een paper die gepresenteerd werd in oktober 2018 op DGINS-Conferentie in Boekarest (Debusschere, Waeyaert, Van Loon, 2018) stipt vier sleutelfactoren aan die vereist zijn voor toegang tot big data zoals mobiele telefoondata door openbare statistieken:

- 1) een duidelijke en gedetailleerde business case;
- 2) engagement en actieve ondersteuning op hoog niveau;
- 3) het bevorderen van vertrouwen door absolute waarborgen op vertrouwelijkheid en privacy;
- 4) specifieke wetgeving.

De door Statbel voorgestelde methodologie voor het MIT-project is uitermate relevant voor het derde punt, de strikte garanties voor databeveiliging, dankzij twee innovatieve benaderingen:

- ▶ individuele data, zowel mobiele-telefoon- als statistische data, hoeven het datawarehouse van de respectievelijke eigenaars niet te verlaten, waardoor privacy en vertrouwelijkheid onder controle zijn en eigendomsrechten verzekerd worden;
- ▶ mobiele-telefoon- en statistische data worden niet individueel gekoppeld, maar wel op een zeer gedetailleerd doch anoniem geografisch niveau: statistische variabelen worden geaggregeerd volgens de dekkingsgebieden van de antennes (via Voronoi 'shapefiles' geleverd door de mobiele netwerkoperator) zodat ze naadloos te combineren vallen met de geaggregeerde mobiele telefoondata voor dezelfde gebieden.

Er moet op gewezen worden dat deze methodologie ook geschikt is voor de commerciële toepassingen van de mobiele netwerkoperatoren, waardoor de mobiele telefoondata significant verrijkt kunnen worden met aangepaste statistische contextinformatie. Dit zou een motief kunnen zijn voor operatoren om tot wederzijds voordeel samen te werken met nationale statistische instituten.

3. STATISTISCHE TOEPASSINGEN

Het huidig artikel legt de focus op de eerste van de vier kritische factoren die in de DGINS-paper vermeld werden, op het ontwikkelen van duidelijke en gedetailleerde use cases, meer bepaald op het domein van mobiliteit en woon-werkverkeer. Op dit moment berekent Statbel woon-werkstatistieken die ofwel beperkt zijn in geografisch detail ofwel uitgebracht worden na een aanzienlijk tijdsinterval, ofwel allebei. De 'Matrix Woonplaats/Werkplaats' die geproduceerd wordt binnen de context van de tienjaarlijkse Census is gebaseerd op administratieve data, op het niveau van de statistische sector (beantwoordend aan 'wijken' van gemeenten), afkomstig van de Belgische personen- en sociale-zekerheidsregisters. De driemaandelijke Enquête naar de Arbeidskrachten (EAK) is een steekproefenquête die ook vragen bevat over woon-werkverkeer, maar omwille van beperkingen in de steekproefomvang worden geen resultaten gepubliceerd met meer geografisch detail dan het niveau van de drie gewesten.

Tijdens de eerste stadia van het onderzoek van het potentieel van mobiele telefoondata voor openbare statistieken werden talloze pilotstudies uitgevoerd met de focus op de karakteristieken van de data, hun statistische validiteit en betrouwbaarheid, de kwaliteit en methodologische kwesties. De resultaten, onder andere van het gezamenlijk Statbel-

Eurostat-Proximus-project, zien er veelbelovend uit: de waarschijnlijkheid is hoog dat mobiele telefoondata kunnen gebruikt worden om openbare statistieken geheel of gedeeltelijk te vervangen, aan te vullen, te verrijken of valideren, in domeinen zoals bevolking en migratie, mobiliteit en woon-werkverkeer, transport, toerisme ... De logisch volgende stap is dan de ontwikkeling van statistische toepassingen, door het identificeren van een bestaand of het uitwerken van nieuw statistisch product, het selecteren van de mobiele telefoondata die nodig zijn als input, en het specificeren van de operaties die erop toegepast moeten worden om uit te komen bij openbare statistieken die minstens deels gebaseerd zijn op mobiele telefoondata.

Het kernelement van een statistische toepassing is een concrete aanvraag van mobiele telefoondata die voldoende gedetailleerd is en realistisch in termen van de complexiteit en doorlooptijd van de query, de omvang van de resulterende dataset en het omgaan met privacykwesties. Bovendien moet de toepassing, om op langere termijn bruikbaar te zijn voor openbare statistieken, duurzaam zijn, namelijk herhaalbaar volgens een afgesproken frequentie zonder bijkomende inspanning zodat een reguliere statistiekproductie mogelijk wordt.

De twee statistische toepassingen die hieronder voorgesteld worden omschrijven in detail welke mobiele telefoondata, aggregaties en berekeningen vereist zijn als basis voor woon-werkverkeerststatistieken. Die kunnen dan uiteraard secundaire doelstellingen dienen, zoals het bepalen van de transportmodus (door bijvoorbeeld deze resultaten te combineren met data over bodemgebruik of het gebouwenregister), het berekenen van reistijden, het inschatten van de milieu-impact, ... Een verder mogelijk gebruik is 'sociale geografie': het bepalen van de overwegend residentiële, werk- of pendelzones binnen een territorium, of het afbakenen van de 'invloedszones' van stadsgebieden, of de stadsexpansie vanuit het perspectief van woon-werkverkeer.

4. BETERE BEVOLKINGSSTATISTIEKEN VOLGENS WOONPLAATS EN WERKPLAATS

4.1. Achtergrond

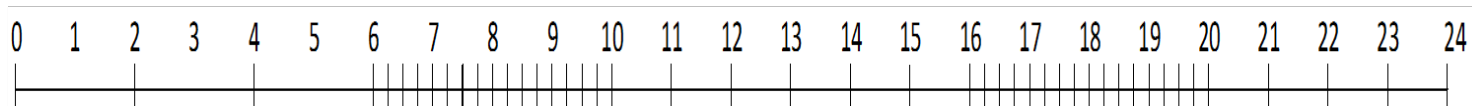
Mobiele telefoondata werden al gebruikt in het Statbel-Eurostat-Proximus-project als replicatie en dus impliciete validatie van Censusresultaten over de bevolking naar woonplaats, gebaseerd op het Belgisch Rijksregister ([De Meersman e.a., 2016](#)). Het veelbelovend resultaat, en met name de hoge correlatie van 0,85 tussen beide databronnen, maakt het mogelijk te preciseren met welke mobiele telefoondata de bevolkingsstatistieken volgens woonplaats en werkplaats accurater gemaakt kunnen worden.

Het Rijksregistergegevens 'domicilie' is een goede maar niet perfecte benadering van de woonplaats, de reële verblijfplaats. Mobiele telefoondata bieden een alternatieve schatting van deze laatste variabele; ook met tekortkomingen, maar dan wel andere. Beide bronnen zijn dus complementair en hiaten in de ene dataset kunnen wellicht met informatie vanuit de andere aangevuld worden. Door beide te combineren, eventueel aangevuld met nog meer datasets (bv. CORINE Land cover⁹), kan de woonplaats waarschijnlijk beter geschat worden en leiden tot Census resultaten die accurater zijn.

Voor de werkplaats kan een gelijkaardige benadering uitgewerkt worden, via het combineren van gegevens afkomstig van de Kruispuntbank Sociale Zekerheid (KSZ) en de Kruispuntbank van Ondernemingen (KBO) met mobiele telefoondata.

4.2. Datavraag

Het aantal mobiele toestellen voor elke Voronoigebied (N=11.000), geëxtrapoleerd volgens het Proximus-marktaandeel naar het totaal, gemeten op de tijdstippen (N=45) op een tijdlijn van 24 uren, voor elke dag van het voorbije kalenderjaar 2016 (N=365) of van de laatste 12 beschikbare maanden; met de topologie (shapefile) van de Voronoigebieden op de eerste dag, en verder voor zover deze wijzigingen ondergaat (om eventuele wijzigingen in rekening te brengen en om te zetten naar andere geografische indelingen zoals het rooster in km²).



⁹ Zie <https://www.eea.europa.eu/publications/COR0-landcover> voor meer informatie.

Deze tijdlijn bevat de volgende onderdelen, met wisselende frequenties van de metingen:

- ▶ Tussen 00:00 en 06:00 (rustige periode) om de 2 uur
- ▶ Tussen 06:00 en 10:00 (ochtendspits) om het kwartier
- ▶ Tussen 10:00 en 16:00 ('normale' kantooruren) om het uur
- ▶ Tussen 16:00 en 20:00 (avondspits) om het kwartier
- ▶ Tussen 20:00 en 24:00 (avondrust) om het uur

Naast de analyse van de verdeling over de dag van de aanwezigheid op een bepaalde plaats, kan door de beschikbaarheid van een volledig jaar ook in detail nagegaan worden wat het differentieel effect is van werkdag versus zaterdag versus zondag, de werkdagen onderling, feestdagen, vakantieperiodes, eenmalige gebeurtenissen en incidenten (zoals de aanslagen van 22 maart 2016, stakingen), weersomstandigheden, de invloed van seizoenen, ...

De dataset zou iets meer dan 180 miljoen records bevatten en de query zou normaal jaarlijks moeten kunnen lopen (bijvoorbeeld in januari voor het voorbije kalenderjaar).

Door 96 tijdpunten te gebruiken, dus om de 15 minuten, in plaats van de voorgestelde 45 zou de omvang van de dataset verdubbelen, maar worden de parameters van de query wellicht eenvoudiger.

5. MATRIX WOONPLAATS/WERKPLAATS (HERKOMST EN BESTEMMING WOONWERKVERKEER)

5.1. Achtergrond

Met de hierboven gevraagde dataset krijgt men een goed globaal beeld van enerzijds de gewone verblijfplaats en anderzijds de werkplaats van de bevolking, maar niet over het verband tussen beide en over de verplaatsingen tussen de woon- en werkplaats; totalen per Voronoigebied zijn daarvoor niet geschikt, meer detail is nodig. Een eerste mogelijkheid is om mobiele telefoons individueel te traceren over tijdstippen en locaties, en deze data dan te aggregeren. Er is echter een alternatief dat elk mogelijk privacyprobleem vermijdt: voor elke mobiel toestel de meest waarschijnlijke woon- en werkplaats bepalen via een algoritme dat kijkt waar het toestel zich het vaakst bevindt tijdens bepaalde periodes van de dag, en die gegevens daarna aggregeren tot een dataset.

Deze mobiele telefoondataset zou dan gebruikt kunnen worden om in het kader van de Census de « Matrix Woonplaats/Werkplaats » te valideren die Statbel nu opstelt op basis van administratieve data, maar ook om de snelheid, accuraatheid en mate van detail in tijd en ruimte significant te verhogen.

De Matrix Woonplaats/Werkplaats wordt onder andere gebruikt door de sociale partners (werknemers en vakbonden) in de Centrale Raad voor het Bedrijfsleven en door het Federaal Planbureau voor hun economische modellen.

5.2. Datavraag

De gevraagde dataset moet het mogelijk maken om een algoritme voor het bepalen van woonplaats en werkplaats te ontwikkelen. Afhankelijk van de resultaten zal die query misschien gemoduleerd moeten worden naar andere combinaties van tijdstippen of periodes, zonder dat dit echter het datavolume significant zal wijzigen.

Concreet wordt gevraagd naar de resultaten van een berekening op gegevens van de meest recente maand oktober; oktober wordt beschouwd als de meest 'normale' 'typische' werkmaand, zonder feestdagen en doorgaans ook zonder schoolvakanties. Hieronder worden twee mogelijke algoritmen voorgesteld, met dezelfde output: een kruistabel voor het aantal mobiele toestellen van ongeveer 11.000 X 11.000 Voronoigebieden. Dat correspondeert met een theoretisch maximum van 121.000.000 records, maar de meeste daarvan hebben waarschijnlijk waarde 0. Een record bestaat uit een woonplaats, werkplaats, aantal mobiele toestellen en topologie (shapefile) van beide Voronoigebieden.

5.2.1. Scenario 1

- ▶ Stap 1: bepalen, voor elk mobiel toestel, van het Voronoigebied dat als de meest waarschijnlijke woonplaats kan beschouwd worden van de persoon die het toestel gebruikt. Voor elk toestel wordt voor elke dag van oktober geregistreerd in welk Voronoigebied het zich bevindt om 04:00. De meest waarschijnlijke woonplaats is dan dit Voronoigebied waar het toestel zich het vaakst om 04:00 bevond gedurende oktober 2016 (of met andere woorden, de modus van de distributieverdeling van Voronoicellen gemeten om 04:00 voor een bepaald toestel in oktober 2016).
- ▶ Stap 2: bepalen, voor elk mobiel toestel, van het Voronoigebied die als de meest waarschijnlijke werkplaats kan beschouwd worden van de persoon die het toestel gebruikt. Voor elk toestel wordt geregistreerd in welk Voronoigebied het zich bevindt om 10u, 11u, 14u en 15u op elke weekdag (maandag t.e.m. vrijdag) van oktober 2016. De meest waarschijnlijke werkplaats is dan het Voronoigebied waar het toestel zich het vaakst bevond op deze verzameling van tijdstippen (of met andere woorden, de modus van de distributieverdeling van Voronoigebieden gemeten om 10u, 11u, 14u en 15u van elke weekdag voor een bepaald toestel in oktober 2016).
- ▶ Stap 3: Nadat voor elk mobiel toestel de meest waarschijnlijke woonplaats en werkplaats bepaald zijn, worden deze geaggregeerd in een kruistabel woonplaats x werkplaats van ongeveer 11.000 X 11.000 cellen, of in totaal zo'n 121.000.000 waarden (waarvan een grote meerderheid wellicht 0 is, als geen enkel toestel dat in Voronoigebied x 'woont' in Voronoigebied y 'werkt').

		Meest waarschijnlijke werkplaats oktober 2016					
		Cel 1	Cel 2	Cel 3	Cel 3	Cel 4	Cel 5
Meest waarschijnlijke woonplaats oktober 2016	Cel 1						
	Cel 2						
	Cel 3						
	Cel 4						
	Cel 5						

5.2.2. Scenario 2

In scenario 2 wordt de woonplaats op dezelfde manier bepaald als in scenario 1, maar waar de meest waarschijnlijke werkplaats voor een bepaald mobiel toestel in scenario 1 altijd wordt bepaald, worden in scenario 2 die gevallen uitgesloten waar het mobiel toestel zich te weinig frequent op één plek bevindt.

Stel dat we n mobiele toestellen hebben en m Voronoigebieden. Laten we het aantal keren (de *frequentie*) dat het i -de toestel zich in de j -het Voronoigebied bevindt (gemeten om 10u, 11u, 14u en 15u op elke weekdag van oktober 2016) noteren als f_{ij} . Laten we het indexnummer van het Voronoigebied waar de i -de mobile unit het vaakst voorkomt noteren als Mod_i (de modus van de Voronoigebieden waar de i -de mobile unit zich bevindt). Met andere woorden,

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

We berekenen de probabilmiteit p dat het i -de mobiel toestel zich op een willekeurig 'werkmoment' bevindt in Voronoigebied Mod_i

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

p_i geeft een idee hoe vaak iemand tijdens de werkdag zich op dezelfde plaats bevindt. We definiëren de ‘meest waarschijnlijke werkplaats’ als de Mod_i -de Voronoigebied indien p_i groter of gelijk is dan een bepaalde minimumwaarde α . Indien $p_i < \alpha$ dan wordt de “meest waarschijnlijke werkplaats” beschouwd als niet gedefinieerd.

α zou gedefinieerd kunnen worden als het laagste deciel van p . Met andere woorden

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

Enmaal voor iedere mobile unit de “meest waarschijnlijke verblijfplaats” en “meest waarschijnlijke werkplaats” zijn bepaald, aggregeren we het bestand. We wensen het aantal mobile units volgens “meest waarschijnlijke verblijfplaats” en “meest waarschijnlijke werkplaats” voor oktober 2016. Zo krijgen we een kruistabel.

		Meest waarschijnlijke werkplaats oktober 2016					
		Cel 1	Cel 2	Cel 3	Cel 3	Cel 4	Cel 5
Meest waarschijnlijke woonplaats oktober 2016	Cel 1						
	Cel 2						
	Cel 3						
	Cel 4						
	Cel 5						

6. BESLUIT

Het ultieme doel van het exploreren van het potentieel van mobiele telefoondata en andere data die het eigendom zijn van privébedrijven als bron voor openbare statistieken, is hun integratie in reguliere statistische resultaten die berekend en gepubliceerd worden volgens een vooraf vastgelegde methodologie, frequentie, tijdigheid, mate van detail en kwaliteitsstandaarden. Het uitwerken van statistische toepassingen, use cases, is de logische tweede stap in dit proces, na de exploratieve fase en voorafgaand aan het ontwikkelen van experimentele statistieken en, uiteindelijk, volwaardige openbare statistieken.

Maar het uitwerken van use cases is ook cruciaal bij het overwinnen van het belangrijkste obstakel om dit doel te bereiken: het ontbreken van toegang tot de data. Beleidsmakers en andere potentiële gebruikers, niet in het minst de netwerkoperatoren zelf, zullen er alleen van overtuigd kunnen worden dat dit essentieel is als aangetoond kan worden dat de kosten laag zijn, de risico's perfect beheersbaar, en de potentiële opbrengsten substantieel. Use cases zijn een concrete operationalisering van deze opbrengsten, doordat ze aantonen welke essentiële kennis uit mobiele telefoondata gehaald kan worden.

De twee toepassingen die in dit artikel gepresenteerd worden schetsen concreet en praktisch hoe zeer gedetailleerde en actuele informatie over woon-werkverkeer gehaald kan worden uit geaggregeerde mobiele telefoondata, zonder enige bedreiging van de privacy of datavertrouwelijkheid, tegen een marginale kost zodra de initiële database query werd opgesteld. Gelijkaardige toepassingen kunnen ontwikkeld worden voor andere domeinen met een hoog belang, enkele concrete voorbeelden:

- ▶ werkelijk aanwezige bevolking, zowel naar aantallen als samenstelling (opsplitsing bewoners, toeristen, werkenden, bezoekers ziekenhuis of school, recreanten via combinatie met spatiale datasets zoals gebouwenregister, bodemgebruik)
- ▶ migratie: arbeidsmigratie en grenspendel via roamingdata.

- ▶ toerisme: identificatie van bezoekersaantallen van (delen van) gemeenten en toeristische locaties, evenementen met origine, duur van verblijf, op basis van lokalisatienetwerkdata en roamingdata.
- ▶ mobiliteit en personenvervoer: verkeersstromen met oorsprong en bestemming, transportmodus door combinatie met spatiale datasets zoals gebouwenregister of bodemgebruik, fijnmazige meting van intensiteit van gebruik van verkeersknooppunten weg, spoor (stations), luchthavens etc.

Een mogelijke verdere stap, die ook voor de netwerkoperatoren interessant zou kunnen zijn, is het combineren van geaggregeerde mobiele telefoondata met op dezelfde wijze geaggregeerde statistische en spatio-temporele datasets, resulterend in totaal nieuwe statistische producten die een antwoord bieden op statistische of commerciële vragen die voorheen niet gesteld werden omdat ze toch niet beantwoord konden worden, zoals: wat is de impact van weersomstandigheden op het autoverkeer tijdens de spitsuren? Vanwaar komen de mensen op een bepaalde locatie en wat zijn de gemiddelde demografische kenmerken, inkomen en opleidingsniveau van die plaatsen van herkomst?

7. REFERENTIES

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 Conference paper, Juni 2016 (pdf download)

M. Debusschere, N. Waeyaert, K. van Loon (2018): *Key Factors for Obtaining Access to Big Data*, DGINS Conference paper, oktober 2018 (geen publicatie, beschikbaar op verzoek bij eerste auteur)

OVER STATBEL

Statbel, het Belgische statistiekbureau, verzamelt, produceert en publiceert objectieve en pertinente cijfers over de Belgische economie, de samenleving en het territorium.

Op basis van administratieve databronnen en enquêtes produceert Statbel wetenschappelijk onderbouwde statistieken. De statistische resultaten worden op een gebruiksvriendelijke manier gepubliceerd, en zijn voor iedereen gelijktijdig beschikbaar.

De ingezamelde gegevens worden door Statbel louter voor statistische doelen gebruikt. Als statistiekbureau garanderen we ten allen tijde de privacy en de bescherming van vertrouwelijke gegevens.

Bezoek onze **website**

www.statbel.fgov.be

of contacteer ons

e-mail: statbel@economie.fgov.be

Statbel (Algemene Directie Statistiek - Statistics Belgium)
North Gate - Koning Albert II-laan 16, 1000 Brussel
E-mail: statbel@economie.fgov.be

Ondernemingsnummer
0314.595.348

Verantwoordelijke uitgever
Nico Waeyaert

North Gate
Koning Albert II-laan 16
1000 Brussel

