

Big data en Statistics Belgium

NPSO, Antwerpen, 2 mei 2017

Marc Debusschere
Coordinator Administrative & Big Data

Statistics Belgium

.be

Overzicht

- **Context**
Big data en de derde datarevolutie in de statistiek
- **Big data en Statistics Belgium**
 - ❖ Overzicht
 - ❖ Project mobiele-telefoondata
Statistics Belgium, Proximus, Eurostat, JRC
 - ❖ Van exploratie naar exploitatie
- **Europese samenwerking**

Big data

- = data niet op 'normale' manier te verwerken
'normaal' is relatief ...
- 3 v's: volume, velocity, variety
- Gevolg van maatschappelijke en technologische evoluties
Satellieten, camera's en sensoren, internet en e-mail, sociale media, mobiele telefoons en tablets, e-business, e-government, machine naar machine (internet van dingen)
- Resultaat: data-explosie, *data deluge*

Big data en statistiek

- Big data = ‘digitale voetafdruk’
- Bevat waardevolle informatie, statistisch bruikbaar (maar ook commercieel ...)
- Gevolg: de derde datarevolutie in statistiek
Na enquêtes (>1846) en administratieve data (>2000), nu: big data!
- Mogelijke databronnen
 - ❖ Scannerdata, elektronische betalingen, kredietkaartdata
 - ❖ Webscraping voor vacatures, bedrijfsgegevens
 - ❖ Verkeerscamera’s en detectielussen
 - ❖ Slimme meters (elektriciteit, gas, water)
 - ❖ Last but not least: mobiele-telefoondata!

De toekomst van de statistiek big data!

Instantstatistieken gebaseerd op big data, aangevuld en/of gevalideerd door administratieve data en kleine en specifieke ad-hoc enquêtes.

Big data en Statistics Belgium

- In productie: voor consumptieprijsindex (CPI)
 - ❖ Scannerdata grootwarenhuizen en winkelketens
 - ❖ Webscraping prijzen (vb. vliegtuigreizen, webshops)
- Exploratie en pilootstudies
 - ❖ Mobiele-telefoondata
 - ❖ Webscraping vacatures
- NIET gepland (voorlopig)
 - ❖ Slimme meters (politieke beslissing om niet uit te rollen => geen data)
 - ❖ Sociale media, zoekresultaten, tekstanalyse ('high-hanging fruit', interpretatie niet evident)

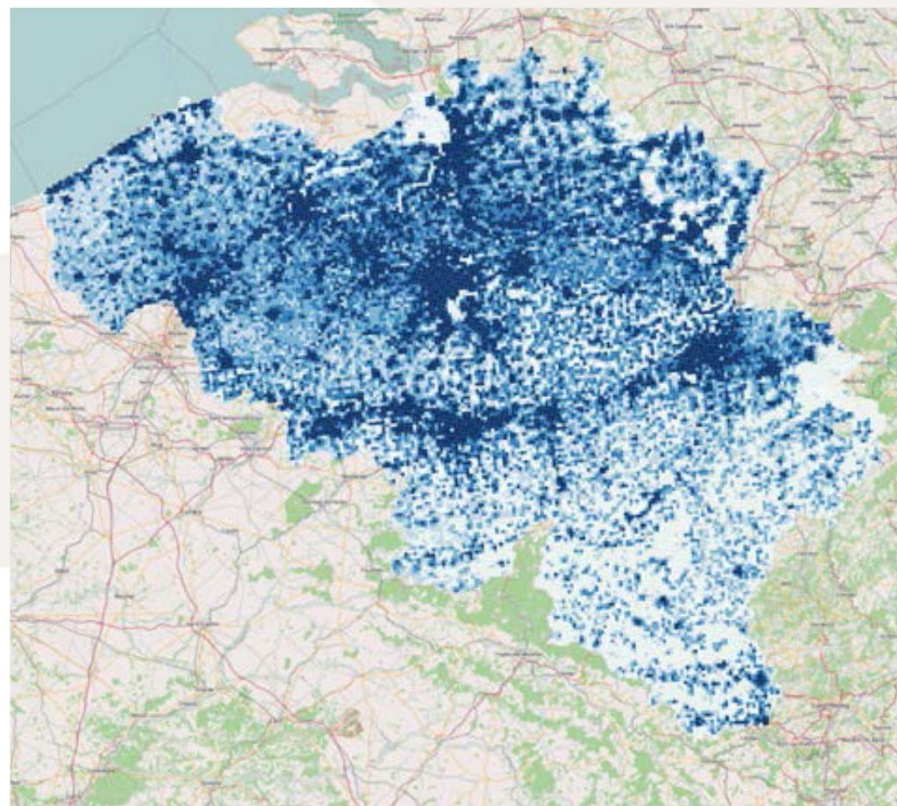
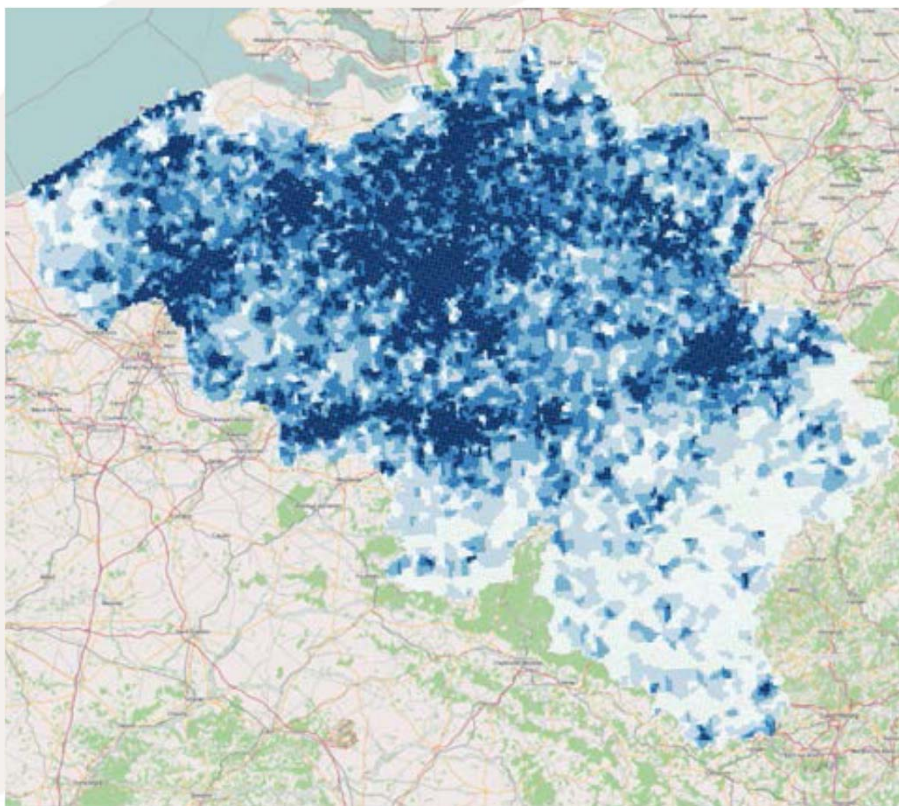
Project mobiele-telefoondata

Statistics Belgium, Proximus, Eurostat, JRC

- Start december 2015, eerste resultaten april 2016
- Stapsgewijs:
 - ❖ Eerst werkelijk aanwezige bevolking
 - ❖ Opstap naar werkelijke bevolking (via woonplaats), werkplaats, 'gewone omgeving', toerisme, arbeidsmigratie, migratie, tijdgebruik
- Innovatief:
 - ❖ Eerste en nog enig samenwerkingsproject in EU => 'echte' data
 - ❖ Geen 'call detail records' maar netwerksignalen: 10 x meer!
 - ❖ Mobiele-telefoondata combineren met statistische datasets
- Via geo-koppeling aggregaten: geen privacy issues!

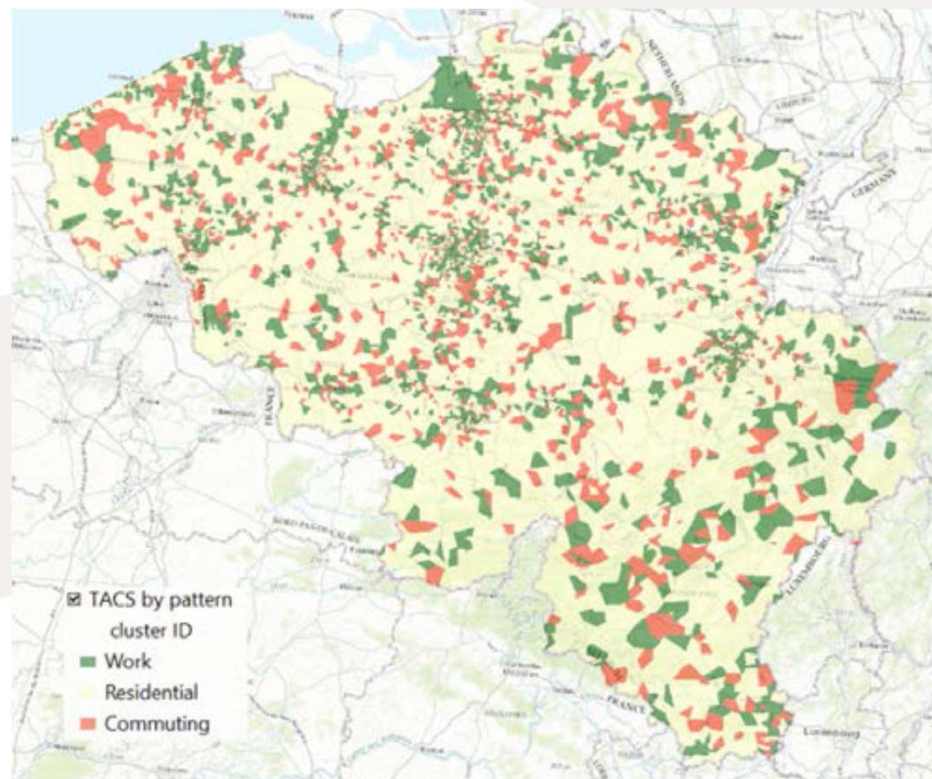
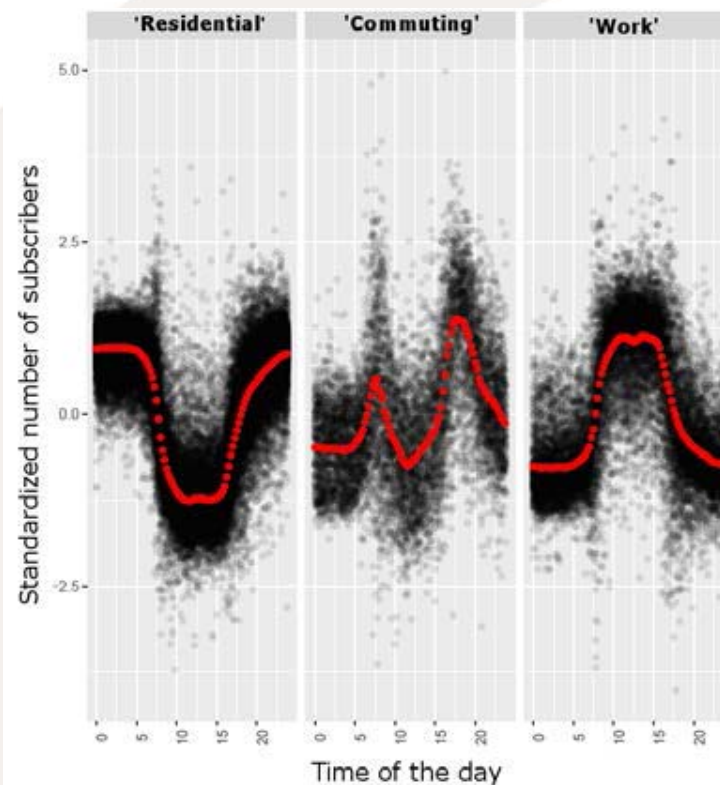
Enkele resultaten

Bevolkingsdichtheid per km²:
mobiele-telefoondata (links) versus Census 2011 (rechts)



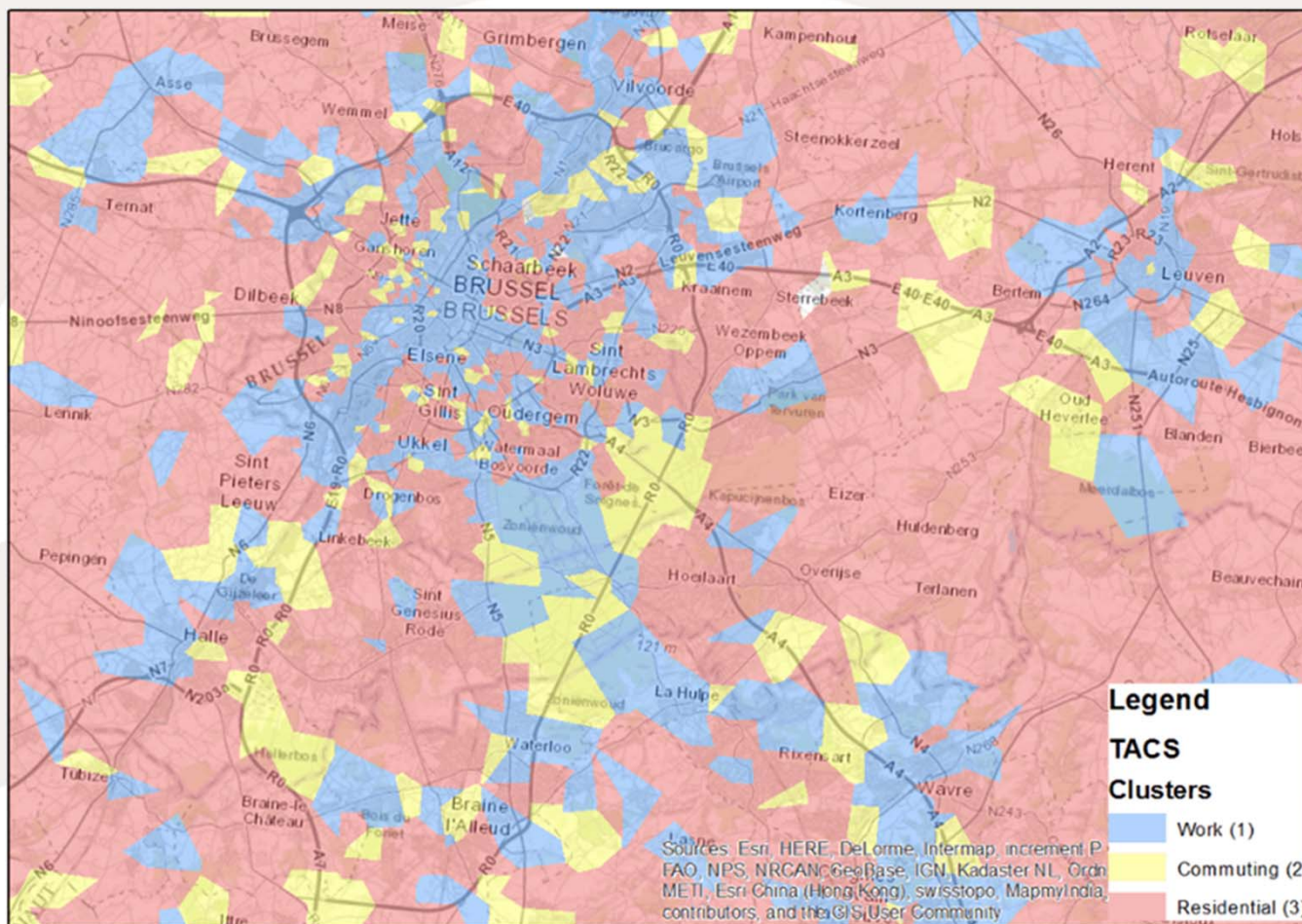
Enkele resultaten, vervolg

Weekdag-cellen geïdentificeerd als 'woon-', 'pendel-' of 'werk', met geografische voorstelling



Enkele resultaten, vervolg

'Woon'-, 'werk'- en 'pendel'-cellen in regio Brussel-Leuven



Volgende stappen

- **Verdere exploratie via gerichte pilootstudies**
Specifieke datasets Proximus gericht geanalyseerd door Statistics Belgium, Eurostat, JRC
- **Concrete use cases met oog op productielijnen**
 - ❖ Op basis van mobiele-telefoondata en statistische datasets
 - ❖ Voor statistische én commerciële toepassingen

Van exploratie naar exploitatie

- **Scannerdata en webscraping voor CPI**
- Webscraping job vacancies
- Validatie woonplaats en werkplaats Census
- Woon-werk matrix
- ... (bevolking, migratie, toerisme, mobiliteit, transport, tijdbesteding, ...)

Big data in het ESS (Europees Statistisch Systeem)

Initiatieven en projecten

- **Steering Group en Task Force Big Data**
Toezicht op big-dataprojecten binnen het ESS
- **ESSnet Big Data**
 - ❖ Zie <https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata>
 - ❖ 22 partners, 20 landen, februari 2016 => mei 2018
 - ❖ 10 *workpackages*, Statistics Belgium actief in:
 - WP1 Webscraping job vacancies
 - WP5 Mobile phone data
 - WP9 Dissemination (WP leader)
- **Weldra op basis van grant DG Regio:**
 - ❖ Grensoverschrijdende activiteit via mobiele-telefoondata e.a.

Vragen? Antwoorden? Commentaren?