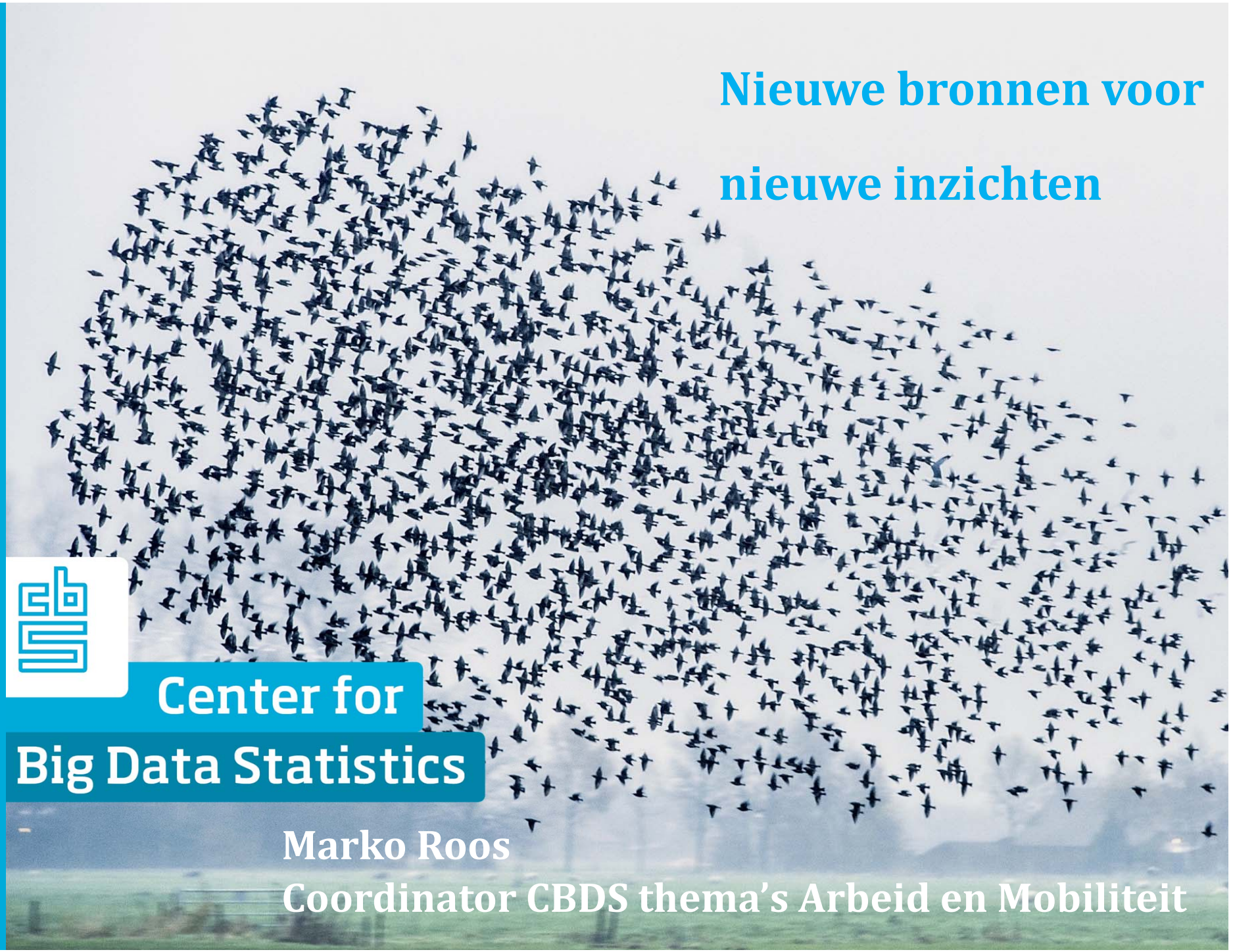
Nieuwe bronnen voor nieuwe inzichten



Center for Big Data Statistics

Marko Roos
Coordinator CBDS thema's Arbeid en Mobiliteit

Paradigm shift: data driven society



- Hoeveelheid geproduceerde data groeit exponentieel
- De samenleving verandert daardoor
- Grote kansen voor nieuwe kennis met economische waarde, nieuwe producten, betere besluitvorming
- Omgaan met data is ons vak

2

Paradigm shift

Surveys
- Labor Force Survey
- Safety Monitor
- International Trade

Admin Data Sources
- Tax Data
- Population Register
- Insurance Register

Big Data
- Sensor Data
- Social Media
- Internet Data



Output
- Health
- Economy
- Safety
- Mobility
- Housing
- Labor force

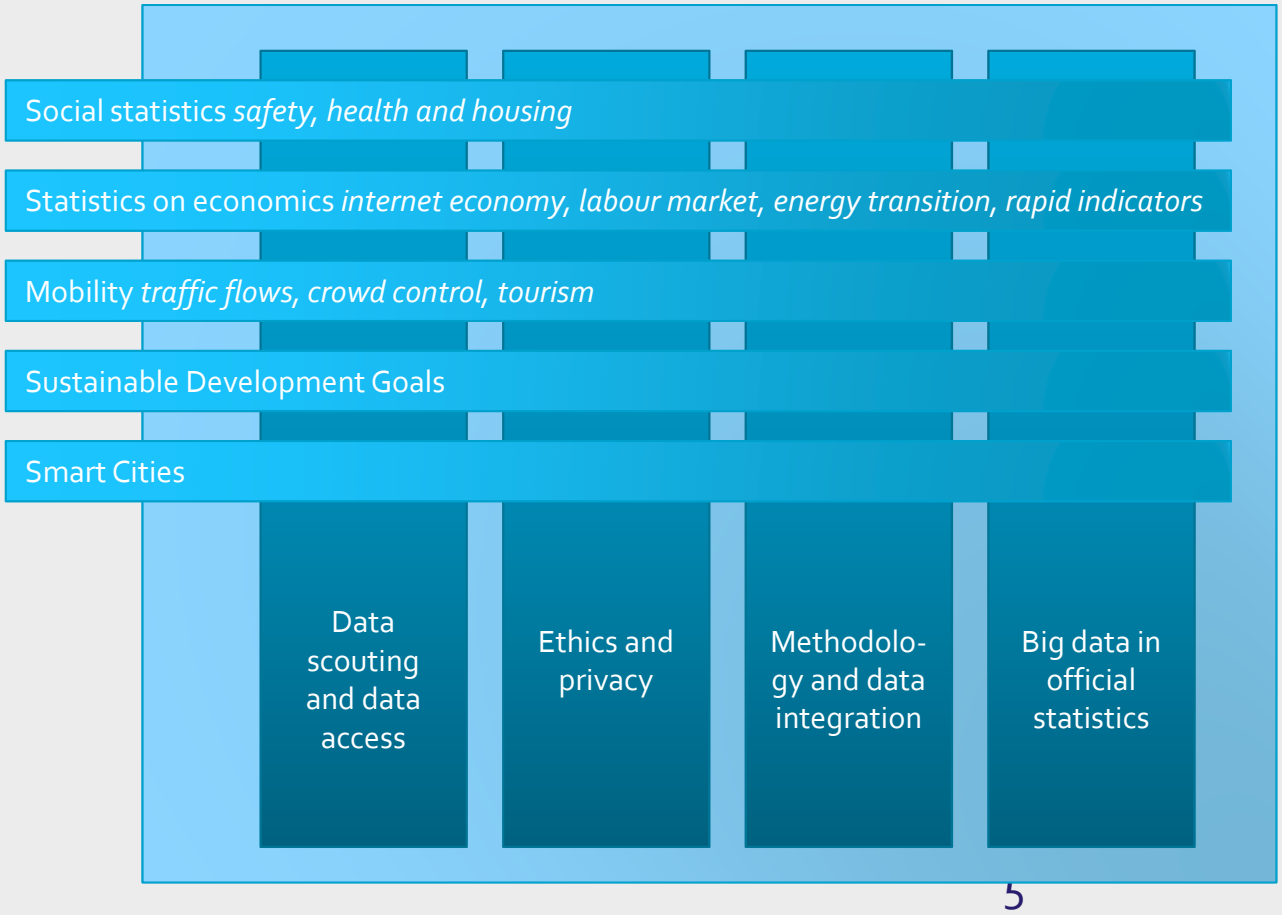
3

CBS Center for Big Data Statistics (CBDS)



- Van denken naar doen: Big data gebruiken ten behoeve van *evidence based policy*
- Nieuw, real time, regionaal
- Safe haven/ data knooppunt
- Kenniscentrum en opleider
- Unieke hotspot met internationale exposure
- Krachtig netwerk

Scope


Social statistics *safety, health and housing*
Statistics on economics *internet economy, labour market, energy transition, rapid indicators*
Mobility *traffic flows, crowd control, tourism*
Sustainable Development Goals
Smart Cities
Data scouting and data access
Ethics and privacy
Methodolo-gy and data integration
Big data in official statistics

Partner netwerk

- CBDS introduceert een internationaal ecosysteem bestaande uit bedrijven, universiteiten, kennis instituten, overheden and National Statistical Institutes (NSI's)

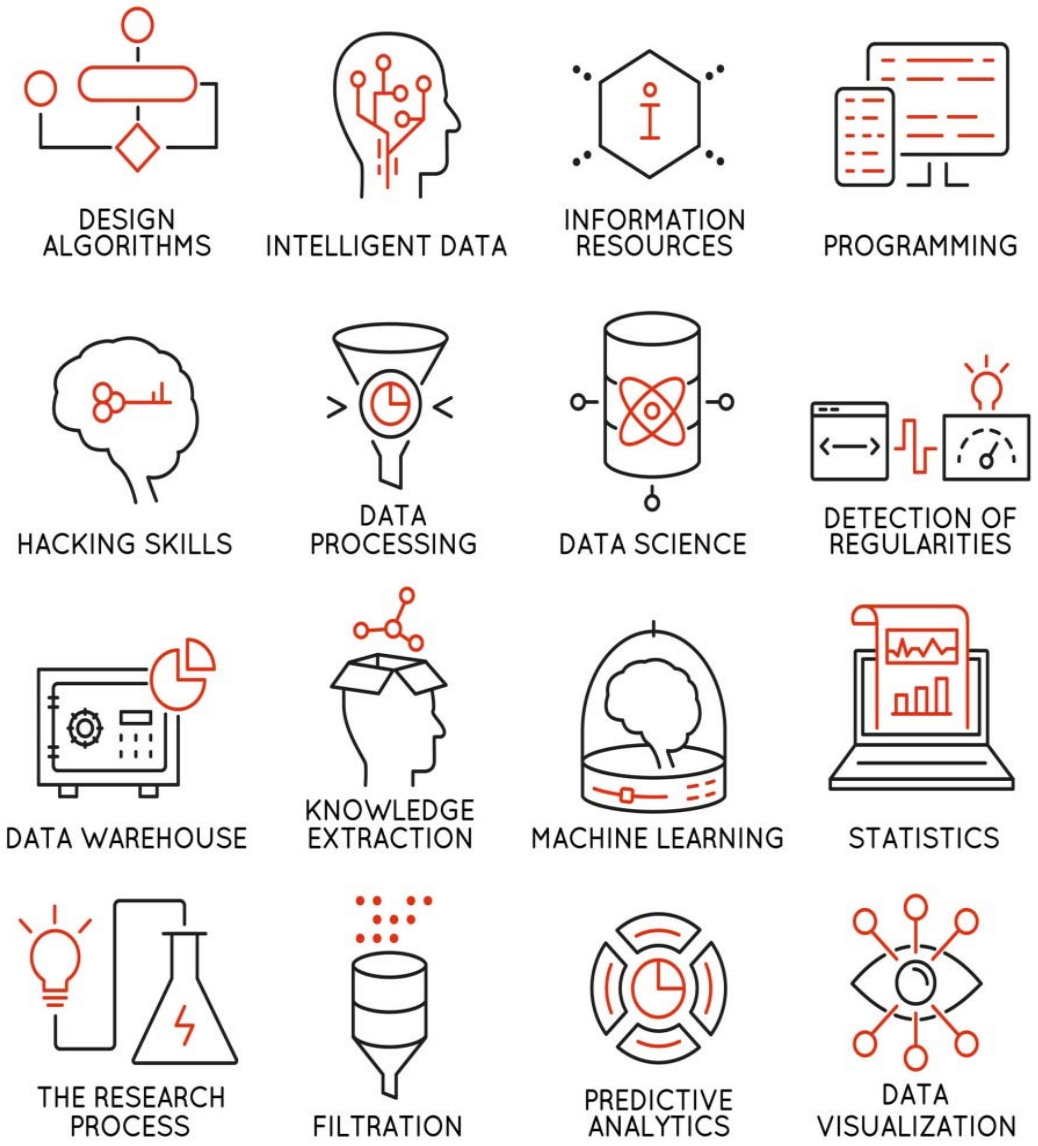
- Dit netwerk ontsluit kennis, data, resources en infrastructuur. Krachtige combinaties mogelijk voor elke fase in product ontwikkeling, research en fund raising

- CBDS brengt veel kennis van en ervaring met data processing, het combineren van bronnen, privacy, en uiteraard de grote hoeveelheid van bronnen die al beschikbaar zijn binnen CBS

Partner netwerk



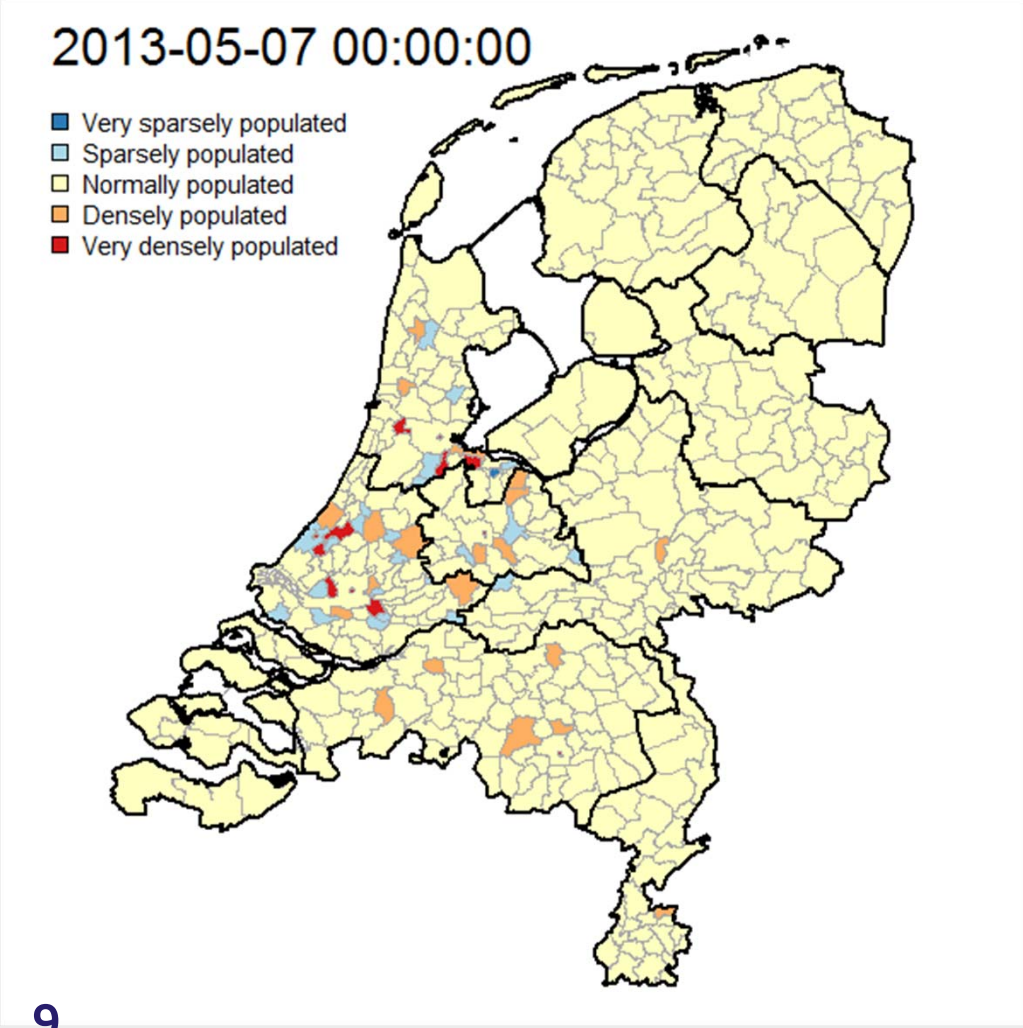
Big data science: profiel

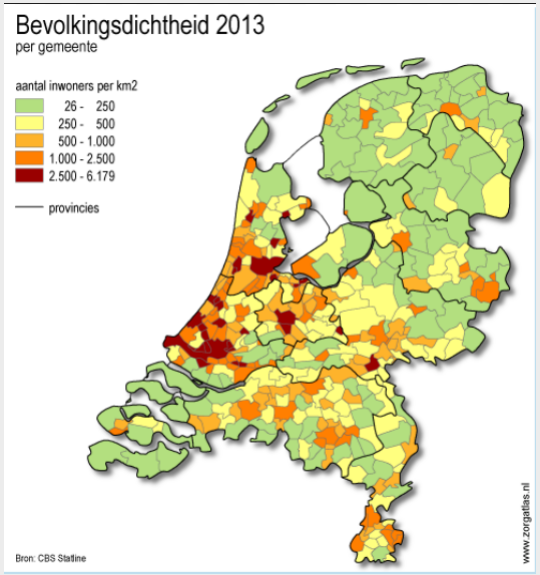


Center for Big Data Statistics

m.roos@cbs.nl

Day time population

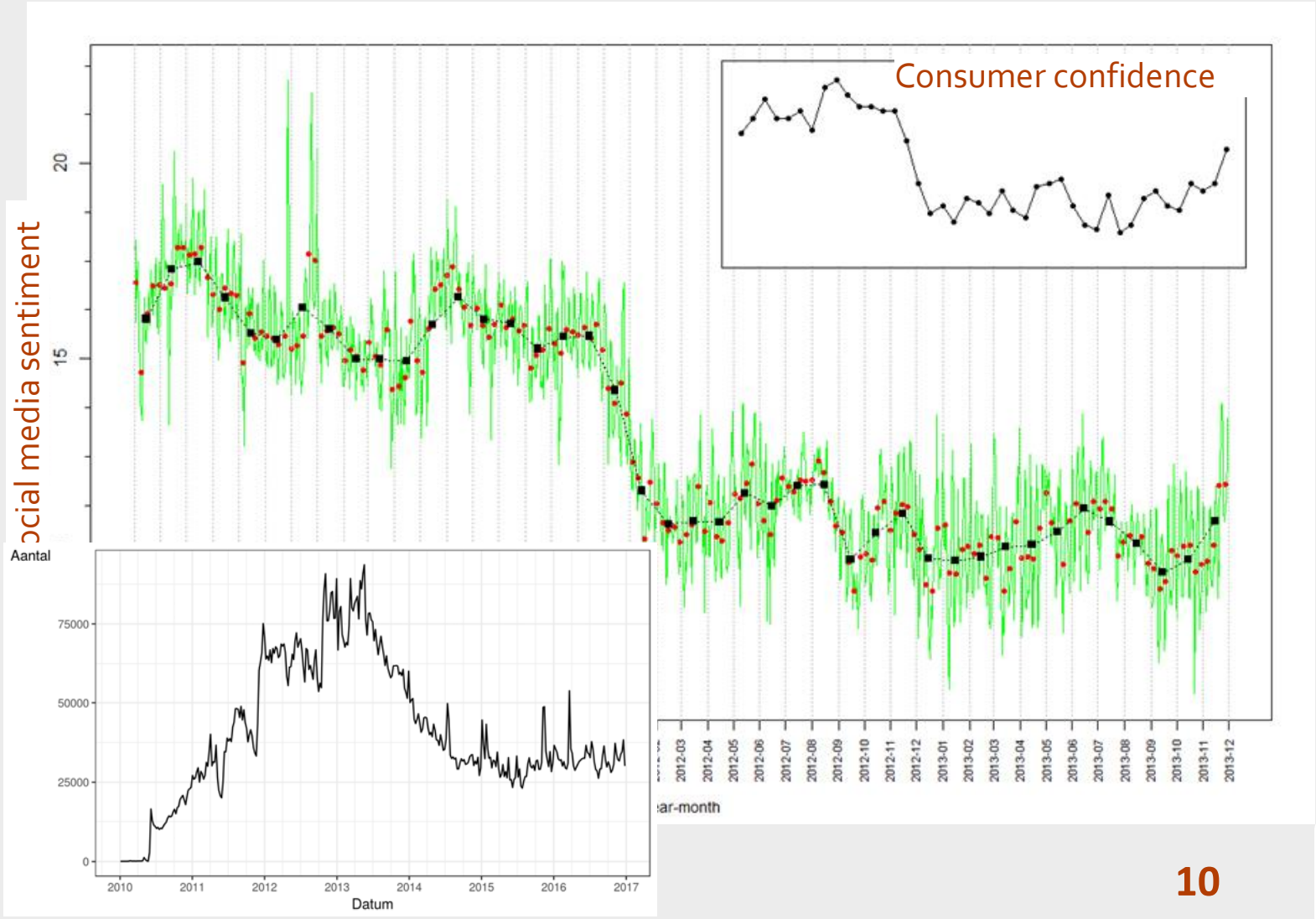




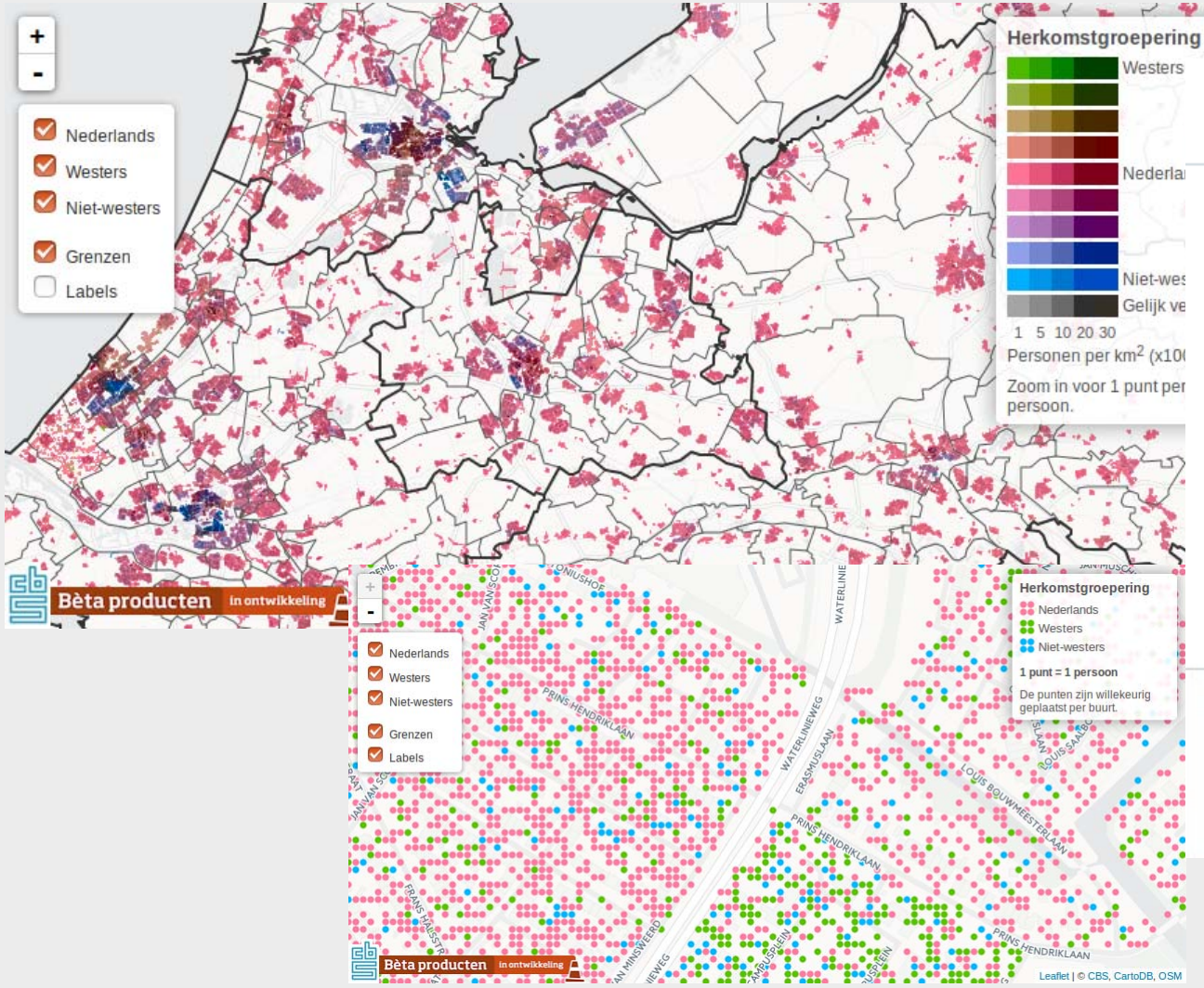
Hourly changes of mobile phone activity

Only data for areas with
> 15 events per hour

Social media en vertrouwen



Visualisatie: de dotmap



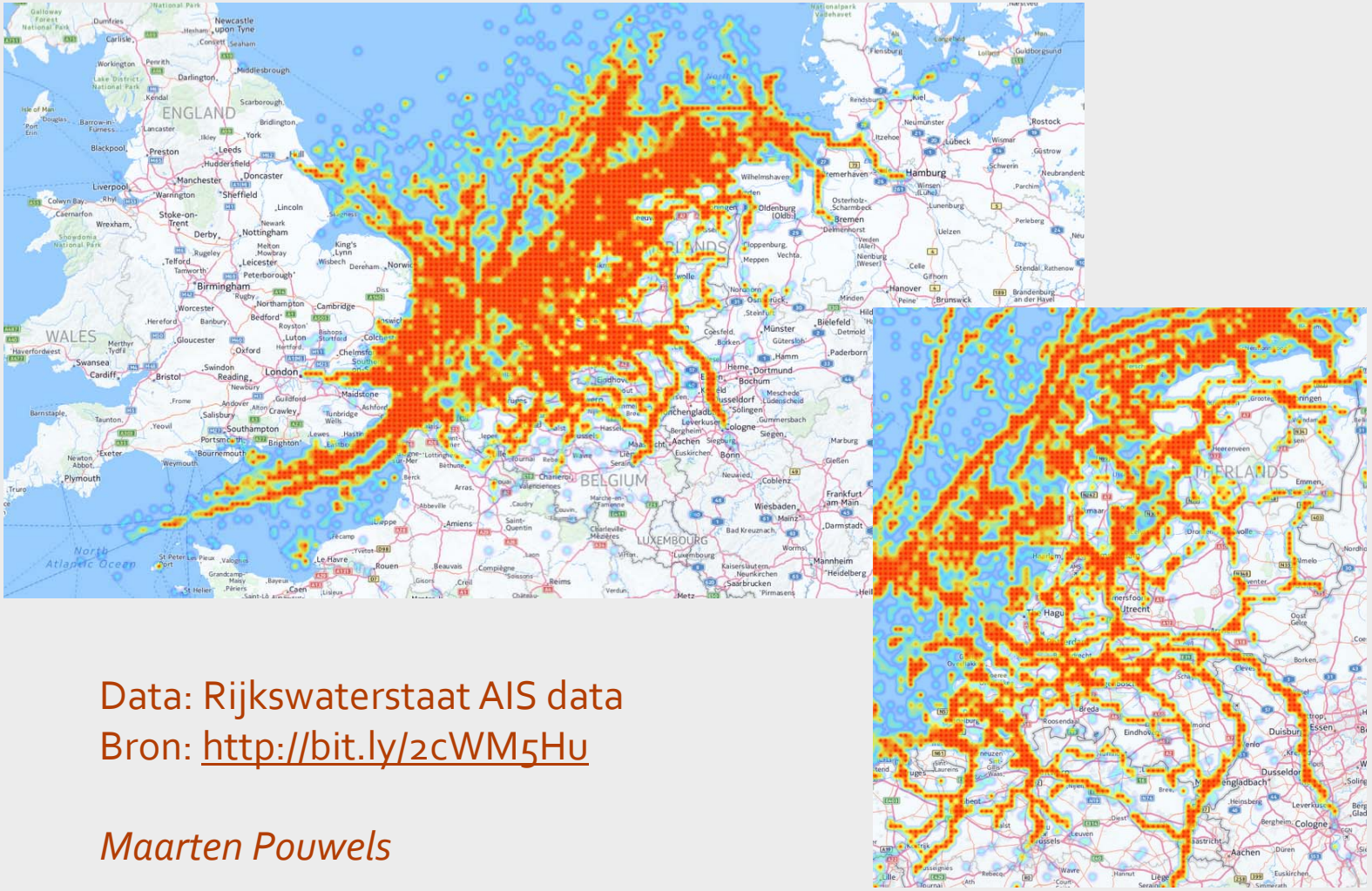
Ton Martens @tonneman86 18h
@statistiekcbs Wow! Indrukwekkend!

Knakdeworst @knakdeworst 16h
Heel mooi @statistiekcbs Suggestie: kaart met aantal dieren NL en locatie op aarde waar deze geconsumeerd worden. twitter.com/statistiekcbs/...

CBS @statistiekcbs
De door CBS ontwikkelde #dotmap toont p. buurt hoeveel inwoners er zijn en wat hun herkomst is. Hoe dat werkt? Zie: cbs.nl/nl-nl/onze-die...

Maciek Jablonski @JablonskiM... 13h
@statistiekcbs mooi en bruikbaar informatiebron. Tip: maak een app - website is niet handig op mobiele toestellen.

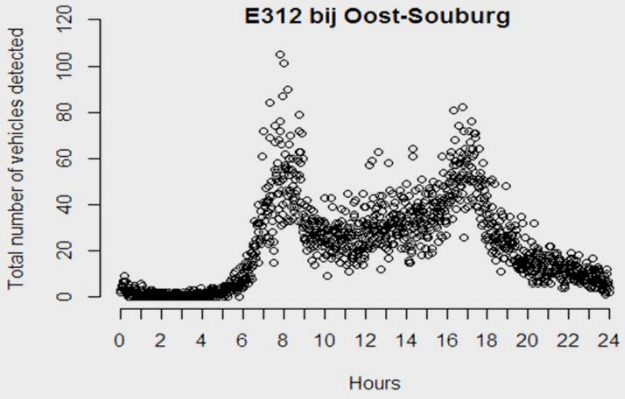
AIS data



Data: Rijkswaterstaat AIS data
Bron: http://bit.ly/2cWM5Hu

Maarten Pouwels

Traffic intensity and GDP



- GDP
- Traffic



Traffic precedes GDP!

- By 1 quarter

Correlation

- 91% from 2011-Q2 till 2014-Q4

Internet economie

Bepaling van de impact van de internet economie, in samenwerking met Google en Dataprovider

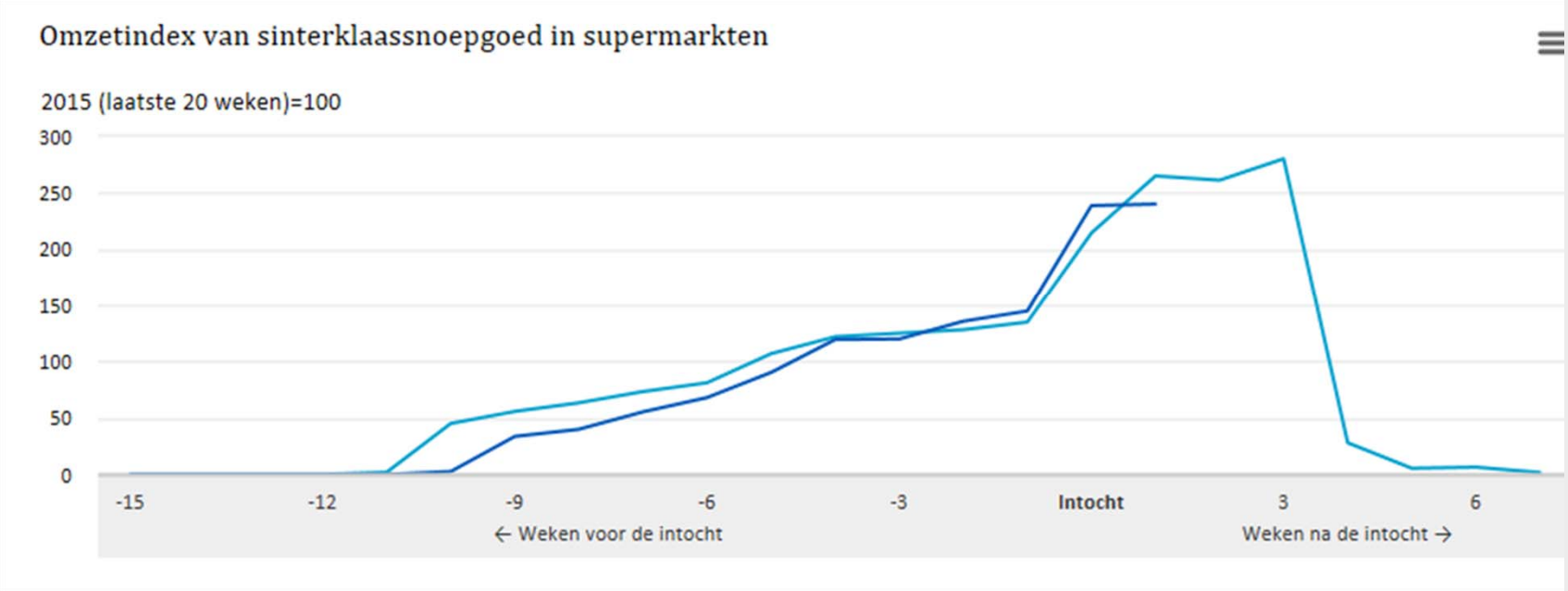


Met gebruik van Big Data van websites en koppeling met het ABR (omzet, aantal werknemers, etc), opdeling van bedrijven in diverse categorieën van de internet economie

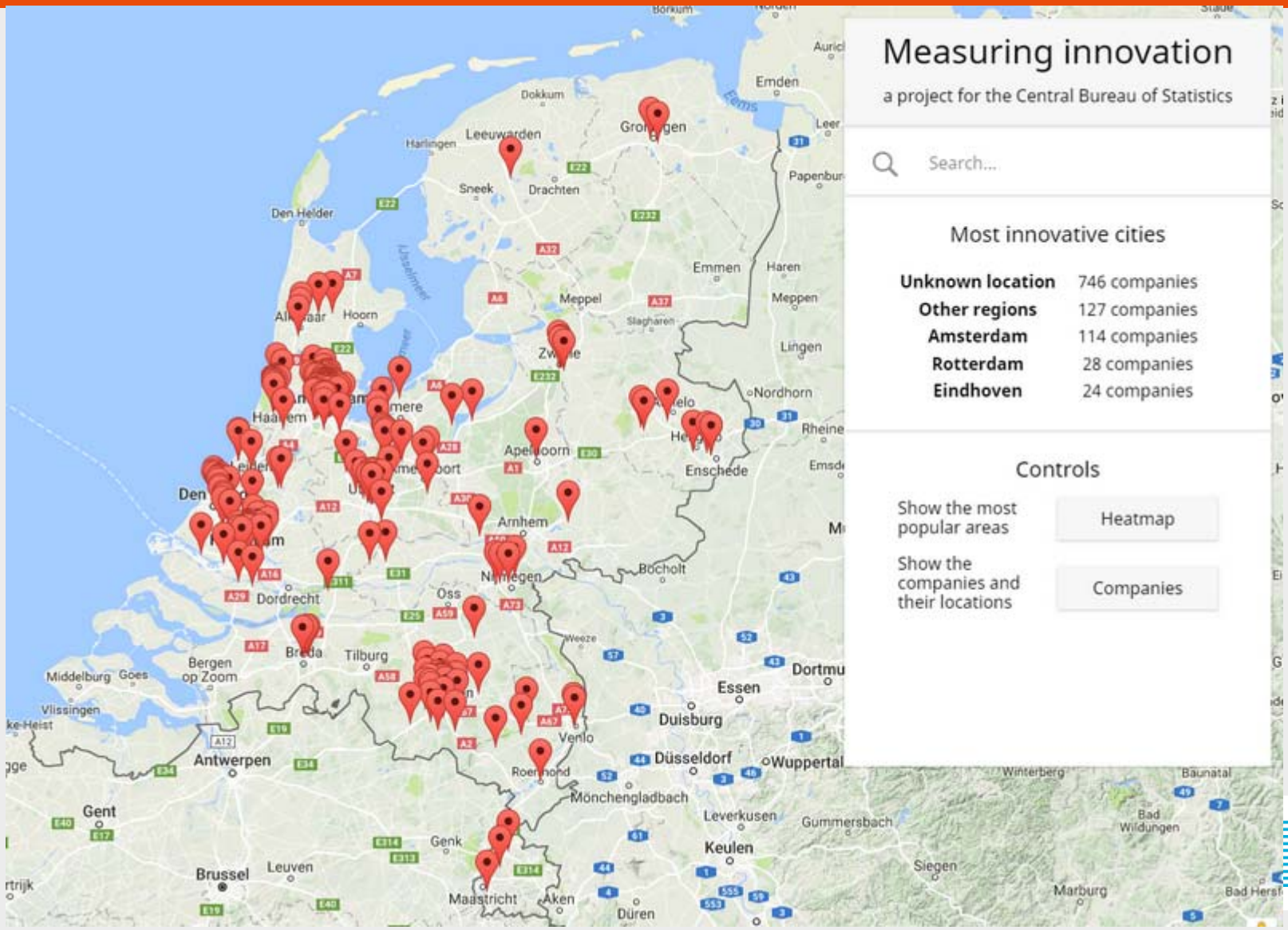
Google

dataprovider

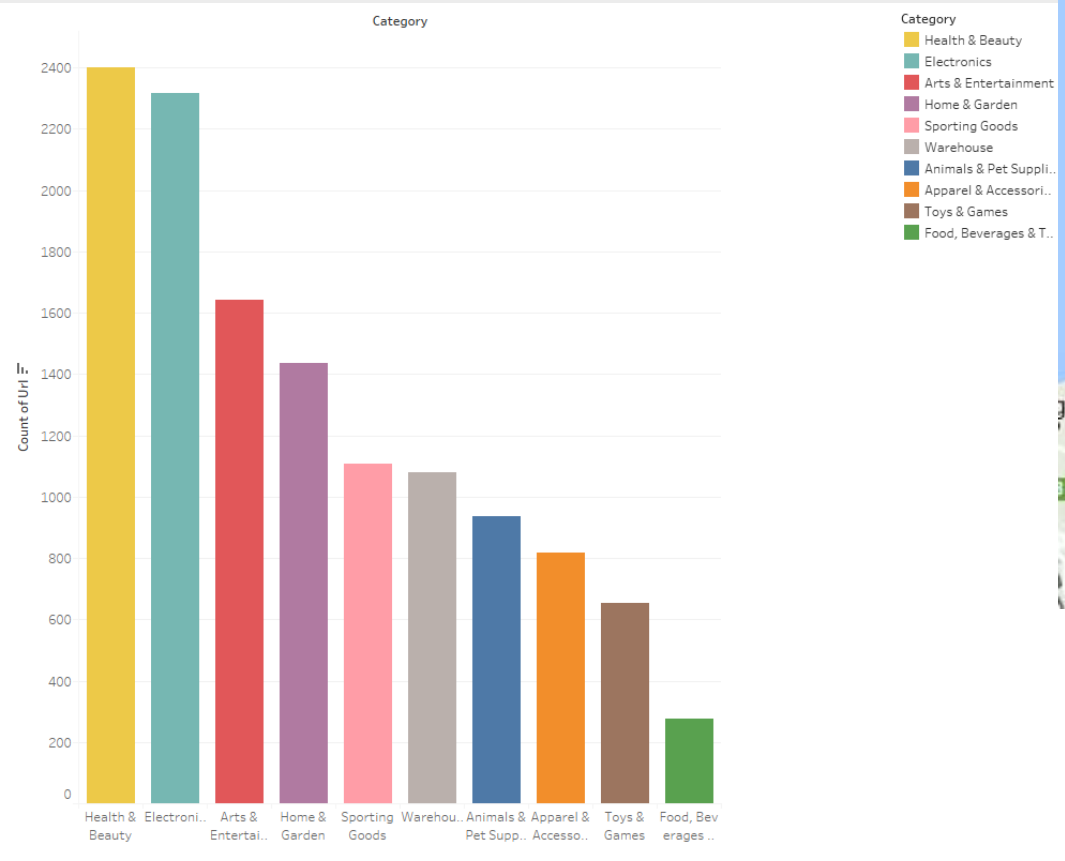
De pepernoten index



Common crawl: innovatie



Common crawl: Dutch webshops





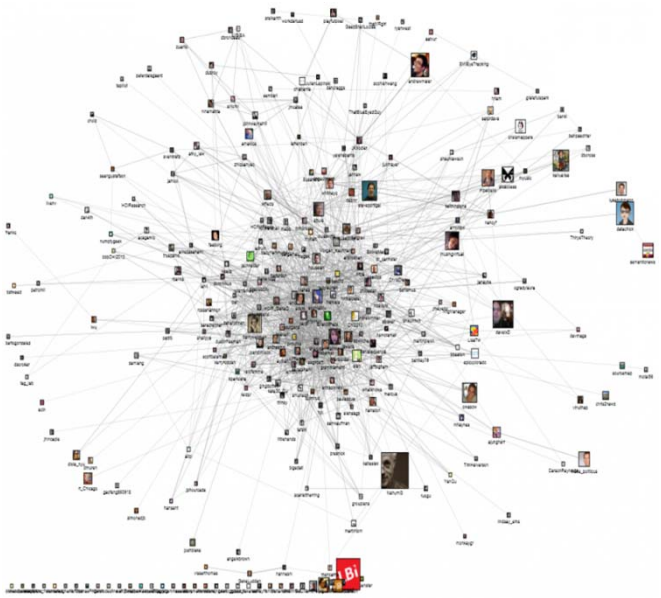
Arbeidsmarkt – Big data bronnen

Mogelijke bronnen:

- Webscraping: vacatures, profielen
- Polisdata, Inkomstenbelasting, Omzetbelastingen, GBA, Nieuw handelsregister, …
- Google trends
- Mobiele telefoongegevens
- …

Center for
Big Data Statistics

Arbeidsmarkt: webscraping



Web Scraping



- In samenwerkingsverband onderzoek naar:
  - Informatie uit openbaar beschikbare profielen
  - Netwerken van profielen
  - Geclassificeerde beroepen, skillsets

- In samenwerkingsverband onderzoek naar:
  - Informatie uit openbaar beschikbare vacatures
  - Geclassificeerd naar beroepssoorten, bedrijfstak, vereiste opleidingsniveaus, skillsets

Center for Big Data Statistics

Arbeidsmarkt – Polisdata, inkomstenbelasting, omzetbelastingen, …

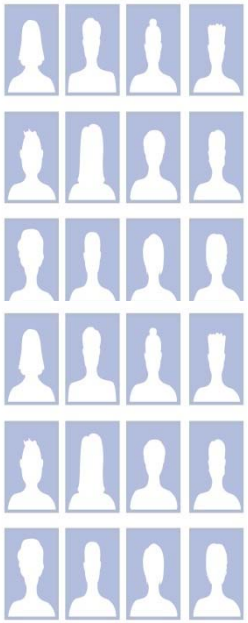
- Al veel gedaan:
  - Gestructureerde datasets en als ‘standen’ dataset goed verwerkbaar en beschikbaar in SSB
  - Al veel producten ontwikkeld (bijvoorbeeld):
    - Arbeidsmarktstromen: stromen van inactiviteit naar werkzaam, stromen tussen bedrijfstakken, …
    - Statistiek wonen-werken
    - Inkomensverdelingen
    - …
- Maar:
  - Nieuwe modelleringstechnieken icm rekenkracht kunnen dataset mogelijk verder verrijken
  - Datasets kunnen met (informatie uit) nieuwe bronnen gecombineerd worden (adressen van niet geregistreerde instituten, afstanden, routes obv Google Street, …)

Center for
Big Data Statistics

Onderzoek: van Big data naar informatie

Verrijken microgegevens:

- Met (nieuwe) modelleringstechnieken en (nieuwe) koppelmethoden informatie toekennen aan eenheden op microniveau







Center for
Big Data Statistics

Onderzoek: van Big data naar informatie

Verrijken classificaties:

- Classificaties verrijken met nieuwe informatie uit big data bronnen
  - Bijvoorbeeld; opleidingen en vaardigheden geassocieerd met beroepen



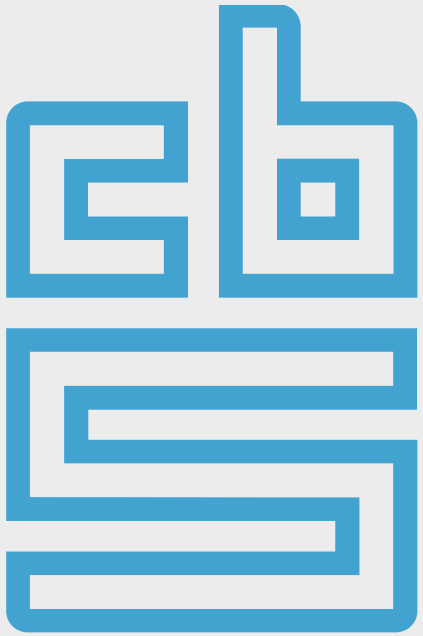
Center for
Big Data Statistics

Onderzoek: van Big data naar informatie

Nieuwe informatieproducten:
- Afhankelijk van kwaliteit verrijkte microgegevens en classificaties!
- In potentie: meer doorsnedes naar achtergrondkenmerken voor arbeidsmarktdynamiek, vacatures, beroepen
- In potentie: meer informatie over dynamiek van classificaties (van bijvoorbeeld bedrijven, beroepen)

Risico's:
- Black box modellen
- Onvolledigheid van de bronnen
- Bewerking van de bronnen door derden
- Privacy!

Center for
Big Data Statistics



Voor wat er **feitelijk** gebeurt