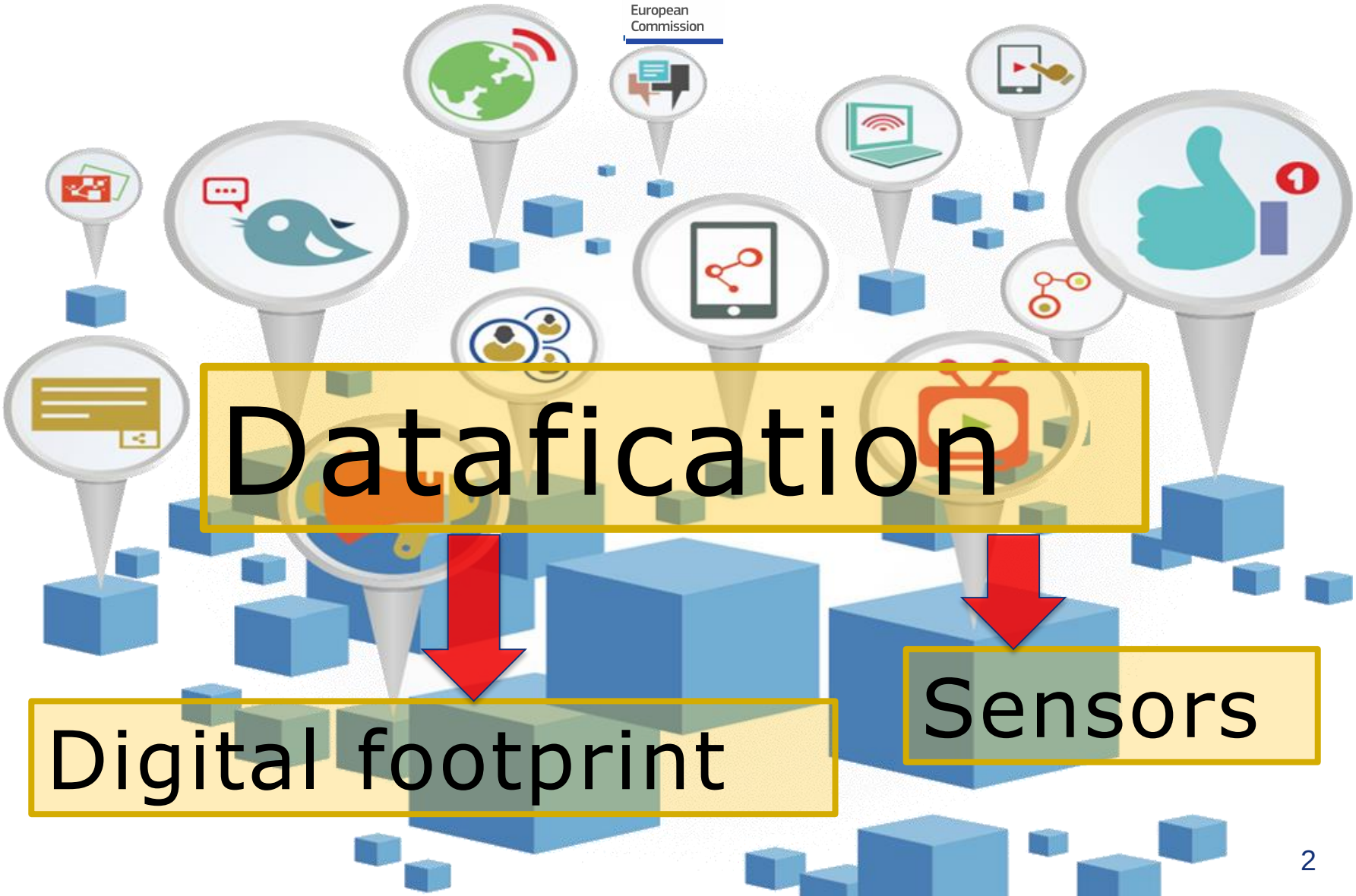
Big Data@Eurostat

UN-GGIM Europe – ESS meeting
Luxembourg, 11 March 2016

Albrecht.Wirthmann@ec.europa.eu


European Commission
Datafication
Digital footprint
Sensors
Eurostat

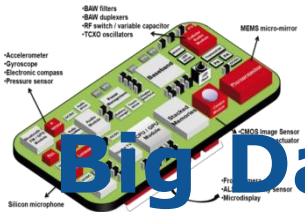
Big Data

What will be the **impact of ubiquitous data collection and networking**

- Mobile Communication
- Internet of [every]Things,
- Social media,
- Wearables,
- Autonomous traffic,
- Smart systems,
- …

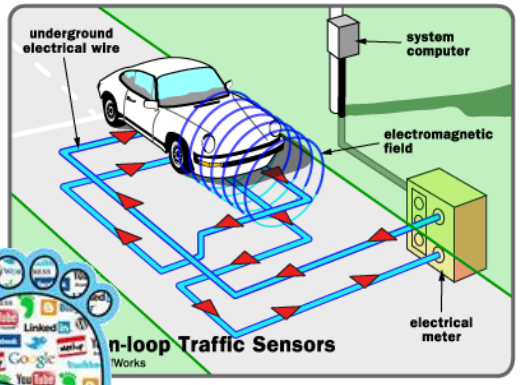
on **official statistics**?

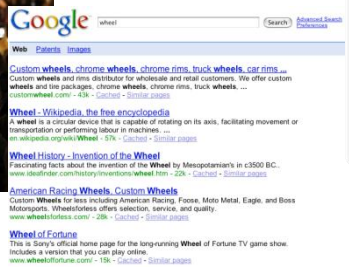
European Commission

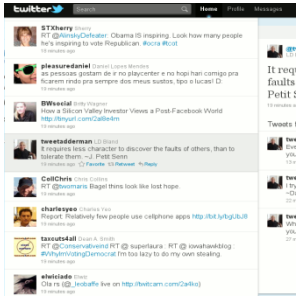


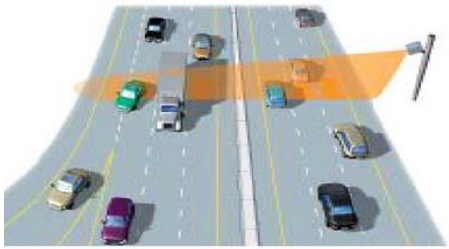
Big Data and Statistics

- National Initiatives
  - **CBS Netherlands**
  - **ISTAT Italy**
  - **ONS UK**
  - **CSO Ireland**
  - **Statistics Finland**
  - **SURS Slovenia**









Eurostat

Big data at Eurostat

ESS (European Statistical System)

- **Scheveningen Memorandum Sep 2013**
  - Task Force Big Data
  - Big Data Roadmap and Action Plan June 2014
    - May 2015 Big Data Business Cases
    - Nov 2015 Big Data ESSnet
    - Jan 2016 Contract on supporting services
- **ESS Vision 2020**

European Commission *Communication*

- Data4Policy Networking Group

International cooperation (UNSD, UNECE, etc.)

- UN/ECE project "Big data in official statistics" (**Sandbox**)
- UNSD Global WG on Big Data

European Commission

Action Plan Themes

Policy	Quality	Skills
Experience sharing	Legislation	IT Infrastructures
Methods	Ethics / Communication	Partnerships

Pilots

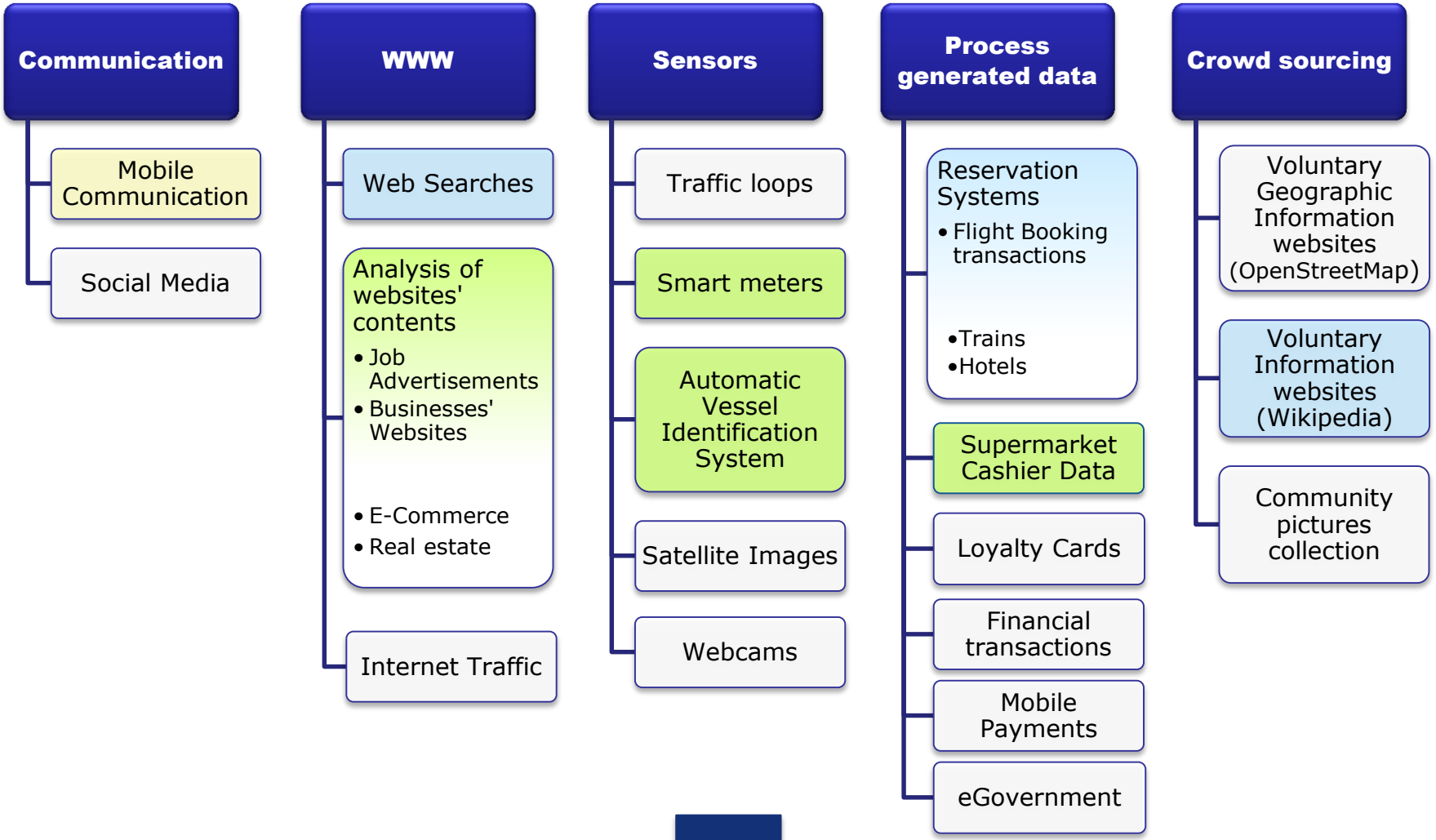
Challenges

- Exploration & tentative implementation
- Scientific Approach
- Cooperation, sharing of know-how
- Access to data, lack of skills, new user demands

Actions

- **Big Data ESSnet: Pilot projects**
  - 2015 – 2019 (FPA/SGA construction)
  - Exploring different big data sources
    - Characteristics of data, potential outputs, Methodology and Quality, IT requirements
  - Partnerships with data providers, academia, users
  - Cooperation with UN
- **Big Data Competition**
- **Workshop Oct 2016**

Big Data Pilots



ESS Big Data Pilots

- **List of pilot projects** (Specific Grant Agreement)
  - Web scraping
    - job vacancies ; enterprise characteristics
  - Smart meters
    - electricity consumption ; temporary vacant dwellings
  - Automatic Identification System (Ships)
    - vessel identification data
  - Mobile phone data
    - Preparing for Access to data
  - Scenario for using multiple inputs

Eurostat Pilot Activities

- Web activity evidence for nowcasting
  - Google Trends
  - Employment
- Wikipedia as source for statistics
  - Cultural and regional statistics
- Mobile communication data
  - Population
  - Urban
  - Tourism

- Flight Reservation Systems for true Origin/Destination
  - Amadeus Database
- Aircraft Tracking System
  - Satellite system to receive flight positions from aircrafts

European Commission

A threat to official statistics?

De Standaard

Kust tevreden met verlengd paasweekend

Vandaag om 11:17 door llo | Bron: BELGA

0 | 12 Tweeten | 1 Aanbevelen | Mail | Print

Foto: Kurt Desplenter

De toeristische sector aan de kust is tevreden over het verlengde openingsweekend van de paasvakantie. Dat blijkt dinsdag uit een monitoring van provinciebedrijf Westtoer in samenwerking met Proximus en het Agentschap Wegen en Verkeer.

Met 150.000 dagtoeristen op Pasen, en 400.000 dagtoeristen verspreid over het hele paasweekend, blikt de kust tevreden terug op het startschot van de paasvakantie.

'De vele evenementen in de kustgemeenten hebben tijdens de voorbije dagen gezorgd voor aangename vakantiesfeer', aldus Franky De Block, gedeputeerde en voorzitter Westtoer.

Ook voor het verblijfstoerisme schoot de paasvakantie goed uit de startblokken. Hotels noteerden een bezettingsgraad van 90 procent of meer. Bij de verhuur van vakantiewoningen worden degelijke cijfers genoteerd met een gemiddelde bezetting van 75 procent en iets meer reservaties in de eerste week, gaf Westtoer vorige week al mee, maar met de aankondiging van het uitstekende lenteweer in de loop van de week verwacht de sector nog tal van last-minuteboekingen.

Article released one day after 2015 Easter weekend about tourism in Belgian coast:

150 000 same-day visitors on Sunday,
400 000 during the entire long weekend

- Data based on a monitoring of the regional tourism board, in cooperation with the main mobile network operator Proximus and the road infrastructure administration;
- In comparison: Eurostat will receive data on same-day visitors for the 2nd quarter of 2015 (not a particular weekend) on 30 June 2016 (not the day after) for the entire country (not a coastal strip within a NUTS2 region);
- Methodology not clear, but it's a nice example of how flash estimates based on big data decreases the relevance of official statistics;

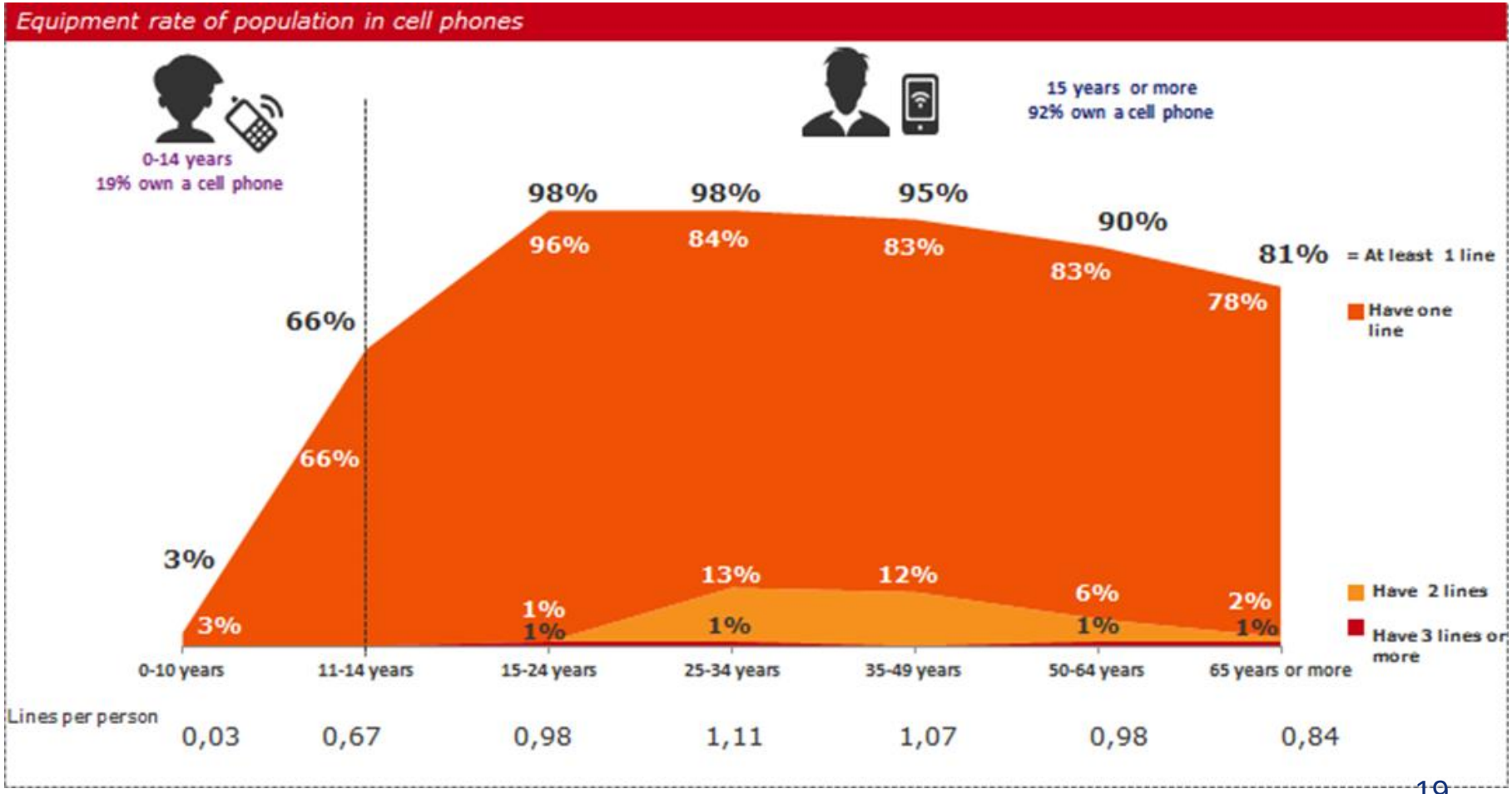
Mobile phone data

Pilot Projects

- Proximus / Statistics Belgium / Eurostat
  - **Number of mobile phones per antenna (tesselation)**
  - **Typical weekday, Sunday**
  - **The location of each mobile phone is recognized after max 3 hours**
  - **Can we use the data to estimate population density by grid-cell?**
  - **Issues**
    - Selectivity
    - Size of cells
    - Place of residence
    - …

Mobile phone data

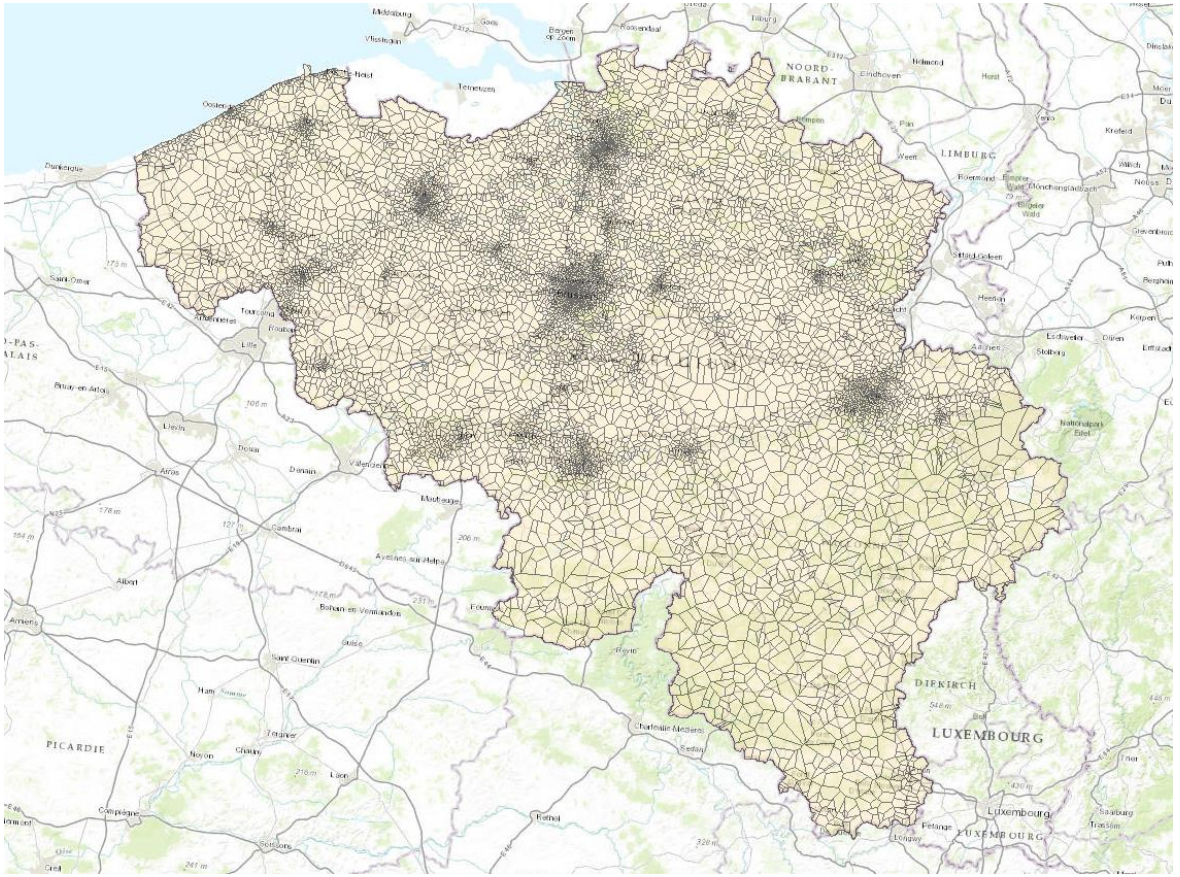
Promising, but some quality issues, e.g. coverage (representativeness?)



Source: DGE, SDP3E, bureau des études sur le tourisme et les catégories d'entreprise (France, Sept 2015)

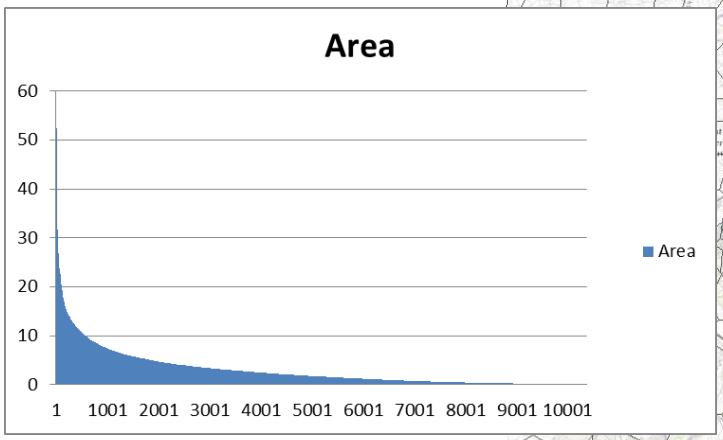
Mobile phone – radio cells

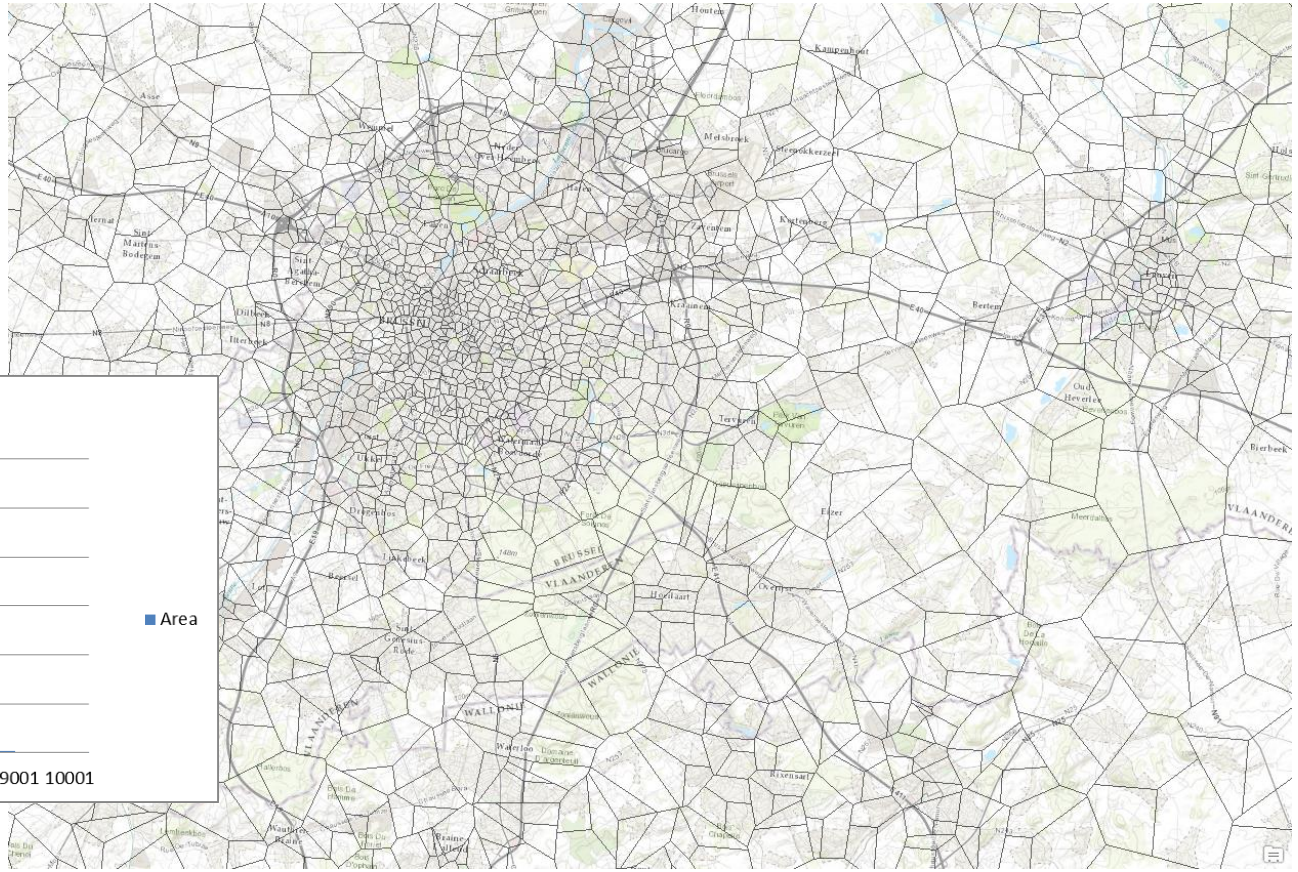
- Tesselation
- > 10,000 cells
- Combining 2,3,4 G cells



Mobile phone – radio cells

Mean	295	ha
Median	166	ha
Maximum	5231	ha
Minimum	0.2	ha



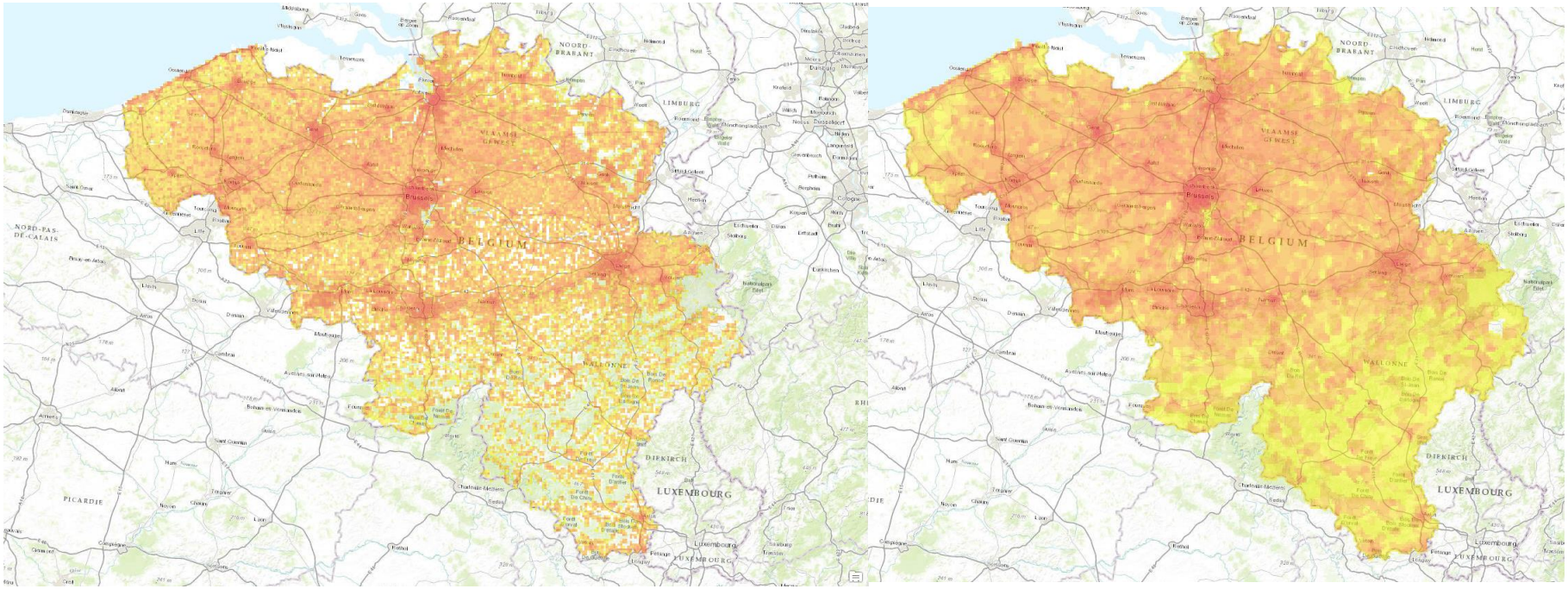


European Commission

Mobile phone – population

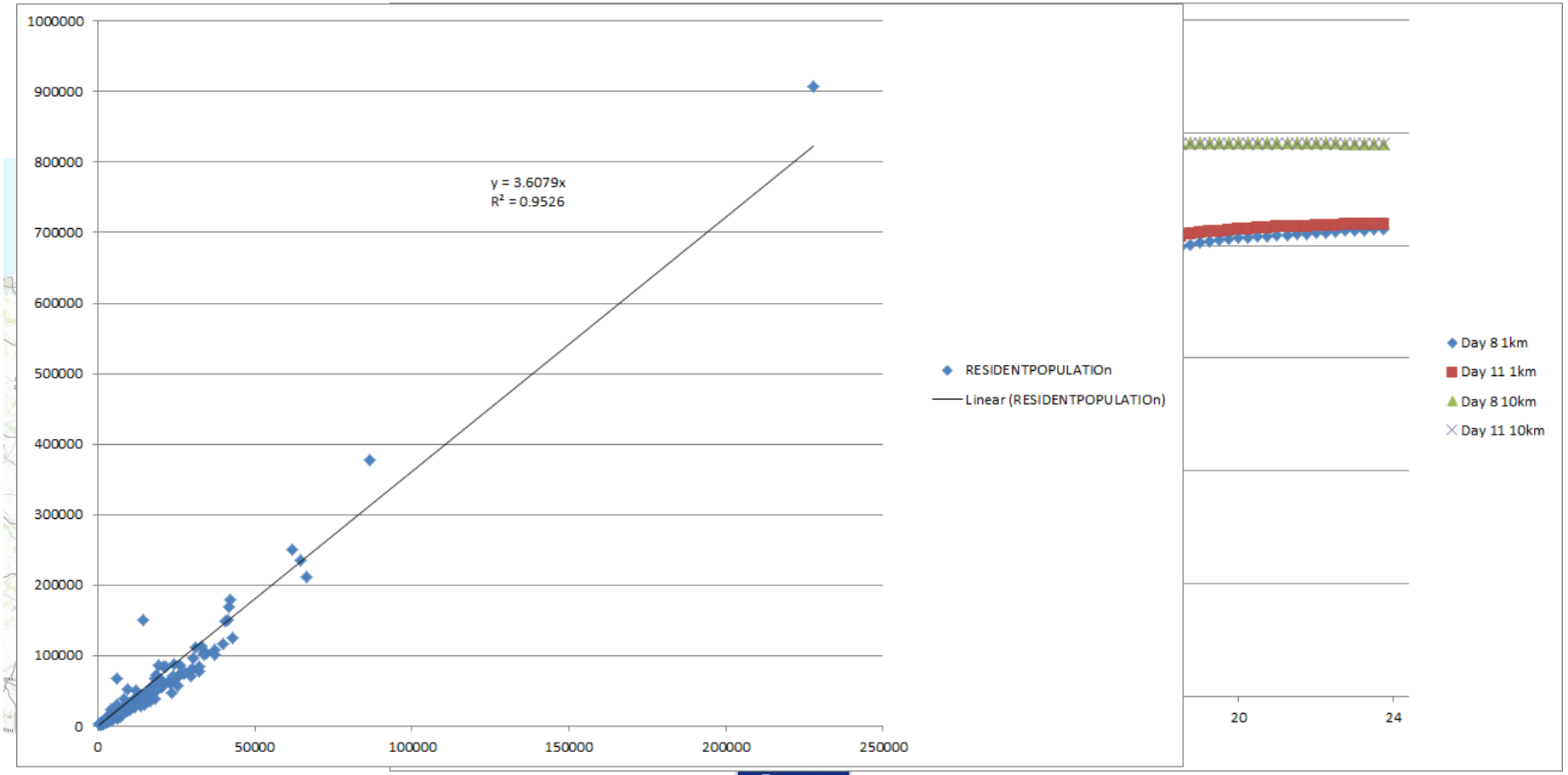
Census 2011

Mobile phones

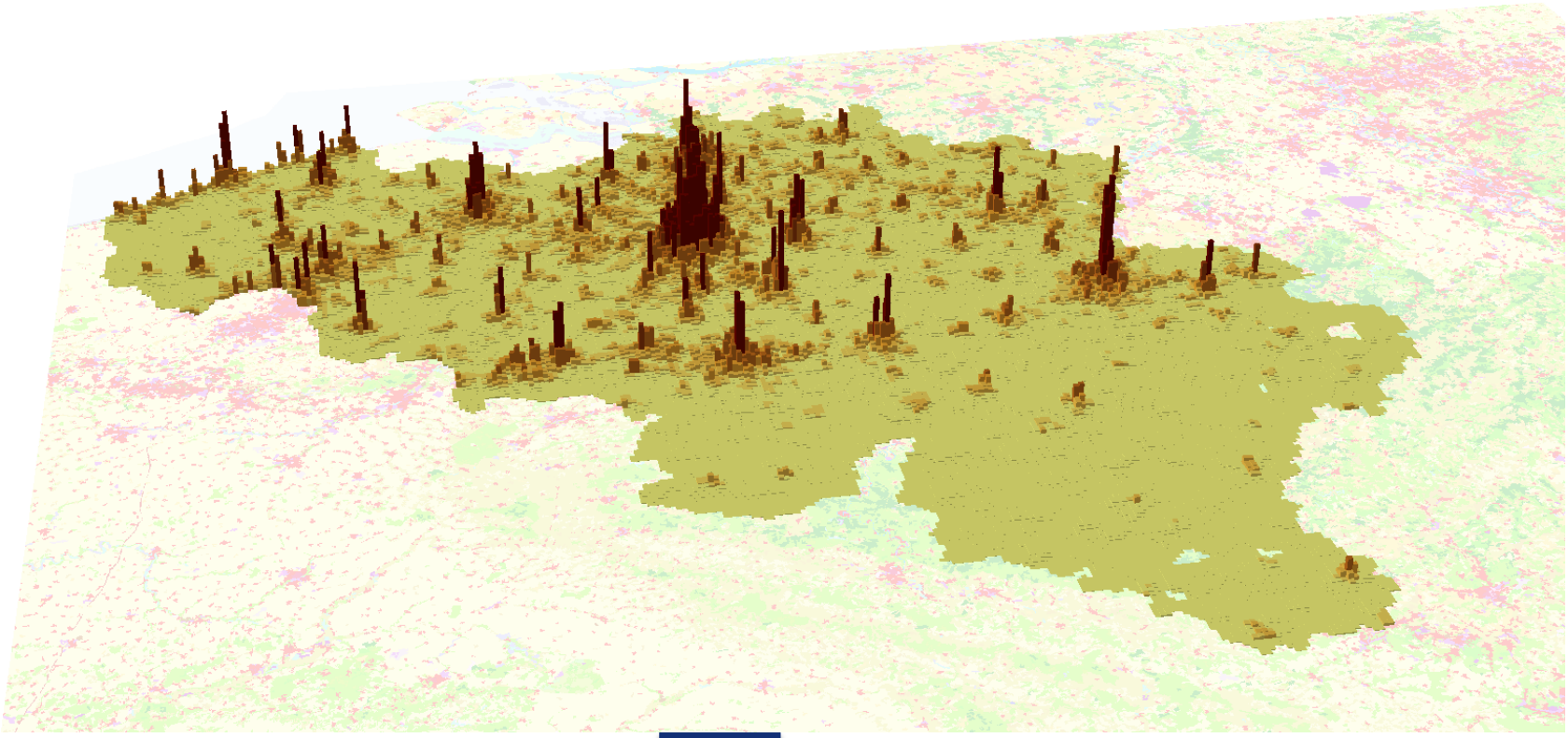


Eurostat

Mobile phone – population

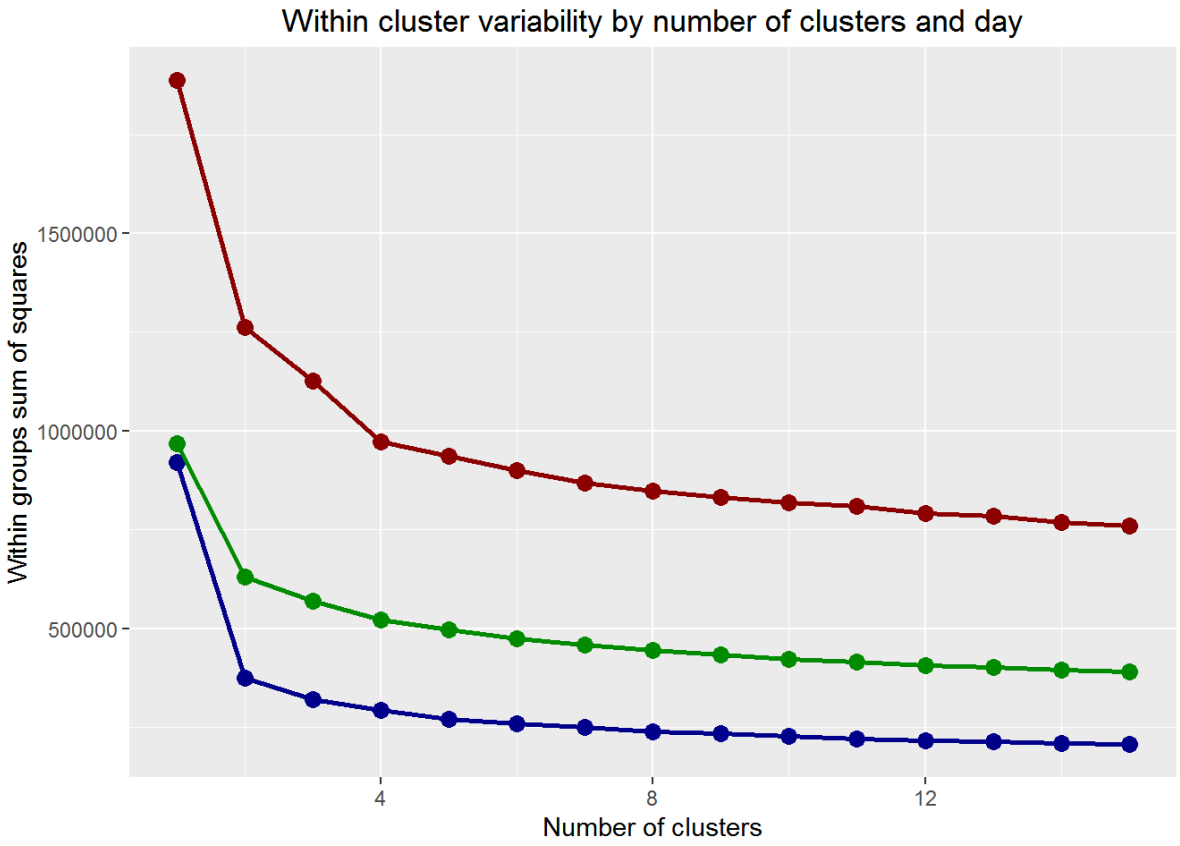


Mobile phone – population

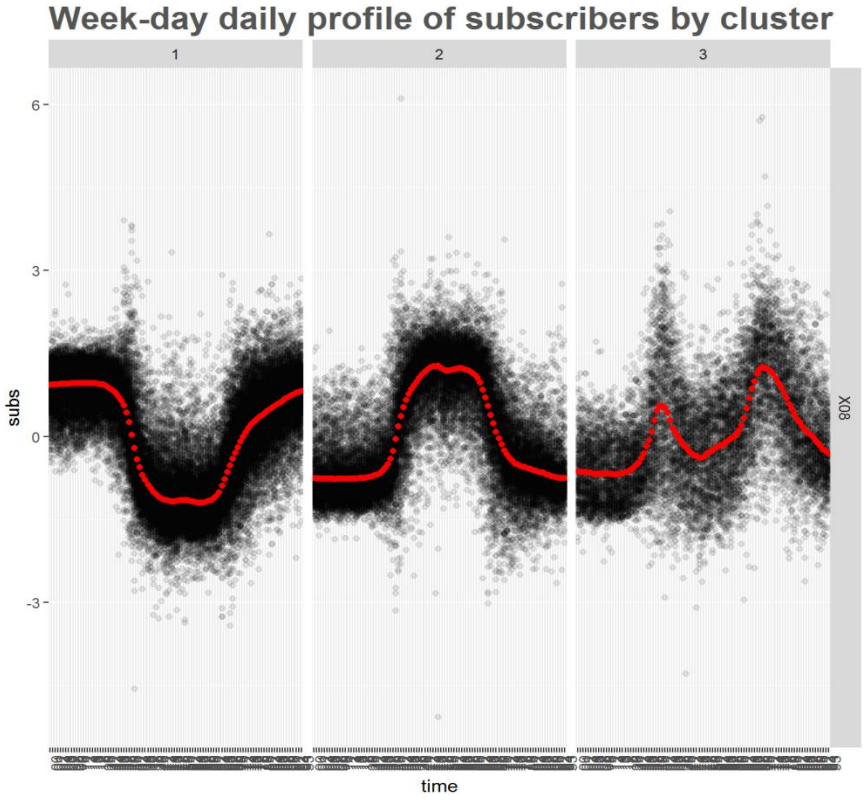


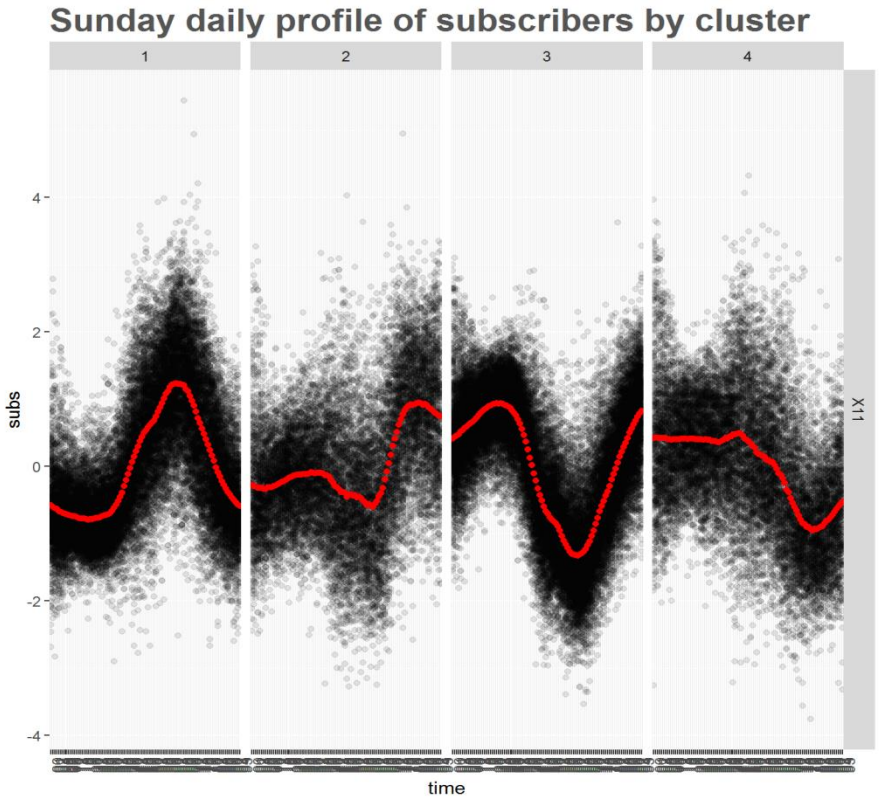
Mobile phone - clustering

- Typical patterns of population density?
- Normalised counts of mobile phones
- Residential, Work, Commuting

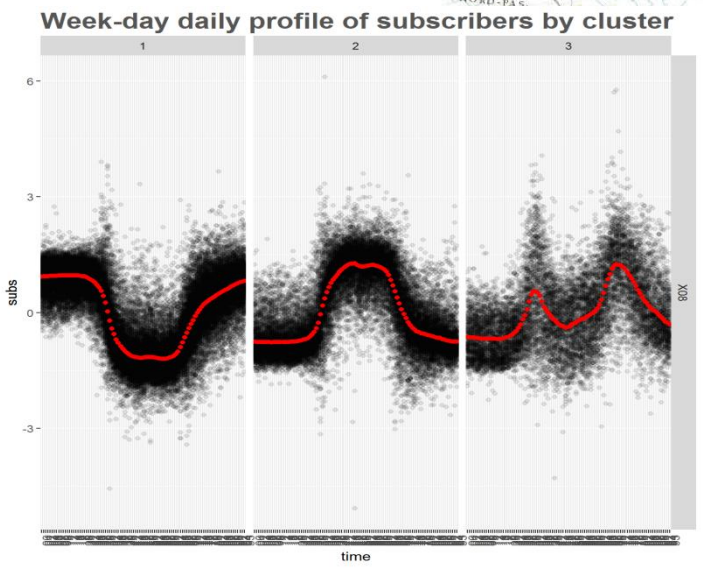


Mobile phone - clustering

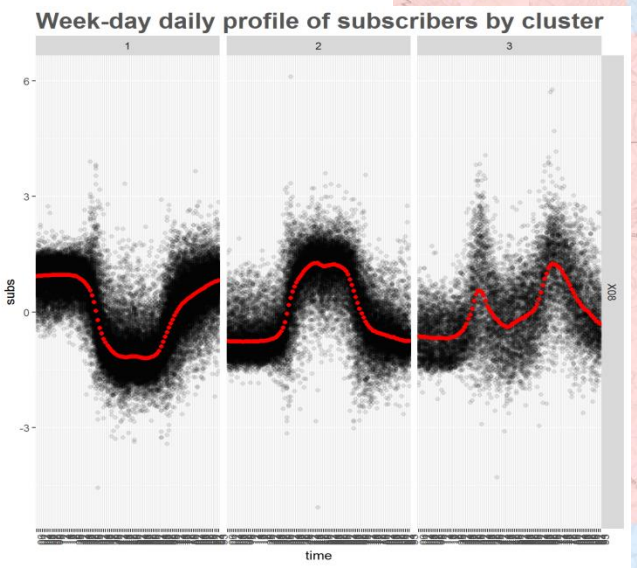


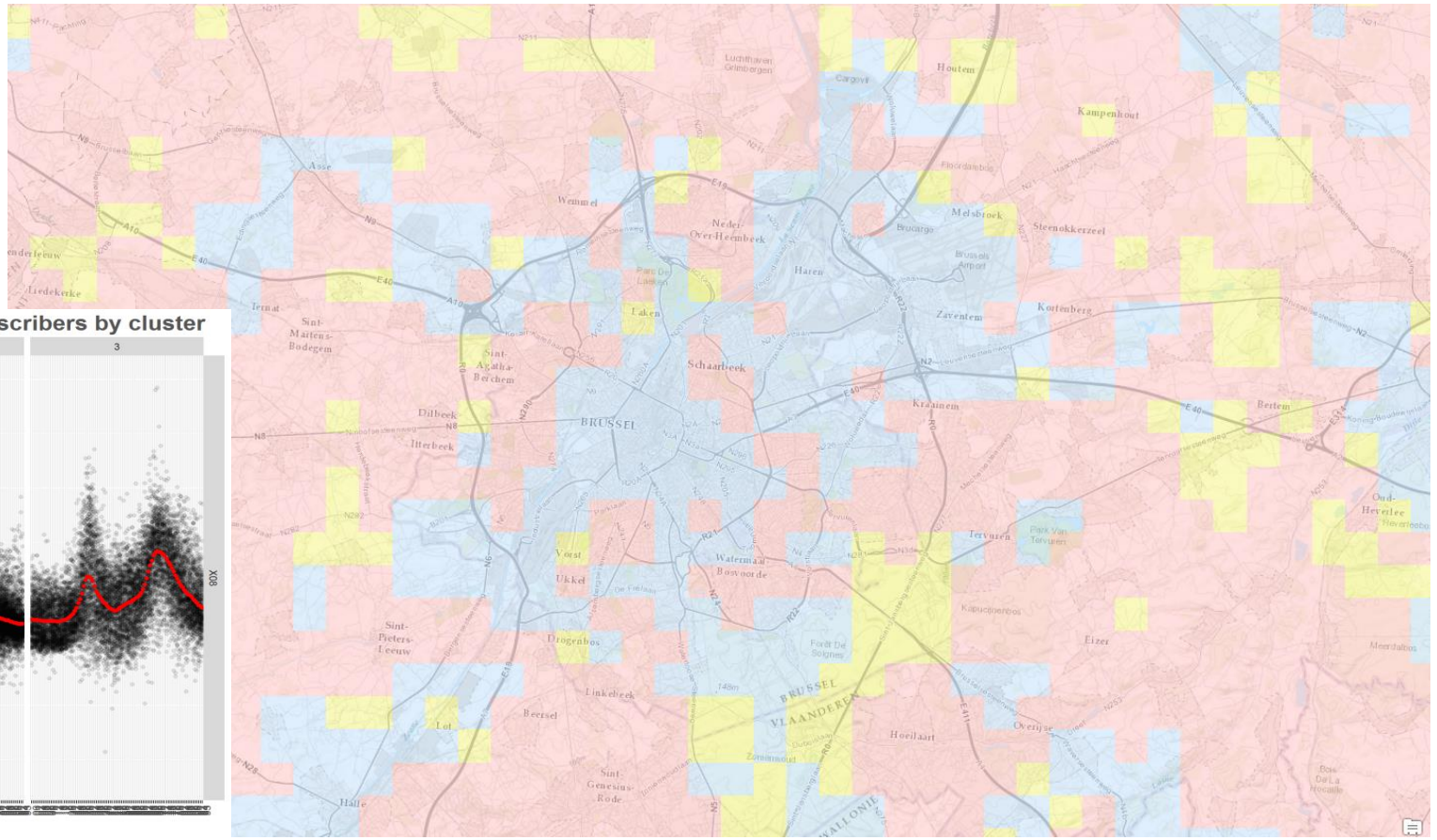


Mobile phone - clustering



Mobile phone - clustering





Mobile phone pilot – continuation

- Continuation of exploration
- Add additional data
  - **Traffic network**
  - **Land use data**
- Mobile phone pilot within ESSnet
- Extension to other countries and operators
- Work on methodology, quality, …
- Use of mobile phone data as complementary information

European Commission

Thank you for your Attention!