Datafication

Digital footprint

Sensors
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?
Big Data and Statistics

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

Communication
- Mobile Communication
- Social Media

WWW
- Web Searches
  - Analysis of websites' contents
    - Job Advertisements
    - Businesses' Websites
    - E-Commerce
    - Real estate
  - Internet Traffic
- Traffic loops
- Smart meters
- Automatic Vessel Identification System
- Satellite Images
- Webcams

Sensors
- Reservation Systems
  - Flight Booking transactions
  - Trains
  - Hotels
- Supermarket Cashier Data
- Loyalty Cards
- Financial transactions
- Mobile Payments
- eGovernment

Process generated data

Crowd sourcing
- Voluntary Geographic Information websites (OpenStreetMap)
- Voluntary Information websites (Wikipedia)
- Community pictures collection
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
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 (tessellation)
  - 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

- Tessellation
- > 10,000 cells
- Combining 2,3,4 G cells
Mobile phone – radio cells

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>295 ha</td>
</tr>
<tr>
<td>Median</td>
<td>166 ha</td>
</tr>
<tr>
<td>Maximum</td>
<td>5231 ha</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.2 ha</td>
</tr>
</tbody>
</table>
Mobile phone – population

Census 2011  Mobile phones
Mobile phone – population

\[ y = 3.6579x \]
\[ R^2 = 0.9526 \]
Mobile phone - population
Mobile phone - clustering

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

Week-day daily profile of subscribers by cluster

Sunday daily profile of subscribers by cluster
Mobile phone - clustering

Week-day daily profile of subscribers by cluster
Mobile phone - clustering

Week-day daily profile of subscribers by cluster

Eurostat
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
Thank you for your Attention!