Item 9
Big data: opportunities and challenges

9.1. Lessons drawn by the Task Force Big Data

Christophe Demunter
European Commission
DG EUROSTAT, Unit G-3 ‘Business cycle, tourism and registers’

Outline of the presentation

- Big data activities & set-up at Eurostat: quick overview
- Useful sources for tourism statistics
- Main risks
- Example on mobile phone data
- Some conclusions
Big data – Eurostat activities

Set-up: Task Force Big Data
- Internal Task Force: some full-time staff, some part-time staff 'seconded' from other areas of work (GISCO, methodology, transport statistics, tourism statistics, etc...)
- ESS Task Force (European Statistical System): regular meetings and projects with big data experts from the countries

Ongoing work on many dimensions of big data
- Partnerships, legal and ethical issues, infrastructure, skills, ...
- Methodological studies (e.g. selectivity bias in big data sources)
- Training program on big data (for statisticians in Europe)
- Data, access, pilots, ...

Pilot projects (2016-2018)
- Source-oriented pilots via an ESSnet project (cooperation network of national statistical offices)
- Some pilots are directly or indirectly relevant for tourism
  - Webscraping of enterprise characteristics
  - Smart meters for energy consumption (identification of seasonal/weekend holiday homes and measurement of their occupancy?)
  - Mobile phone data
  - ...

International dimension
- Cooperation with UNSD: Eurostat involvement in the UN Global Working Group on big data
Big data – main sources

Communication
- Mobile phone data
- Social Media

WWW
- Web Searches
- Businesses’ Websites
- e-commerce websites
- Job advertisements
- Real estate websites

Sensors
- Traffic loops
- Smart meters
- Vessel Identification
- Satellite Images

Process generated data
- Flight Booking transactions
- Supermarket Cashier Data
- Financial transactions

Crowd sourcing
- VGI websites (OpenStreetMap)
- Community pictures collection

Tourism statistics is in the frontline of big data projects!

Plenty of opportunities... but also many risks!

Big data sources come and go
- MySpace ⇔ Facebook ⇔ what’s next?
- SMS ⇔ WhatsApp & FacebookMessenger ⇔ what’s next?
- Huge impact on continuity of data ... vs. stability/continuity as one of the unique selling propositions of official statistics

From 'full control' to being a 'user'
- Big data often held by private companies
- Need for partnerships & business model: development/pilot phase vs. long-term collaboration
- Quality control?
Plenty of opportunities... but also many risks!

**Volume, velocity, volatility**
- "Can't see the wood for the trees"
- Main challenge: keeping the overview and synthesize into a manageable, coherent & sustainable production system

**Concepts and definitions**
- Data sources were not conceived for making statistics
- Big data can cover only part of the user needs (and not fully replace existing methods, at least not in the short/mid term)

**Perception in the public opinion**
- "Efficient modern organisation" or "big brother"?

---

**Where are we now?**

**Gartner Hype Cycle**

![Gartner Hype Cycle Diagram](image)

- **Technology Trigger**
- **Trough of Disillusionment**
- **Slope of Enlightenment**
- **Plateau of Productivity**
- **Peak of Inflated Expectations**

**Visibility**

**Time**
Example: mobile phone data for Belgium

Partnership between mobile network operator and statistical offices
- Eurostat, Statistics Belgium, Proximus
- Explore partnerships & business models; cooperate on concrete pilot projects

Work-in-progress, some preliminary results
- Focus on outbound trips made by residents of Belgium
- Comparison of estimates based on official statistics and estimates based on mobile phone data (April – September 2015/2016)
- Somehow promising (the data makes sense) but lots of open issues

i. Outbound trips by duration: comparison

Comparison of the distribution of outbound trips from Belgium to other EU countries, by duration of the trips
ii. Volume of trips and nights: comparison

Comparison of estimated number of outbound trips, by destination

- Observations: Big differences between the sources, but of a systematic nature.
- Understanding (and solving...) the deviations: differences in scope, parameter setting, selectivity bias, recall bias and non-response in surveys.
- The project continues...

First lessons learnt

- Positive & fruitful experience with the partnership
  - Joining forces (statisticians, data holders, data scientists)
  - Search for a win-win

- Promising results, but lots of homework
  - Mobile phone data clearly captures tourism concepts/definitions
  - Currently: satisfactory for trends, not for estimating volumes
  - How to make the series/sources converge to the unknown true values?
  - Extension to domestic tourism, to same-day visits

- Further research to be encouraged (other countries?)
Some conclusions on using big data

Target: regular data production using big data

⇋ first level: exploring the feasibility of big data sources as auxiliary information for quality checks, calibration, etc.
⇋ second level: using big data to fill current data gaps and to produce experimental 'flash estimates'
⇋ third level: using big data to partially replace 'traditional' sources/surveys
⇋ fourth level: rethinking in a user-oriented way the system of tourism statistics, taking into account the opportunities offered by integrating big data

Essential conditions:

⇋ intensive cooperation (at all levels)
⇋ knowledge sharing

Thank you for your attention!

Eurostat data & publications are available free of charge from the Eurostat website

Just released: tourism statistics leaflet
Order your copies from the EU Bookshop (free)