How "big data" are transforming official statistics

Challenges and opportunities for innovation in official statistics

konstantinos.giannakouris@ec.europa.eu

European Commission Eurostat/B1

Methodology; Innovation in official statistics
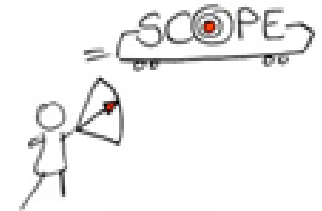
Outline

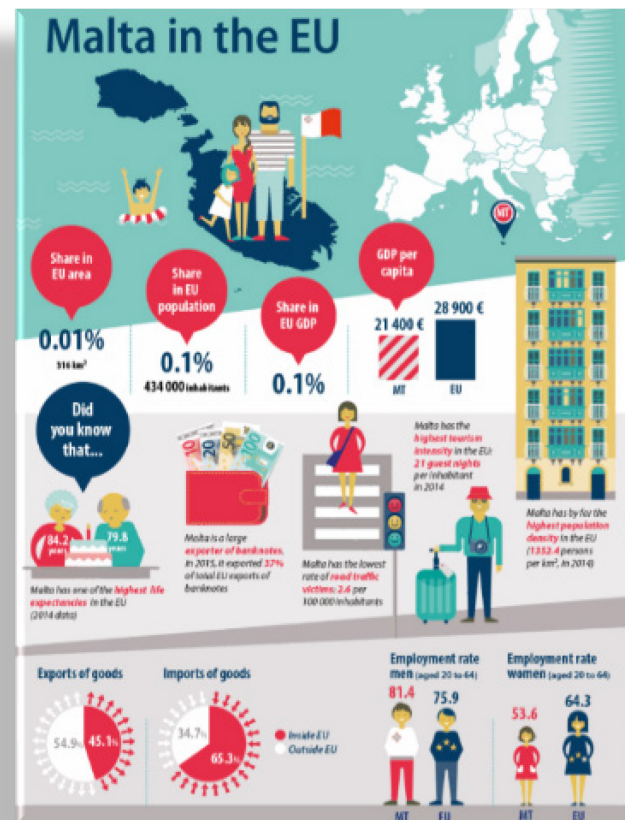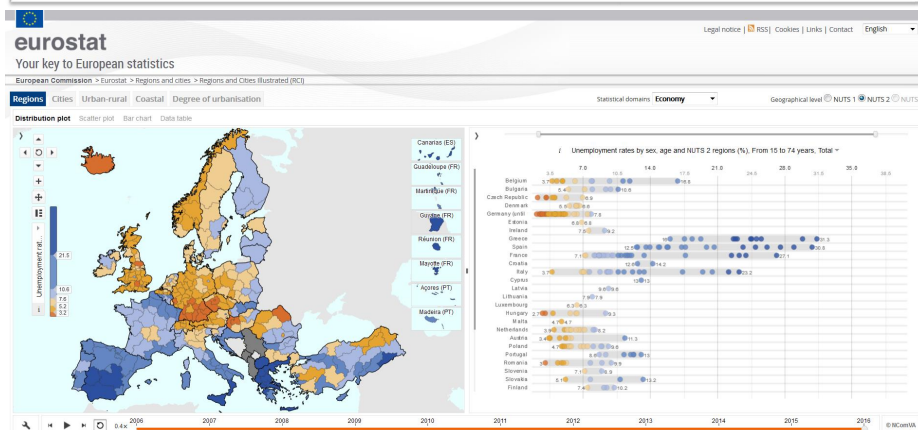*How "big data"* [*] *are transforming official statistics*
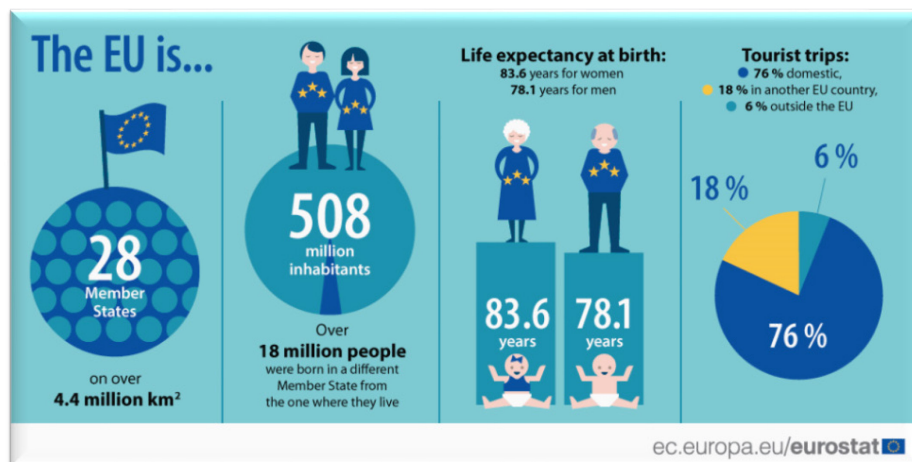
⇒ Eurostat in a nutshell

• The changing landscape for data and statistics

• Lessons learned using "big data" within the European Statistical system

[*] 😉 high dimensional exhaust data left from the use of IT systems or captured by sensors - sensors in a wide context (e.g. satellite sensors, cameras, …)

eurostat

Statistical office of the EU within the European Commission

The European Statistical System

Partnership between Eurostat and the national statistical authorities

Defined in European Statistics Regulation 223/2009

Member States collect and process the national data → Eurostat processes the received data and disseminates European official statistics

eurostat

Producing high-quality statistics is <u>not</u> BUSINESS AS USUAL

In a world of fast-food information and data deluge

we should be trustworthy and our statistics should be of quality and value
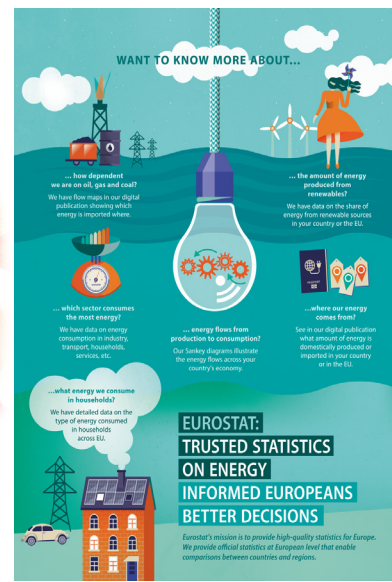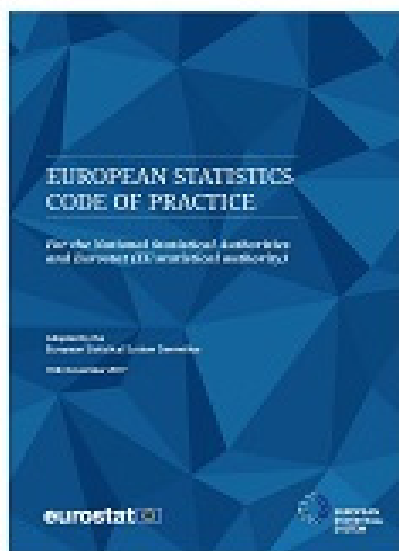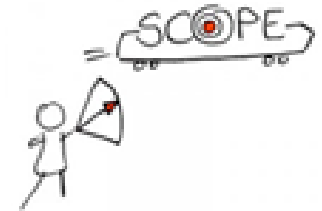
Outline

How "big data" [*] *are transforming official statistics*

- Eurostat in a nutshell

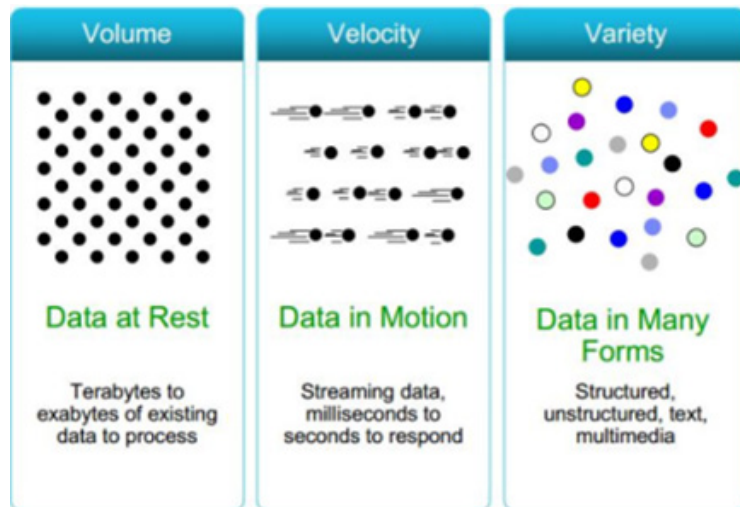⇨ The changing landscape for data and statistics

- Lessons learned using "big data" within the European

  Statistical system

[*] ☺ high dimensional exhaust data left from the use of IT systems or captured by sensors - sensors in a wide context (e.g. satellite sensors, cameras, …)

Official statistics is not a monopoly for statistics anymore

Revolutions in science have often been preceded by revolutions in measurement

Prof. Sinan Aral

| Volume	Velocity	Variety
Data at Rest	Data in Motion	Data in Many Forms
Terabytes to exabytes of existing data to process	Streaming data, milliseconds to seconds to respond	Structured, unstructured, text, multimedia

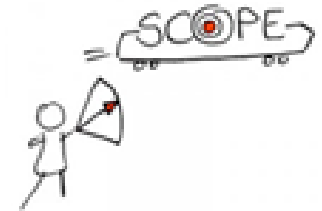*"I'm right there in the room, and no one even acknowledges me."*

eurostat

Official Statistics based on **survey data** and **administrative data** and **new digital data**
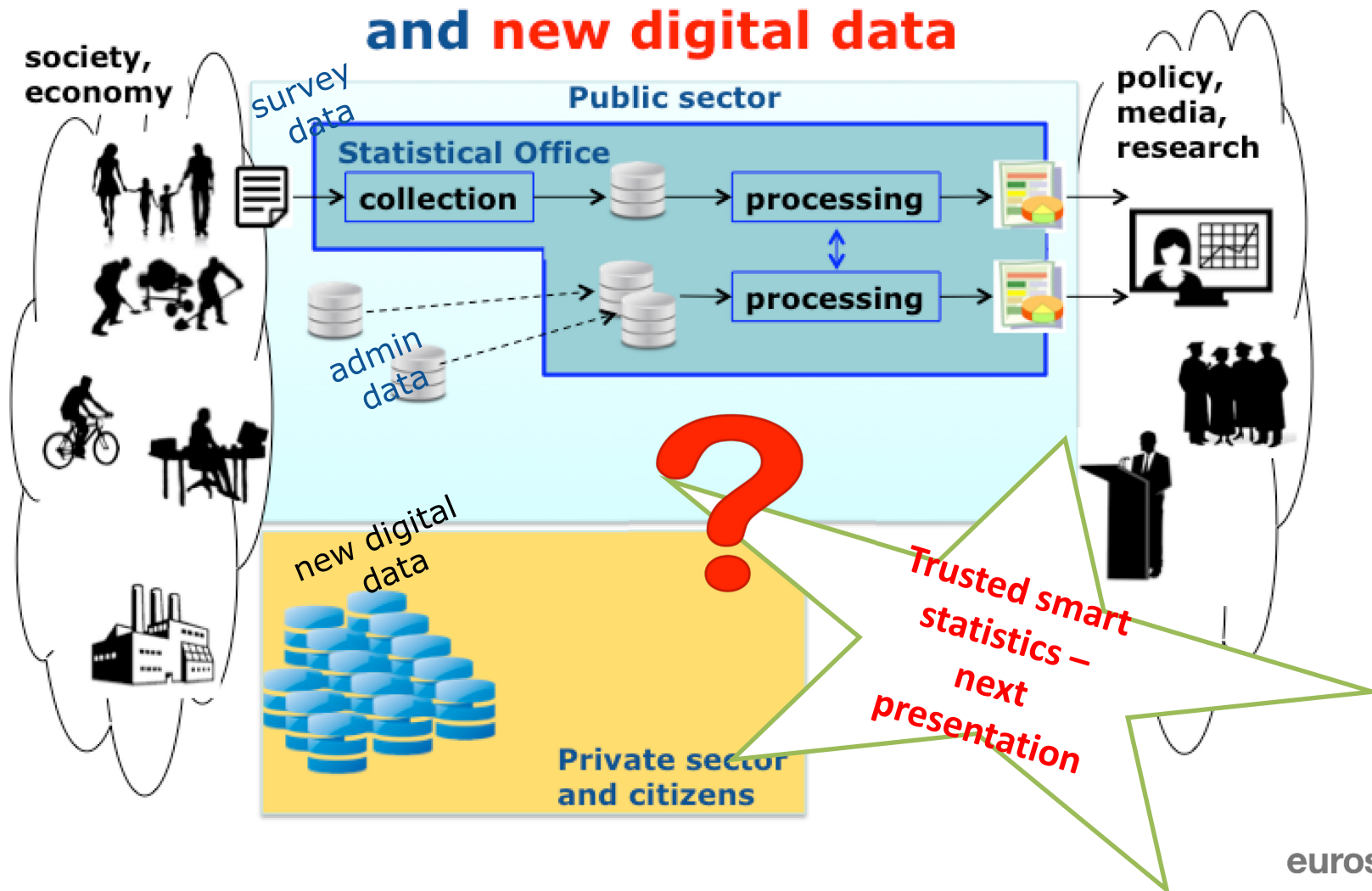
London Transport workers sorting 4 million used #London Underground tickets to identify most and least popular routes in 1939.
Photograph by Gerry Cranham/Fox Photos/Hulton Archive/Getty Images

Outline

How "big data" [*] *are transforming official statistics*

- Eurostat in a nutshell

- The changing landscape for data and statistics

⇨ Lessons learned using "big data" within the European Statistical system

[*] ☺ high dimensional exhaust data left from the use of IT systems or captured by sensors - sensors in a wide context (e.g. satellite sensors, cameras, …)

Scheveningen Memorandum (2013)
Big Data Action Plan and Roadmap @ a glance

Governance		
Policy	Quality	Skills
Experience sharing	Legislation	IT Infrastructures
Methods	Ethics / Communication	Big data sources
Pilots		

Webscraping business' websites

Using data from tracking ships

Webscraping ads of online job vacancies

Using data from smart electricity meters

Using road traffic data as an indicator of GDP growth

Using satellite images for crops identification

Using mobile phone network data

Achievements and lessons learned
Scoping new data types/sources

We identified the **new data sources** created by technological innovation that capture **additional dimensions** of phenomena.
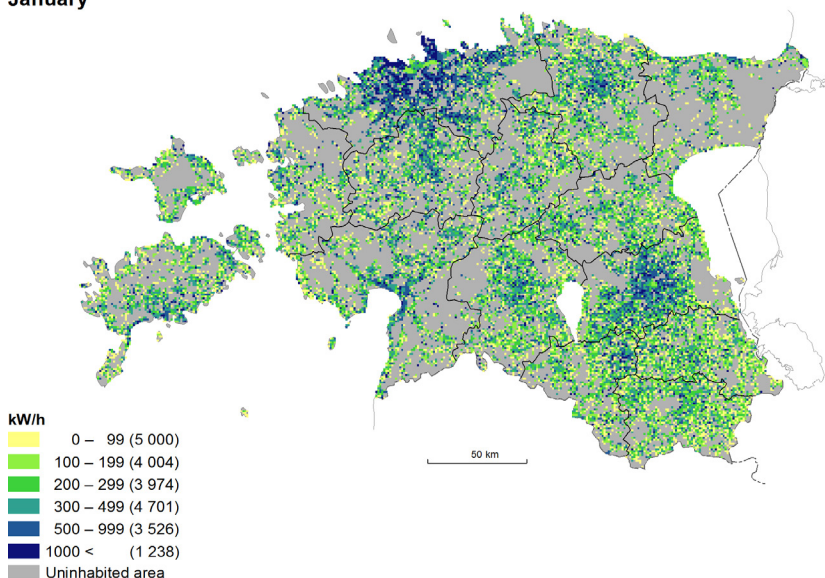
e.g. ships' tracking system, smart meters, traffic loops, smartphones, …

Suez canal, coming from China

Leaflet | Tiles © Esri — Source: US National Park Service

eurostat

Achievements and lessons learned
Complement existing statistics, produce new statistics

January

kW/h
- 0 – 99 (5 000)
- 100 – 199 (4 004)
- 200 – 299 (3 974)
- 300 – 499 (4 701)
- 500 – 999 (3 526)
- 1000 <   (1 238)
- Uninhabited area

50 km

- Create <u>new insights</u>

- Improve <u>relevance</u> (detect trends to be monitored)

- Increase <u>granularity</u> of relevant statistics

- Improve <u>timeliness</u> (decrease time lag between collection, analysis and publication)

Early investigation of the necessary *enablers for using privately held data,* do not linger over the sources, move ahead for PoC and prototypes for *producing statistics*.

eurostat
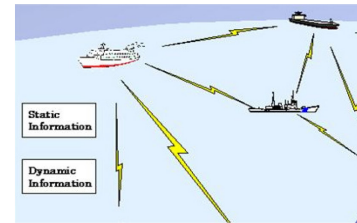
Achievements and lessons learned
Project → *Process: Mind the gap* ☺

- The path of transforming proofs-of-concept or pilot "projects" on big data sources to deploying new production "processes" that will produce statistics involves:

  - **the effective use of new infrastructures,**

  - **developments of new statistical methodologies,**

  - **considerations on the quality of the output, and**

  - **most importantly multi-disciplinary work and development of new skills.**

eurostat

Achievements and lessons learned
Steps towards demystifying big data

- Create a **visible statistical community** in the new data ecosystem; dedicated "big data" community of statisticians in official statistics (ESS)

- Involve at an early stage subject matter experts, and experts from other disciplines

- Acknowledge the potential of **sharing infrastructure** and **big data** as well as **experience** and **skills**

- Develop collectively statistical **methodology** and **quality** dimensions

eurostat

Achievements and lessons learned
## *New processing opportunities – **new skills***

- Technological innovation creates new processing opportunities for existing data and a demand for **new skills**

  - Example: Use of automatic text interpretation, cognitive image processing, deep learning, AI → transforming documents, images, videos, text messages etc. into mineable sources
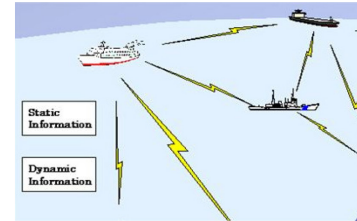

Precision of predicting presence of web shop in website
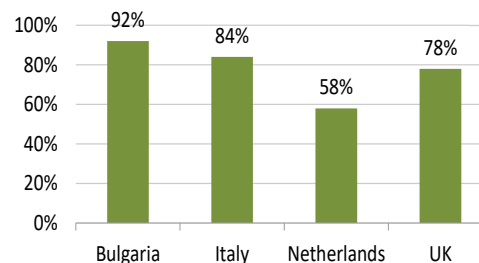
Percentage of enterprises with e-commerce

■ Estimate from web data ■ Survey Estimate
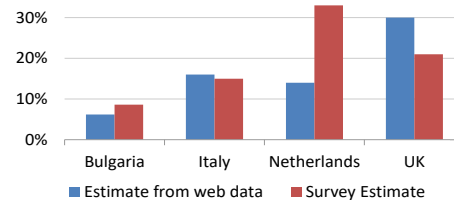
Achievements and lessons learned

Quality assessment in the case of using big data

- No uniformity in data types/data sources

- Big data are not *a priori* administrative data

- No single statistical process pairing « data source » and « statistical application »

- Quality assessment and reporting should be embedded in big data projects

Achievements and lessons learned
Stakeholders – Communication

- Understand needs and objectives of stakeholders – start dialogue

- Investigate enablers for using data held by third parties (legislation, partnership agreements, …)

- Establish partnerships between NSAs and data holders
  - **Within ESS, led by Eurostat**
  - **At national level supported by EU and UN activities**

- More partnerships with academic communities are necessary

eurostat

Conclusions

Same role but enhanced functions (Eurostat & NSAs)

- The mission of official statistics still holds, but the means and the conditions in which this mission is achieved are changing

- This requires a systemic approach where statistical offices will have to rethink their business model (development, production, dissemination)

- ESS is the vocal advocate of quality, providing guidance on quality of data used for public interest

- Sustain and leverage on the reputation of NSAs as regards quality and trustworthiness
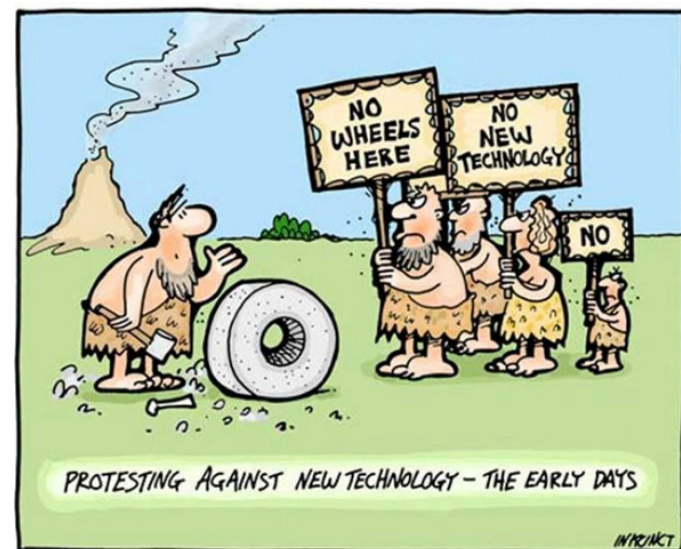
The main difference from traditional official statistics production and the one of the future will be one of <u>culture</u>!

Thank you for your attention

Konstantinos Giannakouris

Eurostat, B1: Methodology; innovation in official statistics

PROTESTING AGAINST NEW TECHNOLOGY - THE EARLY DAYS

Information: Collaboration in Research and Methodology for Official Statistics
European Commission » Eurostat » CROS » Big data » Big Data Initiatives
https://ec.europa.eu/eurostat/cros/content/big-data_en

konstantinos.giannakouris@ec.europa.eu

https://linkedin.com/ Konstantinos Giannakouris

Trusted Smart Statistics: Motivations & principles

eurostat