EUROSTAT and BIG DATA

High Level Seminar on integrating non-traditional data sources in the National Statistical Systems

Santiago, Chile, October 1-2, 2018

Table of contents

What is Eurostat?

The statistical office of the European Union

Mission: to provide high-quality statistics for Europe

A Directorate-General (DG) of the European Commission

Member of the ESS

Responsibilities of Eurostat

- Eurostat defines harmonised methodologies in collaboration with Member States so that data are comparable
- Eurostat receives, treats and publishes data on the EU, euro area, Member States and EFTA
- Eurostat disseminates statistics

Basic facts

Based in Luxembourg

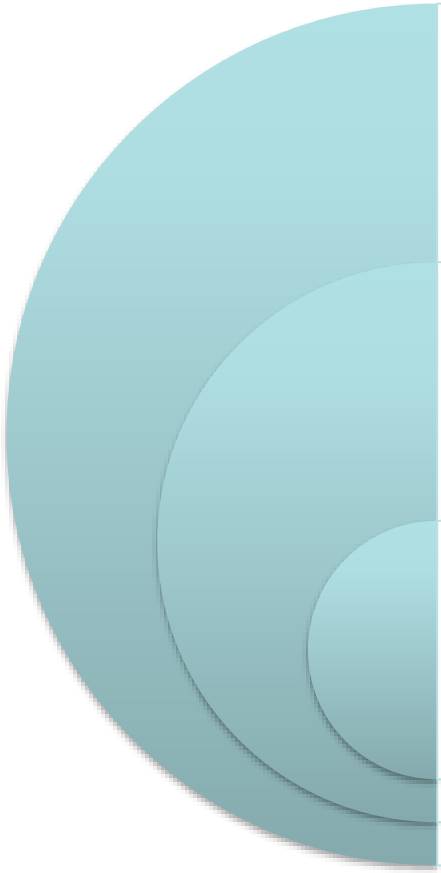
Some 800 staff – officials, national experts, contractual staff

Director-General Mariana Kotzeva

Created in 1953 as a service of the High Authority for Coal and Steel

A long evolution since then…

The European Statistical System (ESS)



A partnership of all National Statistical Institutes (NSIs), other national statistical authorities (ONAs) and Eurostat

Was built up gradually to provide comparable statistics at EU level

The NSIs are the coordinators of the national statistical systems in Member States, including Other National Authorities

European Statistical System Committee - ESSC

Established in 2009

Provides "professional guidance to the ESS for developing, producing and disseminating European statistics"

Composed of Heads of Member States' NSIs, chaired by the Commission

(Director-General of Eurostat)

Meets four times a year

European Statistics Code of Practice

New and improved!

16 principles on institutional environment, production & dissemination

Among them:

- Professional independence
- Commitment to quality
- Impartiality and objectivity

Peer reviews to monitor implementation of the Code

ESGAB

Established in March 2008:

- An independent overview of the ESS as regards the implementation of the Code of Practice
- Opinions
- Annual report to the European Parliament and the Council

European Statistical Governance Advisory Board
ESGAB

ESAC

Established in March 2008

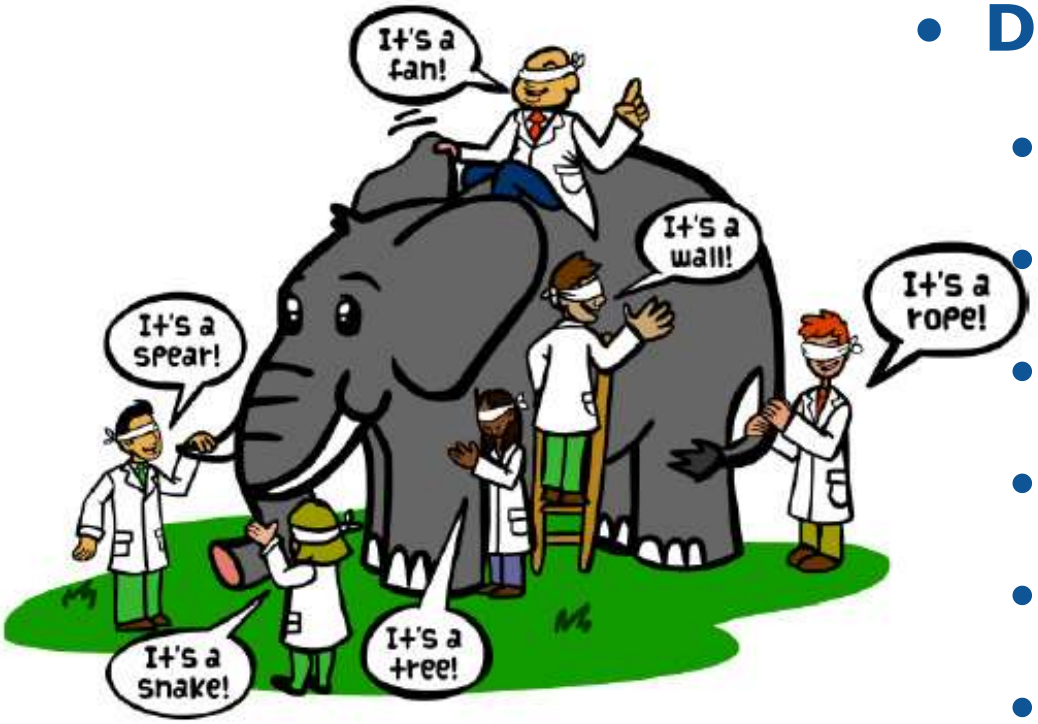
- 24 members representing users, respondents and others
- Makes sure that user requirements and response burden taken into account when making up statistical programmes.
- Gives view on balance (priorities and resources) of different areas of multi-annual statistical programme.

Big Data

Can we use Big Data for official statistics?

Without data, you're just another person with an opinion.
W. Edwards Deming

Describing big data



- **Data deluge**
  - High **V**olume
  - High **V**elocity
  - High **V**ariety …
  - (**V**ariability)
  - **V**eracity
  - (**V**isualisation)
  - **V**alue (…)

http://www.theblindelephant.com/the_blind_elephant_fable.html

Proclamation of pope Benedict XVI

2005

Source: REUTERS/Kimimasa Mayama CRB/DY
http://pictures.reuters.com/archive/POPE-RP6DRMRPZFAB.html

Proclamation of pope Francis

2013

Source: Associated Press

The data deluge (1/3)

The Economist March, 2010"The Data Deluge"
Source: http://www.economist.com/node/15579717

1. **Human-sourced information**
   - Social Networks (Facebook, Twitter, LinkedIn, Pinterest, …)
   - Blogs and posted comments
   - Pictures (Instagram, Facebook…)
   - Videos (Youtube, …)
   - Search engine queries
   - Mobile data content (SMS, …)
   - User-generated maps
   - E-Mails
   - **…**

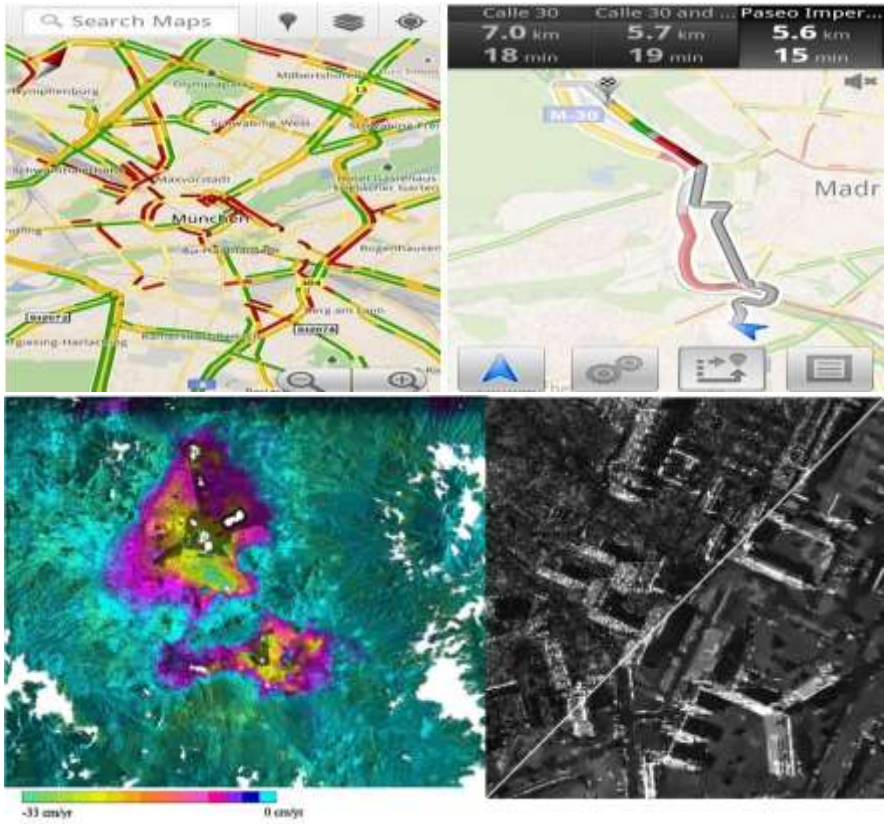
https://statswiki.unece.org/display/bigdata/Classification+of+Types+of+Big+Data

The data deluge (2/3)

2. **Business systems (process-mediated data/transactions)**
   - Commercial transactions
   - Banking/stock prices records
   - E-commerce
   - Telephone Call Detail Records
   - Credit cards
   - Medical records from Public Health
   - …

https://statswiki.unece.org/display/bigdata/Classification+of+Types+of+Big+Data

The data deluge (3/3)



3. Machine generated data (Internet of Things - IoT)

- **Sensor data**
  - ✓ Weather/pollution sensors
  - ✓ Traffic sensors/webcam
  - ✓ Security/surveillance videos/images
  - ✓ …
- **Tracking devices**
  - ✓ GPS systems
  - ✓ Mobile phone location
  - ✓ Satellite images
  - ✓ …
- **Data from computer systems**
  - ✓ Logs & Web logs
  - ✓ …

https://statswiki.unece.org/display/bigdata/Classification+of+Types+of+Big+Data

Implications of big data for official statistics

Data-driven economy

*Official statistics **is no longer** an "almost" statistical monopoly*

Scheveningen Memorandum on big-data and official statistics (**2013**), the general directors of the NSI

"*Acknowledge that the use of Big-data in the context of official statistics requires new developments in **methodology**, **quality assessment** and **IT related issues**. The European Statistical System should make a special effort to supports these developments*".

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

Implications of big data for official statistics

- **Change of paradigm**
  - **From:** finite population sampling methodology
  - **To:** additional statistical modelling and machine learning
  - **From:** designers of data collection processes
  - **To:** designers of statistical products
- **Privacy**
  - Use of digital footprint
  - Data subject has no control of data
  - High data detail and insight from analytics

Official statistics

Census-taking Relief ("Altar of Domitius Ahenobarbus"), Rome, Italy, ca. 100 B.C.E.,

Some things have not changed

- Mandatory census
- Census is official
- The census operation has not changed fundamentally for 2000 years
- It's all about registering and counting people

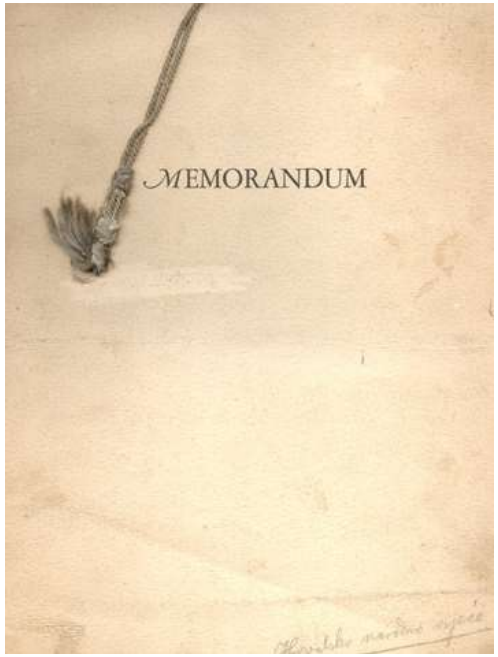
Some other have changed

- People had to go to the censor, now the censors go to the people
- Not everyone was counted, only Roman citizens, now everyone counts regardless social class, gender or age
- Census served the king, now it serves society and the democratic process

Some will soon change

- **Count people vs. re-use digital footprint**

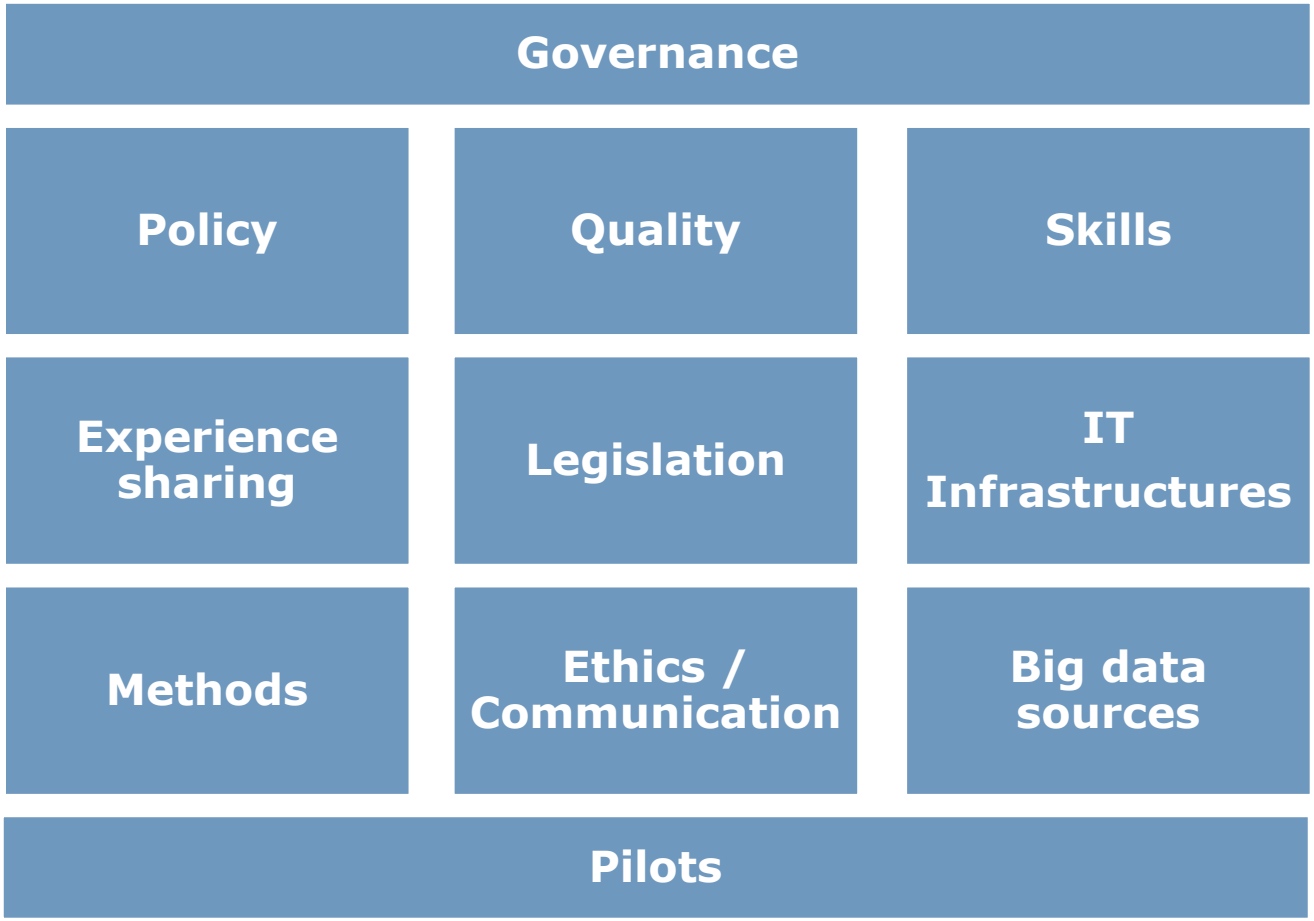
Scheveningen Memorandum on Big Data 09/2013

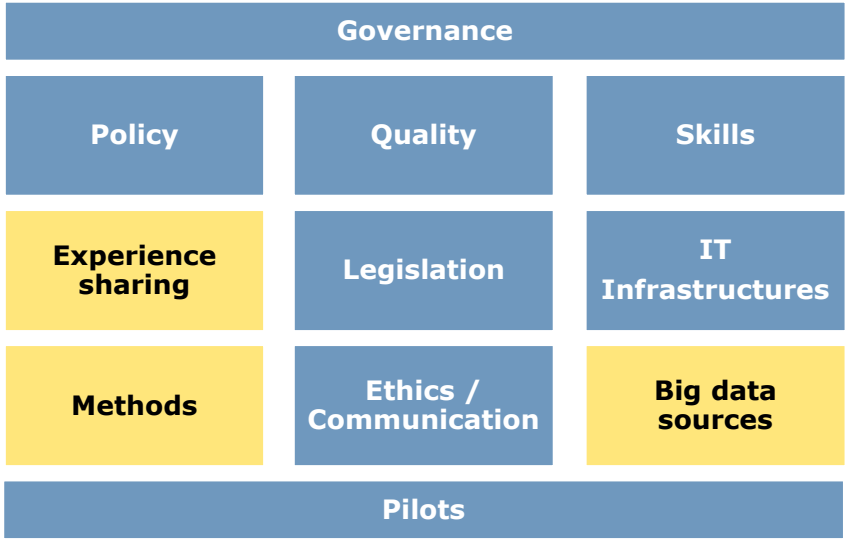


- Examine the **potential** of Big Data sources for official statistics
- Official Statistics Big Data **strategy** as part of wider government strategy
- Address **privacy** and **data protection**
- **Collaboration** at European and global level
- Address need for **skills**
- **Partnerships** between different stakeholders (government, academics, private sector)
- Developments in **Methodology**, **quality** assessment and **IT**
- Adopt **action plan and roadmap** for the European Statistical System

https://ec.europa.eu/eurostat/cros/news/scheveningen-memorandum-big-data-and-official-statistics-adopted-essc_en

ESS big data action plan



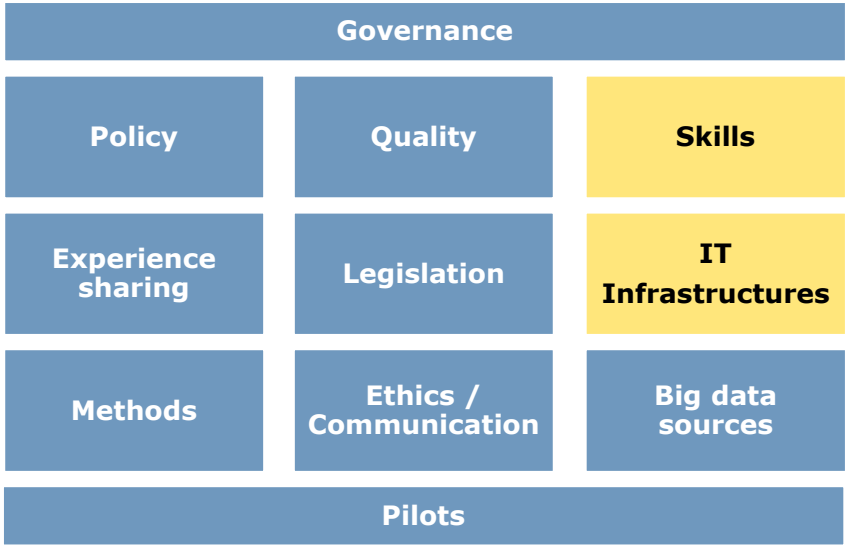


Challenges

- cooperation, sharing of know-how
- development of a sound methodology ("from design-based to model-based approach")
- exploration & tentative implementation
- Looking for partners

Action

- **Pilot projects, Member States** (ESSnet)
  - European Statistical System Network
    - Exploring different big data sources
    - Establish parternships with data providers, research and international organisations
  - Cooperation with UN on Methodological Framework

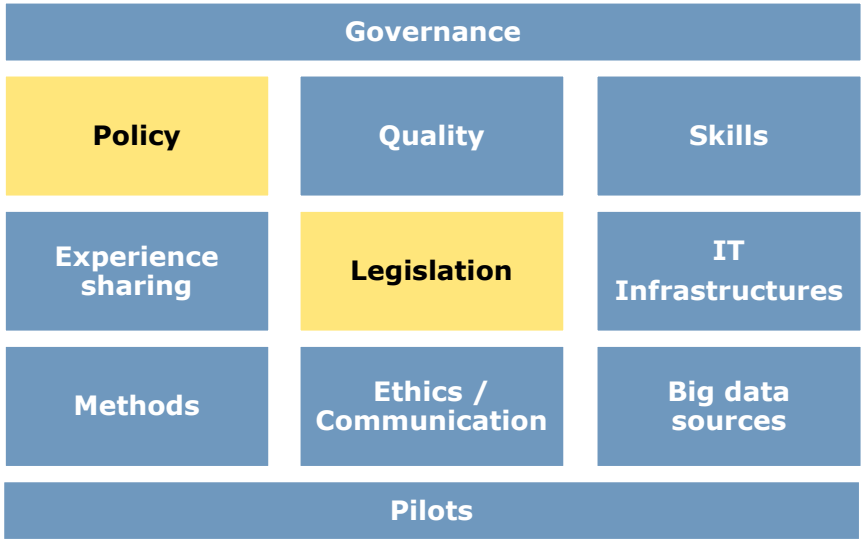


Challenges

- new skills for NSI staff: statisticians *vs.* data scientists ?
- computing capacity, hardware ?
- analytical tools, software?
- storage?

Action

- **Training program for European statisticians** (ESTP)
  - Dedicated courses on big data
  - Focus on big data sources & on big data tools
  - Acquiring the skills needed to assess sources and their quality, the skills to use tools and to explore big data sources

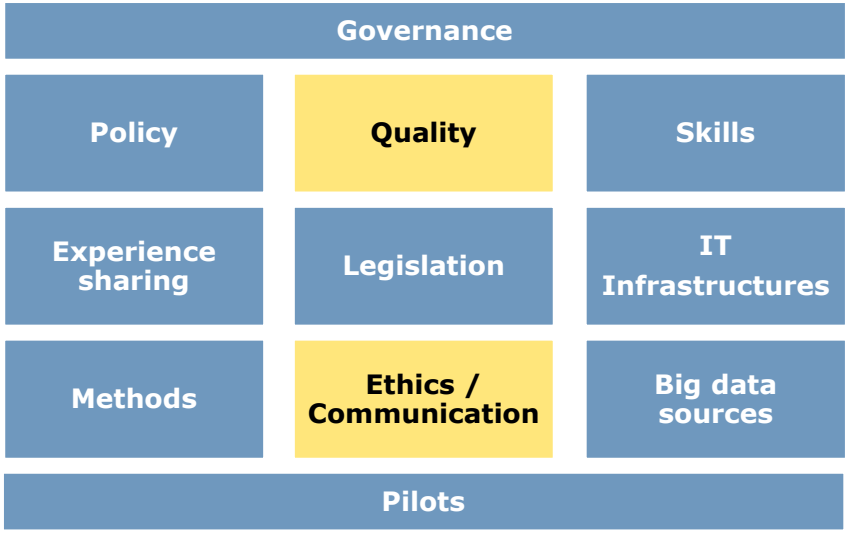


Challenges

- integrating official statistics into big data strategies
- getting access to data & continuity of access
- data security & privacy concerns
- **compensate for the burden** ?

Action

- **Project** on the analysis of legislation and strategy (but also ethics and communication)
  - 2015-2017
  - Analysis for EU and for Member States at national level



Challenges

- transversal challenges to all big data activities: **quality** and **ethics & communication**
- big data *vs.* statistics : "goodness of fit" (concepts, representativeness,…)
- impact on the public opinion of privacy and security concerns?

Action

- Cooperation with UN on a quality framework for big data
- Project on the analysis of ethics and communication (but also legislation and strategy)

ESSnet WP1 Webscraping / Job Vacancies

(UK, DE, EL, IT, SE, SI)

Aim:

"To demonstrate by concrete estimates which approaches (techniques, methodology, etc.) are most suitable to **produce statistical estimates** in the domain of **job vacancies** and under which conditions these approaches can be used in the ESS."

End date: May 2018

ESSnet WP2 Webscraping / Enterprise Characteristics

(IT, BG, NL, PL, SE, UK)

Aim:

To investigate whether webscraping, text mining and inference techniques can be used to collect, process and **improve general information about enterprises**: presence of web sales facilities , profiling information: type of activity, links with other enterprises, etc.

End date: May 2018

ESSnet WP3 Smart Meters (EE,AT,DK,SE)

Aim:

"To demonstrate by **concrete estimates** whether buildings equipped with smart meters can be used to **produce energy statistics** but can also be relevant as a supplement **for other statistics**, e.g. census housing statistics, household costs, impact on environment, statistics about energy production."

End date: May 2018

ESSnet WP4 Automated Identification System Data

(NL, DK, EL, NO, PL)

Aim:

"To investigate whether real-time measurement data of ship positions (via AIS system) can be used

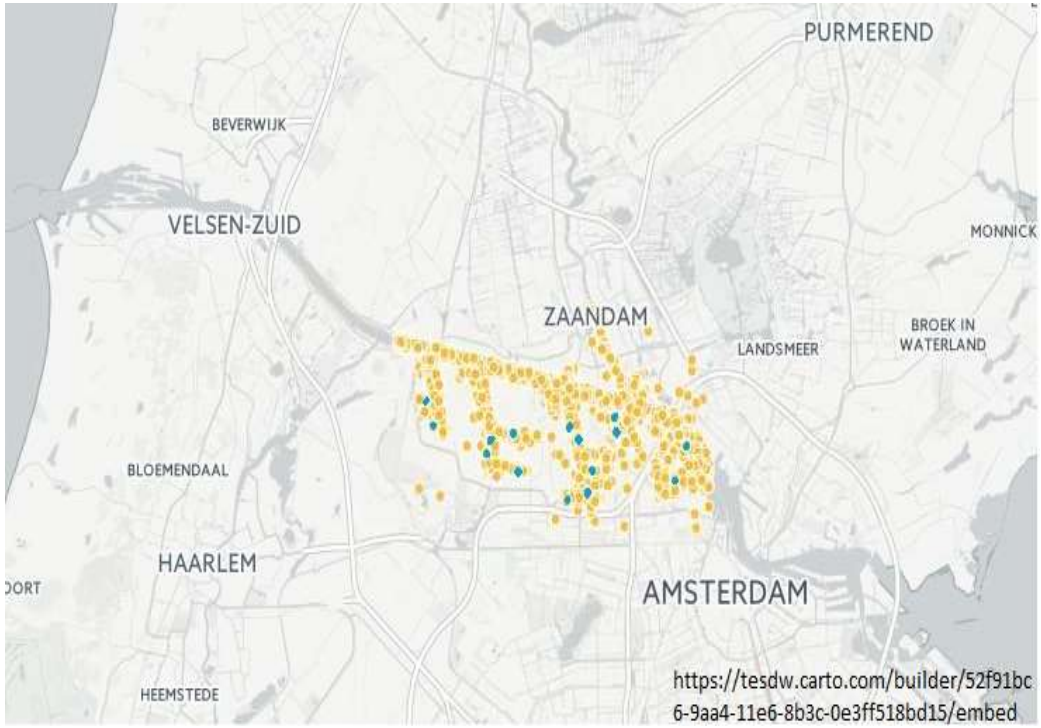
1. **to improve the quality and internal comparability of existing statistics and**
2. **for new statistical products relevant for the ESS."**

End date: May 2018

ESSnet: AIS data Port visits

Possible Advantages:

- Determining **ship routes**
- Improve existing statistics on **fuel consumption** and **emissions**.
- Reduce **respondent burden** for some ports
- Accelerate publishing speed for some maritime statistics
- Experimental: Now-cast economic time series



Moving from Internet of Things …

- A set of sensors, actuators, smart objects,
- Data communications and interface technologies that
  - allow information to be collected, tracked and processed across local and global network infrastructures,
  - enabling the future hyper-connected society

… to Smart statistics

Data capturing, processing and analysis will be embedded in the system itself

Intelligence along data life-cycle enhanced with cognitive processes

Trusted smart statistics

Yes, we can use big data in official statistics

if we approach it carefully.

Thank you for your attention

"Your recent Amazon purchases, Tweet score and location history makes you 23.5% welcome here."

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