Official Statistics in the Age of Big Data

SAS Forum BeLux 2014

Michail.Skaliotis@ec.europa.eu
Albrecht.Wirthmann@ec.europa.eu

Table of Contents / Storyboard

- What is "Official Statistics"?
- Drivers of Big Data
- Recent Development in Statistics and Big Data
- Prototypes
- Challenges and Myths around Big Data
- The statistical Office of the Future
- Concluding Remarks

Eurostat

- Part of the **European Commission** – Commissioner Marianne Thyssen
- The statistical office of the **European Union**
- The central institution of the **European Statistical System**

- Leading provider of high quality statistics on Europe
  - **Statistics at EU level enabling comparisons between countries and regions**

Eurostat and SAS

- > 100 SAS users in Eurostat
- SAS Enterprise Guide connected to a SAS server.

- SAS in some of the statistical production workflows
  - **GSAST (Generic Sas Tool)**
    a metadata driven SAS BI based production tool covering the whole data processing workflow

- Eurostat has presented the architecture and evolution of the GSAST several times at this SAS forums (in 2007 2009 and 2012).

European Commission

Official Statistics

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

What is the role of official statistics today?

*'….To provide an **indispensable element** in the **information system** of a democratic society, serving the government, the economy and the public with data about the economic, demographic, social and environmental situation….'*

*[Fundamental Principles of Official Statistics; principle 1 on **Relevance, impartiality and equal access**]*

What is the Role of official statistics today?

- Selection of data sources with regard to **quality**, costs and the burden on respondents
- Decision on methodology based on professional considerations including **scientific principles and professional ethics**
- Commitment to **confidentiality**
- **Exclusive use for statistical purposes**
- Presentation of information according to **scientific standards** on the sources, methods and procedures of the statistics

[Fundamental Principles of Official Statistics;
Principles 2,3,5,6]

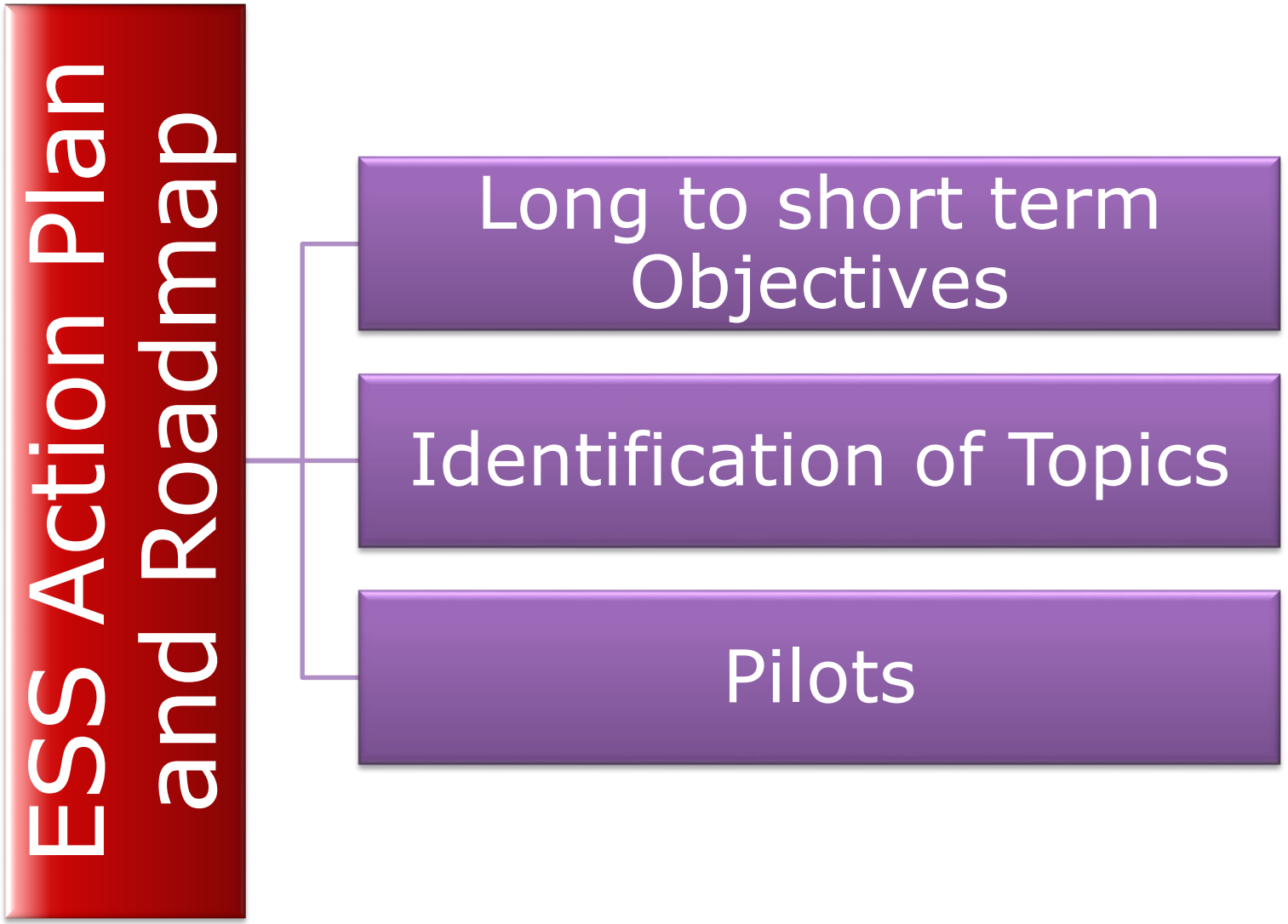

DRIVERS
SOCIAL MEDIA

Recent Developments

- Big Data Project of the UNECE
- Towards a thriving data driven economy
  - **Communication by the Commission**
- Scheveningen Memorandum on Big Data by European Statistical System
  - **Big Data Raodmap and action plan 1.0**
- Projects at Eurostat
- Projects by national statistical offices

European Statistical System

- Scheveningen Memorandum on Big Data
  - 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**


ESS Action Plan and Roadmap
Long to short term Objectives
Identification of Topics
Pilots

Roadmap

"As is" versus "To be"

> 2020	Long term Vision
By 2020	Medium term aims
By 2016	Short term objectives

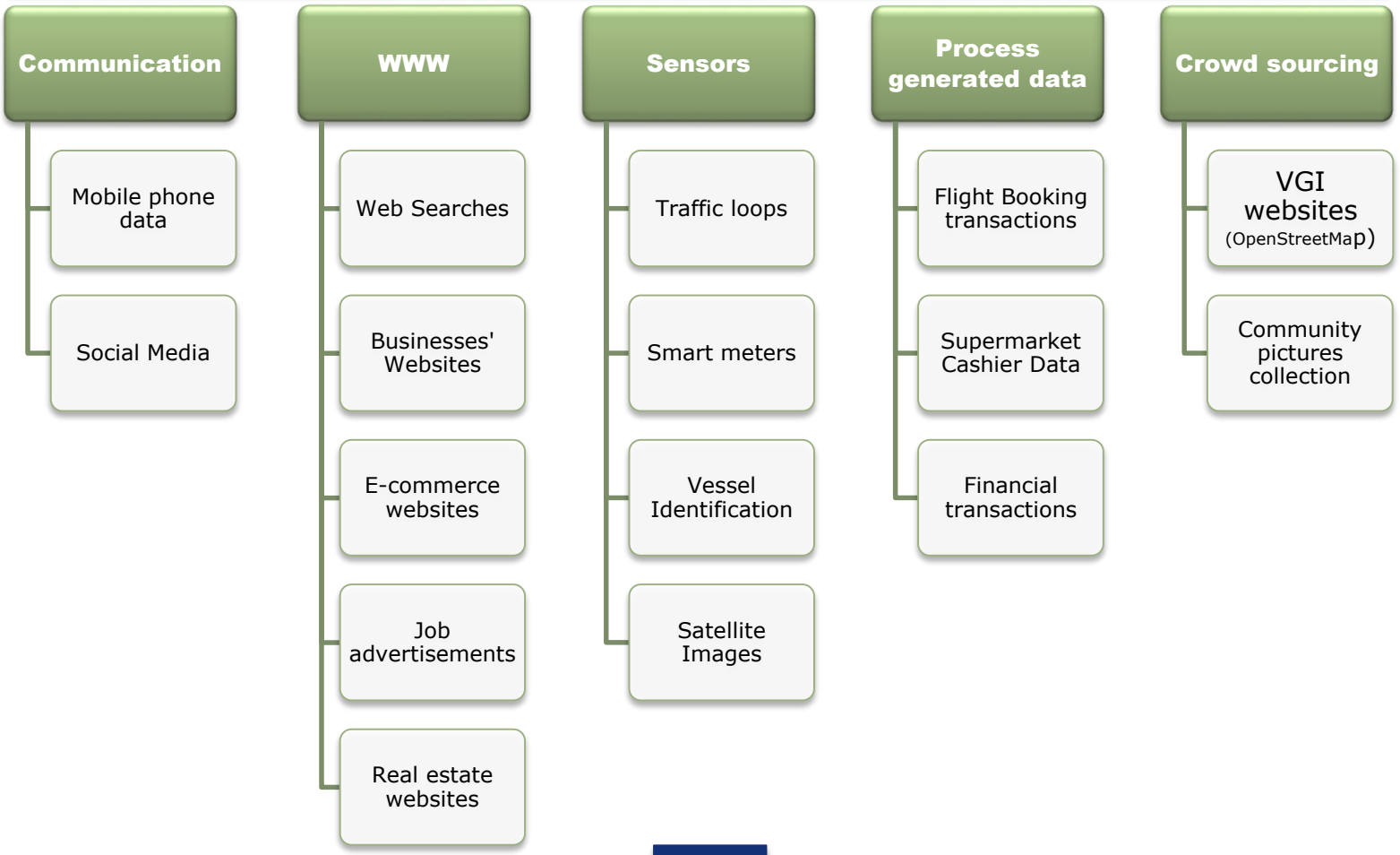
European Commission

TOPICS

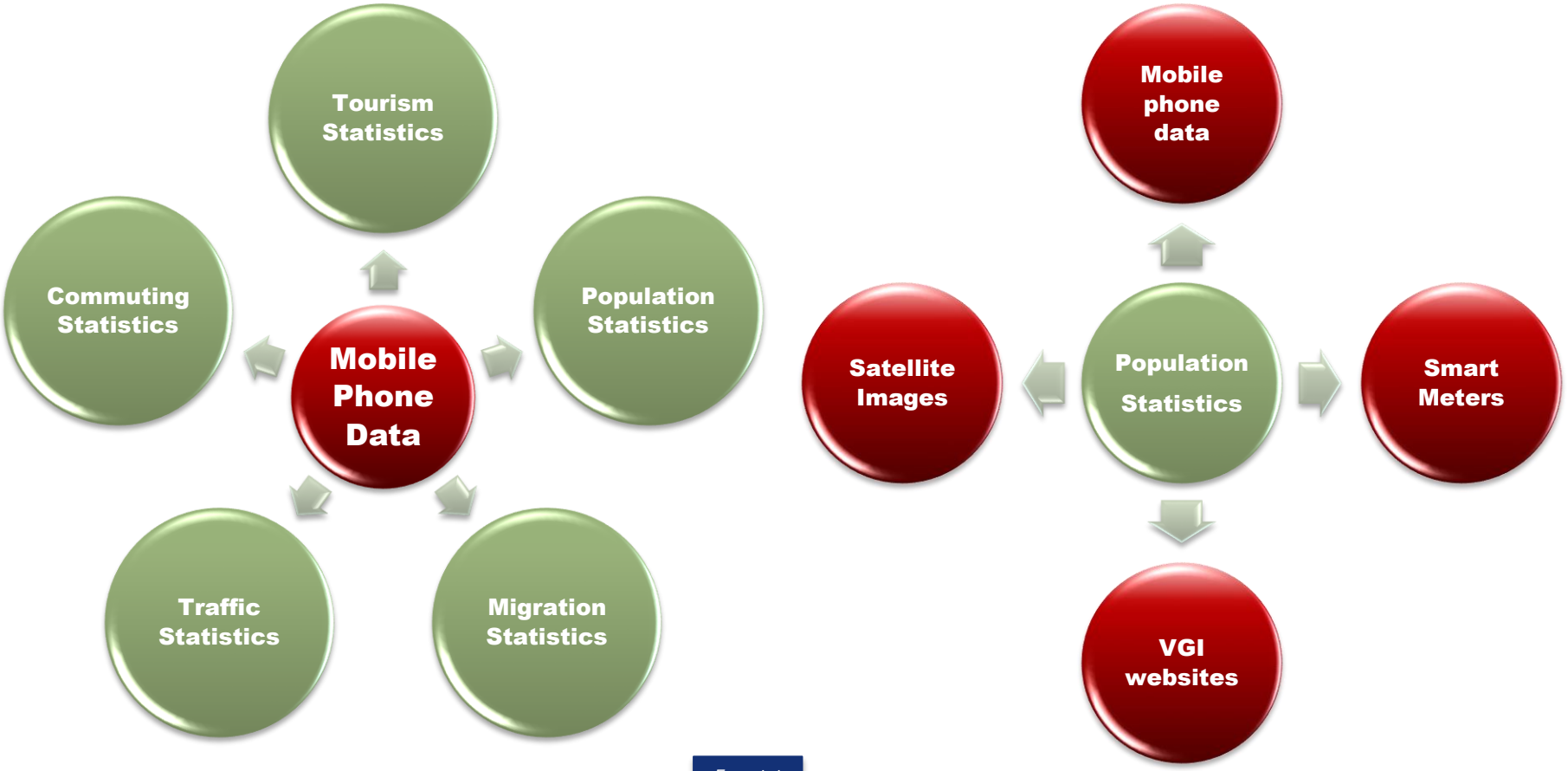
Policy	Quality	Skills
Experience sharing	Legislation	IT Infrastructures
Methods	Ethics / Communication	Pilots

European Commission

Sources of Big Data for Pilots



Blending of Sources and multipurpose Statistics



The Internet as Source (IaD) for Information Society Statistics

- Surveys of Households/Individuals and Enterprises
- Analysis of website functions of enterprises
- Statistics on use of Internet

European Commission

Use of Mobile Phone Data for Tourism Statistics

Eurostat

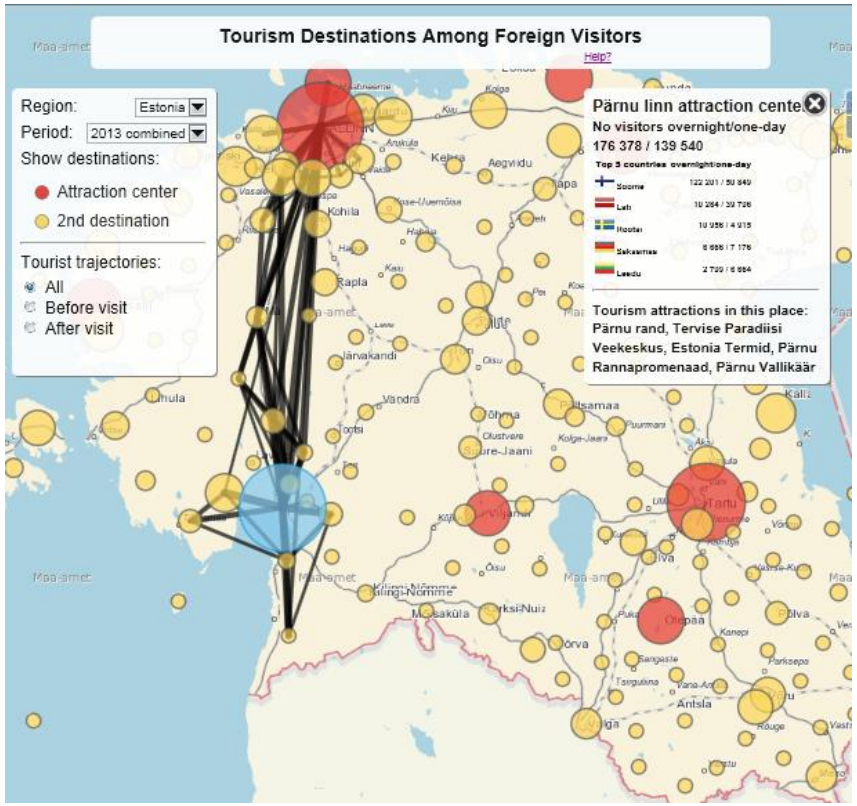
European Commission

Use of Mobile Phone Data for Tourism Statistics

Eurostat

Use of Mobile Phone Data for Tourism Statistics (…)

- Precise and comprehensive data source
- Cost savings
- Description of tourism movements
- Statistics of less visited sites



Prices from Internet

- Potential:
  - **Steep increase of online shopping**
  - **Data is publicly available**
  - **Frequent and automatized data collection**

- Feasibility Study (Netherlands):
  - **Standard software modules (open source license)**
  - **Data collection with Internet Robots**

Challenges and Myths

- Public understanding, perception and trust in statistics
- The end of official statistics monopoly: are NSIs at risk of going out of business?
- Big Data: the end of theory?
- A changing role for official statistics?

Public understanding, perception and trust in statistics

- We do have a serious gap
- The privacy paradox: two opposite faces of trust
- Communicating a value proposition for official statistics

Are NSIs driven out of business?

- It will not be that easy…
- Benchmarking
- Request for information by government will persist
- Unbeatable core values which underpin science, guide public policy and business decisions
- But we need to embrace 'data science' as being part of 'greater statistics'
- Business model for official statistics has to be adapted

The end of theory…or better theory?

- Scientific approach in the era of big data is needed more than ever before
- *Kirk Borne: Statistical Truisms in the Age of Big Data (19 June 2013):*
  -correlation does not imply causation
  -sample variance and bias do not go to zero
  -absence of evidence is not the same as evidence of absence
- A great moment: revisit theory in the age of big data

A changing role for official statistics?

- Accreditation and certification may become core tasks of NSIs
- Statistical modelling will be a main activity
- From descriptive indicators to nowcasting and forecasting
- Re-thinking surveys and censuses in terms of reality mining: blending big data with tradition
- It will be difficult to justify a 'traditional census of population' in the post2020 rounds

What will change?

The statistical office of the future

What will be the impact of ubiquitous data collection and networking

- **Internet of [every]Things,**
- **Cloud services,**
- **Wearables,**
- **Autonomous traffic,**
- **Smart systems,**
- **…**

on **official statistics**?

The statistical office of the future

- Data flows instead of surveys and censuses
- Data customer instead of data provider
- Product designers instead of data collection designers
- New answers related to
  - **Quality and transparency**
  - **Privacy and confidentiality**
  - **Access to third party data sources / data sharing**
  - **Scientific standards and methodology**
  - **Professional ethics**
  - **Skills**
- Accreditation and certification instead of production
- Embedded in data flow – ***statistics 'everywhere'***

Concluding remarks

- **Big Data is here to stay and … grow bigger**
- **Embracing big data and data science into 'greater statistics' is the only way forward**
- **We have much work to do !**

SAS Forum | **Twitter Contest – Tweet to win prizes!**

6. What is not a source of Big Data?
   A. Surveys
   B. Mobile phone data
   C. Smart meters

Tweet your answer:

Example: @spicyanalytics 6D

Start of your tweet | Question # | Your answer

Prizes to win:

1st prize: a ticket for Analytics 2015
2nd prize: a book of Prof Bart Baesens: "Analytics in a big data world"
3rd to 30th prize: chocolates with pepper

Winners will be contacted post-Forum !