Innovation in official statistics and multi- source statistical production

29 March 2017

Emanuele Baldacci
director of Methodology, Corporate Statistical and IT services

Niki Stylianidou; *ESS Vision 2020 Portfolio Management*

Dario Buono *Team Leader, Methodology in official Statistics*

Content of the presentation

- The digital statistical production landscape
  - **Tech trends**
  - **Digital-era statistical production**
  - **Towards smart statistics**

- Modernisation in the ESS
  - **ESS Vision 2020**

- New Methods for New Data
  - **innovation in official statistical production**
  - **key challenges and research areas**

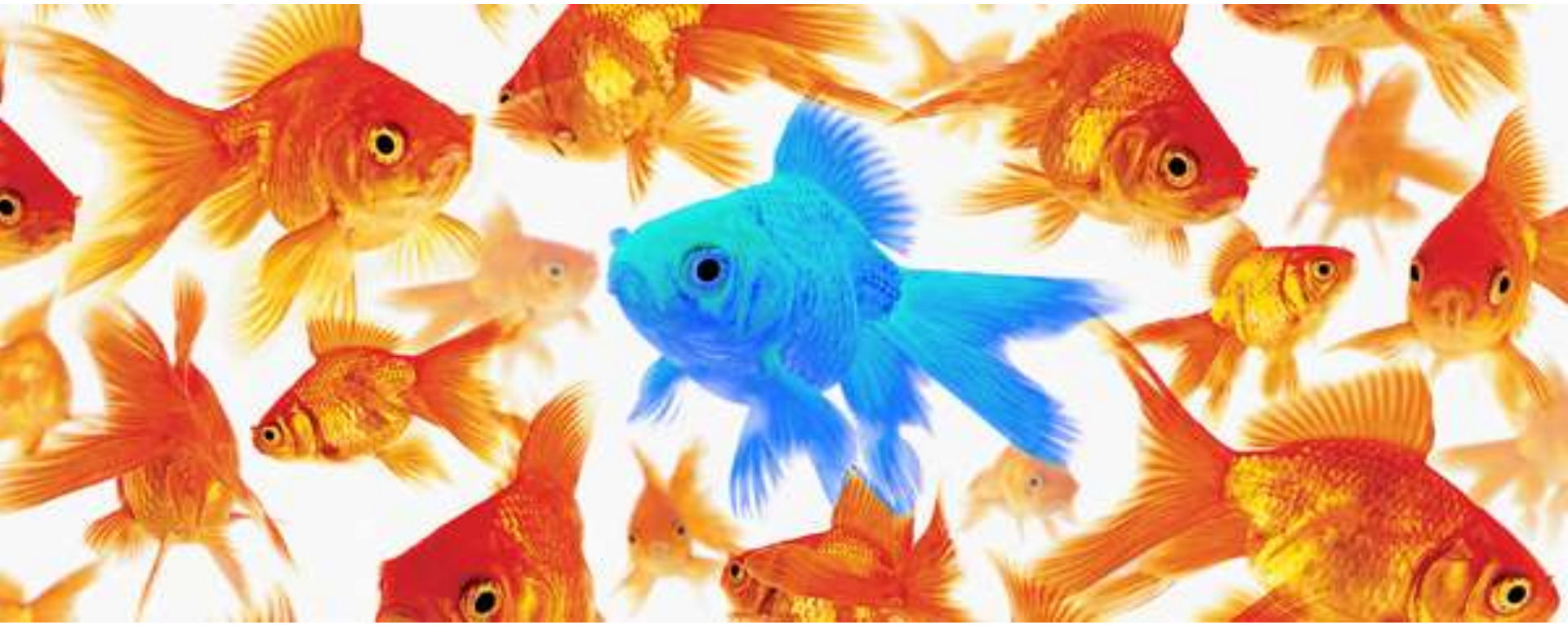
Digital Statistical Production Landscape

Innovation in official statistical production
Key challenges and research areas

by
Emanuele Baldacci, European Commission, Eurostat
emanuele.baldacci@ec.europa.eu @emabal

European Commission

We provide high quality statistics for Europe



...making a difference in the ocean of information

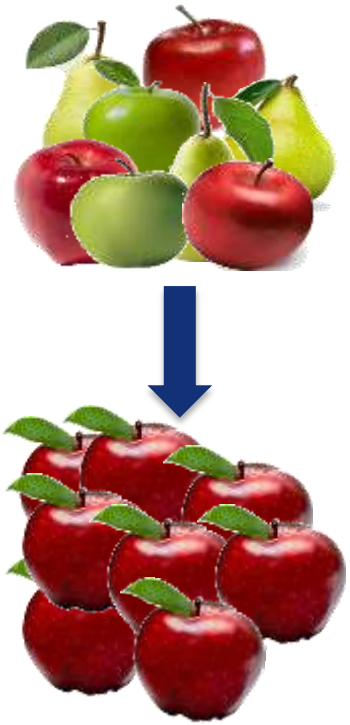
Eurostat

The European Statistical System (ESS)

A partnership of Eurostat and the National Statistical Institutes and other national authorities of the EU, the EEA and the EFTA countries

How does this partnership work?

- **ESS**:
  - Harmonisation of methodologies, concepts and classifications
- **National statistical offices**:
  - Collection of data
- **Eurostat**:
  - Consolidation of the data
  - Production of European aggregates



Data Revolution

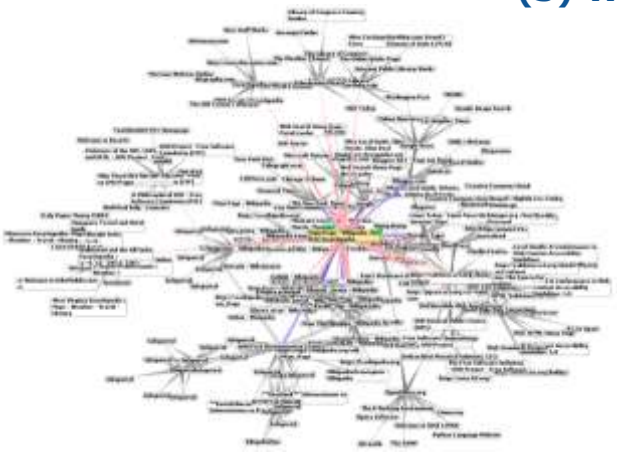
Technology Trends

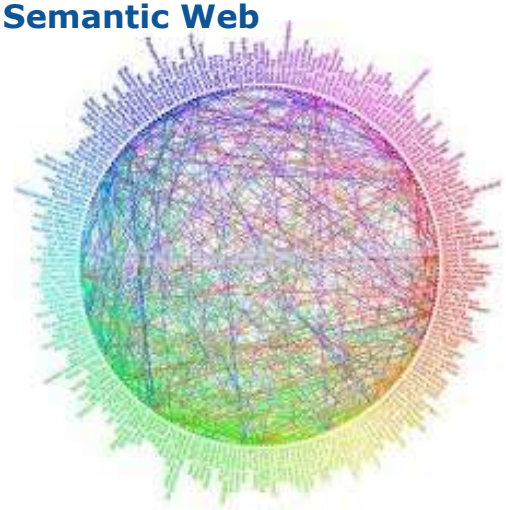
Electronic publishing (1990s) -> e-Business (2000s) -> d-Business (2010s)

(1) Digital Product Re-Mastering:

(2) Information Re-Mastering / Robots:

(3) Web of Documents -> Semantic Web





European Commission

Information & Data Trends

(1) Transfer of Legal Ownership

(2) Misinformation / Political Adverts

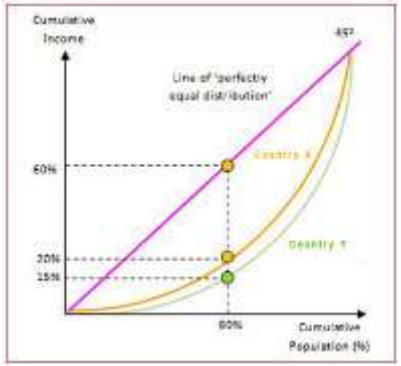
(3) Data Monetization

INDEPENDENT

Facebook knows you better than your friends do - because Likes reveal so much about your character

Study of 86,000 users reveals the power of intelligent machines, and they're getting better

(4) Analytics vs. Statistics



Competitive Ecosystem

(1) Data Giants

(2) New Entrants

(3) Vested Interest Groups

Month to Date View All

#	Make	Volume	Share
1	VOLKSWAGEN	482	13.05%
2	SKODA	348	9.42%
3	BMW	348	9.42%
4	FORD	321	8.69%
5	TOYOTA	308	8.34%
6	HYUNDAI	265	7.18%
7	AUDI	227	6.15%
8	OPEL	188	5.09%
9	KIA	147	3.98%
10	DACIA	145	3.93%

(4) Data as Brand builder

Residential Property Price Register - Home Page
https://www.propertypriceregister.ie/
The Residential Property Price Register is produced by the Property Services ... Price and Address of all residential properties purchased in Ireland since the 1st ...

Property Price Register sold prices | Daft.ie
www.daft.ie/price-register/
218803 results - See the latest house sale prices via the Property Price Register on Daft.ie. ... Price and Address of all residential properties purchased in Ireland since the ... Dublin €235,001 | 18/11/16 | New Dwelling House/Apartment SOLD

Property Prices in Ireland | Irish Real Estate Prices
www.globalpropertyguide.com › Europe › Ireland
A glance at changes in property, house and real estate prices in Ireland.

Irish Property Price Register
propertypriceregisterireland.com/
Search the Property Price Register. Find details of all property sales throughout Ireland.

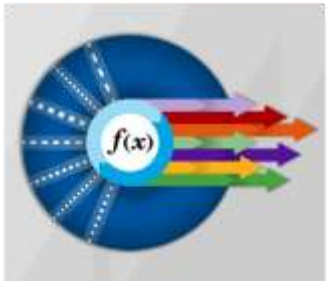
Latest Property Price Changes in Ireland - MyHome.ie
https://www.myhome.ie/priceregister
Find Thousands of House and Apartment price register details across Ireland and much more with MyHome.ie, Ireland's Leading Property Portal.

Impacts on statistics?

WIKIPEDIA
The Free Encyclopedia

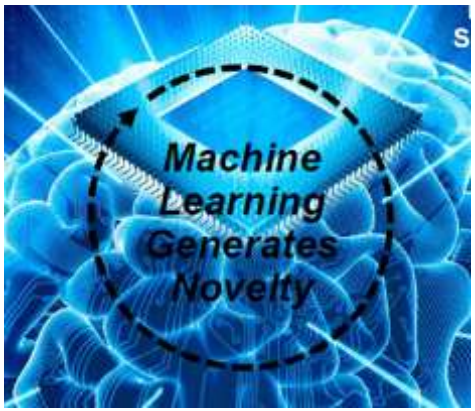
Google België

SMART STATISTICS



IMDb
The Internet Movie Database

skype™



Limitation of traditional surveys:

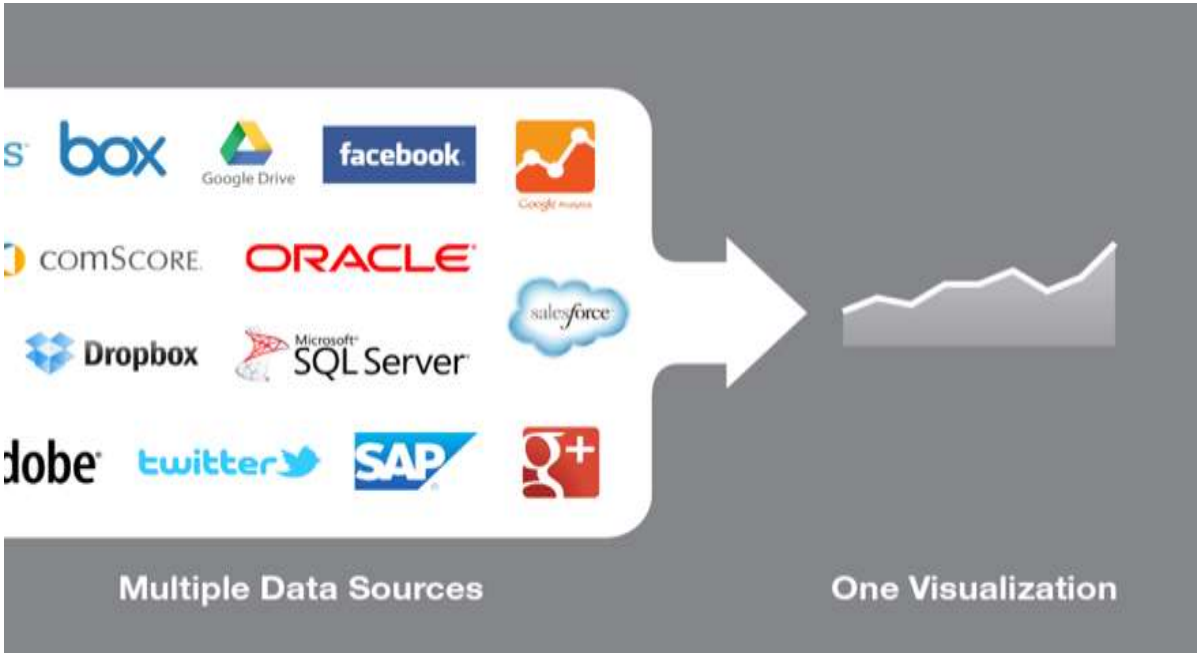
- Increasing non-response rates
- Concerns about response burden
- Lack of flexibility and associated costs

Innovations and changes of the statistical production cycle

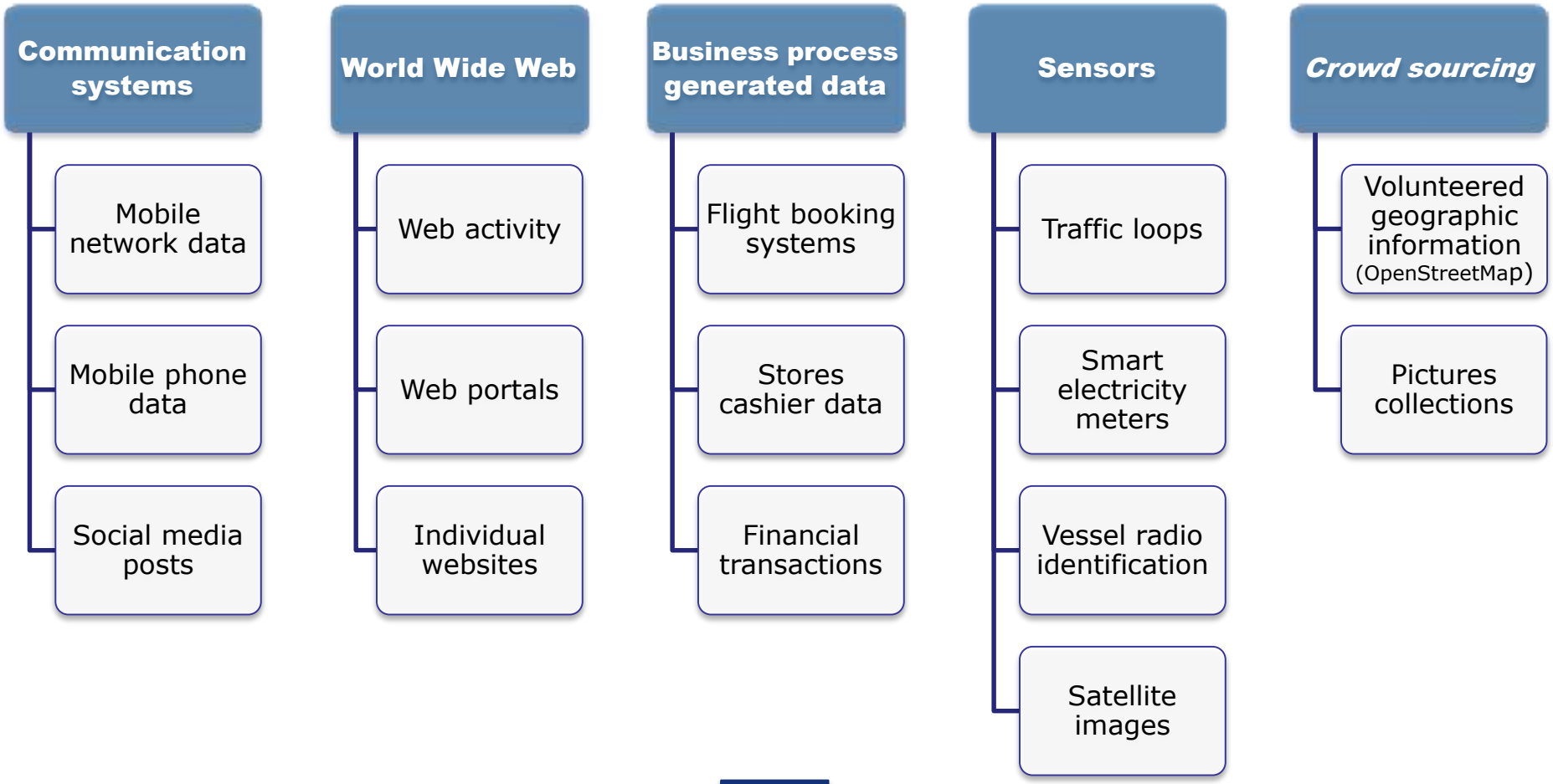
- Use of multiple data sources
- Data mashups
- A new data "factory"
- Data analytics services for "prosumers"

Use of multiple data sources

- Extending traditional data sources to administrative and big data



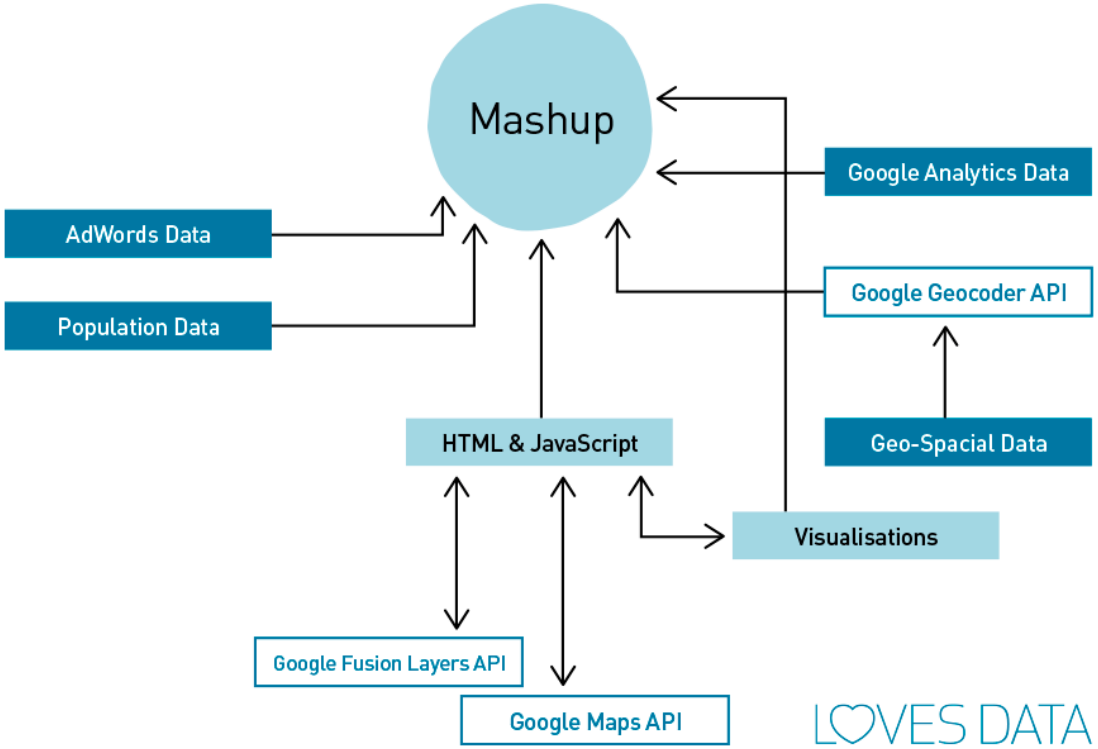
Big data sources



Data mashups

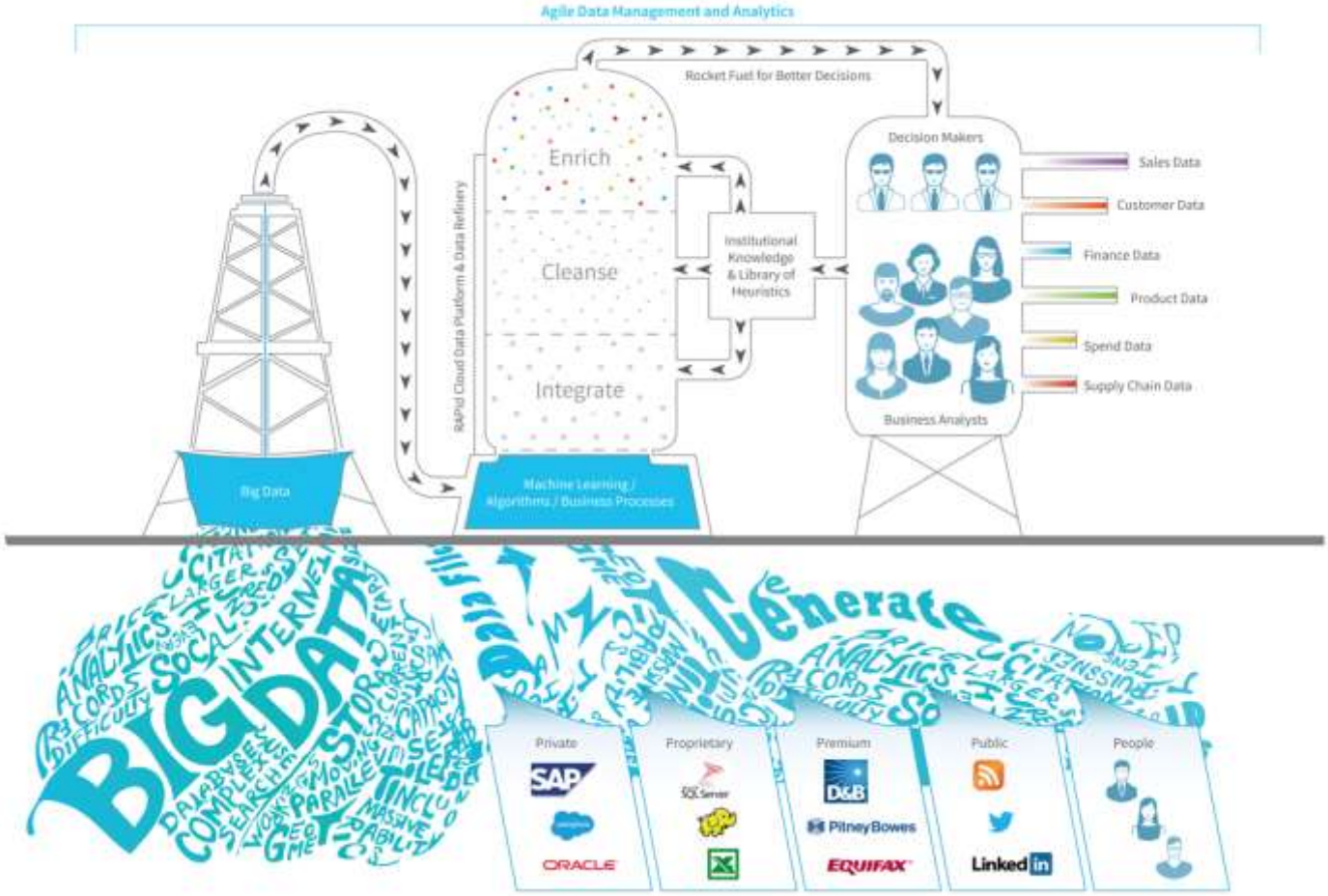
Assembling and reassembling data from multiple sources

Aim is to improve analytics



European Commission

A new data "factory"



Data analytics services for "prosumers"

Statistical organisations need to extend their products according to their users' needs:

- Tailor-made services for customers
- Data-driven surveys
- Economic modelling
- Forecasts and projections

Change without modernising the production processes would not be effective

Enabler: ESS Vision 2020 portfolio

Currently running projects:

ADMIN - Making administrative data more accessible

BIGD - Exploring the use of Big Data

DIGICOM - Tailoring statistical tools and channels for maximum benefit to users

ESBR - National business and Euro-groups registers harmonised and improved to allow sharing of data across borders

ESDEN - Improved data exchange procedures and security

SERV - Sharing statistical services around the ESS

Finalised projects:

SIMSTAT and REDESIGN - Towards a single information system to measure intra-EU trade flows

VALIDATION - Common standards and guidelines to validate data

END OF FIRST PART

Presentation of the ESS Vision 2020 Portfolio

by
Niki Stylianidou, European Commission, Eurostat
niki.stylianidou@ec.europa.eu

Contents

1. *The ESS Vision 2020*
   - **Key Areas - Objectives**
   - **Statistical domains covered**

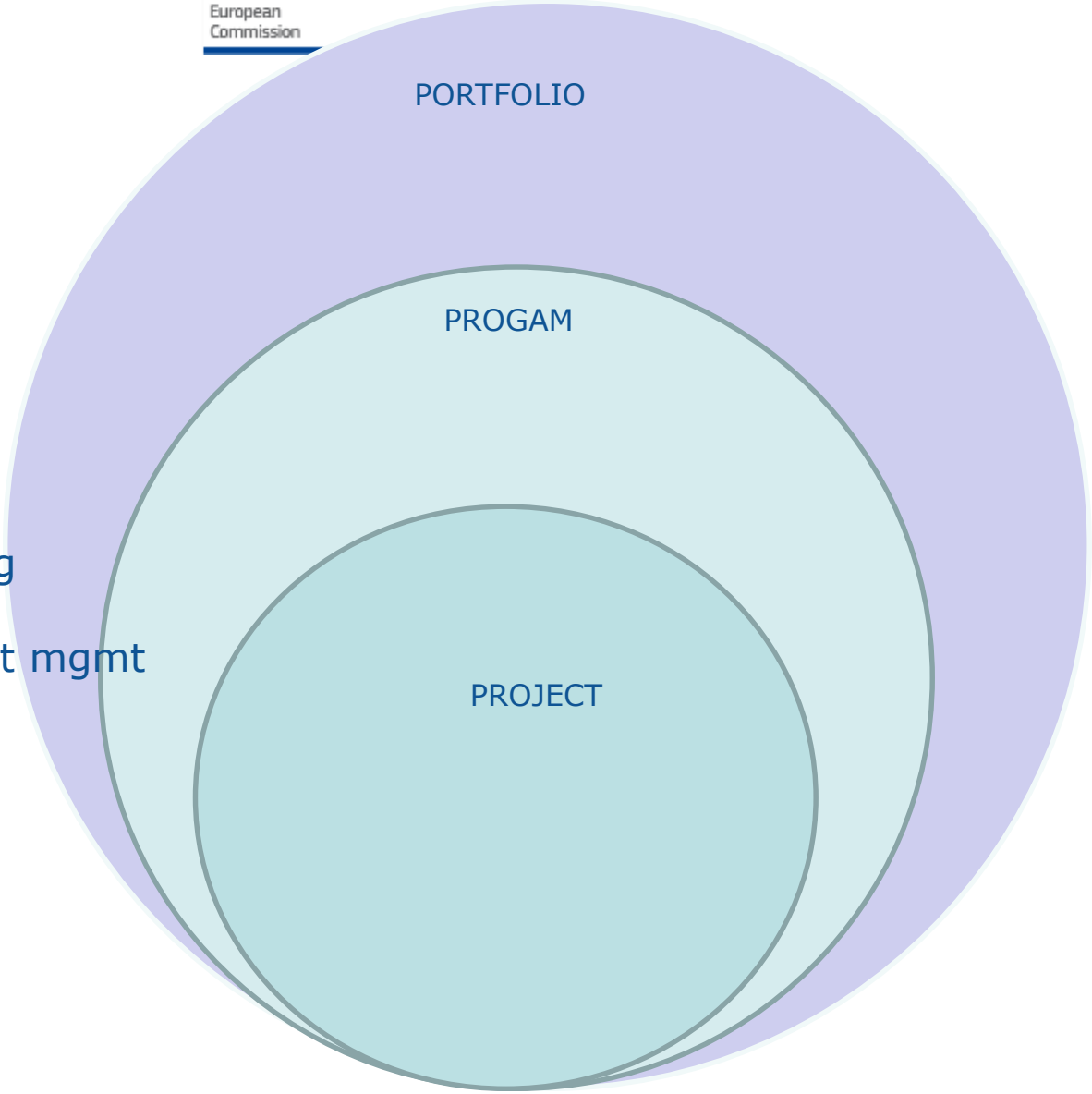
2. *Map of what the current portfolio covers*

3. *Questions & Answers*

PORTFOLIO MGMT
- Led by the Business
- Business Goal Alignment
- Business Value Alignment (risks/benefits)
- Program selection

PROGRAM MGMT
- Sponsored by the Business
- Ownership of Benefits
- multiple projects or working packages
- Compliance with the project mgmt standards

PROJECT MGMT
- Delivery of product or service
- Scope, Cost, Schedule
- Responsible for Quality of Deliverables



European Commission

Key Areas	Objectives	Projects
 Focus on users	>>Better understand and address the user needs >>Enhance quality management >>Assess the usability and quality of source data >>Promote the high quality of European statistics	DIGICOM ADMIN BIGD ESBRs SIMSTAT-REDESIGN-VALIDATION
 Strive for quality	>>Abide with the European Statistics Code Practice >>Apply fit-for-purpose tools to enhance quality assurance >>Assess the usability and quality of source data	
 Harness new data sources	>>Exploit potential of new data sources >>Invest in new IT tools and methodology >>Continue to improve existing data collection methods	ADMIN BIGD
 Promote efficiency in production processes	>>Further intensify the collaborative partnership of the ESS >>Further identify and implement the standards for statistical production >> Adopt enterprise architecture as common reference framework >>Benefit from exchange of (micro)data, while fully respecting statistical confidentiality	ADMIN BIGD ESBRs ESDEN SIMSTAT-REDESIGN-VALIDATION
 Improve dissemination and communication	>>Adopt a new dissemination and communication strategy >>Create a data pool of European statistics >>Optimise ESS portfolio of products and services >>Promote European statistics as a brand	DIGICOM

VIG/VIN role and activities

- Regular overseeing of the portfolio and reporting to the ESSC;
- Prioritisation of actions and initiatives;
- Risk analysis and mitigation;
- Benefits realisation (deployment of ESS.VIPs deliverables);
- ESS skills and capabilities to implement the Vision;
- Communication.

ESS domains

Europe 2020
Economic Governance & Globalization

Economic and Social Performance

Environmental Sustainability

Business


ADMIN


ESBRs


SIMSTAT/
REDESIGN

People's Europe


ADMIN

Geospatial,
Environmental,
Agricultural & Other

Cross cutting


DIGICOM


BIGD


SERV


VALIDATION


ESDEN

End Date: 01.12.2019

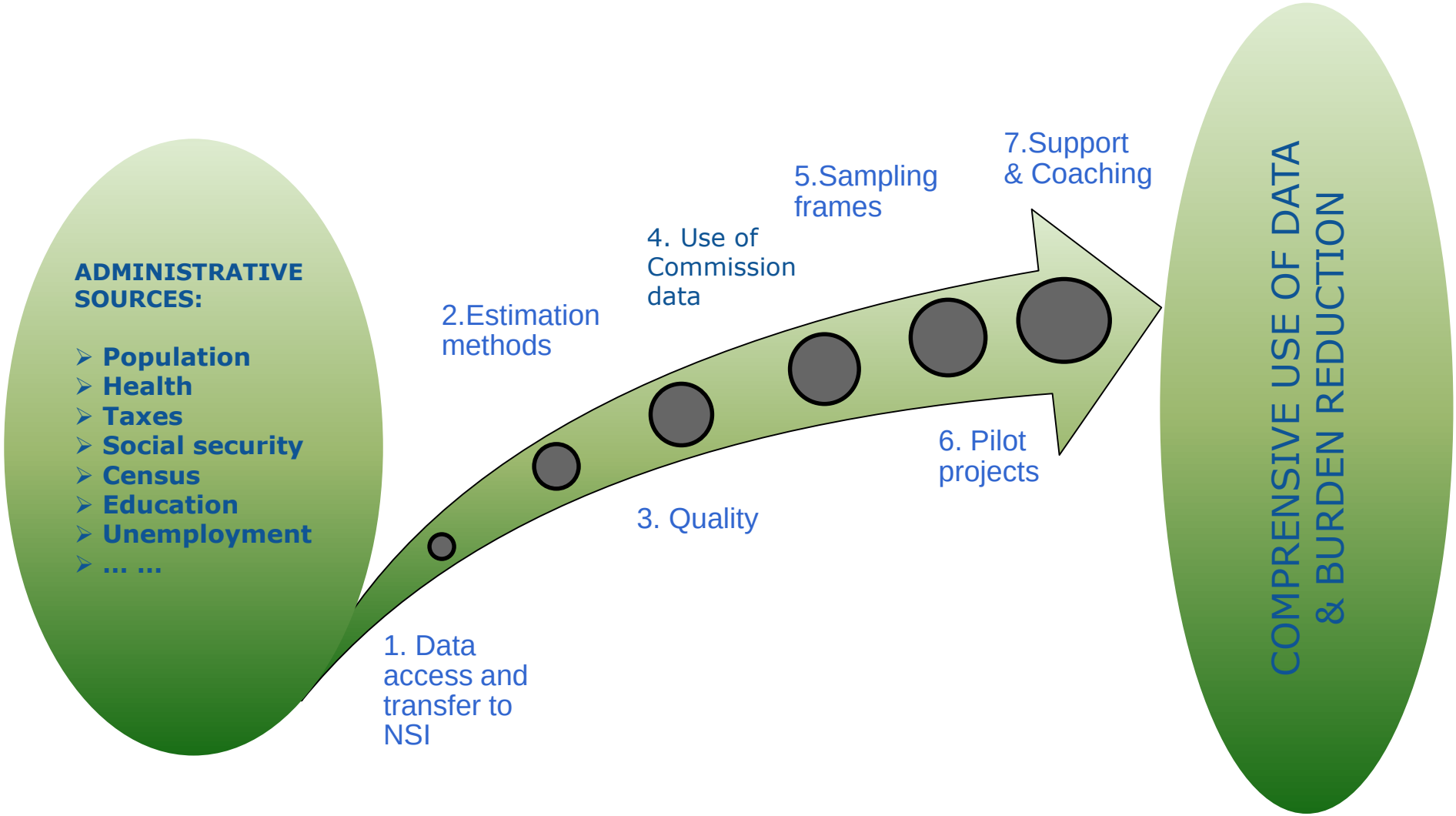
Purpose of the project

The project has a dual purpose:

*- to support the EU Member States to reap the benefits (**decrease costs and burden, increase of data availability** ….)*

*- promote the **quality** of the output produced using administrative sources, in particular the **comparability** of the **statistics***

ADMIN


ADMINISTRATIVE SOURCES:
Population
Health
Taxes
Social security
Census
Education
Unemployment
… …
1. Data access and transfer to NSI
2.Estimation methods
3. Quality
4. Use of Commission data
5.Sampling frames
6. Pilot projects
7.Support & Coaching
COMPRENSIVE USE OF DATA & BURDEN REDUCTION

More information

CROS portal

https://ec.europa.eu/eurostat/cros/content/essvip-admin-administrative-data-sources_en

http://ec.europa.eu/eurostat/cros/content/ess-admin-helpdesk_en

Functional email:

ESTAT-ESSVIP-ADMIN@ec.europa.eu

SIMSTAT/REDESIGN +ESDEN

SIMSTAT: 15.05.2016
REDESIGN: 01.04.2016
ESDEN: 01.12.2018

- Successful pilot exchange of micro-data in SIMSTAT;
- Large amount of data transferred without incidents;
- No security problems concerning the data;
- Connection of 20 Member States with a complicated network of actors involved;
- ESDEN network can be re-used for other secure data exchanges!

Purpose of the project

- *Enable the usage of new data sources*
  - **Increasing security, sensitive data**
  - **Upgrade capacity, high volumes**
- *Creation of efficient and robust statistical processes*
  - **Modernise interfaces, use standard transmission protocols**
  - **Foster automation**

More information

Functional email:

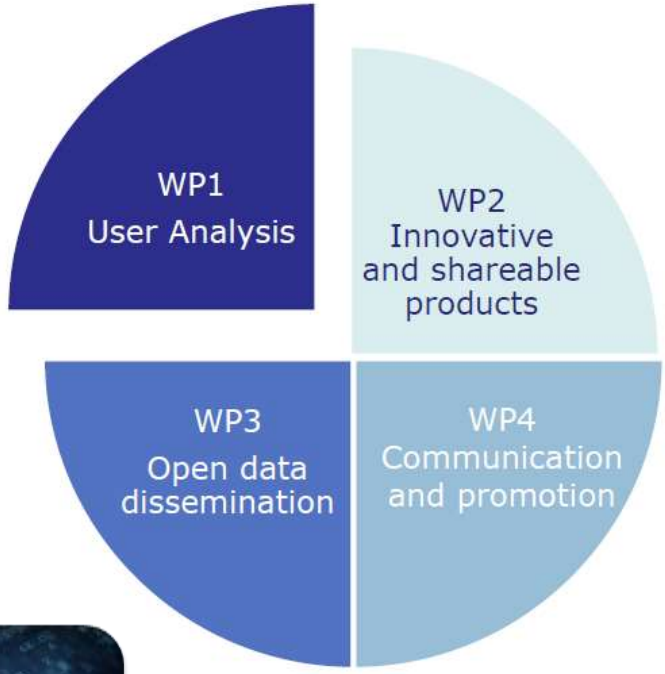
ESTAT-ESS-ESDEN@ec.europa.eu

DIGICOM

End Date: 31.12.2019

Purpose of the project

User engagement through social media



Reaching new users through visualisation

Empowering users to reuse and combine data

Improving users skills
Statistical literacy
Hackathon

Shared infographics within the ESS

- Economic Trends http://ec.europa.eu/eurostat/cache/infographs/economy/desktop/index.html
- Young Europeans *http://ec.europa.eu/eurostat/cache/infographs/youth/index_en.html*
- Quality of Life http://ec.europa.eu/eurostat/cache/infographs/qol/index_en.html
- You in the EU ttp://ec.europa.eu/eurostat/cache/infographs/youineu/index_en.html

User engagement through social media

For more information

Contact and to subscribe to the newsletter

ESTAT-ESS-VIP-DIGICOM@ec.europa.eu

DIGICOM News

Issue 1, September 2016

Editorial

By Emanuele Baldacci, Eurostat and Guillaume Mordant, INSEE

Focus on an event

The ESS Visualisation Workshop held in Valencia

Three questions to...

Karin Hansson, Statistics Sweden, coordinator of Work Package 1 — User analysis

Thematic focus

Work Package 2 — Innovative and shareable products and tools

One NSI, one experience

Statistical literacy at the forefront at Statistics Finland

Events

Overview of upcoming events

Contact

We are pleased to introduce the DIGICOM Project's first Newsletter. Digital communication, User analytics and innovative products (DIGICOM) is one of the eight projects belonging to the ESS Vision 2020 portfolio. It aims to modernize the communication and dissemination of European statistics, by developing innovative products and services, based on new technological opportunities, experiences in the ESS and the concrete needs of users. The project will also contribute to building the ESS's capacities in modern dissemination techniques (visualisation, Linked Open Data), user interaction (such as presence on social media), user knowledge (user analytics) and communication (communication strategy, branding).

DIGICOM was launched in January 2016 for a time-span of four years (see article on DIGICOM in the third issue of the ESS Vision 2020 Newsletter). With its four strands of work, 55 actions and as many deliverables, it is in fact more of a programme than a project and it will be managed in an agile way. Around 50 colleagues from 18 national statistical institutes (NSIs) have volunteered to become involved in the project's implementation, which is intended to benefit the ESS as a whole. Beyond the NSIs' considerable participation in the project's activities, enhancing their capacity to innovate, outputs and deliverables are expected to be of high interest to the ESS since reusable solutions are given priority.

This newsletter aims at informing those who are directly involved in the project, on what is happening in the various strands of work, but also, more broadly, at informing the management and staff of all ESS members, who may not directly take part in the project. You may also be interested as a reader, simply because you are curious, be it about who the users of statistics are or about new visualisation techniques for instance.

The newsletter is designed to illustrate DIGICOM through different angles: zooming on specific actions, asking actors implied for their views, focusing on an event or important progress made, without forgetting to shed light on some interesting NSI experience.

This first issue of the newsletter features a report on the first event organised by the DIGICOM project, the ESS visualisation workshop in Valencia of May 2016, an interview with a key project team member — Karin Hansson from Statistics Sweden who is coordinating Work Package 1 on User analysis — as well as a summary of key decisions taken on Work Package 2 on innovative products. The newsletter also presents the experience of Statistics Finland in increasing statistical literacy. This example was chosen from the ESS inventory of shareable practices, the DIGICOM project's first completed deliverable. We hope that you will find it interesting!

Emanuele Baldacci, Eurostat

Director of Methodology
Corporate statistical and IT services
DIGICOM project owner

Guillaume Mordant, Insee

Head of Department 'Insee Info Service'
Member of DIGICOM Steering Group
Chief editor of the DIGICOM Newsletter — Issue 1

http://ec.europa.eu/eurostat/web/ess/digicom

End Date: 31.10.2018

Purpose of the project

European System of Business Registers *(ESBRs) aims to resolve the issues of:*

- *Inconsistencies in business statistics due to **different production practices**, use and role of national Statistical Business Registers;*
- *Inconsistencies in statistics on globalisation due to missing a shared view on the operational structure of global **enterprise** groups;*
- *Inefficiencies in business register processes as well as statistical production processes due to missing an infrastructure for linking and sharing business register information.*

ESBRs

End Date: 31.10.2018

- Example of a very close collaboration
- EuroGroups Register (EGR) 2.0 in production
- Secure remote access to microdata established
- Interactive Profiling Tool (IPT) in production – first ESS attempt for online collaborative profiling
- 32 countries involved, 300 top groups profiled
- Profiling methodology stabilised

For more information

CROS portal

https://ec.europa.eu/eurostat/cros/content/esbrs-0_en

Functional emails:

ESTAT-EGR@ec.europa.eu concerning the EuroGroups Register (EGR)

ESTAT-IPT@ec.europa.eu concerning Profiling and the Interactive Profiling Tool (IPT)

End Date: 30.12.2020

Purpose of the project

Big Data (BIGD) aims to :

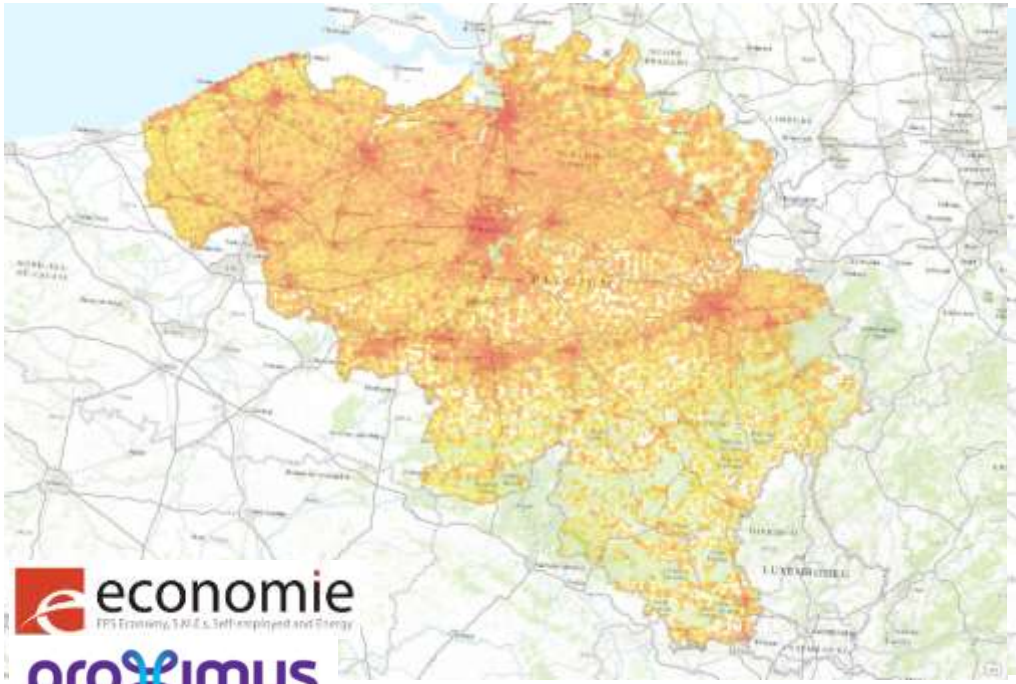
*enable the ESS to gradually **integrate big data sources into the production of European and national statistics** and, in this way, contribute to the broader aims of the ESS Vision 2020.*

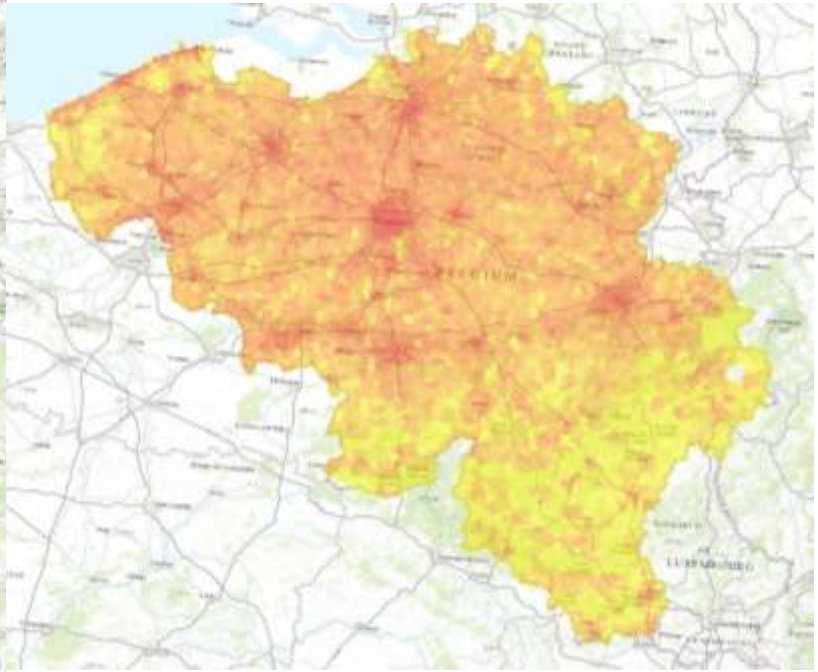
European Commission

Mobile phone data – population

Census 2011

Mobile phones 2015





For more information

CROS portal

- https://ec.europa.eu/eurostat/cros/content/esbrs-0_en

Functional emails:

- Konstantinos.Giannakouris@ec.europa.eu
- Fernando.REIS@ec.europa.eu

End Date: 01.12.2019

Purpose of the project

Shared SERVices *(SERV) aims to :*

- find a solution on ***how to make a service available, by replication in a national production process or to expose a service on a central ESS/Eurostat service oriented architecture****.*

- Contribute to ***standardising the description of business needs*** *by providing Common Statistical Production Architecture (CSPA)*

- Host and maintain a ESS Service Catalogue *at the Commission/Eurostat containing services identified at ESS level*

End Date: 01.12.2019

Purpose of the project

*- Provide **guidelines and recommendations** for the ESS enabling national projects to realise **technological solutions** for **industrialisation and integration of processes** and for the rationalisation of information systems.*

*- Provide an **ESS/Eurostat** SOA environment for sharing those services in ESS for which no replication is necessary (cloud).*

More information

Functional emails:
Pierre.PEYRONNEL@ec.europa.eu

Purpose of the framework QUALity

*….. projects and supporting frameworks contain quality elements ….consisting of the **European Statistics Code of Practice** and the general quality management principles of the ESS.*

*On one hand QUAL seeks to **ensure that developments related to quality within the ESS Vision 2020 projects** and supporting frameworks are in line with the ESS quality framework.*

*On the other hand QUAL aims to **identify new developments** which could have an impact on the ESS quality framework and which **might necessitate updating the framework.***

Purpose of the framework: Enterprise Architecture

*…. In this context the Enterprise Architecture framework acts as an enabler for the ESS collaboration by defining a **common language** to describe both what the **Business does /wants** to do in the future **and the IT systems and services that are needed** to achieve these goals.*

*In particular, the Enterprise Architecture framework, by incorporating the **principles of standardisation**, **interoperability** and service-oriented architecture, seeks to create the conditions for sharing components and for better integrating production systems across the ESS*

Cooperation Models

The ESSC requested in its May 2016 meeting, Eurostat has prepared an action plan to implement the recommendations proposed by the Resource Directors Group (RDG)

- **=>Task Force on Cooperation models.**

The action plan was sent to the RDG on 28 September 2016

Summing Up

The Vision suggests establishing common platforms for data storage, processing, and analysis to render production processes more efficient and effective at the European scale.

More information

ESS Website:

http://ec.europa.eu/eurostat/web/ess/about-us/ess-vision-2020

Functional email:

ESTAT-ESS-VISION-2020-NEWS@ec.europa.eu

Video Clip

https://www.youtube.com/watch?v=i5VRp6mrAjU&t=53s

END OF SECOND PART

New Methods for New Data

Innovation in official statistical production
Key challenges and research areas

by
Dario Buono*, European Commission, Eurostat
dario.buono@ec.europa.eu @darbuo

*The views expressed are the author's alone and do not necessarily correspond to those of the corresponding organisations of affiliation

1. Methodology@Eurostat
2. Data gaps
3. Data Analytics
4. Big Data and Nowcasting

Eurostat, the Statistical Office

- About 700 people with 28 different nationalities
- Statistical Office of European Union, part of EC
- Core business:
  - **Euro-zone (19) & EU (28) aggregates**
  - **harmonization, best practices, guidelines, trainings & international cooperation**
- Methodology team: Time Series, Econometrics, Statistical Disclosure Control, Research & Enterprise Architecture

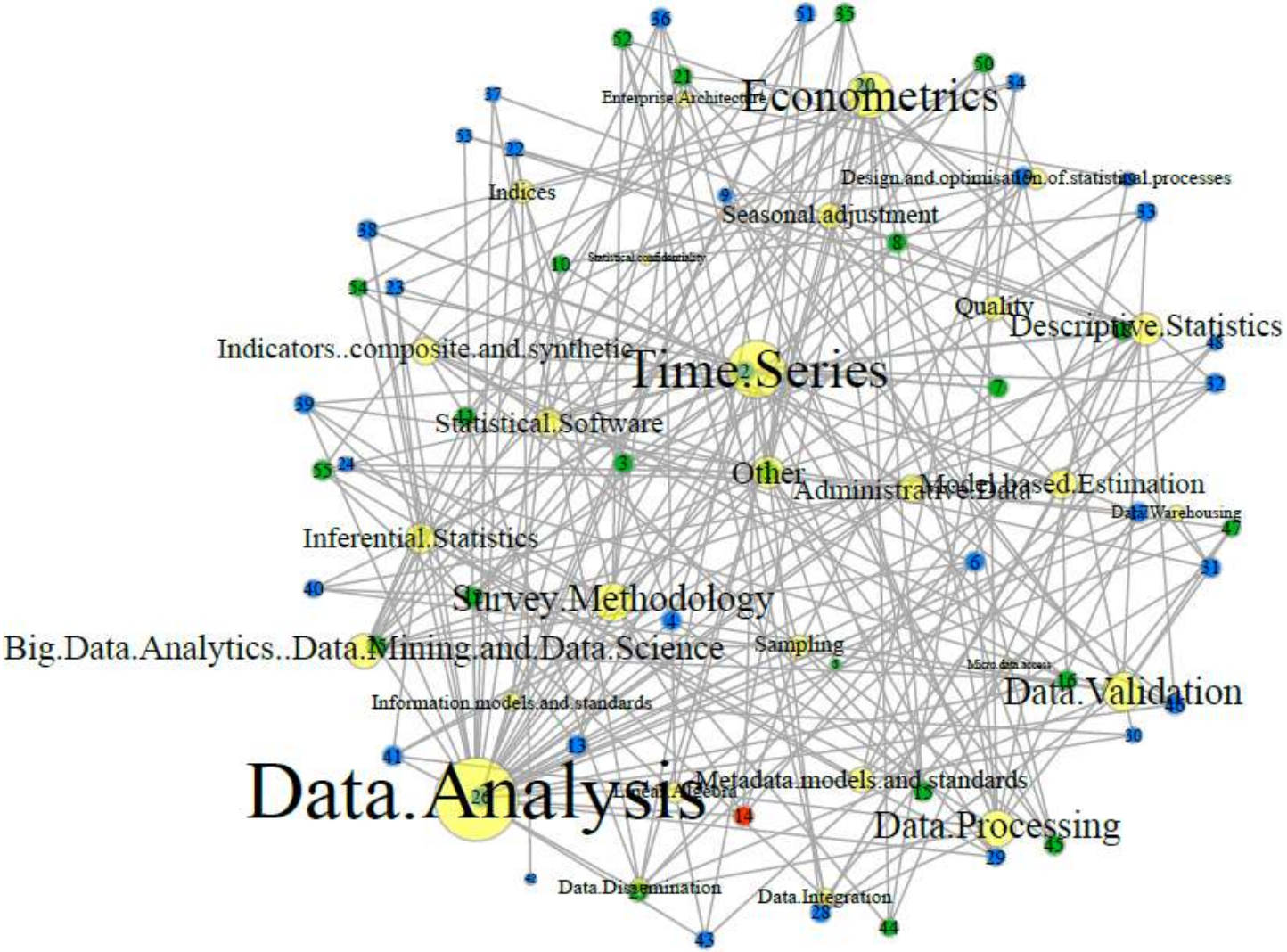
Eurostat dissemination

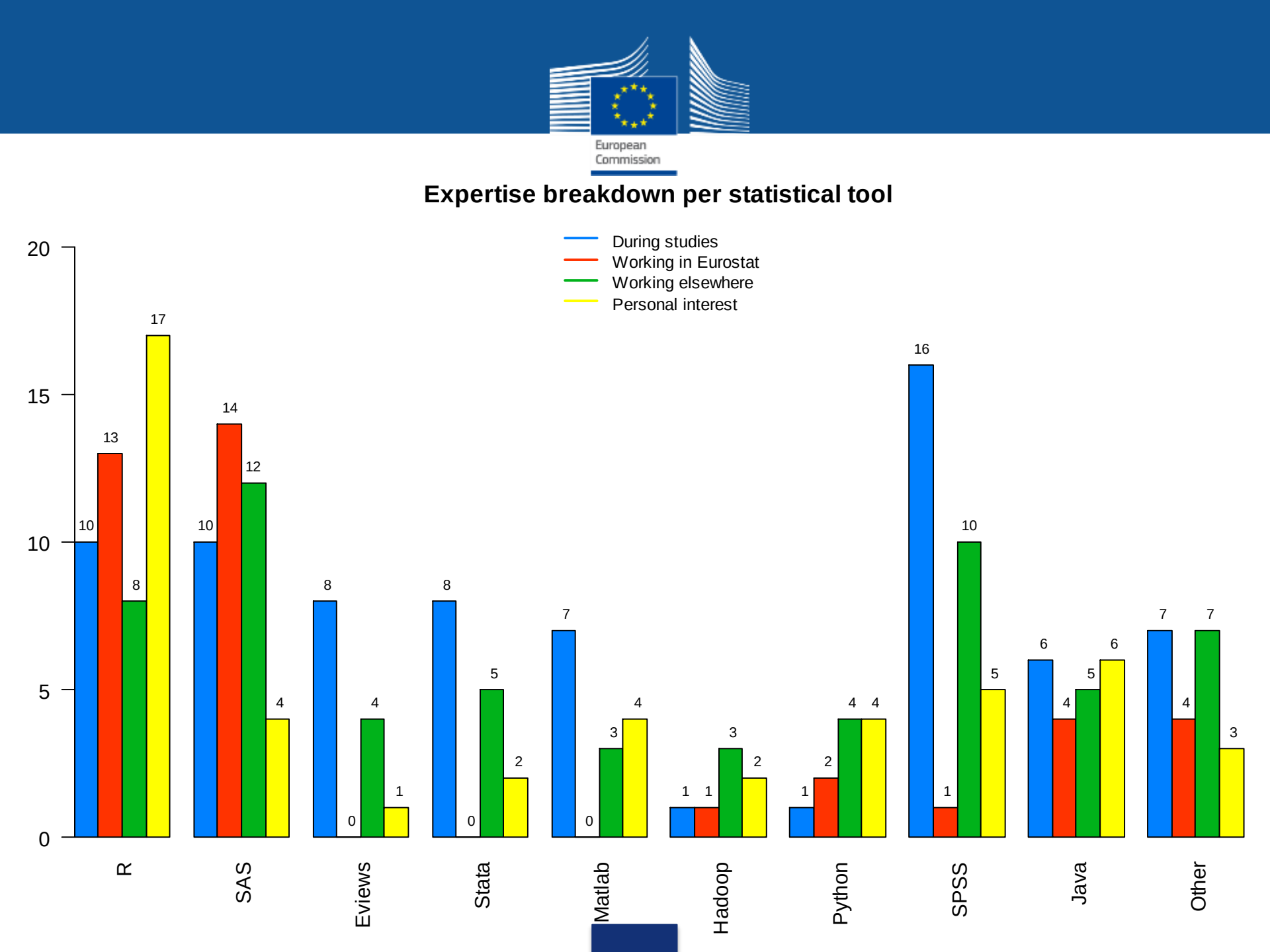
- **Statistics Explained** user-friendly wiki-based
- **Statistical books** data and analysis
- **Manuals and guidelines** applied in the ESS
- **Statistical working papers** research
- **Statistical reports** new or experimental data
- **Leaflets/brochures**
- **Infographics**

http://ec.europa.eu/eurostat/web/main

But we also have Facebook/Twitter accounts

Methodological Network




European Commission
Expertise breakdown per statistical tool
During studies
Working in Eurostat
Working elsewhere
Personal interest
20
15
10
5
0
10 13 8 17
10 14 12 4
8 0 4 1
8 0 5 2
7 0 3 4
1 1 3 2
1 2 4 4
16 1 10 5
6 4 5 6
7 4 7 3
R
SAS
Eviews
Stata
Matlab
Hadoop
Python
SPSS
Java
Other

Other skills

Regional Statistics/ Geographic information

Temporal disaggregation

Asylum migration

Cluster analysis

Labour Market

Dynamic Factor Analysis

Land cover and use

Fortran, Cobol,Pascal, C , C++,

PHP, Octave

ArcGIS

Monte Carlo, sensitivity analysis; Fuzzy logic

Neural network

Nowcasting

Classificat ions and metadata standards

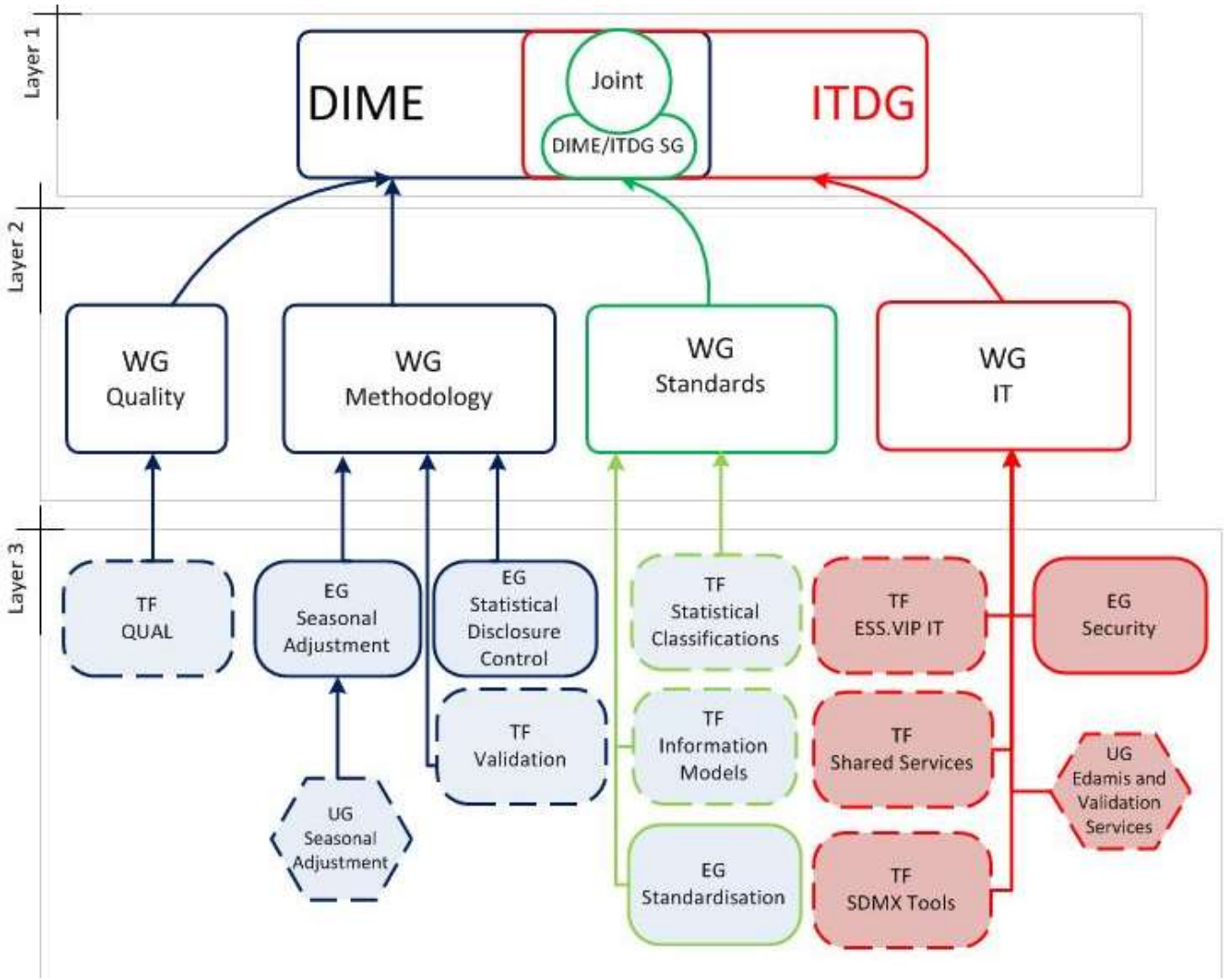
Remuneration and Pensions

Population size; analytical hypotheses testing

Business cycle analysis, S-VARS, G-VARS

OX-metrics

Questionnaire Design


Layer 1
DIME
Joint
DIME/ITDG SG
ITDG
Layer 2
WG Quality
WG Methodology
WG Standards
WG IT
Layer 3
TF QUAL
EG Seasonal Adjustment
EG Statistical Disclosure Control
TF Validation
UG Seasonal Adjustment
TF Statistical Classifications
TF Information Models
EG Standardisation
TF ESS.VIP IT
TF Shared Services
TF SDMX Tools
EG Security
UG Edamis and Validation Services

Filling the data gaps

Better use of the survey data

- **Modelling for more robust estimates**
- **Flash estimates**

Combining survey data with administrative data

Using new data sources such as big data

Methodological Strategy: pillars

1. Data collection profits from **multiple sources** and decisions are taken considering the widest possible evidence base (**Data4Policy**)

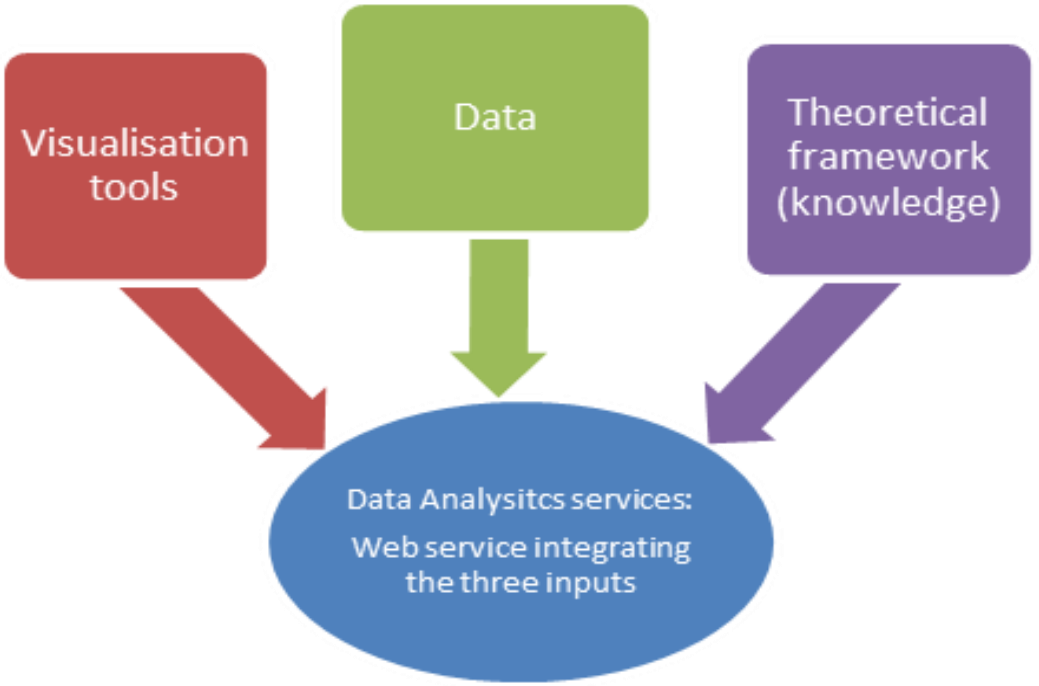
2. **Data are integrated** and provided as service to meet user priorities and needs.

3.Eurostat has greater capacity to improve availability of **data analytics** and **data visualization** tools.

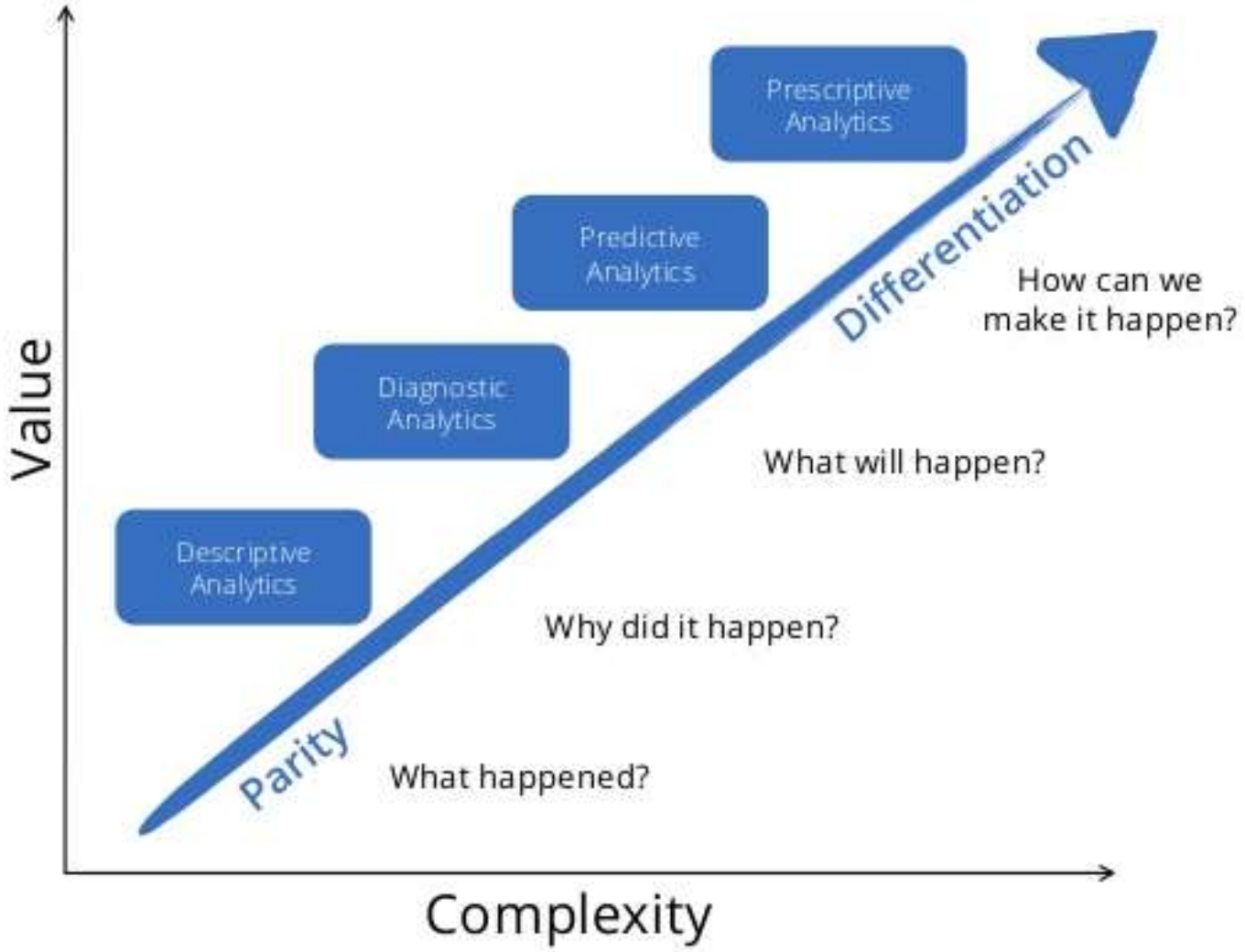
Data Driven World: context

- Demand of data users
- Methodologists need

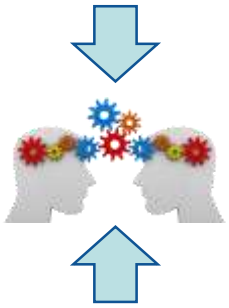
to innovate based on new statistical methods and information technologies



Data Analytics



Big Data Science



Official Statistics

smarter

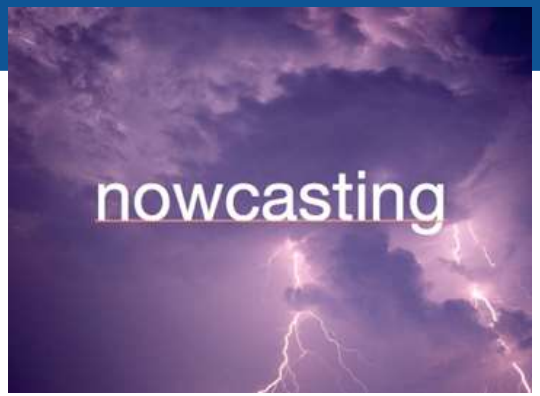
timelier

focused

Big data sources

- Web activity evidence for nowcasting
  - Google Trends
  - Employment
- Wikipedia as source for statistics
  - Cultural, Tourism Statistics
- Mobile communication data
  - Population, Land use, Tourism
- Web scraping
- Smart Meter

Flash estimations &



- Early indicators on Inequality
- t+30 estimates of GDP for Euro area
- HARMONISED INDEX OF CONSUMER PRICES - HICP Flash Estimate
- Euro area business cycle monitoring – rapid estimates
- **Big Data and Macroeconomic Nowcasting**
- **Eurostat's Handbook of Rapid Estimations**

Why interested in Big Data for nowcasting?

- **Big Data** are complementary information to standard data, being based on **different information sets**
- More **granular** perspective on the indicator of interest, both in the temporal and cross-sectional dimensions
- It is **timely** available, generally **not subject to revisions**

Research questions

Can Big Data help for Macroeconomic Nowcasting?

1. **Literature review**
2. **Models/methods** to be used for Big data
3. **Recommendations** on how to handle Big Data
4. **Case study**: IPI, Inflation, unemployment of some EU countries

European Commission

European research project: 2016

Big Data types

- Use of a typology based on Doornik and Hendry (2015):
  - **Tall** data: many observation, few variables
  - **Fat** data: many variables, few observations
  - **Huge** data: many variables, many observations

Models race

- Dynamic Factor Analysis
- Partial Least Squares
- Bayesian Regression
- LASSO regression
- U-Midas models
- Model averaging

255 models tested, **macro-financial** & **google trend** data

Statistical Methods: findings

- Sparse regression (LASSO) works for **fat**, **huge** data
- Data reduction techniques (PLS) helpful when dealing with many variables
- (U)-MIDAS or bridge modelling for mixed frequency

From Data Access to Modelling

Step-by-step approach, accompanied by specific recommendations for the use of big data for macroeconomic nowcasting, guiding to

- **the identification and the choice of Big Data**
- **pre-treatment and econometric modelling**
- **the comparative evaluation of results to obtain a very useful tool for decision about the use or not of Big Data**

Step 1: Big Data usefulness within a nowcasting exercise

Recommendations

1. *Evaluate the **quality** of the existing nowcasts and identify issue (bias or inefficiency or large errors in specific periods), that can be fixed by adding information in Big Data based indicators*
2. *Use of Big Data only when expecting to improve the timeliness and/or the quality of nowcastings*
3. *Do not consider Big Data sources with **spurious correlations** with the target variable*

Step 2: Big Data search
Recommendations

1. *Starting point for an assessment of the potential benefits/costs of the use of Big Data for macroeconomic nowcasting: identification of their source*
   - **Social Networks (human-sourced information)**
   - **Traditional Business Systems (process-mediated data)**
   - **Internet of Things (machine-generated data)**
2. *Choice is heavily dependent on the target indicator of the nowcasting exercise*

Step 3: Assessment of big-data accessibility and quality

Recommendations

1. *Privilege data providers with guarantee of **continuity** and of the availability of a good **metadata** associated to the Big Data*
2. *If a bias is observed a **bias correction** can be included in the nowcasting strategy.*
3. *To deal with possible instabilities of the relationships between the Big Data and the target variables, nowcasting models should be **re-specified on a regular basis** (e.g. yearly) and occasionally in the presence of unexpected events.*

Step 4: Big data preparation
Recommendations

1. *Big data often unstructured: proper mapping*
2. *Pre-treatment to remove deterministic patterns*
   - **Outliers, calendar effects, missing observations, Seasonal adjustment**
3. *Create a **specific IT environment** where the original data are collected and stored with associated **routines***
4. *Ensure the availability of an **exhaustive documentation***

Step 5: Big Data modelling strategy Recommendations

1. *Identification of appropriate econometric techniques*
2. *First dimension: choice between the use of methods suited for large but not huge datasets,*
   - **nowcasting with large datasets can be based on factor models, large BVARs, or shrinkage regressions**
3. *Huge datasets can be handled by **sparse principal components**, linear models combined with heuristic optimization, or a variety of **machine learning** methods such as **LASSO & LARS regression***
4. *In case of **mixed frequency data**, methods such as UMIDAS and, as a second best, **Bridge**, should be privileged.*

Step 6: Results evaluation of Big Data based nowcasting

Recommendations

1. *Run a critical and comprehensive **assessment of the contribution** of Big Data for nowcasting the indicator of interest based, e.g., on standard criteria such as **MSE or MAE**.*
2. *In order to reduce the extent of data and model snooping, a cross-validation approach should be followed:*
   - **various models and indicators, with and without Big Data, estimated over a first sample and selected and/or pooled according to their performance**
   - **then the performance of the preferred approaches re-evaluated over a second sample**

Case study

*- Implementation of all these steps for nowcasting **IP growth, inflation and unemployment in several EU countries** in a **pseudo out of sample context**, using Google trends for specific and carefully selected keywords for each country and variable*

- Big Data specific features: transform unstructured into structured data, time series decompositions, handling mixed frequency data

*- Overall, the **results are mixed** but there are several cases where Google trends, when combined with rather sophisticated econometric techniques, **yield forecasting gains**, though generally small.*

- Gains in term of timeliness or revisions have not been considered

Literature contribution

Eurostat Statistical Working Paper

"Big Data and Macroeconomic Nowcasting: From data access to modelling"

- Methodological finding will be included in 2 chapter of the **Eurostat/UNECE Handbook on Rapid Estimates** currently under 2nd peer review, (forthcoming in 2017)

What's next? Big Data Econometrics

2017, a new project focusing on:

- Econometrics, Filtering issues, advanced Bayesian estimation and forecasting methods
- **Real time** empirical evaluations (including a direct comparison with Eurostat flash estimates),
- **New ways and new metrics** to present nowcasts
- Possible data **timeliness/accuracy gains**
- Big data handling tool developed as **R package**
- Scientific summary for Big Data Econometric **strategy**

… in a nutshell

Methodology

- Multi-mode and multi-source approaches
- Algorithms, data mining
- New inference methods

Data access and sources

- NSIs coordinate administrative data
- legal, financial, ethical issues
- Minimum quality standards

IT environment

- Solutions for handling of Big Data
- Confidentiality
- Security

Skills

- Data analysis
- IT programming skills for unstructured data
- Econometrics

Thank you for your attention!!