Capacity building and data innovation in the ESS

International Conference on Big Data for Official Statistics
Data Science and Capacity Development in Official Statistics
Kigali (Rwanda), 2–3/5/2019
Contents

• Study on training strategy for big data
• Big Data test infrastructure
• European Statistics Training Programme
• European Master in Official Statistics

Never stop learning, because life never stops teaching
Training Strategy for Big Data

To bridge the Big Data skills gap in European official statistics

- Identification of skills required for the use of big data sources
  - Skills framework
- Inventory of existing skills in Eurostat and in the NSIs in Europe
  - Questionnaire to NSIs and Eurostat
- Analysis of the big data training needs
- To define the training objectives and content
  - Competency-Based Education approach
- Develop a training provision strategy to bridge the skill gap using different instruments
Skills classification by groups of skills

- **Skills**
  - Soft skills
  - Tasks and methods
    - Data science tasks
    - Statistical tasks
    - Information technologies (for statistical purposes)
    - Budget tasks (for statistical purposes)
  - Tools and technologies
    - Upper-level data science tools and technologies
    - Statistics and business intelligence tools and technologies
    - Programming languages
    - Architecture tools and technologies
    - Data management
    - Search tools and technologies
    - Visualization tools and technologies
    - Hadoop
    - Databases
    - Data mining tools and technologies
    - Cloud tools and technologies
## Intended outcomes of strategic training plan

<table>
<thead>
<tr>
<th>Time</th>
<th>Change in</th>
<th>Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short</td>
<td>- Knowledge</td>
<td>• Awareness raising</td>
</tr>
<tr>
<td></td>
<td>- Skills</td>
<td>• Change of attitude</td>
</tr>
<tr>
<td></td>
<td>- Attitude</td>
<td>• Skills acquisition</td>
</tr>
<tr>
<td></td>
<td>- Motivation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Awareness</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Awareness raising</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Change of attitude</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Skills acquisition</td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>• Behaviors</td>
<td>• Constant support of level of expertise</td>
</tr>
<tr>
<td></td>
<td>• Practices</td>
<td>• Change of behavior and practices using of big data and other emerging</td>
</tr>
<tr>
<td></td>
<td>• Policies</td>
<td>technologies</td>
</tr>
<tr>
<td></td>
<td>• Procedure</td>
<td>• Adoption of new standards, policies and procedures</td>
</tr>
<tr>
<td>Long</td>
<td>• Situation – environment, social</td>
<td>• Knowledge and new technologies will drive the statistical production</td>
</tr>
<tr>
<td></td>
<td>and economic conditions</td>
<td></td>
</tr>
</tbody>
</table>
The Big Data Test Infrastructure

Big Data use cases in scope

- Predictive Analysis
- Descriptive Analysis
- Network Analysis
- Text Analysis
- Route-Traceability / Flow Monitoring
- Time-Series Analysis
- Web Analysis (Scraping / Monitoring)
- Social Media Analysis
- Population / Customer Segmentation
- IoT Security
- IoT & Smart City
- Image Processing

Legend:
- Use Case Priority: High, Medium, Low
- Use Case in scope
- Use Case not in scope
The Big Data Test Infrastructure

- PaaS for implementing Big Data Use Cases
- Community Building and Innovation Portal
- Big Data and Analytics software catalogue
- Advisory
- Data Catalogue and Data Exchange APIs
- Support for Analytics implementation
- Analytics as a Service

Big Data Test Infrastructure service offering
BDTI and the ESS

**BDTI sandbox test with Wikipedia page views**

**First pilot: European Big Data Hackathon 2019**

**ESSnet big data pilots**

- Online Job Vacancies
- AIS
- Mobile phone
European Statistics Training Program (ESTP)

- The **purpose** of the ESTP is to provide European statisticians with continuous training in new methods, techniques and best practices and integrate the application of European concepts and definitions.

- The programme is tailored to **meet the specific needs** of the European Statistical System (ESS).

- The programme is **managed by Eurostat**, the Statistical Office of the European Union.

- The ESTP **complements national training schemes** and meets the challenges of comparable statistics at European and international level.

- Training **delivered** by external contractors or staff members.

- **Cost** for delivering training ca. 500,000 € over four years.
ESTP - Target audience

Staff of NSI/ONAs in EU Member States, EFTA countries, candidate countries and Eurostat

Applicants from non-European countries may be admitted on individual basis and subject to availability of places
ESTP Programme 2019

- Statistical Methodology
- Economic Statistics
- Environmental Statistics
- Metadata and Classifications
- Quality and Dissemination
- New Data Sources
- Business Statistics
- Social Statistics
ESTP courses supporting big data (2019)

- Introduction to big data and its tools: 18-21 Jun
- Mobile Network Data: 19-21 Feb
- Webscraping: 23-26 Sep
- Social data mining: 16-17 Sep
- AIS vessel data: 14-16 Oct
- Making statistics by combining data from different sources: 18-20 Sep

Other courses:
- Statistics production with the use of GIS: 5-7 Jun
- The use of R in official statistics: model based estimates: 3-6 Jun
- Machine learning econometrics: 25-27 Jun

Big data courses: [Blue]
Methodology courses: [Gray]
ESTP more information

ESTP National Contact Points

ESS Website

CIRCAABC
https://circabc.europa.eu/w/browse/6ade1ca8-6a06-44bd-bff0-498217d0ec05
European Master in Official Statistics (EMOS)

- Label awarded by ESS Committee to Master programmes fulfilling EMOS requirements – valid 4 years
- Coordinated by Eurostat, universities deliver on learning outcomes and mandatory internships with statistical institutes
- Current challenges: recruitment, training & development, long-term retention of employees.
- EMOS graduates – both new students and existing NSI staff - better prepared for work in statistical offices
## EMOS - Label based on learning outcomes

1. **System of official statistics**
   - Official statistics: relevance and sources
   - International and European institutions
   - European Statistics Code of Practice

2. **Production models and methods**
   - Different kinds of data sources!
   - Quality dimensions

3. **Specific themes**
   - Economy & finance, population & social conditions, international trade, environment & energy, science & technology …

4. **Statistical methods**
   - Statistical methods: sampling & estimation, index theory, econometrics …
   - Statistical programmes

5. **Dissemination**
   - Effective communication with different audiences
   - Confidentiality and disclosure control
EMOS- 33 programmes in 19 countries

CALL I – 2015
University of Helsinki
Ensai and Université de Rennes 1
Universität Trier
Università degli Studi di Firenze
Università di Pisa
Sapienza – Università di Roma
Universiteit Utrecht
Universidade NOVA de Lisboa
Örebro University
University of Ljubljana
University of Southampton

CALL II – 2016
KU Leuven
University of Bamberg
FU Berlin
TU Dortmund
LMU Munich
Complutense University of Madrid
University of Bergamo
University of Luxembourg
University of Sopron
JKU Linz
University of Donja Gorica

CALL III – 2019
EMOS label award (10)
University of Economics, Prague
Eötvös Loránd University (ELTE), Budapest
University College Dublin
Leiden University
Warsaw University of Life Sciences
Wroclaw University of Economics
Gdańsk University of Technology
University of Łódź
Bucharest University of Economic Studies
University of Bologna
EMOS
Focus of 3rd call 2018-2019

• Production of statistics by integrating data from different kinds of data sources
• Model-based and algorithm-based estimation techniques
• Understanding of institutional environment
• Dissemination – visualisation techniques
• Communication skills