Smart Statistics Web Intelligence Hub
Focus on data from online job advertisements

Case Study - Eurostat

Big Data for Labour Market Intelligence

ETF Training Programme

Fernando Reis

8, 10 and 15 June 2021

The Web as a statistics data source

- Web scrapping is easy, however…
- You want it to be:
  - **Automated**
  - **Methodologically sound**
  - **Robust**
  - **Transparent**
  - **Reproducible**
  - **Consistent**
  - **Efficient**
  - **Providing time-series**

The Web as a statistics data source

- Web scrapping is easy, however…
- Producing official statistics is difficult!

- The WIH is our tool to take care of the difficult part.

The case for a web intelligence shared system

- Acquisition of Web data in a statistical production context is not easy (e.g. data agreements)
- Infrastructure with big data capabilities is required
- Specialised skills are required
- Web intelligence capabilities spread out the ESS will take very long
- An European system for OJA will exist anyway

A quick history of the exploration of big data in official statistics

- 2013: Scheveningen Memorandum on Big Data
  - **Examine the potential of Big Data sources for official statistics**
- 2015: Big Data Action Plan and Roadmap
- 2016 – 2020: ESSnet Big Data I
  - **Big data pilots (incl. OJA)**
- 2018: Bucharest Memorandum on "Official statistics in a datafied society (Trusted Smart Statistics)
  - **Focus on implementation**

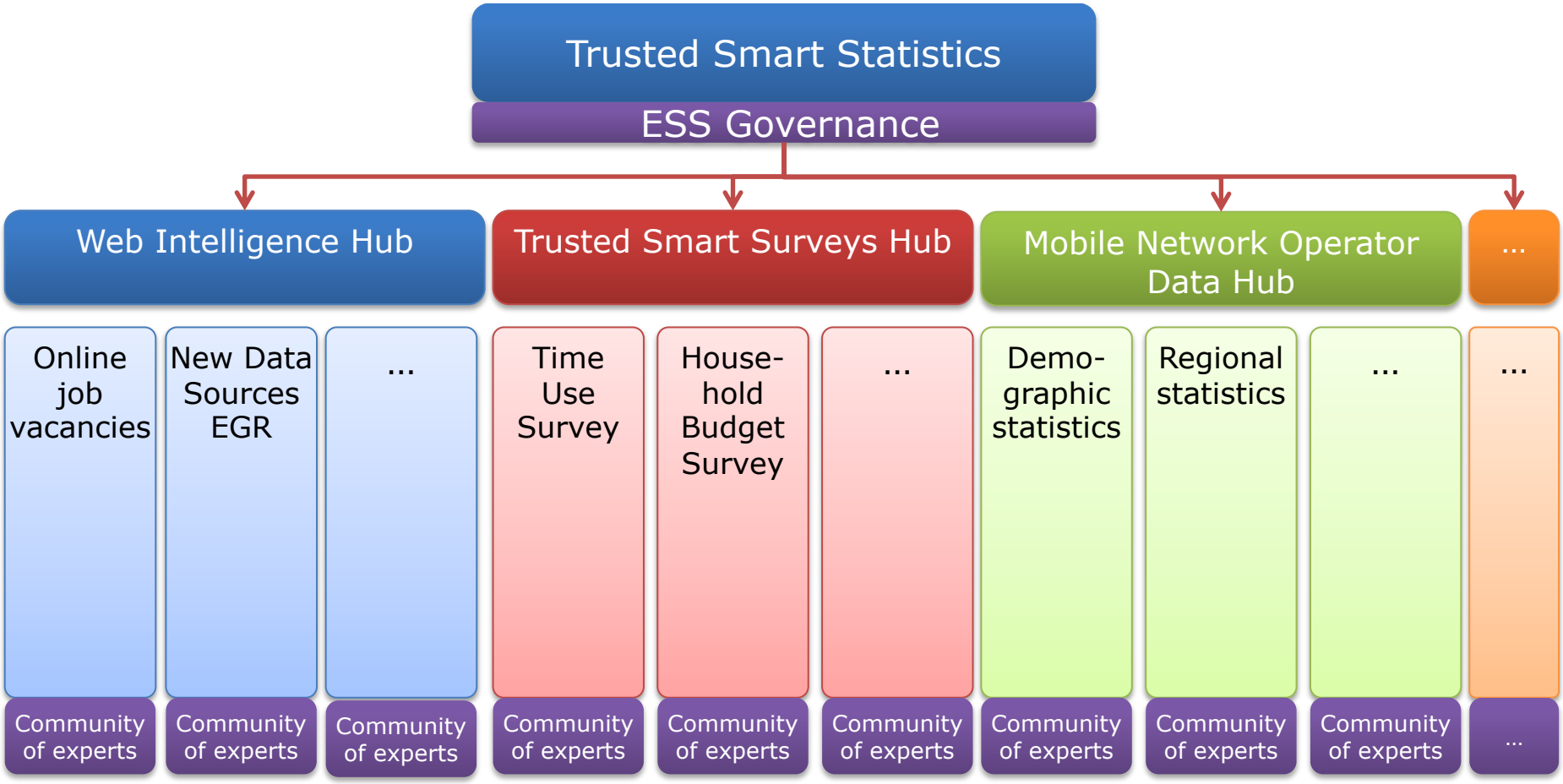
From Big Data to Trusted Smart Statistics

- We can think of Smart Statistics as being the future system of official statistics, where **data capturing, processing and analysis will be embedded in the system itself**, starting with the digital footprints of the human activities
- putting **intelligence to all stages of the data lifecycle** it is expected to enable Official Statistical Agencies to **maintain and reinforce their role as a key providers of data in a digital world**

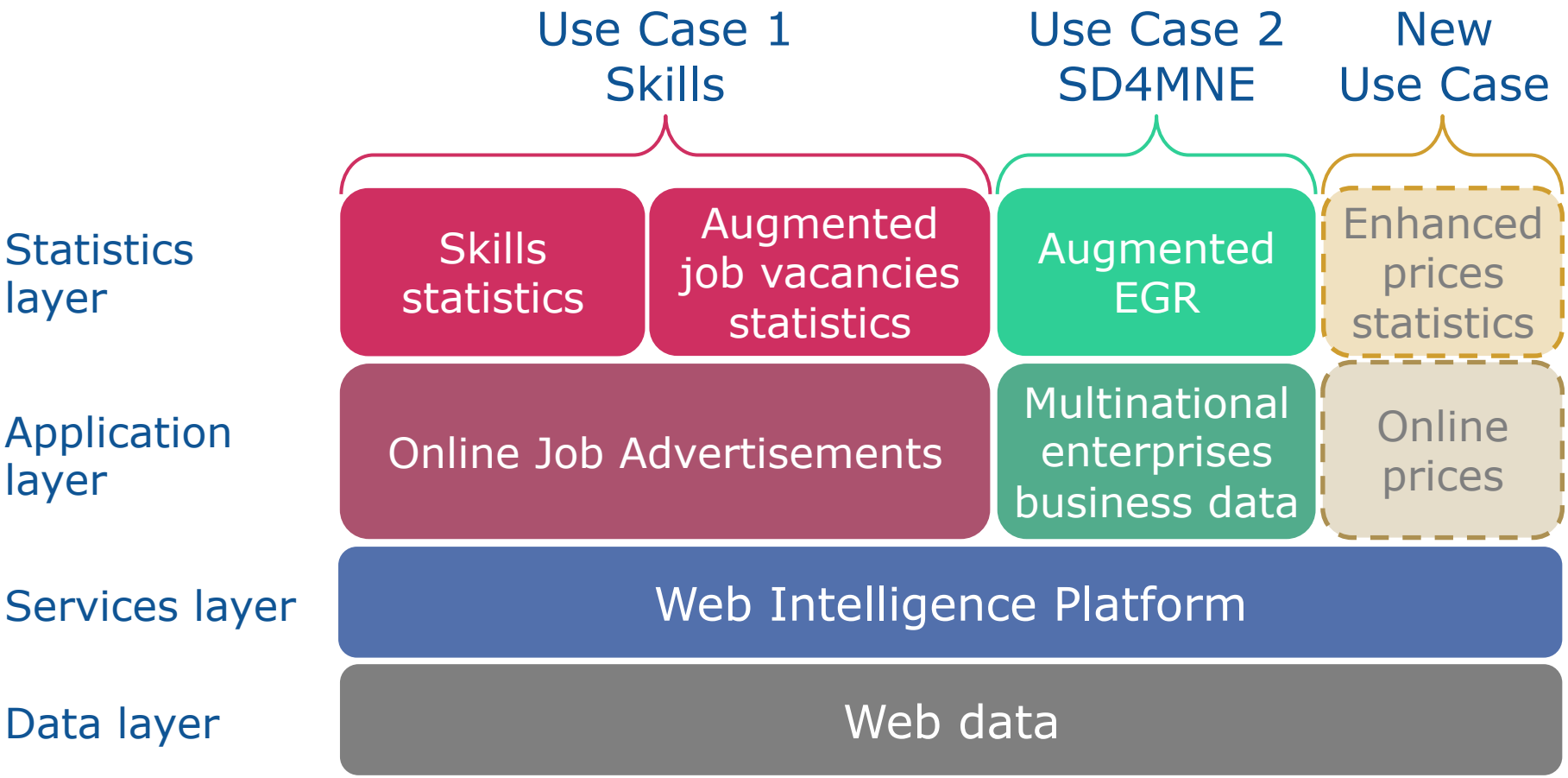
Principles of Trusted Smart Statistics

- Multi-source statistics
- Multi–purpose data sources
- Layered organisation: the hourglass model
- Modular methodological frameworks
- Pushing computation out
- Use data without sharing

Trusted Smart Statistics



Web Intelligence Hub



Web Intelligence Hub - Services

- Provide support to ESS partners in
  - **Data acquisition (web scrapping, APIs)**
  - **Trans-national data agreements**
  - **Partnership models for national data agreements**
  - **IT infrastructure and tools**
  - **Analytical services (e.g. NLP)**
  - **Methodology**
  - **Regulatory aspects**
  - **Skills (training material)**
  - **R&D collaboration**
  - **Governance**

Web Intelligence Hub - Principles

Some principles

- ESS hub
- Serving national and European needs
- Modular structure
- Defined processes and products to be guaranteed
- Priority to working together, possibility to act individually
- Programs should be open source
- Transparency as much as possible
- Common used processes should be certified and audible
- Lineage of data and processes
- Intermediate products usable by all partners

Moving big data to implementation

- ESSnet Big Data I: pilots
- Call for a refocus on the implementation of the most successful pilots
- ESSnet Big Data II: new pilots + trusted smart statistics + implementation
  - **Implementation = producing specifications for implementing, experimental statistics, recommendations for data / process governance**
  - **Online job advertisements**
  - **Enterprises websites**
  - **Smart electricity meters**

OJA data collection systems

- Mostly national approach
  - **ESSnet Big Data**
- European approach
  - **DG-CNECT**
  - **Cedefop**

How to create a WIH

- Build a **platform**
- Create a **community**
- Develop **methodologies**
- Design **learning resources**
- Secure proper **regulatory framework**
- Don’t forget **communication**
- Clarify **governance**

6 steps to build the Web Intelligence Platform

1. Create IT infrastructure for the WIHP

   Done

2. Deploy OJA Cedefop system to WIHP IT infrastr.

   Done

3. Design architecture for reusability for WIHP

   Draft done, discussing with stakeholders

4. Develop components for WIHP (MVP)

   Starting now

5. Launch new use case(s) in WIHP

   2021

6. Transfer OJA to WIHP

   2022

How to create a WIH community

- Web Intelligence Network (WIN)
  - **Promote use of WIH in official statistics production**
  - **Support its methodological development;**
  - **Operationalise the role of the NSIs in the WIH:**
    - ✓ use the services of the WIH;
    - ✓ participate in the running of the use cases;
    - ✓ contribute to the development of the components;

- WIN launched 1 April 2021

How to develop WIH methodologies

- Adopt and further fine tune methodologies developed in the context of OJA Cedefop system
- Gather contributions from official statistics community (WIN)
- Develop methodologies for new use cases
- Develop a OS graded quality framework for Web data

How to design learning resources

- Training courses
  - **European Statistics Training Programme (ESTP) – web scrapping courses since 2015**
  - **Advanced coaching - beginning of next year**
- Training material – available online
- Playground
- Organise workshops

How to secure proper regulatory framework

- Adopt ESS web data retrieval policy
- Identify regulatory needs
- Privately held data regulation

Thank you for your attention

Fernando Reis

Eurostat

fernando.reis@ec.europa.eu

https://github.com/reisfe/

https://twitter.com/reisfe/

https://linkedin.com/in/reisfe/