



Trusted Smart Statistics Centre Web Intelligence Hub

Getting the future right

Towards smarter and people-centred skills intelligence

13 April 2021

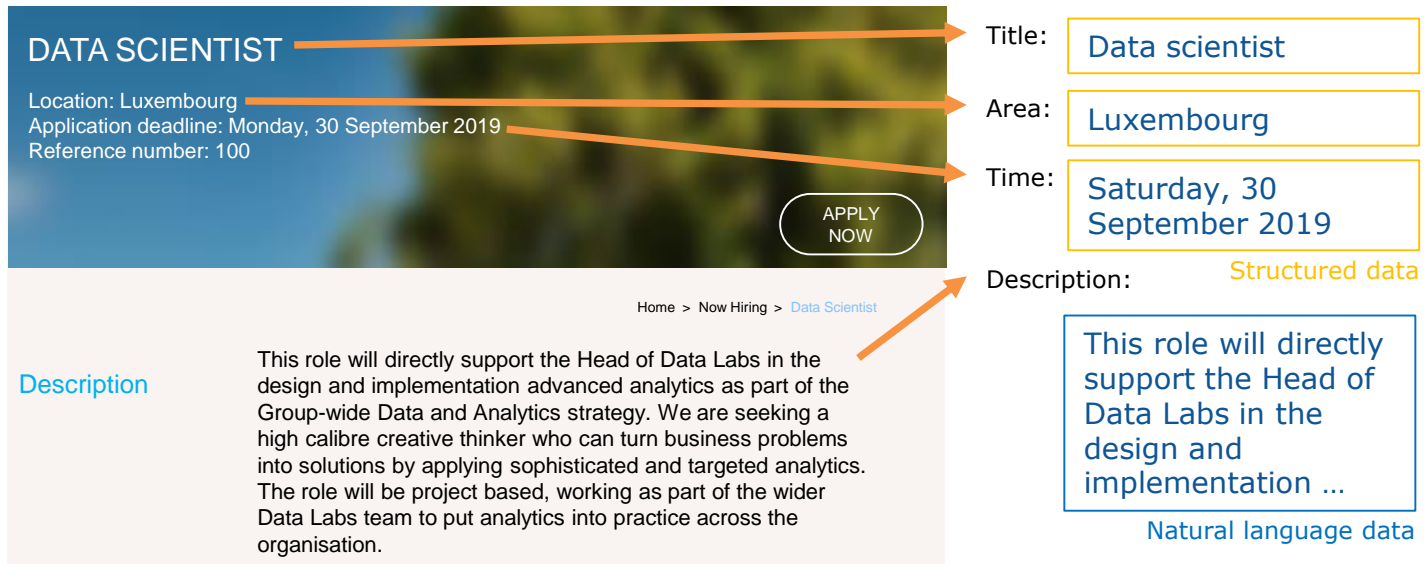
Virtual conference

How Online job advertisement data can add to our understanding on labour market trends?

How to move from research activities or one off exercises to regular production of information for various users?

Extracting data from an OJA

Web data gathering: extracting data from web pages



The image shows a job advertisement for a Data Scientist. The advertisement is divided into two main sections: a header with key details and a description section. Orange arrows point from specific text in the advertisement to corresponding data extraction boxes on the right. The extraction boxes are labeled 'Title:', 'Area:', 'Time:', and 'Description:'. The 'Description:' box is further categorized as 'Structured data' and 'Natural language data'.

DATA SCIENTIST

Location: Luxembourg
Application deadline: Monday, 30 September 2019
Reference number: 100

APPLY NOW

Home > Now Hiring > Data Scientist

Description

This role will directly support the Head of Data Labs in the design and implementation advanced analytics as part of the Group-wide Data and Analytics strategy. We are seeking a high calibre creative thinker who can turn business problems into solutions by applying sophisticated and targeted analytics. The role will be project based, working as part of the wider Data Labs team to put analytics into practice across the organisation.

Title: Data scientist

Area: Luxembourg

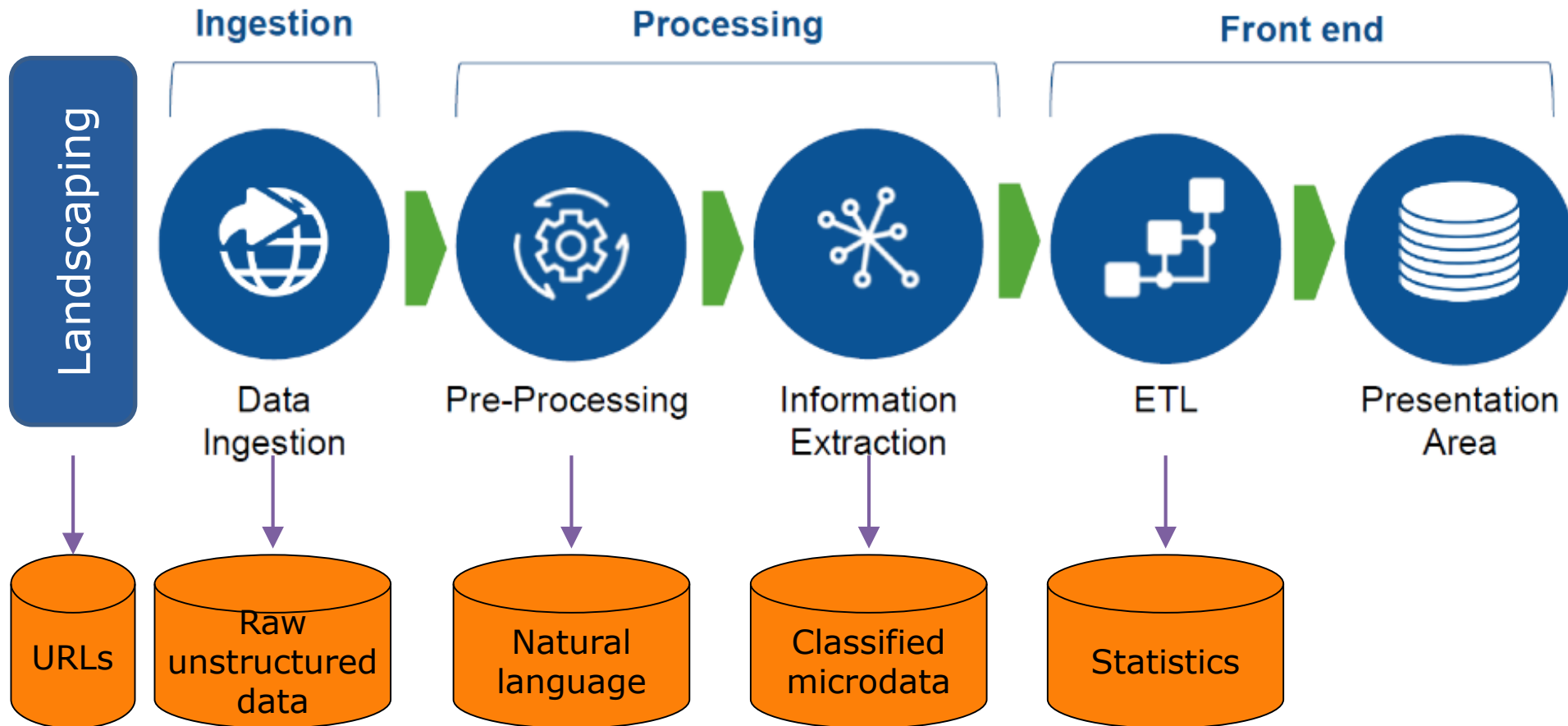
Time: Saturday, 30 September 2019

Description: Structured data

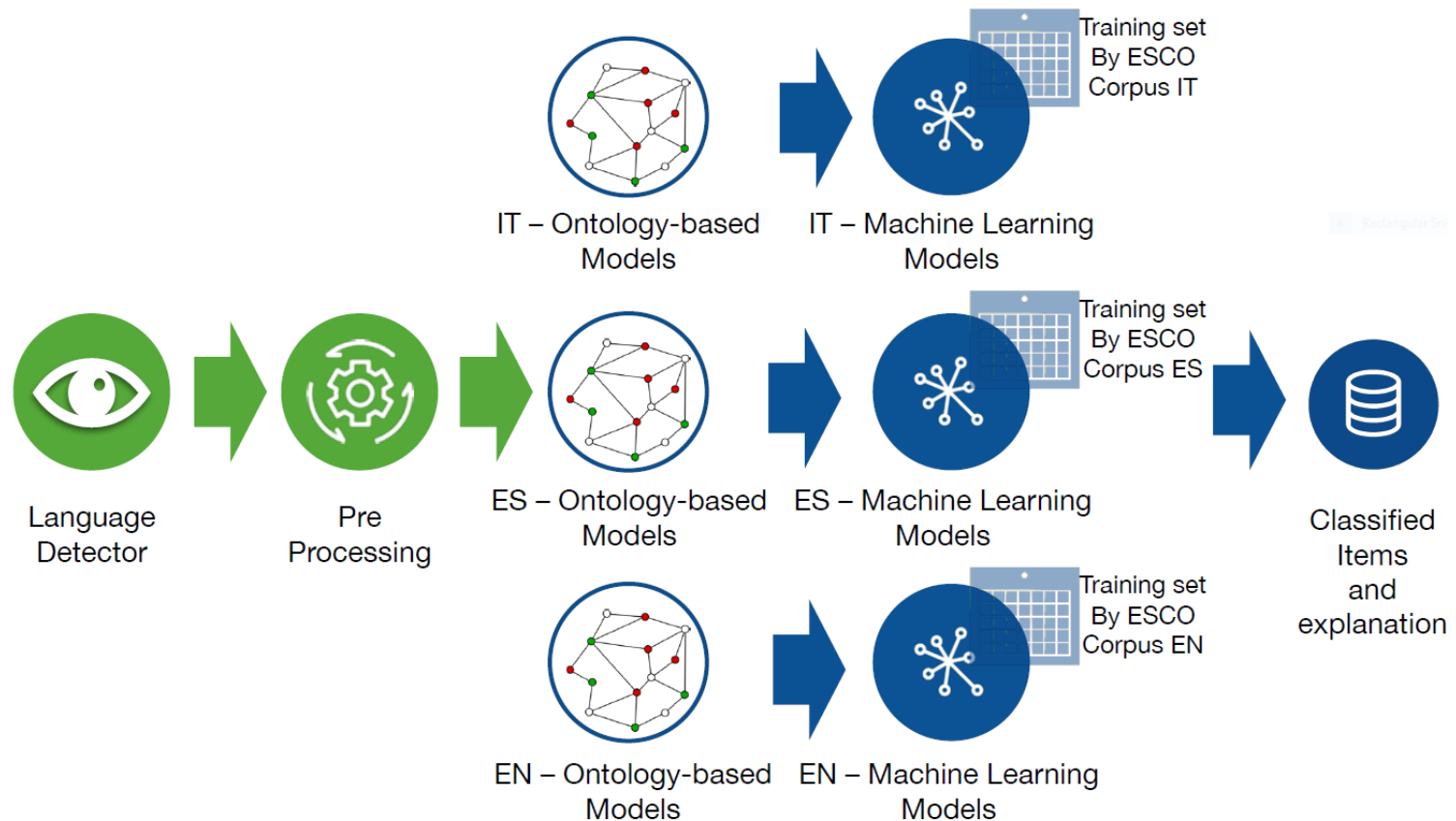
This role will directly support the Head of Data Labs in the design and implementation ...

Natural language data

From Web data to statistics pipeline



Extracting data from an OJA natural language



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**

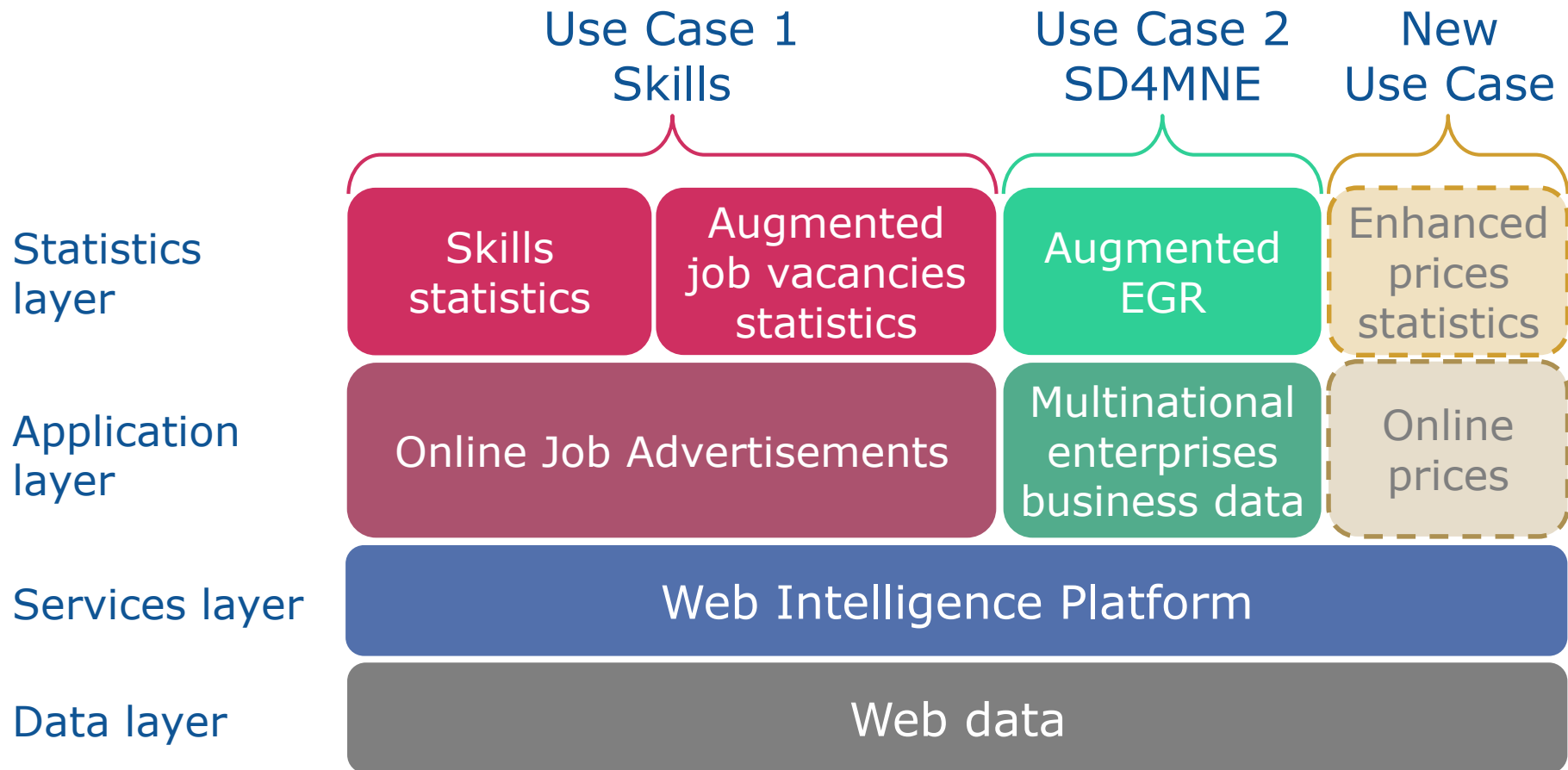
The Web as a statistics data source

- 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

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**



European
Commission

Users

Cedefop

Eurostat

NSIs

IT
experts

Labour market
experts

Network of
National Experts





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/>