BIG DATA
LABOUR MARKET INTELLIGENCE

8, 10 and 15 June 2021
DAY 3
15 JUNE
2021
• Training Programme on Big Data for Labour Market and Skills Intelligence

• Day 3: Presentation area. Use cases. ESCO

• Case study: ESCO – European Skill/Competences Qualifications and Occupations. Taxonomy, novelties of the skills hierarchy, new terms
AGENDA DAY 3: 15 JUNE

9.00-9.20: Opening, introductions, Poll

9.20-11.15 Training (I)
System components: Visualisation
Show-cooking: visualising occupation similarity

11.15-12.00:
ESCO
Mr Dimitrios Pikios, DG EMPL

12.00-12.20: Coffee break

12.00-13.15: TP (II)
Show cooking: Tableau and exercise
Hands-on: analyse Web LM data

12.00-14.15:
Sharing results
Final feedback
Closure
TOPICS OF THE TRAINING

• Visualisation - Presentation area
• Use cases: 10 cases!
• Lab session
• Case Study: ESCO
QUESTIONs

1. Big Data for LMI: what’s the most reliable data source compared with other accessible data sources?
2. Five steps of the approach to developing the OJV system (OJV-AI). Please mark the correct steps from the list.
3. Which are the data sources that we use in the system?
4. We learned about Eurostat Smart Statistics. What is it?
QUESTIONS OF THE POLL

1. Q1: “Data Lake” and “Data Warehouse” – are they the same?
2. Q2: Which features better define “Data Lake”
3. Q3: In the online job vacancy labour market what changes more rapidly?
4. Q4: Skills are the Genome of Jobs. Why is this notion important?
WELCOME TO DAY 2

- Training Programme on **Big Data for Labour Market and Skills Intelligence**
- Day 2: Processing – Build a Pipeline
- **Case study:** Real Time LMI – for career paths and informed choices
AGENDA DAY 2: 10 JUNE

9.00-9.20: Opening, introductions, Poll

9.20-11.15: Training presentation (I)
   Recap
   Data Pipeline
   Storage Layer (i)

11.15-11.30: Q&A

Till 12.00: Coffee break

12.00-13.30: TP (II)
   Spark Foundations
   Databricks
   Design a pipeline

13.45-14.30: Case Study: Data for careers’ paths, sourcing talent.
   Cyberseek
TOPICS OF THE TRAINING

1. Overview & Recap
2. What is a pipeline?
3. Storage layer
   a) Staging area
   b) Metadata
   c) Datalake
4. Spark foundations
5. Databricks (intro)
6. Design a pipelines from zero to solid
   a) The dataset
   b) Build our pipeline with Spark
OUR POLL
QUESTIONS OF THE POLL

1. Big Data for LMI: what is the added-value compared with other approaches and sources?

2. Five steps of the approach to create the Data system (OJV-AI). Please choose the 5 correct steps from the list.

3. Which are the data sources NOT used in the system?

4. We learned about Eurostat Smart Statistics. What is it?
MOTION

TO SESSION 1, WITH MAURO AND ALESSANDRO
DAY 1
8 JUNE 2021
WELCOME

• To the Training Programme on **Big Data for Labour Market and Skills Intelligence**
• Participants from Africa, Europe and Middle East
• Intensive journey: 5h x 3 days
• Knowledge and discovery-rich
• Presentations x discussion x questions and answers x hands-exercises
• 2 case studies
• Team: Alessandro, Mauro, Eduarda, Alina
• Thanks to interpreters’ team and team Pomilio
Big Data for Labour Market Intelligence: Online job vacancy analysis (OJV)

Virtual training programme

Full programme: 8, 10 and 15 June 2021

Multilingual: English-French-Russian

Agenda:

Day 1: 08/06/2021, 09.00-14.30 CET (GMT+2)
- LMI – based on Webdata: from online job vacancies to data and analysis of skills, occupations, sectors. Architecture, data ingestion techniques, data processing pipeline, classification techniques.
- Case study: Smart Statistics - Eurostat

Day 2: 10/06/2021, 09.00-14.30 CET (GMT+2)
- LMI – based on Webdata: Focus on building a data pipeline
- Case study: Big Data – career progression

Day 3: 15/06/2021, 09.00-14.15 CET (GMT+2)
- LMI – based on Webdata: Presentation area, use cases, hands-on exercises and lessons learned
THE 3-DAY JOURNEY HELPS US DISCOVER...

• What is Big Data analytics for Labour Market Information Systems?

• How to explore and create value from large volumes of Online Job Vacancies for real-time LMIS?

• Which dimensions and issues related to skills, occupations, and labour market dynamics questions can be analysed with help of Big Data / Real-time LMI?

• Which are the requirements to harness these novel data sources and systems by ETF Partner countries?

• How can Real-time LMI based on Big Data and AI-aided classification be used and complement established national statistics?
AGENDA DAY 1: 8 JUNE

9.00-9.30: Opening, Introduction, Poll

9.30-9.45: ETF Big Data Project

9.45-11.30: Training – presentation, Q&A (I)

Till 12.00: Coffee break

12.00-13-45: Training – presentation, Q&A (II)

13.45-14.30: Case Study – Eurostat Smart Statistics
GETTING ACQUAINTED…

USING A POLL
ETF PROJECT
BIG DATA FOR LMI

- Data ingested: up 31/03/2021
- Update: up 31/05/2021

Labour Market Intelligence - Ukraine

Number of job vacancies collected:
793,290

Number of job vacancies deduplicated:
458,434

Number of unique Vacancies by Wall Source:

Distribution by Release Date (date of publication of the OIV)
Social media manager, Airbnb host, influencer, SEO specialist, app developer, Uber driver, driverless car engineer, podcast producer and drone operator…

- These are just some of the jobs that did not exist 10 years ago.
- What will happen in the future? What kind of jobs will disappear, what will be created and why?
- Which new skills will be essential in the job market?
- How much is digitalisation and greening changing our occupations, tasks, skills?
Data Torrent + AI algorithms = Computing Anytime, Anywhere

Big Data Era
“Artificial intelligence (AI) and machine learning are not only changing the labour market, but also giving us new tools for analysing the workforce.”

- Terabytes of data from online job vacancies! Job vacancies or job advertisements are published, refreshed, updated in large numbers through websites of different types, size and coverage.
- Exploring the inherent information of a such large data source has become an objective of research centres and public bodies in a number of countries. EU OVATE (Cedefop) is a major breakthrough!
- EUROSTAT: embracing Smart Statistics.
- These vast and fast data sources are essential to understand the transformation of jobs (as bundles of tasks and skills), of changing forms of work and employers’ recruitment practices in the context of Covid-19 (and Beyond…)
ETF PROJECT BIG DATA FOR LABOUR MARKET INFORMATION

2018-2021

- **2019: First application**: Feasibility analysis – Landscaping of Web Labour Markets Tunisia and Morocco
- **2019**: A training programme for experts of the “Make it Match Network” of Eastern Partnership
- **2020**: Creation of the complete data system supported by AI with data visualisation platform (dashboard): Tunisia and Ukraine;
  - Analytical reports: LM and skills Ukraine and Tunisia
- **2021**:
  - New country – Georgia;
  - Green dashboard;
  - ETF works with the data analytics specialists of University Milano-Bicocca and Burning Glass Technologies.
  - The data system is based exclusively on demand – based on job vacancies (OJV) posted on web portals
  - Full comparability with the Real-Time data system of the EU-27 (same methodology)
Let the data speak

AI-aided data system

Data from OJV

Classifications ESCO, ISCO

Volume, Velocity, Variety, Veracity, Value

Dashboard and variables

Main elements
## DATA SOURCES LMI: SOME CHARACTERISTICS

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Type</th>
<th>Data model</th>
<th>Analysis paradigm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical conventional (surveys)</td>
<td>Structured</td>
<td>Relational</td>
<td>Top-down &amp; model based</td>
</tr>
<tr>
<td>Administrative (registers)</td>
<td>Structured or semi-structured</td>
<td>Relational</td>
<td>Top-down &amp; model based</td>
</tr>
<tr>
<td>Web data</td>
<td>Structured, semi-structured or unstructured</td>
<td>Relational and non-relational (NoSQL)</td>
<td>Bottom-up &amp; data driven</td>
</tr>
</tbody>
</table>
PARADIGM SHIFT: FROM TOP-DOWN TO BOTTOM-UP
BIG DATA FOR DEMAND ANALYSIS - DIMENSIONS

SKILLS NEEDS: IDENTIFY AND ASSESS

SKILLS RATES

EMERGING SKILLS AND EVOLUTION OF THE SKILLS MIX

LANDSCAPE OF OJV WEBSITES AND DYNAMICS OF THEIR UTILISATION

SHARPEN THE QUESTIONS THAT CAN BE ANALYSES

EXPLORE THE COMBINATION OF BIG DATA AND CONVENTIONAL STATISTICS
DATA CLASSIFICATION – TAXONOMIES

- **Occupations:** ESCO/ISCO
- **Skills:** ESCO
- **Location:** NUTS and ISO
- **Educational Level:** ISCED
- **Sector:** NACE
**Variables:**
1. Occupation
2. Skills
3. Education level
4. Experience
5. Salary
6. Location
GAME CHANGER?

- New data source: vast, non-structured, inductive, continuous inflow data
- New data analysis and classification
- New paradigms - open our mind for different views on data
THE GOLD MINE OF OJV DATA:

REAL TIME, FAST AND GRANULAR DATA

THE LINKS THAT CAN BE DONE WITH OTHER DATA SETS (STATISTICAL) AND OTHER BIG DATA DATABASES

BUT NOTE! BIG DATA - COVERS PART OF THE LABOUR MARKET (SECTORS, OCCUPATIONS, QUALIFICATIONS)
THANK YOU