

**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

Registration: [link](#)

**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. New data, new analytics. Overview, methodology, uses.
- Case study

**Day 2: 10/06/2021, 09.00-14.30 CET (GMT+2)**

- LMI – based on Webdata: Focus on data ingestion, processing and classification, with AI-aided processes.
- Case study

**Day 3: 15/06/2021, 09.00-14.15 CET (GMT+2)**

- LMI – based on Webdata: use cases, hands-on exercises and lessons learned



ETF working with Burning Glass Europe

ETF Working together Learning for life

[Link to access](#)  
[Link of the Webpage](#)

## ETF project Big Data for Labour Market Intelligence (LMI)

Started in the second half 2018 the ETF innovation project elaborated first a brief methodological handbook “[Big Data for labour market intelligence: an introductory guide](#)” (2019). A [training programme](#) for experts of the “Make it Match Network” of Eastern Partnership was a success in November 2019. In 2020 the full data system for analysis of OJV was established, with data dashboards available: [Tunisia](#) and [Ukraine](#). The dashboards are updated with data from online job vacancies collected until 31 March 2021. The data system is focused on demand (online job vacancies – OJV).

The ETF Big Data for LMI project continues in 2021, with addition of a new country and the continuation of Online Job Vacancies (OJV) data ingestion, classification and analysis for Tunisia and Ukraine. Many new questions and queries can be analysed on the basis of this growing database of OJVs, as predicted by the paradigm “let the data speak”. New insights will be identified from the vast data set, and a “Green Dashboard” developed.

**The three-day journey of this virtual training programme is conceived to respond to these questions.**

- But 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?

**Participants:** representatives from research and analytical departments of ministries (education, training and labour) and from specialized research centres from all ETF partner countries. Participation in all activities planned in the three-day programme is recommended (mandatory) for all registered participants.

The three-day journey of this virtual training programme mobilises participants 4,5 hours a day and intercalates days for review and preparation of the thematic blocks of the training programme.

## Agenda of training programme

**Day 1: 08/06/2021: 09.00 – 14.30 (CET) – 07.00-12.30 GMT**

Time (CET)	Session	Speaker
09.00-09.30	Opening, welcome remarks. Objectives of the training workshop. Introductions of participants	Euarda Castel-Branco, ETF
09.30-09.45	Brief overview of ETF Project Big Data Project for LMI	Euarda Castel-Branco
09.45-11.15 Session 1	<b>Big Data for Labour Market Intelligence LMI</b> System Presentation and outcomes <ul style="list-style-type: none"> <li>• General overview</li> <li>• Challenges</li> <li>• Dataflow overview</li> </ul>	Alessandro Vaccarino Mauro Pelucchi
11.15-11.45	Poll. Questions and answers. Discussion	Euarda Castel-Branco Alessandro Vaccarino, Mauro Pelucchi
11.45-12.00	<b>Coffee break</b>	
12.00-13.30 Session 2	<b>Big Data for LMI</b> System description <ul style="list-style-type: none"> <li>• Ingestion</li> <li>• Pre-processing</li> <li>• Classification</li> </ul>	Alessandro Vaccarino Mauro Pelucchi
13.30-13.45	Questions and answers. Discussion	Alessandro Vaccarino Mauro Pelucchi
13.45-14.30	Case study: Eurostat – Smart Statistics. Focus on data from online job advertisements	Fernando Reis Eurostat

**Day 2: 10/06/2021: 09.00 – 14.30 (CET) – 07.00-12.30 GMT**

Time (CET)	Session	Speaker
09.00-09.15	Opening. Impressions from day 1.	Euarda Castel-Branco, ETF
09.15-11.00 Session 1	<b>Big Data for Labour Market Intelligence (LMI)</b> System Components <ul style="list-style-type: none"> <li>• Recap and practical application (Data ingestion and processing)</li> <li>• Show-cooking: scraping data from a corporate website</li> <li>• Questions</li> </ul>	Alessandro Vaccarino Mauro Pelucchi
11.00-11.30	Questions, clarifications. Poll.	
11.30-12.00	<b>Coffee break</b>	
12.00-13.30 Session 2	<b>Big Data for LMI</b> Lab <ul style="list-style-type: none"> <li>• Show-cooking: build a classification pipeline</li> </ul>	Alessandro Vaccarino Mauro Pelucchi
13.00-13.15	Questions and answers. Discussion	Alessandro Vaccarino Mauro Pelucchi
13.45-14.30	<b>Case study:</b> Real time LMI applications: skills adjacency, careers' progression and informed choices.	Speaker: Davor Miskulin, Burning Glass Technologies

**Day 3: 15/06/2021: 09.00 – 14.15 (CET) – 07.00-12.15 GMT**

**Technical requirements: Tableau Public installed on the workstation ([download link](#))**

Time (CET)	Session	Speaker
09.00-09.15	Opening. Recap from days 1 and 2.	Euarda Castel-Branco, ETF
09.15-11.00 Session 1 (EN)	<b>Big Data for Labour Market Intelligence (LMI)</b> System Components: data presentation <ul style="list-style-type: none"> <li>• Visualisation of data</li> <li>• Show-cooking: visualizing occupation similarity</li> </ul>	Alessandro Vaccarino Mauro Pelucchi
11.00-11.15	<b>Coffee break</b>	
11.15-12.00 <b>Special session</b>	<b>ESCO</b> – the taxonomy of skills and occupations that supports labour and learning mobility, and links with qualifications and employment.	Dimitrios Pikos ESCO – Directorate General Employment Social Affairs and Inclusion – European Commission
12.00-13.15 Session 2	<b>Big Data for LMI</b> <ul style="list-style-type: none"> <li>• Show-cooking: use Tableau Public and exercise presentation</li> <li>• Hands-on: analyse Web Labour Market data</li> </ul>	Alessandro Vaccarino Mauro Pelucchi
13.15-14.15	Sharing results and findings from session 2. Final feedback Closure of the training workshop	Euarda Castel-Branco Alessandro Vaccarino Mauro Pelucchi

## Background information

### Training programme: themes and outcomes

Day 1: 8 June	Day 2: 10 June	Day 3: 15 June
<p><b>Title:</b> Real-time Labour Market Intelligence (LMI): Overview, goals, advantages and use cases.</p> <p><b>Type of session:</b> Lecture (presentation)</p> <p><b>Case study:</b> additional session (30 min)</p> <p>Eurostat and smart statistics, use of web data (with focus on OJV) - perspectives, scope and requirements. Potentialities for other national statistical agencies?</p> <p>Speaker: EUROSTAT</p>	<p><b>Title:</b> Real-time LMI: Focus on data Ingestion, processing and classification</p> <p><b>Type of session:</b> Show cooking <u>Optional requirements:</u> Tableau Public, Google Colab, Databricks Community Edition</p> <p><b>Case study:</b> additional session (30 min)</p> <p>Real time LMI applications: skills adjacency, careers and informed choices.</p> <p>Speaker: Davor Miskulin, Burning Glass Technologies</p>	<p><b>Title:</b> Skill Intelligence: presentation of results; use cases, hands-on exercises and lessons learned.</p> <p><b>Special session:</b> ESCO – the European taxonomy of skills and occupations supporting mobility.</p> <p><b>Type of session:</b> Workshop</p> <p>Participants analyse data to deeply understand the kind of insights / information provided by LMI System based on Big Data.</p> <p><b>Description:</b> on the basis of practical examples, participants are invited to make brief data analyses using Tableau, or the preferred analysis tool.</p>

		Lessons learned and conclusions shared with the group. <u>Requirements:</u> Tableau Public
<b>Outcome:</b> participants acquire a full overview of LMI System based on Big Data (OJVs). From theory of Knowledge Discovery in Databases (KDD) - to data collection and processing phases.	<b>Outcome:</b> participants deepen understanding on the LMI System, showing how the components work.	<b>Outcome:</b> participants are able to: <ul style="list-style-type: none"> <li>- discuss and explore requirements for Big Data for LMI systems;</li> <li>- interpret and question data dashboards</li> <li>- propose questions to explore the full potentialities of Big Data for LMI</li> <li>- analyse skills and occupational features and trends based on Big Data.</li> </ul>

## Big Data for Labour Market Intelligence (LMI)

### A new world of data analytics...

Skills intelligence as business as usual is not enough to understand the direction and extent of the transformation of tasks, jobs, skills and qualifications prompted by a wave of drivers of change, which boosted the digitalisation of most processes in our societies. New data analytics have emerged to advance skills intelligence and complement conventional statistics, surveys and administrative data.

Data is being called the new oil. Digitalisation of processes, services, businesses, personal and social interactions generates a growing mass of data across the globe. Creating knowledge out of large volumes of data, available with high velocity and variety is the major goal of Big Data analysis.

### ...can be applied for labour market information

Artificial intelligence (AI) and machine learning are not only changing the labour market, but also giving us new tools for analysing the workforce. 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. These vast data sources are essential to understand the dynamics and functioning of Web Labour Markets, and of changing employers' recruitment choices.

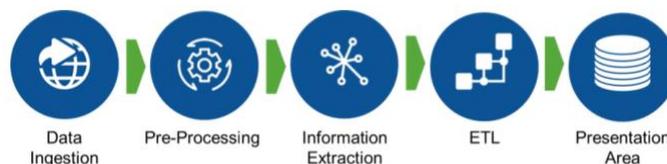
Big Data analytics can be used to map skills by occupations, to identify obsolete skills, to do predictive analysis of demand for new occupations and skills, and to better capture skills interactions - based on granularity of data and quasi in real time.

In the European Union, since 2016 Cedefop is leading a breakthrough project in this area and created a vast data system based on the analysis and classification of millions of online job vacancies (OJV) of European Union (EU) Member States. In the platform [OVATE](#) the results are presented in interactive dashboards of combined variables, and different geographic coverage. In 2021 Cedefop renewed and upgraded the OVATE dashboards and will continue analysis of occupation-skills-sector relationships. On the other side Eurostat takes over the data infrastructure part, to focus on statistics and detailed time trends and focus on territorial location. This new phase of the project is jointly steered by Cedefop and Eurostat.

### ETF project Big Data for LMI

ETF project started practical application in 2019 with a feasibility analysis of the web labour markets of Morocco and Tunisia, resulting in a comprehensive report assessing and ranking **online job vacancy** (OJV) portals. The establishment of an integrated system for data collection, processing, classification, analysis and visualisation was the core of the work in 2020, in two pilot countries (Ukraine and Tunisia).

### Schematic overview of the workflow and method of Big Data for LMI

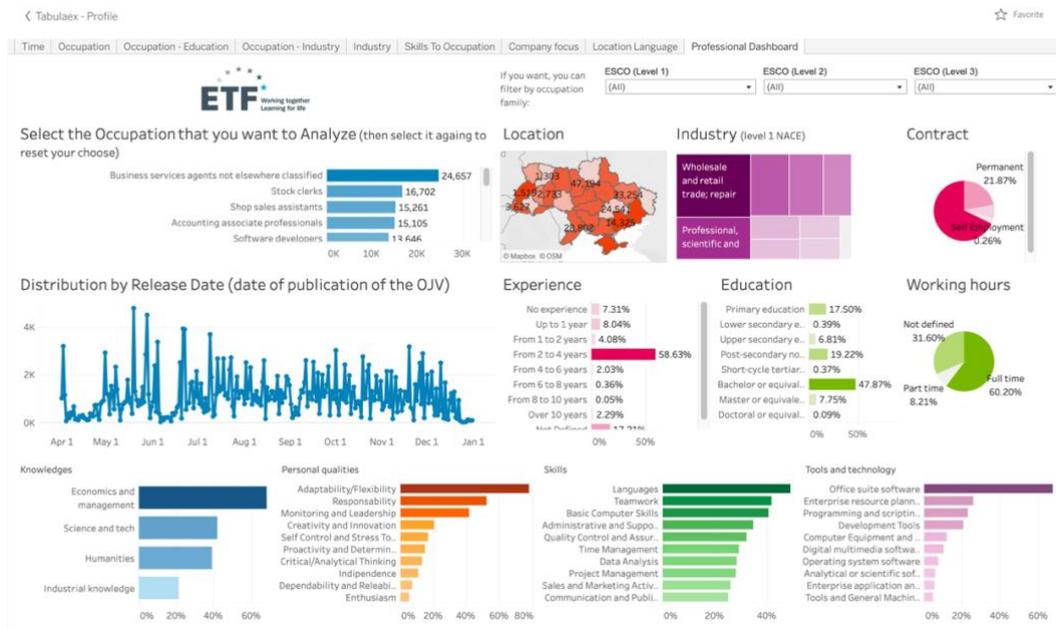


Key outcomes of ETF Big Data for LMI activities, including **data dashboards, training programmes, methodological handbook and analyses**, have been published at:

- Open Space webpage: [LMI in transformation – Focus on Big Data](#)
- Data dashboards: [Ukraine](#) and [Tunisia](#)
- Brief methodological handbook (English) “[Big Data for labour market intelligence: an introductory guide](#)” (2019).
- Brief Methodological Handbook, 2019 ([Russian](#))
- A specific [training programme](#) for data analysts and experts, November 2019
- Report: [Big Data for LMI, Web Labour Market Landscaping Ukraine](#), 2020.
- Report: [Big Data for LMI in Ukraine: analytical report](#). 2020.
- [Webpage of webinar on Big Data for LMIS](#), 10 December 2020

Working with the data science team of *University Milano Bicocca* and *Burning Glass Europe* (Italy), the European Training Foundation (ETF) has completed in December 2020 a deciding phase of its innovation project “Big Data for Labour Market Intelligence”. Hundreds of thousands of online job vacancies collected over 8 consecutive months in 2020 (April-December), processed and automatically classified against such international classifications / taxonomies as ISCED 2011, NACE, NUTS / ISO and ESCO provide unique insights on skills and occupational features and dynamics of the [Tunisian](#) and [Ukrainian](#) labour markets. We say “unique insights” – because of their granularity and real time nature. Some of the many possible angles of analysis are visualized in two countries’ dashboards.

**Figure 1: [Ukraine professional dashboard](#): a multi-dimensional view of variables in one snapshot**



The data system uses ESCO as the reference for machine-classification of skills identified in the hundreds of thousands of OJVs. For the case of Tunisian OJV data we used **ESCO** – in French and English versions. But for the case of the Ukraine, an additional step was indispensable: translation of **ESCO skills into Russian and Ukrainian** languages (over 4,000 terms).

The particular advantage of OJVs as sources for LMI lies with the fact that they express / represent the employers' determination of the profiles they need for the purposes of the business or activity in a given period. The machine processing and classification of employers' own terms and descriptions of skills shows cases of OJVs skills without a direct ESCO correspondence. No surprise: technological and digital transformation of work and skills is much faster than the pace of alignment and update of ESCO. What to do then? The data science team involved in the project applied machine-learning techniques (e.g., Word2Vec) to enhance ESCO skills, creating a correspondence between a new 'non-ESCO' skill with a close (approximate) ESCO-skill. This process and the machine-proposed correspondence is discussed and validated by (human) professionals in the given sector, occupation and technology. Can this technique and approach have a wider application in the context of ESCO updates?

27/04/2021. Contact: Eduarda Castel-Branco - [ecb@etf.europa.eu](mailto:ecb@etf.europa.eu)

### Technical requirements

## Databricks Access procedure

Follow steps below to create a free account to access Databricks *Community Edition*:

1. Go to <https://databricks.com/try-databricks>
2. Compile the form. Insert a valid mail address, where the activation link will be sent

# Try Databricks

An open and unified data analytics platform for data engineering, machine learning, and analytics

From the original creators of Apache Spark™, Delta Lake, MLflow, and Koalas

Tell us a little about yourself to get started.

* First Name: Mauro	* Last Name: Pelucchi
* Company Name: Unimib	* Work Email: mauro.pelucchi@unimib.it
* How would you describe your role?: Data Scientist	* What is your intended use case?: Education - Learning & Teaching
Phone Number: +	

By Clicking "Sign Up", you agree to the [Privacy Policy](#)

Keep me informed with occasional updates about Databricks and Apache Spark™.

SIGN UP

### 3. Select "Community Edition"

## COMMUNITY EDITION

For students and educational institutions

- Single cluster limited to 6GB and no worker nodes
- Basic notebooks without collaboration
- Limited to 3 max users
- Public environment to share your work

GET STARTED

By clicking "Get Started" for the Community Edition, you agree to the [Databricks Community Edition Terms of Service](#).

4. Check your mailbox for the activation mail. Within it, click on the activation link and follow instructions.

5. Check access to Databricks using the following url:

<https://community.cloud.databricks.com/>

Login with provided mail and password

6. Verify that the user is correctly activated and the system is up and running, as below:



# Welcome to databricks

## Featured Notebooks



[Introduction to Apache Spark on Databricks](#)



[Databricks for Data Scientists](#)



[Introduction to Structured Streaming](#)

## New

-  Notebook
-  Job
-  Cluster
-  Table
-  Library

## Documentation

-  [Databricks Guide](#)
-  [Python, R, Scala, SQL](#)
-  [Importing Data](#)

## Open Recent

Recent files appear here as you work.  
Get started with the welcome guide.

## What's new?

- Jobs support concurrent runs
- GPU instance types are available
- Github commit messages support international characters

[Latest release notes](#)

[Send Feedback](#)