Big data for Labour Market Intelligence (LMI)

A new world of data analytics…

'Business-as-usual' skills intelligence is not sufficient to understand the direction and the extent of the transformation of tasks, jobs, skills and qualifications. This transformation has been prompted by a wave of drivers of change that have boosted the digitalisation of most processes in our societies.

Data is referred to as 'the new oil'. The digitalisation of processes, services, businesses, and personal and social interactions generates a growing mass of data across the globe. Extracting knowledge from large volumes of data, generated at high velocity and with great variety, is the major goal of big data analysis.

New data analytics have emerged to advance skills intelligence and complement conventional statistics, surveys and administrative data. EUROSTAT is implementing Trusted Smart Statistics and starting up the Web Intelligence Hub with a number of partner institutions. Smart Statistics can be seen as the future extended role of official statistics in a world impregnated with smart technologies. Smart technologies involve real-time, automated, interactive technologies that optimise the physical operation of appliances and consumer devices. Statistics themselves would then be transformed into a smart technology embedded in smart systems that would transform 'data' into 'information'. However, using web data for official statistics is not easy, as it requires: infrastructure with big data capabilities, specialised skills, data agreements, spreading web intelligence capabilities among statistical services. Read more here.

…can be applied to obtain labour market information

Artificial intelligence (AI) and machine learning are not only changing the labour market, but are also giving us new tools to analyse the workforce. Job vacancies or job advertisements are published, refreshed and updated in large numbers through websites of different types, size and coverage. Exploring the inherent information in such a large data source has become an objective of research centres and public bodies in a number of countries. These vast data sources are essential for understanding the dynamics and functioning of Web Labour Markets and how employers' recruitment choices are changing.

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

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

ETF project: Big Data for LMI

The practical application of the ETF project began in 2019 with a feasibility analysis of the web labour markets in 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 the Big Data for LMI project

Data Ingestion → Pre-Processing → Information Extraction → ETL → Presentation Area
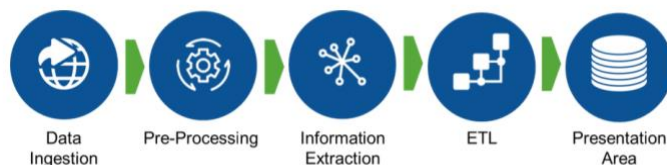
The key outcomes of the ETF's 'Big data for LMI' activities, including **data dashboards, training programmes, methodological handbook and analyses**, have been published at:

- 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 the University of Milano-Bicocca and Burning Glass Europe (Italy), the European Training Foundation (ETF) completed a deciding phase of its innovation project 'Big data for Labour Market Intelligence' in December 2020. Hundreds of thousands of online job vacancies collected over eight 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 can be visualised in the two countries' dashboards.

Ukraine professional dashboard: a multi-dimensional view of variables in one snapshot
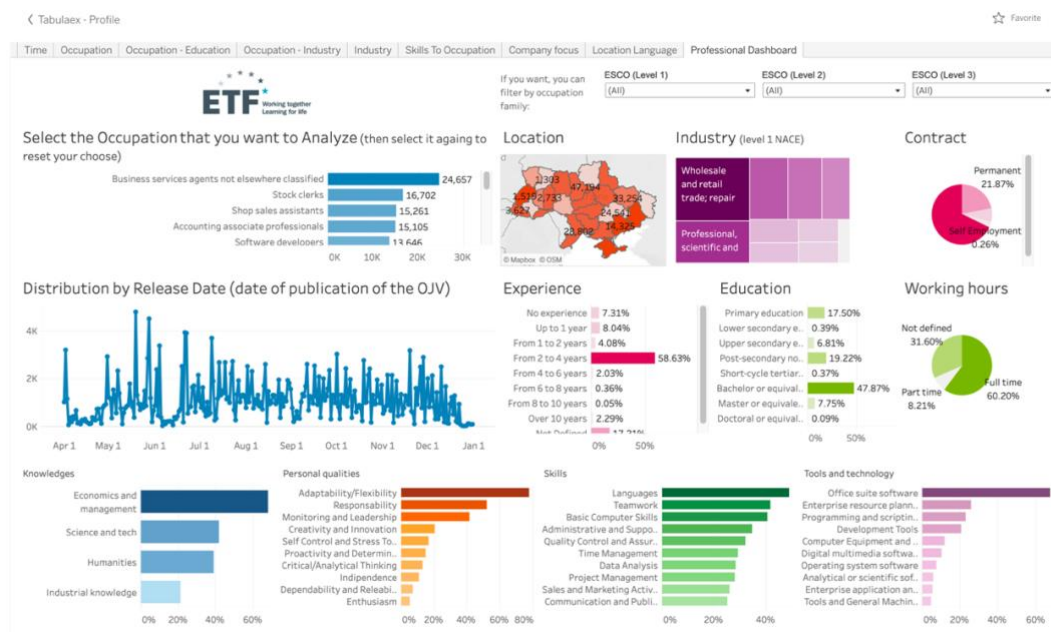
The data system uses ESCO as the reference for the machine classification of skills identified in the hundreds of thousands of OJVs. For the Tunisian OJV data we used **ESCO**, in French and English versions, but for Ukraine, an additional step was indispensable: the translation of **ESCO skills into Russian and Ukrainian** (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: the technological and digital transformation of work and skills is much faster than the pace of alignment and updating 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?

Contact: Eduarda Castel-Branco - ecb@etf.europa.eu