LABOUR MARKET INTELLIGENCE – SOME NEW IDEAS BUILDING ON INNOVATION, WEBDATA AND DATA SCIENCE

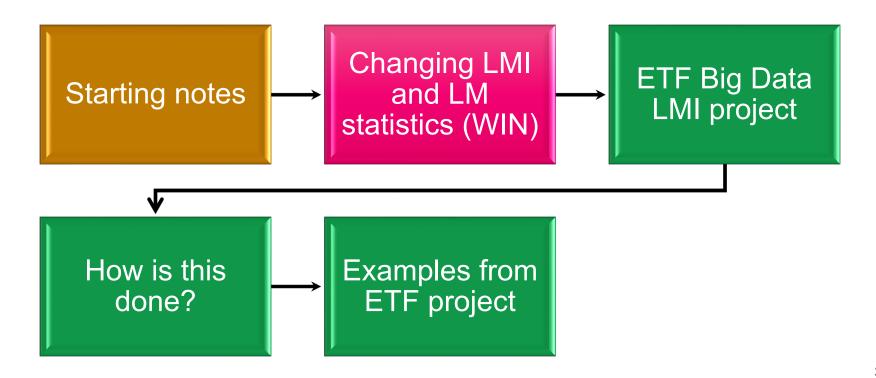
EDUARDA CASTEL-BRANCO - ETF

17 OCTOBER 2022





TOPICS...

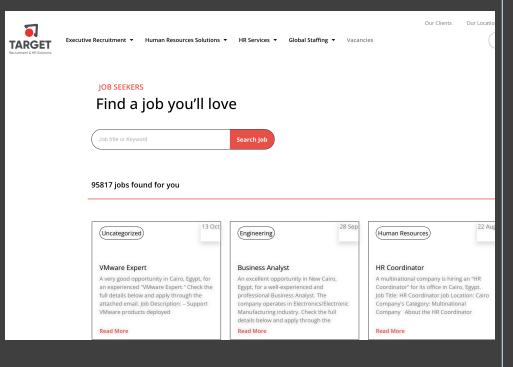




STARTING NOTES

- The current activities of ETF Big Data LMI project are focused on OJV analysis (demand analysis).
- But...Al-aided data analytics and webdata offer wider perspectives for analysis, insights and trend-discovery related with supply side, which ETF is exploring.
- Aspects of the data system can be tailored, adjusted to your needs, e.g.: key variables, data visualization and data delivery (concept & tools)
- Building and maintaining the OJV analysis system: skills and technologyintensive
- To ensure quality, stability and credibility of the databases and dashboards ETF works with an experienced data science team, and a tested and continuously monitored methodological framework.
- Access of relevant national organisations to the databases can be planned and organised, and ETF will provide the necessary technical support to ensure clarity, quality and efficiency.
- Capacity development, dissemination and debate is important: annual programme (next: Nov-Dec 2022: 2 modules x 3 webinars).

ONLINE JOB ADVERTISEMENTS



- Online job advertisements (OJAs) refer to advertisements published on the internet (WWW) revealing an employer's interest in recruiting workers with certain characteristics for performing certain work.
- This could be motivated by the employer's need to fill a current vacancy, by an exploration of potential opportunities, or other reasons. OJAs usually include data on the characteristics of the job (e.g. occupation and location), characteristics of the employer (e.g. economic activity) and requirements (e.g. education/skills).
- Part of this information is available only as natural language textual data.
- This type of big data requires specific methodologies for processing and analysis but also provides much more detailed information (compared to alternative data sources) and avoids pre-conceived classifications (e.g. important to identify emerging skills).

INNOVATION IN LMI AND IN LABOUR MARKET STATISTICS

DRIVERS OF CHANGE



Digitisation - Automation - Al

Covid-19

Greening of economy and society

Societal movements: rights, democracy, information, innovation, geopolitics

- The future of work arrived ahead of schedule
- "2 years of digital innovation in 2 months"
- Hybrid / Remote work: lasting post-Covid19
- Changing mix of jobs x skills mix
- Reskilling: 50% of working population 2025 (WEF)
- Learning anytime, anywhere
- Hybrid skills rising
- New types anf forms of qualifications: microcredentials, digital certificates
- Wave of innovation
- Digitally enabled productivity gains accelerate 4th Industrial Revolution

'Double-disruption' scenario for workers

Tandem Covid-19 recession x automation & digitalisation

LABOUR MARKET / SKILLS INTELLIGENCE MORE IMPORTANT THAN EVER

Wide range of users LMI: Macro-meso-micro levels

- National, regions, economic-professional sectors, education and training community, employers, learners
- Research and analysis (departments, observatories, analytical centres)

Combination of methods: ST-MT-Long Term / macro-meso-micro levels

Benefits / questions:

- Inform wider public and inform policy (re)-orient policies and related implementation measures
- Trends in demand, supply: what is happening now and what are the directions of change
- New / changing occupations / jobs profiles new / changing skills
- Serious bottlenecks, shortages, gaps in the **crisis conjuncture** how to solve...
- LLL: upskilling, reskilling, company-training, validation of acquired skills (skills adjacency)
- Vulnerability in LM transitions due to skills / qualifications gaps

LMI IS CHANGING...

"Traditional"

Innovative / Changing

Granular x Real

1. Data sources

Structured conventional statistical data

+ new sources ("datified society") administrative and Big Data (Internet) – Web scraping

2. Data analysis

Statistical; specialised studies (sectoral, occupational)

+ Al-aided methods, data science analytics

3. Data delivery

Tables, Databases, Reports

+ Web intelligence shared systems / API / online interactive dashboards / tailored for users; regional / multi-country platforms

Working together
Learning for life

European Training Foundation

Eurostat "Trusted Smart Statistics" + Cedefop – Web Intelligence Hub

TRANSFORMATION OF LABOUR MARKET INFORMATION / INTELLIGENCE

- We live in a "datified" world
- "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 / resources of different types, size and coverage
- These vast and fast data sources are useful 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...)
- Exploring the inherent information of a such large data source has become an objective of research centres and public bodies in the EU and other regions.
 - > EU Skills OVATE (Cedefop): https://www.cedefop.europa.eu/en/tools/skills-online-vacancies
 - ➤ **EUROSTAT:** building "Trusted Smart Statistics" established the Web Intelligence Network: https://ec.europa.eu/eurostat/cros/WIN_en







EUROSTAT

Source: Presentation by F. Reis at ETF "Big Data LMI capacity development programme", June 2021



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
- Big Data II project
- 2020-2021: Intelligence Hub launched Web Intelligence Network (WIN)

EUROSTAT

Source: Presentation by F. Reis at ETF "Big Data LMI capacity development programme", June 2021



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

EUROSTAT: TRUSTED SMART STATISTICS

• The Eurostat Trusted Smart Statistics – Web Intelligence Network (WIN) project has evolved from previous activities implemented between 2016-2020 (ESSnet Big Data I and ESSnet on Big Data II projects (multi-purpose statistics based on non-traditional data sources). The project contributes to the establishment of the Web Intelligence Network (WIN) across the ESS and beyond. It aims at a comprehensive and coordinated approach to building the web intelligence system at the ESS level, providing a greater chance to generate the right conditions for the integration of web data into official statistics, rather than via individual NSI efforts. https://ec.europa.eu/eurostat/cros/WIN en

4 work packages:

- Work Package 1: coordination, support, dissemination;
- Work Package 2 Advancement of the WIHub, and moving the online job advertisements (OJA) and online-based enterprise characteristics (OBEC) use cases into statistical production stage
- Work Package 3 Exploration of the potential to extend the WIH by new data sources and use cases
- Work Package 4 The development of solid methodological and quality foundations for generating statistics within the WIHub

EUROSTAT: TRUSTED SMART STATISTICS

Why exploring online job advertisements?

- Innovation in official statistics has been further stimulated by the recent health crisis, which has generated new opportunities and requests to provide new indicators to monitor the economic, social and environmental impacts of the pandemic. OJAs, which provide a real chance to support and deepen labour market statistics, offer one of these opportunities, created through collaboration with other governmental and private institutes.
- Eurostat is doing applied research **demonstrating possible uses of this data source** to show its potential and possibly inspire future use cases. Online job ads are available at Eurostat and the European Statistical Systems thanks to an administrative **agreement reached in 2020 with CEDEFOP**.
- OJAs are a powerful source of information on job requirements, which are not available and is difficult to gather by the current official statistical sources and methods.
- While providing new analytical opportunities, online job advertisements do not cover all job offers.
 Data on OJAs cannot replace other sources of labour market information, but they can complement
 them by providing comprehensive, detailed, and timely insights into labour market trends and
 allowing the early identification of new emerging jobs and skills.
- At least two applications of OJA data that may support and deepen labour market statistics are
 especially promising: (i) using job offers as a leading indicator of the labour market situation, (ii) providing
 structural and qualitative information at a highly granular level, for example on skills.

EUROSTAT: STUDY ON URBAN LABOUR MARKET CONCENTRATION (A USE CASE OF OJA DATA) – 2021 (1)

Study "Competition in urban hiring markets: evidence from online job advertisements" (2021)

- Study on urban labour market concentration index calculated using online job ads. There are currently no statistics available on labour market concentration at European level. But, there is a growing interest in measuring concentration of labour markets since it may imply limited competition among firms, which in turn can drive down workers' bargaining power and deteriorate job conditions. The study provides evidence of the level of labour market concentration across nearly all occupations and for every functional urban area (FUA) of the 27 EU MS, using OJA data for 2019 and 2020. The Herfindahl-Hirschman Index (HHI) is used to calculate the concentration for labour markets at the occupation (ISCO level 4), functional urban area and quarterly level.
- The results indicate that the largest urban areas in Europe tend to have lower level of concentration of the hiring market, indicating more competition among employers and more job opportunities forworkers across all occupations. This is also confirmed by migration trends that show how theseurban areas attract more people in search for better job conditions.
- Some occupation types appear to be more concentrated than others on average, but this may be partly due to the fact that some occupations are more frequently advertised online than others. With respect to time series, an average increase in labour market concentration can be seen in the second quarter of 2020, when the pandemic crisis hit Europe stronger.

EUROSTAT: STUDY ON URBAN LABOUR MARKET CONCENTRATION (A USE CASE OF OJA DATA) – 2021 (2)

- This paper presents the first experimental results on the use of Eurostat's Online Job Advertisements
 (OJAs) data for estimating an urban labour market concentration index, together with the underlying
 methodology.
- In the coming years, a further improvement of the quality of the OJA data is foreseen. This will be combined with more detailed work on quality assessment and reporting in the use of web data for statistical production (including a set of quality indicators), with the intention of aligning the OJA database with the standards of official statistics.
- In conclusion, the contribution of this paper is twofold:
 - It describes a new source of data, addressing some of the challenges related to its application to policy-relevant work.
 - 2. It provides the first Europe-wide evidence on demand-side labour market competition in urban labour markets. The paper also provides directions for future work to improve the accuracy and reliability of both the data sources and the statistical outputs.

BIG DATA FOR LMI

ETF PROJECT



Main elements of ETF OJV analysis system

4. Dashboard and variables

> 2. Al-aided data system

Big Data Era

5. Let the data speak

International Torrent algorithms Comp ESCO, ISCO Anytime

> 1. Data from OJV

BIG DATA FOR LABOUR

MARKET INTELLIGENCE

INTRODUCTORY GUIDE

Classifications

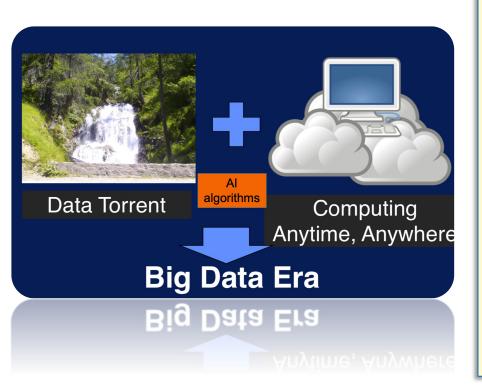
THE CHANGING NATURE OF WORK

6. Volume, Velocity, Variety, Veracity, Value



ETF PROJECT

BIG DATA FOR LMI 2018-2021 (PILOT PHASE)



- 2018-2019: Methodology: first step brief methodological handbook "Big Data for labour market intelligence: an introductory guide" (published in 2019).
- 2019: First application: Feasibility analysis Landscaping of Web Labour Markets Tunisia and Morocco
- **2019-2021:** 3 main training programmes for experts of the partner countries and other regions(Asia, Africa)
- 2020: Creation of the complete OJV analysis system and dashboards: Tunisia and Ukraine
 - Analytical reports: LM and skills Ukraine and Tunisia
- 2021:
 - New country Georgia;
 - Green dashboard;
- 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)
- ETF works with the data analytics specialists of University Milano-Bicocca and LightCast

ETF PROJECT BIG DATA FOR LABOUR MARKET INFORMATION

2022-2024: NEW PHASE

- Continue, maintain, improve the 3 existing country OJV systems
- Expand to new countries: Egypt, Kenya; Algeria, Morocco...
- Egypt:
 - Create the system architecture, data inflow, classification, visualization tool
 - New: OJVs in Arabic language, integration with national LMI platforms; green and digital skills analysis – for Future of Work forum 2023

General:

- Improved data delivery and dashboards
- Analysis of new phenomena / features on demand side: green skills, remote work, dynamics of web LM...
- 2 new concept papers: a) OJV data in the wider LMI statistics elements for a methodological approach; b) webdata – to analyse the supply side
- Capacity development, dissemination of results, debate





ONLINE JOB VACANCIES: WHY?

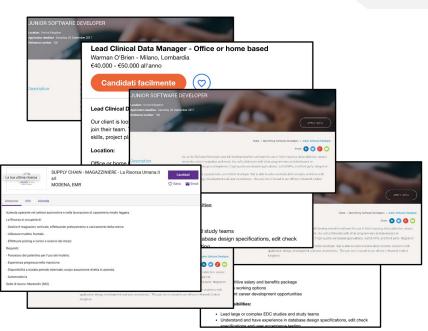
- OJV represent demand of employers / enterprises
- OJV express / represent employers' determination of the profiles they need for the purposes of the business or activity in a given period.
- Up-to-date: employers publish job vacancies when they need to recruit workers
- Detailed: most OJV describe the need as specifically as possible in terms of:
 - Occupation / job
 - Requirements (skills, experience, level of education / qualification)
 - Work context (place, contract, sector, working hours...)
- Close to reality: labour market / sector terms are used for the jobs and the required skills.
 These terms can contribute to identify new developments in the LM, such as emerging jobs and skills.

Use this information / data to better understand the evolution of the labour market in a country and compare to other countries. Compare, complete with other data

THE GOAL IS OF OJV ANALYSIS IS...

To transform this...

...into value







DATA CLASSIFICATION - TAXONOMIES

- Occupations: ESCO/ISCO
- Skills: ESCO
- Location: NUTS and ISO
- Educational Level: ISCED 2011
- Sector: NACE





HOW IS THIS DONE?



KEY COMPONENTS OF THE APPROACH

Data ingestion:
collect raw data from
OJV in both structured
and unstructured (raw
text) formats

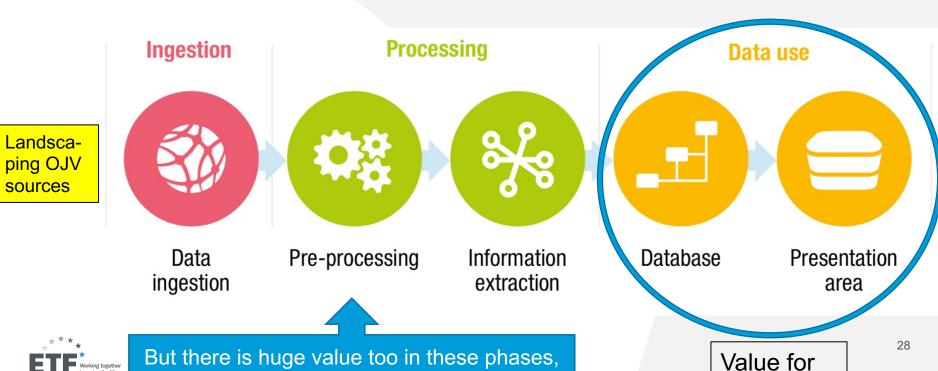
Data processing:
clean, deduplicate,
classify data through
machine learning
techniques

Data analysis: extract information from data and make it accessible through visualisation tools



THE METHODOLOGY

Knowledge Discovery in Databases

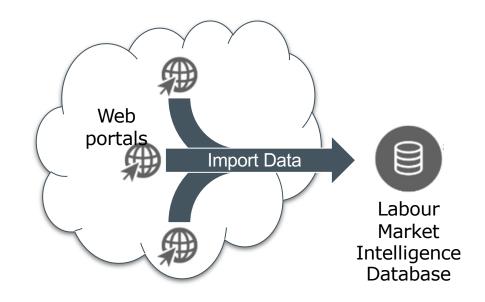


loops and steps (not to be neglected)

end-users

NEW SOURCE OF DATA

Web Data ingestion is the process of obtaining and importing data from web portals and storing in a database





COLLECTING REAL-TIME LABOUR MARKET DATA



Visit Online Job Sites



Collect & Deduplicate Job Postings

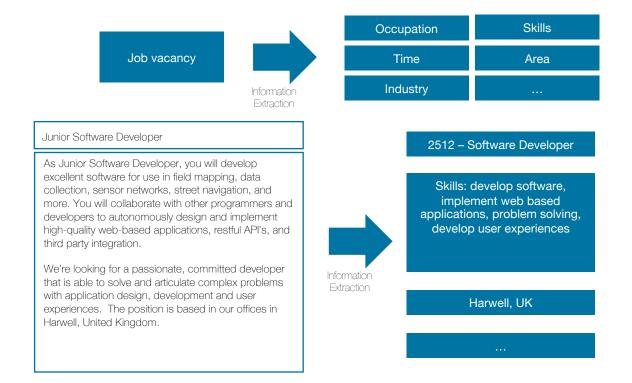


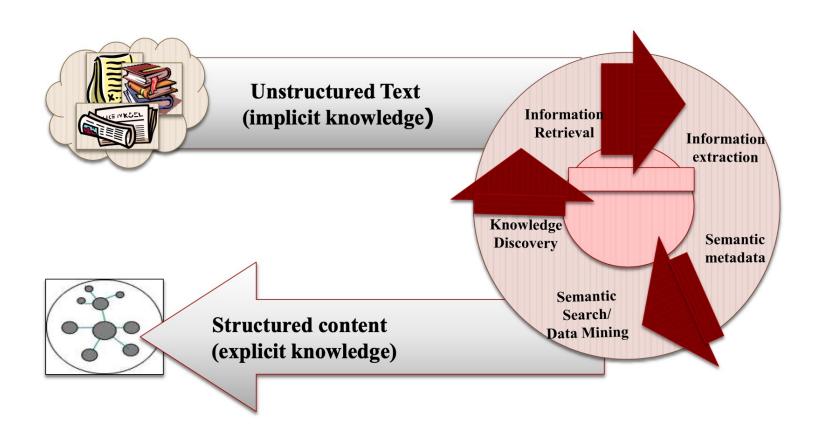
Tagging & Normalising Postings to Generate Detailed Data

- Job Title & Occupation
- Employer & Industry
- Technical Skills
- Foundational Skills
- Certifications
- Educational Requirements
- Experience Levels
- Salaries



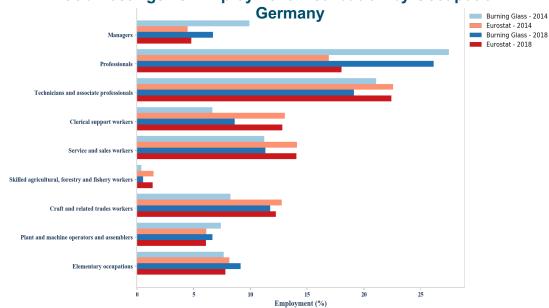
Data Classification - An example





UNDERSTAND & CONTROL FOR VARIANCES

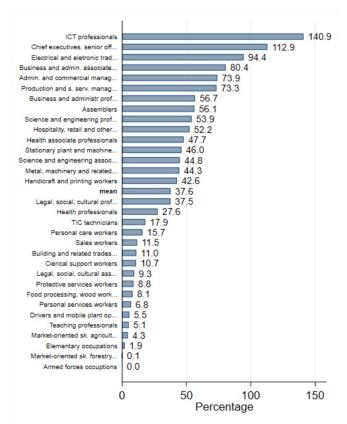




Bertelsmann Stiftung

Source: Lightcast

Figure 5. Ratio between vacancies (BGT) and employment contracts (SPES) by occupation. 2018-2021 **NEW DATA**







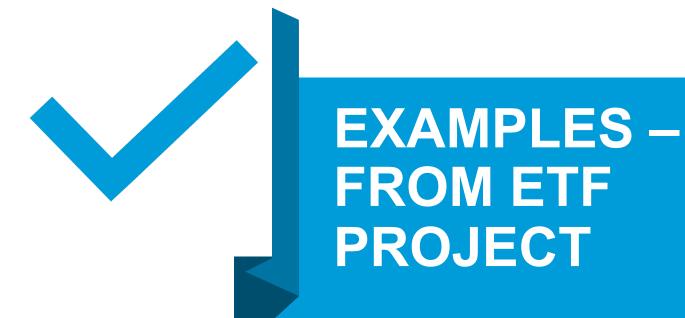
Data coming from web → Need to detect and reduce noise

Accuracy of automated classification

Multi language environment

- Need to relate to classification standards
- Find a way to summarize and present a wide and complex scenario
- Use of datasets by key institutions
- Need to relate to other databases and statistics

CHALLENGES AND OPPORTUNITIES





ETF OVJ DATA DASHBOARDS

Tunisia:

https://public.tableau.com/app/profile/tabulaex/viz/ETF-BigDataLMI-Tunisia/Time

Ukraine:

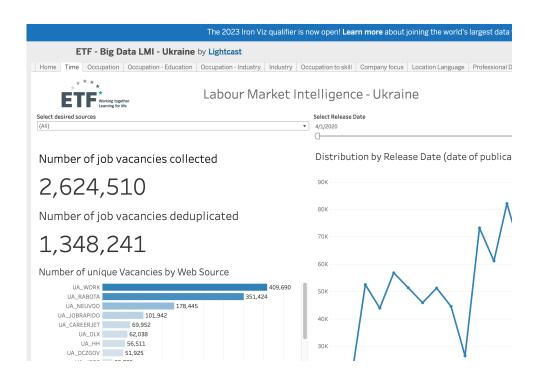
https://public.tableau.com/app/profile/tabulaex/viz/ETF-BigDataLMI-Ukraine/Time

Georgia:

https://public.tableau.com/app/profile/tabulaex/viz/ETF-BigDataLMI-Georgia/Time?publish=yes

Egypt:

https://public.tableau.com/app/profile/tabulaex/viz/ETF-BigDataLMI-Egypt/Home



ETF OVJ DATABASES

Country	Total collected	Total deduplicated	Time series
Tunisia	684,209	178,368	From 04/2020
Ukraine (general)	2,624,510	1,348,241	From 04/2020
Georgia	131,936	86,635	From 04/2021
Egypt	1,388,809	404,567	From 01/2021
Kenya	Approx. 15,000 (1st release soon)	1st release soon	From 08/2022

VARIABLES – ANALYSED, COMBINED AND VISUALISED



Occupation



Education level



Sector / industry



Knowledge, skills, attitudes



Contract type



Experience



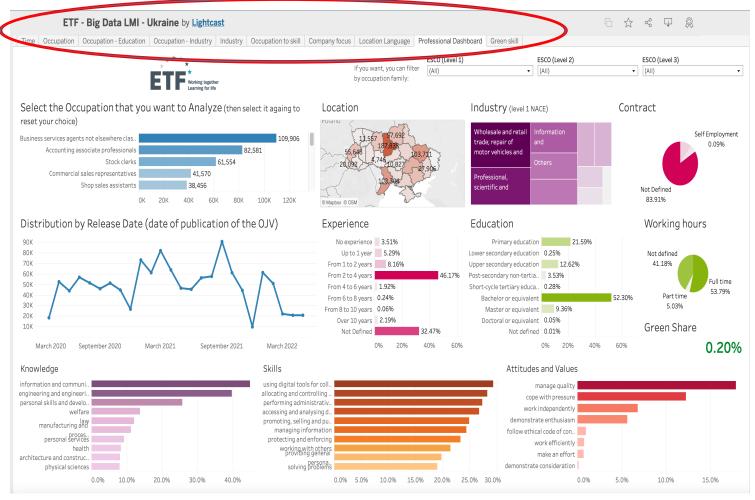


Variables and features of the demand for jobs and skills can be added, customised

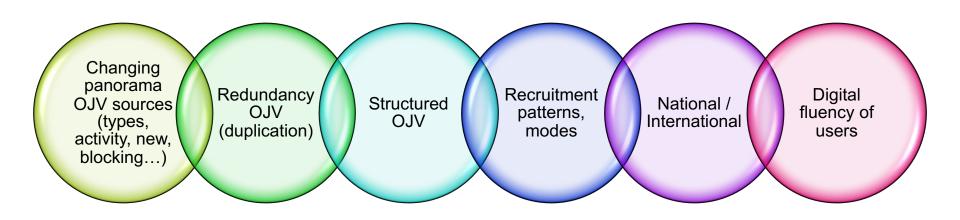
DASHBOARD: SKILLS AND OCCUPATIONS

Variables:

- 1. Occupation
- 2. Skills
- 3. Education level
- 4. Experience
- 5. Salary
- 6. Location
- 7. Green skills



VALUE FROM LANDSCAPING OF WEB LM



ANALYSIS

2020: OJV outlook Ukraine, Tunisia

2021: OJV in the context of Covid-19 (Ukraine, Tunisia, Georgia)

2022: OJV outlook new countries; digital and green skills Egypt

OJV DATABASES: USE FOR NEW THEMES

GREEN: SKILLS BY OCCUPATIONS AND SECTOR



SET OF GREEN TERMS – GREEN DASHBOARDS: 7 CLUSTERS OR CONTEXTS

Environment and sustainable tourism

Sustainable agriculture

Sustainable construction

Sustainable economy

Sustainable energy

Sustainable production

Sustainable transport



GREEN TERMS – SO FAR: 224...A GLIMPSE

Construction

airtight construction asbestos removal building air tightness building enclosure building envelope building performance energy efficient building Green Architecture green building green building nanocoating green building practice evaluation green building standards green certified construction practices green construction low energy buildings near zero energy building nzeb retrofitting smart thermostat sustainable building sustainable building materials sustainable installation materials thermal insulation zero energy buildings

Energy

alternative energy alternative fuels biodiesel biofuels biogas systems biomass biomass systems biorefinery carbon-neutral fuel Carbon negative fuel clean energy cogeneration combined heat and power district heating and cooling efficient energy use energy conservation energy conversion energy efficiency energy efficient operations energy reduction energy saving energy storage energy-from-waste geothermal energy geothermal engineering **Geothermal Heat Systems** Geothermal Production green energy

Transport

alternative fuel vehicles clean fuel clean vehicles bicycle sharing bike sharing carpooling services carsharing Car pooling Car sharing compare alternative vehicles electric drive system electric vehicle **Energy Efficient Transportation** Flexible fuel vehicles Green Automotive Technologies Green Transportation green vehicle hybrid operating strategies hybrid vehicle Hydrogen vehicle mobility as a service plug-in hybrid powertrain suitability sustainable transport vehicle ecological footprint **Hybrid Buses**





GREEN SKILLS ANALYSIS: NEXT



Compare and approximate to ESCO skills labelled "green" – in cooperation with ESCO-Support team

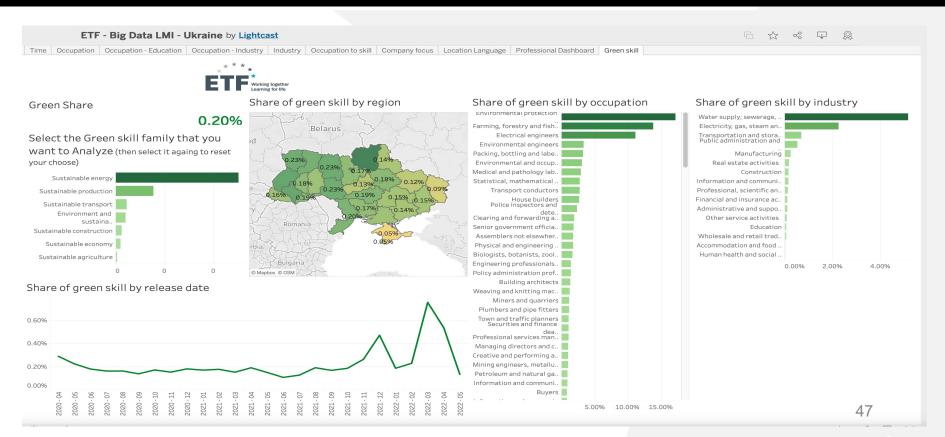
2

Update / review the classification (consider ongoing Cedefop work on green skills taxonomy)

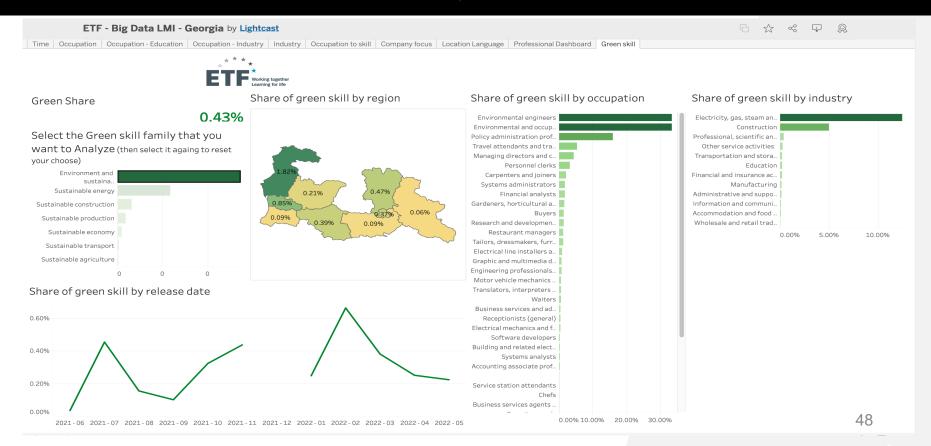
3

Add new countries: Egypt, Kenya

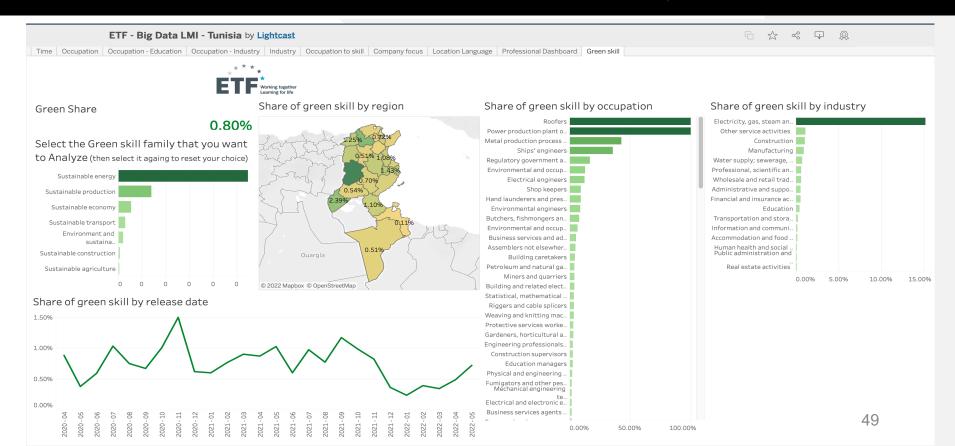
THE SHARE OF OJV REQUIRING SKILLS OF GREEN CATEGORY IS LOW! UKRAINE...0,20%



GEORGIA: ONLY 0,43% GREEN SHARE



TUNISIA...THE GREENEST: 0,80%



THANK YOU

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