ETF Working together Learning for life

European Training Foundation

SKILLS LAB

European Commission

EaP | Eastern Partnership

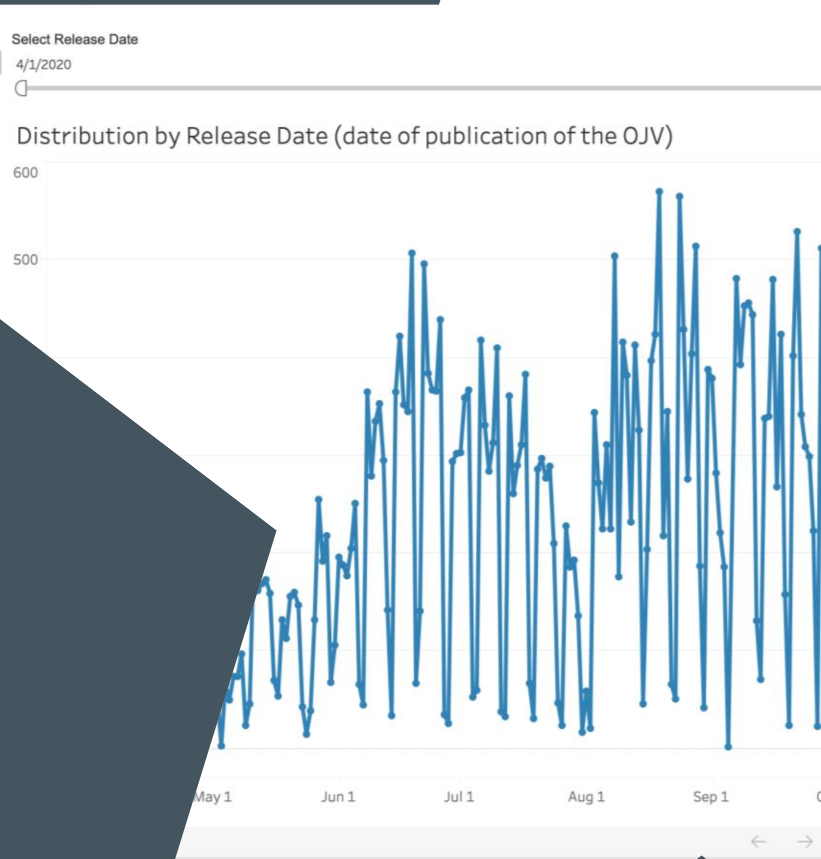
EUROPEAN TRAINING FOUNDATION

Session 2: Let the Data speak. Labour market information in transformation – Big Data analytics in application: Tunisia and Ukraine. Main conclusions. Visualisation of the results in interactive Dashboards.

Alessandro Vaccarino, Burning Glass Europe

Labour Market Information in Transformation | 10 December 2020

THE EU AGENCY SUPPORTING COUNTRIES TO DEVELOP THROUGH LEARNING



ES ACTIVITY ABOUT SIGN U

essional Dashboard

Select Release Date

4/1/2020

llected

Distribution by Release Date (date of publication of the OJV)

CONTENTS

1. Context and Goals
2. Methodological approach
3. Ukrainian and Tunisian experiences
4. Interactive Dashboard

ETF Working together Learning for life

European Training Foundation

1. Context and Goals

CONTEXT

Continuously evolving Labour Market:

- Digitalization of professions
- Relevance of Soft skills
- Internationalisation
- New professions and skills emerging
- Smart and Remote working
- Impact of Covid-19 pandemic
- …

We need *something* that can help us monitor and analyze **how** LM is evolving, to support Decision Makers taking **the right decisions at the right time**

WHAT WE HAVE / WHAT WE NEED

We already have **official statistics**, that are:

- *Representative*
- *Strong* in terms of value

But we can benefit of additional, complementary information that could be:

- *Fast*, to track what’s happening now (e.g. Covid-19 Impact analysis)
- *Granular* and *adherent* to real and current market terms, to capture emerging trends analyzing what companies are actually looking for

How to find a similar, complementary source of information?

Using **Web Labour Market**

ETF Working together Learning for life

European Training Foundation

WHY WEB LABOUR MARKET

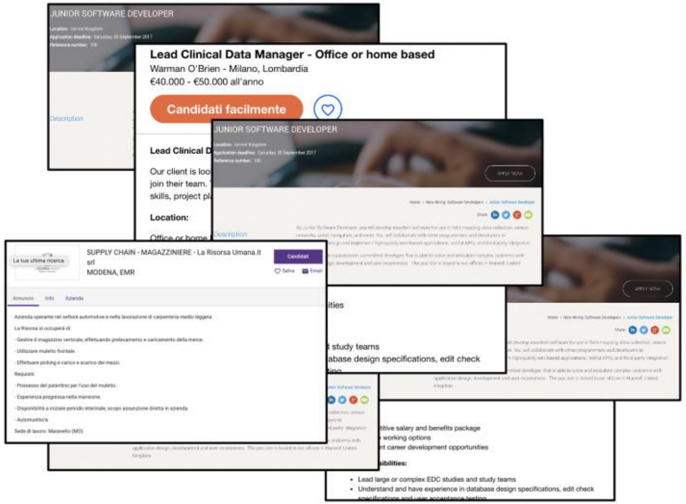
It’s the **exact representation** of what companies are looking in a given period:

- Up to date: companies publish an announcement **when** they actually need to hire
- Detailed: an announcement describes **as well as possibile** the specific need, in terms of:
  - Profession needed
  - Requirements (skills, experience, educational level,…)
  - Working context (place, contract, sector, working hours,…)
- Adherent to reality: **market terms** are used, both for occupation and skills. This helps identify emerging terminology adopted by Market

It would be great to use those information in addition to better and deeper understand how Labour Market is evolving in a given country, even compared to other countries

OUR GOAL

Transform those…



JUNIOR SOFTWARE DEVELOPER

Description

Lead Clinical Data Manager - Office or home based

Warman O'Brien - Milano, Lombardia

€40.000 - €50.000 all'anno

Candidati facilmente

Lead Clinical D

Our client is loo
join their team.
skills, project pl

Location:

Office or home

SUPPLY CHAIN - MAGAZZINIERE - La Risorsa Umana.it srl

MODENA, EMR

La tua ultima ricerca

Candidati

Salva Email

Annuncio Info Azienda

Azienda operante nel settore automotive e nella lavorazione di carpenteria medio leggera

La Risorsa si occuperà di:

- Gestire il magazzino verticale, effettuando prelevamento e caricamento della merce.
- Utilizzare muletto frontale.
- Effettuare picking e carico e scarico dei mezzi.

Requisiti:

- Possesso del patentino per l'uso del muletto.
- Esperienza pregressa nella mansione.
- Disponibilità a iniziale periodo interinale, scopo assunzione diretta in azienda.
- Automunito/a

Sede di lavoro: Maranello (MO).

ities

study teams

abase design specifications, edit check

titive salary and benefits package

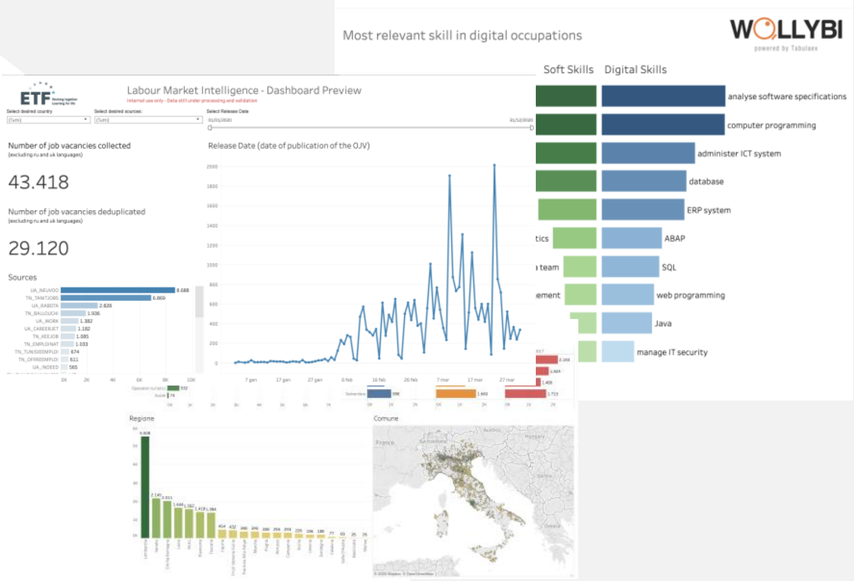
working options

nt career development opportunities

sibilities:

- Lead large or complex EDC studies and study teams
- Understand and have experience in database design specifications, edit check specifications and user acceptance testing

…to this

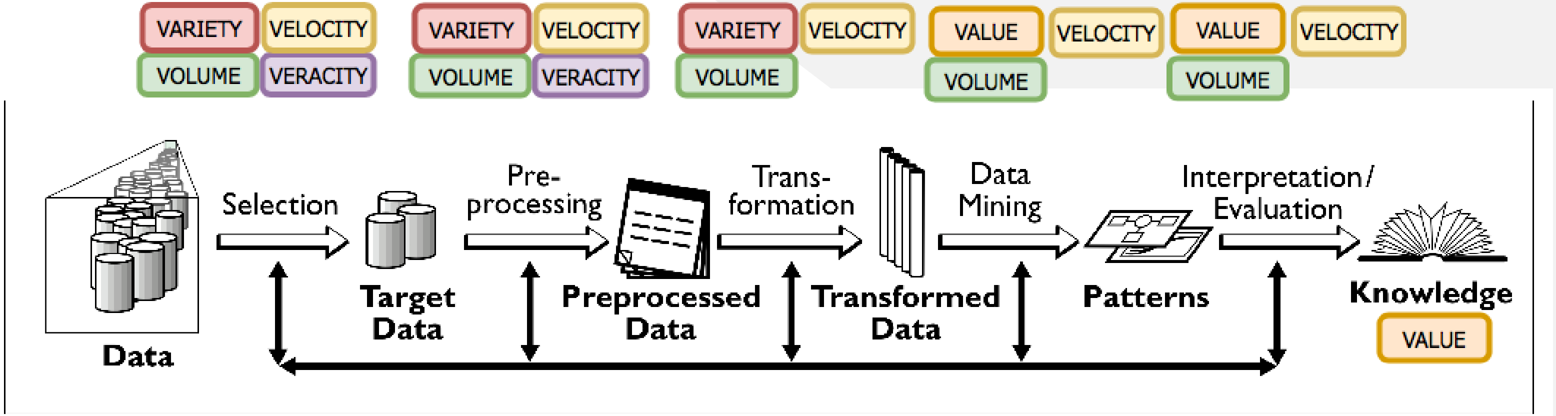


European Training Foundation

2. Methodological approach

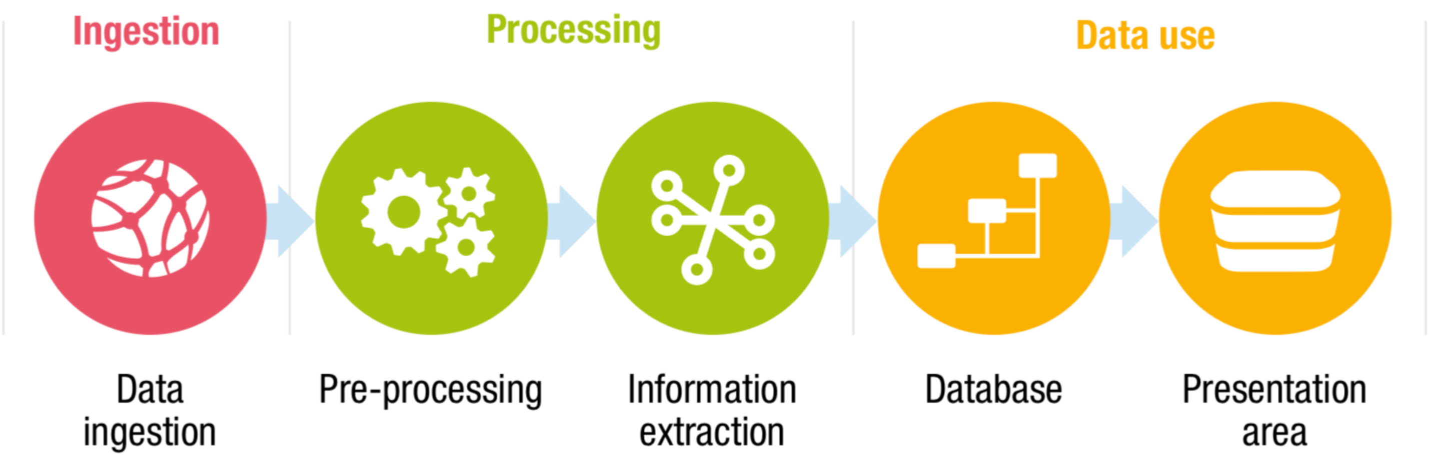
METHODOLOGICAL BACKGROUND

KDD – Fayyad, 1997



OUR APPROACH

KDD 4 LMI



Ingestion

Processing

Data use

Data ingestion

Pre-processing

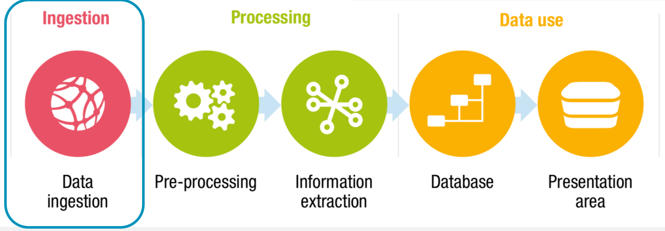
Information extraction

Database

Presentation area

Let’s take a deeper look on this framework

DATA INGESTION



The process of obtaining and importing data from web portals and storing them in a Database

Focus on volume

Coverage augmentation

Balance between quality and effort

ETF Working together Learning for life
European Training Foundation

DATA INGESTION - GOALS

Robustness
of process

Quality
of data collected

Scalability and
governance

ETF Working together Learning for life

European Training Foundation

DATA INGESTION – ROBUSTNESS

Issue: potential technical problems when gathering data from a source (unavailability, block, changes in data structure)

Risk: loss of data

Solution: redundancy

- Have the most important sites (by volume and/or coverage) ingested from two or more sources
- Avoid loss of data in case of troubles with a source
- Collect data from both primary and secondary sources

DATA INGESTION – QUALITY

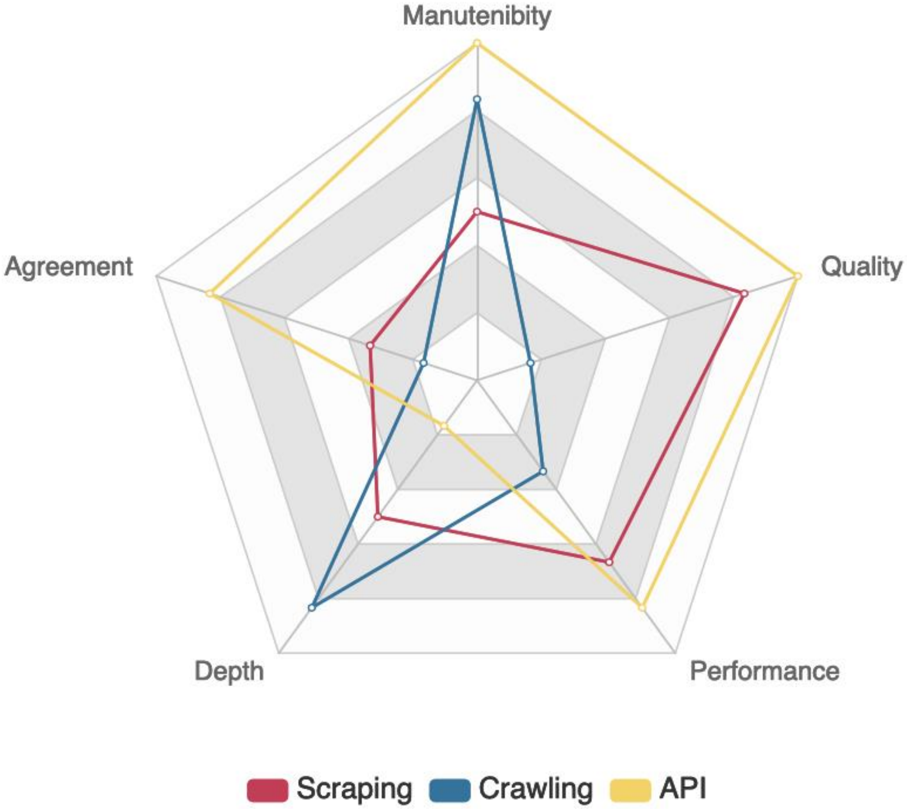
Issue: need to obtain data as clean as possible, detecting structured data when available

Risk: loss of quality

Solution: tailored ingestion. We collect data using a specific approach based on the single source:

- API
- Scraping
- Crawling

DATA INGESTION – QUALITY FRAMEWORK



DATA INGESTION – SCALABILITY AND GOVERNANCE

Issue: need to handle a real and complex Big Data environment, simultaneously connecting to thousands of websites

Risk: Loss of Process control and loss of OJVs due to slowness of the process

Solution:

- A scalable infrastructure
- A monitoring and governance custom tool

DATA INGESTION – RECAP

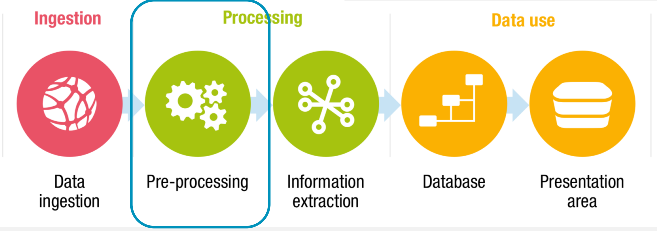
After this phase, we have web pages, most likely Online Job Advertisments.
But they are:

- Noisy
- Duplicated
- Unstructured

As discussed, a proper source selection is strategic: it's mandatory to identify the most relevant web portals, in terms of numbers, quality and informational value. How to ensure a proper selection?

With a Landscaping phase

DATA PRE-PROCESSING



The process of cleaning ingested data and deduplicating OJVs, to guarantee that analytical phase’ll work on data at the highest quality possible

Language detection

Noise reduction

OJVs Deduplication

DATA PRE-PROCESSING – LANGUAGE DETECTION

Why:

- Each language has different keywords, stopwords,…
- It can reflect different cultures and Labour Market scenarios…

How:

- We trained 60+ specific classifiers based on Wikipedia corpus
- Models are accurate (~99% of precision) and fast to adopt

What we obtain:

- A fast and strong classification of the language used in each OJV
- A way to archive OJVs for which we don’t have language support

DATA PRE-PROCESSING – NOISE DETECTION

Why:

- In a Big Data environment, we must deal with noise
- Information gathered from the web, one of the most noisy place avilable

How:

- AI based models, similar to mail spam filters

What we obtain:

- Identification of:
  - Web pages explicitly not related to OJVs
  - Web pages disguised as OJVs

European Training Foundation

DATA PRE-PROCESSING – DEDUPLICATION

Why:

- Companies post several advertisments for each vacancy → Visibility
- It cannot affect analysis: no over-estimation due to multiple postings

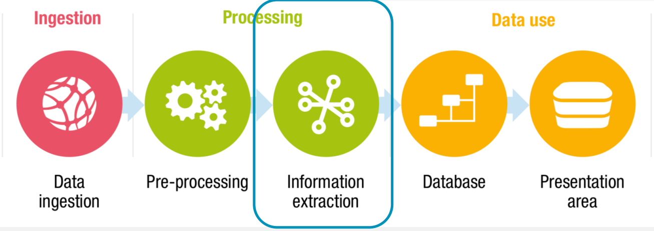
How:

- Statistical-based approach: identification of the standard duration of an OJV
- Text-analysis to detect similar/identical advertisments

What we obtain:

- Unique OJVs, to ensure coherent analyses

DATA CLASSIFICATION



Extract and structure information from data, with respect to the most proper taxonomy

Artificial Intelligence

Taxonomy selection

Information Linkage

DATA CLASSIFICATION – AN EXAMPLE

Junior Software Developer

2512 – Software Developer

As Junior Software Developer, you will develop excellent software for use in field mapping, data collection, sensor networks, street navigation, and more. You will collaborate with other programmers and developers to autonomously design and implement high-quality web-based applications, restful API's, and third party integration.

We're looking for a passionate, committed developer that is able to solve and articulate complex problems with application design, development and user experiences. The position is based in our offices in Harwell, United Kingdom.

Software development, application design, UX,…

Harwell, UK

DATA CLASSIFICATION – TAXONOMY

Why:

- We need to formalize all our inforation, to make it consistent and enable analyses
- Occupations/Skills/Places/… must be related to a proper taxonomy
- A unique taxonomy for each dimension enables analyses across countries and projects

How:

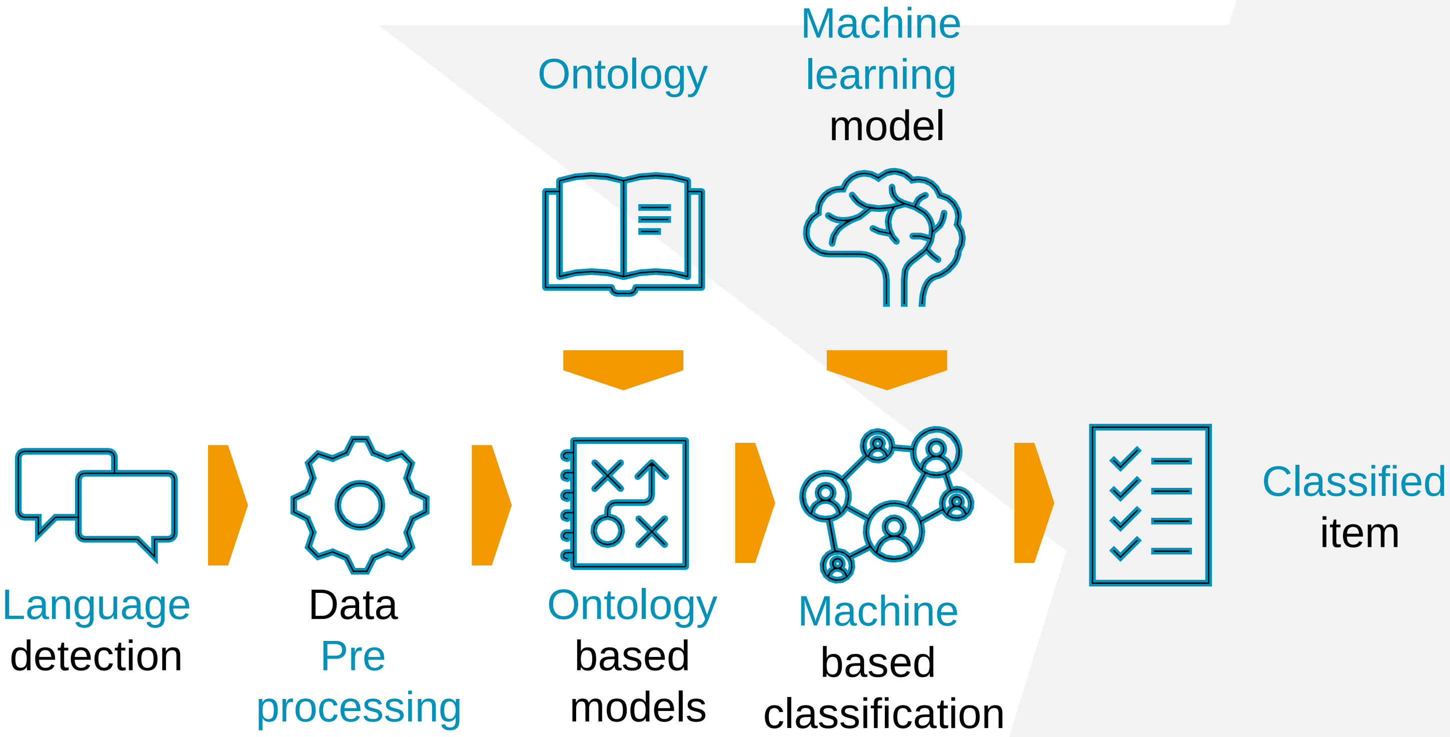
- Selection on international and custom taxonomies, that fit Labour Market terms and enable

DATA CLASSIFICATION – TAXONOMY

Most relevant taxonomies adopted:

- Occupation: ESCO/ISCO
- Skills: ESCO
- Places: NUTS and ISO
- Educational Level: ISCED
- Sector: NACE
- Senority/Working hours/Contract type/…: custom taxonomies

DATA CLASSIFICATION – APPLICATION



Ontology

Machine learning model

Language detection

Data Pre processing

Ontology based models

Machine based classification

Classified item

ETF Working together Learning for life

European Training Foundation

ETF Working together Learning for life

European Training Foundation

3. Ukrainian and Tunisian experiences

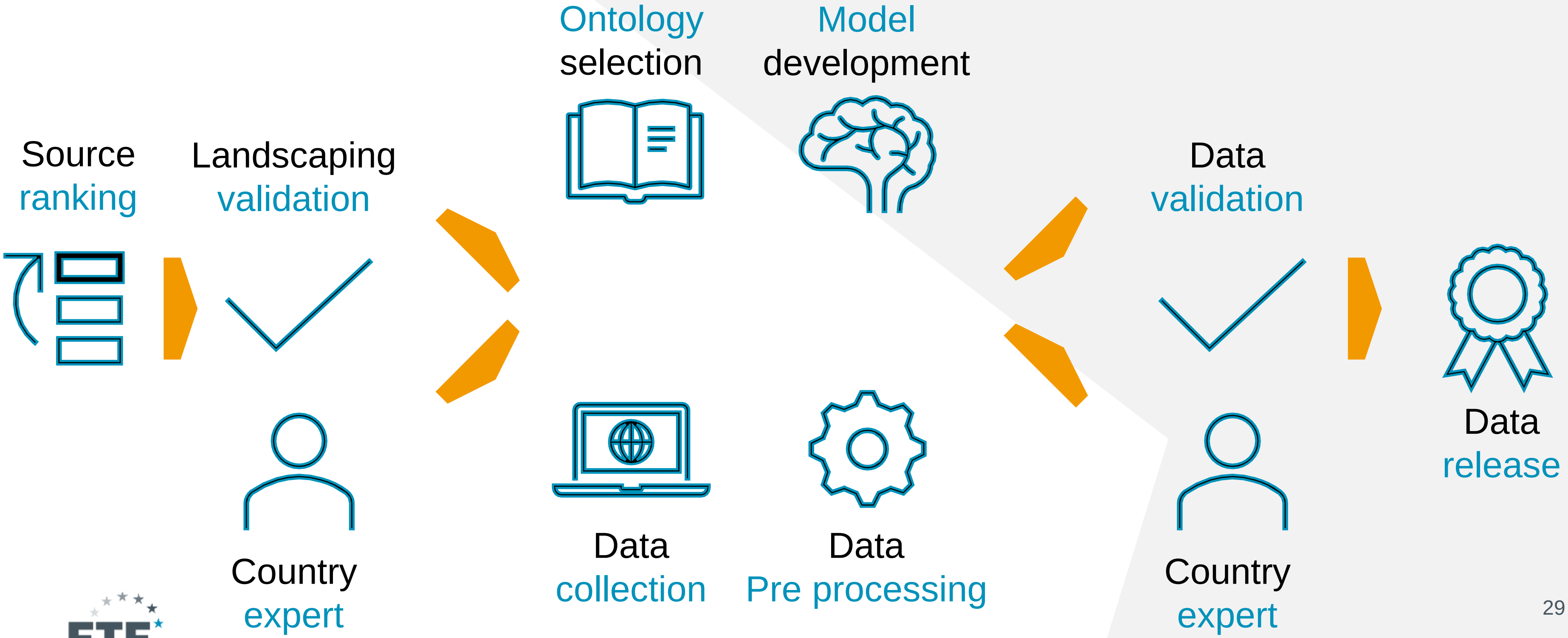
UKRAINIAN AND TUNISIAN EXPERIENCES – INTRODUCTION

In January 2020, we started a project to collect, classify and analyze data regarding Web Labour Market in both Tunisia and Ukraine

The project followed the same methodology presented in the previous section.

- A Source Selection was performed and validated by our Country Esperts
- Data were collected, cleaned and classified on their own languages
  - Specific classifiers were developed for both Urkainian and Russian languages
- Information collected was shared with Country Experts, to identify possible issues in the process and validate it

UKRAINIAN AND TUNISIAN EXPERIENCES – WORKFLOW



Source ranking

Landscaping validation

Country expert

Ontology selection

Model development

Data collection

Data Pre processing

Data validation

Country expert

Data release

ETF Working together Learning for life

European Training Foundation

European Training Foundation

Working together
Learning for life

4. Interactive Dashboard

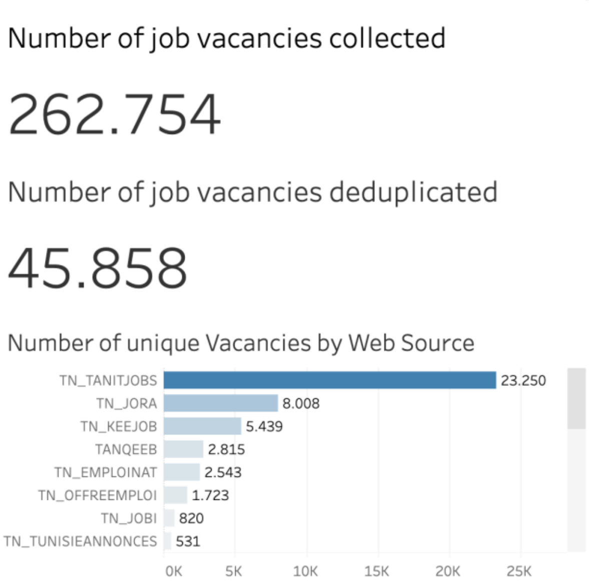
INTERACTIVE DASHBOARDS

You can find dashboards at:

- Tunisia:
  https://public.tableau.com/profile/tabulaex#!/vizhome/ETF-BigDataLMI-Tunisia/Time

- Ukraine:
  https://public.tableau.com/profile/tabulaex#!/vizhome/ETF-BigDataLMI-Ukraine/Time

INTERACTIVE DASHBOARDS – SOME INSIGHTS



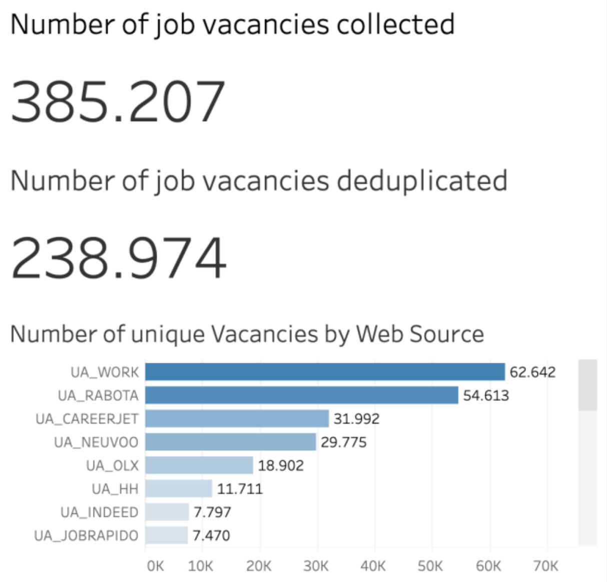
Number of job vacancies collected

262.754

Number of job vacancies deduplicated

45.858

Number of unique Vacancies by Web Source



Number of job vacancies collected

385.207

Number of job vacancies deduplicated

238.974

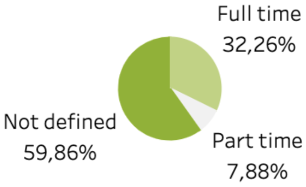
Number of unique Vacancies by Web Source

ETF Working together Learning for life

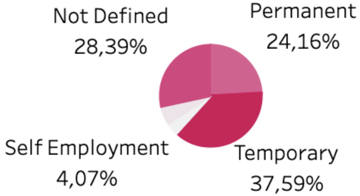
European Training Foundation

INTERACTIVE DASHBOARDS – SOME INSIGHTS

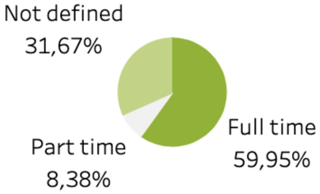
Working hours



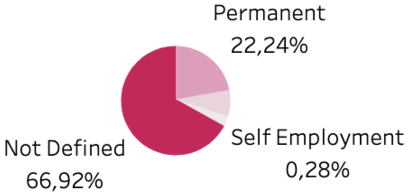
Contract



Working hours

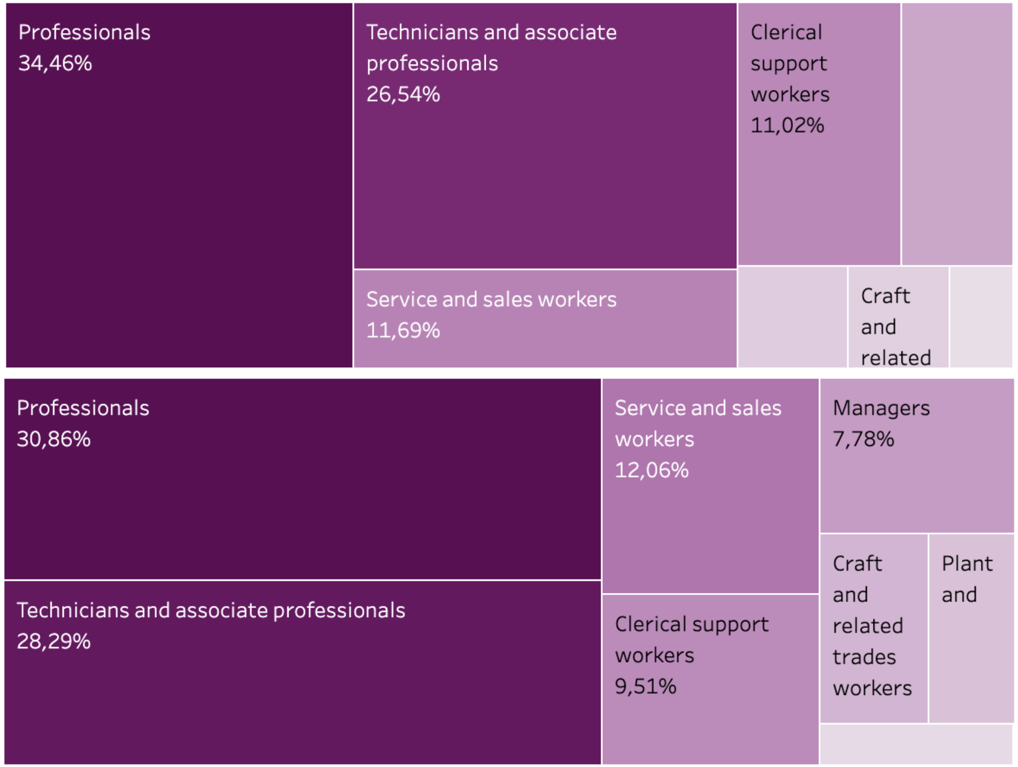


Contract



INTERACTIVE DASHBOARDS – SOME INSIGHTS

Occupation (level 1)



Professionals
34,46%
Technicians and associate professionals
26,54%
Clerical support workers
11,02%
Service and sales workers
11,69%
Craft and related
Professionals
30,86%
Service and sales workers
12,06%
Managers
7,78%
Technicians and associate professionals
28,29%
Clerical support workers
9,51%
Craft and related trades workers
Plant and

ETF Working together Learning for life

European Training Foundation

ETF Working together Learning for life

European Training Foundation

European Commission

EaP | Eastern Partnership

EUROPEAN TRAINING FOUNDATION

OPEN SPACE
Sharing knowledge on skills
ETF Working together Learning for life
European Training Foundation

Thank you very much

Alessandro Vaccarino, Burning Glass Europe

Labour Market Information in Transformation | 10 December 2020

THE EU AGENCY SUPPORTING COUNTRIES TO DEVELOP THROUGH LEARNING