

# Big Data for Labour Market Intelligence

Capacity development programme 2022

**Module 2: dissemination and analysis**

**Session 2**

Landscaping, ranking and monitoring of OJV sources: value for the wider LMI

# Overview

- What is a landscaping and why do we need it?
- ETF Landscaping methodology
  - List of potential sources and their characteristics
  - Country profile – OJV market and its context
  - Ranking sources
  - Output
- Conclusions

# Landscaping: a definition

**Def** The set of procedures used to list and survey the websites that post OJV relevant in representing the online labour market in the country of interest.

**Goal** Build a Labour Market Information system to inspect labour demand structure at the occupation-skill level.

# Why do we need a Landscaping?

To extract value from OJV data sources, we need to manage their complexity.

Sources should be:

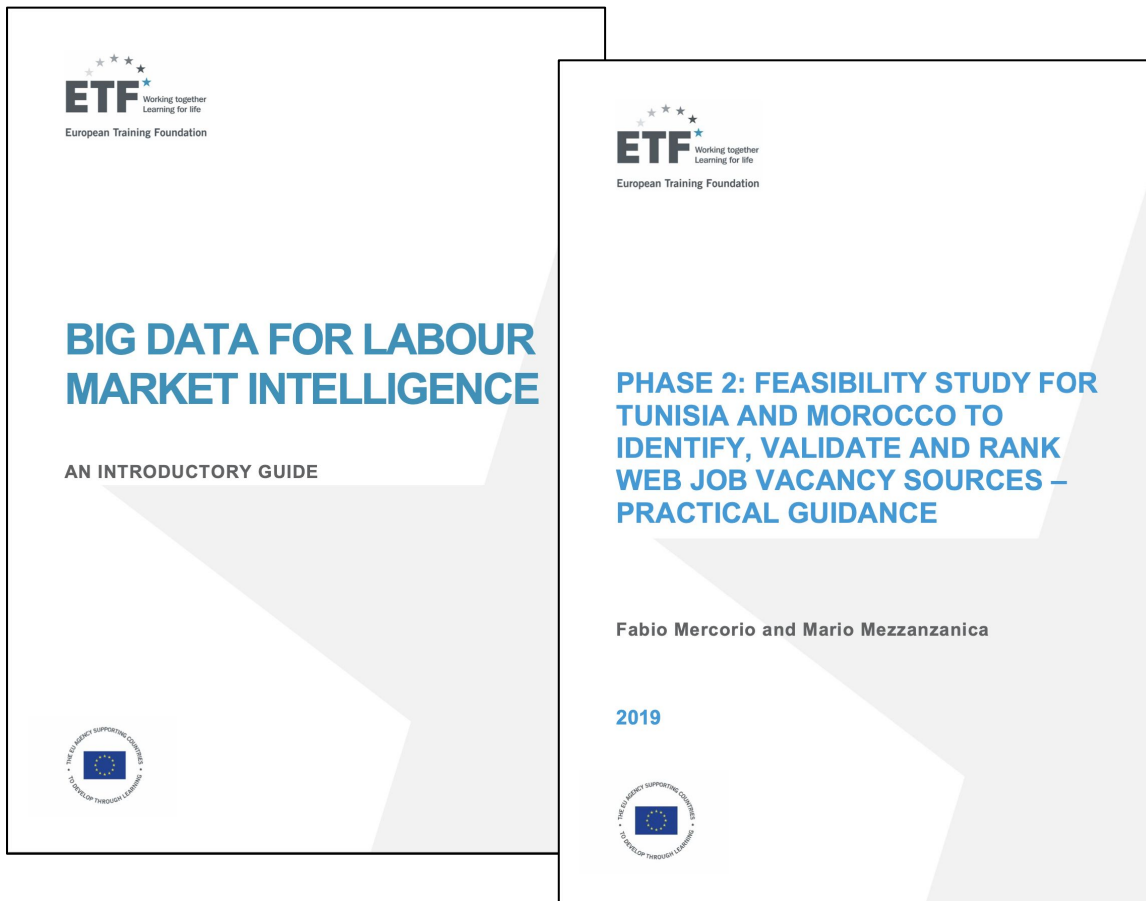
- **reliable** □ regular, timely, cover the aspects of interest
- **trustable** □ high-quality on the aspects of interest

To be technically sustainable and trustable, the system must be

- **Efficient** minimise deduplication, minimise maintenance costs
- **Effective** high-quality information which we can classify, high direct coverage of relevant information

# ETF Methodology

## Context



The framework for the use of big data for Labour Market Intelligence (LMI) was presented in 2019.

In the same year, a feasibility study for Tunisia and Morocco was published. This report contained a first exposition of the methodology used to select the sources to extract data to develop an LMI system.

Today we present an updated methodology. It is developed in the framework of the former and it improves the methods exposed in the latter.

# ETF Methodology

## The process

### Step 1

Populate a list of potential OJV sources

### Step 2

Country profile on labour market (supply, demand, institutions), internet access, use of internet in recruitment practices and job search

### Step 3

Ranking sources

# ETF Methodology

## Step 1 List of sources

How do we look up websites?

Use Google Search + input keywords related to job postings

- Job search <country>
- Online job search sites <country>
- Find job <country>
- Job ads <country>
- Job recruiting websites <country>

Why...

... use Google Search? OJA portals *want* to be found + Google is by far the most used search engine

... input so general keywords? Minimise the risk of listing specialised portals that bias representativeness

# ETF Methodology

## Step 1 List of sources

Which characteristics matter?

Both characteristics of the websites and OJV ads' pages posted on the websites



# ETF Methodology

## Step 1 List of sources

### Characteristics of the websites [1]

Variable	Value	Description
<b>Google rank</b>	• first page	The website is listed on the first page of the Google Search results
	• second page	The website is listed on the second page of the Google Search results
<b>Job portal (type)</b>	• primary job-portal	Portals that advertise vacancies for which the user can apply for on the portal itself and do not redirect to another website.
	• secondary job-portal	Portals that advertise vacancies that are published on other websites.
	• combination of primary and secondary function	Portals that advertise both vacancies directly and vacancies ads published elsewhere.
<b>Operator (type)</b>	• classified ads portal	General ads portals containing ads for job vacancies.
	• job search portal	Websites collecting OJA from different sources and presenting them in an integrated search engine (aggregators).
	• recruitment agency	The website of a private employment agency that connects employers with jobseekers.
	• public employment service	The website of portals of national employment services.
	• national newspaper	Job advertisement pages of online newspapers.
	• company website	The website is owned by a company that advertises internal jobs.
<b>Number of OJV</b>		Number of vacancy posted on the website. If unknown, please leave it blank.
<b>Language</b>		List the languages used to advertise vacancies, using the ISO 639-1 code.

# ETF Methodology

## Step 1 List of sources

### Characteristics of the websites [2]

Variable	Value	Description
<b>Geographical scope</b>	international	The source advertises job vacancies for more than one country.
	national	The source advertises job vacancies inside a country.
	regional	The source advertises job vacancies for a bounded area of a country, such as a province, a metropolitan area or a city.
<b>Economic activity</b>	one economic activity	The website advertises job vacancies related to a single industry. If so, please move to next. Otherwise skip.
	more than one economic activity	The website advertises job vacancies related to more than one industry.
<b>Economic activity [d]</b>		NACE Sections
<b>Publication date</b>	present	A publication-date structured field is visible on the page.
	not present	The publication date is not provided in a structured field.
<b>Expiry date</b>	present	A publication-date structured field is visible on the page.
	not present	The publication date is not provided in a structured field.
<b>Update frequency</b>	daily	Entries in the ads list are posted at a time distance of a day.
	other	Entries in the ads list are posted at a time distance of more than a day.
<b>Paid advertisement</b>	free advertisement	The user can advertise a job vacancies free of charge.
	paid advertisement	The user pays a fee to advertise a job vacancy.
	both	The user may either post an advertisement free of charge or by paying a price (e.g. for additional services).

# ETF Methodology

## Step 1 List of sources

### Characteristics of OJV ads' pages

Structured fields

Free-text field

**Customer Sales Executive / Customer Service Advisor** Title

📍 Bristol Area, [South West](#)  
£ £16640 - £20000 per annum + £30-40k OTE + benefits  
🏢 [Proactive Solutions Group Ltd](#)  
🕒 Permanent  
 Today  
📄 [Check your FitScore](#) Description

[Activate one-click apply.](#) [What is one-click apply?](#)

Customer Sales Executive / Customer Service Advisor  
Customer Sales Executive / Customer Service Advisor. A Bristol based mobility firm are looking for a new Customer Sales Executive / Customer Service Advisor to join the team. The Customer Sales Executive / Customer Service Advisor will work 9:00-17:00 Monday to Friday and every other Saturday for which you receive a day off in the week.  
Although this role is a phone based telesales type position, you do not need telesales experience, just a bright and enthusiastic approach to call handling, kindness, empathy, understanding, a caring approach, ability to care about getting the right result for the customer.

The Customer Sales Executive / Customer Service Advisor Executive responsibilities are:

# ETF Methodology

## Step 1 List of sources

### Characteristics of OJV ads' pages

For each variable, we collect whether the information is exposed in a structured field, in free text or it is not available

- Occupation type mapped on ESCO Occupation (unit, IV digit)
- Contract type mapped to Permanent, Temporary, Apprenticeship and Internship
- Working time mapped to Full-time/Part-time
- Economic activity mapped to NACE (sections, I digit)
- City
- District
- Region
- Education level
- Salary

# ETF Methodology

## Step 2 Country profile

How to choose the best sources to set up a data production system?

Know the context in which the portals operate. The activity is run by the country expert with the support of ETF/CRISP/Lightcast

Areas of interest

- **Review of labour market dynamics** Population, labour supply, key sectors of employment
- **General and employment-related use of the web**
- **Collection and characterisation of the data** Identifying the job vacancy sources and their characteristics, List of job vacancy sources identified

This research informs about

- Priority of surveyed sources in terms of relevance for the online LM in the country
- Representativeness of the data
- Possible uses

# ETF Methodology

## Step 3 Ranking

How to choose the best sources using the findings of Step 1 and 2 and to address the technical needs and the need for valuable knowledge?

Rank the sources according to two dimensions:

- Relevance for the OJA market in the country
- Quality of the source

Then, synthesise a single indicator taking into account the two dimensions

# ETF Methodology

## Step 3 Ranking

**Ranking 1** Relevance of the sources for the OJA market in the country

The country expert is asked to rank sources based on the importance for recruitment practices and job search.

ETF provides support information on:

- Sources popularity
- Sources stability and status in the DPS

Feedback:

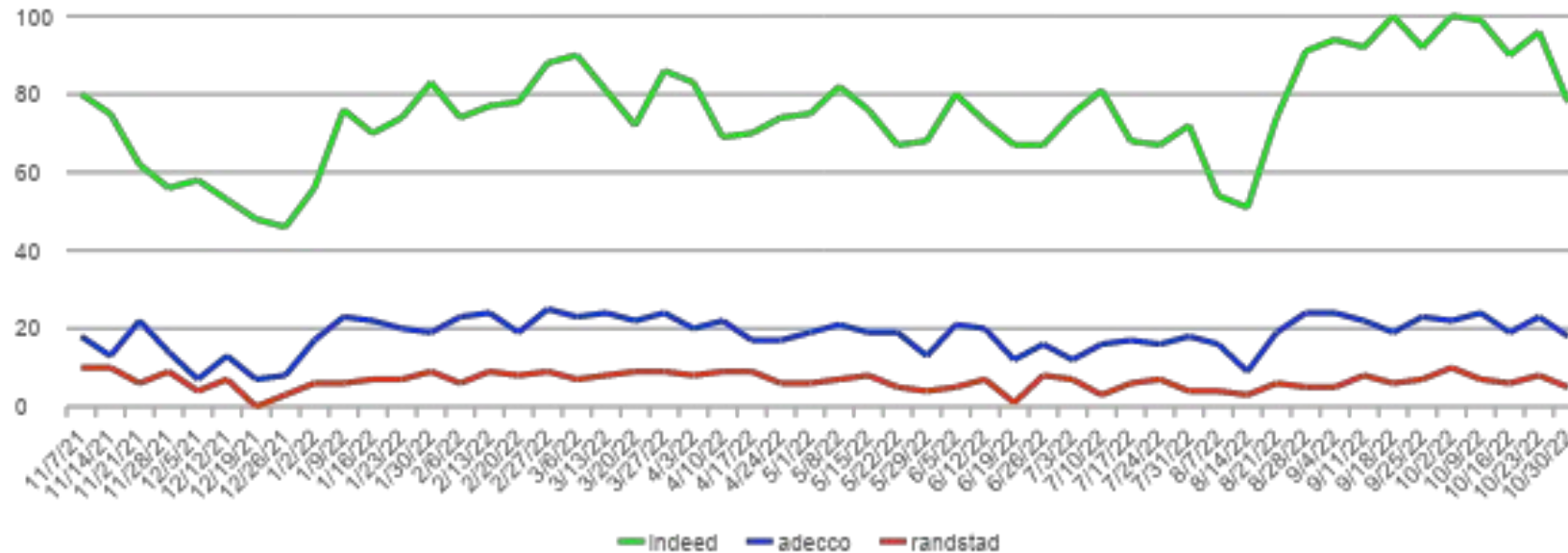
- Ranking
- Comments on the source relevance

# ETF Methodology

## Step 3 Ranking

### Ranking 1 Popularity

Using the names of the sources collected in Step 1 as inputs, run searches in Google Trends to compare the relative interest in the sources in the year of the survey.



Category job ads



# ETF Methodology

## Step 3 Ranking

### Ranking 1 Stability

If the country is in the DPS already, the expert receives information on

- **Sources' status in the DPS** whether scrapers are working or not, if the source owner blocked them, the mode of data ingestion (crawlers, scrapers, API)
- **The time series of the ads collected from the sources** quality of the analytics depends on the regularity of the dataflow

# ETF Methodology

## Step 3 Ranking

### Ranking 2 Quality of the source

We collect information on the sources based on the list of Step 1. Not all information is equally important.

How to define the importance of variables and values?

The Analytic Hierarchical Process\* is a technique to deal with multi-criteria decision-making.

It reduces the complexity of assigning a priority to a large set of criteria distributed on a hierarchy, by running pairwise comparisons at all levels.

\* Saaty, T. L. (1990). How to make a decision: the analytic hierarchy process. European journal of operational research, 48(1), 9-26. Chicago

# ETF Methodology

## Step 3 Ranking

### Ranking 2 Quality of the source - AHP

#### Level 1

- Website features
- Job ads webpages features

#### Level 2

- Google rank
- Job portal (type)
- Operator (type)
- Geographical scope
- Economic activity
- Publication date
- Expiry date
- Update frequency
- Paid advertisement
- Occupation type
- Contract type
- Working time
- Economic activity
- City
- District
- Region
- Education level
- Salary

#### Level 3

All variables' values

Example of decision hierarchy				
Level 1	Level 2	Level 3	Glb Prio.	
websiteFeatures	google_ranking_position	first page	0.720	1.4%
		second page	0.280	0.5%
	job_portal_type	primary job-portal	0.492	2.4%
		secondary job-portal	0.138	0.7%
		combination of primary and secondary function	0.370	1.8%
	operator_type	classified ads portal	0.104	0.3%
		job search portal	0.235	0.7%
		recruitment agency	0.150	0.5%
		public employment service	0.296	0.9%
		national newspaper	0.099	0.3%
		company website	0.116	0.4%
		geographical_scope	international	0.446
		national	0.347	1.1%
		regional	0.207	0.7%
	economic_activity_scope	one economic activity	0.314	0.9%
		more than one economic activity	0.686	1.9%
		A-B Agriculture and fishery	0.158	0.2%

# ETF Methodology

## Step 3 Ranking

### Ranking 2 Quality of the source - AHP

Top priority for:

- Occupation (structured or free-text)
- Publication date
- Region
- Economic activity

Rank	Variables	Weight
1	Occupation - both	0.093650
2	Occupation - structured	0.067961
3	Publication date - present	0.052958
4	Education level - structured	0.044442
5	Economic activity - structured	0.042600
6	City - structured	0.038425
7	Region - structured	0.035621
8	Education level - both	0.033045
9	Economic activity - both	0.031046
10	Salary - structured	0.028143

# ETF Methodology

## Step 3 Ranking

### Ranking 2 Quality of the source - AHP

#### Features

- Output a vector of numeric weights
- Check the consistency of individual judgements
- Evolution from a single judgement to an AHP performed by many stakeholders
- Possibility to integrate additional stakeholders' judgements ex-post

#### How to use it?

Map the values on the Step 1 list and sum to obtain a source-specific AHP score.

# ETF Methodology

## Step 3 Ranking

### Source selection

First map the Expert score and the AHP score to their quartiles of belonging. Then set a decision rule to combine the two rankings in one score.

	Priority	Cases
Prioritised for development	High	Sources in Q1 of <u>Expert score</u> and Q1 of <u>AHP score</u> .
	High-medium	Sources in Q1 or Q2 of <u>Expert score</u> and Q1 or Q2 of <u>AHP score</u> .
	Medium	Sources in Q2 or Q3 of <u>Expert score</u> and Q2 or Q3 of <u>AHP score</u> , except for (Q3, Q3).
	Medium-Low	Sources in Q3 or Q4 of <u>Expert score</u> and Q3 or Q4 of <u>AHP score</u> , except for (Q3, Q3).
Futher inspection	Mixed	Sources with distance between position in <u>Expert score</u> distribution and <u>AHP score</u> distribution larger than 1 quartile.

# ETF Methodology

## Output

ETF publishes a landscaping report for each country

## CONTENTS

### TABLE OF ABBREVIATIONS

#### 1. INTRODUCTION

Landscaping methodology overview

#### 2. REVIEW OF LABOUR MARKET DYNAMICS

Population characteristics

Labour supply characteristics

Key sectors of employment

#### 3. GENERAL AND EMPLOYMENT-RELATED USE OF THE WEB IN <COUNTRY>

General use of the web

Employment-related use of the web

#### 4. COLLECTION AND CHARACTERISATION OF DATA

Identifying the job vacancy sources and their characteristics

List of job vacancy sources identified

#### 5. RANKING MODEL

5.1. Source evaluation

5.2. Source Selection

#### 6. CONCLUSIONS

### REFERENCES

# Conclusion

OJV data are a unique source of information on labour market demand, i.e., high-frequency, detailed information at the vacancy level as exposed by employers in online job postings.

The complexity of these data sources must be managed to obtain meaningful and valuable information, which is the basis of a trustable LM intelligence system.

The landscaping methodology is a fundamental step in the process, as it provides a policy to restrict the input sources of the DPS to the most promising ones.



**Thank you for your attention**

Francesco Trentini  
francesco.trentini@unimib.it