

ONLINE JOB AD LANDSCAPING: METHODOLOGICAL GUIDE

Author: Fabio Mercorio, Mario Mezzananza and Francesco Trentini

Reviewers: Eduarda Castel Branco

Date: October 1, 2022



European Training Foundation

The contents of this draft report are the sole responsibility of the authors and do not necessarily reflect the views of the ETF or the EU institutions.

© European Training Foundation, 2022

Reproduction is authorised, provided the source is acknowledged.



CONTENTS

TABLE OF ABBREVIATIONS	5
1. PREFACE	6
2. ONLINE JOB ADS LANDSCAPING METHODOLOGICAL GUIDE	7
1. Identification of online-job-ads sources	8
2. Country-specific landscaping report	11
2.1. Labour market dynamics and impact on the online-job-ads market	11
2.2. Survey of potential sources	11
2.3. Evaluation of sources	11
2.3.1. Quantitative assessment of websites characteristics	11
2.3.2. Qualitative assessment of websites relevance	13
2.4. Selection of sources for development	15
REFERENCES	17

TABLE OF ABBREVIATIONS

OJV	Online job vacancy
OJA	Online job advertising
OJP	Online job posting
PES	Public Employment Services
AHP	Analytic Hierarchy Process
API	Application Programming Interface
ESCO	European Skills/Competences, qualifications and Occupations
GDP	Gross domestic product
KDD	Knowledge discovery in databases
KPI	Key performance indicators
LMI	Labour market information
NEET	Not in employment, education or training
NGO	Non-governmental organisation
OJV	Online Job Vacancy
DPS	Data Production System
MENA	MENA, an acronym in the English language, refers to a grouping of countries situated in and around the Middle East and North Africa.
Q&A	Questions and answers

1. PREFACE

In the last few years, we have witnessed a massive burst in the availability of online data. This data is generated from several sources, often without control from the user or the organisation, and in different forms, both structured and unstructured (e.g., text and images). This has two main implications for the way data is managed and analysed.

On the one side, there has been a paradigm shift in the way of reasoning, integrating the top-down deductive approach that starts from a series of hypotheses, collects and analyses data useful to verify it with a bottom-up inductive approach that aims to discover hidden patterns. and to support the decision-making process in social domains. This data-driven approach allows one better to exploit the hidden informative power of Big Data, distilling relevant insights among it and, eventually, reconciling it with administrative data and official statistics to infer claims about the entire observed population. On the other side, those enormous flows of data with different forms need new management paradigms to cope with their complexity.

In such a scenario, in 2019 the European Training Foundation (ETF) published *[Big data for labour market intelligence. An introductory guide](#)*¹. The document stated the general framework for the use of big data in the activities of the institution, giving an account of the advantages and drawbacks in the use of this novel source of data and sketching some possible applications. Since then, ETF has started working on collecting online job ads (OJAs) in different countries, such as Tunisia, Morocco, Ukraine and Georgia. The methodology used in these initial implementations of the collection of OJV from the web was presented in *[Phase 2: feasibility study for Tunisia and Morocco to identify, validate and rank web job vacancy sources – practical guidance](#)*. The document presented the landscaping activity, the first step in data collection who includes the survey, the evaluation, and the selection of relevant sources of online job vacancies to be used as inputs for the Labour Market Intelligence system. This step is critical for implementing a reliable and trustworthy analysis of insights provided because, by selecting the most relevant and high-quality sources, it makes the system efficient by controlling the volume of ingested pages and by minimising the risk of duplications, while improving its effectiveness, by maximising the potential coverage of occupations represented in the system.

The present guide builds on this latter work and improves it in many respects. First, the new methodology has introduced new variables to survey the completeness of information on potential sources. Second, the assessment of the quality of the sources is done by assigning to each source characteristic a numeric weight that represents its relative relevance compared to other source characteristics. In the previous version, a single expert defined the weights, which are now collected among a larger group of experts, including labour market experts and data scientists. The aggregation of the various preferences are consistently done by an algorithm. Third, the final ranking of prioritised sources is now obtained as a combination of the country expert ranking and the quality assessment, reducing the discretion in choosing the sources to start data collection from. Finally, this guide also introduces the topic of assessing the quality of sources that are part of the data production system. Addressing this latter point is important for the sustainability of the system and to set up counter-measure to possible issues in data retrieval that may impinge on the quality of the output statistics.

This publication follows.

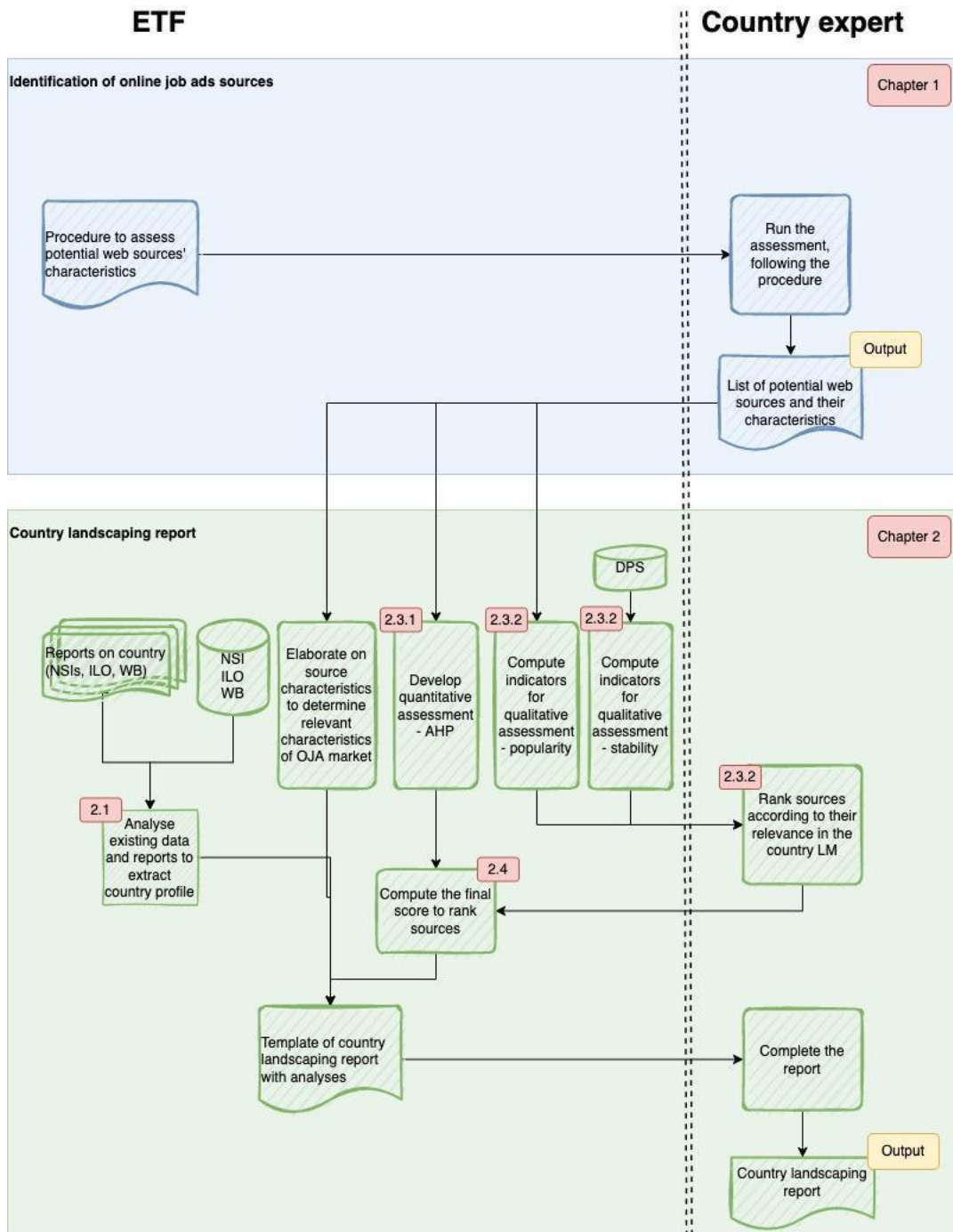
¹ <https://www.etf.europa.eu/en/publications-and-resources/publications/big-data-labour-market-intelligence-introductory-guide>

2. ONLINE JOB ADS LANDSCAPING METHODOLOGICAL GUIDE

The methodology presented in this guide aims at establishing a replicable and accountable procedure to select the best candidates for scraping online job ads in a country. The relevance of sources to include in the production system is therefore evaluated on different dimensions, i.e., the information that sources can potentially provide on occupations and skills, the relevance of these features in terms of desirability for the involved stakeholders, the role the source plays in the country online job ad market in relation to other players, the accessibility of the website in terms of its structural characteristics and the ability to ingest and process it.

The process is composed of four main parts. The first is the identification of potential web sources of online job ads (Section 1). At this stage, country experts use a standard procedure to collect a list of websites that post online job ads, so to identify, select and assess new and known sources. Companion to the survey of sources is the national report, which collects information on the context in which online job ads providers operate, with a focus on socio-demographic characteristics of the country, labour market institutions and features and an analysis of public and private websites that host online job ads (Section 2). The country profile helps understanding the representativeness of online job ads with respect to the national labour market and inform on the possible uses that can be made of the information collected from the web. Finally, the information gathered in the previous steps is harmonised and centrally processed to obtain a ranking model of sources based on quantitative and qualitative information. The aim of this stage is the integration of the knowledge obtained on multiple domains to select, on the base of a range of criteria, the most relevant online job ad sources (Section 2.3).

Figure 1 Workflow of the production of the country landscaping report. Red labels refer to the chapter of the methodological guide that deals with the topic.



Note: Data Production System (DPS), National Statistical Institute (NSI), International Labour Organisation (ILO), World Bank (WB),

1. Identification of online-job-ads sources

The first stage is aimed at identifying websites that advertise job vacancies. Structural characteristics of the websites are collected and analysed according to a matrix of relevant

features, that assesses their presence and their accessibility on job-portals web pages. The whole process is standardised in four main steps to grant replicability.

Step 1: Translation of search keywords in national languages The search for online job portals starts with the assignment of a list of keywords to country experts. Keywords are disseminated in English to maximise the standardisation of the search activity while leaving to the country expert the task of finding the translation that more accurately describes the same content in their national language, therefore increasing the consistency of the results.²

Step 2: List of online job portals Each keyword is used by the country expert to run a query on google.com.³ The list of returned websites is analysed by the expert, who registers all the search results produced by the search engine that refer to a web page that posts job ads.

Step 3: Features of new sources Websites that are categorised as sources of job vacancies are analysed in depth by the country expert. This activity is necessary to prevent the well-known “garbage-in, garbage out” phenomenon that might happen by selecting all the sources without assessing their quality, thus affecting the overall quality of the analysis of Web sources. The standardisation of data gathering is ensured by a data validation procedure that constrains the possible values. A definition of each variable and its metadata is provided to the country expert, as well as the operative steps that are required to accomplish the activity, through contextual information and guidance. The list of variables has been enriched to include new important aspects of web pages, while some attributes have been dropped because they have shown irrelevant according to previous experience. Figure 2 lists the values and description of each characteristic that is collected.

² The keywords are “Job search <country>”, “Online job search sites <country>”, “Find a job <country>”, “Job ads <country>”, “Job recruiting websites <country>”. The choice of these keywords is such that they maximise the emergence of job portals which are not specialised on some sectors or occupations, so as to have the highest coverage possible of the spectrum of occupations.

³ The reason for the choice of google.com is manifold. First, Google is almost a monopolist in web search, as about 90% of queries are processed by it. Since our aim is to simulate the typical search behaviour of a jobseeker, Google is likely to be the selected search engine by most of them. Second, this makes the standardisation of the process of website listing free from any ex-post computations, which would be required in case of use of more than one search engine. Third, the calculation of popularity can take advantage of Google Trend which uses the same underlying algorithm that lists web sources, increasing the coherence of the procedure by minimising the private data generating processes.

Figure 2 Survey of potential online job ad sources: variables, values and description

Variable	Value	Description
google_ranking_position	first page second page from third page on	The website is listed on the first page of the Google Search results The website is listed on the second page of the Google Search results The website is listed on the third page of the Google Search results or later.
job_portal_type	primary job-portal secondary job-portal combination of primary and secondary function	Portals that advertise vacancies for which the user can apply for on the portal itself and do not redirect to another website. Portals that advertise vacancies that are published on other websites. Portals that advertise both vacancies directly and vacancies ads published elsewhere.
operator_type	classified ads portal job search portal recruitment agency public employment service national newspaper company website	General ads portals containing ads for job vacancies. Websites collecting OJA from different sources and presenting them in an integrated search engine (aggregators). The website of a private employment agency that connects employers with jobseekers. The website of portals of national employment services. Job advertisement pages of online newspapers. The website is owned by a company that advertises internal jobs.
oja_volume		Number of vacancy posted on the website. If unknown, please leave it blank.
language		List the languages used to advertise vacancies, using the ISO 639 code.
geographical_scope	international national regional	The source advertises job vacancies for more than one country. The source advertises job vacancies inside a country. The source advertises job vacancies for a bounded area of a country, such as a province, a metropolitan area or a city.
economic_activity_scope	one economic activity more than one economic activity	The website advertises job vacancies related to a single industry. If so, please move to economic_activity_1d The website advertises job vacancies related to more than one industry.
economic_activity_1d	A — Agriculture, hunting and forestry B — Fishing C — Mining and quarrying D — Manufacturing E — Electricity, gas and water supply F — Construction G — Wholesale and retail trade; repair of motor vehicles, motorcycles and personal and household goods H — Hotels and restaurants I — Transport, storage and communication J — Financial intermediation L — Public administration and defence; compulsory social security M — Education N — Health and social work O — Other community, social and personal service activities P — Activities of households Q — Extra-territorial organizations and bodies	
publication_date_data_type	present not present	A publication-date structured field is visible on the page. The publication date is not provided in a structured field.
expiry_date_data_type	present not present	A publication-date structured field is visible on the page. The publication date is not provided in a structured field.
update_frequency	daily other	Entries in the ads list are posted at a time distance of a day. Entries in the ads list are posted at a time distance of more than a day.
advertising_payment_option	free advertisement paid advertisement both	The user can advertise a job vacancies free of charge. The user pays a fee to advertise a job vacancy. The user may either post an advertisement free of charge or by paying a price (e.g. for additional services)
occupation_data_type	structured not structured both	The information is displayed in a standard field. (highlighted in green in the example) The information is displayed in a text box which is not standardised. (highlighted in blue in the example) The information is displayed both in a standard field and in a text box which is not standardised. (as in the example)
contract_data_type	structured not structured both	
working_time_data_type	structured not structured both	
economic_activity_data_type	structured not structured both	
city_data_type	structured not structured both	
district_data_type	structured not structured both	
region_data_type	structured not structured both	
education_level_data_type	structured not structured both	
salary_data_type	structured not structured both	

Customer Sales Executive / Customer Service Advisor



Training and support

Training and support are provided to country experts before and along the whole duration of the activity by staff members via email or video calls.

Transmission of the documents and performance monitoring

Country experts fill a dedicated worksheet on a cloud drive service. This arrangement has several advantages: transmission issues and compatibility of editors are avoided; support can be timely; monitoring of the progress can be done live by the administrators, by means of dashboards and analytics tools.

Feedbacks and iterations

Feedback constitutes an important means of improvement of the procedure that may require adjustments based on some specific features of each country, which may have not been included in the first place. The use of cloud services allows for implementing changes in the procedure on the go, e.g., modifying the structure of the deliverable while preserving the filled content.

2. Country-specific landscaping report

The survey of sources is complemented by an analysis of the online job ads market and the socio-economic context in which it is embedded. This knowledge is organised in a country-specific report. The report covers labour market dynamics and their impact on the online job ads market and presents the survey results of potential sources. It exposes the evaluation criteria and the outcome of the selection process. Country experts receive a template and a table of content to grant standardisation in the type of information that is collected and the output format of the document.

2.1. Labour market dynamics and impact on the online-job-ads market

This section covers topics that are useful to depict the economic profile of the country. The goal of the section is twofold. On the one hand, it is useful to understand the relevance of the internet channel in recruitment practices and job search. On the other hand, to frame the possible uses of online job ads to inform policy-making in the country and to inform the source selection that bounds the data ingestion. The topics that are covered are the socio-demographic profile of the country, human capital and skills, a general overview of the country's economic activities and a detailed account of the labour market in terms of inactivity, unemployment and employment, with a focus on the composition in terms of occupations. To do that, the consortium and the country experts collect and elaborate reports and data from National Statistical Institutes (NSIs) and trusted international institutions, such as the International Labour Organisation (ILO) and the World Bank (WB).

2.2. Survey of potential sources

This section of the report summarises the main characteristics of online job portals gathered from the first step of landscaping (Section 1) and the knowledge of country experts. The main focus is on the presence of public online job portals and their mission and the characteristics of private job portals.

2.3. Evaluation of sources

The fieldwork conducted by the country experts produces information on every source and a general overview of the online job ad market in the country. After this assessment, the next stage is the evaluation of the sources themselves.

The specificity of the evaluation framework is the involvement of many methods that cover many aspects of each source. Indeed, two assessments are performed. The first is a qualitative assessment of sources' relevance in the OJV market, which produces a quantitative indicator based on criteria set by involved stakeholders through an Analytic Hierarchy Process (AHP) model (Section 2.3.1). The second is a qualitative assessment of website characteristics by country experts (2.3.2) based on two criteria: popularity and stability. This step produces a rank based on qualitative judgements by country experts.

Finally, the outputs of this assessment are then used as inputs for the calculation of a final ranking, described in Section 2.4.

2.3.1. Quantitative assessment of websites characteristics

The goal of this step is to prioritise the selection of new online job advertisement sources for the ingestion and processing phases. This activity aims to use experts' judgement to set weights in

a ranking system that will be used to select sources for further ingestion and processing phases. In practice, categorical variables identified in the protocol described in Section 1 will be assigned numerical values as weights depending on their importance. This activity has been evolving since the feasibility study produced in 2019. In that case, a single expert was involved in the definition of the importance of websites' characteristics. Extending the number of participating stakeholders was therefore important to preserve the exercise from biases in the evaluation. Nonetheless, this choice introduces complexity in the system as these many preferences need to be aggregated in a single one. To summarise, the two criteria that we need to satisfy under these new conditions are the following:

- Numerical values are assigned following the relative importance that each variable's value bears in relation to all other variables
- Include the preferences of all the involved stakeholders

With these two characteristics in mind, we want to use a model that can integrate different points of view on the relative importance of every variable's value in a transparent and accountable way.

The Analytic Hierarchy Process (AHP) model handles such a scenario and guarantees trust and explainability of the model results. The AHP ⁴ is an effective technique for dealing with multi-criteria decision-making problems that allow decision-makers to set priorities to variables integrating the preferences of many stakeholders. By reducing complex decisions to a series of pairwise comparisons and then synthesising the results, the AHP helps to capture both subjective and objective aspects of a decision. Moreover, the AHP can translate the evaluations (both qualitative and quantitative) made by many decision-makers into a single score, and the process can be repeated at higher levels of the structure and assigning a score to variables and groups of variables. Note that the AHP model presents two distinct features useful in our context. First, it highlights inconsistencies in evaluating participants' preferences, enabling them to revise their judgments accordingly. Second, it allows computing the consensus among participants at each taxonomy level. This means the decision maker can have a fine-grained analysis of the degree of agreement for each variable value identified and decide accordingly.⁵

An example is shown in Figure 3, where the hierarchy of criteria – a group of variables, variables and variable values – are evaluated by a stakeholder. The procedure requires each stakeholder to pairwise compare all the elements belonging to a level of the hierarchy, which are then translated into a score – such that the sum of the scores for each level is equal to 1. Moreover, the procedure checks the consistency of preferences for each level – namely that the preferences expressed by each stakeholder must satisfy the axiom of transitivity. The column “Level 3” shows the scores of the categorical values of each variable. Besides attributing numerical values to categorical variables, this method allows feedback on variables or groups of variables' importance from the point of view of each stakeholder.⁶

The list of web sources provided by country experts will allow the consortium members to draw a ranking of web sources. The goal of the ranking model is twofold: the identification of sources that are not eligible for being included in the final list of sources, due to a relatively low quality-score and the priority of sources to be engaged first for data ingestion.

⁴ Saaty, T. L. (1990). How to make a decision: the analytic hierarchy process. *European journal of operational research*, 48(1), 9-26. Chicago

⁵ Intuitively, as AHP deals with independent criteria, this means that if criterion A is better than B, and B is better than C, we expect a consistent judgment requires A to be better than C (transitivity)

⁶ The specific tool that is used in this case is: Goepel, K.D. (2018). Implementing an Online Software Tool for the Analytic Hierarchy Process (AHP-OS). *International Journal of the Analytic Hierarchy Process*, Vol. 10 Issue 3 2018, pp 469-487, <https://doi.org/10.13033/ijahp.v10i3.590>

Figure 3 AHP model on the complete hierarchy. Level 3 corresponds to variable values

Decision Hierarchy				
Level 0	Level 1	Level 2	Level 3	Glb Prio.
RelevanceOfCriteria AHP	CriteriaOnWebsite 0.125 AHP	PosGoogle 0.067 AHP	firstpage 0.900	0.8%
			secondpage 0.100	0.1%
		TypeJobPortal 0.444 AHP	primary 0.770	4.3%
			secondary 0.068	0.4%
			combination 0.162	0.9%
			classified 0.275	1.7%
		TypeOfOperator 0.489 AHP	company 0.116	0.7%
			jobPortal 0.152	0.9%
			newspaper 0.152	0.9%
			pes 0.152	0.9%
		recruitmentAgency 0.152	0.9%	
	CriteriaOnOJA 0.875 AHP	Occupation 0.761		66.6%
		Wage 0.191		16.7%
		Region 0.048		4.2%
			1.0	

2.3.2. Qualitative assessment of websites relevance

This part of the procedure aims at producing a rank of sources based on their relevance in the country's online job vacancy market from the point of view of country experts. While the assessment in the previous paragraph evaluates mainly the accessibility of relevant information in the potential sources, following technical criteria, this second evaluation is conducted in terms of the relevance of the sources for the OJA market in the country. Each country expert assigns a score to each source and comments on the specificities of each source.

To support country experts' evaluation, the consortium elaborates indicators over two dimensions: popularity and stability. The latter evaluation requires some historical online job ad data on which to be performed and introduces in the ETF methodology the notion of quality assessment for sources in the data production system. Then experts are asked to evaluate the importance of each source, considering the evidence provided by the consortium and their personal knowledge of the labour market in their country. For instance, country experts may decide to flag a source as relevant even if it is not stable or does not rank high in popularity because of its importance in the country landscape. In such a case, country experts are required to motivate their personal judgement.

The dimension of popularity captures the knowledge that jobseekers have of surveyed sources, by evaluating the volume of direct searches generated towards them. The computation is done by assessing the frequency of direct web searches. This is done through Google Trends, which produces an index of search interest based on the volume of search normalised by major region (NUTS1) and time range. Each query produces a weekly index by basic regions (NUTS2). The output produced by Google Trend is computed on a sample of all the queries that are conducted through the Google search engine in the country and time interval of interest. The representativeness of the sample is granted by the large number of queries administered by the provider. Individual searches data anonymized, categorised and aggregated.⁷ Another feature

⁷ Following Google documentation, "[s]earch results are normalised to the time and location of a query by the following process: (i) Each data point is divided by the total searches of the geography and time range it represents to compare relative popularity. Otherwise, items with the most search volume would consistently be ranked highest. (ii) The resulting numbers are

of the service is the availability of categories each keyword can be associated with. This feature is important to address the ambiguity of keywords and refine searches.⁸ For example, searches for “Indeed” would incorrectly count queries for the English adverb among the relevant searches of the job portal. Among the available categories, “960 – Job listings” is the one of interest in the case of LMI.

We have developed a script in Python that automatically generates the ranking of sources based on Google Trends. The system uses Pytrends⁹ that allows a direct download of the data from Google Trends in an organised structure, i.e., a Pandas data frame. Pytrends allows specifying all the parameters available through the GUI of Google Trend, namely localization (Country), time interval (daily, from 2004 to 36 hours before current time), categories (the topic of searches), and type of search (Google Search, Google Images, Google News, Google Shopping and YouTube search). The service allows one to compare up to 5 keywords and extract the relative interest among these words. The index is additionally standardised in relation to one of the provided keywords. To have comparable results, we need to identify a common benchmark word to be used in every chunk that satisfies the property of being the most popular word in all the queries. We adopted an iterated procedure that uses the most popular website name or, in case it is not possible to identify it, the translation of a common word that scores highest.¹⁰ Google Trends measure is weekly and is then further aggregated at the quarter level, computing the maximum, and then averaged to obtain the mean value over the year.

The second dimension is stability in the provision of OJVs data. Each source–country pair is flagged as unstable if it does not meet at least one of a set of conditions, namely, on the maximum number of months without observations, on the variability of the time series and the number of outliers. Stability is enriched by additional information for each source, in cases in which it is not developed or currently not present in the final dataset.

We use a time series of online job ads to estimate sources’ stability over time. Source stability is a common criterion used to evaluate the reliability and trustworthiness of a data source, providing a robust time series.

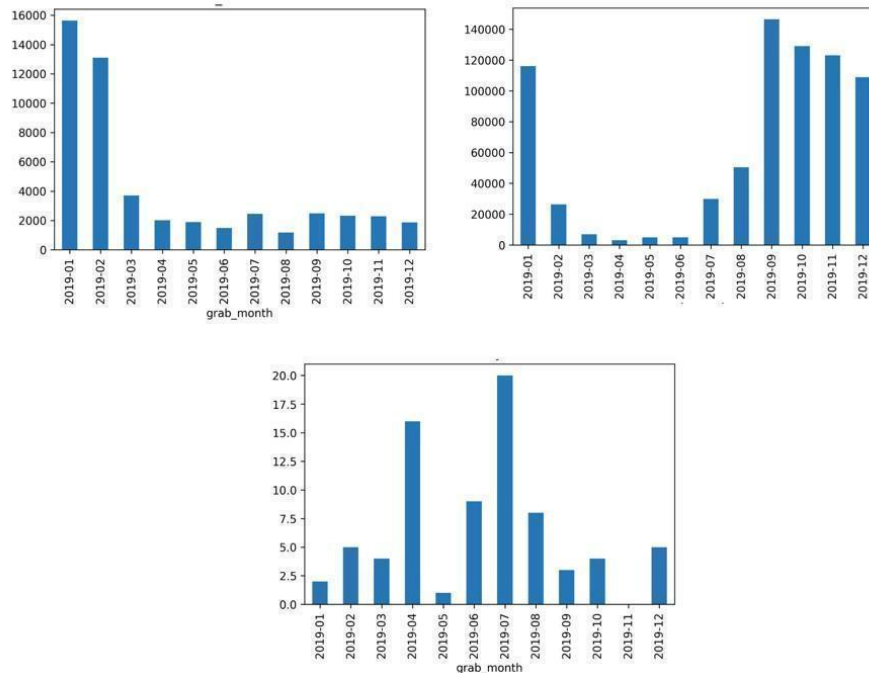
then scaled on a range of 0 to 100 based on a topic’s proportion to all searches on all topics. (iii) Different regions with the same search interest for a term don’t always have the same total search volumes.”

⁸ <https://support.google.com/trends/answer/4359597?hl=en&topic=4365530>

⁹ An unofficial open-source API for Google Trends

¹⁰ Google guidelines stress the importance of punctuation to refine searches. Avoiding the use of punctuation allows including associated queries implicitly (for example searching for Careerjet automatically includes Careerjet jobs in Milan and all other associated queries; on the contrary searching for “Careerjet” would exclude all other associated queries). The drawback of this method is that “[n]o misspellings, spelling variations, synonyms, plural, or singular versions of your terms are included”.

Figure 4 Example of three different Country's source time series.



The stability assessment is performed following a set of criteria:

1. Percentage of missing online job ads in the time series
2. Unavailability of online job ads for at least three months (even not consecutive) in the time series
3. The *diffrange* criterion is the ratio $(\max - \min) / \max$ of the number of online job ad for each time point¹¹
4. The mean of the variation for each time point of the series;
5. The relative standard deviation of the time series;
6. The presence of outliers, computed as those observations that fall outside of the interval $(Q1 - 3 \cdot IQR, Q3 + 3 \cdot IQR)$, being IQR the interquartile range.

For each criterion, a threshold is identified and fine-tuned by following optimality criteria. We provide some examples to explain the approach. The source in the upper right panel of Figure 4 is marked as unstable as it is above the threshold identified for the *diffrange* criterion sets to 90% (above 96% for this source). The upper left panel shows the case of a source that is marked as unstable due to (i) the presence of an outlier and (ii) the value of the relative standard deviation 1.14 above the threshold, i.e., 1. Finally, the lower central panel shows the case of a source that is classified as unstable because of the high mean of the variation, 67% against a threshold of 45%. In the above exercise setting the threshold is a crucial aspect.

2.4. Selection of sources for development

The final step of the landscaping is the selection of sources. The final indicator is based on the combination of the qualitative and quantitative assessments of the sources to establish a decision rule to identify sources to be enrolled in the development phase. Prioritised sources should be engaged first through agreements with the owners to access the data.

¹¹ Notice we plan to use the second min, and max value as this does not penalise sources having a one-time issue with missing a (or very low) number of online job ads.

Table 1 Score definition.

Score	Definition	Cases (AHP score quartile, expert's rank quartile)
1	Sources with position in Q1 of expert's rank and Q1 of AHP score.	(Q1,Q1)
2	Sources with position in Q1 or Q2 of expert's rank and Q1 or Q2 of AHP score. Exclude the case (Q1,Q1).	(Q1,Q2),(Q2,Q1) and (Q2,Q2)
3	Sources with position in Q2 or Q3 of expert's rank and Q2 or Q3 of AHP score. Exclude the case (Q2,Q2).	(Q2,Q3),(Q3,Q2) and (Q3,Q3)
4	Sources with position in Q3 or Q4 of expert's rank and Q3 or Q4 of AHP score. Exclude the case (Q3,Q3).	(Q3,Q4),(Q4,Q3) and (Q4,Q4)
5	Sources with distance between position in expert's rank distribution and AHP score distribution larger than 1 quartile.	All the others, e.g. (Q1,Q4), (Q3,Q1)

The final score is computed following a simple decision rule. The outputs of the qualitative and quantitative ranks are mapped to the quartile of belonging in the respective distribution of values. Then a five-levels score is defined considering the joint distribution of the two ranks. The rule is provided in Table 1. The lowest the score the highest the position in the joint distribution, meaning a higher priority assigned to the source. The general proposal, which can be adapted to specific country conditions, is to prioritise development sources with scores 1 and 2 and to inspect those with a score of 5. While the former sources are those exhibiting the highest quality and relevance scores, the latter have a high discrepancy between the two assessments and therefore require further analysis to understand what led to this outcome.

REFERENCES

Goepel, K.D. (2018). Implementing an Online Software Tool for the Analytic Hierarchy Process (AHP-OS). *International Journal of the Analytic Hierarchy Process*, Vol. 10 Issue 3 2018, pp 469-487, <https://doi.org/10.13033/ijahp.v10i3.590>

Mezzanzanica, M. and Mercurio, F., European Training Foundation, 'Big data for labour market intelligence - an introductory guide', ETF, Turin, 2019: <https://www.etf.europa.eu/en/publications-and-resources/publications/big-data-labour-market-intelligence-introductory-guide>

Mercurio, F. and Mezzanzanica, M. [PHASE 2: FEASIBILITY STUDY FOR TUNISIA AND MOROCCO TO IDENTIFY, VALIDATE AND RANK WEB JOB VACANCY SOURCES – PRACTICAL GUIDANCE, ETF](#), Turin, 2019

Volodymyr Sarioglo, Olexandr Cymbal, [Big Data for LMI, Web Labour Market Landscaping Ukraine](#), ETF. 2020.

LaValle, I. H. and Bard, J. F., Review of Golden, B.L., Wasil, E.A. and Harker, P.T., The analytic hierarchy process: applications and studies, reviewed in *Interfaces*, Vol. 21, No. 2, 1991, 94–98.

European Centre for the Development of Vocational Training (Cedefop). "Online job vacancies and skills analysis: a Cedefop pan-European approach." (2019). <https://www.cedefop.europa.eu/en/publications/4172>