



Online job vacancies and skills analysis

A Cedefop pan-European approach



Online job vacancies and skills analysis

A Cedefop pan-European
approach

Please cite this publication as:
Cedefop (2019). *Online job vacancies and skills analysis:
a Cedefop pan-European approach*. Luxembourg:
Publications Office. <http://data.europa.eu/doi/10.2801/097022>

A great deal of additional information on the European Union
is available on the Internet.
It can be accessed through the Europa server (<http://europa.eu>).

Luxembourg:
Publications Office of the European Union, 2019

Copyright © European Centre for the Development
of Vocational Training (Cedefop), 2019
All rights reserved.

PRINT

ISBN: 978-92-896-2851-8
doi:10.2801/567701
TI-04-19-203-EN-C

PDF

ISBN: 978-92-896-2850-1
doi:10.2801/097022
TI-04-19-203-EN-N

**The European Centre for the Development
of Vocational Training** (Cedefop) is the European Union's
reference centre for vocational education and training.
We provide information on and analyses of vocational
education and training systems, policies, research and practice.
Cedefop was established in 1975
by Council Regulation (EEC) No 337/75.

Europe 123, 570 01 Thessaloniki (Pylea), GREECE
PO Box 22427, 551 02 Thessaloniki, GREECE
Tel. +30 2310490111, Fax +30 2310490020
Email: info@cedefop.europa.eu
www.cedefop.europa.eu

Mara Brugia, *Acting Executive Director*
Tatjana Babrauskiene, *Chair of the Management Board*

Foreword

The supply of the right skills at the right time is essential for competitiveness and innovation, as the European Commission's 2016 *New skills agenda* ⁽¹⁾ emphasised. In today's dynamic world, understanding job requirements and being able to react fast to emerging skill needs is key to addressing skills mismatch and success on the labour market. Understanding this is crucial for citizens, employers, training providers, career guidance staff and policy-makers.

Faster and more detailed evidence on skills requirements can help in designing education and training programmes or guiding people's career paths within and across sectors, regions and countries. Real-time information is particularly relevant for upskilling and reskilling of adults at all levels, as continuing training and active labour market initiatives can respond more rapidly than initial vocational education and training for young people.

Unfortunately, timely information on employers' skill requirements at EU level is scarce. Traditional statistical methods based on surveys do not provide just-in-time and detailed data comparable across the EU. Scrutinising job vacancies published online in several countries can be an alternative way to offer insights into skills required and their dynamics.

However, freely accessible sources of such data, covering all 28 EU Member States, do not yet exist. As generating comprehensive evidence on skills demand to inform vocational education and training (VET) is one of its main strands of activity, Cedefop has decided to develop a pan-European system for online vacancy analysis. This task is highly complex and challenging. Not only does the use of online platforms for job search and recruitment vary widely across the EU; multilingual job vacancy notices in some Member States require the capacity to capture and analyse from more than the 24 official EU languages.

⁽¹⁾ European Commission (2016). *A new skills agenda for Europe: working together to strengthen human capital, employability and competitiveness: communication from the Commission to the European parliament, the Council, the European economic and social committee and the Committee of the regions*. COM2016 (381) final.

<https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A52016DC0381>

With this work, Cedefop extends its well-established battery of instruments to produce skills and labour market intelligence. The new real-time information system complements our regular long-term Europe-wide skills forecasts, our European skills and jobs survey, our European skills index, and the information available through the Skills Panorama.

Accompanying the early release of data in March 2019, this booklet marks the first major steps towards providing timely information on current and emerging skills needs. It presents key features of our new system and gives a concise overview of Cedefop's work, including the methodology and analytical approach we have used so far, to help understand the results. It also points to contextual issues and limitations that must be kept in mind when using the data.

The early data release aims primarily at showcasing the potential of analysing online vacancy notices in seven countries (Czech Republic, Germany, Ireland, Spain, France, Italy and the UK). In 2020, the results for all Member States will become available and will then be regularly updated.

In its vision for VET beyond 2020 ⁽²⁾, the European Commission's Advisory Committee for Vocational Training has stressed the need to underpin national VET and employment policy design with skills anticipation and foresight mechanisms and faster detection of trends. With this new, EU-wide data source, we hope to enrich the skills and labour market intelligence currently available.

Mara Brugia
*Acting Executive
Director*

Alena Zukersteinova
*Acting Head of department
for skills and labour market*

⁽²⁾ European Commission (2018). *The future of vocational education and training post 2020: opinion of the Advisory Committee on Vocational Training*.
<https://ec.europa.eu/social/BlobServlet?docId=20479&langId=en>

Acknowledgements

This publication was drafted by experts of Cedefop's department for skills and the labour market Vladimir Kvetan and Jiri Branka, under the supervision of the Acting Head of department, Alena Zukersteinova. Special thanks go to Cedefop experts Stelina Chatzichristou, Jasper van Loo and Steven Bainbridge for reviewing the manuscript.

The publication is a result of work undertaken within Cedefop's project *Real-time labour market information on skill requirements: setting up the EU system for online vacancy analysis* (AO/DSL/VKVET-GRUSSO/Real-time LMI2/006/16). It is based on results and background reports delivered by a consortium of research institutes led by CRISP, the Interuniversity Research Centre for Public Services at the University of Milano-Bicocca.

Table of contents

Foreword	4
1. Introduction	9
2. Online job vacancies and portals in the EU	12
2.1. What drives the use of OJVs?	14
2.2. Types of job portals and their features	14
2.3. How do employers choose recruiting channels?	17
3. How to collect online job vacancies	18
3.1. Data ingestion	18
3.2. Data pre-processing	21
3.3. Information extraction	22
3.4. Data quality limitations	24
4. What information and for whom	26
4.1. Information and knowledge produced	26
4.2. Potential users of vacancy-based analysis	27
5. Way forward	30
Abbreviations and acronyms	31
References	32

Figures

Figure 1. Job openings, OJVs and OJV portals	13
Figure 2. Most common globally oriented sectors and employers	16
Figure 3. OJV data collection and production process	18
Figure 4. Structure of vacancy sources by country	19
Figure 5. Data ingestion process	20
Figure 6. Information extraction process	22

CHAPTER 1.

Introduction

Over the past decade, the use of the internet for posting job vacancies has significantly increased. While job vacancies posted online were at first predominantly for highly skilled workers, nowadays – due to widespread internet access and use and increased ICT literacy – these platforms contain job offers for almost all occupations and skill levels. Apart from making it easier to match employers and jobseekers, the increasing use of online job vacancy (OJV) portals has great potential for labour market and skills analysis.

OJVs are a rich source of information about skills and other job requirements which is difficult to gather via traditional methods. Access to this information can provide several opportunities: help labour market actors understand better skill demand and its dynamics; enable individuals to make better career and skill development choices; support employers to develop or adjust human resources (HR) policies; help policy-makers make more informed decisions; and improve the targeting of employment services, guidance counsellors and learning providers.

OJVs do not replace other types of labour market information and intelligence; on the contrary, the full potential of the data can only be unlocked by combining OJVs with conventional sources. By analysing skills and job requirements typically requested in occupations, OJV provide additional comprehensive, detailed and timely insights into labour market trends and enables new and emerging jobs and skills to be identified early. This fills an important gap in current EU evidence on employer skill demands.

It must be noted, though, that using OJVs as a data source for labour market analysis has several limitations:

- (a) vacancies in some sectors and occupations are over-represented in OJV portals;
- (b) the use of OJV portals differs across and within the countries due to the digital divide and different employment structures;
- (c) skills listed in a vacancy notice do not reflect the full job profile; employers tend to list only critical skills and qualifications to ‘filter’ job applicants;

- (d) vacancy notices have to be machine-readable and use a standardised vocabulary and, given the quantity of data, some simplifying assumptions have to be made;
- (e) the same vacancy notice may be published on several web sites and not necessarily correspond to an actual job opening.

To investigate the potential of OJVs as a data source, and taking account of their limitations, Cedefop decided to develop a pan-EU system to gather and analyse data in OJVs ⁽³⁾. This booklet presents the key features of this system. It accompanies the early data release, the first major step in an effort to provide timely information on current and emerging skills needs in seven countries (Czech Republic, Germany, Ireland, Spain, France, Italy and the UK). The aim is to raise awareness of the data's nature and their value by:

- (a) defining what OJVs are, how employers use OJV portals, and what trends are influencing the OJV market across the EU (Chapter 2);
- (b) explaining how to collect OJVs and transform the data into useful evidence, as well as outlining quality control tools and methods (Chapter 3);
- (c) providing an overview of the type of information that the Cedefop system will produce and its potential value to users (Chapter 4);
- (d) outlining next steps (Chapter 5).

The fully operational system, working across all EU Member States, will be available by the end of 2020. As increasing data quality and general acceptance of the data is an important part of our effort, Cedefop actively cooperates with other EU institutions and initiatives. The need for a multilingual taxonomy for occupation and skills allows Cedefop to use, test and eventually enrich the European skills/competences, qualifications and occupations (ESCO) classification developed and managed by the European Commission (DG Employment). There is also continuing cooperation with Eurostat's big data task force and network of European statistical systems (ESSnet), exploring the potential for Cedefop's system to produce official statistics.

Additional information to be found on the Cedefop website includes a synthesis report giving an overview of the OJV market and the trends

⁽³⁾ For a brief project description see: <http://www.cedefop.europa.eu/en/events-and-projects/projects/big-data-analysis-online-vacancies>

influencing it, as well as background reports for each EU Member State that provide detailed insights into the OJV market in that country ⁽⁴⁾.

Key milestones in the development of Cedefop's online job vacancy analysis system

- 2015: a feasibility study showed that a pan-EU system can provide detailed and unique information on skill demand.
- 2017: mapping of the OJV landscape across Member States provided the insight needed to understand the context of OJVs and their role in employer recruitment practices; it also identified the most important OJV providers.
- March 2018: the methodology was discussed and endorsed by leading researchers and data users.
- April 2018: data collection from more than 1 000 vacancy sources began; many more sources have been added since then.
- September 2018: downloaded and classified vacancies were validated by a team of experts; the classification mechanisms of Cedefop's system and data quality mechanisms were substantially improved.
- December 2018: about 55 million vacancies had been downloaded, processed and analysed (since July 2018).
- March 2019: early data release, showcasing the potential of results for seven countries.

⁽⁴⁾ Cedefop (2019). *The online job vacancy market in the EU: driving forces and emerging trends*. www.cedefop.europa.eu/en/publications-and-resources/publications/5572
Dedicated reports on OJVs posted on EURES and global OJV portals are also available.

CHAPTER 2.

Online job vacancies and portals in the EU

This chapter provides key highlights of the OJV portal landscape in Europe. It identifies the trends driving their use; presents different types of portal and their features; and reviews the links between employer characteristics and their use of OJV portals.

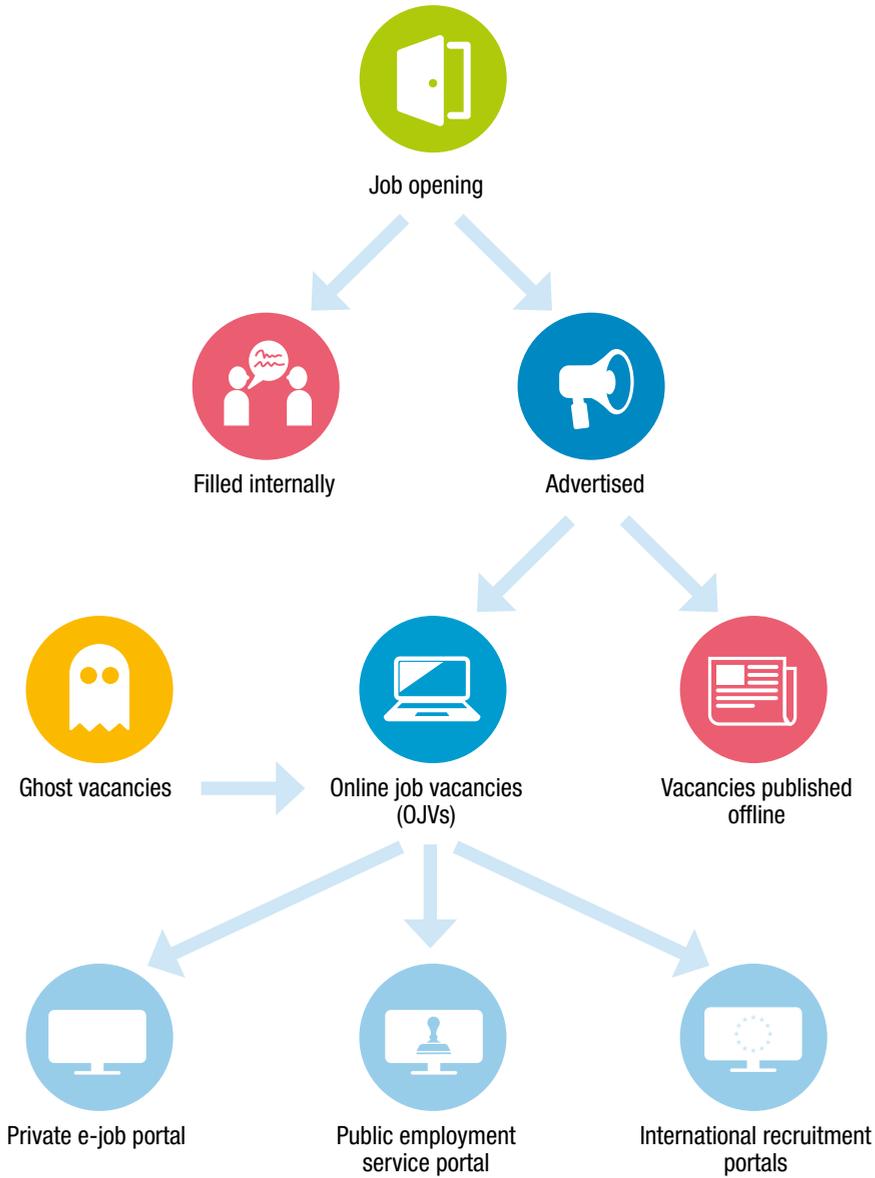
A job vacancy is a job opening that is newly created, unoccupied, or about to become vacant ⁽⁵⁾. To fill that job opening, the employer seeks a suitable candidate either internally or on the labour market. Vacancies can be advertised in various ways: posters on the enterprise's premises, advertisements in newspapers, or online posting on various corporate, private, public websites and portals or even on social networks. Job openings in specific sectors, for example in the oil industry, may be advertised on specialised global portals.

Not all job advertisements posted online are linked to an actual vacancy related to a job opening. Often employment agencies or enterprises will place 'ghost vacancies' to test the market, attract spontaneous applications and even create reserve lists of suitable candidates ⁽⁶⁾. For occupations with high turnover, enterprises will permanently advertise vacancies to replace employees leaving. Identifying these types of vacancy is difficult; however, we believe that the share of such vacancies does not significantly distort the data and analysis of employer requirements.

⁽⁵⁾ As defined by Eurostat glossary: [https://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Job_vacancy_rate_\(JVR\)](https://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Job_vacancy_rate_(JVR)) [accessed 18.12.2018].

⁽⁶⁾ Eurostat; Swier, N. et al. (2017). *Final technical report: ESS net Big data specific grant agreement SGA-1*. https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata/images/5/5d/WP1_Deliverable_1.3_Final_technical_report.pdf [accessed 18.12.2018].

Figure 1. Job openings, OJVs and OJV portals



2.1. What drives the use of OJVs?

Cedefop mapped and analysed the OJV landscape in all 28 EU Member States to find out how OJVs are used in different countries and portals. Two major factors drive the use of OJV (7):

- (a) digitisation and rising computer literacy increase the possibility of using the internet for job search. Digitisation also changes skill needs across occupations and digital tools are increasingly used by public employment services and OJV portals to match jobseekers with employers. The internet becomes a more suitable channel for advertising jobs, more attractive for both jobseekers and employers;
- (b) economic growth and changes in work organisation may lead to skill shortages, affecting employers' HR practices and use of recruitment channels to attract suitable candidates. The internet allows employers to expand their reach to other regions or countries at relatively low cost. Employers can use specialised websites to target specific professions, people with specific employment needs (such as the elderly), or to find people for short-term/project-based work (such as gig jobs).

2.2. Types of job portals and their features

In most EU countries, national public employment services (PES) operate a public portal. Some countries, however, have more than one public portal. These may serve jobseekers at various administrative levels of the country (federal, regional), cater for jobs for specific target groups (disadvantaged groups or highly skilled professionals), or serve as recruitment platforms for public sector jobs to ensure transparency.

PES portals primarily offer employment reactivation and intermediation services to the unemployed and low-skilled; some countries grant access to PES portals only to registered jobseekers. Although, in most EU countries, job advertisements in public portals target a specific part of the workforce, a shift can be observed in some, with the PES taking a proactive role in helping jobseekers with wide range of 'high-level jobs'.

(7) For more details see: <https://skillspanorama.cedefop.europa.eu/en/blog/understanding-online-labour-market-eu> [accessed 18.12.2018].

Moving away from the approach where employers send their vacancies directly to the PES, some public portals have set up partnerships with private ones and offer a wider range of job advertisements. Other public portals have invested in automated systems that ease skills-based matching of jobseekers to employers. In some countries, employers are legally obliged to post vacancies on the PES online job portal. While not always enforced, it signals the aim to diversify the qualification levels of vacancies on PES portals.

In contrast to free-of-charge public portals, private portals are businesses operating on a for-profit basis. In the past these portals have primarily targeted high-skilled workers in selected sectors or occupations, but the most popular ones aim to cover all vacancies.

‘Primary job portals’ are those where employers directly publish their vacancies; enterprise websites can also be considered primary sources. However, the most important market players tend to be ‘aggregators’: these are job search engines that can repost job vacancies gathered from various OJV portals.

Many nationally owned and operated portals have developed directly out of or remain connected to national newspapers. The role of social media in online recruiting, mainly for higher skilled, specialist and management posts, is expanding, but such channels act to promote vacancies already published on OJV portals.

Global labour mobility and the specific nature of skills required for certain jobs or occupational groups led to the creation of cross-national portals, advertising jobs in several countries. Such portals cater for globally oriented sectors or globally oriented employers (Figure 2).

An example of a cross-national portal is EURES⁽⁹⁾, an EU-level job mobility portal that hosts advertisements for jobs in all EU countries. Although it has primarily acted as an aggregator, bringing together vacancies published on national PES portals, it can now also be used by private sector employers directly to place their job advertisements.

⁽⁹⁾ EURES portal: <https://ec.europa.eu/eures/public/homepage>

Figure 2. **Most common globally oriented sectors and employers**



2.3. How do employers choose recruiting channels?

The use of recruiting channels is at the discretion of individual employers and their HR department. However, it is possible to identify various factors affecting preferences for using particular channels:

- (a) regulation: in some EU countries, legislation obliges employers to advertise their vacancies on PES portals, though in many cases this is not enforced. Instead and/or in addition, employers in these countries prefer to use private portals;
- (b) type of jobs: jobs requiring higher qualifications or in fast-growing sectors are more often advertised on private portals;
- (c) type of jobseeker: PES portals are mainly used by employers to target unemployed jobseekers. For 'career movers' employers tend to advertise on private portals;
- (d) type of employer: larger companies, firms in urban areas and/or foreign-owned companies are more likely to advertise on private portals;
- (e) level of labour shortages: labour shortages linked to declining unemployment in the EU, increasing needs to replace retiring workers and other trends ⁽⁹⁾ result in growing recruitment difficulties. This causes employers to diversify recruiting channels and so increase the number of jobs advertised online.

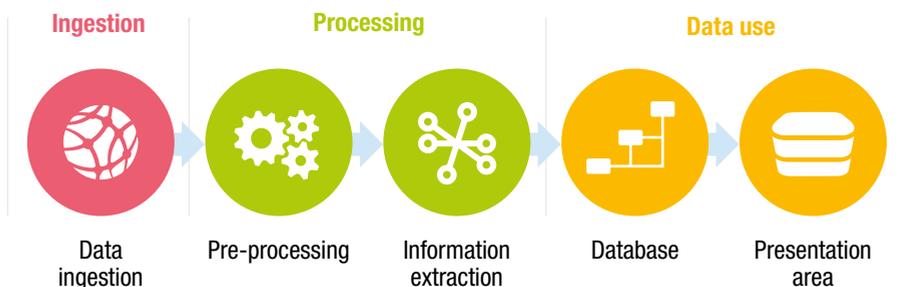
⁽⁹⁾ For more information see: Cedefop (2018). *Skills forecast: trends and challenges to 2030*. Luxembourg: Publication office. Cedefop reference series; No 118.
http://www.cedefop.europa.eu/files/3077_en.pdf

CHAPTER 3.

How to collect online job vacancies

The process of collecting online job vacancy data involves several steps (Figure 3). It starts with selecting the information source (Section 2.2). The identified sources enter the ‘data ingestion’ phase, which involves identifying vacancies available in the source and downloading their content. During next ‘pre-processing’ phase, the downloaded data are cleaned of irrelevant content. Following a process of ‘information extraction’, it translates the relevant content of the OJV into a database organised into taxonomies and classifications. This final database feeds various automatically updated dashboards or analysis to produce labour market and skills intelligence.

Figure 3. **OJV data collection and production process**



Source: Cedefop.

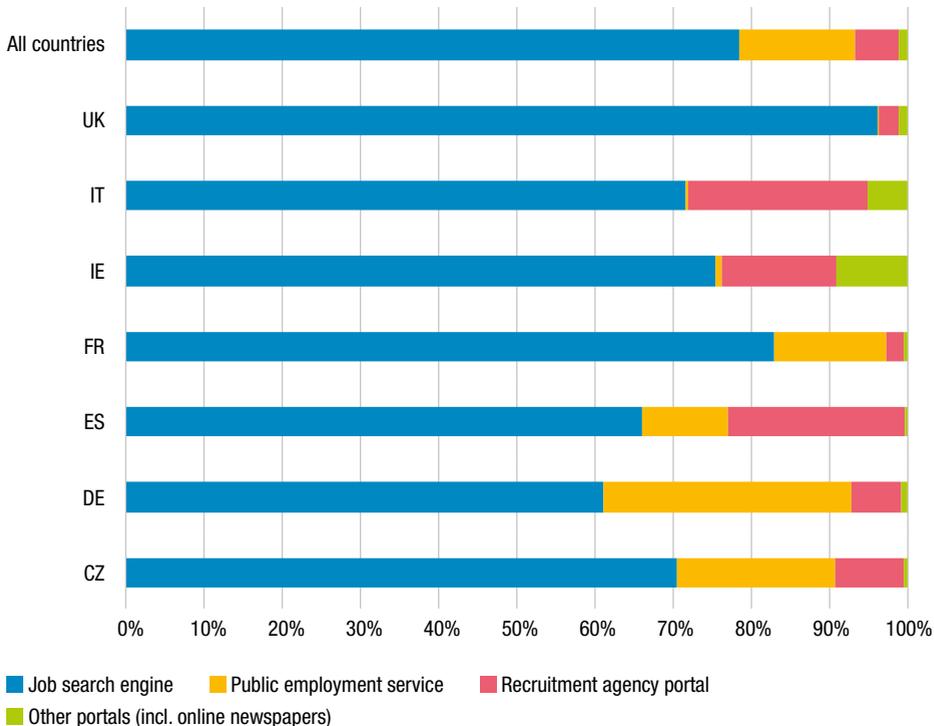
3.1. Data ingestion

To maximise the quality of the information extracted from vacancies, websites were prioritised and ranked by information provided: vacancy release dates; the frequency and regularity of vacancy updates; territorial and sectoral coverage of vacancies; and structured fields in the vacancy notice. About

300 sources were selected (covering more than 60 000 websites ⁽¹⁰⁾), ranging from two sources in Malta to over a hundred in Germany ⁽¹¹⁾.

A wide variety of sources operate within countries (Figure 4). In Ireland, Italy and the UK, the role of PES portals is minimal in contrast to Czech Republic, Germany, Spain and France where it is important. The importance of online newspapers in Czech Republic in the vacancy market is noteworthy.

Figure 4. **Structure of vacancy sources by country**



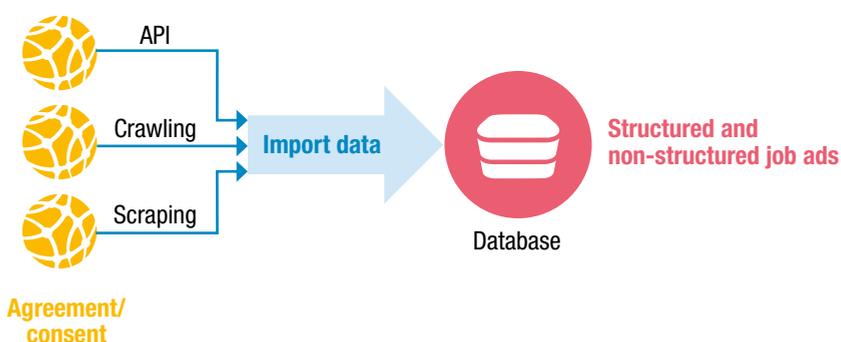
Source: Cedefop.

⁽¹⁰⁾ One source, such as an aggregator website, may contain links to several hundreds or even thousands of websites (such as company websites) where the actual vacancy is placed.

⁽¹¹⁾ The exact number of actual sources is subject to dynamic change and is available via the Cedefop web portal: <http://www.cedefop.europa.eu/skills-online-vacancies>

All selected sources are considered in the data ingestion process. Data from the sources can be accessed from website front ends (the part of the website visible for users) and/or back ends (databases and systems powering the website, which its operator can provide access to). Various techniques can be used:

Figure 5. **Data ingestion process**



Source: Cedefop.

- (a) scraping is used to extract structured data from websites. Using web scraping implies that data are already structured on the web page and can be extracted precisely by knowing the exact position of each field on the web page. As specific web scrapers must be programmed for each website, this is ideal for sites which contain many vacancies;
- (b) crawling uses a programmed robot to browse web portals systematically and download their pages. Crawling is much more general compared to scraping and is easier to develop. However, crawlers collect much more website noise (irrelevant content) and more effort is needed to clean the data before further processing;
- (c) direct access via application programming interface (API) allows download of vacancy content directly from OJV portal databases. This direct access requires a formal agreement from the website operator and is subject to maintenance and agreement costs. Data collected in this way have the highest quality of the different methods and can be downloaded much faster.

OJV portal owners were informed about the intended data collection and, where possible, written agreements were made to formalise cooperation between Cedefop and the portal. In many cases owners granted direct access to data via API. Where this was not possible, scraping and crawling were used. Rejection of access to data sources was negligible (around 0.1% of indexed vacancies) and came mainly from small websites with low market coverage.

After the ingestion phase, the language of each vacancy is detected, independent of the website it originated from. A vacancy from one country may be in a language different from the official one of the country. Cedefop's system can recognise and process all EU official languages as well others that are widely used in some countries, such as Catalan, Basque, Gaelic and Russian.

3.2. Data pre-processing

Vacancy sources vary in quality and content. To develop a database suitable for subsequent analysis, the pre-processing involves the following actions:

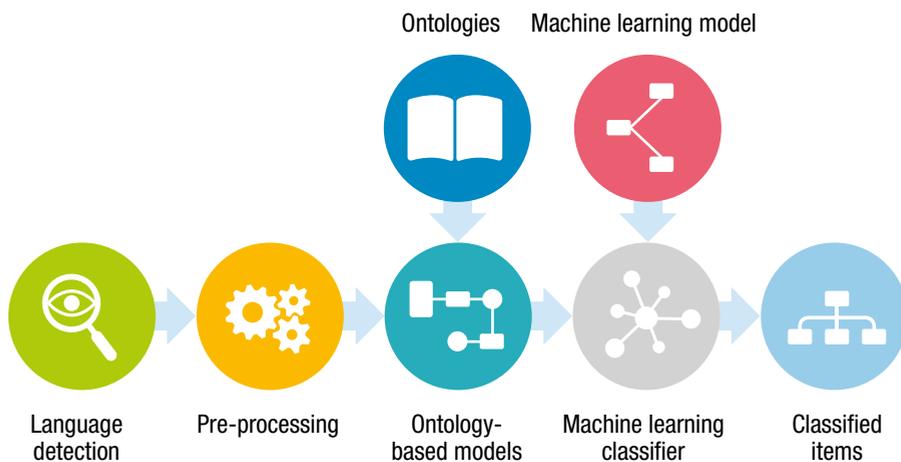
- (a) cleaning: OJVs are designed to attract the most suitable candidate, not to provide clean data for labour market analysis. Alongside analytically useful information, OJVs can also download 'noise' (such as advertisements, unticked options in drop-down menus, company profile presentations). Cleaning is a sequence of activities to remove 'noise' from the data to improve their quality and prepare them for the following phases;
- (b) merging: employers often post a vacancy on more than one portal. Aggregators increase the chance of finding duplicates. These are not desirable for the final analysis but, in the initial part of the pre-processing phase, duplications can enrich the data by combining the information on the vacancy posted at different places;
- (c) de-duplicating: once data for the same vacancy have been merged, it is necessary to remove duplicates from the analysis. An OJV is considered a duplicate if the description and job location are the same as another job advertisement in database. Vacancy metadata ⁽¹²⁾ (such as reference ID, page URL) are also used to identify and remove job vacancy duplicates on aggregator websites.

⁽¹²⁾ Metadata are defined as the data providing information about one or more aspects of the data; the term is used to summarise basic information about data which can make tracking and working with specific data easier.

3.3. Information extraction

The pre-processed data pass to the information extraction phase. This has two important features: ontologies and a machine learning model (Figure 6).

Figure 6. **Information extraction process**



Source: Cedefop.

Ontologies create a framework for processing and analysis of online vacancies. The Cedefop system uses both standard ⁽¹³⁾ and custom ⁽¹⁴⁾ ontologies related to the skills and jobs market. Using the power and flexibility of machine learning algorithms, the content of job advertisements is matched to available ontology terms, such as occupation, industry, region of the workplace or type of contract. Each vacancy is processed in its original language, which has its own unique logical framework.

The process first tries to classify the vacancy using text matching and/or similarity with relevant ontology (such as vacancy title with ISCO occupation

⁽¹³⁾ ESCO for occupations and skills, NACE for industry, NUTS for places of work, ISCED for education level.

⁽¹⁴⁾ Developed from information available in vacancies, such as type of contract, experience, salary or working hours.

titles). If no result is obtained, a machine learning algorithm ⁽¹⁵⁾ decides on the classification.

To improve the accuracy of the classification, ontologies are continuously updated and enriched. The results of the machine classification process are regularly validated by experts. Outcomes of this process (proposed corrections) are used to improve machine classification accuracy. The semi-automatic augmentation process adds new terms and synonyms not yet included in ontology through machine learning algorithms; it is approved by expert checks. Ontologies can also be updated manually, to reflect new information (new trends in occupations or even updating the whole underlying ontology such as ESCO).

The machine learning process must be applied separately for each variable (such as occupation or region) and for each language. Each model must be trained to fit best the variable and the language, beginning with creation of the training set ⁽¹⁶⁾. Training sets are defined separately for every language pipeline ⁽¹⁷⁾ depending on available vacancies, country size and composition of vacancies: each training set includes over 70 000 vacancies. The training set is then divided in three parts:

- (a) 60% cases for training of the machine;
- (b) 20% for testing of performance;
- (c) 20% for evaluation of accuracy (quality measure).

The training itself combines different methods. Experts can classify vacancies manually to create a sufficient basis for initiation of machine learning algorithm. The outcomes of the expert validation (corrections of misclassified vacancies) can also be incorporated into the training set. Once ready, the set can be used to train the machine learning model and subsequently to test its accuracy. Currently, the Cedefop system uses machine learning only to classify occupations; application of machine learning models to other variables is under development.

⁽¹⁵⁾ The machine learning algorithm uses statistical techniques to give computers the ability to 'learn' (progressively improve performance on a specific task) without being explicitly programmed.

⁽¹⁶⁾ Training set: in machine learning, a set of examples used to fit the parameters of the model.

⁽¹⁷⁾ Language pipeline denotes the set of data processing elements referring to a specific language.

3.4. Data quality limitations

Online vacancy data have limitations in terms of representativeness and must be treated with appropriate caution. Key considerations regarding the reliability of vacancy data include:

- (a) OJVs represent only part of the jobs demand; not all job vacancies are advertised on-line and some jobs are more likely to be advertised on-line than others. It can be expected that data are subject to occupational or qualification bias;
- (b) in most countries, the OJV market comprises multiple actors with different business models. There is usually no single source of all online job ads; the volume, variety and quality of the data depend on the portals selected for analysis;
- (c) penetration of OJV markets varies in and across countries and changes over time. Low internet penetration and lack of basic digital skills across the population are the key parameters influencing employers' decisions on extent of use of OJV portals as a recruitment channel;
- (d) vacancies and the information they contain need to be processed, using the various tools described, to produce viable data. Even the most up-to-date techniques are still subject to errors;
- (e) all ontologies are developed to sort and organise a diverse and complex universe. Despite enormous effort in developing them, they are still imperfect and may contain systemic errors that can only be corrected over time.

A detailed and robust quality assurance process has been designed and adopted for Cedefop's project to tackle these challenges. A set of tools and checks are being used to mitigate potential data bias, with data quality covering three major aspects:

- (a) documentation of the process, suggestions and revisions; ontologies and training sets will be published under creative common licences to enable potential contributors to evaluate and improve them;
- (b) validation of data and information is a key pillar of the project. Its objective is to identify data errors and anomalous patterns; this is not to measure overall data quality, but to improve dictionaries and fine-tune the classification and machine learning process. The outcomes of the validation are also used as an input to the machine learning process (for example, by adding terms to dictionaries to map educational titles correctly

- or by defining wrong associations between occupations and skills). This ensures permanent development and improvement of the system;
- (c) validation is executed manually over smaller data sets (about 1 000). The results are then translated into improvements in the model or ontology enrichment. The process is carried out again and quality measured after fixing. This iterative process (execution - validation - fixing - execution) can be repeated until an appropriate level of quality is reached;

Ensuring high data quality will position OJV-based analyses among the most relevant sources of labour market information. Cedefop intends to combine OJV analyses with conventional sources, such as household and employee surveys and skills forecasts to gain a comprehensive view of trends in skills demand and supply in Europe.

CHAPTER 4.

What information and for whom

4.1. Information and knowledge produced

Obtaining information from OJV analysis is primarily driven by the need to understand employers' skill requirements. It would augment already available information to produce more detailed labour market and skills intelligence and contribute to better policy-making. OJVs generally have the potential to add more clarity to questions such as:

- (a) for which occupations is demand increasing most? In which sectors or regions?
- (b) what profiles are employers seeking to recruit in these 'top jobs'? What new types of job are emerging? For which type of jobs and skills are employers facing recruitment difficulties?
- (c) what are employers' demands for specific skills in specific jobs? How does this differ across countries, regions or sectors? What new skills are employers demanding? In which jobs?
- (d) considering the core set of skills required in different jobs, what possible career moves are there for jobseekers? Which jobs, although different, require a similar set of skills?

The data set will reflect the type and structure of the information available in OJV portals. Cedefop's mapping of the OJV market across the Member States, carried out as preparatory to developing the system described in Chapter 2, shows significant differences in the content of OJVs across web platforms, between and within countries. Nevertheless, a number of key common variables can be identified and classified using the information available from all sources:

- (a) occupations: using the job title and the job description, it is possible to identify the occupation and map it to ISCO classification at detailed level (4 digit). Such detailed information is usually not available in conventional data sets;

- (b) countries and regions: the place of work is typically well described in vacancies, allowing classification of the data by region. However, these data should be interpreted taking into account the digital divide between regions and countries, as well as the industrial composition of regions, which will influence hiring practices among local employers;
- (c) skills: the highest value-added information in OJVs is the description of employers' requirements for jobs, and skills in particular. Although employers rarely specify a full job profile in the vacancy, the skills mentioned can be considered as critical to assessing and selecting the right applicant for the post. In the Cedefop system, skills will be classified using the ESCO taxonomy. The data will also allow for analysis of the following skills-related phenomena:
 - (i) new and emerging skills: ESCO covers most skills identified in OJVs, but the data allow identification of new skills that are not yet included, such as new skills required to work with emerging ICT tools or technologies;
 - (ii) diffusing skills: in the ESCO skills pillar, a set of skills is attributed to a particular 'typical' occupation where these are usually present. It can happen that skills attributed to this typical occupation may appear in the different one, for which these skills were not yet assigned;
 - (iii) new synonyms for existing skills: new expressions defining particular skill already attributed to the occupation in ESCO;
- (d) time dimension: the high frequency of data collection will enable Cedefop to derive time trends and analyse the dynamics of skills requirements within occupations as well as short-term skills demand projections;
- (e) other variables: there are also other variables in the OJVs that can be analysed, such as wages, sectors, contract types, non-skill requirements or required experience. Cedefop will investigate gathering, analysing and presenting these variables, as they are dependent on the quality and cross-country comparability of the data.

4.2. Potential users of vacancy-based analysis

Cedefop OJV data provide valuable information for many users, including policy-makers, education and training providers, employment services, guidance practitioners, researchers and citizens. All of these groups

can potentially benefit from OJV- based analysis, but their needs differ substantially.

4.2.1. Policy-makers

Policy-makers, policy experts and policy analysts, along with trade unions, professional bodies and employers' associations, form a heterogeneous group with different needs. Potentially, OJV data bring the most added value to policy-makers wanting information about skills and other requirements. This information can, for example, help education and training systems to respond promptly to changing labour market needs, designing measures to help people acquire the skills employers demand. It can also support labour mobility and address skills mismatches. The eventual better matching of the EU workforce can increase economic growth and social inclusion. In the future, the data will enable assessments of how the demand for skills in occupations changes over time, identifying 'hot (or policy relevant) skills or technologies' that require greater attention.

4.2.2. Education and training providers

The education and training sector comprises various public and private organisations engaged in developing individuals' knowledge, skills and competences. Given ageing populations and highly dynamic labour markets, speedy access to information on current and expected skill demand is highly valuable in designing and updating training programmes and curricula. As the labour market relevance of education and training becomes increasingly important, education and training providers are challenged to keep up with changing skills requirements; the ability to of the system to provide a picture of skills required on the labour market can help in this. Moreover, the potential to identify vacancies at entry level may be highly useful in identifying the basic skill sets of new graduates.

4.2.3. Employment services and career guidance practitioners

Employment services aim at quick but sustainable (re-)entry of jobseekers into the labour market. Career guidance counsellors help students, the unemployed or employees to choose the right career, education, learning and skills development options. Information on jobs with higher than average demand among employers and proximity between occupations (indicating possible career moves and learning needs) can support well-informed career advice.

4.2.4. Citizens

Citizens are the ultimate target group, even though they benefit indirectly through improved skills policies, education programmes and guidance services. For direct data use, it is important to take account of the diversity of this group of potential users; they range from well-informed and skilled individuals to those with low or no (ICT) literacy. However, citizens are keen to access information that may improve their job and career prospects.

Data from online vacancies has specific qualities. It may be difficult for some users to interpret them without assistance from a qualified intermediary, or without understanding the right context or combining them with additional information.

It should be possible to create tools or apps that present OJV data in combination with other relevant labour market intelligence; as yet, at EU level there is no such tool. Therefore, it is not feasible, in the short term, to provide data to citizens. Consideration of the role of OJVs in the new Europass ⁽¹⁸⁾ or any other similar tool such as EURES lay outside the scope of the current Cedefop project.

4.2.5. Research community

Research into relations between occupations and skills in vacancies in Europe is still in its infancy. So far, researchers have relied on OJV data provided or sold by private entities. These data only cover some Member States or come from other countries, primarily the US. In many ways, the Cedefop project is ahead of anything else done in Europe in terms of its aim and scope. The research community can be seen as both user and partner in exploring the potential of this unique vacancy data set. Access to a data set covering all Member States, using the same methodology and taxonomies, will be a huge asset, not available elsewhere.

Granting access to the data set may enable Cedefop to steer research on OJV data in topics and issues of high policy relevance. This should also help Cedefop to exploit fully the data's potential and develop innovative, high-value added products for various end-users.

⁽¹⁸⁾ More information on the new Europass is available at: <https://ec.europa.eu/social/main.jsp?catId=1266&langId=en> and at: <https://ec.europa.eu/social/main.jsp?langId=en&catId=1223&newsId=2638&furtherNews=yes> [accessed 18.12.2018].

CHAPTER 5.

Way forward

Work on Cedefop's system continues. In March 2019, Cedefop will release first results, covering seven countries ⁽¹⁹⁾; the full EU coverage is expected by the end of 2020. The early data release will showcase the system's potential to provide unique information that is not available elsewhere. It will cover only a subset of variables suitable for wider dissemination but will include, in relation to vacancies posted on line, information about the professions/occupations/job titles and skills (based on ESCO). Geographic distribution of OJVs by occupation or skills required may also be available.

Sufficient data will be gathered to enable Cedefop to explore the full potential. Going beyond identifying the importance or intensity of skills as they appear in vacancies for various occupations, it should be possible to identify patterns in skill demand for occupations, skills, or groups of skills as they change across occupations, sectors, countries and over time.

Cedefop will continue to assess data quality and ability to provide interesting and robust messages. Understanding the needs of potential users, Cedefop will make data available as soon as they are of sufficient quality. Cedefop is working with Eurostat and ESSnet (network of national statistical authorities) to obtain the best possible quality data and be able to trust their results.

Recognising the data limitations as well as their original purpose, the results, gathered from online job vacancies in this project will provide an important piece of the EU skills intelligence puzzle. Data produced will provide an unprecedented level of detail on occupations and skill needs, far beyond that provided by traditional data collection methods. Understanding the immediate labour market situation and skills employers require will help various users develop strategies to mitigate future skills mismatches and contribute to better life chances for all EU citizens.

⁽¹⁹⁾ Czech Republic, Germany, Ireland, Spain, France and the United Kingdom.

Abbreviations and acronyms

API	application programming interface
ESCO	European skills/competences, qualifications and occupations
ESSnet	network of European statistical systems
OJV	online job vacancy
PES	public employment services
VET	vocational education and training

References

[URLs accessed 18.12.2018]

- Cedefop (2018). *Skills forecast: trends and challenges to 2030*. Luxembourg: Publication office. Cedefop reference series; No 118. <http://www.cedefop.europa.eu/en/publications-and-resources/publications/3077>
- Cedefop (2019). *The online job vacancy market in the EU: driving forces and emerging trends*. www.cedefop.europa.eu/en/publications-and-resources/publications/5572
- European Commission (2016). *A new skills agenda for Europe: working together to strengthen human capital, employability and competitiveness: communication from the Commission to the European parliament, the Council, the European economic and social committee and the Committee of the regions*. COM2016 (381) final. <https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A52016DC0381>
- European Commission (2018). *The future of vocational education and training post 2020: opinion of the Advisory Committee on Vocational Training*. <https://ec.europa.eu/social/BlobServlet?docId=20479&langId=en>
- Eurostat; Swier, N. et al. (2017). *Final technical report: ESS net Big data specific grant agreement SGA-1*. https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata/images/5/5d/WP1_Deliverable_1.3_Final_technical_report.pdf

Online job vacancies and skills analysis

A Cedefop pan-European approach



Over recent decades, online job portals have become important recruitment and job search tools. Beyond assisting skills matching, the job vacancies these portals gather can also be used to analyse labour market trends in real time, generating evidence that can inform education and training policies and help ensure that people's skills meet the needs of rapidly changing workplaces. These insights can complement skills intelligence based on information collected via traditional methods, such as Cedefop's Europe-wide skills forecasts, the European skills and jobs survey, and the European skills index. This booklet outlines the main features of online job vacancies and the key characteristics of Cedefop's new system to collect and analyse them. It accompanies the first release of results based on the collection and analysis of online job vacancies in seven EU Member States.



CEDEFOP

European Centre for the Development
of Vocational Training

Europe 123, 570 01 Thessaloniki (Pylea), GREECE
PO Box 22427, 551 02 Thessaloniki, GREECE
Tel. +30 2310490111, Fax +30 2310490020, Email: info@cedefop.europa.eu

visit our portal www.cedefop.europa.eu



Publications Office

