ANALYSING THE LABOUR MARKET THROUGH A WEB DATA SOURCE: THE SKILL DYNAMIC BASED ON ONLINE JOB ADVERTISEMENTS

Elena Catanese, Francesca Inglese, Annalisa Lucarelli, Giuseppina Ruocco

Abstract. Online advertisements in job portals and company sites have great potentials for job market analyses, providing detailed information on the jobs and skills required by employers. This new source of data – available thanks to advances in web crawling technologies, machine learning and big data techniques – represents an important use case of the ESSnet Web Intelligence Network project, whose main objective is to foster the integration of web data in official statistics. At European level, Online Job Advertisements (OJAs) are centrally collected, and stored in a Data Lab jointly developed by European Centre for the Development of Vocational Training (CEDEFOP) and Eurostat for producing new statistics on job vacancies. OJAs data provide a lot of information on job position characteristics and requirements by enterprises not collected through the official surveys on job vacancies. The objective of this work is to present the potentialities of OJA data in the analysis of the labour demand. In particular, the paper shows some analyses on Italian skill demand, over a period of four years (2019-2022), focussing on relevant changes at national and macro-regional level, by economic activity sectors and major occupational groups.

1. Introduction

The increasing amount of Online Job Advertisements (OJAs) in job portals and company sites has great potential to get detailed and timely insights into labour market trends. Advances in web crawling technologies, machine learning, and big data techniques, have given a boost to the growing interest of the National Institutes of Statistics in data acquired from the web for their use for statistical purposes. OJAs represent an innovative data source for the analysis of job market as it provides relevant information on job characteristics (e.g. occupation, location, type of contract, working time and salary), characteristics of the employer (e.g. economic activity sector), job requirements (e.g. education, skill and experience) and concerning the advertisement itself (e.g. job portal, publishing and expiring date of the ads). To exploit the potentialities of the information contained in OJAs, some considerations are necessary for their correct use: job vacancies and job advertisements cover different populations; they refer to different statistical concepts and to different measures of job openings. Therefore, it is a new data source that does not replace traditional surveys, but can

complement them to produce additional indicators, thus enriching the current official statistical production. The use of this new data source for the analysis of the labour market can be, in part, a response to overcoming the information gaps in existing statistics. More information and granularity represent today a necessity for EU policies and for decision-making in democratic societies. The Italian National Institute of Statistics (Istat) cooperates with several European countries on a project for the development of a Web Intelligence Network (WIN). The WIN project started from the results of the two previous projects (the ESSnet Big Data and ESSnet on Big Data II) implemented between the year 2016-and 2020. The main goal of this initiative within the European Statistical System (ESS) is to foster the integration of web data in official statistics. OJAs is one of the use case of the project, based on the joint work of Eurostat and CEDEFOP, finalized to collect, explore, process and analyse the OJAs for statistical purposes. CEDEFOP OJA system and ESS DataLab are the reference frameworks to analyse job market trends, based on the information contained in the OJA data. The objective of this work is to present the OJA potentialities for analysing in deep the labour demand. To this scope, it focuses on skill analysis, in particular on skill changes observed over a period of four years – from the first quarter 2019 to the fourth quarter 2022 – and broken down by geographical area, economic activity and occupation. Skill and occupation, as well as other variables, are extracted from OJAs through text analysis and classified using keyword lists (ontologies) and machine learning models. The keywords lists are based on the existing standard classification, namely ESCO (European Skills, Competences, Qualifications and Occupations). The paper is organised as follows. Section 2 reports some information on OJAs source compared with traditional sources (Job Vacancy Survey and the Excelsior Information System). Section 3 highlights the definition of OJAs and the conceptual differences between OJAs and Job Vacancies Statistics (JVS) based on official surveys. Section 4 describes OJA data pipeline and the standard classifications used for classifying the skill and occupation variables. Section 5 is dedicated to the skill analysis. In this section, we first highlight the importance of this type of data in providing an up-to-date picture of the skill requirements by employers, then, we present some results on skill demand dynamics broken down by macro-regional area, economic activity sector and major occupational groups. Finally, section 5 concludes the work, with main findings and challenges.

2. OJA: a non-traditional data source

OJAs are web data collected from many online sources, such as: job portals, company sites, social networks, employment websites, employment agencies, job search engines, online newspapers, public employment services, and employers organizations. Unstructured data in textual form – extracted from online sources through complex procedures – are transformed into information that can potentially be

used for statistical purposes. This new data source allows to overcome some limitations of traditional data (limited timeliness and granularity) and to enrich the current official statistical production. In Italy, in the context of the official statistics covering the demand side of the labour market, two main official sources supply information on the unmet labour demand. In particular, Istat has carried out since 2003 the quarterly Job Vacancy Survey (VELA), which produces estimates of the job vacancy rate. The survey covers all enterprises with one or more employees in all economic activities, from sections B to S, according to the international Classification of Economic Activities (NACE Rev.2). Activities in agriculture, forestry and fishing, those of households as employers, as well as activities of extraterritorial organizations and bodies and public administration and defence are excluded (NACE sections A, T, U and O). Data on job vacancies represent the stock of vacant posts at a given reference date. In particular, the time of recording is the last calendar or working day of the quarter. The quarterly job vacancy rate estimates are published at national and NACE section level. Istat's job vacancy rate production is based on the requirements of an EU framework regulation (No 453/2008). The European job vacancy rate is included in the list of the Principal European Economic Indicators (PEEIs) on the labour market. The other official source on labour demand is the Excelsior Information System produced by Unioncamere and ANPAL. Data are collected through the monthly surveys and with web techniques. They provide updated, systematic and reliable knowledge of the consistency and territorial, dimensional and economic activity distribution of the labour demand expressed by companies, as well as the main characteristics of the professional figures required (type of contract, level of education and qualifications, professions and elementary figures). Although it captures a third of the total number of companies present in the Business Register (the Statistical Archive of Active Companies - ASIA), it manages to cover the majority of the country's stable employment stock. Among the most relevant information collected by the Excelsior System: the number of contract activations expected for the following quarter detailed by economic activity, territorial area occupation, type of contract and education. Compared to the official Istat survey, the Excelsior system does not collect information on vacancies for which a concrete search is in progress (as defined in the EU regulation, see section 3). It focuses only on the possible future intentions expressed by companies to hire new workers, which is a different and less strong measure of the actual search of personnel by the enterprise. OJA web data source enables an enrichment of the labour demand information with respect to the official surveys under EU Regulation (as the Istat one) as the Table 1 shows. OJAs from web data collect high detailed information on the characteristics of the position to be filled by the employers (profession, education, skill, territorial area, type of contract, working hours and salary), in addition to the economic activity section – the only detail required by the EU Regulation –. Information is also at high frequency (daily basis) and not only

at a quarterly basis. Furthermore, while the count of job vacancies is measured at a point in time, making it a stock variable, the count of job advertisements is measured per unit of time, making it a flow variable.

Table 1 – *JVS from EU official survey vs OJAs from web data source - enrichment from OJAs.*

JVS from OFFICIAL SURVEY on the basis of EU Regulation	OJAs from WEB DATA
National, by NACE Rev.2 economic activity sections	National, by NACE Rev.2 economic activity sections
Stock of vacancies	Stock of vacancies
Not available	Flow of OJAs collected on each day of the month
Region on voluntary basis	Geographical area/ region / province / city
Occupation on voluntary basis	Occupation, education, skill, contract type, working hours, salary
Quarterly	Daily

3. OJA definition: a different concept from Job Vacancies

OJAs are one of the methods, increasingly widespread, to notify a job vacancy. OJAs refer to advertisements published on job portal revealing an employer's interest in recruiting workers with certain characteristics for performing a specific work. OJAs cannot be assimilated to Job Vacancies (JV) as defined by the EU Regulation. On the basis of the EC Regulation No 453/2008, a job vacancy is defined as: "a paid post that is newly created, unoccupied, or about to become vacant: a) for which the employer is taking active steps and is prepared to take further steps to find a suitable candidate from outside the enterprise concerned; and (b) which the employer intends to fill either immediately or within a specific period of time." The active search for a suitable candidate must have already started but not yet ended at the reference time at which job vacancies are collected. The active steps to find a suitable candidate include: (i) notifying the job vacancy to the public employment services; (ii) contacting a private employment agency/head hunters; (iii) advertising the vacancy in the media (for example the Internet, newspapers, magazines); (iv) advertising the vacancy on a public notice board; (v) approaching, interviewing or selecting possible candidates/potential recruits directly; (vi) approaching employees and/or personal contacts; (vii) using internships. The above definition of a job vacancy in the EU regulation framework makes it clear that the concept of OJAs does not correspond to that of JV, due to different factors. In particular, online job vacancies could cover all those job vacancies, as defined by the EU Regulation, for which the active steps carried out by the employers to find a suitable candidate include also the advertising on internet job portals. But, not all job vacancies are advertised on-line: although there is a general trend towards more job vacancies being advertised on-line, many of them continue to be filled through traditional channels, such as newspapers, employment agencies,

noticeboards, or personal contacts. Moreover, job portals used by employers may not be fully covered by web scraping and other data ingestion activities, and thus not all jobs advertised online may be fully captured. OJA data are generally drawn from multiple job portals: job boards, which publish original ads uploaded by employers; job search engines, which republish ads from other portals; hybrid job portals, which are a combination of both. In addition, new job portals may appear while existing portals may decline in importance. However, even if the whole set of advertisements published on all online portals were considered, they do not refer to total job vacancies that actually exists at a specific reference date. This is due to different reasons: for example, delay in the communication between enterprises and the online portals could cause the presence on portals of job vacancies that are closed (because, for example, a suitable candidate has already been found). In addition to this, either some OJAs may not represent a job vacancy in the scope of the official survey (these include non-existent vacancies, referred to as "ghost vacancies", international jobs, and non-payed student internships) or may be published without any direct correspondence with existing vacancies. Such "fictional" ads may be posted for example for attracting new talents as a recruiting strategy, or improving the image of a company. An additional and a background fact to take into account is that there is no one-to-one correspondence between advertisements and vacancies, as one advertisement may contain more than one vacancy. All the above factors imply that the stock of job vacancies measured by the OJAs could be higher or lower than the corresponding JV data measured by the current official job vacancy surveys. Furthermore, the number of vacancies collected by the official survey is the stock of vacancies, for which businesses are actively seeking suitable candidates at a specific reference date. OJAs represent a flow of new vacancies but usually do not contain information on when the recruitment activity started and ended.

4. OJA data: pipeline and variables classification

At European level, OJAs data are centrally managed: the European Center for the Development of Vocational Training (CEDEFOP) has implemented a system to collect and extract statistical information from OJAs for all European countries [1].

OJA data pipeline starts with web data scraping and includes all phases of processing, cleaning, standardizing, and classifying. A set of validation rules applied to the OJA dataset checks the consistency and plausibility of the statistical output. Consistency is checked in relation to Eurostat's standard classifications (ESCO and NACE Rev. 2). In addition, the stability of the ads distribution within categories of a classification is analysed over time and across data releases. These rules, applied to validate each record or a variable distribution, are complemented by other structural validation rules, to check the correct naming of datasets and variables, empty fields, etc. The statistical output produced by CEDEFOP is integrated and stored in a Data

Lab and disseminated on a quarterly basis with data available at a daily level. Starting from the third quarter of 2018, a new issue of OJA data is released each quarter. All Member States involved in the ESSnet can access the Data Lab, explore OJAs for statistical analysis and contribute to the quality assessment. This cooperative approach fosters the production of experimental indicators, as well as the improvement of the accuracy of the information produced. The text analysis of OJAs provides several details about the job position to fill, such as the salary, the workplace, the reference period and the main classification variables describing the occupation, skills, education level and economic activity. ESCO classification has been conceived to harmonise the core concepts and language about occupations and skills available for different stakeholders, dealing with employment education and training issues. ESCO is divided in two main pillars: occupations, skills/competences. The latest version allows describing 3008 occupations. The skill pillar provides a comprehensive list of knowledge, skills and competences relevant to the European labour market. It contains 13,890 concepts structured in a hierarchy, which contains four sub-classifications: Knowledge, Language skills and knowledge, Skills, Transversal skills. In addition, ESCO is linked to other official classifications and frameworks, such as the International Standard Classification of Occupations (ISCO), corresponding to the top four levels for the occupations pillar. In this case, ESCO describes the occupations classified at level 5 and lower.

5. Skill demand: needs, analyses and results

5.1. Skills from OJAs: an enrichment for the labour demand analysis

OJAs can provide valuable information on labour market dynamics, like the changing demand in skills and new occupation profiles, for instance for data scientist and computer scientist or for highly qualified profiles, covering important informative gaps. These could lead to the production of new timely indicators on the changes in the labour demand and so monitor the labour market dynamic for some sectors at least. Since not all new jobs are advertised (online, or at all), online data may also provide a distorted picture of the labour market. This type of data can provide an up-to-date picture of the skills that employers require, at higher detail than traditional surveys. The collection of reliable online job advertisement data is part of "Skills intelligence", one of the actions of the European Skills Agenda. A new European Regulation – that establishes a common framework for European Labour Market statistics on Businesses (regarding earnings, labour costs, gender pay gap and job vacancies) – is in the process of being approved, allowing the use of innovative sources to obtain detailed information on labour shortages by occupation and region. Generally speaking, collection of this type of information through surveys would entail additional costs for NSIs and increased statistical burden for enterprises. Experimental statistics may reduce this burden and should be based on a harmonized framework amongst

European countries, at least for the main variables: skill, occupation, territorial areas, type of contract and education. In addition, policy-makers and researchers by means of OJA based indicators may support the National Recovery and Resilience Plan (NRRP), to promptly monitor the offer of professional trainings and to analyse the causes of the mismatch between labour supply and demand. Furthermore, skill demand variations allow monitoring many phenomena such as: job creation and destruction, employment shifts, emerging occupations and skills needs, and new working patterns/relations.

Table 2 – *Description of the skills included in the sub-classification "Skill" of the ESCO classification skills/competences pillar.*

Skill	Description
Assisting and caring	Providing assistance, nurturing, care, service and support people, and ensuring compliance to rules, standards, guidelines or laws
Communication, collaboration and creativity	Communicating, collaborating, liaising, and negotiating with other people, developing solutions to problems, creating plans or specifications for the design of objects and systems, composing text or music, performing to entertain an audience, and imparting knowledge to others
Constructing	Building, repairing, installing and finishing interior and exterior structures
Handling and moving	Sorting, arranging, moving, transforming, fabricating and cleaning goods and materials by hand or using handheld tools and equipment. Tending plants, crops and animals
Information skills	Collecting, storing, monitoring, and using information; conducting studies, investigations and tests; maintaining records; managing, evaluating, processing, analysing and monitoring information and projecting outcomes
Management skills	Managing people, activities, resources, and organisation; developing objectives and strategies, organising work activities, allocating and controlling resources and leading, motivating, recruiting and supervising people and teams
Working with machinery and specialised equipment	Controlling, operating and monitoring vehicles, stationary and mobile machinery and precision instrumentation and equipment

In general, the labour demand changes can by analysed at ISCO (occupation) or ESCO (skill) level, disaggregated at NUTS (Nomenclature of Territorial Units) and NACE level. The skill considered in this work are those included in the skills/competences pillar of the ESCO classification and, in particular, in the sub-classification "Skill" (as described in Table 2).

5.2. Skill dynamic

In this section, we present descriptive analyses on Italian skill demand changes based on CEDEFOP OJA data, for the period from the first quarter 2019 to the fourth quarter 2022. The amount of skill, demanded by the employers by means of the online channels and collected in the CEDEFOP OJA system, was measured at the last day of

each quarter. The skill demand changes was assessed by utilizing raw data, i.e. by means of the year on year variations (that is, the percentage changes between the skill demand of each quarter compared to that of the same quarter of the previous year). CEDEFOP OJA time series start from the third quarter of 2018. Therefore, they are too short to be seasonally adjusted. The following figures show changes in the demand for skills at Italian national level and by macro-regions in the period 2019-2022.

Figure 1 – *Skill demand from OJA year on year percentage changes – Italy 2020Q1-2022Q4 – Source: CEDEFOP.*

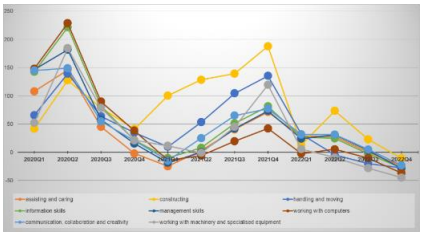
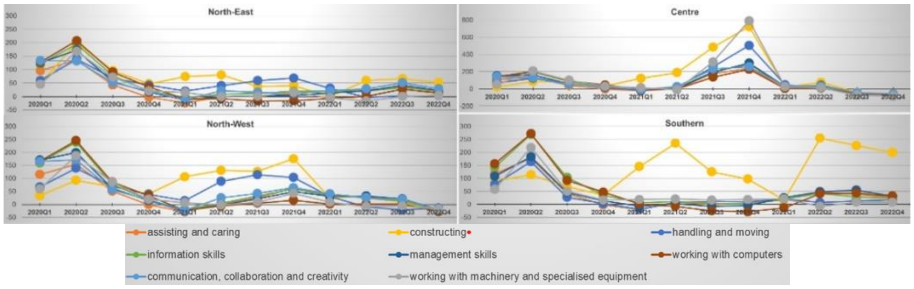


Figure 2 – *Skill demand from OJA year on year percentage changes – Italian macro regions 2020Q1-2022Q4 (skill colours same as Figure 1). Source: CEDEFOP.*

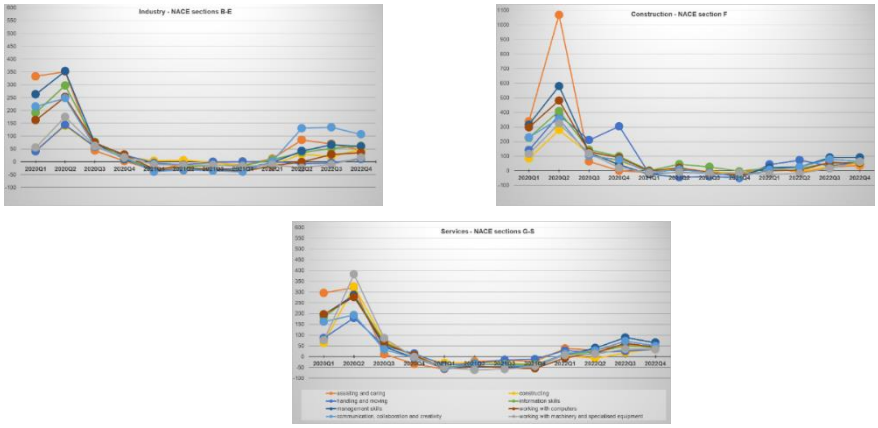


The national level analyses (Figure 1) show a first peak in demand for skills "working with computers" and "information skills" in the second quarter 2020 (see Table 2 for skill description). Moreover, the upward trend in the demand for "constructing skills" (building, repairing, installing and finishing interior and exterior structures), during the year 2021, could be attributed to the bonus effect in the construction NACE economic activity sector. Consistently with the period under analysis characterized by the pandemic crisis, a similar trend is also observed in the demand for "handling and moving skills".

The dynamic of the skill demand shows marked differences at macro-region break down (Figure 2). In the Centre of Italy, the demand for skills in the construction sector shows a higher increase than the national average, approximately four times higher. In Southern Italy, the demand for skills in the construction sector shows a higher increase

too – even if to a lesser extent than in the Centre – with a peak occurring two quarters earlier. A peak is also observed in the second quarter of 2022, with an upward trend. In the North-East, on the other hand, the signal for an increase in constructing skills is much weaker than the national average and concerns the first two quarters of 2021. While in the North-West on the other hand, the signal for an increase in constructing skills is halfway between that of the South and the North-East with a peak in the fourth quarter 2021 later than that one observed for the South. Furthermore, while at national level the increase in demand for constructing skills is also followed by an increase in other kind of skills, this is not the case in the Southern Italy, where an increase in constructing skills is the only one observed. At national level, the increase in demand for constructing skills is also followed by a rise in other kind of skills, this is not the case in the Southern Italy, where an increase in constructing skills is the only one observed. The evidence emerging from the analysis gives information consistent with the period under analysis characterized by the pandemic crisis.

Figure 3 – *Skill demand from OJA year on year percentage changes – Italian NACE economic activity sectors 2020Q1-2022Q4. Source: CEDEFOP.*



5.3. Skill and NACE economic activity sections

This section analyses the dynamic of skill demand by economic activity sector (according to NACE Rev.2 classification). In particular, the dynamic of skills is analysed separately for the industry sector (NACE sections from B to E), for the construction sector (section F) and for the service sector (sections from G to S). As Figure 3 shows, similar trends are observed in the three sectors in the central observation period from the first quarter 2021 to the first quarter 2022, with a slight prevalence of skills with negative trend in the services sector during the year 2021. This could be due to the long term effect of the Covid emergency, which affected economic activities in the service sector the most. In all the three sectors, the research

for skill "assisting and caring" shows a very high increase, reaching almost the maximum[1], which could be due to the emergency period effect too.

In the tails of the observation period, the differences between the three sectors are greater than those observed in the central period. A peak is observed in both industry and construction sectors in the second quarter 2020 in relation to the demand for "management skill", in line with the needs emerged during the Covid crisis of reorganizing work activities and resources and developing new strategies. While in the service sector, in the same quarter, the skill demand for "working with machinery and specialised equipment" shows the highest positive changes. Furthermore, during the last three quarters of 2022, the specialization in "communication, collaboration and creativity" is of particular interest for the industry sector. In this case too, the increasing requirement of this skill should be considered as a long term effect of the emergency period.

Figure 4 – *Skill percentage composition increase by occupation – 2019Q3 and 2022Q3. Source: CEDEFOP.*

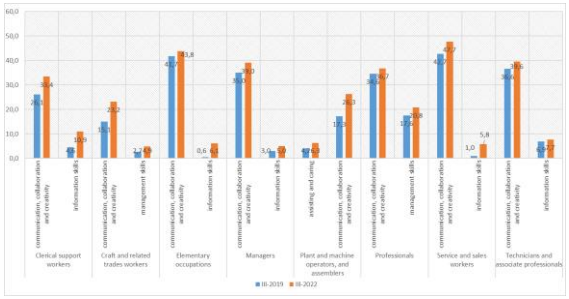
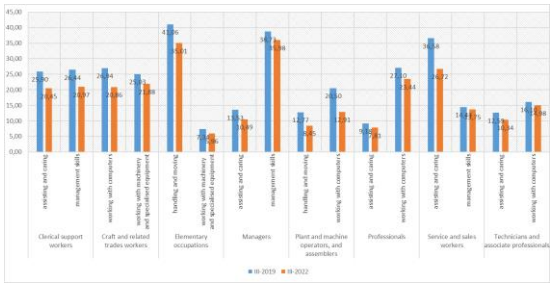


Figure 5 – *Skill percentage composition decrease by occupation – 2019Q3 and 2022Q3. Source: CEDEFOP.*



[1] In the construction sector, in the second quarter 2020, a strong peak is observed, which could be due to changes in the recruitment channels used by enterprises in this sector.

5.4. Skill and occupation

The analysis took also into account changes in the percentage composition by skill of each of the major occupational groups, which occurred between the third quarter 2019 (prior to the health emergency period) and the correspondent quarter in 2022. For the construction of the next figures, we select two skills in each occupation with the higher increase (Figure 4) and with the higher decrease (Figure 5) in percentage point.

As Figure 4 shows, the skill "communication, collaboration and creativity" was the skill with the largest increase across all occupational groups, except for the Elementary occupation and Professional occupation groups. For these two groups, the greatest increase was for the information skill and management skill. These last two skills represent the second largest increase in the percentage composition of all other occupational groups. The widespread increase in the weight of the skill "communication, collaboration and creativity" should be also due to the impact of the pandemic period. As described in Table 2, it implies an increase in the search for candidates with ability in communicating, collaborating, relating and negotiating with other people, developing solutions to problems, creating plans or specifications for the design of objects and systems, etc.

The demand for skills that showed a decrease between the two periods across occupations (Figure 5), were "assisting and caring" and "working with computers". This evidence is in line with the effects of the health emergency phase, which led to an increased demand for these skills and a subsequent return to normal levels in the years after the health emergency.

6. Conclusions

In this paper, we have highlighted the potentialities of OJA data by analysing a core variable in the analysis of the labour demand: changes in the skill demand by enterprises. OJAs is a new data source that does not replace traditional surveys but can be used to enlarge the availability of information in the labour market. OJA data represent a source with great potentialities for the production of new indicators, enriching the official job vacancies statistics. In particular, the demand for skill indicators broken down by geographical area, occupation and economic activity, analysed in this paper, could improve the matching between the two sides of labour market – labour demand, on the one hand, and labour supply on the other – by means of a better orientation of the professional training programme. Furthermore, new granular and detailed information on changes in labour demand could address the funds provided by the National Recovery and Resilience Plan (NRRP) for requalification and professional training in a more efficient way. The results presented showed a consistency of the OJA indicators with the economic framework during the health emergency phase. For example, the upward trend in the demand for "construction skills" during 2021 due to the bonus effect in the construction sector; the

peak both in industry and construction sectors in the second quarter of 2020 in the demand for "management skills", in line with the needs that emerged during the Covid crisis to reorganise activities and labour resources and to develop new strategies; finally, the widespread increase in the weight of the "communication, collaboration and creativity" skill among the main occupational groups in the whole period. This evidence highlights the usefulness of these indicators in describing the dynamics of labour demand, given their timeliness and granularity. Although the potential of the OJA data source is evident, many methodological challenges need to be addressed to ensure good quality statistics based on it. The WIN project is moving in this direction by working on various aspects that can help improving the quality of OJA data, such as: selection of reliable websites and improvement of their stability over time; data cleaning; definition and application of validation rules; and improvement of classified variable accuracy. The quality of OJA data is an essential prerequisite for the production of reliable statistics and for analyses as accurate as possible. Moreover, when using OJA data for statistical purposes some specific critical aspects must be taken into account, because OJA data are a non-probability sample and suffer from coverage issues.

References

CEDEFOP, 2019 "Online job vacancies and skills analysis: a Cedefop pan-European approach".

BERĘSEWICZ M. R. PATER ,2021, "Inferring job vacancies from online job advertisements", *Statistical Working Papers*, Eurostat. 2021.

BERĘSEWICZ M, BIAŁKOWSKA G, MARCINKOWSKI K, MAŚLAK M, OPIELA P, PATER R, ZADROGA K,2019, "Enhancing the Demand for Labour survey by including skills from online job advertisements using model-assisted calibration". https://arxiv.org/abs/1908.06731

BRANKA J., V. KVETAN, J. NAPIERALA, 2022. From the online job advertisements to official statistics – the aspects of quality assurance. *Q2022*, Vilnius.

BUELENS B., P. DAAS, J. BURGER, M. PUTS, J. VAN DEN BRAKEL,2014, "Selectivity of Big data" *Discussion Paper 11,* . Statistics Netherlands

TURRELL A., BRADLEY JS, DJUMALIEVA J, COPPLE D, THURGOOD J 2019. Transforming naturally occurring text data into economic statistics: the case of online job vacancy postings. *No. w25837*. National Bureau of Economic Research

---

Elena CATANESE, Istat, catanese@istat.it; Francesca INGLESE, Istat, fringles@istat.it; Annalisa LUCARELLI, Istat, anlucare@istat.it; Giuseppina RUOCCO, Istat, giruocco@istat.it