Development of a Labour Shortage Indicator by Occupation from OJA data

ISTAT

Annalisa Lucarelli
(*Department for social and demographic statistics*)

M. Amarone, F. Inglese, A. Masi, R. Magistro and G. Ruocco
(*Department for the Development of Methods and Technologies for the Production and Dissemination of Statistical Information*)

Contents

Job Vacancy official Statistics (JVS)

- In the context of the official statistics covering the demand side of the labour market, there are official **job vacancy surveys** that supply quarterly information on the **unmet labour demand**
- As job vacancies measure employment intentions that have materialized in candidates' searches, they can give "early warnings" on the dynamics of jobs in the near future
- It connects vacancies to short-run economics and makes job vacancy statistics **leading indicators of the economic cycle**
- Data on job vacancies are used by the European Commision and the European Central Bank to monitor short-term developments in the business cycle and the labour market
- The vacancy indicator currently used at the European level – the job vacancy rate – is one of the **Principal European Economic Indicators PEEIs** on the labour market
- The production by the NSIs takes place on the basis a **EU framework Regulation** (No 453/2008)

Online Job Advertisements (OJAs)

- OJAs refer to advertisements published on web-sites/job portals
- A valuable and **innovative source to complement the official statistics** on labour demand
- OJAs offer more possibilities for **analysing labour market trends** and **early capturing new emerging needs of employers** than traditional surveys:
  - ✓ **Granularity** of the information contained in the advertisements:
    - job characteristics (e.g. occupation, location, type of contract, working hours and pay),
    - employer characteristics (e.g. economic activity),
    - job requirements (e.g. education, skills and experience)
    - on the advertisement itself (publishing and expiring date from the website/job portal)
  - ✓ **High frequency**
  - ✓ **Timeliness**
- High frequency, timeliness, as well as the fact that the OJAs represents an immediate channel for search for personnel, make **OJA-based statistics even a more effective leading indicators of the business cycle than the JVS**
- The production of OJA-based statistics is at an **experimental stage** and would need an **harmonized framework between the different European countries**

JVS *vs* OJAs

JVS from OFFICIAL SURVEY on the basis a EU Regulation	OJAs from WEB DATA
National, by NACE Rev.2 economic activity section, quarterly	National, by NACE Rev.2 economic activity section, quarterly
Stock of vacancies on the last calendar day of the quarter	Stock of vacancies on a specific day of the reference quarter
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**
not available	**Daily (weekly, monthly, quarterly)**

ENRICHMENT

Main OJA quality issues

Representativeness

- OJAs over-represent certain occupations and skills and under-represent other
  - over-representation for workers with higher level of education and large size enterprises
  - the opposite for small enterprises (in Italy word-of-mouth is frequent for seasonal small enterprises)

Selection bias

- Not all job vacancies are advertised online. Employers use methods other than an advertisement on the websites

Under- coverage

- **Job portals/web-sites are not necessarily totally covered by data ingestion activities**

Over coverage

- job ads not removed from portals (delays in the communication)
- ghost vacancies
- duplicates: job offers published on several sources or several times

Classification errors

- Measurement errors due to the classification process of variables and ML algorithms

Eurostat OJA-based experimental statistics

Experimental statistics: using a new data source, OJA web-scraping, and natural language processing

Statistics on labour demand for ICT specialists

- Provide detailed and timely information on the impact of the digitalisation process, which is changing many occupations and work tasks
- Statistics:
  - shares of OJAs for ICT specialists, by NUTS-2 region and by occupation sub-groups of the ICT specialists
  - percentage growth of the number of OJAs for ICT specialists in a quarter compared to the same quarter of the previous year, by NUTS-2 region

Labour shortage indicator (OJAR)

- OJAR lists the most sought occupations on the web, indicating where recruiters might face potential challenges to recruit staff
  - helping policy makers, employers and jobseekers knowing which sectors need more labour force
- OJAR is the ratio of the number of OJAs divided by the number of employees (from the EU-LFS)
  - Occupation 3 digit (ISCO-08)
- it complements the JV rate with breakdowns by occupations
- It fills the information gaps of the official JV rate
  - covering this gap through surveys is not envisaged due to cost and burden

An attempt to (construct) validate the OJAR for Italy

We tried to construct a labour shortage indicator, **similar to the Eurostat one**, and to define **a possible validation strategy**

Sources

- OJA (CEDEFOP DataLab)
- LFS

Variables

- **number of OJAs**
- **number of employed persons**

Measures

- Total **flow of new OJAs** posted over each quarters of the year 2023
- **Flow** of employed persons in 2023 **who started their current job during each quarter of the year 2023**

Level of detail

- Occupation 3 digit (ISCO-08)

Indicator

- ratio of the flow of OJAs divided by the flow of employees from LFS

How to evaluate the quality of the indicator?

Source coverage for Italy

- Our starting point was **Eurostat's selection model for OJAs**
  - *Landscaping of Websites for Webscraping with Focus on Selection Models,* ESSnet-WIN, WP4 M.Six, A. Kowarik, J. Gussenbauer, October 2023
- **Two building blocks**: **quantitative assessment** and **qualitative assessment** of the sources' relevance
- The **first building block** involved indicators such as: type of the job-portal; the OJA volume displayed on the website; the displayed form for variables of interest
- The **second building block** is based on three dimensions:
  - **popularity** of the website
  - **stability** (of the access to the website and of the time series based on the scraped data)
  - **coverage** (of all classes belonging to a classification of interest)

Representativeness of OJA distribution by occupation

- Starting from the work
  - *From the online job advertisements to official statistics – the aspects of quality assurance,* CEDEFOP, J. Branka, V. Kvetan, J. Napierala, Q2022, Vilnius
- Comparison between OJA and LFS distributions by occupation based on homogeneity test:
  - Total flow of new OJAs posted over each quarters of the year 2023
  - Flow of employed persons in 2023 who started their current job during each quarter of the year 2023
  - Occupation 3 digit level (ISCO-08)

Source coverage for Italy (1/2)

A web-scraping Istat expert has identified the most important job portals currently available in Italy

- Indicators **similar** to those used by Eurostat

- **Proxy**
  - no information from previous scraping rounds available

Indicators	Description
Total visit [1]	Total site visit (Million) from August to October 2023 - Source Similarweb (https://pro.similarweb.com)
MozRank [2]	MozRank (https://www.checkergooglerank.com/it/mozrank-checker), developed by SEOmoz, is a 10-point scale measurement **similar to Google Pagerank**. It gauges the linking authority and popularity of a specific webpage on the internet. **Essentially, MozRank serves as a score indicating a webpage's importance in comparison to others online (0 min / 10 max)**
Alexa Rank [3]	Alexa Rank (https://www.sitexpired.com/) is a widely recognized global ranking system that assesses the popularity of millions of websites. **The ranking is determined by analyzing the estimated average daily unique visitors and the number of page views a site receives over the past three months**. A lower Alexa rank indicates a higher level of poularity. **Websites are assigned ranks between 1 and 100,000, with lower rankings signifying greater popularity**. Typically, any site with an Alexa Rank under 30,000 is considered to be in the top tier of popularity (lower max / higher min).
Open PageRank [4]	The Open PageRank (https://www.domcop.com/openpagerank/) initiative aims to reinstate Page Rank metrics, allowing for straightforward comparisons between various domains. **PageRank, which allocates a score ranging from 0 to 10, serves as a measure of a website's relative value to users.** A PageRank score of 0 generally indicates a low-quality website, while a score of 10 is reserved for the most authoritative sites on the web (0 min / 10 max)
Stability (proxy) [5]	The website "https://who.is/whois" provides a WHOIS lookup service. WHOIS is a protocol used to query databases and obtain information about the registration of domain names, IP addresses, and autonomous system numbers. When you enter a domain name or IP address into the search bar on "https://who.is/whois", it retrieves publicly available information about the domain name registrant, registrar, **registration and expiration dates**, name servers, and more. This information can be useful for identifying the owner of a domain, checking domain availability, investigating potential abuse, and other related purposes
Web scraping compatibility [6]	**Evaluating Cascading Style Sheets (CSS) for Efficient Web Scraping of Web Page Information**. OJA classification: NACE, NUTS, ISCO, ESCO, salary, contract type, work time. Web scraping score: CSS=2, Text=1, Missing=0. Min score 0 to Max score 14
OJA coverage [7]	**Refers to the extent of coverage regarding the key classification systems of interest, such as ISCO, NUTS, NACE, etc. within the scraped OJA**. OJA classification: NACE, NUTS, ISCO, ESCO, salary, contract type, work time. Coverage score: Present=1, Not present=0. Min score 0 to Max score 7

Source coverage for Italy (2/2)

List of most important job portals currently available in Italy

Ranking:

- **Total visits** [1]
- **Popularity** [2],[3],[4]
- **Stability** [5]
- **Coverage** [6],[7]

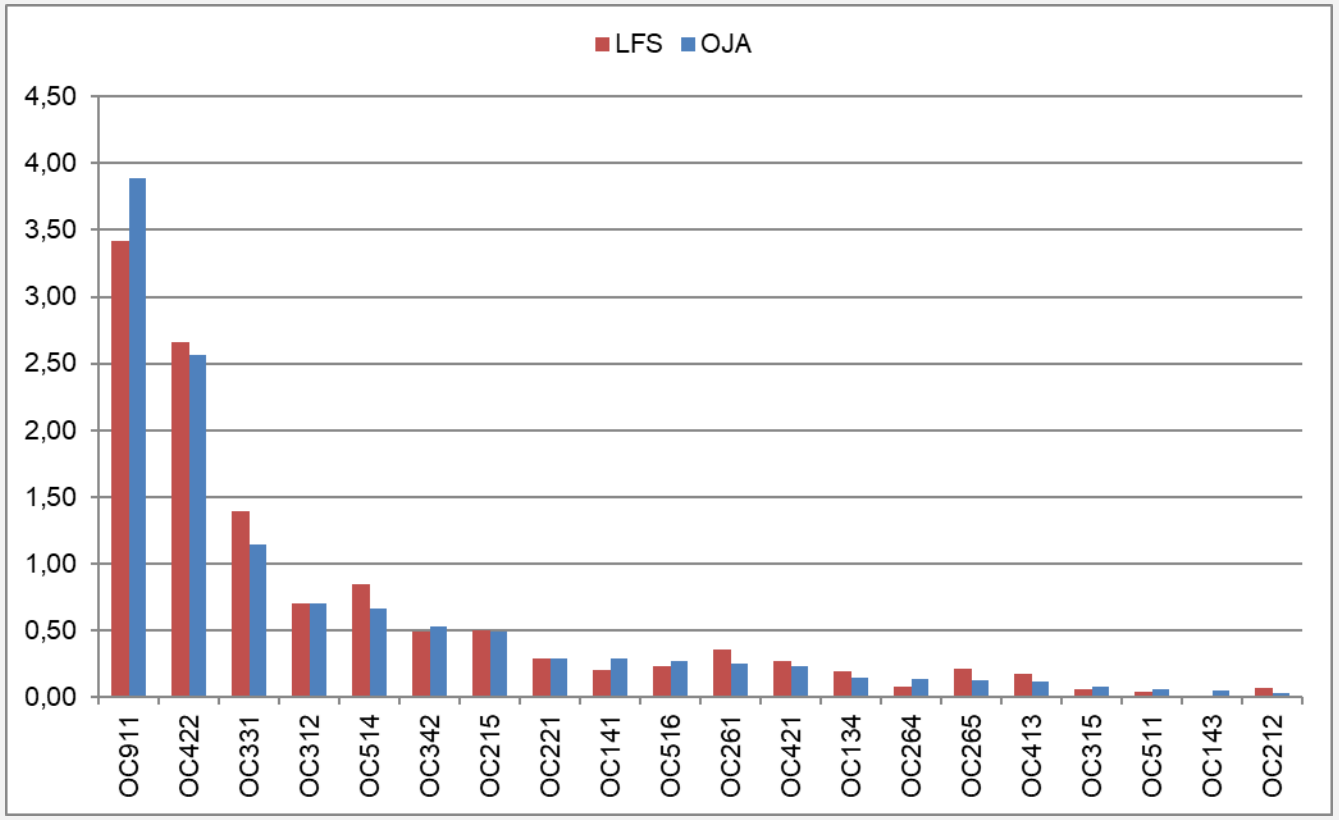
Source	CEDEFOP	Source type	Total visit [1]	MozRank [2]	Alexa Rank [3]	Open PageRank [4]	Stability (proxy) [5]	Multi country	Web scraping compatibility [6]	Coverage [7]	Registration needed	Link redirect
Indeed	**No**	Job Search	0,31	6,5	-	3,46	25	Yes	9	6	No	No
Banca Lavoro	**No**	Job Aggregator	0,6	4,1	89.008	3,54	14	No	7	4	No	Yes
Kit Lavoro	**No**	Job Search	0,67	3	226.606	1,71	11	No	7	6	No	No
Jobbydoo	**No**	Job Aggregator	1,3	4,8	73.326	3,77	12	Yes	8	4	No	Yes
Sercanto	**No**	Job Aggregator	1,8	3	301.013	0,98	12	Yes	8	4	No	Yes
Help Lavoro	**No**	Job Search	2	-	31.683	-	17	No	10	6	No	No
Talent	**No**	Job Search	6	4,5	-	-	31	Yes	8	6	No	No
Jooble	**No**	Job Search	6,4	-	-	3,49	14	Yes	8	6	No	No
Ti Consiglio	**No**	Job Search	12	-	56.126	4,2	23	No	7	5	No	No
JobCrawler	Yes	Job Aggregator	0,13	4,7	150,67	1,62	21	No	5	6	No	Yes
CareerJet	Yes	Job Aggregator	0,182	5	445.705	3,28	22	Yes	7	6	No	No
Careerjet	Yes	Job Search	0,3	4,4	79.444	4,07	22	No	7	6	No	No
Cerca Lavoro	Yes	Job Search	0,3	-	78.050	4,16	25	No	10	6	No	No
CliccaLavoro	Yes	Job Aggregator	0,4	5	289,34	2,95	24	No	10	5	No	No
Umana	Yes	Job Search	0,46	-	397.137	4,35	26	Yes	7	6	No	No
SimplyHired	Yes	Job Search	0,8	4,5	240.760	1,47	20	Yes	8	6	No	No
Euspert	Yes	Job Aggregator	0,9	4,1	230,12	1,95	11	No	7	6	No	No
Randstad	Yes	Job Search	4,3	4,6	74.094	4,09	24	Yes	8	6	No	No
InfoJobs	Yes	Job Search	4,7	5,4	56.264	4,31	25	Yes	10	6	No	No

Representativeness of OJA distribution by occupation (1/3)

Test results summary:

	Number of Minor Groups (3-digit codes)	Weight in the OJA distribution (%)
OJA=LFS	20	12,1
OJA<LFS	54	24,8
OJA>LFS	33	62,6

Distribution of occupation OJA vs LFS (%) – No significant differences (OJA=LFS)



OC911	Domestic, hotel and office cleaners and helpers
OC422	Client information workers
OC331	Financial and mathematical associate professionals
OC312	Mining, manufacturing and construction supervisors
OC514	Hairdressers, beauticians and related workers
OC342	Sports and fitness workers
OC215	Electrotechnology engineers
OC221	Medical doctors
OC141	Hotel and restaurant managers
OC516	Other personal services workers
OC261	Legal professionals
OC421	Tellers, money collectors and related clerks
OC134	Professional services managers
OC264	Authors, journalists and linguists
OC265	Creative and performing artists
OC413	Keyboard operators
OC315	Ship and aircraft controllers and technicians
OC511	Travel attendants, conductors and guides
OC143	Other services managers
OC212	Mathematicians, actuaries and statisticians

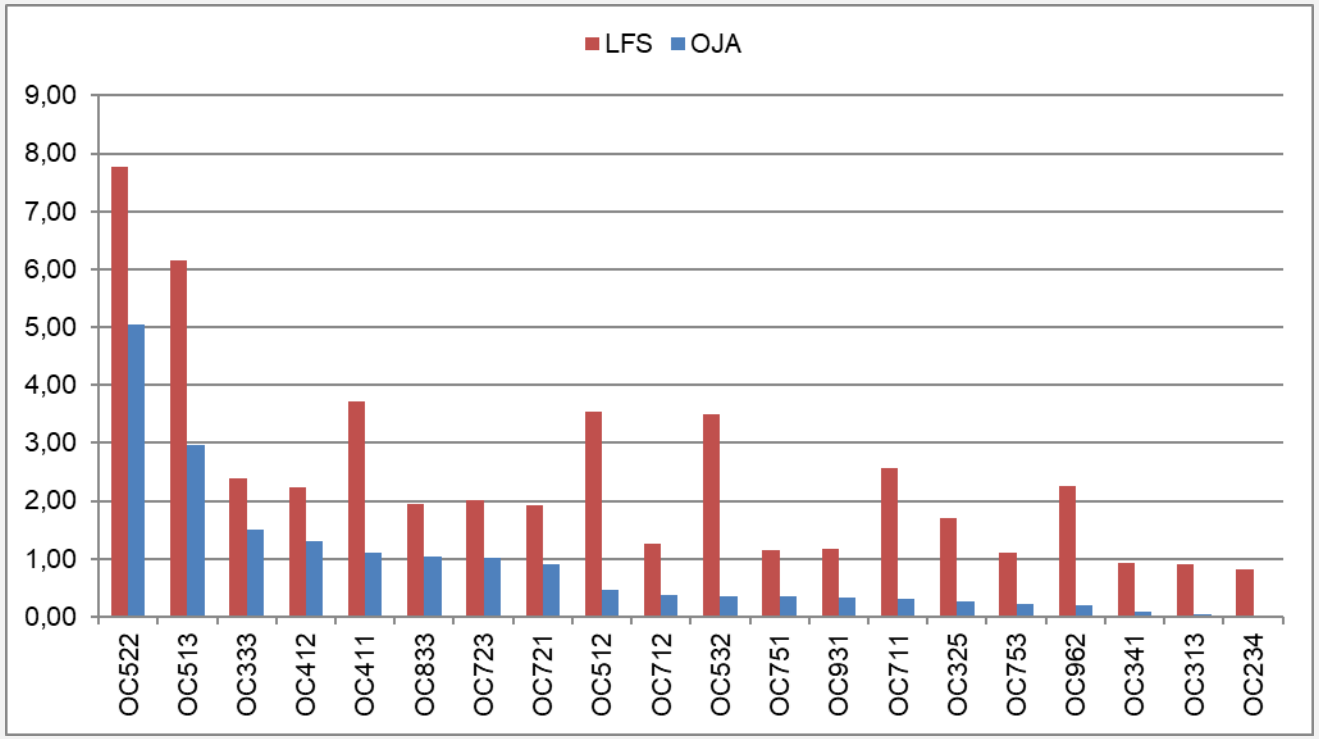
Representativeness of OJA distribution by occupation (2/3)

Test results summary:

	Number of Minor Groups (3-digit codes)	weight in the OJA distribution (%)
OJA=LFS	20	12,1
OJA<LFS	54	24,8
OJA>LFS	33	62,6

Distribution of occupation LFS vs OJA (%) – significant differences, under-representation (OJA<LFS)

only the first 20



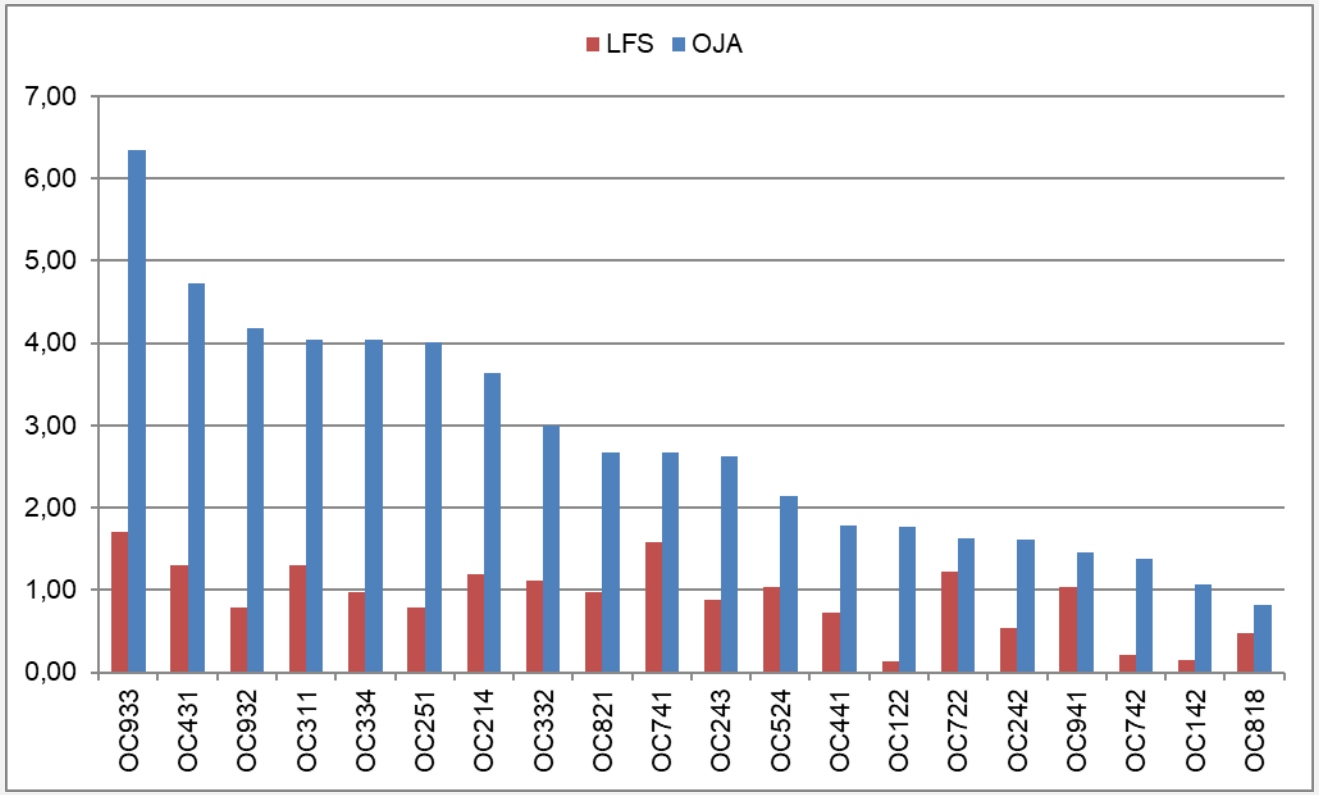
OC522	Shop salespersons
OC513	Waiters and bartenders
OC333	Business services agents
OC412	Secretaries (general)
OC411	General office clerks
OC833	Heavy truck and bus drivers
OC723	Machinery mechanics and repairers
OC721	Sheet and structural metal workers, moulders and welders, and relate
OC512	Cooks
OC712	Building finishers and related trades workers
OC532	Personal care workers in health services
OC751	Food processing and related trades workers
OC931	Mining and construction labourers
OC711	Building frame and related trades workers
OC325	Other health associate professionals
OC753	Garment and related trades workers
OC962	Other elementary workers
OC341	Legal, social and religious associate professionals
OC313	Process control technicians
OC234	Primary school and early childhood teachers

Representativeness of OJA distribution by occupation (3/3)

Test results summary:

	Number of Minor Groups (3-digit codes)	weight in the OJA distribution (%)
OJA=LFS	20	12,1
OJA<LFS	54	24,8
OJA>LFS	33	62,6

Distribution of occupation LFS vs OJA (%) – significant differences, over-representation (OJA>LFS)



OC933	Transport and storage labourers
OC431	Numerical clerks
OC932	Manufacturing labourers
OC311	Physical and engineering science technicians
OC334	Administrative and specialised secretaries
OC251	Software and applications developers and analysts
OC214	Engineering professionals (excluding electrotechnology)
OC332	Sales and purchasing agents and brokers
OC821	Assemblers
OC741	Electrical equipment installers and repairers
OC243	Sales, marketing and public relations professionals
OC524	Other sales workers
OC441	Other clerical support workers
OC122	Sales, marketing and development managers
OC722	Blacksmiths, toolmakers and related trades workers
OC242	Administration professionals
OC941	Food preparation assistants
OC742	Electronics and telecommunications installers and repairers
OC142	Retail and wholesale trade managers
OC818	Other stationary plant and machine operators
OC343	Artistic, cultural and culinary associate professionals
OC252	Database and network professionals
OC754	Other craft and related workers
OC226	Other health professionals
OC241	Finance professionals
OC263	Social and religious professionals
OC811	Mining and mineral processing plant operators
OC121	Business services and administration managers
OC132	Manufacturing, mining, construction, and distribution managers
OC112	Managing directors and chief executives
OC531	Child care workers and teachers' aides
OC314	Life science technicians and related associate professionals
OC515	Building and housekeeping supervisors

Comparison with a benchmark official source

- Monthly surveys of the **Excelsior Information System** carried out by **Unioncamere** in agreement with the **National Agency for Active Labour Policies (ANPAL)**
- Cover the demand side of labour market: **demand for labour from enterprises**

		OJAs *stock* at 31/12/2022		Excelsior *expected inflow of workers* in January 2023	
	TOTAL	***671.716***	***100,0***	***503.670***	***100***
OC1	Managers	45.524	6,8	1.840	0,4
OC2	Professionals	137.084	20,4	40.540	8
OC3	Technicians and associate professionals	119.812	17,8	86.020	17,1
OC4	Clerical support workers	89.021	13,3	45.880	9,1
OC5	*Service and sales workers*	*86.479*	*12,9*	*97.900*	*19,4*
OC7	*Craft and related trades workers*	*58.809*	*8,8*	*84.410*	*16,8*
OC8	Plant and machine operators, and assemblers	36.960	5,5	75.490	15
OC9	Elementary occupations	*98.027*	*14,6*	*71.610*	*14,2*

Conclusions

Opportunities

- Extremely useful **indicators on skill mismatch**
  - ✓ for the orientation of professional training programme
  - ✓ resolution of **job matching problems**
- Meet **new requirements from EU Regulations** for unsatisfied/emerging needs of information
  - ✓ new EU Regulation on LMB statistics allows the use of innovative sources
- (**Italy**) Exploit the experience gained in the use of OJAs **to fill the information gap** on job vacancies in the **public sector**
  - ✓ Web-craping on the jop portal which collect ads in the public sector (INPA)

Open issues

- **Extending** job portals coverage
- Having a **precision measures** of the variables of interest (occupation and skill)
- **Assess the quality of the results** using a multi-factor analysis (coverage, representativeness, etc.)
- Improve the comparison with official benchmark sources

Thank you

anlucare@istat.it