

Big Data for Labour Market Intelligence

Capacity development programme 2022

Module 2: Dissemination and analysis

Session 9

Supply side analysis: data sources, analytical framework. Issues. Perspectives.

Speaker: Fabio Mercorio

Roadmap

- The main issues to get the supply side of the web labour market
- Preliminary - Getting data from Digital Labour Platforms (IT)
- Paradigm shift: relevant skills define jobs

Goal: Explore the state of the art about the acquisition of data related to the supply side of the Web labour market.

Supply and Demand: A neverending story!

- The Demand-side is referred to the set of skills, competencies and requirements that a company is looking for;
- The Supply refers to information related to people that are available to work in terms of education, qualification, skills, experiences, etc.

Supply and Demand... about **Job Roles/titles**

Demand (OJA)

- Always indicated within job ads, usually through free text
- Can be recognised and classified according to a standard taxonomy

Supply (CV, resume, etc)

- Always indicated within CV or resume, usually referring either to the desired position or to the last position covered.

Both ISCO International Standard Classification of Occupations and ESCO multilingual classification of European Skills, Competences, and Occupations provides a taxonomy of job occupations and expected duties (i.e., roles)

Supply and Demand... about **Qualification**

Demand (OJA)

- Rarely requested in terms of minimal requirements

Supply (CV, resume, etc)

- It can be expressed either as raw text or using a standard, eg., Europass asks for EQF (European qualification framework).
- In addition, OECD's PIAAC survey measures adults' proficiency in key skills to perform a specific job role (ISCO II digit).

PIAAC is usually considered a proxy of the supply side

Supply and Demand... about **Education**

Demand (OJA)

- Often requested in terms of minimal requirements

Supply (CV, resume, etc)

- Always present. It can be expressed as raw text. Europass asks for ISCED (International Standard Classification of Education)

ISCED can be used as a reference taxonomy to match educational levels

Supply and Demand... about **Training Programmes**

Demand (OJA)

- Never requested by companies unless a specific training programme is needed to perform a specific job (e.g., certifications)

Supply (CV, resume, etc)

- It can be expressed either as raw text. Europass asks for ETER (European Tertiary Education System), a database of information on higher education institutions (HEIs) in Europe

Getting data related to demand (OJA) is a well-established and consolidated approach, rich in examples and fully fledged systems. Conversely, collecting the supply side still represents a challenging and open issue in the research community.

Why? (tentative)

Privacy related issues

1. A **job posting** is free of privacy issues by construction. OJAs do not include any sensible or personal data, and companies are likely to have privacy disclaimers in the footer to allow the job ads to be distributed without limitations.
2. **Curriculum Vitae (CV) or resume must contain personal data.** GDPR clarifies that personal data can be processed in case the written consensus of the data subject or – alternatively – whether certain conditions occur (e.g., necessary for legitimate interest, etc). If none of these cases apply, **processing any document containing the subject's data is against the GDPR regulation.**

The presence of personal data within any offer side information strictly limits their usage for activities outside the written subject's consensus.

It depends on the purposes

1. Notably, The EU clarified that *“processing covers a wide range of operations performed on personal data, including by manual or automated means. It includes the collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction of personal data. The GDPR applies to the processing of personal data wholly or partly by automated means and non-automated processing if it is part of a structured filing system”*

So, does a resume - deprived of personal data – still contain helpful information?

1. **No**, if the goal is to perform job matching and find the right candidate for the right job;
2. **Yes**, If the goal is to perform analytics, that is, to understand jobs, skills within jobs, job similarities, and skills mismatch to produce insights for decision-making

Personal data is all around us

1. According to a study by European Commission*, the data economy represented about 2.6% of the GDP in 2019 over the whole EU-27 (i.e., 325 billion euros). The EU estimates this value to grow to over 550 billion euros by 2025, representing 4% of GDP, generating 3.2 million jobs for data-related professionals.

Collecting labour market demand data **is more practical than the supply** – despite the goal of the analyses - and **this explains the growing number of applications**, prototypes and projects aimed at analysing labour market demand.

* (2022) Giorgio Micheletti; Nevena Raczko. Lisbon Council: Cristina Moise; David Osimo for the European Commission: DATA Market Study 2021–2023 CNECT/LUX/2020/OP/0027–VIGIE 2020-0655, contract number: LC-01568518

“data economy” usually refers to all economic activities in which data plays a central role as a business asset.

... and what about aggregators of CVs and resumes (including social media platforms)?

Getting resumes from social media

- There are a number of **intermediaries** such as Indeed, LinkedIn, and Monster that aggregate data on workers from resumes uploaded to the internet, **social profiles** created on job boards, and **online job ads** posted by employers.
- **Pros:** They capture **daily changes in the supply** (*skillset, and educational attainment of talent, and occupational demand, skill demand, wage offerings, and educational and experience requirements across industries and employers*) enabling for real-time macro- and micro-level analysis of labour market trends and shocks across and within economies, often down to the city level.
- **Cons:** Getting data via scraping might be difficult (blocking): one can decide to buy data instead; Assessing the quality and representativeness might be impracticable, due to the impossibility to compare data across official statistics due to the impossibility to perform a landscaping activity (as we did in ETF project on the demand side);

...is there any online source / “platform” that publicly discloses supply-side data that can be collected by third-party software for analytics freely?

Idea: Look at Digital Labour Platforms

Digital Labour Platforms

- **Digital Labour Platforms (DLPs)** are held private companies that act as intermediaries for on-demand labour services requested by individuals or corporate firms. Three parties are involved:
 1. the **client** demanding work,
 2. the **platform** which manages the algorithm
 3. the **person** who provides the work through the platform.

Two benefits for analysing the demand side

The match happens from the demand side → the employer selects and reviews professionals (GDPR compliance guaranteed by default)

Platforms make visible information on skills of professionals publicly available, including personal information collected through explicit consent to perform job matching.

1. Defined criteria for “evaluating” DLP (like a landscape...)
2. Tested to scrape supply-data for Italian DLPs

Qualitative Criteria for evaluating DPLs

- **Availability:** Check whether the professional resume is present or not. Allowed values: *Open, free registration, match, payment*
- **Composition:** Check how resumes are composed. Allowed values: *only description, description + skill, description + structured data*
- **Quality of Professional Qualitative:** Check if there is a qualitative review of the professional
Allowed Values: *yes, no*
- **Geo:** Check whether the location mentioned in the resume. Allowed values: *Mandatory, optional, possibility of remote working*

Qualitative Criteria for evaluating DPLs (preliminary)

Platform Name	Professional resume	Professional skills	Quality of Professional Qualitative	Geo
Cronoshare	open	no	yes	province
yoopies	register	no	no	30km
Go lance	open	yes	no	country
Starofservice	match by email	/	/	/
Prontopto	match (by email)	/	/	/
toptata	register	no	yes	50km
sitterlandia	open	yes	no	country

Comments

This feasibility test showed that **scraping activities would be feasible for the identified sources**

- each website allows getting data, requiring from 1 to 20 days to retrieve all the data

However, we have to note that

- we do not have any insights about the quality of the collected information
- a stable and rigorous collection process from those platforms would require performing a landscaping exercise to identify which sectors/professionals are covered, which platforms deserve to be scraped and to investigate how professional can be “classified” over ESCO occupation taxonomy to allow supply, demand skill mismatch through ESCO as well.

Getting Supply-side from social media or scraping from DLPs are both feasible solutions

If the goal is analytics - rather than job matching - asking people to declare top-skills used to do a job might be more effective than analysing CVs (see PIAAC....)

Skill2job: Support citizens and operators for upskilling and reskilling

Key elements:

1. *“skills define a job, and not the vice-versa”*
2. *“Asking people to indicate the three skills that they consider as fundamental to do their job, rather than processing an historical CV where the knowledge of a skill might be obsolete or not related with the job at all”*



Tell me your skills: #wordpress #design_graphics #adobe photoshop

Which is your current role? Marketing Specialist

Skill2job: Support citizens and operators for upskilling and reskilling

Tell me your skills: #wordpress #design_graphics #adobe photoshop

Which is your current role? Marketing Specialist



Importance of skills you hold within your job

#wordpress ★★★★★

#design_graphics ★★★

#adobe photoshop ★

Reskilling: you should also acquire:

#webanalytics ★★★

#searchengine ★★★

#graphical_design_patterns ★

Upskilling: You might also be interested

#HTML5 ★★

#responsive design ★

Career path: similar jobs

Digital Media Specialist

Web and Multimedia Developer

.....

Concluding remarks

We discussed about the **issues** (both technical and juridical) to **get the supply-side information from the Web (social and specific platforms)**

A preliminary result on Italy showed structured **information on professionals can be retrieved by Digital Labour Platforms** (GDPR compliance)

Meanwhile, we discussed about the **opportunity to change the mind**: moving from collecting and processing CVs to collect top-skills through an AI-based system that allow people compare with the labour market, reducing the noise and effort needed to process CVs

Notably: this is the same approach followed by OECD in realising PIAAC, the survey on adult competencies in which a stratified sample of workers is asked to estimate if - and to what extent - a (general) competence is used to do the declared job (from 1, rarely, to 5, daily). An EU-27 update of PIAAC is expected to be released in 2023/2024, representing a good benchmark to analyse the mismatch between the supply and demand sides.

Thank you!!!