
Establishing objective information about the labour market is important for policy and decision makers to ensure Europe's competitiveness and that the skills of the workforce meet demand.

With a significant share of job offers advertised online, the data contained within them hold valuable pieces of information about the current job market. Tapping into this data with big data technologies can help to understand, for example, trends in existing and emerging skills required and the number of job vacancies. Establishing objective information about the labour market is important for policy and decision makers to ensure Europe's competitiveness and that the skills of the workforce meet demand. When working on new data-driven solutions, the European Commission's big data testing tool helps to experiment with real data before starting system development.

Quick facts

- **Parties involved:** The European Centre for the Development of Vocational Training (https://www.cedefop.europa.eu/) (CEDEFOP) and the statistical office of the European Union (https://ec.europa.eu/eurostat) (Eurostat), representing the European Statistical System (ESSnet) Big Data project (https://ec.europa.eu/eurostat/cros/content/essnet-big-data_en) and participating National Statistical Institutes (NSIs)
- **Challenge:** Use Online Job Advertisement (OJA) data to provide timely information about European labour markets
- **Solution:** Establish requirements for a pan-European system to produce official statistics and dedicated policy analysis based on OJA data
Perfect partners

About a half a decade ago, CEDEFOP (https://www.cedefop.europa.eu/) and Eurostat (https://ec.europa.eu/eurostat) realised that they were working hard on the same data to produce similar insights into the European labour market. On the one hand, CEDEFOP is an agency that needs labour market data to help policymakers and other labour market actors on EU and national levels to understand trends in the supply and demand of vocational labour and skills, as well as any imbalances between the two. On the other hand, Eurostat, the Statistical Office of the European Union, helps policy and decision makers by gathering and providing the official statistics on the EU society, economy and environment, including the European labour market. Both wanted to use Online Job Advertisements (OJAs) to explore and extract new kinds of information using big data technologies.

In March 2019, CEDEFOP launched the Skills Online Vacancy Analysis Tool for Europe (Skills-OVATE), which already in its first limited-scope implementation proved the system’s potential to process and analyse OJA data for skills demand. Eurostat has been promoting the use of OJA data by the NSIs to enhance the labour market official statistics, supporting the ESSnet Big Data (https://ec.europa.eu/eurostat/cros/content/essnet-big-data_en) since 2015. CEDEFOP and Eurostat have taken steps towards expanding CEDEFOP's Skills-OVATE to adapt it to the production of European official statistics; and to extend it to the exploration of many other sources of web data with the creation of the European Web Intelligence Hub.

Benefits of BDTI

Before taking on the European Commission's big data tool, BDTI, NSIs tried to collect and analyse OJA data by themselves, but they ran into technical issues with regards to insufficient computing power and memory. The big data nature of OJAs required a more robust infrastructure, which BDTI provided free of charge. BDTI provided a readily available testing environment with customisation possibilities and support services. Its virtual environment templates work with various data sources, software tools and big data techniques. BDTI allowed the agencies to focus on gathering knowledge, insight and value from their data, instead of putting effort into setting up a complex experimental test space. This enables the creation of quick prototypes to verify and test data hypotheses, methodologies and visualisations.

In 2019, the BDTI team set up a data testing environment for a four-month period and organised a training session for selected NSIs. The objective was to teach and empower NSIs to explore real OJA data from their national points of view. This data had been collected by CEDEFOP over the course of more than one year. BDTI helped to save time, costs and effort in innovating and validating which statistical data products are feasible and worth developing in Skills-OVATE. As part of this implementation, BDTI handled 15 terabytes of OJA data and millions of queries.

BDTI helped us tremendously by allowing us very quickly to have an appropriate environment for the training event and the following exploration of the OJA data.

Fernando Reis, Big Data Statistician, Eurostat
What’s new in OJAs?

OJA data enable the creation of unique and innovative indicators to complement traditional official statistics. They will not be used to replace existing statistics, but to bring new dimensions to labour market intelligence. The traditional means for achieving statistics involve more time-consuming processes through surveys, which explore the labour force (supply side) and job vacancies (demand side). OJA data brings to the table near real-time statistics and more details about skills and skill demand.

Here are some more benefits to statistics based on OJA big data:

- Can be produced more often and faster than traditional statistics (released quarterly)
- Provides information about jobs omitted in traditional channels e.g. international jobs
- Effective way to study even sub-national labour markets
- Possibility to extract more detailed information about title, occupation, skills and location

How it works

Extracting data from OJAs requires converting the digital footprint left by companies in the online labour market into 10 relevant statistical variables. Since none of the data collected and analysed relate to personal information, the system falls outside the scope of EU data protection legislation (GDPR). The two types of data collected are:

- **Structured data** contained in data fields, such as job location and publication date, quickly collected by data scrapers.
- **Natural language** found in free text. Here, Artificial Intelligence (AI), combining Natural Language Processing (NLP) and machine learning algorithms, are used to clean the free text and extract relevant data on the occupation, education and skills required, salary and so on.

**BDTI** helped to plan for more targeted use of resources in extending Skills-OVATE to statistics, and to refine the technical requirements for producing them. For example, NSIs knew that OJAs do not always translate to job vacancies. A job vacancy can be advertised on several platforms, and there may be multiple job vacancies behind a single OJA. Thanks to BDTI, the workgroup has a better understanding for establishing a method to calculate indicators without bias, before taking the time and effort to code functionalities in Skills-OVATE.

Next steps

Skills-OVATE now collects OJA data from all EU Member States and more than 94,000 webpages, including Public Employment Services (PES), private job boards, recruitment agencies, company websites and online newspapers. Skills-OVATE is expected to be fully operational later on in 2020 and provide a single, centralised pan-European system for all stakeholders on all levels: EU, national and regional.

The next steps for the Web Intelligence Hub include setting up data quality monitoring procedures, defining a governance model and expanding the system for the large-scale production of official statistics in labour market intelligence and beyond. The Hub will be based on the Skills-OVATE system and leverage the collaboration between the NSIs kick-started by the BDTI environment and training.
Related topics


Source URL: