

ESSNET Big Data project II (2018-2020)

https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata/index.php/WPB_Overview

WPB Overview

Contents

- [1 Objectives](#)
- [2 Summary description of the workpackage](#)
- [3 Tasks](#)
 - [3.1 Task 1 – Methodological framework](#)
 - [3.2 Task 2 - Statistical output](#)
 - [3.3 Task 3 - Implementation requirements of prototypes in the relevant statistical production processes at European and national level](#)
- [4 Milestones and deliverables](#)
 - [4.1 WPB milestones](#)
 - [4.2 WPB deliverables](#)

Objectives

The aim of implementing this pilot is to produce statistical estimates in the statistical theme of online job vacancies. Suitable techniques and concrete methodologies were developed during the pilot phase of the project (see [WP1 Online job vacancies](#)). Implementation will be based on work that was carried out regarding the conditions that web scraping techniques can be used as far as the quality of the scraped data is concerned, as well as the use of mixed sources including job portals and job adverts on enterprise websites, and job vacancy data from third party sources. Relationships with the project of CEDEFOP (European Centre for the Development of Vocational Training) will be consequently explored. Within the same statistical theme, the combination of existing data from multiple sources will be promoted and embedded in the methodology. The final aim of the activities is to develop and test the methodology and prototypes as well as capacity building to facilitate their integration into production at the level of individual NSI and at the level of the ESS.

Summary description of the workpackage

Taking into account the developments of the successfully concluded work during 2016-2018, the aim is to develop a functional production prototypes including setting up procedures and developing technical solutions to promote and support the collection, processing and analysis of online job vacancies (OJV) data for statistical production. Along the production of the relevant methodologies, recommendations, specifications and statistical software, the production of experimental statistics demonstrating the capabilities to produce statistics is one of the objectives. Subject matter experts of the relevant statistical production units of the NSIs will be involved in all phases of the work. Participation to relevant Eurostat Working groups and Task Forces will be part of the work of this WP. This includes presentation of findings and work to the relevant working groups at an early stage. Web scraping, text mining and inference techniques will be tested and used in order to collect, process, calculate indicators on OJV and eventually supplement the existing JV statistics. An important factor is the role of third party data sets providers (e.g. CEDEFOP) and others as partners and likely future sources of web-scraped OJV data. The cooperation with CEDEFOP will ensure that

the data meets the needs of official statistics as far as possible and NSIs may be able to reduce their activities around data access and data handling and focus more on the challenges around further methodological development. Already at the time of the execution of the pilot phase some concerns were raised. A major concern is that the methods for producing sufficiently robust statistical outputs based on OJV data are not sufficiently mature. OJV data cannot be used to directly replace the existing job vacancy statistics that are required by EU regulation. Indeed, the quality issues are such that it is not clear if these data could be integrated in a way that would enable them to meet the standards expected of official statistics. On the other hand, OJV data can provide more granular insights that official estimates usually offer. It was concluded that the role of OJV data within official statistics is more likely to be as the basis for producing supplementary indicators. For this reason, the implementation project will be framed in terms of something that is achievable. This WP has links with WPC regarding methods for web scraping, data processing and analysis and WPF on designing and adopting application and information architectures. A combined meeting with WPC and back-to-back meeting with WPF will be held.

Tasks

Task 1 – Methodological framework

Performed by SURS (Statistics Slovenia), NSI (Statistics Bulgaria), FSO (Statistics Switzerland), DARES (Ministry of Labour, FR), Destatis (Statistics Germany), LSD (Statistics Lithuania), GUS (Statistics Poland), INS (Statistics Romania) and ONS (Statistics UK)

This task will produce generalised and extended methods, procedures and implementation requirements for using online job vacancy (OJV) data in statistical production. The work will be based on results of use cases developed in the pilot phase ([ESSnet Big Data I WP1 Online job vacancies](#)). The general sub-tasks will be: Identification of statistical production processes and capabilities that may be affected at national level and definition of the conceptual production processes at national level and at the level of the ESS. Developing and evaluating scenarios for data governance and data management in the wider aspect of data sharing and collaboration processes management. Protection of privacy and access, compliance with statistical regulatory requirements should be an integral part of the scenarios (no sharing, partial/conditional sharing, and full sharing). Cooperation with WPC in the development of the ESS webscraping policy. Development and testing of fully functional prototypes including the adaptation and consolidation of the methodology, the procedures and the tools for collecting, processing and analysis for producing new statistics or enhancing existing ones. Output will include software for obtaining and processing third party data sets, transforming web data from job portals into the structure for analysis (data cleaning, correction of formatting errors, evaluation and treatment of missing data, de-duplication of job adverts within and across portals, classification of data) and computing statistical outputs. The cooperation with CEDEFOP will ensure that NSIs may be able to reduce their activities around data access and data handling and focus more on the challenges around further methodological development. Developing and evaluating scenarios for linking identified portals, company, profession etc., to the business register and established statistical classifications, for example classification of occupations. Addressing issues related to sustainability of data sources, data use by NSIs, as well as data sharing between NSIs.

Task 2 - Statistical output

Performed by SURS (Statistics Slovenia), NSI (Statistics Bulgaria), FSO (Statistics Switzerland), DARES (Ministry of Labour, FR), Destatis (Statistics Germany), ISTAT (Statistics Italy), GUS (Statistics Poland), INS (Statistics Romania), SCB (Statistics Sweden) and ONS (Statistics UK)

Most of the results of the pilot phase were intermediate analyses and cannot be classified as statistical outputs, or even experimental statistics. It was shown that the role of OJV data within official statistics is more likely to be as the basis for producing supplementary indicators. This task will identify and produce these indicator(s) using OJV data. The general sub-tasks will be: Investigation and selection of potential use cases and define indicators(s) in the field of job vacancies statistics. Investigation of the methodology for the calculation of the indicator(s) in the field of job vacancies statistics. Calculation of the indicator(s) in the field of job vacancies statistics. Quality assesment of statistical outputs (e.g. accuracy, sensitivity, specificity) based on existing quality frameworks (for example UNECE Framework for the Quality of Big Data). Publication of indicators on OJVs which can supplement existing JV statistics as experimental statistics, to be published at the Eurostat website and at the national level. Investigation of other possible statistics beyond job vacancies statistics which could be produced with OJV data, such as new statistics on the international labour market that are currently not produced, identification of new job titles (help in updating classifications) and skills statistics.

Task 3 - Implementation requirements of prototypes in the relevant statistical production processes at European and national level

Performed by SURS (Statistics Slovenia), NSI (Statistics Bulgaria), DARES (Ministry of Labour, FR), GUS (Statistics Poland), INS (Statistics Romania) and ONS (Statistics UK)

This task will produce documentation of the integration process and the methodologies at national level, the incremental implementation of software and production of statistical indicators based on OJV data. The general sub-tasks will be: Definition of the implementation requirements of prototypes in the relevant statistical production processes at European and national level. Outline of the architecture, processes and infrastructure for future production of statistical outputs in other statistical domains, including the aspects of scalability and evolution of IT resources. The issue of common infrastructure will be examined in detail in order to achieve important economies of scale. Interaction with WPF on designing and adopting application and information architectures (WPF Task 1) and providing input on WPB solution architectures. Definition of the necessary material for the integration process: transfer of knowledge, methodologies, IT infrastructure needed, toolbox of methods, software development, testing, and maintenance.

BM1 (M9) Report on the WP meeting mid 2019

Milestones and deliverables

See [here](#) for an overview of available milestones and deliverables.

WPB milestones

BM1	Report on the WP meeting mid-2019	Month 9
-----	-----------------------------------	---------

BM2	Report on the WP meeting mid-2020	Month 20
-----	-----------------------------------	----------

WPB deliverables

B1	Interim technical report
B2	Methodological Framework V.1
B3	Methodological Framework V.2
B4	Report on the statistical output, required quality and definition of the necessary metadata at European
B5	Technical report on the implementation requirements of prototypes in the relevant statistical production

Categories:

- [Overview](#)
- [WPB Online job vacancies](#)