

ESSnet Big Data is a project within the [European statistical system \(ESS\)](#) jointly undertaken by 22 [partners](#). Its objective is the integration of big data in the regular production of official statistics, through pilots exploring the potential of selected big data sources and building concrete applications.

ESSnet Big Data has started in February 2016 and is to run for 28 months until May 2018; it consists of 10 workpackages: eight of these are content-oriented, while the other two, *Coordination* and *Dissemination*, support the overall project.

General objectives

Big data: from exploration to exploitation.

The ESSnet BIG DATA is part of the Big Data Action Plan and Roadmap 1.06 (BDAR below) and it was agreed to integrate it into the ESS Vision 2020 portfolio. The related business case Big Data received the support of the ESSC at its meeting on 20 June 2015 in Luxemburg. The overall objective of the project is to prepare the ESS for integration of big data sources into the production of official statistics. The award criteria mentioned that the project has to focus on running pilot projects exploring the potential of selected big data sources for producing or contributing to the production of official statistics. Aim of these pilots is to undertake concrete action in the domain of big data and obtain hands-on experience in the use of big data for official statistics.

Taking into account these objectives the slogan of this ESSnet is: "BIG DATA: from exploration to exploitation for official statistics" The slogan has been chosen in order to make clear the difference of this European-funded international 'big data' project compared to more scientific and policy-making projects in the domain of big data. The second reason to choose this slogan is to clarify the choices of activities which are included in this ESSnet or not.

Specific objectives

A consortium of 22 partners, consisting of 20 national statistical institutes and 2 other statistical authorities has been formed in September 2016 to meet the objectives of the project. According to the Framework Partnership Agreement (FPA) between the consortium and [Eurostat](#), the project runs between February 2016 and May 2018. To concentrate the work as much as possible on the pilots, the consortium has organised its work around the pilots. More specifically, the consortium has subdivided its work into workpackages (WP), each WP dealing with one pilot and a concrete output.

This is the list of workpackages:

- WP0: Coordination
- WP1: Webscraping job vacancies
- WP2: Webscraping enterprise characteristics
- WP3: Smart meters
- WP4: AIS Data
- WP5: Mobile phone data
- WP6: Early estimates
- WP7: Multiple domains
- WP8: Methodology
- WP9: Dissemination

Three workpackages are not dealing with pilots. The aim of WP8 (Methodology) is to generalise the findings of the pilots in order to relate them to the conditions for future use of big data sources within the ESS. WP0 and WP9 deal with the coordination of the action and the dissemination of the results, respectively.

The ultimate aims of each workpackage at the end of the project can be briefly described as:

WP1 Webscraping job vacancies

The aim of this pilot is to demonstrate by concrete estimates which approaches (techniques, methodology etc.) are most suitable to produce statistical estimates in the domain of job vacancies and under which conditions these approaches can be used in the ESS. The intention is to explore a mix of sources including job portals, job adverts on enterprise websites, and job vacancy data from third party sources.

WP2 Webscraping enterprise characteristics

The aim of this pilot is to investigate whether webscraping, text mining and inference techniques can be used to collect, process and improve general information about enterprises. Challenges compared to the Webscraping / Job Vacancies pilot are application of more massive scraping of websites and collecting and analysing more unstructured data. In particular, the pilot intends to demonstrate whether business registers can be improved by using webscraping techniques (kind of activity / key financial variables etc. / structure of enterprises etc.). Further possibilities for statistical outputs will be analysed toward the end of the action.

WP3 Smart meters

The aim of this pilot is to demonstrate by concrete estimates whether buildings equipped with smart meters (= electricity meters which can be read from a distance and measure electricity consumption at a high frequency) can be used to produce energy statistics but can also be relevant as a supplement for other statistics e.g. census housing statistics, household costs, impact on environment, statistics about energy production. Challenges ahead with this dataset are: representativity issues, linking to other datasets, privacy concerns. Another challenge with smart meters data is that these are currently available in a few countries only, but will be available in several countries before 2020. Second aim of this workpackage is to relate the results of this pilot to future use in other countries.

WP4 AIS data

Aim of this work package is to investigate whether real-time measurement data of ship positions (measured by the so-called AIS-system) can be used 1) to improve the quality and internal comparability of existing statistics and 2) for new statistical products relevant for the ESS. Improvement of quality and internal comparability can be obtained e.g. by developing a reference frame of ships and their travels in European waters and then linking this reference frame, by ship number, to register-based data about marine transport from port authorities. These linked data can then be used for emission calculations. New products can be developed for e.g. traffic analyses. The added value of running a pilot with AIS-data at European level is that the source data are generic world-wide and data can be obtained at European level. Challenges ahead with this dataset are: obtaining the data at European level, processing and collecting the data in such way that they can be used for multiple purposes, and visualising the results. A part of this work package is also to look into AIS analyses done by others and to investigate the possibility of obtaining already processed data as input for creating comparable official statistics. Especially it is important to make contact

with other public authorities.

WP5 Mobile phone data

The aim of this workpackage is to investigate how NSIs may obtain more or less 'stable' and continuous access to the data. In the current situation, most NSIs face complications to get access to these data due to legal, privacy and contractual issues. On the other hand, the potential of mobile phone data as a data source for official statistics is beyond any debate. This workpackage will describe concrete statistical outputs - relevant for official statistics - based on mobile phone data (and previous studies) and discuss the pro's and con's of using aggregated or microdata from mobile phone providers for these actions. The 1st year of the action will be used to get access to the data based on this plan. It is envisaged that in a possible SGA-2 an appropriate pilot based on the plan (if the data are available) will be run. Challenges ahead are: data access, representativity issues, linking to other datasets, linking the outputs to other statistical estimates.

WP6 Early estimates

The aim of this pilot is to investigate how a combination of (early available) multiple Big Data sources and existing official statistical data can be used in order to create existing or new early estimates for statistics. For those combinations which will be determined to have the greatest potential, the WP team will describe the data collection, data linking, data processing and methodological issues. As a maximum two pilots will be carried out on quick wins. Challenges ahead are: representativity issues, linking to other datasets, metadata. The result of the pilot will be guidelines and recommendations regarding using Big Data sources for early estimates.

WP7 Multiple domains

The aim of this pilot is to investigate how a combination of Big Data sources and existing official statistical data can be used to improve current statistics and create new statistics in statistical domains. The work package focusses on the statistical domains : Population, Tourism/border crossings, Agriculture. The WP team will describe the data collection, data linking, data processing and methodological aspects when combining data in statistical domains. Challenges ahead are: representativity issues, linking to other datasets, metadata, international comparability.

WP8 Methodology

The aim of this workpackage is laying down a generally applicable foundation in areas such as methodology, quality and IT infrastructure for future use of the selected big data sources from the pilots within the European Statistical System. WP8 will therefore start with the creation of a literature overview of all papers, presentations or webpages relevant to the application of Big Data for official statistics and link this literature overview with the findings of the pilots in the second phase of the project, SGA-2.