

Description of the workpackage

The aim of this pilot is to investigate multiple big data, administrative and other existing sources in order to produce early estimates for statistical purposes. For those sources, which will be determined to have the greatest potential, the WP partners will determine the business cases related to statistics and propose them for the second SGA. The focus will be on two concrete domains which have potential of getting “quick wins”. The expected output of the pilot will be guidelines and recommendations regarding using big data sources in the area of early estimates. The most promising sources are:

- social media data, newsfeeds and survey data for the aim of consumer confidence index;
- web-based sales inquiries for the aim of nowcasts of turnover indices

Since the workpackage considers many crosscutting issues, such as methodology, quality and technical requirements, care will be taken that its outputs can be used as inputs for WP8 Methodology, foreseen for SGA-2, the second specific grant agreement. The participation of NL, which is to lead WP8 of SGA-2, is to ensure this (see also [WP7](#)).

Tasks

Task 1 - Data access

- Prepare an inventory of relevant social media sources in view of the CCI in participating countries.
- Prepare an inventory of relevant sources of Web-based sales inquiries of the nowcasts of turnover indices in participating countries.
- Investigation of possible sources relevant to other early estimates for statistics.
- Qualitative assessment of the information available.
- Evaluate the role of other existing sources (e.g. VAT data, surveys) to support the collection of data and creating the statistics.
- Investigation of needed IT infrastructure for storing and processing of data.
- Prepare an interim feasibility report.

Task 2 - Data handling

- Study the technical aspects of collecting the data for CCI and nowcasts of turnover indices.
- Identify technical requirements for collecting the data for CCI and nowcasts of turnover indices, prepare a suitable IT environment and design and build a database for storing data.
- Deployment of the system of collecting the data for purposes of CCI and nowcasts of turnover indices.
- Modification of Newsfeed application for purposes of consumer confidence index (in SGA-2).

Task 3 - Methodology and techniques

- Exploration of the methodology of collecting and processing the data for CCI and nowcasts of turnover indices.
- Exploration of feasibility of linking other administrative and existing sources for CCI and nowcasts of turnover indices.
- Initial quality assessment of the input (and throughput) phase of statistical process

Task 4 - Future perspectives (SGA-2)

- Definition of methodology of collecting the data for CCI and nowcasts of turnover indices.
- Calculation of the CCI and nowcasts of turnover indices based on big data sources.
- Quality assessment of calculated CCI and nowcasts of turnover indices.
- Execution of pilots which combine two or more sources for the aim of "early estimates".

Deliverables (SGA-1 only)

6.1	List of potential big data sources together with the business cases for the aim of early estimates (general)	month 13
6.2	Recommendations about IT tools for collection of data for purposes of consumer confidence index and nowcasts of turnover indices	month 13
6.3	Recommendations about methodology for processing the data for purposes of consumer confidence index and nowcasts of turnover indices	month 13

Milestones (SGA-1 only)

6.4	Interim feasibility report	month 4
6.5	Progress and technical report of internal WP-meeting	month 6