26th Meeting of the Wiesbaden Group on Business Registers
- Neuchâtel, 24 – 27 September 2018

Dimitar Nenkov, David Broska (Eurostat),

Johannes Frey, Sebastian Hellmann (Leipzig University)

Session 5 - New Data Sources

New Data Sources

Abstract

Eurostat governs the EuroGroups Register (EGR), the statistical business register of multinational enterprise groups in the European Union.

In order to create the EuroGroups Register Eurostat collects enterprise group information from the national statistical business registers of the EU Member States, participating EFTA countries and commercial data sources. The EU part of the legal units, enterprises and enterprise groups and their characteristics are therefore well-covered by the EGR.

Additional sources of information such as crowdsourcing platforms, web crawling and different open data projects are seen as further opportunities to increase the quality of the EGR, its completeness and accuracy namely with the units outside of the EU and EFTA as well as on the whole group level.

Under the umbrella of Eurostat BIG DATA project, the EGR Team is investigating these additional data sources. Eurostat is collaborating with the Leipzig University to explore the possibility of using DBPedia as new additional source of data of multinational enterprise groups. DBpedia is a project which extracts structured information from Wikipedia to make it publically available in a format that allows to ask sophisticated queries against Wikipedia and to link different data sets to Wikipedia data.

The paper will present the results of the feasibility study carried out by EGR Team and Leipzig University. The objective of the Prove of concept (POC) is to automatize at a large extend the collection of aggregated whole group figures using as input the names of the enterprise groups. Currently final checks and updates of enterprise group figures are done manually by looking into the data published in the annual accounts or websites of the enterprise groups. A population of 70 enterprise groups (EGs) was selected based on size and geographical diversity and provided to DBPedia for matching. The main attributes which were targeted to be collected were persons employed, turnover and assets. The following indicators were analysed:

- *Coverage – number of successful matched enterprise groups names*
- *Completeness – number of received values for the different attributes*
- *Accuracy – quality of the returned values when compared to the figures published by the enterprise itself*

Keywords: Statistical business registers, Crowd sourcing data, Alternative data sources for multinational enterprise groups.