

Enhanced Access to Micro Data of Official Statistics - Integrated Data and Meta Data Management System of the German Research Data Centres

Topic 4 – Getting the statistics out

Keywords: Meta data, research data centre,

Introduction

The two Research Data Centres of the German Federal Statistical Office and the Statistical Offices of the Federal States offer micro data from more than 100 official statistics for scientific purposes. They provide different ways of access to this data like safe centres, remote execution, scientific use files, public use files and campus files. These ways of access are limited to institutions of higher education or other institutions with tasks of independent scientific research.

The availability of detailed documentation for those datasets is crucial, to provide secondary users with a full understanding to enable them to use data effectively, efficiently, and accurately. Due to an upcoming change in the statistics law and a planned new meta data system, the future access to micro data will be improved. The new meta data system should reduce the work of maintenance of the meta data and will lead to more up to date information for the user. This is due to the fact that it will be included in the new meta data management system of the common statistics portal that is now under development.

The first features will be released at the end of 2016. It also will support the retrieval and access to micro data and their usage. The future possibility of an access to only formally anonymised micro data inside the safe premises of the statistical offices by domestic and foreign guest researchers will especially improve promptness of data access and a better handling of confidentiality.

Methods / Problem statement

User surveys and a direct on-going feedback from users about currentness, completeness and overall quality of the existing meta data system of the research data centres revealed a dissatisfaction with the structure and quality of meta data presented on the website. This is the reason why researcher mostly use only pdf documents with descriptions of the surveys. The maintenance and administration of the old meta data system caused a lot of work in the research data centres, but did not contribute to a satisfaction of meta data users.

All this led to the decision to install a new meta data system responding to data requests often mentioned by the users about survey descriptions, definitions, interpreting assistance and other meta data. The proliferation of statistical micro data and their broad use led to more complex analysis with an increased demand on combining data of different surveys. This has raised the issue of consistency, comparability and in particular the problem of protecting confidentiality of the micro data.

The new meta data system should help users better to discover, understand, interpret and interrogate the micro data they need. Complex micro data sets combined from different surveys are used in the safe centres and that has to become more comfortable and efficient.

Results / Proposed solution

The novel meta data system of the Research Data Centre will be integrated in the new meta data management system of the common statistics portal of the German Statistical Offices. In statistical organizations meta data exist throughout the entire statistical production process. A prerequisite of the new meta data management system therefore is that as many meta data as possible are captured automatically, as a result of a computer process. Via different application programming interfaces selected information should be extracted from all available sources and data bases to fill the meta data repository.

The implementation of a data sharing model will lead to improved efficiency through central meta data repositories that are organized to facilitate reuse of existing data. The automatic management should ensure maintenance a meta data history and updating links between meta data in the corporate repository. The new meta data management system will be an in-house development of the statistical offices in Germany. The core of the meta data system of the research data centres will be a so called meta data report.

On the level of the statistic it will contain detailed information on the single statistic, the methods, quality and publications. On the level of the single data set information on data processing, the variables etc. will be included. Every data set that is used will get a digital object identifier (DOI), that enables researchers to cite the data sets unambiguously.

Conclusions

German Research Data Centres in official statistics will introduce a meta data system that reacts flexible and responsive to rapidly evolving requirements for statistical micro data and will, in the first phase, meet the demands articulated for the most used data sets. After an evaluation of user satisfaction and implementing recommended changes, the meta data reports will be expanded to all statistics with active demand in the next step. The attention will be focussed on the improvement of user support with regard to finding data and the assistance of their analysis to provide accurate results.

Due to the improved data access in the safe centres of the research data centres this way of analysis of official micro data has to be extended. Besides the two research data centres of the German Federal Statistical Office and the Statistical Offices of the Federal States there is a decentralized network of 31 data centres that provides access to a wide range of relevant research data. All data centres are accredited by the German Data Forum. The efficiency of this network could be increased by harmonizing meta data, standards, work tasks, documentation and operational procedures to ensure data protection.

Conclusively the design of future structures of the research data infrastructure in Germany will be addressed in this contribution.