Access to official statistical (OS) microdata: What are the ideas of the researcher?

For the proposal of the DwB portal resource discovery functionality we interviewed researchers who have experience with access to microdata. What are their experiences and expectations?

Search strategy

There is no single search strategy; some researchers prefer to start with a simple search, whereas others want to start with browsing. Often the choice of search strategy is dependent on the research phase, from orientation (simple search) to searching for specific variables. A lot of the researchers mentioned that they often make use of the expertise of colleagues and journal articles to find information about data. Alongside the search for data, the suggestion was to disclose (academic) journals and publications through the portal as well.

A large number of interviewees wished to search or browse on 3 dimensions: topic, time, and location. All the researchers wished to search for variables.

Quick overview

Researchers, generally, would like to get a quick overview of all resources by country, year, topic, accessibility and type of data. At the same time, it should be transparent to the user why the results fit their search terms. Some researchers indicated that they don't want to search in all resources,

whereas another wanted to specify the number and the kind of collections.

Good documentation

In addition to resource discovery, good documentation is essential for the understanding of the data. Almost all the researchers indicated that they would like to have more extensive documentation reports. At the moment information about aspects of the data collection (sampling procedure, mode of collection, etc.), weighting of the data, and other methodology characteristics are often lacking. The interlocutors stated links to codebooks, methodology reports and quality reports are very important. Furthermore they would like to have accurate information about the way a concept is measured, precise information about the variables and the question wordings. At least the survey questionnaire should be made available. Each kind of data needs its own type of documentation; in the case of register data, information about the process the registers use to collect their data would be very useful.

Comparability

National Statistical Institutes (NSI) tend to publish only the documentation on recent surveys. Several researchers stated that they would like to have information about previous surveys as well. In addition, information about trend breaks, changes over time in methodology, changes in data collection mode, question wordings, etc. is essential for the comparability of longitudinal data. Access to this type of documentation is seen as very essential.

Another type of comparability is the comparison between countries. An issue mentioned during

the interviews is the difference across data providers in the use of topics, when searching for comparative studies. Data providers may use different topics, or topics in different languages for the same phenomena. The use of a multi-lingual thesaurus would overcome some of the language problems.

Some interviewees indicated a more fundamental issue with comparative research. A basic knowledge of the national context of a country is needed to be able to interpret the results of the analysis. Reports with key indicators are essential in this respect.

Information about procedures

Some interviewees indicated that the whole process to obtain access to data should be accelerated. This is way beyond the goal of the DwB resource discovery portal. However, the portal should redirect the researcher to the appropriate web pages with information about procedures on how to access the data. Information about availability, access conditions and procedures should be available. Alternatively if the Eu-RAN (remote access network), as proposed by DwB work package 4, is available then researchers will be directed here.

User generated content

Some researchers indicated that they would like to annotate their search results and share these with their colleagues. They would like to share documentation, work on harmonisation, expert knowledge, literature, papers, and sometimes even syntax of the analyses. They spoke of user groups, Wikis on OS data and crowd sourcing of expertise. OS data is in many cases only available in a secure environment, the incorporation of user created content should therefore be accompanied by strict security regulations. This form of collaborative online environment for research is called a VRE - Virtual Research Environment, and as such this class of functionality would be a very practical feature of the portal.

Background information

This information was gathered in 2012 during 26 interviews with researchers in Germany, Spain, the Netherlands and Romania, and a focus group meeting with several European researchers.

Read more

Organisations involved in this research were: KNAW-DANS, CNPS-INS, CIS, Destatis, UEssex and MT. The underlying report (DwB Deliverable 8.4) is available at
http://www.dwbproject.org/about/deliverables.html

Wordle made of reports of interviews