

Editor's notes

Data, the whole Data, and nothing but the Data ... and the Metadata, and the Access to Data

Welcome to the third issue of volume 38 of the IASSIST Quarterly (IQ 38:3, 2014). This issue is unquestionably about data. There are three papers on projects for improving delivery of data to users.

The first paper is 'Distributing Access to Data, not Data' by David Schiller from the Institute for Employment Research (IAB) at Nuremberg (Germany) and Richard Welpton at UK Data Archive, University of Essex (UK). They focus on the problem that access to European microdata for researchers is restricted by national borders and the barriers for performing comparative analyses between the member states. The 'Data without Boundaries' project now has an initiative to build a 'European Remote Access Network' (EuRAN). The problem is that prevention of identifying respondents in the microdata conflicts with the importance for modern research methods of access to detailed data. Some control is necessary and the paper describes remote access as the appropriate answer in the forms of job submission, remote execution, and remote desktop. As an example, one version of secure remote desktop access encrypts pictures of the desktop screens to make secure the transport over the Internet. The authors reference a set of principles for access, e.g., that it is not desirable to physically move data and that access should come through a single point that can access multiple sources of data. The researchers' need to analyse the data is supported by a 'Virtual Research Environment' that includes software for generating and presenting results through the EuRAN project.

The next paper presents a two-year metadata project based upon two well-known series of studies: the American National Election Study (ANES) and the US General Social Survey (GSS). The goal is to improve their metadata and build demonstration tools to illustrate the value of structured, machine-actionable metadata as reported in 'Creating Rich, Structured Metadata: Lessons Learned in the Metadata Portal Project'. The authors are Mary Vardigan (Inter-university Consortium for Political and Social Research (ICPSR)), Darrell Donakowski (American National Election Studies (ANES), University of Michigan), Pascal Heus (Metadata Technology North America (MTNA)), Sanda Ionescu (ICPSR), and Julia Rotondo (NORC at University of Chicago). The article reports on their experiences, and also includes recommendations. The National Science Foundation funded the project under the 'Metadata for Long-standing Large-Scale Social Science Surveys' (META-SSS) program. ICPSR and ANES are co-distributors of most of the ANES studies while the GSS is co-distributed by NORC, the Roper Center, and ICPSR. In the project metadata tools revealed small differences between supposed identical datasets, for instance in study titles, variable names, etc. The project also decided which types of content to include. Both of the the series are huge collections - as the 58 ANES surveys contain 79,521 variables and the cumulative GSS has 5,558 variables. Marking up this legacy documentation is laborious and time-intensive

and the future naturally lies in capturing the metadata at the source. In conclusion, the project learned a great deal about converting legacy documentation and identified several steps for documentation development, including the areas of paradata and versions of datasets. The concept of versions of datasets relates to the solution described in the first paper of not bringing data but access to data to the users.

The third paper demonstrates further work in the project described above. In the paper 'Mapping the General Social Survey to the Generic Statistical Business Process Model: NORC's Experience' the three authors - Scot Ausborn, Julia Rotondo, and Tim Mulcahy - all from NORC at the University of Chicago - present how they carried out the mapping of the GSS workflow to the Generic Statistical Business Process Model (GSBPM). An analysis of the business processes for the production of survey data was carried out with the intention of direct capture of survey cycle DDI-based metadata, thus avoiding the need to generate it retroactively. The work is based upon an internal survey of GSS staff, asking them to explicate their respective roles on the survey in terms of the GSBPM. Connecting aspects of the GSS workflow to elements of the GSBPM produced a comprehensive and integrative view of the individual efforts that together produce the survey. Of the lessons learned, I noticed that they later found that it may have been more fruitful to have held a workshop in which GSS staff could discuss the workflow processes together, rather than having a survey with each person providing his or her input in isolation. They mention that they think an expert in GSBPM could have conducted the mapping of the workflow; however they did identify points for improvement in the workflow relating to both metadata and paradata.

Articles for the IASSIST Quarterly are always very welcome. They can be papers from IASSIST conferences or other conferences and workshops, from local presentations or papers especially written for the IQ. When you are preparing a presentation, give a thought to turning your one-time presentation into a lasting contribution to continuing development. As an author you are permitted 'deep links' where you link directly to your paper published in the IQ. Chairing a conference session with the purpose of aggregating and integrating papers for a special issue IQ is also much appreciated as the information reaches many more people than the session participants, and will be readily available on the IASSIST website at <http://www.iassistdata.org>.

Authors are very welcome to take a look at the instructions and layout:
<http://iassistdata.org/iq/instructions-authors>

Authors can also contact me via e-mail: kbr@sam.sdu.dk. Should you be interested in compiling a special issue for the IQ as guest editor(s) I will also be delighted to hear from you.

Karsten Boye Rasmussen
March 2015
Editor