

Editor's Notes

When things get digital and huge. Doing the things right and doing the right things.

Welcome to the fourth issue of Volume 40 of the IASSIST Quarterly (IQ 40:4, 2016).

There is a lot of management involved in the data management carried out at data archives and with data collections. The phrase 'Doing the things right and doing the right things' belongs to fathers of modern management and is used to distinguish management vs. leadership, efficiency vs. effectiveness, and tactics vs. strategy. The winning authors of the 2016 IASSIST paper competition used the article 'More Product, Less Process: Revamping Traditional Archival Processing' (Mark A. Greene and Dennis Meissner, 2005) as their starting point for investigating the 'More Product, Less Process' (MPLP) approach for digital data. The winning paper 'More Data, Less Process? The Applicability of MPLP to Research Data' is written by Sophia Lafferty-Hess and Thu-Mai Christian. The authors work at the Odum Institute for Research in Social Science, University of North Carolina at Chapel Hill as Research Data Manager and Assistant Director of Archives. The paper was presented in the session 'Data Management Archiving/Curation Platforms' at the IASSIST 2016 conference in Bergen.

In their paper Lafferty-Hess and Christian set out to apply the principles and concepts formulated in MPLP to the archiving of digital research data. They discuss data quality, usability, preservation and access, leading to the question: What is the 'golden minimum' for archiving digital data? In terms of data archiving, spending too much effort on doing the things right may bring the trade-off problem that the resources are not sufficient to do all the things. Users in the digital world retrieve and consume lots of information by themselves, but digital data comes in forms that are seldom directly consumable without additional processing. When the authors also relate the phrase 'golden minimum' to the phrase 'good enough', management is again brought into the discussion. In my view, the short formulation of Herbert Simon's 'satisficing' concept in his theory of bounded rationality is 'good enough is best'. Lafferty-Hess and Christian are aware that shifting responsibility for certain data curation tasks from the data archive to the data producer and to the data user can present problems. Their best advice and hope for the future is that additional 'future research will help us build better understanding of the connection between user needs and data curation processes'.

The following paper 'MMRepo - Storing qualitative and quantitative data into one big data repository' is authored by Ingo Barkow, Catharina Wasner and Fabian Odoni, working at University of Applied Sciences Eastern Switzerland HTW Chur where Barkow is Associate Professor and Wasner and Odoni are research associates. They describe a prototype of their MMRepo project that addresses the problem of storing qualitative large binary objects with regular quantitative data in order to achieve the advantage of storing mixed mode data in the same

infrastructure, whereby only one system needs to be provided and maintained. Linking to the first paper they are looking into the efficiency problem of doing the things right. When you are efficient you can do more things right. The project is trying to achieve this by combining CERN's Invenio portal with a Hadoop 2.0 cluster and DDI 3.3. The prototype was successful and the project continues. The paper was presented at the IASSIST 2016 conference in the session 'Technical Data Infrastructure Frameworks'.

Aidan Condron works with the Big Data Network Support team at the UK Data Service. At the IASSIST 2016 conference he presented 'Data Science: The Future of Social Science?' at the session 'Big Data, Big Science', and has submitted this presentation as the paper 'Servicing New and Novel Forms of Data: Opportunities for Social Science'. These 'new and novel' forms are, for example, social media data that present potential resources for researchers but also pose challenges for access provision and analysis. The paper introduces Data Service as a Platform (DSaaP), which is a project to establish technological infrastructure support. As with the MMRepo project, the DSaaP project will include both familiar and new and novel forms of data. The novel forms of data are often huge, and 'Hadoop' solutions are also at play here using a data lake built through use of open source software. The article also gives several demonstrations through graphs of energy consumption based on 3.7 billion datapoints. After the presentation and the paper, the Big Data Network Support team will standardise and generalise the procedures developed from their DSaaP project.

Submissions of papers for the IASSIST Quarterly are always very welcome. We welcome input 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. We permit authors 'deep links' into the IQ as well as deposition of the paper in your local repository. 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.iasistdata.org>.

Authors are very welcome to take a look at the instructions and layout:

<http://iasistdata.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
July 2017
Editor