DDI DIRECTIONS

A NEWSLETTER OF THE DATA DOCUMENTATION INITIATIVE

Volume XV, Number 1, February 2017

From the Director

In this issue, we report on the recent 8th Annual European DDI Users Conference and announce the upcoming 5th Annual North American DDI Users Conference. We invite you to attend these conferences and, especially, to present about your own DDI activities. We also invite DDI Alliance members to attend our annual meeting on May 22nd in Lawrence, Kansas, where we'll discuss the state of the Alliance, planning for a new strategic plan, and DDI technical developments.

Jared Lyle, Director, DDI Alliance, lyle@umich.edu

In This Issue

Save the Date: DDI Alliance to Meet in Lawrence, Kansas in May

The DDI Alliance will hold its annual meeting on Monday, May 22, 2017, in the Big 10 Room of the University of Kansas Memorial Union in Lawrence, Kansas (the day before the start of the IASSIST conference). The morning will be devoted to the Meeting of Members and the afternoon to the meeting of the Scientific Board with lunch provided in between. As noted in previous years, in most cases it will be the same person attending both meetings, but do feel free to send different people. More details about the meeting will be available soon.

New Member Elected to DDI Alliance Executive Board

The Alliance held a January election to fill an opening on the Executive Board. Dana Muller, head of the Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB) in Nuremberg, Germany, was elected to complete the 2015-2017 term of David Schiller.

Sincere thanks are owed to David Schiller for his excellent service and dedication during his time on the Executive Board.

Call for Papers for NADDI 2017, April 6-7 (Pre-Conference workshop April 5)

A Call for Papers is now open for the 5th Annual North American DDI Users Conference hosted by the Cornell Institute for Social and Economic Research (CISER) and The Roper Center for Public Opinion Research. The theme for this year's conference, “Metadata Across the Research Data Lifecycle,” emphasizes the benefits of metadata creation early in the research data lifecycle, as well as its subsequent re-use throughout.

Aimed at individuals working in and around data and metadata, NADDI 2017 seeks submissions of presentations and posters that highlight the use of DDI and other metadata standards within research projects, survey operations, academic libraries, and data archives. Also invited are proposals that address the increasing need for interoperability of standards in research data management and the leveraging of DDI to facilitate data discovery and data integration. Presentations of an applied nature are encouraged – how are you working with DDI and metadata generally within the larger framework of research data management and the research data lifecycle? Submissions of a more technical nature are also encouraged, as well as presentations on the DDI standard itself and its continued development.

For further details about the conference, please visit http://naddiconf.org. The deadline for submissions is February 17, 2017. We hope to see you there!

8th Annual EDDI Conference Held in Cologne, December 2016

EDDI16, the 8th Annual European DDI Users Conference, took place December 6-7, 2016, in Cologne, Germany. The conference was hosted by GESIS - Leibniz Institute for the Social Sciences, and organized jointly by GESIS - Leibniz Institute for the Social Sciences and IDSC of IZA - International Data Service Center of the Institute for the Study of Labor.

There were 91 participants from 50 organizations in 17 countries in attendance. The conference program included 37 presentations and 6 posters, 3 tutorials, and 7 side meetings. Keynote addresses were given by Keith Jeffery ("Metadata: Foundation, Philosopher's and Rosetta Stones"), and Stefan Winkler-Nees ("Funding, Policies, Community Building - Data Sharing from a Funders Perspective").

The presentation files are now online. Presentations with a Creative Commons license also are available in Zenodo.

Save the date for the next EDDI!

EDDI2017 will be hosted by FORS - Swiss Center of Expertise in the Social Sciences in Lausanne, Switzerland on December 5-6, 2017.

Public Review of XKOS

The DDI Alliance recently conducted a Public Review of XKOS, an RDF Vocabulary which extends the Simple Knowledge Organization System (SKOS) for the needs of statistical classifications. XKOS extends SKOS in two main directions. First, it defines a number of terms that enable the representation of statistical classifications with their structure and textual properties, as well as the relations between classifications. Second, it refines SKOS semantic properties to allow the use of more specific relations between concepts. Those specific relations can be used for the representation of classifications or for any other case where SKOS is employed. XKOS adds the extensions that are desirable to meet the requirements of the statistical community. More information can be found on the DDI website.

RWX: Top Impact Publications from the Last 20 Years

By Knut Wenzig

The DDI Community has produced a rich store of DDI and metadata-related publications. Read-Write-Execute (RWX) will highlight some of these existing publications as well as new work as it is produced. The first column featured some of the foundations of DDI in scientific literature. This second column will revisit some of the top impact publications related to DDI from the last 20 years.

It is not surprising that the DDI publications with many citations cover more high level discussions rather than specific technical details. But revisiting conceptual fundamentals or policy goals, comparing standards, and evaluating approaches should also be done if one is currently planning the next project. So, let's take a look at some of the top cited DDI publications over the last 20 years.

When Ryssevik and Musgrave (2001) write about their social science dream machine, they were thinking about the distributed NESSTAR system, which is based on DDI. But there is nothing wrong with the idea of an “integrated resource discovery gateway and search system to identify and locate these resources” which consists of not less than “all existing empirical data” (what is today called federated search). And being able to convert an “extensive amount of metadata ... totally integrated with the data as such” to a number of formats and copy them to a local machine is a reasonable wish. The same holds true with “an efficient feedback system to the body of metadata, allowing the user to add to the collecting memory of a data set”. Even “The FAIR Guiding Principles for scientific data management and stewardship” (doi:10.1038/sdata.2016.18) from 2016, which are considered to be state of the art, do not cover the range of features Ryssevik and Musgrave describe.

The most cited publication in 2004 contains an important reminder: “Technology itself, however, will not fulfill the promise of e-science, Information and communication technologies provide the physical infrastructure. It is up to national governments, international agencies, research institutions, and scientists themselves to ensure the institutional, financial and economic, legal, and cultural and behavioural aspects of data sharing are taken into account.” (Arzberger et al. 2004: 137) The use of DDI, especially at ICPSR, serves as a use case for the technological domain where access and usability and multiple use of the data must be assured by interoperability.

While Arzberger et al. look at use cases from different disciplines in the different identified domains, Willis, Greenberg and White (2012) compare nine metadata standards in order to understand similarities and differences. They consider DDI as the standard to describe social science statistical data from experimental, observational, and statistical studies. The objective to cover the whole data lifecycle is unique to DDI. DDI is one of two standards

which “are intended to be comprehensive, yet support instances of description using a minimal number of required elements.” They conclude that metadata scheme creation depends more on the goals than on the discipline or type of data described (p.1517). At the same time the common discipline specific approach contributes “to artificial boundaries between disciplines and impede interdisciplinary and transdisciplinary reuse” (p. 1516).

For Jeffrey et al. (2014), who describe the CERIF approach to design a research information management system, domain specific metadata standards build the lowest of three levels of information. The first level consists of information on research output (organized by flat metadata like Dublin Core similar to a catalogue card). The second level is built by contextual metadata, which can generate the discovery metadata of level one and point to the domain metadata of level three (which could be DDI). The contextual metadata hold information about base entities (e.g., persons and publications) and connect them using a semantic layer with flexible link entities, which can express roles (defined by a term which captures the semantics and a controlled vocabulary to which the term belongs (p. 10) and have a start and end date). Using this semantic layer a publication can have an author, a publication date, and even a country of publication (using so called localisation entities).

This small list of four top publications related to DDI:

- shows us that looking more than 15 years back might yield new insights into new products from old ideas,
- reminds us that technology does not solve social problems,
- reveals different perspectives on the discipline specific fragmentation of metadata standards,
- and gives an insight into a concept of a flexible and expressive linking mechanism.

References (also available at Bibsonomy)

Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L., Moorman, D., Uhlir, P. & Wouters, P. (2004). Promoting Access to Public Research Data for Scientific, Economic, and Social Development. Data Science Journal, 3, 135-152. doi:10.2481/dsj.3.135

Jeffery, K., Houssos, N., Jörg, B. & Asserson, A. (2014). Research Information management: the CERIF approach. International Journal of Metadata, Semantics and Ontologies, 9, 5-14. doi:10.1504/ijmso.2014.059142

Ryssevik, J. & Musgrave, S. (2001). The Social Science Dream Machine: Resource Discovery, Analysis, and Delivery on the Web. Social Science Computer Review, 19, 163-174. doi:10.1177/089443930101900203

Willis, C., Greenberg, J. & White, H. (2012). Analysis and Synthesis of Metadata Goals for Scientific Data. Journal of the American Society for Information Science and Technology, 63, 1505--1520. doi:10.1002/asi.22683

A bibliography of DDI articles, working papers, and presentations is being built and is available at Bibsonomy.org with easily reusable bibliographic metadata. This metadata will also be made available on the DDI Alliance website. Suggestions for papers and topics for RWX, or the bibliography, are appreciated and can be sent to: Knut Wenzig, kwenzig@diw.de