2nd Annual European DDI Users Group Meeting

December 8-9, Utrecht, The Netherlands

Abstracts

as of 4. December 2010 (final version)

Alerk Amin (CentERdata - Institute for Data Collection and Research)
Keeping up with Questasy

Questasy is a data dissemination website application based on DDI 3. It was primarily developed for the LISS Data Archive (www.lissdata.nl), but is freely available for other organizations. This presentation will relate the new developments in Questasy since the 2009 EDDI.

William Block (CISER - Cornell Institute for Social and Economic Research)
If you build it, they will come: the case for creating DDI 3 metadata and the advanced search and discovery tools that will follow

One of the most compelling strengths of Version 3 of the Data Documentation Initiative is the thoroughness with which it is possible to document social science data all through the data lifecycle. Yet, the most richly documented data in the world isn't much good unless search and discovery tools are built to take advantage of the metadata that DDI makes possible. This presentation will review some of the advanced search functions that are already available on the Internet and make the argument that what is currently lacking is a wealth of metadata for these advanced tools to compute upon. As a result, this presentation will argue that next generation search and discovery itself can serve as a justification for local investment in the creation of DDI 3 metadata: if you build it (metadata), they (advanced Internet search tools) will come.

Tito Castillo (Institute of Child Health, University College London)
The use of DDI tools and standards in epidemiology & public health

This presentation will describe the experiences of the SERPent project (Secure Epidemiology Research Platform at University College London) with DDI tools and standards, predominantly DDI 2. The aim of this project, funded by JISC as a fast-track Virtual Research Environment project, was to begin develop a metadata catalogue for epidemiology and public health research studies to promote effective information governance and support the documentation, preservation, discovery, access, and use of the underlying data.

A small project management team has work closely with researchers from 5 selected use cases, representative of a range of public health and epidemiology research studies within UCL, to document the data and build a corresponding metadata catalogue. Under the guidance of Pascal Heus (Metadata Technology) we have employed a combination of the IHSN Microdata Management Toolkit, the NADA web catalogue and bespoke software utilities for metadata extraction and transformation.

A significant element of this project has involved cultural change and shared learning of the potential of standards-based metadata representation. Although significant limitations of DDI 2 certainly do exist, the disciplined approach to the description of project metadata has proved popular with all participants.

The project is due to be completed by the end of October 2010 at which time a retrospective session will be conducted to learn lessons and formulate an ongoing strategy for metadata management. We anticipate that the next phase of this work will involve the development of a viable migration path to DDI 3 and integration with existing data management tools.

Louise Corti (UK Data Archive, University of Essex)
Progress from the DDI Working Group on Qualitative Data
We will provide a report on work undertaken by the Group which held its first meeting in April 2010. The membership of the group comprises 20 people across 10 countries, all actively engaged in archiving or wanting to archive qualitative data.

The first meetings have aimed to collect a set of broad use cases which can then be refined, prioritised and translated into a set of technical use cases. Standards and tools development can follow from that with input from the DDIs TIC. The use cases have covered most areas of the life cycle - from raw and enhanced data relationships and ingest, management and storage of data storage, basic and rich data and contextual description, through to exposure of data for exploration and visualisation and user needs. They have also looked beyond purely qualitative data to include linking to numeric data and handling more complex social science data collections.

We present an overview of the use cases and also report back from related work happening on DDI, QuDEx (the qualitative data exchange schema proposed by Corti and Gregory in 2008) and FEDORA ingest.

Nadine Dulisch (GESIS - Leibniz Institute for the Social Sciences)
Converting general Microsoft Word questionnaires to DDI
In the social sciences questionnaires are commonly available in form of office documents. They are mostly created with word processing programs, especially Microsoft Word. In these documents data is presented formally unstructured and thus can't be processed by other software, because the contained text can't be recognized in regard to its semantics. The structure of Word documents is just oriented on paragraphs supplemented with layout.

In addition, a questionnaire is often subject to frequent revisions during its creation process. These revisions are rarely documented to a sufficient extent due to the high number of versions being created in that process and the lack of supporting software systems.

The software QDDS allows editing of questionnaires together with a detailed documentation of the whole process of questionnaire development. This makes analysis and archiving of data possible, also with regard to further development.

In QDDS data is stored according to the DDI-specification and is therefore highly structured. To make use of the possibilities of structured data concepts are needed how to convert questionnaires from office documents.

This talk focuses on the development of an import-interface for office documents that allows to document questionnaires available in Microsoft Word format with QDDS.

Johan Fihn, Olof Olsson, Iris Alfredsson, Hans Jørgen Marker (all SND - Swedish National Data Service)
DDI at the Swedish National Data Service
This session is an overview of the work done with DDI involvement at the Swedish National Data Service (SND). We intend to present some of the tools we have developed used on DDI, as our in-house documentation system, currently being upgraded from supporting DDI2 to supporting DDI3, our question data bank and the stylesheets for DDI documents, with which we build our codebooks. We also intend to discuss the possible use of DDI as an intermediary standard for transferring documentation between SND and other agencies as Statistics Sweden.

Christian Gerhards (Bielefeld University, Faculty of Sociology)
Data Service Center for Business and Organizational Data – Specifics concerning DDI dealing with organizational data
The Data Service Center for Business and Organizational Data (DSZ-BO) at Bielefeld University collects, archives and distributes business and organizational data from the social sciences. The goal is to bridge the gap between data producers and data users by providing an institutional and technical framework for the acquisition, standardization, preservation, and dissemination of data for scientific usage.

Therefore a fully web based user interface is planned, which contains the following functionalities:

- data entry and data visualization by the DSZ-BO staff
- searching and browsing functions for data users
- access to microdata (restricted via scientific use files and remote access etc.)

In cooperation of the Faculty of Sociology with the Bielefeld University Library a technical infrastructure for the storage, visualization, management, and retrieval of organizational research data is created in order to facilitate efficient access to existing study information. Most of the data is documented using the DDI 3.1 standard, such as general information about the study, data collection, inquiries in a study, questions and variable information. Our presentation will focus on certain problems that specifically arise from modeling business and organizational data with DDI 3.1. The DSZ-BO has to find ways to describe important information about the unit "organization" and solve the problem of linked multilevel data, e.g. on employers and employees (so called LEE-data). Another difficulty is that a lot of organizational studies are based on qualitative research methods and therefore need another focus of documentation. A more general problem is the lack of controlled vocabularies (especially for organizational data, e.g. classification of employers) for the interoperability between different data centers. Connected with this problem, a further issue will

comprise question databases and thesaurus management systems which should be globally available, editable, and accessible among different data centers.

Jeremy Iverson (Algenta Technologies)
Metadata Driven Survey Design

In current survey practice, the creation of a data collection instrument involves two distinct steps. The first is survey design, in which a researcher defines the questions and flow of a survey. The second is survey implementation, in which a researcher or programmer turns the design into an electronic or paper survey instrument.

This paper presents an alternative approach to survey development. Metadata-driven survey design means that the actions taken to define a survey are the same actions that create the survey instrument. Using a metadata-driven approach to survey design provides several benefits: less redundant work; better, easier data documentation; reusability of key survey components; increased data harmonization potential; and greater research integrity.

Stefan Kramer (CISER - Cornell Institute for Social and Economic Research)
Development of the next DDI Tools Catalog

In its meeting of May 31, 2010, the DDI Alliance Expert Committee decided to establish a new Tools Catalog Group to develop and maintain the next version of the DDI tools listing (current version at: http://www.ddialliance.org/resources/tools). This presentation will provide an update on the work of the group to date, including a demonstration of the new version of the tools catalog in development, solicit feedback on the usability of this developing online resource, and provide an opportunity for interested volunteers to become involved in the ongoing effort.

Monika Linne, Alexander Mühlbauer and Wolfgang Zenk-Möltgen (all GESIS - Leibniz Institute for
The STARDAT Project: Integrating DDI Tools at the GESIS Data Archive

The GESIS Data Archive for the Social Sciences provides high quality data and documentation of survey datasets. Tools to create standardized documentation on study level and on dataset level have been developed already for some time. With DBKEdit, a web-based editing system for bi-lingual study descriptions, and DSDM (Dataset Documentation Manager) for language specific documentation on variable level there are two tools available which are compliant to DDI2. However, the challenges of the DDI Version 3 and the collaboration needs at different stages of the data life cycle leaded to the awareness that an integrated management system for metadata is needed. Therefore GESIS started a project to develop the STARDAT application. It will be based on DDI 3 and will contain modules for structured metadata capture, management, and dissemination. STARDAT will be integrated into the workflow between data depositors, data managers, and data users. STARDAT will support multi-language documentation on study and variable level, and provision of further information, e.g. about related publications, classifications, continuity guides, scales or trends. The talk will give an overview about the challenges of going into the DDI 3 world and will provide information about the current initial state of the project.

Don McIntosh (STR - Space-Time Research) / presented by Arofan Gregory
A query-based access system using DDI and SDMX

In a recent project with the Australian Bureau of Statistics, titled Remote Execution Environment for Microdata (REEM), Space-Time Research built a system to ingest DDI 3.1 for weighted survey

microdata and then provide both ad hoc queries through a browser-based UI, plus a RESTful SDMX API for querying metadata and constructing aggregated output directly from microdata. The result is a query-based access system, which delivers greater flexibility than regular dissemination based access, while still preserving respondents' privacy through use of new robust confidentiality techniques.

This presentation covers an outline of the overall system, our experiences and impressions of the DDI standard, and the extensions to SDMX to enable it to support ad hoc tabulation queries.

Meinhard Moschner (GESIS - Leibniz Institute for the Social Sciences)
Controlled vocabularies for DDI and other metadata structures - an update

The presentation will describe ongoing work on the publication of a first set of controlled vocabularies by the DDI Alliance. It will resume the current scope of vocabularies and technical formats for publication.

Issues of purposes and advantages for DDI and other metadata structures will also be addressed: lexical control, consistency and efficiency, interoperability, support for machine-actionability, as well as retrieval precision and potential multi-linguality.

Ørnulf Risnes (Nesstar)
DDI + API: building services on top of your existing DDI holdings

A Nesstar Server is an example of a web-based container that can store and make available DDI metadata to consuming applications.

Once a DDI-document is published to a Nesstar Server, the contents of the document is made available via a RDF API implemented in Java.

Developers of third party clients may use this API to connect to any such server, and programmatically navigate and harvest collections of DDI-documents held there, or just to extract information from specific DDI-fields relevant to the consuming application.

For many years, most clients developed to interact with the Nesstar API was developed in-house, and integrated as part of the Nesstar Software Suite.

Recently, however, third party clients have started to emerge, and they use the DDI-container and the API for very different purposes, including:

- Variable "shopping carts" for simplified downloads of complex data sets
- Harvesters taking snapshots for persistent archiving in DataVerse
- Automated harvesters for indexing in Solr-powered search systems, including searchable study- and question/variable databases
- Search engine optimized rendering of DDI content for exposing DDI-holdings to generic search engines

The presentation will demonstrate the basic architecture of the DDI-driven API, and show examples of some of the current services built to interact with the API.

Dan Smith (Algenta Technologies)

Colectica: A Platform for DDI 3-based Metadata Management

This demonstration will show Colectica, a set of fully supported, commercial DDI 3 tools specifically designed for survey design and data documentation. Colectica desktop applications are used to create and manage your study's metadata. The software enables data and documentation to be published to the web and to paper documentation formats. The tools can generate CAI source code, paper questionnaires, and statistical source code.

Colectica Repository is an ISO 11179 based metadata repository, backed by DDI 3, which enables collaborative workflows for the entire research process.

The entire data life cycle can be easily visualized using the free Colectica Reader application.

The Colectica Platform is an ideal solution for statistical agencies, survey research groups, public opinion research, data archivists, and other data centric collection operations that are looking to increase the expressiveness and longevity of the data collected through standards-based metadata documentation.

Samuel Spencer (ABS - Australian Bureau of Statistics)

Putting DDI in the driver's seat: Using Metadata to control data capture

The internet being the largest communications network in use, and numerous statistical agencies are looking at using the technology as a data capture source. With this in mind we will look at the data capture possibilities contained within DDI and see how these map to existing technologies to examine the ways well-documented metadata can control the flow of data capture and how this can offer benefits to users of DDI and the web.

Along with looking at DDI and XForms based example from the ABS, we will examine other uses of DDI metadata to format, present and control statistical data capture.

Steven Vale (UNECE - United Nations Economic Commission for Europe)

Exploring the relationship between DDI, SDMX and the Generic Statistical Business Process Model

The UNECE and the Conference of European Statisticians Steering Group on Statistical Metadata (better known as "METIS") have recently developed the "Generic Statistical Business Process Model" (GSBPM). This model has already been widely adopted by national statistical organisations around the world, and is intended to facilitate the convergence of statistical production processes, both within and between organisations. There are certain obvious similarities between the GSBPM and the DDI 3 Combined Life-cycle Model.

At the same time, there is growing interest in official statistics in using DDI 3 in the earlier phases of the statistical production process (particularly for microdata), perhaps in combination with SDMX (Statistical Data and Metadata eXchange) standards, which are seen as more appropriate for macro-data. This paper highlights the work so far on exploring the relationships and interoperability between DDI, SDMX and the GSBPM.

Joachim Wackerow (GESIS - Leibniz Institute for the Social Sciences)

Locating objects identified by DDI3 Uniform Resource Name

DDI3 Uniform Resource Names (URN) are great as persistent identifiers for DDI3 objects. An application would need not just an unique name (URN) but a physical location of the DDI3 object which is identified by the DDI3 URN. A resolution system for DDI URNs will be presented which provides pointers to DDI services delivering DDI objects. The resolution system is based on the Domain Name System (DNS).

Wolfgang Zenk-Möltgen, Brigitte Hausstein (both GESIS - Leibniz Institute for the Social Sciences), Jan Brase (TIB - German National Library of Science and Technology)

DataCite: Making Data Citable

DataCite is an international consortium to establish easier access to scientific research data on the Internet, increase acceptance of research data as legitimate, citable contributions to the scientific record, and to support data archiving that will permit results to be verified and re-purposed for future study. DataCite promotes data sharing, increased access, and better protection of research investment. Just as science is global, with individual researchers working and publishing, DataCite with 12 members from 9 countries is global, with individual regional member institutions offering services and advice directly where they are needed by the scientists. The talk gives an insight view on the DataCite network and introduces the efforts to standardise Metadata for DOI registration. It reflects on synergies between DataCite Meta Data Kernel and the DDI standard and discusses the potential of back and forth referencing in both standards.

As a member of DataCite, GESIS operates the non-commercial registration agency da|ra to make social science primary data permanently identifiable and available with DOI names. da|ra pursues the goal of promoting and establishing global, uniform standards for acceptance of research data as independent, citable scientific entities. The talk will introduce the main features of the da|ra DOI registration service for research data.