1st Annual European DDI Users Group Meeting: DDI - The Basis of Managing the Data Life Cycle

Deutsche Post DHL, Bonn
December 4, 2009

Organizers: **Joachim Wackerow** (GESIS - Leibniz Institute for the Social Sciences)
Nikos Askitas (IZA)
Georgios Tassoukis (IZA)

Abstracts

"Implementation of Questasy: Online data information and dissemination using DDI 3"
Alerk Amin (CentERdata, Institute for Data Collection and Research, Tilburg University)

Questasy is a web application developed to manage the dissemination of data and metadata for panel surveys. It was primarily developed for the LISS Data Archive, but was designed to be repurposed for other surveys. The structure of the application, from the underlying database to the generated web pages, is based on DDI 3. This presentation will describe the design and implementation of Questasy, with emphasis on the mapping of DDI elements to a relational database. The advantages and disadvantages with implementing DDI will also be discussed.

"Usage of an DDI Structure in the NEPS Data Warehouse"
Ingo Barkow (German Institute for International Educational Research (DIPF)), Melanie Kirchner (University of Bamberg), Andreas Thalheimer (University of Bamberg)

The National Educational Panel Study (NEPS) is a longitudinal survey with a long running time and high complexity (5 pillars and 8 stages combined with five waves till 2013). This session will introduce the database structure of the panel as well as give a first outlook on the tools planned to support it (e.g. NEPS Metadata Editor and DDI Im- and Export Webservice).

“The case of CHARMCATS: Use of DDI3 for publishing harmonisation routines”
Martin Friedrichs (GESIS - Leibniz Institute for the Social Sciences), Alex Agache (GESIS - Leibniz Institute for the Social Sciences), Markus Quandt (GESIS - Leibniz Institute for the Social Sciences)

Cessda HARMonisation of CATegories and Scales (CHARMCATS) is a desktop prototype application (developed at GESIS) designed to support in depth documentation of harmonisation routines from the level of concepts to the actual re-coding process of survey data. The presentation will address issues of DDI3 implementation (that is still in its early stage) within CHARMCATS under two main perspectives: 1) Conceptually: it will show insofar does DDI3 as published cover the structural elements required as input and produced in the CHARMCATS workflow and what additional metadata should be considered by DDI3 for the use case of harmonisation; 2) Technically: it will present how the DDI3 elements were mapped to the SQL DB within a Java environment and what are the main benefits and technical pitfalls when employing DDI3 for an efficient implementation. The presentation will focus on the interplay between the Variable and Question Schemes with other DDI3 schemes and will illustrate the use of grouping and of the Comparison Module within CHARMCATS.

“The EVS and NKO Subject Portal for Enhanced Publications: Connecting Archives and Libraries with DDI”
Rob Grim (Tilburg University)

The DataplUS project aims at the creation of two subject portals with enhanced publications that are based on survey data: one for publications on the European Values Study (EVS), and one for publications on the Dutch Parliamentary Election Study (NKO). This project specifically aims at operational subject portals that are ready to use for the researchers of these two large research groups. EVS (Tilburg University) and NKO (University of Twente) will be responsible for the content of the subject portals. The publications will be enriched with the codes of the constructs used and with links to the data and variables, which are stored at the data archives GESIS and DANS. The Library and IT services of Tilburg University will create and maintain the technical features of the portals, as they have gained much experience during the former Samen in Delen project. The library of the University of Twente will be involved in this part as well. CentERdata, Tilburg University, will create a DDI3.0 Enhanced Publication Editor (a data entry form), which enables researchers to enter their enhanced publication in an intuitive manner. The EP-editor will be used to add content to the subject portals during this project, but it will also play an important role to the continuation of the portals afterwards. Finally, collaboration with GESIS, the data archive for the EVS data, and DANS, the data archive for the NKO data, is aimed to result in ‘deep linking’ to variables that are used in the publication and stored in their archives.

"Developments of the Research Data Infrastructure in Germany since the end of the 90's"
Denis Huschka (German Council for Social and Economic Data (RatSWD))

This presentation is an introduction into the work of the German Council for Social and Economic Data (RatSWD). The Council's main purpose is to advise in the development of the German data infrastructure for empirical research in the social and economic sciences. The Council is working not only to increase access to microdata and to sustainably improve data quality, but increasingly also in the development of long-term data surveys, together with both official government institutions (official statistical offices, social insurance institutions, government research units, etc.) and non-governmental institutions (universities and non-university research institutes, e.g., Leibniz Society institutions). All of the Council's work is fundamentally based on the objective of fostering constructive dialogue between the research community and data production facilities.

"Editing DDI"
Jannik V. Jensen (Danish Data Archive)

Presenting an Open Source project for editing DDI. The presentation will lay out how to reuse developed software components and include a demo. The editor application is based on Eclipse RCP, XmlBeans and Oracale DBXML.

"DDI3 and Re-use of Metadata in Archival Processes"
Uwe Jensen (GESIS - Leibniz Institute for the Social Sciences)

DDI3 provides new features to support consequent re-use of metadata. It is of great interest for data archives to capitalize these potentials in future workflows for data documentation and managing collections of complex survey data. The presentation will focus on present considerations and outstanding questions to apply top level modules Grouping, Resource Package and Comparison in documenting comparative survey data.

"Exploring DDI3 in an Archive"
Mari Kleemola (Finnish Social Science Data Archive), **Matti Heinonen** (Finnish Social Science Data Archive)

This presentation will describe the DDI3 evaluation process at FSD. We currently document our data in Finnish and in English using DDI2.1, which has worked well. However, several reasons are leading us to take steps towards DDI3, for example multilinguality issues and the need to avoid repeating information when documenting series of studies. We are also planning to change the way we manage our metadata (i.e. from xml files to database) so careful planning is needed. Furthermore, FSD participated in CESSDA PPP work packages that dealt with DDI3. We decided to use ISSP 2006 Finnish data as our use case. We have marked-up a subset of this data from scratch, using an XML editor and available software. Special attention was paid to multilinguality and comparability issues as well as preservation metadata. The paper will describe in detail how the mark-up was done, what programs were used, and the results of this exercise. We will also make the mark-up available and discuss FSD's future plans re: DDI3.

"DDI for large scale assessments in education – data archiving and question banking"
Martin Mechtel (Institute for Educational Progress (IQB), Berlin)

The IQB – Institute for Educational Progress in Berlin – produces ca. 6,500 test items per year and performs studies to estimate the quality of these items. In order to provide sufficient long term preservation and to enable scientists and the public to reanalyze or just to use the data as a source of information, IQB extends its in-house item database applications towards processing, archiving and reusing datasets. The use of an international XML standard for documentation will ensure that data will be understood outside the IQB systems. Using DDI in the context of assessment of student's performance requires the addressing of several topics which are discussed in the presentation: complex structure of questions, need for storing several attributes to the questions in a flexible way, storing of scoring instructions for each code, storing all information about the test instruments incl. multi-matrix design, and markup of complex generation processes in order reproduce each derived dataset.

"Controlled Vocabularies for DDI3 - a Work in Progress"
Meinhard Moschner (GESIS - Leibniz Institute for the Social Sciences)

The presentation will cover the following topics: 1) General description of how the DDI-CVG (Controlled Vocabulary Group) worked and works to compile controlled vocabularies needed for an appropriate implementation of DDI3, including an overview of the vocabularies drafted so far, the kind of problems faced (e.g. purpose of and place for the respective vocabulary in DDI3) and the limitations due to the variety of kinds of data potentially covered by DDI3. 2) The plans of the Council of European Social Science Data Archives (CESSDA) regarding controlled vocabularies in the context of the forthcoming European Reserach Infrastructure (ERIC), heading towards more harmonized metadata, and in which the adoption and maintenance of controlled vocabularies play an important part. 3) CESSDA plans regarding multi-lingual data documentation (local language + English) and the integration of the controlled vocabularies in the multilingual search functionality provided by the CESSDA data portal (e.g. a search interface which would include filter based on methodology vocabularies).

"How to Present an Application Based on DDI3 to the World Out There? - Experiences from a project in which we have built an online survey data dissemination application based on DDI3"
Marika Puumala (CentERdata, Institute for Data Collection and Research, Tilburg University)

During this presentation we look back at the project in which we have built an online application based on DDI 3. We discuss our experiences from the project as well as the main challenges we faced when applying the DDI 3 structure into practice, such as: "How to create a user interface that is understandable and easily approachable to data users unfamiliar with DDI 3?" Our application, called Questasy, has been operational since February 2009. Questasy consists of a DDI 3 compatible database with a web interface for internal (data entry) and external users. It supports longitudinal studies and contains elements such as questions, questionnaires, datasets, variables, other material, and concepts.

"Creating DDI Metadata for a Diverse Body of Aggregate Statistics: Experience with the GB Historical GIS"
Humphrey Southall (University of Portsmouth)

The Great Britain Historical GIS combines boundary maps for the main British statistical reporting units with a wide range of demographic, economic and social statistics from the last 200 years. Although the original system held statistics in many separate database tables, a new architecture was developed in 2001-4 holding a large subset of the statistics collection in just one column of a single large database table, facilitating graphical presentation via the public and very popular web site "A Vision of Britain through Time". In this architecture, the meaning of each data value – what it measures, as distinct from where and when – is recorded via a metadata framework based fairly closely on DDI2. This paper focuses not on the overall architecture but on the decisions and trade-offs involved in holding a diverse library of statistics in this single data structure. The original system was mainly focused on census statistics and vital registration statistics. More recent additions include labour market data, a large body of statistics taken from British Agricultural Censuses between 1867 and 1971, and election results: the number of votes for each party in each constituency in each British parliamentary election since 1832. Almost no architectural changes have been needed to accommodate these extensions.

"Using a DDI3 Based Single Source Approach to Increase the Efficiency of Social Science Research Processes"
Karsten Stephan (Higher Education Information System (HIS), Hannover)

The Higher Education Information System (HIS) is the central service provider for institutions of higher education as well as ministries and administrative bodies in the German scientific landscape. HIS supports German institutions of higher education (universities and universities of applied sciences) and their administrations as well as higher education policy-makers in their efforts to fulfil their tasks effectively. The focus of the work is on activities as a software house for higher education administration, within the area of higher education research through empirical studies, and in the area of higher education development with the central topics of higher education organisation, construction and building. One of the major functions of the division "Research on Higher Education" is to complement official statistics on higher education and to provide essential information for higher education policy in Germany by means of empirical social research. In more than 50 research projects more than 100.000 interviews are collected annually. The working processes cover the production of measuring instruments, data collection, data preparation, data analysis and reporting. Beside that datasets are made available for external scientists as public use files, as scientific use files, in terms of remote data processing and by providing HIS internal computer workstations. A wide range of software products is used to support all these processes. The efficiency of the research processes can be improved highly by using DDI. In the current situation the interoperability of the software products is suboptimal. Different file formats are used which are not compatible to each other. Thus high transformation costs accrue. A further problem consists of the fact that text processing files are used for research documentation. One consequence is the limited possibility of automatic content processing. Both problems will be solved by changing the IT-infrastructure to a DDI3 based single source approach. This will be implemented in five steps: Firstly a XML database will be installed to store one DDI instance for each research project. Secondly DDI interfaces will be developed for the software products mentioned. Thirdly a web interface will be provided for the editing and administration of documentation. Fourthly transformation routines will be provided (Java, XSLT) to transform the DDI coded information to various output formats (e.g. a scientific use file documentation). Lastly a web interface will be developed to query the documentation. The first two steps will be completed until the end of 2010.

“DDI Specification: Current Status and Outlook”
Wendy Thomas (Technical Implementation Committee, DDI Alliance), **Joachim Wackerow** (Technical Implementation Committee, DDI Alliance)

This short presentation will give an overview on current development work on the DDI specification and on future steps.

“Ask the Experts”
Wendy Thomas (Technical Implementation Committee, DDI Alliance), **Joachim Wackerow** (Technical Implementation Committee, DDI Alliance)

Questions/answers session, discussion with TIC experts.

“Implementing DDI3.0 : a Case Study of the German Microcensus”
Andias Wira-Alam (GESIS - Leibniz Institute for the Social Sciences), Oliver Hopt (GESIS - Leibniz Institute for the Social Sciences)

This paper shares our experience in developing a software application for the metadata of the German Microcensus on the variable level. First, we developed an editor acting in compliance with the DDI 3.0 standard as the documentation software which improves and simplifies the process of the data documentation. Second, we developed a web information system in order to present the end users with various looks at the metadata. The scope of our work depicts the development cycle of a software application based on DDI 3.0 standard.