3rd Annual European DDI Users Group Meeting (EDDI11)

December 5-6, Gothenburg, Sweden

Book of Abstracts

Final version as of 27. November 2011

Name: **Randy Banks**

Affiliation Institute for Social and Economic Research (ISER)

Understanding Society, Processing Data, Challenging DDI

Understanding Society (UKHLS) is a large-scale household panel study that has been running since 2008. It's total sample size is approximately 40k households, comprised of 5 major subsamples including that migrated from the British Household Panel Survey (BHPS) sample. Interviewing is continuous throughout the year and at anyone point 4 separate surveys are in planning, and up to 4 different interview schedules are in the field.

While similar in many respects to BHPS, the differences between it and Understanding Society were sufficient to require a complete retooling of our data processing operations, starting with the way in which specify questionnaires to our fieldwork contractor, and ending with the means by which we repurpose and document data for the secondary data analysis.

A major requirement of the retooling was to ensure that we both collected and made maximum (re)use of metadata at all points in the operational cycle, and DDI is an integral component in it. Nevertheless, DDI is not the only metadata standard or tool we use and nor have we adopted DDI lifecycle even though in many respects it would be well suited to our needs.

This presentation will discuss the Understanding Society data processing cycle and consider:

- the ways in which metadata is harvested and used
- the role of DDI codebook
- the challenges to the take-up of DDI lifecycle suggested by the Understanding Society experience

---

Name: **Ingo Barkow**

Affiliation DIPF - German institute for international educational research

Co-author(s): David Schiller (IAB - Institute for Employment Research of the German Federal Employment Agency)

Using DDI Lifecycle in Relational Database Systems (RDBs) - Opportunities and Challenges

Though DDI Lifecycle is a standard basing on a XML scheme some agencies decided to represent the model within a relational database. This presentation basing on an upcoming paper to be released at EDDI describes the pros and cons of using a RDB structure for DDI as well as implementation examples and problems. Furthermore also some best practices are featured.

Name: **William Block**

Affiliation CISER - Cornell Institute for Social and Economic Research

The Cornell Census-NSF Research Node: Integrated Research Support, Training and Data Documentation

This presentation will describe a recently-funded $3 million NSF project at Cornell University dedicated to integrating research support, training, and DDI documentation of restricted access data within the United States Census Research Data Center (RDC) network. In addition to a brief overview of the project, this presentation will go into detail on the aspects of the proposed work that rely heavily upon the DDI Lifecycle: 1) a DDI-based metadata schema that will permit synchronization between public and confidential instances of documentation; 2) disclosure avoidance review compliance and improvement; and, 3) a shareable toolkit to be used in the construction of a Comprehensive Census Bureau Metadata Repository (CCBMR).

Name: **Thomas Bosch**

Affiliation GESIS - Leibniz Institute for the Social Sciences

Co-author(s): Brigitte Mathiak (GESIS - Leibniz Institute for the Social Sciences)

A Generic Multilevel Approach for Designing Domain Ontologies based on XML Schemas

Designing an ontology for a specific domain is a time-consuming process. In many cases, information sources like XML Schemas serve as a basis for ontology engineers to conceptualize the intended ontologies. The ontology design process is sped up significantly when XML Schemas are transformed automatically into generated ontologies. An XML Schema Metamodel Ontology has been designed to represent the components of the XML Schema abstract data model. The generated ontologies' classes are defined as sub classes of this ontology. The classes specified for the generated ontologies are intended to be further supplemented with additional semantic and domain specific information defined in domain ontologies. The resulting ontologies are as usable as ontologies that were constructed completely manual, but with a fraction of necessary effort. A complete use case designing a DDI domain ontology using the devised multilevel approach based on already existing XML Schemas will be described in detail.

Name: **Arofan Gregory**

Affiliation   ODaF - Open Data Foundation

SDMX and DDI: How Do They Fit Together in Practical Terms?

This presentation will examine in detail how the DDI standard and the SDMX standard are being implemented together in various institutions. Further, it will explore some of the ideas about how this combined use of the standards might be implemented in the near future. There has been an on-going dialogue between the DDI and SDMX communities, and this work has identified several business cases for how the standards can be combined to support statistical systems, and for use in describing statistical microdata as well. The theoretical business cases and actual on-going and planned implementations are covered.

This presentation will be in a long format, and assumes a general knowledge of DDI and SDMX.

Name: **Marcel Hebing**

Affiliation   SOEP - German Socio-Economic Panel, DIW Berlin

Co-author(s):   William Block (CISER - Cornell Institute for Social and Economic Research)

Developing a Generic Process Model for Longitudinal Data Collections

At the 2011 Dagstuhl Workshop on longitudinal data, one of the working groups focused on producing a reference model for the process of conducting longitudinal and repeat cross-sectional data production. This work was based on the Generic Statistical Business Process Model (GSBPM) produced by METIS, the UN's working group on statistical metadata. (The GSBPM was based in part on the DDI 3.0 combined lifecycle model.)

This talk describes the Generic Longitudinal Business Process Model which came out of this workshop. We will present our approach, building on the high-level lifecycle model which emerged from the preceding year's workshop on longitudinal data. The presentation will focus on an overview of the resulting model, along with a review of the steps and sub-steps used to describe each wave of a longitudinal data collection effort. The model is intended to enable better communication about how processes across longitudinal studies can be compared, and to provide information for each process step for using DDI-Lifecycle. Further, the model helps to identify where metadata is used across the waves of an ongoing data collection.

Name: **Larry Hoyle**

Affiliation   IPSR - Institute for Policy & Social Research, University of Kansas

Co-author(s):   Joachim Wackerow (GESIS - Leibniz Institute for the Social Sciences)

DDI as a Common Format for Export and Import from Statistical Packages

This paper will review the information on metadata commonalities and differences presented in the DDI Working Paper Series -- Use Cases, No. 4 paper: DDI 3 - EXTRACTING METADATA FROM THE DATA ANALYSIS WORKFLOW in the light of the recent inclusion of DDI as an import and export

format in data conversion tools like StatTransfer. One import question addressed will be: which metadata are preserved or lost in round-trips among the major statistical packages.

Name: **Jeremy Iverson**

Affiliation Colectica

Co-author(s): Barry Radler (University of Wisconsin)
Dan Smith (Colectica)

Documenting and Harmonizing MIDUS with DDI 3 and Colectica

Midlife in the United States (MIDUS) is a large, longitudinal study operated by the University of Wisconsin. MIDUS researchers want to provide a comprehensive, canonical source of documentation for the research project. To accomplish this we took the diverse set of sources that previously documented the MIDUS study and created a standardized, DDI 3-based set of documentation that better enables researchers to discover and use the MIDUS data.

This talk will outline the process used to create the DDI 3 documentation, and will demonstrate the resulting documentation and dissemination tools provided by Colectica.

The project is a joint effort between MIDUS and Colectica.

Name: **Jeremy Iverson**

Affiliation Colectica

Co-author(s): Joan Corbett (National Centre for Social Research)
Sophia Kuan (Booz Allen Hamilton)
Abdul Rahim (Metadata Technology)
Wendy Thomas (Minnesota Population Center)

Integrating DDI Metadata into your Research Process

Given the wide scope of the DDI standard, it can be intimidating for those who would like to begin using the specification. This paper provides a simple roadmap for getting started with the standard.

At a basic level, using DDI means that you are able to produce an XML file that validates against the DDI XML schema. This paper shows you how to do that.

The paper should be useful for two groups of people: content-focused people responsible for documenting statistical data and the data lifecycle, and technically-focused people who need to produce the XML.

The paper discusses the basic structure of a statistical study's lifecycle; shows which DDI content can be used to document and drive the processes and output of a study; and provides practical examples of DDI XML.

Name: **Jannik Jensen**

Affiliation   DDA - Danish Data Archive

Future Roadmap for DDI-Lifecycle in DDA

The Danish Data Archive (DDA) will be releasing version 1 of the DDA DdiEditor in the 4th quarter of 2011. Version 2 is planned to be released 4th quarter of 2012.

The presentation will demo version 1 and outline the main development activities for version 2. With a layout of each of the three milestones in version 2 we will be pointing to areas of collaborative work opportunities. Additionally the presentation will touch on DDI-Lifecycle discovery aspects.

DdiEditor is an Open Source project facilitating editing of DDI-Lifecycle, for further information see project homepage: http://www.samfund.dda.dk/dditools/default.htm

---

Name: **Marianne Johnson**

Affiliation   ReTki - Finnish Information Centre for Register Research

Co-author(s):   Irma-Leena Notkola (ReTki - Finnish Information Centre for Register Research)

Documenting Register Data for Research Purposes

In Finland, a federated micro data remote access system for research use of administrative register data is being planned. Administrative data that has been gathered into registers over a long time period make for valuable research data. For researchers to use and interpret the data and for them to take into consideration changes that have happened to the data over time, these registers need to be documented more extensively for this secondary usage than for the primary use. As part of the federated remote access project, plans are to find a common metadata standard for the different administrative registers. We are looking into how DDI Lifecycle answers to our requirements.

Name: **Jack Kneeshaw**

Affiliation   UK Data Archive, University of Essex

Employing DDI at the UK Data Archive - Now and Next

This presentation will describe how the UK Data Archive plans to move its metadata holdings from (broadly) DDI-C to (broadly) DDI-L over the next 12-18 months.

The presentation also focuses on a specific DDI-L use case - serving the UK's Survey Question Bank with DDI-L metadata and the advantages that will bring to users.

Name: **Steven McEachern**

Affiliation   ADA - Australian Data Archive

Co-author(s):   Deborah Mitchell (ADA - Australian Data Archive)
Ben Evans (ANU - Australian National University Supercomputing Facility)
Olaf Delgado-Friedrichs (ANU - Australian National University Supercomputing Facility)

DDI at the Australian Data Archive

The Australian Data Archive has recently been redeveloping a new website content management system to provide new search, browse and visualisation tools, based on DDI metadata managed through Nesstar. This case study presentation will provide an overview of the role of DDI in supporting the new capabilities in the ADA website, tools developed by ADA for creating and managing DDI metadata, and a discussion on our experiences in working with DDI and Nesstar for metadata management.

Name: **Alexander Mühlbauer**

Affiliation   GESIS - Leibniz Institute for the Social Sciences

Creating a Prototype Application Compatible with DDI 3.1 for the STARDAT Project

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 already been developed. However, the challenges of the new DDI versions 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 the project STARDAT to integrate existing DDI tools to one web-based application suite.

The talk will give an overview about the basic architectural ideas and first prototypical implementation results: In particular, I report on our experiences to map DDI 3 to relational database management systems and to create a suitable, object-oriented class model upon that mapping. Furthermore, I present our approach for comprehensive audit logging and DDI versioning mechanisms in a distributed environment. The prototypical implementation supports the shortened use case of editing metadata on study level.

Name: **Olof Olsson**

Affiliation   SND - Swedish National Data Service/Swedish Language Bank

Co-author(s):   Leif-Jöran Olsson (Swedish Language Bank)
Johan Fihn (SND - Swedish National Data Service)

How Can the DDI Community Take Advantage of the Native XML Database eXist-db

This paper will discuss some of the advantages and use cases of the deployment of eXist-db as a DDI storage and application platform. SND has tools in development for eXist-db; a question bank, transformations to codebooks and web pages from DDI Lifecycle and DDI Codebook, web services for searching and resolving resources in the collection, as well as harvesting of metadata from external resources using OAI-PMH. We will present a demo of xForms generation for the DDI-model inside eXis

and also present some news for eXist-db regarding the release of 1.4.2, the upcoming versions and the new indexing of binary resources like PDF documents.

Name:	**Hilde Orten**
Affiliation	NSD - Norwegian Social Science Data Services
Co-author(s):	Joachim Wackerow (GESIS-Leibniz Institute for the Social Sciences)

The ISCED Standard for Coding of Education and DDI-Lifecycle

Education is a core explanatory factor social research. Educational attainment is, however, difficult to measure cross-nationally, since educational programmes and qualifications vary to a large extent between countries, and educational reforms take place over time. An increasingly common approach is to measure educational attainment by country-specific questions where the resulting variables are coded into an international standard, of which UNESCO's ISCED (International Standard for Coding of Education) is the most advanced.

This presentation gives a brief introduction to ISCED 2011 and its application in cross-national surveys, and point at challenges related to documentation of survey variables based on ISCED. The presentation will have a main focus on the advantages of including ISCED metadata elements in a DDI resource package, which would allow for reuse of metadata by reference by official and non-official survey programmes and other data producers.

Name:	**Marco Pellegrino**
Affiliation	Eurostat

The Ongoing Work for a Technical Vocabulary of DDI and SDMX Terms

The SDMX Technical Vocabulary project, run by Eurostat, is documenting the specific terminology used in the SDMX 2.1 technical specifications (e.g. data structure definition, concept scheme, maintainable artefact, or constraint). Eurostat is discussing the first draft of the SDMX Technical Vocabulary with the SDMX Secretariat and its technical working group. A second iteration is foreseen before delivering the first release by Spring 2012. The second step will be, then, an integrated DDI-SDMX Vocabulary which aims at documenting both SDMX and DDI technical terminologies, in order to ensure a better knowledge and a possible mapping between the two standards. This work is a result of several meetings of the SDMX DDI Dialogue. The structure and content of the Vocabulary is discussed during working groups on technical issues and international work sessions such as UNECE/Eurostat/OECD/METIS. The tentative schedule for delivering the first DDI-SDMX Vocabulary is mid-2012.

Name: **Nicole Quitzsch**

Affiliation   GESIS-Leibniz Institute for the Social Sciences

Co-author(s):   Erdal Baran (GESIS-Leibniz Institute for the Social Sciences)

The Registration Agency da|ra, DDI and Linked Open Data

Our presentation is divided into two parts. It begins with a general introduction regarding the development of the da|ra metadata schema 2.2 and a brief overview of the main fields of the schema. The first part deals with the DDI 3 compliant standard used for the documentation of research data which includes general as well as detailed information on a study such as temporal coverage, collection mode, related publications, etc. Furthermore, the mapping of the Version 2.2 of the metadata schema to the DDI 3 standard demonstrating the compliance of the Version 2.2 to the requirements of the DDI standard will be explained.

Another focus of the presentation is on the set up of the connection of da|ra Metadata to the Linked Open Data (LOD) - cloud by using an automated ontology generating tool based on the DDI Schema. Particularly with regard to the Linked Open Data network the approach is to publish data and metadata in form of a standard based exchange format like the Resource Description Framework (RDF). An ontology has to be built based on the conceptual model of DDI 3. This ontology should encompass the most relevant DDI 3 components, not all available elements of DDI 3. Initially defined generated ontologies are linked to an ontology of the appropriate domain used to specify supplementary semantic information not covered in the XML Schemas of the named Metadata standards. To take the advantage of the Linked Open Data - cloud to create ontology-based Metadata the convertion of the da|ra Metadata schema to the DDI 3 compliant schema is required.

Name: **Samuel Spencer**

Affiliation   Open-source developer

Single-serve Software - Meeting the Needs of a New Generation of Users

The DDI information model promotes evidenced-based research through management of an exhaustive array of social science and statistical metadata. However, this is can only be achieved through software tools capable of meeting users need to find, access and manipulate metadata across the lifecycle. These needs are complicated by users changing desires for how they interact with software. Recent shifts in mobile development have altered users expectations from complex, all-encompassing tools towards simpler, more task-oriented apps. The challenge comes from ensuring the DDI community can support task-oriented wants while still meeting their needs of easy access to the breadth of metadata available.

By examining the benefits to users, developers and the wider community, this talk aims to recommend working towards smaller, more task-focused software to support users requirements, while simultaneously encouraging sustainable, ongoing development. This is followed by an examination of the use of DDI profiles as way to manage user expectations and encourage diversity in the DDI developer community. These recommendations are followed with a practical example - Virgil UI: a DDI 3.1 compliant classification and codelist editor - to demonstrate the advantages of task-oriented development.

Name:	**Karsten Stephan**
Affiliation	HIS-HF - (HIS-HF - Higher Education Information System / Institute for Research on Higher Education)
Co-author(s):	Daniel Buck (HIS-HF)

Challenges to Implement DDI for the Whole Lifecycle in a Social Science Research Institution

In the HIS-Institute for Research on Higher Education (HIS-HF Institut für Hochschulforschung), you can find the complete research and data lifecycle. The institute founded in 1969 has a grown organizational structure and uses some in-house developed applications. Now the institute is aiming for the organizations' further development first by setting up a data research center and secondly as a long-term objective by implementing DDI 3 in the whole information infrastructure. From todays' perspective the DDI implementation process is planned to follow these steps: identification of required DDI elements and modules regarding our research processes, data and documentation standard, programming of export/import routines for yet existing applications, set up of a DDI data repository, the development of a data and system management application as well as a tool to search through the metadata. The outcome should be improvements in documentation standard, working processes and interoperability within the institute and potentially with external users and applications. First plans had been presented on the EDDI 1 in 2009.

In 2009 we started our first steps in DDI by initiating two open source software development projects that were planned to be based on DDI 3: first implementing DDI 3 import-/export routines for "HIS-Code", an existing tool to support coding of open questions in surveys, and secondly the development of HIOB, an online survey software. In the presentation we would like to focus on the challenges in the application of DDI. One of them was the implementation of DDI for the highly complex data and processes in the online survey process. Another challenge was that some parts of DDI were still work in progress. In the end we even had to skip the plan to use DDI in HIOB due to financial and temporal restraints. But still we plan to program DDI import/export routines and enable HIOB to communicate with other DDI applications. We stick to the objective to use DDI for the whole data lifecycle – though after two years this seems to be far more challenging than we thought before.

Name:	**Maarten Streefkerk**
Affiliation	CentERdata
Co-author(s):	Suzan Elshout (CentERdata)
	Alerk Amin (CentERdata)
	Edwin de Vet (CentERdata)

Dissemination of Survey (Meta-)Data Using Questasy

This presentation will relate the new developments in Questasy since the 2010 EDDI. During presentation, we will show support for variable baskets, multiple panels, resource packages, and more.

Questasy is a data dissemination website application based on DDI 3. It was primarily developed for the LISS Data Archive (www.lissdata.nl) to disseminate both data and metadata, but is freely available for other organizations.

With Questasy, one can easily enter survey (meta)data i.e. questions, variable information, data collection events and concepts. Using the LISS data archive one can download the survey data and easily browse topics and concepts.

Name: **Bo Sundgren**

Affiliation Stockholm University

Communicating in Time and Space - How to Overcome Incompatible Frames of Reference of Producers and Users of Archival Data

When we put data into an archive, we do not know very much about who will be the future users of these data. We may know a lot about contemporary users, especially if we adopt the modern view of an archive as an integrated part of the information system of an organisation (or a society), consisting of data generated by the organisation's (or society's) processes, and providing feedback to these processes. But at the other end of the spectrum, we have future researchers living centuries away from now, in a society which will almost certainly be very different from ours.

Name: **Marion Wittenberg**

Affiliation DANS - Data Archiving and Networked Services
Co-author(s): Merja Karjalainen (SND - Swedish National Data Service)

Structured Metadata for Reuse

This paper focuses on the reuse of metadata in the case where references are made from one agency's metadata to another agency's metadata.

Generally, there are benefits of reusing metadata including better support for comparability between different studies and facilitating the management of references to classifications and other resources. The reuse of metadata also has many advantages within studies e.g. for longitudinal studies, between different waves. This paper especially addresses the referencing of de facto metadata that are agreed as standards, here called Foundational Metadata.

Name: **Marion Wittenberg**

Affiliation DANS - Data Archiving and Networked Services

Question Bank Cultural Changes in the Netherlands

The survey Cultural Changes in the Netherlands is a bi-annual survey from 1975 on, to measure the attitudes of the Dutch population to social, ideological and political issues. The Netherlands Institute for Social Research (SCP), a government agency which conducts this research, maintains the question wordings in a local question database. In the project Question bank Cultural Changes we want to move the longitudinal information of this Study into DDI Life cycle. For this we make use of Questacy, the online data and metadata management application of Centerdata. This presentation will focus on the issues involved in the conversion of this collection metadata into DDI3 and the new features of Questacy we want to develop.