



4th Annual European DDI User Conference (EDDI12)

December 3-4, 2012, Bergen, Norway

<http://www.eddi-conferences.eu/eddi12>

Schedule and Program with Abstracts

Final version as of 15. January 2013 with links to presentations

The conference will open on Monday, Dec 3rd at 9:00 and close on Tuesday, Dec 4th at 17:00.

| | |
|--|----|
| Schedule | 2 |
| <i>Plenary</i> : Keynote Speech..... | 7 |
| <i>Plenary</i> : The Generic Statistical Information Model (GSIM) and DDI - A Collaborative Relationship | 8 |
| Presentations in Concurrent Sessions | 9 |
| Posters / Demonstrations..... | 18 |
| Concurrent Tutorials..... | 21 |
| Side Meetings | 23 |

Schedule

Venue

| | |
|--|--|
| Main venue | Scandic Hotel Bergen City, Hakonsgaten 2 |
| Special meetings and some tutorials | NSD – Norwegian Social Science Data Services, Harald Hårfagres gate 29 |

Wednesday to Friday, November 28-30, 2012

| | |
|--------------|--|
| 9:00 - 17:00 | <p style="text-align: center;">Special Meetings</p> <p style="text-align: center;">M1: Modelling Qualitative Metadata Structures and the DDI (description)</p> <p style="text-align: center;">Location: NSD – 3rd floor, room 344</p> <p style="text-align: center;">Chair: Louise Corti</p> <hr/> <p style="text-align: center;">DDI Alliance Qualitative Data Exchange Working Group</p> |
|--------------|--|

Sunday, December 2, 2012

| | |
|-------|---|
| 18:00 | <p style="text-align: center;">Informal Get-together</p> <p style="text-align: center;">Location: Cafe Opera, Engen 18</p> |
|-------|---|

Monday to Tuesday, December 3-4, 2012

| | |
|---------------|---|
| 9:00 - 17:00 | <p style="text-align: center;">Conference Sessions</p> <p style="text-align: center;"><i>Details are below</i></p> |
| 18:50 - 20:00 | <p style="text-align: center;">Reception by the City Council of Bergen in Schøtstuene, Øvregaten 50 (map)</p> |

Wednesday to Thursday, December 5-6, 2012

| | |
|--------------|--|
| 9:00 - 17:00 | <p style="text-align: center;">Special Meetings</p> <p style="text-align: center;">M2: Software Development (description)</p> <p style="text-align: center;">Location: NSD – ground floor, room 104</p> <p style="text-align: center;">Chair: Johan Fihn</p> <hr/> <p style="text-align: center;">DDI Developers Meeting</p> |
|--------------|--|

Monday, December 3, 2012

| | | | | |
|---------------|---|---|--|---|
| 8:15 | Starting Registration in Scandic | | | |
| 9:00 - 12:30 | Concurrent Conference Tutorials | | | |
| | Concurrent T1: Introduction to DDI with a Business Perspective Location: Scandic – 3rd floor, room Muséplass Chair: Oliver Hopt | Concurrent T2: Introduction to DDI with a Technical Perspective Location: Scandic – 2nd floor, room Sydneshaugen Chair: Olof Olsson | Concurrent T3: Using DDI Software Location: NSD – ground floor, room 104 Chair: Hilde Orten | Concurrent T4: Building DDI Applications Location: NSD – 3rd floor, room 344 Chair: Wolfgang Zenk-Möltgen |
| | Data and Metadata Management Using DDI (abstract, presentation) Wendy Thomas (Minnesota Population Center), Marcel Hebing (DIW - German Institute for Economic Research) | Create your own information systems on the basis of DDI-Lifecycle (abstract, presentation) Thomas Bosch, Matthäus Zloch (both GESIS - Leibniz Institute for the Social Sciences) | Hands on DDI 3 with Colectica (abstract) Jeremy Iverson, Dan Smith (both Colectica) | Building Applications with the Nesstar API (abstract, presentation) Archana Bidargaddi and Nesstar Development Team (all NSD - Norwegian Social Science Data Services) |
| 12:30 - 14:00 | Lunch in Scandic Restaurant on ground floor | | | |

Monday, December 3, 2012 (cont.)

| | | |
|---------------|---|--|
| 14:00 - 15:30 | Conference Plenary 1: Welcome and Keynote Location: Scandic – 2nd floor, room Sydneshaugen Chair: Joachim Wackerow (EDDI program committee) | |
| | Welcome Bjørn Henrichsen (executive director NSD) | |
| | Keynote: Ideas about Statistical Data Archives and Services (presentation) Svein Nordbotten (University of Bergen) | |
| 15:30 - 16:00 | Break in foyer on ground floor | |
| 16:00 - 18:00 | Concurrent Conference Sessions | |
| | Concurrent A1: DDI Usage in National Statistical Institutes and with Census Data Location: Scandic – 2nd floor, room Sydneshaugen Chair: Arofan Gregory | Concurrent A2: Facilitating Sharing, Integration, and Discovery Location: Scandic – 3rd floor, room Muséplass Chair: Johanna Vompras |
| | DDI-L in the Production of Official Statistics abstract , presentation Mogens Grosen Nielsen (Statistics Denmark), Jannik Jensen (Danish Data Archive) | Sharing Category Systems for Survey Items with Others in DDI-Lifecycle Format (abstract , presentation) Wolfgang Zenk-Möltgen (GESIS - Leibniz Institute for the Social Sciences) |
| | Powering Official Statistics at Statistics New Zealand with DDI-L and Colectica: A Case Study abstract , presentation Adam Brown (Statistics New Zealand), Jeremy Iverson (Colectica), Dan Smith (Colectica), Sally Vermaaten (Statistics New Zealand) | DDI Resource Packages and How They Can Be Used in the Data Production Process (abstract , presentation) Hilde Orten (NSD - Norwegian Social Science Data Services), Joachim Wackerow (GESIS - Leibniz Institute for the Social Sciences) |
| | An Early Prototype of the Comprehensive Extensible Data Documentation and Access Repository (CED²AR) (abstract , presentation) William Block, Jeremy Williams, John Abowd, Lars Vilhuber, Carl Lagoze (all Cornell University) | REDCap and DDI (abstract , presentation) Larry Hoyle (Institute for Policy & Social Research, University of Kansas) |
| | | Semantic Categorization of DDI Metadata abstract , presentation Vasily Bunakov (Science and Technology Facilities Council, United Kingdom) |
| 18:20 | Group walk to reception venue from Scandic hotel | |
| 18:50 - 20:00 | Reception by the City Council of Bergen in Schøtstuene, Øvregaten 50 (map) | |
| 20:00 | Informal Get-together Location: Dickens , Kong Olav V's Plass 4 | |

Tuesday, December 4, 2012

| | | |
|---|---|--|
| 9:00 - 10:00 | Conference Plenary 2: The Generic Statistical Information Model (GSIM) and DDI – A Collaborative Relationship Location: Scandic – 2nd floor, room Sydneshaugen / Chair: Larry Hoyle | |
| | Generic Statistical Information Model – Overview and Introduction (abstract , presentation) Thérèse Lalor, Steven Vale (both UNECE - United Nations Economic Commission for Europe) | |
| | DDI and GSIM – Impacts, Context, and Future Possibilities (abstract , presentation) Arofan Gregory (Metadata Technology) | |
| 10:00 - 10:30 | Break in foyer on ground floor | |
| 10:30 - 12:30 | Concurrent Conference Sessions | |
| | Concurrent B1: Infrastructure for Data Collection, Research, and Archiving Location: Scandic – 2nd floor, room Sydneshaugen / Chair: David Schiller | Concurrent B2: Software Solutions Location: Scandic – 3rd floor, room Muséplass / Chair: Johan Fihn |
| | 70 Years of UK Birth Cohort Data into DDI Lifecycle? (abstract , presentation) Claude Gierl, Jon Johnson (both Centre for Longitudinal Studies Institute of Education London) | Modeling the Lifecycle into a Combination of Tools (abstract , presentation) Ingo Barkow (DIPF - German Institute for International Educational Research) |
| | Issues around Implementing DDI in a Survey Organisation (abstract , presentation) Joan Corbett (ScotCen Social Research), Roger Stafford (NatCen Social Research) | Implementation of the DdiEditor in DDA (abstract , presentation) Anne Sofie Fink (DDA - Danish Data Archive) |
| | DDI-Lifecycle Infrastructure at DDA (abstract , presentation) Jannik Jensen (DDA - Danish Data Archive) | Managing Private and Public Views of DDI Metadata Repositories (abstract , presentation) Dan Smith (Colectica) |
| | RODA Infrastructure for Data and Metadata (abstract , presentation) Cosmin Rentea (RODA - Romanian Social Data Archive) | Nesstar – from Tool to Toolkit: A Programmable DDI-based Platform (abstract , presentation) Archana Bidargaddi, Ricco Førgaard, Ørnulf Risnes (all NSD - Norwegian Social Science Data Services) |
| | Data Management, DDI-based Documentation, and Visualization of Business and Organizational Research Data at the DSZ-BO (abstract , presentation) Johanna Vompras (University of Bielefeld - University Library) | |
| 12:00 - 14:30 | Poster / Demonstration Session Location: Scandic – in foyer on ground floor | |
| simultaneously with lunch, and partly with previous and next sessions | Colectica: Data Management with DDI 3 (abstract) Jeremy Iverson, Dan Smith (both Colectica) | DdiXslt (abstract , poster) Jannik Jensen (Danish Data Archive) |
| | DdiEditor (abstract , poster) Jannik Jensen (Danish Data Archive) | The Generic Statistical Information Model (GSIM) (abstract , poster) Thérèse Lalor (UNECE - United Nations Economic Commission for Europe) |
| | Nesstar Toolkit (abstract) Ørnulf Risnes (NSD - Norwegian Social Science Data Services) | Towards a DDI Backend Architecture - Sharing DDI-related Software Modules (abstract , poster) Matthäus Zloch, Thomas Bosch, Dennis Wegener (all GESIS - Leibniz Institute for the Social Sciences) |
| 12:30 - 14:00 | Lunch in Scandic Restaurant on ground floor | |

Tuesday, December 4, 2012 (cont.)

| | | |
|---------------|--|--|
| 14:00 - 15:15 | Concurrent Conference Sessions | |
| | <p>Concurrent C1: Metadata within the Project Data without Boundaries (DwB) Supporting a European Service Center for Researchers</p> <p>Location: Scandic – 2nd floor, room Sydneshaugen Chair: Nikos Askitas</p> | <p>Concurrent C2: Answers by Questasy</p> <p>Location: Scandic – 3rd floor, room Musépllass Chair: Ingo Barkow</p> |
| | <p>Data without Boundaries: A European Project to Enhance Access to Official Microdata (abstract, presentation)</p> <p>Roxane Silberman (Réseau Quetelet), Christof Wolf (GESIS - Leibniz Institute for the Social Sciences)</p> | <p>Update on Questasy, a Data Dissemination Tool Based on DDI3 (abstract, presentation)</p> <p>Edwin de Vet (CentERdata)</p> |
| | <p>Resource Discovery for Official Statistics (abstract, presentation)</p> <p>Arofan Gregory (Metadata Technology), Marion Wittenberg (DANS - Data Archiving and Networked Services)</p> | <p>Implementing a Longitudinal Data Portal with a Customized Version of Questasy (abstract, presentation)</p> <p>Toni Erik Sissala (FSD - Finnish Social Science Data Archive)</p> |
| | <p>CIMES: A Tool for Describing European Official Statistics Microdata (abstract, presentation)</p> <p>Raphaëlle Fleureux, Cyril Jayet, Arnaud Szulek (all Réseau Quetelet)</p> | <p>Question Bank Cultural Changes in the Netherlands, Moving from DDI Codebook to DDI Lifecycle (abstract, presentation)</p> <p>Marion Wittenberg (DANS - Data Archiving and Networked Services), Edwin de Vet (Centerdata)</p> |
| 15:15 - 15:45 | Break in foyer on ground floor | |
| 15:30 - 17:00 | Conference Plenary 3: Current Status | |
| | <p>Location: Scandic – 2nd floor, room Sydneshaugen Chair: William Block</p> | |
| | <p>DDI Lifecycle: Moving Forward – Outcome of the Recent Workshop in Dagstuhl (presentation)</p> <p>Joachim Wackerow (GESIS - Leibniz Institute for the Social Sciences)</p> | |
| | <p>DDI Specification: Current Status and Outlook (presentation)</p> <p>Wendy Thomas (Technical Implementation Committee, DDI Alliance)</p> | |
| | <p>Questions and Grumbles? Answers from TIC</p> <p>Arofan Gregory, Wendy Thomas, and Joachim Wackerow (all Technical Implementation Committee, DDI Alliance)</p> | |
| | <p>Announcement of 1. NADDI, Invitation to EDDI13 and Goodbye (presentation, presentation)</p> <p>Larry Hoyle (NADDI team), Nikos Askitas, Joachim Wackerow (both EDDI core team), and Next Year's Host (name will be disclosed in session)</p> | |
| 18:00 | Informal Get-together | |
| | <p>Location: Kafe Kippers, Georgernes Verft 12</p> | |

Plenary: Keynote Speech

Ideas about Statistical Data Archives and Services

Svein Nordbotten (University of Bergen)

Half a century ago, the emerging electronic technology and the implementation of central registers initiated new visions about official data collection and statistical services. Organized data archives were proposed as a base for preparation of statistics and research. The proposals were followed by a debate on the risk of data misuse and the introduction of privacy legislations. The visions were gradually realized during the following years by implementation of statistical data bases, register-based statistics and researchers' access to microdata. So far, access to statistical information has mainly been limited to standard publications and retrieval services from immobile devices. Most people already carry their cell phones 24/7. They will soon be able to get/check statistical facts by a few clicks on their smart phones independent of time and location. Electronic footprints represent a rapidly growing and potential data source to be included in official data collection, organized and used by the statistical services. But extended statistical services will also create new challenges as to preserving statistical confidentiality.

Plenary: The Generic Statistical Information Model (GSIM) and DDI - A Collaborative Relationship

Generic Statistical Information Model – Overview and Introduction

Thérèse Lalor, Steven Vale (both UNECE - United Nations Economic Commission for Europe)

Across the world statistical organizations undertake similar activities albeit with variation in the processes each uses. Each of these activities use and produce similar information (for example all agencies use classifications, create data sets and publish products). Although the information used by statistical organizations is at its core the same, all organizations tend to describe this information slightly differently (and often in different ways within each organization). There is no common means to describe the information we use. GSIM is a conceptual model that provides a set of standardized, consistently described information objects, which are the inputs and outputs in the design and production of statistics.

GSIM must be implementable. In order to support the implementation of GSIM, known standards and tools have also been examined, to ensure that the reference framework is complete and useful in this respect. The relationship between GSIM and other models and standards is two-fold. The standards and models serve as inputs to the creation of GSIM, and also act as targets for the use of GSIM within organizations.

This presentation will inform participants on the latest status of this important standardisation work.

DDI and GSIM – Impacts, Context, and Future Possibilities

Arofan Gregory (Metadata Technology)

The Generic Statistical Information Model (GSIM) represents the agreed information requirements for national statistical production, and DDI is a key standard in both the development of GSIM itself, and as an implementation tool for organizations using GSIM. Beyond that, it also will influence the future directions of DDI development, attracting a larger number of data producers into the DDI community. This presentation looks at the interaction between GSIM and DDI (and other related standards), and provides an update on a rapidly-evolving vision around the use of DDI within the statistical institutes in Europe and elsewhere. With closer cooperation emerging between data archives and statistical institutes, the adoption of DDI among data producers promises to help realize standard, up-stream data documentation coming into the archives. This presentation will cover both the direct interaction between DDI and GSIM, and also provide a broader context for understanding what that dynamic may mean in the future.

Presentations in Concurrent Sessions

Modeling the Lifecycle into a Combination of Tools

Ingo Barkow (DIPF - German Institute for International Educational Research)

During the last years several different tools for DDI Lifecycle have been published. Nevertheless none of the current tools is able to cover the full lifecycle from beginning to end. This presentation wants to show how a survey process from creating a study from the scratch, designing the instruments, performing the data collection, handling the administrative processes, curating the data, disseminating the data, publication and at last data archiving for secondary usage could be handled with individual tools. Next to well-known programs like Colectica or Questasy this presentation will also introduce a first outlook into Rogatus QMMS - an open source toolset currently in development at DIPF with support of GESIS, TBA21 and OPIT. Rogatus consists of different DDI compliant applications (e.g. Questionnaire Builder, Translation Builder, Metadata Builder, Rogatus Portal) to support a multitude of survey processes.

Nesstar - from Tool to Toolkit: A Programmable DDI-based Platform

Archana Bidargaddi, Ricco Førgaard, Ørnulf Risnes (all NSD - Norwegian Social Science Data Services)

For many years, Nesstar has been developed and released as an integrated tool with applications for editing metadata, publishing, data analysis and visualization.

An integrated tool can serve many purposes, and NSD will continue to release Nesstar as an integrated solution. However, there is an increasing demand from the user-community (including national and international initiatives and infrastructure projects) for building alternative applications for data publishing, dissemination and analysis, and to integrate Nesstar with in-house production systems.

Responding to these changes in the data and metadata landscapes, NSD started developing a Java based Nesstar API in 2012. Version 0.6 of the API will be released in December 2012, and will support all functionalities offered by the underlying Nesstar 4.x-architecture, including:

- Publishing/uploading
- Metadata harvesting and -repurposing
- Microdata analysis (cross-tabulation, correlation, regression)
- Downloading
- Searching
- Cubes/aggregate data
- Content structuring and User management

This talk will demonstrate Nesstar as a programmable DDI-based dissemination platform, and present applications that have been built with the API. Furthermore we will discuss the future plans for the Nesstar API and Nesstar, and give an updated view on NSD/Nesstar's roadmap regarding DDI-C/2.5 and DDI-Lifecycle and other data and metadata related developments and initiatives.

An Early Prototype of the Comprehensive Extensible Data Documentation and Access Repository (CED²AR)

William Block, Jeremy Williams, John Abowd, Lars Vilhuber, Carl Lagoze (all Cornell University)

This presentation will demonstrate the latest DDI-related technological developments of Cornell University's \$3 million NSF-Census Research Network (NCRN) award, dedicated to improving the documentation, discoverability, and accessibility of public and restricted data from the federal statistical system in the United States. The current internal name for our DDI-based system is the Comprehensive Extensible Data Documentation and Access Repository (CED²AR).

CED²AR ingests metadata from heterogeneous sources and supports filtered synchronization between restricted and public metadata holdings. Currently-supported CED²AR "connector workflows" include mechanisms to ingest IPUMS, zero-observation files from the American Community Survey (DDI 2.1), and SIPP Synthetic Beta (DDI 1.2). These disparate metadata sources are all transformed into a DDI 2.5 compliant form and stored in a single repository. In addition, we will demonstrate an extension to DDI 2.5 that allows for the labeling of elements within the schema to indicate confidentiality. This metadata can then be filtered, allowing the creation of derived public use metadata from an original confidential source. This repository is currently searchable online through a prototype application demonstrating the ability to search across previously heterogeneous metadata sources.

Powering Official Statistics at Statistics New Zealand with DDI-L and Colectica: A Case Study

Adam Brown (Statistics New Zealand), Jeremy Iverson (Colectica), Dan Smith (Colectica), Sally Vermaaten (Statistics New Zealand)

Statistics New Zealand is the national statistical office of New Zealand, the country's major source of official statistics, and leader of New Zealand's Official Statistics System. Statistics New Zealand's goal is to give New Zealand the statistical information it needs to grow and prosper.

Statistics New Zealand has adopted DDI-L to document its statistical products. For the past year, the Statistics New Zealand has been using Colectica to build its standards-based metadata infrastructure.

This presentation will highlight how Statistics New Zealand is using DDI and Colectica to document and drive its statistical production processes. It will look at the software components, metadata standards, and workflows used at Statistics New Zealand, and will review the accomplishments made and future plans.

Semantic Categorization of DDI Metadata

Vasily Bunakov (Science and Technology Facilities Council, United Kingdom)

ENGAGE project (<http://www.engage-project.eu>) aims at supplying the open data infrastructure and is largely oriented on the data re-use by research communities, data journalists, and other user categories. ENGAGE has identified DDI as an important framework to cater for the ENGAGE user needs, specifically for researchers in social science. One of the fundamental use case scenarios in ENGAGE is data discovery for which Linked Data is an important means, and linking to DDI metadata is a gateway to the actual data.

The work presents a modelling and implementation effort in semantic categorization of DDI metadata samples against a bespoke ontology. The categorization output is then linked with a few common ontologies and semantic data sources that are of interest to ENGAGE. The potential for linking open DDI data with other open data is emphasized.

Issues around Implementing DDI in a Survey Organisation

Joan Corbett (ScotCen Social Research), Roger Stafford (NatCen Social Research)

NatCen Social Research is an independent social research institute in the UK. Research datasets for secondary analysis are one of the most important research outputs that we produce. This requires a substantial amount of documentation, which is usually collated at the end of the project as part of the archiving process.

Whilst we see the benefits of DDI the key for adoption is finding the correct software solution. It has to be adaptable to all our surveys, compatible with the software we currently use and create minimum disruption to our current work. We presently use Blaise as software for computer assisted interviewing and SPSS for data management and analysis. Software with DDI outputs such as MQDS offers integration with Blaise and Questasy integrates with SPSS but we are not sure whether they entirely match our needs. We are aware that other organisations have often had to utilise the DDI by creating or commissioning a bespoke system, a major commitment for any organisation. We are exploring what we can do with our existing software and welcome the opportunity that the EDDI conference offers to discuss these issues and to find out what colleagues elsewhere are doing.

Implementation of the DdiEditor in DDA

Anne Sofie Fink (DDA - Danish Data Archive)

This year DDA has implemented the DdiEditor for data curation processing. This has significantly changed work processes for ingest and migration. The presentation will focus on new and future developments and the lessons we have learned from continuous delivery and integration of new software and user feedback. Further information see project homepage: <http://code.google.com/p/ddieditor>.

CIMES: A Tool for Describing European Official Statistics Microdata

(Part of session on Data without Boundaries)

Raphaëlle Fleureux, Cyril Jayet, Arnaud Szulek (all Réseau Quetelet)

This paper aims at presenting the documentation of European Official Statistics Microdata thanks to the CIMES software, recently developed under the Work Package 5 of the EU-funded FP7 project DwB (Data without Boundaries). In particular, we will give an insight into the challenges that had to be overcome when processing this documentation and using DDI. First, we will describe the metadata schema and the database model which will allow importing and exporting in DDI2-C and in DDI3-L. Then, we will present the web-based tool CIMES, developed to document the existing data on the basis of both this schema and model. Finally, we will give an overview of the current state of play and expected achievements of the documentation work carried out under DwB.

70 Years of UK Birth Cohort Data into DDI Lifecycle?

Claude Gierl, Jon Johnson (both Centre for Longitudinal Studies Institute of Education London)

CLOSER (Cohorts and Longitudinal Studies Enhancement Resources) is a five year programme which aims to maximise the use, value and impact of these studies both within the UK and abroad. The programme is run by a network of nine of the UK's leading studies (8 birth cohorts and one Panel study), with participants born between 1911 and 2007.

A major strand will be documenting these surveys and data (over 250 survey instruments and around 250,000 data variables) in DDI-L. The surveys cover a wide range of survey collection methods, paper questionnaires to CAI, biomedical and administrative linked data. The presentation will cover the Centre for Longitudinal Studies experience so far of meta-data creation and management using a variety of software tools.

This includes an in-house application for questionnaire capture written in Ruby on Rails, Python based data quality tools to interface with SPSS and Colectica for overall data management.

The presentation will also highlight how some recent changes to the DDI specification will assist in the management of these projects.

Resource Discovery for Official Statistics

(Part of session on Data without Boundaries)

Arofan Gregory (Metadata Technology), Marion Wittenberg (DANS - Data Archiving and Networked Services)

The aim of Work Package 8 of the Data without Boundaries project is to propose portal resource discovery functionality for a pan-European search and browse portal for Official Statistics.

In the first part of this presentation we look at the requirements for this portal and the functionality. What does the user want of such a portal? What are the constraints in metadata collection from archives and NSI's?

The second half of the presentation will focus at the DwB metadata model and its relationship to DDI and other standards. The strategies for collecting metadata from data archives and statistical institutes will also be presented, which will be implemented in Work Package 12 as part of the DwB portal prototype.

REDCap and DDI

Larry Hoyle (Institute for Policy & Social Research, University of Kansas)

The REDCap (Research Electronic Data Capture) consortium is a group of over 450 institutions that supports a web application which supports data capture for research studies (see <http://project-redcap.org/>). The application allows interactive survey instrument development and data collection. Data and scripts for SPSS, SAS and R can be exported from REDCap. Survey metadata including question text and flow control can be also exported as a csv file.

This paper describes code in the R language to convert the REDCap survey metadata from csv to DDI 3.1. It includes a discussion of which REDCap instrument attributes can be represented by DDI 3.1 elements other than Note.

DDI-Lifecycle Infrastructure at DDA

Jannik Jensen (DDA - Danish Data Archive)

The Danish Data Archive (DDA) is moving a collection of app. 2000 survey data sets into DDI-Lifecycle in the fourth quarter of 2012. The presentation will focus on the challenges of this process and present implemented research infrastructure –Indexing Platform –around the collection of data sets annotated in DDI-L. The presentation will demo discovery mechanisms and DDI-L URN functionality.

DDI-L in the Production of Official Statistics

Mogens Grosen Nielsen (Statistics Denmark), Jannik Jensen (Danish Data Archive)

Statistical organizations disseminate statistics to an extent never seen before. However according to in-depth analysis of user needs, it is an urgent task to give end-users better assistance when they use statistics or wish to find relevant statistics. The paper suggests that metadata should be defined and implemented targeted at providing help to end-users.

Statistics Denmark strives for an integrated metadata approach with quality-, concepts-, variables- and classification-elements. To fulfill this, standards and tools used in Sweden, Portugal, Canada etc. were investigated. It was apparent that all solutions could be used to fulfill our needs, but still a lot of work on development was required. In 2011 DDI-L was discovered. The standard looked like as an excellent way forward. Since then Statistics Denmark, together with Danish Data Archive as reviewer, has carried out a pilot-project and other tests of the standard using various software implementations - Colectica being the main tool.

The paper will present models, DDI-L structures and issues and thereby contribute to a common understanding of the use of DDI-L in the production of official statics that would benefit both the

international DDI-community and the international statistical community. We hope this paper can contribute to this process.

DDI Resource Packages and How They Can Be Used in the Data Production Process

Hilde Orten (NSD - Norwegian Social Science Data Services), Joachim Wackerow (GESIS - Leibniz Institute for the Social Sciences)

DDI Resource Packages structure study independent metadata that are intended for reuse in a project or department, institution, network of institutions and in public. Typical examples of metadata that usefully can be structured in DDI Resource packages are standard classifications, high quality survey questions and variables etc.

The following presentation first introduces DDI Resource Packages at a conceptual level. Two examples of possible Resource Packages will then be shortly introduced. The first example is of a Resource Package for the Standard Classification of Education (ISCED), for which work is ongoing with the aim of realising a real example. The second example represents some ideas for a Resource Package for some Background Variables from the International Social Survey programme (ISSP). Based on these examples, the presentation will show how a DDI Resource Package can be useful in the day-to-day work throughout the data production process of an institution or project, and also point at some tools using the Resource Package that could be of help. A set of guidelines for developing and managing a Resource Package will be outlined in the end.

RODA Infrastructure for Data and Metadata

Cosmin Rentea (RODA - Romanian Social Data Archive)

In the following years, RODA (Romanian Social Data Archive) plans to enhance its capabilities in the field of social data archiving and information management.

We propose an architecture for handling both new DDI 3.x and legacy DDI 2.x metadata, alongside other statistical software formats for data, while improving and maintaining RODA's existing RDBMS schemas. The web application will be available online and will have multiple access levels, treating more stages in the metadata and data life-cycle: allowing meta-data editing, metadata and data importing and exporting, complex search queries, data processing and archiving, statistical analysis and visualization. An architectural overview and preliminary results will be presented in order to obtain feedback and stimulate debate in the growing DDI community.

RODA has the goal of becoming a full-fledged member of CESSDA (Council of European Social Science Data Archives) by focusing on interoperability with other members – both technical (e.g. using a common authentication system, compatible authorization systems) and semantic (e.g. controlled vocabularies, question pools), while respecting national regulations and individual constraints.

Data without Boundaries: A European Project to Enhance Access to Official Microdata

(Part of session on Data without Boundaries)

Roxane Silberman (Réseau Quetelet), Christof Wolf (GESIS - Leibniz Institute for the Social Sciences)

Data without Boundaries – DwB – is a project funded by the European Commission under the 7th Framework Programme. The aim of the project is to develop standards, tools, and services to support equal and easy access to official microdata for the European Research Area.

DwB reaches this goal through coordination of existing infrastructures, the Council of European Social Science Data Archives (CESSDA) and the European Statistical System (ESS) in collaboration with representatives of the research community.

The project's activities include harmonizing and enhancing legal frameworks, information security standards and procedures for researcher accreditation to increase transnational access to data. Training courses, user conferences and access forums enhance researcher's competencies to successfully work with official microdata. This is further supported by building extensive information systems for existing official data in Europe and other resource discovery tools.

The presentation will highlight the main features and goal of the project, as well as first outputs focusing on metadata issues.

Implementing a Longitudinal Data Portal with a Customized Version of Questasy

Toni Erik Sissala (FSD - Finnish Social Science Data Archive)

FSD has an ongoing four- year project, FSD Upgrade. One of its requirements is to create a longitudinal data portal to facilitate access to variable level data across multiple waves of a longitudinal study. The portal should have a web interface and be compatible with DDI.

Questasy is a web application developed by CentERdata which manages data and metadata for panel surveys. We evaluated Questasy and came to the conclusion that with some customization it would fit our needs.

FSD documents metadata in DD12 files. This metadata needs to be harvested into Questasy. We also wanted to be able to express statistical measures derived from the answers.

Questasy's database model is based on DDI3 and its source code is open for modifications. This gave us the benefit of not having to "reinvent the wheel", while we were still able to develop import functionality to automate the harvesting process, and expand the database to include statistical measures. The resulting portal will be interoperable with other FSD systems.

This presentation will demonstrate the developed features and point out the advantages we get from using Questasy in terms of accessibility, data discovery and dissemination.

Managing Private and Public Views of DDI Metadata Repositories

Dan Smith (Colectica)

DDI has traditionally been used to document published datasets in a standardized way. With the rapid adoption of the DDI Lifecycle version 3, organizations have also started to use the DDI metadata to drive their data collection, production, and cataloging processes. It is simple to publish DDI describing these datasets and processes if all the information recorded in the metadata is disclosable to the public. It is a challenge to publish DDI describing only specific types of the recorded information or specific fields in a metadata item.

The Colectica Workflow Service addresses this challenge by providing a publication workflow for DDI metadata. Users may submit publication requests and Administrators may optionally approve publications. Filters can be applied as whole sets of DDI items are moved from one Repository to another, such as between a private and public repository view. The Workflow Service can apply filters to specific DDI item types and even individual fields, as configured by an Administrator, to prevent the disclosure of non-public information in the public metadata.

Data Management, DDI-based Documentation, and Visualization of Business and Organizational Research Data at the DSZ-BO

Johanna Vompras (University of Bielefeld - University Library)

The main aim of the DSZ-BO (Data Service Center for Business and Organisational Data) at Bielefeld University is both to collect, archive, and distribute data from business and organizational social research but also to provide – and keep up-to-date – a catalog of existing relevant studies in this research domain. The technical infrastructure for these services consists of three strictly separated layers: the information portal with query processing as the presentation layer, XML/SQL databases and file systems as content layer, and the administration backend for the DSZ-BO staff to manage the study collection, upload DDI-XML-files, and to create XML instances. In this paper we focus on our web-based solution for a DDI 3.1-based information portal which visualizes the collected studies and provides fundamental search functionalities based on XQuery. One main issue here is to illustrate the intermediate steps of the process chain: from investigation of a study to its visualization at the portal; in other words we examine our strategies for comprehensible metadata management and present our scientific/technical workflows which are needed to bring together different research artefacts (e.g. study descriptions, materials, external resources, administrative data, and conceptualizations) into one 'publishable' DDI instance using various applications and editors.

Update on Questasy, a Data Dissemination Tool Based on DDI3

Edwin de Vet (CentERdata)

Questasy is a data dissemination tool based on DDI3. It is written in CakePHP and uses a MySQL database. In version 4.0.1 a basket was introduced which allows the researcher to select variables from various study units and create a data export from those selected variables in SPSS/SAS format. We extended this functionality by allowing the use of long format data in addition to the use of wide format.

In the 4.1 release, we will also include the possibility to version question items. Every question item has a mandatory version number field and question items are considered different versions of one another if they have the same name, are in the same question scheme but have a different version number. Another feature will be the direct inclusion of images in the question texts. The search function has been improved by adding the use of filters and pagination of the search results.

Question Bank Cultural Changes in the Netherlands, Moving from DDI Codebook to DDI Lifecycle

Marion Wittenberg (DANS - Data Archiving and Networked Services), Edwin de Vet (Centerdata)

The Netherlands Institute for Social Research (SCP) has a long tradition in research on cultural changes in the Netherlands. The (bi)annual survey Cultural Changes (CV) examines the opinions among the Dutch population on culture opinions and how these change over time.

The documentation and data of this survey is kept in several databases and systems. For the general public the data and metadata are published on the DANS NESSTAR server. In a joint project of SCP, Centerdata and DANS, the documentation and data of this survey will be converted into an instance of Questasy, a web application developed by Centerdata. Therefore Questasy has to be adjusted to the specifications of the survey Cultural Changes.

Our presentation will focus on the versioning of questions and the conversion of NESSTAR metadata into DDI lifecycle.

Sharing Category Systems for Survey Items with Others in DDI-Lifecycle Format

Wolfgang Zenk-Möltgen (GESIS - Leibniz Institute for the Social Sciences)

The CodebookExplorer is a GESIS tool to manage and print documentation for survey datasets, as well as to do searches and analyses of the datasets. Part of some CodebookExplorer databases are explorer views that hold hierarchical category systems for survey items. The categories may consist of topics, trends, scales, or other information according to the specific documentation needs. They may also be nested in as much levels as necessary. The variables of all studies in the database may be assigned to one or many categories in the different explorer views, enabling comparison of the variable documentation and question texts in each category. The presentation will investigate possibilities to share these category systems and the assignments of variables to them with others by using DDI-Lifecycle exports. The example used is the database "Childhood, Adolescence and becoming an Adult 1991-1997", which holds all documentation parts in German and English language and has information about waves/populations, comparative questions, and scales in its category systems. By making this documentation available in the structured format of DDI-Lifecycle, the re-use of the topics, trends or scales of this and other databases becomes possible.

Posters / Demonstrations

Colectica: Data Management with DDI 3

Jeremy Iverson, Dan Smith (both Colectica)

Colectica is a DDI3-based platform for creating, documenting, managing, distributing, and discovering data. Colectica aims to create publishable documentation as a by-product of the data management process. This booth will provide live demonstrations of the various components of the Colectica platform.

- Colectica Repository is a centralized storage system for managing data resources, allowing multiple people to document data together, and providing automatic version control.
- Colectica Portal is a web-based application which enables data and metadata publication and discovery. Colectica Portal integrates with several social networking technologies to provide enhanced collaboration and discovery.
- Colectica Designer interacts with Colectica Repository to provide advanced data management and documentation functionality. Colectica Designer can import data and metadata from a variety of formats, and can generate documentation and source code in a variety of formats.

DdiXslt

Jannik Jensen (Danish Data Archive)

Within the open source project DdiXslt, see: <https://code.google.com/p/ddixslt/> the Danish Data Archive (DDA) has contributed to the general visualization tool in the form of XML transformation technology producing classic code book information with added survey design routing.

DdiEditor

Jannik Jensen (Danish Data Archive)

The DdiEditor developed by the Danish Data Archive (DDA) is the key tool in a framework of data processing tools and processes composing data processing of survey datasets. The end product is data documentation in accordance with international metadata standards. The DdiEditor produces metadata documentation in DDI-Lifecycle.

A key objective for development of the DdiEditor is to provide users with a tool which is configurable, extendable and customisable allowing users to customise their personal work environment to their needs.

The DdiEditor is aimed at data processing for curation purposes (accommodating the requirements for a data archive) Development of the framework move forward to invite user groups ranging from researchers documenting a survey to re-users of data supplying additional metadata to the original study. For further information see project homepage: <http://code.google.com/p/ddieditor/>.

The Generic Statistical Information Model (GSIM)

Thérèse Lalor (both UNECE - United Nations Economic Commission for Europe)

The Generic Statistical Information Model is a reference framework of information objects, which enables generic descriptions of the definition, management, and use of data and metadata throughout the statistical production process.

GSIM provides a set of standardised, consistently described information objects, which are the inputs and outputs in the design and production of statistics. As a reference framework, it helps readers understand significant relationships among the entities involved in statistical production, and supports the development of consistent standards or specifications.

GSIM is one of the cornerstones for modernising statistical production and moving away from traditional silos. By defining and grouping objects common to all statistical production, regardless of subject matter, GSIM enables statistical organisations to rethink how their business could be organised to generate economies of scale.

The need for a Generic Statistical Information Model (GSIM) was first agreed at the 2010 Meeting on Management of Statistical Information Systems (MSIS).

The development of GSIM forms a key part of the strategic vision of the High Level Group on Strategic Business Architecture for Statistics (HLG-BAS) - a group of heads of National Statistical Institutes and International Agencies that support a common vision to modernise statistical production.

Nesstar Toolkit

Ørnulf Risnes (NSD - Norwegian Social Science Data Services)

Nesstar is a DDI-Codebook based software platform for enhancing, publishing and disseminating data on the Web. The Nesstar-toolkit enables you to:

- Find, browse, visualize and analyse data online
- Publish different kinds of data, from survey data to multidimensional tables
- Build custom applications over the Nesstar API to meet your specific needs.

This poster-session will provide live demonstrations of various components of the Nesstar toolkit.

- Nesstar Publisher is a data documentation tool. It consists of data and metadata conversion and editing tools, and enables users to enhance datasets by combining a wide range of catalogue and contextual information, which can be published to Nesstar Server or exported in required formats.
 - Nesstar WebView is a web-based data dissemination application. It can be used to view data, including tabular (cube) data, and metadata published on a Nesstar Server. Users can search, browse, analyse, and download a wide variety of statistical and related data.
 - It is possible to build custom applications to harness or to do tailor made operations on the rich metadata & data stored in a Nesstar server over the Nesstar API. Many applications built over the API are already in production.
-

Towards a DDI Backend Architecture - Sharing DDI-related Software Modules

Matthäus Zloch, Thomas Bosch, Dennis Wegener (all GESIS - Leibniz Institute for the Social Sciences)

In addition to using DDI as data exchange format, the development of DDI based systems can profit from sharing software modules that implement the handling of DDI metadata. We show a case study of how to implement an abstract DDI based data model that is instantiated and used by the projects Missy and StarDat.

We plan to have a paper poster visualizing the technical architecture and a computer demonstration showing the implementation of our prototype in order to have in-depth discussions with other developers, software engineers and architects.

Concurrent Tutorials

The tutorials take place in parallel on Monday, December 3rd at 9:00-12:30 am. Please indicate your participation when registering for the conference on-line if you are interested in attending one of them.

Data and Metadata Management Using DDI

Wendy Thomas (University of Minnesota, Minnesota Population Center), Marcel Hebing (DIW - German Institute for Economic Research)

Introduction to DDI with a business perspective.

Formalizing the process of data and metadata management has become increasingly important. The DDI metadata standard was designed to support metadata management from simple stand-alone studies to major statistical production systems. This workshop will look at how DDI supports the data and metadata management process from a high level business perspective. Use cases covering different organizational structures and processes will be used to provide a checklist of options for introducing DDI into an organization.

Introductions and Workshop Outline

Business goals:

What are the internal and external goals for your organization? Do they differ between types of organizations?

Features of DDI that can help address those goals:

What is the coverage of DDI? What content can be used to drive production and what is just informational? What additions are under development? What are the basic structural components of DDI and how does it interact with other standards used by data organizations?

How to integrate DDI into an existing system:

Where do you start within different organizational types and in different situations? How do you develop buy in from management, funders, and staff?

Create your own information systems on the basis of DDI-Lifecycle

Thomas Bosch, Matthäus Zloch (both GESIS - Leibniz Institute for the Social Sciences)

Introduction to DDI with a technical perspective.

DDI-Lifecycle is designed to document and manage data across the entire data lifecycle, from conceptualization to data publication, analysis and beyond. This tutorial offers a technical view on the core components of DDI-Lifecycle. The Microdata Information System (MISSY), which purpose is to document microdata like the German Microcensus, EU-SILC, and EU-LFS, serves as a representative use case of how to deal with DDI-Lifecycle in own software projects.

The DDI Ontology represents the most important parts of DDI-Codebook and DDI-Lifecycle in the Web of Data. In various software projects in the statistical domain the base concepts of DDI-Lifecycle are re-used to a large extent. Thus, the DDI Ontology could serve as a common abstract data model for all of these software projects. As software products have individual requirements, the common abstract data model has to be customized in form of individual data models.

Through the MISSY use case, it is shown how a well-structured software architecture, based on the model-view-controller software design pattern, might look like, how certain software project layers interact, and how to implement different persistence formats like DDI-XML, DDI-RDF, and relational databases.

Hands on DDI 3 with Colectica

Jeremy Iverson, Dan Smith (both Colectica)

Colectica is a platform for creating, documenting, managing, distributing, and discovering data. Colectica is built on open standards including DDI 3. This training course covers the following topics:

- Introduction to Colectica
- Introduction to DDI 3
- Documenting concepts and general study design
- Designing and documenting data collection instruments
- Creating and documenting data products
- Ingesting existing resources
- Publishing resources
- Hands-on: use Colectica to publish documentation for a sample study

Bring your laptop.

Building Applications with the Nesstar API

Archana Bidargaddi and Nesstar Development Team (NSD - Norwegian Social Science Data Services)

Get started building your own DDI-driven microdata and cube applications.

The Nesstar API gives simple Java-based access to your DDI-C based metadata and data collections. Furthermore, supports tabulation, correlation, regression and downloads of microdata - as well as cube functionality such as querying, slicing and dicing.

This hands-on tutorial will get you started using the Nesstar API, and show how you can build powerful applications around it. Participants that already have developed such applications are welcome to demonstrate their tools during this informal tutorial.

Participants are encouraged to experiment with the publicly available API before the tutorial. Bring your laptop.

Side Meetings

Two side meetings take place in the margins of EDDI12.

Modelling Qualitative Metadata Structures and the DDI

Organizers: Louise Corti (UK Data Archive), Arofan Gregory (ODaF - Open Data Foundation), Larry Hoyle (Institute for Policy & Social Research, University of Kansas), Joachim Wackerow (GESIS - Leibniz Institute for the Social Sciences)

November 28-30, 2012

This meeting is a continuation of the work already done by the DDI Alliance Qualitative Data Exchange Working Group and will provide an opportunity to produce a detailed technical report including a model on recommendations for modifying and improving the current DDI, plus an associated position paper. This meeting is by invitation only.

DDI Developers Meeting

Organizer: Johan Fihn (SND- Swedish National Data Services)

December 5-6, 2012

The meeting is an opportunity for software developers who work with the DDI Lifecycle standard to gather together. Attendees will:

- Find the latest work others in the field are doing
 - Learn about upcoming changes to the DDI standard
 - Increase interoperability among our applications
 - Get feedback and advice regarding tool implementation
 - Exchange experiences and come up with new development ideas
-