European DDI User Conference

6th Annual European DDI User Conference (EDDI14)

December 2-3, 2014, London, United Kingdom

Hosted by the Institute of Education(IOE), University of London

http://www.eddi-conferences.eu/eddi14

Schedule and Program with Abstracts

Version as of 25 November 2014

The conference will be start on Tuesday, December 2 at 9:00 am and will end on Wednesday, December 3 at 4:15 pm.

Tutorials will take place on Monday, December 1 from 9:00 am to 5:00 pm.

Side meetings will take place in the week before EDDI and on Thursday, December 4 from 9:00 am to 5:00 pm.

Venue

**Institute of Education, University of London
20 Bedford Way, London WC1H 0AL, United Kingdom**

Registration	Crush Hall
Main conference	Jeffery Hall (Level 1), Room 728 (Level 7)
Tutorials	Rooms 728 and 731 (Level 7)
Side meetings:	
DDI Lifecycle Moving Forward	Monday: Room 541 (Level 5), Tuesday: Room 541 (Level 5), Wednesday: Room 777 (Level 7), Thursday: Room 639 (Level 6), Friday: Room 603 (Level 6)
Developers Meeting	Room 731 (Level 7)
International Colectica User Conference (ICUC)	Room 728 (Level 7)
Conference Dinner December 2, 2014	Bloomsbury House 2-3 Bloomsbury Square, London, WC1A 2RL
Informal Get-Together	See detailed schedule below

Schedule

November 24-28, 2014

9:00 - 17:00	**Side Meeting**
	M1: Specification Development
	Location: Monday: room 541, Tuesday: room 541, Wednesday: room 777, Thursday: room G39, Friday: room 603
	DDI Lifecycle Moving Forward
	(By invitation only. For more information see related page at DDI Alliance.)

Sunday, November 30, 2014

19:00	**Informal Get-together**
	Location: The Lady Ottoline, 11A Northington Street, London WC1N 2JF

Monday, December 1, 2014

8:30	Starting Registration	
9:00 - 10:30	**Concurrent Tutorials**	
	Concurrent T1 (full day)	Concurrent T2 (half day)
	Introduction to DDI and Exercises with a Tool	**DDI for the Semantic Web**
	Location: Room 728	Location: Room 731
	A Hands-On, Practical Introduction to DDI Using Colectica (abstract)	**Let's Disco - Publish Your DDI Metadata as Linked Data** (abstract)
	Jeremy Iverson, Dan Smith (Colectica)	Thomas Bosch, Benjamin Zapilko (GESIS - Leibniz Institute for the Social Sciences)
10:30 - 11:00	**Break** - Location: Crush Hall	
11:00 - 12:30	Block 2 of T1	Block 2 of T2
12:30 - 13:30	**Lunch** - Location: Crush Hall	
13:30 - 15:00	Concurrent T1 (continued) Location: Room 728	Concurrent T3 (half day) **Software Development** Location: Room 731
	Block 3 of T1	**Working with the STARDAT DDI-Lifecycle Library** (abstract)
		Alexander Mühlbauer (GESIS - Leibniz Institute for the Social Sciences)
15:00 - 15:30	**Break** - Location: Crush Hall	
15:30 - 17:00	Block 4 of T1	Block 2 of T3
18:00	**Informal Get-Together**	
	Location: Princess Louise, 208 High Holborn, London WC1V 7EP	

Tuesday, December 2, 2014

8:30	Registration opens
9:00 - 10:15	**Conference Plenary P1: Welcome and Keynote** Location: Jeffery Hall Chair: Joachim Wackerow (EDDI Conference Chair)

	Welcome Jude England (Head of Research Engagement, British Library, and Acting Director, CLOSER – Cohort and Longitudinal Studies Enhancement Resources)
	Keynote Peter Knight (Deputy Director, Head of Research Information and Intelligence, Research and Development Directorate, Department of Health, United Kingdom Government) (about)

10:15 - 10:45	**Break** - Location: Crush Hall
10:45 - 12:15	**Concurrent Conference Sessions**

Concurrent A1: **Infrastructure for Data Collection, Research and Archiving** Location: Jeffery Hall Chair: Tito Castillo	Concurrent A2: **DASISH Software for Questions and Variables** Location: Room 728 Chair: Mari Kleemola
Enhancing Discoverability of Public Health and Epidemiology Research Data (abstract) Arofan Thomas Gregory (Open Data Foundation)	**A Common Metadata Understanding for the Three DASISH Survey Tools** (abstract) Hilde Orten (NSD - Norwegian Social Science Data Services); Taina Jääskeläinen (FSD - Finnish Social Science Data Archive); Edwin de Vet (CentERdata); Brita Dorer (GESIS - Leibniz Institute for the Social Sciences)
Facilitating Shared Survey Metadata Across the Life-Cycle in Longitudinal Studies - A UK Perspective (abstract) Jon Johnson (Centre for Longitudinal Studies, Institute of Education); Louise Corti (UK Data Archive, University of Essex)	**The DASISH Questionnaire Design Documentation Tool: Keeping Track of the Questionnaire Design Process** (abstract) Håvard Venge Bakkmoen, Hilde Orten (NSD - Norwegian Social Science Data Services); Yvette Prestage (City University, London)
Shifting from the Codebook Model to the Lifecycle Model of DDI: The Specifications of the French Center of Socio-Political Data (CDSP) (abstract) Alina Danciu (CDSP - Center of Socio-Political Data, Sciences Po Paris); Alexandre Mairot (CDSP, CNRS - The National French Center for Scientific Research)	**The DASISH Translation Management Tool: Adapting Existing Software to DDI** (abstract) Maurice Martens (CentERdata); Brita Dorer (GESIS - Leibniz Institute for the Social Sciences); Taina Jääskeläinen (FSD - Finnish Social Science Data Archive)
	The DASISH Question Variable Data Base: Developments and Outlook (abstract) Håvard Venge Bakkmoen, Hilde Orten, Benjamin Beuster (NSD - Norwegian Social Science Data Services)

Tuesday, December 2, 2014 (cont.)

12:15 - 13:45	**Lunch** - Location: Crush Hall	
13:45 - 15:15	**Concurrent Conference Sessions**	
	Concurrent B1: **Official Statistics** Location: Jeffery Hall Chair: Uwe Jensen	Concurrent B2: **Metadata Specifications** Location: Room 728 Chair: Olof Olsson
	Creating a Model for a Central Metadata Repository (abstract) Klas Blomqvist (Statistics Sweden)	**DDI 3.2 Best Practices - Consistency and Interoperability** (abstract) Dan Smith (Colectica)
	Use of DDI at INSEE (abstract) Guillaume Duffes (INSEE - National Institute of Statistics and Economic Studies, France)	**XKOS (Extended Knowledge Organization System) and How to Use it** (abstract) Franck Cotton (INSEE - National Institute of Statistics and Economic Studies, France)
	Issues, Current Status, and Perspectives about DDI in the National Institute Of Statistics and Geography of Mexico (INEGI) (abstract), *short presentation* Silvia L. Fraustro (INEGI)	**Generating Database and Domain Models from DDI 4** (abstract) Oliver Hopt, Brigitte Mathiak (GESIS - Leibniz Institute for the Social Sciences)
15:15 - 15:45	**Break** - Location: Crush Hall	
15:45 - 16:45	**Concurrent Conference Sessions**	
	Concurrent C1: **Official Statistics** Location: Jeffery Hall Chair: Klas Blomqvist	Concurrent C2: **Software / Tools** Location: Room 728 Chair: Marcel Hebing
	The Copenhagen Mapping - GSIM 1.1 and DDI 3.2 (abstract) Dan Smith (Colectica); Jannik Vestergaard Jensen (Danish Data Archive); Mogens Grosen Nielsen (Statistics Denmark)	**Collaborative Editing and Versioning of DDI Metadata: The Latest from Cornell's NCRN CED²AR Software** (abstract) Benjamin Perry, Venkata Kambhampaty, William Block (Cornell University); Kyle Brumsted (McGill University)
	Progress in the Use of DDI, SDMX, GSBPM, and Other Standards Inside and Outside of Statistics Denmark (abstract) Mogens Grosen Nielsen (Statistics Denmark); Anne Sofie Kjeldgaard, Jannik Vestergaard Jensen (Danish Data Archive)	**Creating a Joint Metadata Domain for the Social Sciences and Humanities in Europe** (abstract) Catharina Wasner (GESIS - Leibniz Institute for the Social Sciences); Stephanie Roth, Olof Olsson (SND - Swedish National Data Service)
18:30	**Conference Dinner** Drinks from 18:30, dinner served at 19:00 Location: Bloomsbury House, 2-3 Bloomsbury Square, London, WC1A 2RL	

Wednesday, December 3, 2014

9:15 - 10:15	**Conference Plenary P2** Location: Jeffery Hall Chair: Nikos Askitas **Incentivising Uptake of Metadata in the Social and Behavioural Sciences** (abstract) Louise Corti (UK Data Archive, University of Essex)
10:15 - 10:45	**Break** - Location: Crush Hall
10:45 - 12:15	**Concurrent Conference Sessions**

Concurrent D1: **Infrastructure for Data Collection, Research, and Archiving** Location: Jeffery Hall Chair: Hilde Orten	Concurrent D2: **Software / Tools** Location: Room 728 Chair: Barry Radler
Protocol Development for Large-Scale Metadata Archiving using DDI-Lifecycle (abstract) Will Poynter, Jennifer Spiegel (CLOSER, Institute of Education)	**An Integrated, Open Source Data Archiving System** (abstract) Adrian Dușa, Sorin Milutinovici, Cosmin Rentea, Letitia Velcescu (University of Bucharest)
Metadata Requirements to Document Data Analyses and Syntax Files in a Virtual Research Environment (VRE) (abstract) Uwe Jensen (GESIS - Leibniz Institute for the Social Sciences)	**DDI on Rails - Version 1.0** (abstract) Marcel Hebing (SOEP - German Socio-Economic Panel Study, DIW Berlin)
An Open Source, DDI-Based Data Curation System for Social Science Data (abstract) Limor Peer (Yale University); Ann Green (Digital Lifecycle Research & Consulting); Jeremy Iverson, Dan Smith (Colectica); Niall Keleher (Innovations for Poverty Action); Stephanie Wykstra (Innovations for Poverty Action)	**Lessons Learned from Using DDI-RDF Discovery Vocabulary as a Backend Data Model** (abstract) Matthäus Zloch (GESIS - Leibniz Institute for the Social Sciences)

Wednesday, December 3, 2014 (cont.)

12:15 - 13:45	**Posters and Software Demonstrations**
simultaneously with lunch	Location: Jeffrey Hall Foyer
	Chair: Jon Johnson
	Work Towards a Common Metadata Model for the Three DASISH Tools / The DASISH Questionnaire Design and Documentation Tool (QDDT) and the Question Variable Data Base (QVDB) – First Developments and Outlook (abstract)
	Håvard Venge Bakkmoen, Hilde Orten (NSD - Norwegian Social Science Data Services)
	DDI on Rails (abstract)
	Marcel Hebing (DIW Berlin - German Institute for Economic Research)
	Colectica for Excel: Bringing DDI to Excel (abstract)
	Jeremy Iverson, Dan Smith (Colectica)
	Protocol Development for Large-Scale Metadata Archiving Using DDI-Lifecycle (abstract)
	Will Poynter, Jennifer Spiegel (CLOSER, Institute of Education)
	New Features in EpiData - Among Others Exporting from EpiData to DDI 3.1 and Easy Archiving (abstract)
	Bodil Stenvig, Henrik Sejersen (Danish Data Archive); Torsten Bonde Christiansen, Jens Lauritsen (EpiData Association and Odense University Hospital)
12:15 - 13:45	**Lunch** - Location: Crush Hall

Wednesday, December 3, 2014 (cont.)

13:45 - 14:45	**Concurrent Conference Sessions**	
	Concurrent E1: **Data Harmonization** Location: Jeffery Hall Chair: Guillaume Duffes	Concurrent E2: **Software / Tools** Location: Room 728 Chair: Johan Fihn
	Leveraging DDI 3.2 to Power a Harmonized Data Extraction Tool for MIDUS (abstract) Barry Radler (University of Wisconsin); Jeremy Iverson, Dan Smith (Colectica)	**The IAB Metadata Portal: Consolidating Different Data Sources in One Application** (abstract) Iris Sonja Dieterich (The Research Data Centre (FDZ) of the Federal Employment Agency (BA) at the Institute for Employment Research (IAB))
	Application of DDI Comparison Capabilities to a Multi-Site Sexual Behaviour Data Harmonisation Exercise (abstract) Chifundo Kanjala, Gareth Knight, Jim Todd, Basia Zaba (London School of Hygiene and Tropical Medicine); Tito Castillo (University College London); David Beckles (Independent IT consultant)	**Exporting DDI 3.2 from IBM SPSS Data Collection using Colectica Designer 5** (abstract), *short presentation* Robert Pratt (Ipsos MORI) **Machine Actionable Integration of DataCite and DDI Metadata** (abstract), *short presentation* Wolfgang Zenk-Möltgen (GESIS - Leibniz Institute for the Social Sciences)
14:45 - 15:15	**Break** - Location: Crush Hall	
15:15 - 16:15	**Conference Plenary P3: Current Status and Outlook**	
	Location: Jeffery Hall Chair: William Block	
	DDI Lifecycle: Status of the Development of DDI 3 Wendy Thomas (Technical Committee, DDI Alliance)	
	DDI Lifecycle: Moving Forward – Status of the Development of DDI 4 Joachim Wackerow (Technical Committee, DDI Alliance)	
	Questions and Grumbles? Answers from the Technical Committee Arofan Gregory, Wendy Thomas and Joachim Wackerow (all Technical Committee, DDI Alliance)	
	Announcement of 3rd NADDI, Invitation to EDDI15 and Goodbye Barry Radler (NADDI team), Nikos Askitas, Joachim Wackerow (both EDDI core team), and Next Year's Host (name will be disclosed in session)	
18:00	**Informal Get-Together**	
	Location: The Jeremy Bentham, 31 University Street, London WC1E 6JL	

Thursday, December 4, 2014

9:00 - 17:00	Side Meetings	
	Concurrent M2: **International Colectica User Conference (ICUC 2014)** Location: Room 728 Organizers: Jeremy Iverson and Dan Smith (both Colectica) with CLOSER (Cohort and Longitudinal Studies Enhancement Resources)	Concurrent M3: **DDI Developers Meeting** Location: Room 731 Chair: Johan Fihn
	(for registration and program, see webpage of ICUC 2014)	(for contact see related email group)
18:00	**Informal Get-together** Location: The Lamb, 94 Lamb's Conduit Street, London WC1N 3LZ	

Plenary Presentations

Keynote
Peter Knight (↑ schedule)

Peter is a Deputy Director, Head of Research Information and Intelligence in the Research and Development Directorate at the Department of Health of the United Kingdom government with responsibility for policy and delivery of research information and intelligence. Peter joined the Department in April 2010 having established the Research Capability Programme in 2008 that established a secure research data service now operated by the MHRA called the Clinical Practice Research Datalink. Prior to these roles Peter was a Managing Director and interim Chief Executive at Winchester and Eastleigh Health NHS Trust.

Plenary Talk

Incentivising Uptake of Metadata in the Social and Behavioural Sciences
Louise Corti (University of Essex, UK Data Archive) (↑ schedule)

To some metadata is a complete way of life; to others it is a swear word. The meaning has numerous interpretations, but everyone who uses DDI knows exactly what the term means.

In this talk I will explore some of the great work being done to help incentivise those working in the domains of the Social and Behavioural Sciences to embrace metadata. I will provide examples from the UK that focus on: persuading survey producers and survey software vendors to appreciate and smooth the metadata journey; the work being done to help researchers at the coalface understand why good metadata matters; work to retrospectively fit a range of complex study documentation into structured metadata for a common search portal; and outreach work with data repository managers to show how DDI can be used as the model for describing most research projects and resulting data. Finally I will give an example of a new cohort study that is planning a lifecycle approach to metadata by considering sustainable metadata from its inception.

The trick is to keep it simple, using community narratives to gain trust and compliance.

Presentations in Concurrent Sessions

In alphabetical order by the last name of the first author

The DASISH Question Variable Data Base: Developments and Outlook

Håvard Venge Bakkmoen (NSD - Norwegian Social Science Data Services), Hilde Orten (NSD - Norwegian Social Science Data Services), Benjamin Beuster (NSD - Norwegian Social Science Data Services) (↑ schedule)

The DASISH Question Variable Data Base is a searchable metadata base with a broad public profile, under development for the Data Service Infrastructure for the Social Sciences and Humanities (DASISH) project. The primary aim of the tool is to serve the European Social Survey and other survey projects in their archive business processes, but the tool will also allow users from within the research community to browse the database in order to design new research or to explore existing projects. When populated, the DASISH Question Variable Data Base will give access to research questions in original languages, the concepts they are based on, the resulting variables, coding classifications and more.

The aims and requirements of the DASISH Question Variable Data Base were presented in EDDI 2013. Further developments on this DDI based tool have been going on since then, we now want to present its current functionalities, as well some ideas for possible future work.

The DASISH Questionnaire Design Documentation Tool: Keeping Track of the Questionnaire Design Process

Håvard Venge Bakkmoen (NSD - Norwegian Social Science Data Services), Hilde Orten (NSD - Norwegian Social Science Data Services), Yvette Prestage (City University, London) (↑ schedule)

The DASISH Questionnaire Design Documentation Tool is developed for the Data Service Infrastructure for the Social Sciences and Humanities (DASISH) project with the aim of assisting large-scale survey projects in the processes related to questionnaire development and documentation of the questionnaire design process from the first conceptualization to the final questionnaire. In EDDI 2013, the DASISH Questionnaire Design Documentation Tool was presented with focus on its purposes, usages, requirements and DDI usages. The tool has been further developed since then, and real life usage of its first functionalities will soon take place. We now want to present the current functionalities of the tool, its DDI based model, as well as an outlook for possible further developments.

Collaborative Editing and Versioning of DDI Metadata: The Latest from Cornell's NCRN CED[2]AR Software

Benjamin Perry (Cornell University), Venkata Kambhampaty (Cornell University), Kyle Brumsted (McGill University), William Block (Cornell University) (↑ schedule)

This presentation reports on the latest innovations of the Comprehensive Extensible Data Documentation and Access Repository (CED[2]AR), part of Cornell's NSF-Census Research Network Project. CED[2]AR is an online repository for metadata on surveys, administrative microdata, and other statistical information. CED[2]AR uses DDI 2.5 and supplemental XML to document these studies without the use of any relational databases, making the entire repository is extremely portable and lightweight. The user interface not only provides the means to navigate and search across multiple codebooks, but the ability to edit DDI through simple and intuitive web forms. This allows us to collaboratively curate DDI metadata.

However, in addition to the study it describes, the DDI document itself evolves over time. Tracking and documenting these changes is an essential part of a researcher's job. Rather than modifying the existing DDI Codebook schema, the CED[2]AR project uses Git, a distributed version control system traditionally built for software development. Git allows CED[2]AR to show how DDI metadata changes over time, and can retrieve old versions or parallel branches of the metadata. Git runs passively in the background, so

CED[2]AR begins versioning its codebooks without any user intervention, however the more advanced features of Git are still accessible.

Creating a Model for a Central Metadata Repository

Klas Blomqvist (Statistics Sweden) (↑ schedule)

In 2010 Statistics Sweden decided on a register and data warehousing vision and strategy. A GSIM based central metadata repository has a key role in providing information on data and controlling GSBPM processes.

Today metadata applies to locally stored and managed microdata. In the future the microdata is stored in structured data warehouses and the metadata makes it possible to connect microdata and macrodata with the statistics that are being produced, i.e. all the stages in the GSBPM.

In a changing society the NSIs are facing an increase in demand on both current compiled information sets but also an increase for new information.

The development of a structured data warehouse involves a number of production related issues; efficiency improvements, flexible data extractions and less redundancy. Reduced risk of errors, consistency, standardization of populations and variables, traceability and documentation will give quality improvements.

In the initial implementation phase the focus is on activities directed towards structured metadata of good quality in order to identify relevant variables, populations and objects in the data store. The development of a GSIM based conceptual model for a central metadata repository covering GSBPM has been a key step.

XKOS (Extended Knowledge Organization System) and How to Use it

Franck Cotton (INSEE - National Institute of Statistics and Economic Studies - France) (↑ schedule)

XKOS is an RDF vocabulary (or semantic web data model) published by the DDI Alliance and dedicated to the representation of statistical classifications.

XKOS is aligned on SKOS, the semantic web standard for thesauri, on the Neuchâtel Model (a data model for statistical classifications created by the official statistics community), and on the GSIM, the Generic Statistical Information Model.

The presentation will relate the inception and development of XKOS and present the main features of the specification, like classification levels, correspondences between classifications, typology of explanatory notes, etc.

It will also describe how XKOS fits in the classifications life-cycle. In particular, it will present:

- how to create XKOS data from existing classifications available in various formats (spreadsheets, text files, databases or web pages),
- how to publish XKOS on the web of data and link it to other classifications or concept schemes,
- how to manage classifications with XKOS in the Linked Data paradigm,
- how to use XKOS Linked Data.

Some typical use cases will be detailed: the dissemination of the classifications themselves, the organization of statistical data into cubes and the representation of associations between different data cubes.

Shifting from the Codebook Model to the Lifecycle Model of DDI: The Specifications of the French Center of Socio-Political Data (CDSP)

Alina Danciu (Center of Socio-Political Data (CDSP), Sciences Po Paris), Alexandre Mairot (Center of Socio-Political Data (CDSP), The National French Center for Scientific Research (CNRS)) (↑ schedule)

The French Center of Socio-Political Data has been disseminating surveys provided by external researchers since 2005. We have been using DDI-C to document the databases in Nesstar Publisher. Since 2012, the CDSP has been the data producer of the ELIPSS panel. Panel members have been administered a monthly web questionnaire. To date, this has resulted in 15 databases that are to be disseminated to the scientifically interested public. What are the best practices for using DDI effectively in this context?

Our first challenge was to identify an approach in which the entire lifecycle of the data collection, processing and archiving of the early-phase project ELIPSS is addressed. To meet its longitudinal information needs, we took DDI-L under consideration. At the same time, shifting to the Lifecycle model raised the issue of the migration of the existing databases of the CDSP catalogue into the new standard.

We started by making a list of specifications: level of metadata detail, possibility of upgrading DDI-C documented databases into DDI-L, user-friendliness of the interface, etc. We then compared different tools (Colectica, Questasy, DDI Editor, etc.) and selected the one that best fits our needs. We also identified good documentation practices. Our paper will present this strategy.

The IAB Metadata Portal: Consolidating Different Data Sources in One Application

Iris Sonja Dieterich (The Research Data Centre (FDZ) of the Federal Employment Agency (BA) at the Institute for Employment Research (IAB)) (↑ schedule)

The Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute of Employment Research (IAB) provides different types of data for the scientific community. There are research data on individuals, households and establishments and linked data (e.g. linked employer-employee, linked survey and register data) originating from different sources like the notification process of the social security system or surveys conducted by IAB research departments. Available metadata tools do not comply with the requirements for managing metadata of such complexity from different data sources. The presentation is based on the development of a software tool to handle and standardize our metadata in cooperation with TBA21 Assessment System GmbH (Germany), OPIT Consulting Kft. (Hungary) and Colectica (USA). The application will support the cooperation of different IAB internal research departments in the whole workflow of data documentation using a role based user management system. The metadata will be available publically through a web application which allows users to generate customized reports. The focus of this presentation lies on the challenges of handling data from multiple sources and the role DDI could play in this respect.

Use of DDI at INSEE

Guillaume Duffes (INSEE - National Institute of Statistics and Economic Studies - France) (↑ schedule)

There have been so far a significant number of initiatives around DDI at Insee, with two major projects involved: Coltrane ("transverse data collection for business surveys") and RMéS (Statistical Metadata Repository).

Efforts were shared by both projects to set up a statistical questionnaire generator driven by active metadata. RMéS defined these metadata (including a DDI profile) and Coltrane developed tools to transform them into different output formats.

In parallel, RMéS defined a generic information model based on GSIM, which underpins a Resource Oriented Architecture (ROA) relying on three main principles:

- Distinction between informational and non-informational resources

- Identification of resources by Uniform Resource Identifiers (URI)
- Access to resource representations by ReSTful Web Services

RMéS has already defined together with other projects a handful of metadata around statistical surveys and specified related web services to access them.

DDI 3.2 and additional RDF vocabularies will be the main representation formats for resources. Thus, it is crucial that the DDI 4 implementation model be aligned with GSIM, as well as compliant with DDI 3.2.

Insee pins also high hopes on the future RDF representation of DDI, which could be complementary with other RDF vocabularies already in use and offer great opportunities.

An Integrated, Open Source Data Archiving System

Adrian Dușa (University of Bucharest), Sorin Milutinovici (University of Bucharest), Cosmin Rentea (University of Bucharest), Letitia Velcescu (University of Bucharest) (↑ schedule)

During the past few years, RODA (Romanian Social Data Archive) enhanced its capabilities in the field of data archiving and information management, and focuses on technical and semantic interoperability with the CESSDA network, while obeying national regulations and constraints imposed by different data-owners.

The primary objective was to implement an online data archiving platform capable of handling all SIP, AIP and DIP components of the OAIS reference model, for both DDI 3.x and legacy DDI 2.x metadata. We have validated this system using real-world data from RODA's existing studies/codebooks.

This online platform has multiple access levels established through a Content Management System, allows importing /exporting /editing /searching of DDI metadata and various data formats, handles controlled vocabularies and simple questionnaire design, has a data browser/editor which offers statistical analysis using R as a statistical engine, followed by visualization thus enabling workflows for data archiving and processing.

An extensive array of open-source tools have been used, and he final aim is to release this data archiving system under an open-source license, and present the status of the current implementation to the community in order to obtain feedback, testing, bug reports and suggestions for future features.

Issues, Current Status, and Perspectives about DDI in the National Institute Of Statistics and Geography of Mexico (INEGI)

Silvia L. Fraustro (INEGI National Institute of Statistics and Geography of Mexico) (↑ schedule)

In this short presentation we will make a review of the process we followed to implement DDI for 77 statistical projects at INEGI, including population and economic censuses; national censuses of government, public safety, and prison system; social, labour, economic and services surveys; as well as several statistical projects from administrative records. We will also give an idea of where we are at this moment and how we plan to go deeper in the adoption of this standard. We envision promotion of DDI among all Mexican institutions to encourage documentation of the statistical projects that are part of the National Statistical Information System. We also plan to consider the possibility of using DDI to document production of those projects which result from processing Big Data sources.

Enhancing Discoverability of Public Health and Epidemiology Research Data

Arofan Thomas Gregory (Open Data Foundation) (↑ schedule)

In January 2014 the Wellcome Trust commissioned research on behalf of the Public Health Research Data Forum to appraise ways in which Forum partners could make research datasets easier for potential users to discover.

The research was undertaken by a team led by Dr Tito Castillo, Director of the NIHR BioResource at the University of Cambridge, with partners from the Farr Institute (at UCL Partners, University College London), the London School of Hygiene and Tropical Medicine, the Open Data Foundation, Ubiquity Press and the UK Data Archive.

The research involved a review of key research datasets, an online survey of key stakeholder groups, in-depth interviews with researchers, and an analysis of existing approaches for data discoverability.

Based on the findings of the research, three key models emerged through which the discoverability of public health and epidemiology research data could be enhanced – all of which depend critically on the wider adoption of best practices and data standards. The report sets out a series of recommendations for how funders could work with the community to address this key challenge.

DDI on Rails - version 1.0

Marcel Hebing (German Socio-Economic Panel Study (SOEP), DIW Berlin) (↑ schedule)

DDI on Rails is a web-based data portal optimized for the documentation of panel surveys. It accompanies researchers throughout the entire course of their research projects from conception to publication. After 3 years of development, we are glad to release DDI on Rails as an open source product. The presentation gives an overview of the software design, the use of DDI in the relational database model, and additional services around DDI on Rails.

Generating Database and Domain Models from DDI 4

Oliver Hopt (GESIS - Leibniz Institute for the Social Sciences), Brigitte Mathiak (GESIS - Leibniz Institute for the Social Sciences) (↑ schedule)

One of the reasons for DDI 4 to switch to a model-driven approach was to enable automatic generation of deliverables, such as the XML Schema and the documentation. We use the same pipeline to feed into the generation of domain models. Such domain models can be used by applications to handle import and export and database storage routines and may also serve as a basis for some of the functionality an application may provide. We demonstrate the current state of our JAVA-based prototype and discuss some of the open issues concerning code generation, such as under-defined data types.

Metadata Requirements to Document Data Analyses and Syntax Files in a Virtual Research Environment (VRE)

Uwe Jensen (GESIS - Leibniz Institute for the Social Sciences) (↑ schedule)

Metadata supporting collaborative data analyses within a virtual research environment demands more elaboration within the DDI standard. The ongoing German project "VRE soeb3" (3rd. Report on socio-economic development of Germany) is taken as use case to outline sense and objectives in developing a metadata set that support data documentation and retrieval in a joint research project.

The presentation focuses on needs and recommended solutions regarding four issues:

1. Metadata on used social science research data files from different data provider
2. Metadata to describe research questions and particular use of data sets in soeb3 work packages
3. Metadata to document the substance of syntax files and output files
4. Metadata of original and created variables

Of particular interest is first which metadata for syntax files researcher consider necessary, and secondly how they support syntax sharing across different work-packages in particular.

The framing focus is roughly to evaluate options to use exiting DDI metadata and to outline missing elements that might be considered in future DDI work. A short note informs about certain technical aspects of VRE soeb3.

Facilitating Shared Survey Metadata across the Life-Cycle in Longitudinal Studies - A UK Perspective

Jon Johnson (IOE - Institute of Education, Centre for Longitudinal Studies), Louise Corti (University of Essex, UK Data Archive) (↑ schedule)

Longitudinal studies in the UK involve a range of stakeholders:

1. University or Government Departments as commissioners and study design and specification;
2. Commercial market research agencies in the survey development and implementation;
3. Software vendors who provide the platform on which the survey is implemented
4. University or Government Departments as data managers and producers of research data
5. The UK Data Archive as disseminators

The presentation will describe recent developments in the UK which has brought together these stakeholders to create a common understanding of what the benefits of shared metadata are and an agreement on a common way forward.

Application of DDI Comparison Capabilities to a Multi-Site Sexual Behaviour Data Harmonisation Exercise

Chifundo Kanjala (London School of Hygiene and Tropical Medicine), Tito Castillo (University College London), Gareth Knight (London School of Hygiene and Tropical Medicine), Jim Todd (London School of Hygiene and Tropical Medicine), David Beckles (Independent IT consultant), Basia Zaba (London School of Hygiene and Tropical Medicine) (↑ schedule)

Harmonised datasets provide unique opportunities to analyse existing data obtained from disparate sources using new methods. Preparation of harmonised datasets involves examination of the source datasets in order to determine how the source datasets can be harmonised. The process of deciding the appropriate sources to use is often carried out by experienced researchers who are clear on what concepts they want to analyse in the harmonised dataset. When ad hoc comparisons are made during the harmonisation process, it becomes difficult to replicate the process. In this paper, we propose to investigate how DDI Lifecycle can improve the harmonisation of multi-site sexual behaviour data collected as part of demographic and HIV surveillance studies and allow systematic comparison of source and target datasets. Specific aspects considered include its group and comparison modules and the Conceptvariable and RepresentativeVariable features. We will discuss the findings of our explorations of the DDI comparison capabilities including how they can be used as a basis for harmonisation processes.

A more systematic approach facilitates replication of the decisions made in selecting the original data sources to use and what harmonisation choices to make.

The DASISH Translation Management Tool: Adapting Existing Software to DDI

Maurice Martens (CentERdata), Brita Dorer (GESIS - Leibniz Institute for the Social Sciences), Taina Jääskeläinen (FSD - Finnish Social Science Data Archive) (↑ schedule)

The DASISH Translation Management Tool is an online service for supporting translation processes for large multilingual surveys. It has been available since 2004 for the renowned Survey of Health, Ageing and Retirement in Europe (SHARE), but in order to comply with other software developed in the DASISH project it had to be adapted to support multiple surveys and therefore called for the use of DDI.

Challenges were including multiple surveys, versioning, referencing identifying languages and locales, and attribute pairs defining status of the overall translation process.

In EDDI 2013, the DASSISH Translation Management Tool was introduced with focus on user experience and interface. We now want to present the underlying architecture.

Progress in the Use of DDI, SDMX, GSBPM, and Other Standards Inside and Outside of Statistics Denmark

Mogens Grosen Nielsen (Statistics Denmark), Anne Sofie Kjeldgaard (Danish Data Archive), Jannik Vestergaard Jensen (Danish Data Archive) (↑ schedule)

At the end of 2014 existing Quality Declarations for all surveys at Statistics Denmark will be replaced using a number of international standards including Code of Practice (CoP) and Quality Assurance Framework (QAF), Single Integrated Metadata Structure (SIMS), GSBPM, DDI and SDMX. The implementation uses the standard DDI-tool Colectica with addition of SDMX-standards and user-interfaces following the phases in GSBPM. Parallel projects at Statistics Denmark are focusing on completing a general metadata system containing metadata on quality and methodology, concepts, variables and classifications. The work on classifications will be based on input from the Copenhagen Mapping of Neuchatel to DDI. At the national level a project on creating a metadata portal is launched together with Danish Data Archive and other external parties. At the international level an EU wide architecture on reporting quality information is being established based on SDMX and the European Interoperability Framework (EIF). This implies detailed agreements on interoperability at legal, organization, semantic and technical level. The paper will present various implementations, models and challenges.

A Common Metadata Understanding for the Three DASISH Survey Tools

Hilde Orten (NSD - Norwegian Social Science Data Services), Taina Jääskeläinen (FSD - Finnish Social Science Data Archive), Edwin de Vet (CentERdata), Brita Dorer (GESIS - Leibniz Institute for the Social Sciences) (↑ schedule)

Three tools are under development under the Data Service Infrastructure for the Social Sciences and Humanities (DASISH) project: a questionnaire design and documentation tool, a translation tool and a question/variable metadata base. The primary aim of the three tools is to assist large-scale survey projects in their business processes related to questionnaire design and documentation, translation and metadata production, archiving and dissemination.

The DASISH tools are developed as individual tools that should be able to communicate with each other. To facilitate this, a common metadata understanding between the three tools is needed. Requirements for a common metadata model for the three tools were presented in EDDI 2013. This presentation will focus on the further decisions made regarding the common metadata model, as well as their rationale. Topics like DDI usage, the communication between the tools, as well as the identification and versioning system and policy will be discussed.

An Open Source, DDI-Based Data Curation System for Social Science Data

Limor Peer (Yale University), Ann Green (Digital Lifecycle Research & Consulting), Jeremy Iverson (Colectica), Niall Keleher (Innovations for Poverty Action), Dan Smith (Colectica), Stephanie Wykstra (Innovations for Poverty Action) (↑ schedule)

The Institution for Social and Policy Studies (ISPS) at Yale University and Innovations for Poverty Action (IPA) partnered to develop a repository for research data from randomized controlled trials in the social sciences. The repository is an expansion – and major upgrade – of the existing ISPS Data Archive.

Together with Colectica, the partners have developed a software platform that leverages DDI Lifecycle, the standard for data documentation. The software structures the curation workflow, which also includes checking data for confidentiality and completeness, creating preservation formats, and reviewing and

verifying code. The software enables a seamless framework for collecting, processing, archiving, and publishing data.

This data curation software system combines several off-the-shelf components with a new, open source, Web application that integrates the existing components to create a flexible data pipeline. The software helps automate parts of the data pipeline and unifies the workflow for staff. Default components include Fedora Commons, Colectica Repository, and Drupal, but the software is developed so each of these can be swapped for alternatives.

This presentation will include a live demonstration of the data curation software.

Protocol Development for Large-Scale Metadata Archiving using DDI-Lifecycle
Will Poynter (IOE - Institute of Education), Jennifer Spiegel (IOE - Institute of Education) (↑ schedule)

A substantial part of CLOSER (Cohorts and Longitudinal Studies Enhancement Resources, www.closer.ac.uk) is to electronically document over 100,000 questions from 379 instruments for use in a metadata discovery platform (MDP). A subset profile of the DDI-Lifecycle 3.2 is the specification used by CLOSER.

During the metadata ingest process, a set of 4 entry principles were developed to outline and protect the overall standard by which the metadata is being recorded. To supplement the 4 principles, an entire protocol manual is being organically developed, including real examples of procedure. However there are situations where the entry principles conflict and therefore one or more must be broken in order to follow the more significant principle.

This paper also discusses how the selected DDI profile (the use of Question Grids and Interview Instructions) and the tools used for ingest (CADDIES) effect the principles and methodology used to document the metadata.

Exporting DDI 3.2 from IBM SPSS Data Collection using Colectica Designer 5
Robert Pratt (Ipsos MORI) (↑ schedule)

IBM SPSS Data Collection is one of the most widely used platforms for survey research. It is used by large market and social research agencies who often collaborate with government and academic institutions on social research projects. Currently IBM provides no DDI export and existing DDI tools cannot import from IBM SPSS Data Collection files. As commissioning bodies make more use of DDI there will be an increasing need to convert between these formats and converting through an intermediate, such as SPSS, is possible but resulted in lossy conversions. Colectica have recently added IBM SPSS Data Collection compatibility to version 5 of their Designer software. This presentation outlines how Colectica, IOE & Ipsos MORI are collaborating to allow rich conversions between IBM SPSS Data Collection and DDI.

Leveraging DDI 3.2 to Power a Harmonized Data Extraction Tool for MIDUS
Barry Radler (University of Wisconsin), Jeremy Iverson (Colectica), Dan Smith (Colectica) (↑ schedule)

Many longitudinal studies of health and aging contain thousands of variables and pose particular challenges for researchers who wish to analyze the data themselves or replicate others' research. MIDUS (Midlife in the United States) is a national longitudinal study of approximately 10,000 Americans designed to study aging as an integrated bio-psychosocial process. The study has a broad and unique blend of social, health, and biomarker data collected over several decades through a variety of modes. Recently the study has added new longitudinal cohorts that make data management, interpretation, and analysis even more challenging. A tool that allows researchers to easily create documented and citable data extracts that are directly related to their research questions would allow more time to be spent on public health research questions instead of data management.

In late 2013, the United States National Institutes of Health funded MIDUS to create a DDI-based, harmonized data extraction system. This presentation is an update on progress towards creation of a DDI-based tool that will facilitate identification and harmonization of similar MIDUS variables, while enhancing the MIDUS online repository with a data extract function. This will accomplish something unprecedented: the ability to obtain customized cross-project downloads of harmonized MIDUS data that are DDI-compliant. Doing so will greatly enhance efficient and effective public use of the large longitudinal and multi-disciplinary datasets that comprise the MIDUS study.

DDI 3.2 Best Practices - Consistency and Interoperability

Dan Smith (Colectica) (↑ schedule)

DDI 3.2 represents five years of development and is a vast improvement over DDI 3.1. This presentation will describe the approach used by the Technical Committee while creating DDI 3.2, suggestions on using DDI 3.2 to ensure interoperability, and future proofing DDI 3.2 implementations for eventual upgrades to new versions of the DDI standard.

While some new content was added to DDI 3.2, a major focus of the technical committee was on consistency throughout the entire model. This work included an Open Source tool to check consistency and interoperability rules and ensure their proper application to the XML schema, which will be described.

With the improvements in DDI 3.2, new approaches for using the standard are possible that take advantage of the consistency in the 3.2 model. These best practices eliminate the need for "searching" various locations in a DDI XML document in order to locate an item and simplify DDI profile creation. This approach is also being used in the XML serialization for DDI version 4. Using these best practices described in this paper will prepare your DDI 3.2 implementation for the future.

The Copenhagen Mapping - GSIM 1.1 and DDI 3.2

Dan Smith (Colectica), Jannik Vestergaard Jensen (Danish Data Archive), Mogens Grosen Nielsen (Statistics Denmark) (↑ schedule)

This presentation introduces the Copenhagen Mapping, an effort to implement GSIM 1.1 Statistical Classifications using the new DDI 3.2.

The Generic Statistical Information Model (GSIM) is an internationally endorsed reference framework for statistical information. This overarching conceptual framework will play an important part in modernizing, streamlining and aligning the standards and production associated with official statistics at both national and international levels. Statistical Classifications in GSIM are based on the previous the Neuchâtel model.

The Copenhagen Mapping describes the classification model in DDI 3.2 along with a mapping to the GSIM classification model. The Copenhagen Mapping also identifies gaps in the DDI standard's coverage of the GSIM classification model, and provides a controlled vocabulary for use with DDI 3.2 to cover these gaps.

The Generic Statistical Information Model (GSIM) is the first internationally endorsed reference framework for statistical information. This overarching conceptual framework will play an important part in modernizing, streamlining and aligning the standards and production associated with official statistics at both national and international levels.

Creating a Joint Metadata Domain for the Social Sciences and Humanities in Europe

Catharina Wasner (GESIS - Leibniz Institute for the Social Sciences), Stephanie Roth (SND - Swedish National Data Service), Olof Olsson (SND - Swedish National Data Service) (↑ schedule)

Task 5.4 of the DASISH project (Digital Services Infrastructure for Social Sciences and Humanities) aims to create a joint metadata index which will offer a publicly available faceted search for research data from the Social Sciences and Humanities. To achieve these objectives, the following technical solutions have

been developed and applied: Metadata was harvested from the participating research infrastructures CESSDA, CLARIN and DARIAH. Due to the heterogeneity of the involved partner communities, the XML files were collected in a wide range of formats like DDI-C, DDI-L, DataCite, CMDI and DC records. A set of facets based on existing catalogues and metadata schemas was created for the metadata index. The mappings to this simple flat format consisting of facets like keyword, type, access, language, country etc. were extracted using XPath expressions. The presentation introduces the work on the mappings and the normalization of field values. It will discuss the technical solutions used, issues encountered during the project, different solutions we could - and in some cases perhaps should have used - as well as where to go from here.

Machine Actionable Integration of DataCite and DDI Metadata

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

DataCite DOI persistent identifiers are used by the GESIS Data Archive to provide an easy way for researchers to identify and cite research datasets from the archive. By assigning a DOI, the basic cataloguing metadata from the Data Catalogue DBK is transformed from DDI-Codebook format to the DataCite metadata schema (via the da|ra service). For users of data citations it is appropriate to click on a DOI name and being redirected to a landing page with lots of information about the data, including links to additional rich metadata like DDI codebooks. But for machine actionable processing of data citations this is a show stopper. Computers need structured information to get to the additional files with the rich metadata. For this reason DataCite has included a structured way to document additional metadata in the schema: For each resource there may be related resources with the type "IsMetadataFor" or "HasMetadata". This makes it possible to process DataCite metadata and access DDI codebooks for selected studies. This presentation will show the implementation of an example at the GESIS Data Archive, describe lessons learned, and provide ideas for further improvement.

Lessons Learned from Using DDI-RDF Discovery Vocabulary as a Backend Data Model

Matthäus Zloch (GESIS - Leibniz Institute for the Social Sciences) (↑ schedule)

Designing a data model that covers "all" functional requirements of a specific domain is not a trivial task. The Social Science and the Linked Data communities developed the DDI-RDF Discovery Vocabulary, an ontology of the Data Documentation Initiative, in order to support the discovery of person-level data and its metadata. Software developers are faced with the problem to implement these data models and they should provide APIs to enable re-usage, extension, and persistence of such models.

Within the project Microdata Information System MISSY, we implemented a backend data model based on DDI-RDF. MISSY provides structured and accurately documented metadata, on variable and series level, for microdata from official statistics online, used by a wide community in Germany. In the current project we extended MISSY for the documentation of integrated microdata from Eurostat, where the European Social Sciences community is addressed.

In this presentation, we will show how we implemented DDI-RDF as an API with full support for the implementation of different persistence models, such as XML, RDF, and the relational model. By using these implementations in our MISSY applications, we will focus on code reuse and extension. We will depict some lessons-learned from obstacles we encountered, providing some usage statistics.

Posters and Software Demonstrations

In alphabetical order by the last name of the first author

Work Towards a Common Metadata Model for the Three DASISH Tools / The DASISH Questionnaire Design and Documentation Tool (QDDT) and the Question Variable Data Base (QVDB) – First Developments and Outlook

Håvard Venge Bakkmoen (NSD - Norwegian Social Science Data Services), Hilde Orten (NSD - Norwegian Social Science Data Services) (↑ schedule)

Three tools are under development under the Data Service Infrastructure for the Social Sciences and Humanities (DASISH) project: a questionnaire design and documentation tool, a translation management tool and a question/variable metadata base. The tools are being developed as individual tools that should be able to communicate with each other using DDI. Work towards a common metadata model for the three tools has been carried out in order to facilitate their interoperability. This paper poster will present the main decisions made regarding the common metadata model, as well as plans for further work.

The Questionnaire Design and Documentation Tool (QDDT) and the Question Variable Data Base (QVDB) are two software tools under development under the DASISH project. The aim of the QDDT is to assist the European Social Survey and other large-scale survey projects in their processes related to questionnaire development and documentation of the questionnaire design process. The QVDB will be developed to serve the archive business processes of surveys. When populated the tool will also allow users from within the research community to browse the data base for questions in original languages, the concepts they are based on, the resulting variables, coding classifications and more. This paper poster will present the first developments of these DDI-based tools, as well as an outlook for further work.

DDI on Rails

Marcel Hebing (DIW Berlin - German Institute for Economic Research) (↑ schedule)

The data portal DDI on Rails accompanies researchers throughout the entire course of their research projects from conception to publication/citation. The system offers researchers the possibility to explore the data, to compile personalized datasets, and to publish results on the publication database. DDI on Rails is study-independent and open-source, is able to document data with multiple versions/distributions and the specific characteristics of a longitudinal study, and is easy to use.

Colectica for Excel: Bringing DDI to Excel

Jeremy Iverson (Colectica), Dan Smith (Colectica) (↑ schedule)

Traditionally, data in spreadsheets and plain text formats do not contain rich documentation. Often, single-word column headers are the only hint given to data users, making it difficult to make sense of the data. Colectica for Microsoft Excel is a free tool to document your spreadsheet data using DDI, the open standard for data documentation. With this Excel addin, users can add extensive information about each column of data. Variables, Code Lists, and the datasets can be globally identified and described in a standard format. This documentation is embedded with the spreadsheet, ensuring the information is available when data are shared. The addin also adds support for SPSS and Stata formats to Excel. When opening an SPSS or Stata file in Excel, standard metadata is automatically created from the variable and value labels. Colectica for Excel can create print-ready reports based on the data documentation. The information can also be exported to the DDI standard, which can be ingested into other standards-based tools. This booth will include live demonstrations of the latest version of the Colectica for Excel tool,

showing how to document the contents of a spreadsheet, publish the information, and use the documentation to access data in an informed way.

Protocol Development for Large-Scale Metadata Archiving Using DDI-Lifecycle

Will Poynter (IOE - Institute of Education), Jennifer Spiegel (IOE - Institute of Education) (↑ schedule)
A substantial part of CLOSER (Cohorts and Longitudinal Studies Enhancement Resources, www.closer.ac.uk) is to electronically document over 100,000 questions from 379 instruments for use in a metadata discovery platform (MDP). During the metadata ingest process, a set of 4 entry principles were developed to outline and protect the overall standard by which the metadata is being recorded. However there are situations where the entry principles conflict and therefore one or more must be broken in order to follow the more significant principle. This poster will outline the 4 entry principles in order of importance and describe how the careful breaking of one or more principles can be applied in order to preserve the remaining principles and produce a standardised approach.

New Features in EpiData - Among Others Exporting from Epidata to DDI 3.1 and Easy Archiving

Bodil Stenvig (Danish Data Archive), Henrik Sejersen (Danish Data Archive), Torsten Bonde Christiansen (EpiData Association and Odense University Hospital), Jens Lauritsen (EpiData Association and Odense University Hospital) (↑ schedule)
EpiData is a free software program for data management and analysis in epidemiology, demographic and health research and social sciences. The program can be downloaded http://www.epidata.dk. The conceptual understanding, design and functions in EpiData are linked to the research working cycle. It offers you possibilities to define, enter, manage, analyze and export or archive data in the workflow of your research's project. People involved with EpiData and DDA have continually shared experiences and knowledge in the field of data documentation, preservation and dissemination, especially knowledge about the Data Documentation Initiative (DDI). Researchers, archivists and developers from EpiData and DDA Health has since 2010 been cooperating to changes the data format of EpiData Manager EpiData Entry to xml and to make it possible to export data to DDI 3.1. As a result it is now possible to export data to DDI 3.1 and also to support data archiving from EpiData Manager and EpiData Entry. New versions of the program are available from the homepage. EpiData Manager and EpiData Entry are used worldwide by approximately 200.000 worldwide.

Tutorials

In alphabetical order by the last name of the first instructor

Let's Disco - Publish Your DDI Metadata as Linked Data

Thomas Bosch (GESIS - Leibniz Institute for the Social Sciences), Benjamin Zapilko (GESIS - Leibniz Institute for the Social Sciences) (↑ schedule)

This half-day tutorial introduces Disco. The DDI-RDF Discovery Vocabulary (Disco) is a RDF Schema vocabulary that supports the discovery of microdata sets and related metadata using RDF technologies in the Web of Linked Data. Disco can be used to discover datasets by searching for specific questions, topics, and geographical coverage. Disco is intended to provide means to describe microdata by essential metadata for the discovery purpose. Existing DDI-XML instances can be transformed into this RDF format and therefore exposed as Linked Data. This workshop aims to present the Disco vocabulary and its practical appliance with DDI metadata in detail.

For a publication of DDI metadata as Linked Data, widely accepted and adopted RDF vocabularies (e.g. RDF Data Cube, DCAT, and PHDD) are reused to a large extend. The workshop organizers show how Disco is interwoven with these vocabularies. The Data Cube vocabulary is a W3C standard for representing data cubes representing multidimensional aggregate data derived from microdata which is represented by Disco. DCAT is a W3C standard for describing catalogs of datasets. Physical data description (PHDD) represents data (tables) in a rectangular format. The data could be either represented in records with character-separated values (CSV) or in records with fixed length. The combined usage of PHDD, Disco, and DCAT enables the creation of data repositories which provide metadata for the description of collections, for data discovery, and for processing of the data.

In different real world use cases, the workshop organizers demonstrate the adoption and inclusion of Disco in existing information systems. Workshop participants are encouraged to present their own use cases where Disco has been applied or will be used. Together with the workshop organizers, workshop participants will have the possibility to elaborate an RDF representation of their use cases and to formulate typical queries which are necessary to solve use case related problems.

The workshop has a technical perspective and addresses software developers as well as data publishers who want to publish DDI metadata as Linked Data. No prerequisites are necessary, since a short introduction to RDF and Linked Open Data will be given.

A Hands-On, Practical Introduction to DDI Using Colectica

Jeremy Iverson (Colectica), Dan Smith (Colectica) (↑ schedule)

This full-day workshop offers a practical approach to getting started with DDI. It starts with a high-level overview of the DDI content model, and then teaches how to create DDI XML, both manually and with Colectica. Finally, participants will learn how to publish DDI metadata. Attendees may **bring their laptops** to follow the hands-on exercises.

The DDI Information Model

This session provides a detailed look at the content described by DDI, and explores how different organizations' information relates to the DDI terminology.

In-depth with DDI

This session provides technical training on the DDI standard. Participants will learn how to create XML that complies with the DDI XML schemas.

What's New in DDI 3.2

If you are familiar with DDI 3.1 or DDI 3.0, but haven't yet looked into DDI 3.2, this session gives an overview of the big improvements and simplifications in the new version.

- Schema Improvements
- Simplified packaging
- Clearer identification
- More consistency
- Bug fixes
- Content additions and improvements
- Data Collection Instruments
- Harmonization
- Groups
- Data

Version Management and Propagation in DDI

Learn how to manage changes to information over time, both in DDI and in Colectica. Learn how to keep historical information accurate, while allowing for changes to actively managed information.

Publishing DDI Metadata

We will discuss methods of publishing information, including XSLT transforms and custom tools. Participants will also use Colectica to create publishable materials from DDI metadata.

Working with the STARDAT DDI-Lifecycle Library

Alexander Mühlbauer (GESIS - Leibniz Institute for the Social Sciences) (↑ schedule)

This half-day tutorial introduces the DDI-Lifecycle library of STARDAT. With the project STARDAT the GESIS Data Archive for the Social Sciences develops an open source, Java based, extensible domain model library with object-relational persistence to support DDI Lifecycle metadata documentation.

Based on a documentation example, this workshop discusses the experiences in extracting and implementing the model and introduces the usage and extension possibilities of the library. Core features like identification and uniform resource name resolution, historization and versioning and staged reuse of metadata are illustrated by code snippets. As participant with Eclipse IDE, Git and Apache Maven running on your notebook, you will be able to produce DDI XML 3.1 and 3.2 for the provided example.

Target audience are metadata documentation specialists with programming skills preferably in Java/Spring and software developers who have started or plan to develop a metadata documentation system on DDI Lifecycle and think about reusing or forking existing developments.

Authors Index

Organizations Index