5th Annual European DDI User Conference (EDDI13)

December 3-4, 2013, Paris, France

Hosted by Réseau Quetelet

http://www.eddi-conferences.eu/eddi13

Schedule and Program with Abstracts

Final version as of 4. December 2014

The conference will open on Tuesday, Dec 3rd at 9:00 pm and close on Wednesday, Dec 4th at 5:30 pm.

Tutorials will take place on Monday, Dec 2nd at 2:00 pm - 5:30 pm.

Schedule

Venue

Main venue	Cité Internationale Universitaire de Paris
	17 Boulevard Jourdan, 75014 Paris
	Cambodia House ("Maison du Cambodge")
	   Room "Salon d'honneur" (plenaries and concurrent sessions)
	Maison Internationale - west wing
	   Room "Salon Gulbenkian" - 1st floor (concurrent sessions)
	   Room "Salle Nathan" - 1st basement floor (concurrent session and posters/software demonstrations)

Tutorials	ENSAE Building
	3 Avenue Pierre Larousse, 92240 Malakoff
	   Room F9 (1st floor)
	   Room 26 (2nd floor)

Side meetings:
   Developers Meeting and **DDI Lifecycle Moving Forward Sprint 2**
	EHESS Building
	190 avenue de France, 75013 Paris
	   Room "Salle du conseil B" (lower ground floor)
	   Room "Salle Jean Pierre Vernant" (8th floor)

Conference Dinner	L'Autobus Impérial - 14 Rue Mondétour, 75001 Paris
Informal Get-together	Le Café Tournesol - 9 Rue de la Gaité, 75014 Paris

Sunday, December 1, 2013

18:00	**Informal Get-together**
	Location: Le Café Tournesol

Monday, December 2, 2013

9:00 - 17:00	**Special Meeting**	
	M1: Software Development	
	Location: EHESS Building - Room "Salle du conseil B"	
	Chair: Johan Fihn	
	DDI Developers Meeting	
14:00 - 17:30	**Concurrent Tutorials**	
	Concurrent T1:	Concurrent T2:
	Introduction to DDI with a Business Perspective	Building DDI Applications
	Location: ENSAE Building - Room F9 (1st floor)	Location: ENSAE Building - Room 26 (2nd floor)
	Data and Metadata Management Using DDI (abstract, presentation)	**Learn How to Use the Nesstar REST API. Use Javascript and HTML to Bring Your DDI Metadata and the Power of Nesstar to Your Website** (abstract, presentation)
	Wendy Thomas (University of Minnesota, Minnesota Population Center), Marcel Hebing (DIW Berlin - German Institute for Economic Research)	Ørnulf Risnes (NSD - Norwegian Social Science Data Services)
18:00	**Informal Get-together**	
	Location: Le Café Tournesol	

Tuesday, December 3, 2013

8:30	Starting Registration

9:00 - 10:15	**Conference Plenary P1: Welcome and Keynote**
	Location: Cambodia House - Room "Salon d'honneur"
	Chair: Joachim Wackerow (EDDI Team Chair)

	Welcome
	Roxane Silberman (Head of Réseau Quetelet)

	Keynote: Why National Statistical Institutes are Increasingly Interested in Standards
	Philippe Cuneo (Director of Methodology, Statistical Coordination and International Relations at Insee, French National Institute for Statistics and Economic Studies)

10:15 - 10:45	**Break** in front of room "Salon d'honneur" of Cambodia house

10:45 - 12:30	**Concurrent Conference Sessions**	

	Concurrent A1:	Concurrent A2:
	Official Statistics	Software / Tools
	Location: Cambodia House - Room "Salon d'honneur"	Location: Maison Internationale - Room Gulbenkian
	Chair: Raphaëlle Fleureux	Chair: Johanna Vompras
	GSIM and DDI: Working Together (abstract, presentation) Therese Lalor (UNECE), Guillaume Duffes (INSEE)	**An Update on the Rogatus Platform** (abstract, presentation) Ingo Barkow (German Institute for International Educational Research), David Schiller (IAB - Institute for Employment Research, Germany)
	Common Statistical Production Architecture (abstract, presentation) Therese Lalor (UNECE), Arofan Gregory (Metadata Technology)	**Update on Questasy, a Data Dissemination Tool Based on DDI3** (abstract, presentation) Edwin de Vet (CentERdata)
	Standardized Quality Declarations with DDI, SDMX, and Colectica (abstract, presentation) Mogens Nielsen (Statistics Denmark), Jeremy Iverson (Colectica), Dan Smith (Colectica)	**Historization and Versioning of DDI-Lifecycle Metadata Objects - Findings in the STARDAT Project** (abstract, presentation) Alexander Mühlbauer (GESIS - Leibniz Institute for the Social Sciences)
	Using DDI to Automate Blaise Instrument Generation (abstract, presentation) Simon Wall (Australian Bureau of Statistics)	**Metadata Management for Research Data with DBKfree** (abstract, presentation) Wolfgang Zenk-Möltgen (GESIS - Leibniz Institute for the Social Sciences)

12:30 - 14:00	**Lunch** in Bar du Théâtre (Maison Internationale – east wing, ground floor)

Tuesday, December 3, 2013 (cont.)

14:00 - 15:30	Concurrent Conference Sessions	
	Concurrent B1: Infrastructure for Data Collection, Research, and Archiving Location: Cambodia House - Room "Salon d'honneur" Chair: Hilde Orten	**Concurrent B2:** Software / Tools Location: Maison Internationale - Room Gulbenkian Chair: Alexander Mühlbauer
	Proposing a Metadata Solution over Multiple RDCs in the German Context (abstract, presentation) David Schiller (IAB - Institute for Employment Research, Germany), Ingo Barkow (German Institute for International Educational Research)	**Using DDI within a Data Archive - the SND Case Study** (abstract, presentation) Johan Fihn, Olof Olsson (both Swedish National Data Service)
	The Medical Research Council Gateway (abstract, presentation) Philip Curran (MRC Unit for Lifelong Health & Ageing at UCL), Peter Dukes (Medical Research Council), Caroline Shriver (Medical Research Council), Catherine Jones (Science and Technology Facilities Council), Alastair Duncan (Science and Technology Facilities Council), Alex Addyman (MRC Unit for Lifelong Health & Ageing at UCL)	**RODA's Open-Source Web Platform for DDI** (abstract, presentation) Adrian Dușa, Cosmin Rentea (both RODA - Romanian Social Data Archive)
	Building a Harmonized Data Market for Longitudinal Data with MIDUS and DDI (abstract, presentation) Barry Radler (University of Wisconsin - Madison), Jeremy Iverson (Colectica), Dan Smith (Colectica)	**DDI on Rails** (abstract, presentation) Marcel Hebing (DIW Berlin - German Institute for Economic Research)
	The Questionnaire Generation in the French National Statistical Institute (abstract, presentation) Eric Sigaud (INSEE - French NSI)	**Using Extended Attributes in Data Analysis Software- Controlled Vocabularies, Tools and DDI** *[full paper]* (abstract, presentation) Larry Hoyle (Institute for Policy & Social Research, University of Kansas)
15:30 - 16:00	**Break** in front of room "Salon d'honneur" of Cambodia house	

Tuesday, December 3, 2013 (cont.)

16:00 - 17:00	Concurrent Conference Sessions	
	Concurrent C1: DwB - Data without Boundaries / Virtual Research Environment Location: Cambodia House - Room "Salon d'honneur" Chair: Marcel Hebing	**Concurrent C2:** Software / Tools Location: Maison Internationale - Room Nathan Chair: Olof Olsson
	A New CESSDA Portal for European Research Data Discovery (abstract, presentation) Ørnulf Risnes (NSD - Norwegian Social Science Data Services), John Shepherdson (UK Data Archive), Pascal Heus (Metadata Technology)	**Colectica 5: A New Generation of Open Metadata Tools** (abstract, presentation) Jeremy Iverson, Dan Smith (both Colectica)
	The Next Generation of the Microdata Information System MISSY: An Integrated Solution for the Documentation of European Microdata (abstract, presentation) Jeanette Bohr, Alexander Mack, Florian Thirolf, Thomas Bosch (all GESIS - Leibniz Institute for the Social Sciences)	**DataForge:SledgeHammer: Your Data/Metadata Management Tool** (abstract, presentation) Pascal Heus (Metadata Technology North America)
	Metadata Requirements to Document Data Analyses and Syntax Files in a Virtual Research Environment (abstract), *cancelled* Uwe Jensen (GESIS - Leibniz Institute for the Social Sciences)	**REST API for Nesstar. Present Your Data and Metadata in New Ways - Fast.** (abstract, presentation) Ørnulf Risnes (NSD - Norwegian Social Science Data Services)

Tuesday, December 3, 2013 (cont.)

17:00 - 18:00	**Posters and Software Demonstrations (PSD)**
	Location: Maison Internationale - Room Nathan
	Chair: Catharina Wasner
	Early Implementation of DDI Lifecycle on the Integrated Version of the Cornell National Social Survey (abstract, presentation)
	Florio Arguillas, Jeremy Williams, William Block, Warren Brown (all Cornell Institute for Social and Economic Research - CISER)
	The New Microdata Information System (MISSY): Integration of DDI-based Data Models, an Open-Source Software Architecture, and Independent Persistence Service Implementations (abstract, presentation)
	Thomas Bosch, Jeanette Bohr, Matthäus Zloch (all GESIS - Leibniz Institute for the Social Sciences)
	DataForge:SledgeHammer: Your Data/Metadata Management Tool (abstract, presentation)
	Pascal Heus (Metadata Technology North America)
	Colectica for Excel: Increasing Data Accessibility using Open Standards (abstract, presentation)
	Jeremy Iverson, Dan Smith (both Colectica)
	SPSS and DDI-L (abstract)
	Jannik Jensen (Danish Data Archive)
	From Curation to Publication of DDI-L Metadata (abstract)
	Jannik Jensen, Anne Sofie Fink (both Danish Data Archive)
	Modelling an Evolving, Geospatial SKOS Code List (abstract, presentation)
	Ørnulf Risnes, Ole Voldsæter (both NSD - Norwegian Social Science Data Services)
	Managing and Sharing Data within the Collaborative Research Center SFB 882 in Practice: Data Documentation within a Virtual Research Environment for the Social Sciences. (abstract, presentation)
	Johanna Vompras (University of Bielefeld LibTec, Library Technology and Knowledge Management), Christoph Broschinski (Collaborative Research Center (SFB 882), Data and Information Infrastructure Project)
	Easy DDI Organizer (EDO) - Open Source Survey Planning and Metadata Management Software (abstract, presentation)
	Yuki Yonekura (Center for Social Research and Data Archives, Institute of Social Science, The University of Tokyo)
19:30	**Meeting at Restaurant** – Welcome Drink
20:00	**Conference Dinner**
	Location: L'Autobus Impérial - 14 Rue Mondétour, 75001 Paris

Wednesday, December 4, 2013

9:00 - 10:30	**Conference Plenary P2: DDI Specification** Location: Cambodia House - Room "Salon d'honneur" Chair: Simon Wall
	To 3.2 or Not to 3.2? That is the Question. (abstract, presentation) Wendy Thomas (University of Minnesota Minnesota Population Center), Jon Johnson (Centre for Longitudinal Studies, London)
	Proposal for a DDI Data Discovery Model (abstract, presentation) Jay Greenfield (Booz Allen Hamilton)
10:30 - 11:00	**Break** in front of room "Salon d'honneur" of Cambodia house
11:00 - 12:30	**Concurrent Conference Sessions**

	Concurrent D1: Reusing and Sharing Metadata Location: Cambodia House - Room "Salon d'honneur" Chair: Larry Hoyle	Concurrent D2: Software / Metadata Standard Location: Maison Internationale - Room Gulbenkian Chair: Arofan Gregory
	Harmonizing Between Different Agencies Using DDI Profiles (abstract, presentation) David Schiller (IAB - Institute for Employment Research, Germany), Ingo Barkow (German Institute for International Educational Research), Marcel Hebing (DIW Berlin - German Institute for Economic Research)	**Open Source DDI-L Search** (abstract, presentation) Jannik Jensen (Danish Data Archive)
	Implementing DDI Profiles at a National Statistical Office: Experiences at INSEE and ABS (abstract, presentation) Guillaume Duffes (Insee - Institut National de la Statistique et des Etudes Economiques), Simon Wall (Australian Bureau of Statistics)	**Use Cases and Vocabularies Related to the DDI-RDF Discovery Vocabulary** (abstract, presentation) Thomas Bosch (GESIS - Leibniz Institute for the Social Sciences)
	Describing Qualitative Surveys at beQuali: a Crosswalk between Documentation, Archiving and Dissemination Metadata (abstract, presentation) Guillaume Garcia, Sarah Cadorel, Julien Rault (all Centre de Données Socio-Politiques - Réseau Quetelet)	**Mapping DDI-RDF Discovery - a Hackathon Story** (abstract, presentation) Olof Olsson (Swedish National Data Service), Jannik Jensen (Danish Data Archive)
	Integrating DDI with Metadata for Qualitative Data: Fully Implementing the QuDEx Standard (abstract, presentation) Louise Corti (UK Data Archive), Darren Bell (UK Data Archive), Agustina Martinez (Liverpool John Moores University)	**Data Model and Data Standard - a Happy Marriage?** (abstract, presentation) Oliver Hopt, Brigitte Mathiak (both GESIS - Leibniz Institute for the Social Sciences)
12:30 - 14:00	**Lunch** in Bar du Théâtre (Maison Internationale – east wing, ground floor)	

Wednesday, December 4, 2013 (cont.)

14:00 - 15:30	Concurrent Conference Sessions	
	Concurrent E1: DASISH - Data Service Infrastructure for the Social Sciences and Humanities Location: Cambodia House - Room "Salon d'honneur" Chair: Thérèse Lalor	Concurrent E2: Reusing and Sharing Metadata Location: Maison Internationale - Room Gulbenkian Chair: Johan Fihn
	Data Service Infrastructure for the Social Science and the Humanities (DASISH): Improving Survey Quality in Cross-national Research (abstract, presentation) Eric Harrison (City University London)	**Encoding Provenance of Social Science Data: Integrating PROV with DDI** *[full paper]* (abstract, presentation) Carl Lagoze (University of Michigan School of Information), Jeremy Williams, Lars Vilhuber, William Block (all Cornell Institute for Social and Economic Research - CISER)
	Developing a Transportable, Standardised System of Monitoring Employing Harmonised Metadata Files Which Can Aid Central Field Supervision, Control and Monitoring (abstract, presentation) Sally Widdop (City University London), Johanna Bristle (Max Planck Institute for Social Law and Social Policy), Yvette Prestage (City University London)	**How Do We Manage Complex Questions in the Context of the Large-Scale Ingest of Legacy Paper Questionnaires into DDI-Lifecycle?** *[full paper]* (abstract, presentation) Claude Gierl, Jon Johnson (both Centre for Longitudinal Studies, Institute of Education, London)
	New Tools for Complex Surveys: the DASISH Questionnaire Design Documentation Tool and Question Variable Data Base (abstract, presentation) Hilde Orten, Håvard Bakkmoen (both NSD - Norwegian Social Science Data Services), Yvette Prestage, Sally Widdop (both City University London)	**Generating Blaise Surveys from the Data Documentation Initiative's Metadata Standard using Colectica** (abstract, presentation) Dan Smith, Jeremy Iverson (both Colectica)
	Making Custom DDI 3.1 Exports from the Language Management Utility (LMU) (abstract, presentation) Edwin de Vet (CentERdata), Eric Balster (CentERdata)	**DDI in Denmark** (abstract, presentation) Jannik Jensen (Danish Data Archive), Mogens Nielsen (Statistics Denmark)
15:30 - 16:00	**Break** in front of room "Salon d'honneur" of Cambodia house	

Wednesday, December 4, 2013 (cont.)

16:00 - 17:30	**Conference Plenary P3: Current Status and Outlook**
	Location: Cambodia House - Room "Salon d'honneur"
	Chair: William Block
	DDI Lifecycle: Moving Forward – Outcome of the Recent Workshop in Dagstuhl (presentation)
	Joachim Wackerow (GESIS - Leibniz Institute for the Social Sciences)
	Questions and Grumbles? Answers from the Technical Committee
	Arofan Gregory, Wendy Thomas, and Joachim Wackerow (all Technical Committee, DDI Alliance)
	Announcement of 2. NADDI (presentation), Invitation to EDDI14 (presentation) and Goodbye
	Larry Hoyle (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: Le Café Tournesol

Thursday, December 5, 2013

9:00 - 17:00	**Special Meeting**
	M2: Specification Development
	Location: EHESS building - Room "Salle du conseil B" (lower ground floor)
	DDI Lifecycle Moving Forward Sprint 2

Friday, December 6, 2013

9:00 - 17:00	**Special Meeting**
	M2: Specification Development
	Location: EHESS building - Room "Salle Jean Pierre Vernant" (8th floor)
	DDI Lifecycle Moving Forward Sprint 2

Plenary Presentations

Keynote

Philippe Cuneo (Director of Methodology, Statistical Coordination and International Relations at Insee, French National Institute for Statistics and Economic Studies)

Proposal for a DDI Data Discovery Model

Jay Greenfield (Booz Allen Hamilton)

The next generation DDI specification will take the form of a UML data model. As envisioned it will include a core that will be the nexus of several modules. Each module will realize one or perhaps a family of use cases. So far two modules have been envisioned: a survey description module and a data description module. Here we propose a third module - the data discovery module.

The UML model architecturally has a base, a core and extensions. The definition of the base is a work in progress. The core consists of DDI study objects that are pivotal to the several modules. The modules themselves extend the core. Extensions either consist of additional DDI study objects not in the core, new study objects and/or "bindings" that relate DDI study objects to other models like OWL and RDF.

In our presentation additions to the core and several extensions are proposed. The core addition is the DataElement (DDI 3.2) and related objects. One extension is already in play: the DDI-RDF discovery vocabulary. Another extension is faceted classification like we see Amazon and other vendors employ for the presentation and discovery of product information. A third extension is for hypothesis testing.

To 3.2 or Not to 3.2? That is the Question.

Wendy Thomas (University of Minnesota Minnesota Population Center), Jon Johnson (Centre for Longitudinal Studies, London)

As DDI-Lifecycle develops new versions there will always be the question of whether to adopt the new version or to continue with what works now. This question is particularly perplexing when work is already taking place on an improved model based approach for future versions of DDI. Is it worth the effort to change to 3.2? Like most DDI application questions, the answer is "That depends". We already know that some users are leaping into 3.2 because of new functionalities that they need. Others are sticking to 3.1 because it does what they need it to do and they have invested in development and programming work.

…

This session will look at 3.2 from a user's perspective to answer the following questions:

- What needs are met by 3.2 that cannot be handled well in earlier DDI versions?
- Complex Questionnaires and data processing
- Conceptual modeling
- Quality control and reusable content
- What's been cleaned up in 3.2 to function more smoothly or completely
- Multi-language content
- Tracking changes in Organizations, Individuals, and Geographic content Consistency
- Archiving submitted content
- I'm sticking to 3.1 so what do I need to know about moving forward in the future and skipping 3.2? Managing string content
- Managing identification content
- Resource packages
- Representations and missing values

Presentations in Concurrent Sessions

An Update on the Rogatus Platform

Ingo Barkow (German Institute for International Educational Research), David Schiller (IAB - Institute for Employment Research, Germany)

Rogatus is an open source questionnaire and metadata solution basing on the DDI 3.2 and SDMX standard and using the Generic Longitudinal Business Process Model (GLBPM) to specify its tool chain. Currently the project is supported by DIPF, TBA21, OPIT, IAB and GESIS and creates more and more interest especially with NSIs and data collection agencies. This presentation gives an update on new developments since NADDI 2013 including the data management portal, coding support for ISCED, improvements on the case management system, compatibility to other platforms like Colectica or MMIC plus an outlook on the mobile sampling client.

The Next Generation of the Microdata Information System MISSY: An Integrated Solution for the Documentation of European Microdata

Jeanette Bohr (GESIS - Leibniz Institute for the Social Sciences), Alexander Mack (GESIS - Leibniz Institute for the Social Sciences), Florian Thirolf (GESIS - Leibniz Institute for the Social Sciences), Thomas Bosch (GESIS - Leibniz Institute for the Social Sciences)

MISSY is an integral part of the service infrastructure for official microdata offered by the GESIS German Microdata Lab. This DDI based system provides structured metadata online in order to facilitate and encourage the use of official microdata for scientific purposes. Up to now, MISSY focuses on metadata for the German Microcensus, but an extension of MISSY for the documentation of integrated European microdata from Eurostat is under work. Within the data documentation process we are tightly cooperating with the Data without Boundaries project.

In addition to detailed variable level metadata MISSY provides users with a comprehensive study description which includes country specific information on a yearly basis. Given the complexity of the documented data, efficient and reliable technical solutions are required to facilitate the metadata preparation process and to ensure that documentation meets social science standards. For these purposes, we have developed a new editing tool for metadata entry which also allows for off-site use by external project partners.

The presentation will highlight the continuous development of MISSY as a long term DDI case study, benefits of the new MISSY Editor (via a live demonstration) and the cooperation between the MISSY project and Data without Boundaries.

Use Cases and Vocabularies Related to the DDI-RDF Discovery Vocabulary

Thomas Bosch (GESIS - Leibniz Institute for the Social Sciences)

In order to support the discovery of person-level data and its metadata, the Social Science data and the Linked Data communities developed the DDI-RDF Discovery Vocabulary, an ontology of the Data Documentation Initiative. We reuse widely adopted and accepted vocabularies like DCMI, FOAF, ORG, ADMS, PROV-O, SKOS, XKOS, DCAT, and Data Cube to a large extend. We focus on the further development of this ontology, on the description of how other vocabularies are reused reasonably as well as on the description of use cases which are associated with the usage of the DDI-RDF Discovery Vocabulary.

Integrating DDI with Metadata for Qualitative Data: Fully Implementing the QuDEx Standard

Louise Corti (UK Data Archive), Darren Bell (UK Data Archive), Agustina Martinez (Liverpool John Moores University)

The UK Data Archive is a mature archive that has been using DDI and its predecessors for over 40 years and is currently transitioning to a DDI-L compliant data management system.

In this paper we discuss how we have implemented a digital data browsing system for qualitative data based on highly structured data and metadata. The crux of this exciting project has been the incorporation of object and sub-object level metadata using the **QuDEx** metadata schema in addition to DDI study-level metadata, and using the Text Encoding Initiative (TEI) for encoding textual data.

The QuDex schema was initially released in 2006 following a project undertaken by the UK Data Archive and Metadata Technologies. The schema has been updated following testing and has formed the basis for the grander DDI Qualitative Working Group model released only recently. QuDEx enables simple description of collections, data objects, parts of data objects, captures formal relationships between them, and analytical elements such as categories, codes and memos. We will showcase how QuDEx has been implemented in the **UK Quali Bank** browsing system, our use of an XML database (namely Base X) to store and deliver both metadata and textual data, and Solr for powerful searching.

The Medical Research Council Gateway

Philip Curran (MRC Unit for Lifelong Health & Ageing at UCL), Peter Dukes (Medical Research Council), Caroline Shriver (Medical Research Council), Catherine Jones (Science and Technology Facilities Council), Alastair Duncan (Science and Technology Facilities Council), Alex Addyman (MRC Unit for Lifelong Health & Ageing at UCL)

The Medical Research Council (MRC) Research Data Gateway ("the Gateway") is a central repository for metadata from longitudinal and cohort studies. The goals of the Gateway are to support high quality research across a range of studies with facilities to allow researchers to discover, search, and browse study metadata down to the variable level. The Gateway contains study and variable level metadata with information on over 45,000 variables.

In 2010, five major UK cohort studies provided metadata to the Gateway. Original metadata was ingested from a variety of sources. This bespoke ingestion process consumed a significant portion of the original project resources. In 2011 the project successfully investigated using DDI3 as a medium for metadata ingestion.

When evaluating the Gateway in 2012, it became clear that a much more efficient system was needed to allow study teams to publish updated metadata to the Gateway. At the same time, the emerging DDI-L standard was being evaluated by the research community for expressing longitudinal and cohort study metadata, and tools were appearing to support DDI-L use in study host units. The next phase of the MRC Gateway project will make full use of DDI-L for ingesting and exporting studies.

Update on Questasy, a Data Dissemation 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. With the new release, we added two new major features: multilingual support and DDI import. Until recently, Questasy was only able to deal with two languages. Now we added the possibility to have question texts and categories in an unlimited amount of languages. A DDI import (based on version 3.1) was written that can import single wave studies as well as longitudinal studies consisting of multiple waves. The import can deal with standard DDI elements such as question

items, question constructs, variables, code schemes and categories. The existing Questasy DDI export was improved by including more DDI elements in the export and allowing caching of created DDI XML files. These new features allow efficient and robust transfer of questionnaire metadata using the DDI 3.1 standard.

Making Custom DDI 3.1 Exports from the Language Management Utility (LMU)
Edwin de Vet (CentERdata), Eric Balster (CentERdata)

The Language Management Utility (LMU) is a web based translation tool developed at CentERdata. We use it to translate the SHARE questionnaires (a multilingual longitudinal study accross Europe) into more than 20 different languages. We developed a DDI 3.1 export from the LMU database containing the translated texts of questions, interviewer instructions and answer categories. All waves of SHARE present in the LMU database were exported in a single DDI instance. This file was used to import the SHARE metadata into Questasy 4.3 which supports DDI import.

Implementing DDI Profiles at a National Statistical Office: Experiences at INSEE and ABS
Guillaume Duffes (Insee - Institut National de la Statistique et des Etudes Economiques), Simon Wall (Australian Bureau of Statistics)

Practical experience using DDI in large statistical organisations has led to the conclusion that DDI profiles are an essential tool for both modernising the statistical production process and facilitating sharing of technical tools and processes.

Both INSEE (Institut National de la Statistique et des Etudes Economiques) and ABS (Australian Bureau of Statistics) are part of the DDI profiles team (sub-group of the DDI/GSIM - Generic Statistical Information Model - mapping group) under the auspices of UNECE (United Nations Economic Commission for Europe), which aims at creating profiles for particular parts of the DDI lifecycle.

- INSEE has been modelling survey questionnaires using the DDI lifecycle information model for more than two years. INSEE originally used version 3.1, but due to lack of some objects specific to business statistics, such as question grids, has been moving towards adopting version 3.2.
- The ABS has adopted DDI as a key enabling standard for describing metadata, for use in the development of the next generation of statistical infrastructure. The ABS is implementing version 3.1. of the standard, with extensions where necessary.

This presentation will explain the DDI profile experience for Survey Instruments at INSEE, and for Variables at ABS.

RODA's Open-Source Web Platform for DDI
Adrian Dușa, Cosmin Rentea (both RODA - Romanian Social Data Archive)

In the next years, RODA (Romanian Social Data Archive) plans to enhance its capabilities in the field of social data archiving and information management. Our organization is focusing on interoperability with other CESSDA members – both technical and semantic, while respecting national regulations and constraints imposed by data-owners.

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 electronic archive.

The platform has multiple access levels, addressing multiple stages in the metadata and data life-cycle. The web application is available online and has a data browser, importing and exporting of DDI

metadata and various data formats, handles controlled vocabularies, allows meta-data editing and searching, flows for data archiving and processing, and basic statistical analysis and visualization. The architectural overview and current implementation status are presented in order to obtain feedback, to clarify future features (such as federated authentication and authorization, meta-search, RDF / DISCO) and to stimulate debate in the growing DDI community.

Using DDI within a Data Archive - the SND Case Study
Johan Fihn (Swedish National Data Service), Olof Olsson (Swedish National Data Service)

Swedish National Data Service, SND is a service organisation for Swedish research within the humanities, social sciences and health sciences. This diversity of scientific domains creates difficulties when different types of data and metadata standards get involved. SND uses DDI as its primary standard for documentation, however other standards are used as well depending on the situation.

This presentation is about how SND currently handles its data and metadata and how the tools interacts today and how we plan to use DDI in the future. Tools presented in this presentation will cover transformation between standards using XSLT, archive management tools, web services and much more.

Describing Qualitative Surveys at beQuali: a Crosswalk between Documentation, Archiving and Dissemination Metadata
Guillaume Garcia, Sarah Cadorel, Julien Rault (all Centre de Données Socio-Politiques - Réseau Quetelet)

This presentation aims to describe the implementation of DDI 2.5 in the French qualitative archive named beQuali. We will discuss the limitations of DDI for documenting qualitative surveys with regard to describing their methodological context, data and materials, and the archive ordering.

We have applied DDI following conventional practices, but supplemented these with other standards to meet our needs: documenting the specificities of our materials, inter playing with archiving technologies and digitisation process, and data dissemination and harvesting.

Indeed none of the existing metadata formats (DDI, EAD, PREMIS, METS) can fully handle the multiple aspects of each purpose: qualitative dataset description, materials tree view ordering, preservation of metadata and compound object aggregation. For simplicity and performance we had to design an in-house metadata format named MetaJSON that acts as a pivot or switchboard between imported or disseminated data. We are now developing an open source tool that edits and crosswalks this internal format with others standards. This tool would be designed to help researchers in deposing data and to help us to process materials for archiving and dissemination.

How Do We Manage Complex Questions in the Context of the Large-Scale Ingest of Legacy Paper Questionnaires into DDI-Lifecycle? *[full paper]*
Claude Gierl (Centre for Longitudinal Studies, Institute of Education, London), Jon Johnson (Centre for Longitudinal Studies, Institute of Education, London)

The Centre for Longitudinal Studies (CLS) and the CLOSER (Cohorts and Longitudinal Studies Enhancement Resources) project are in the early phase of a massive metadata ingest programme, largely from historic paper questionnaires. The ultimate goal of the project is a Unified Search Platform (USP) which will make the metadata of some of the most significant cohort studies in the world available and searchable online.
Part of the building blocks of the project is an in-house editor used to capture the metadata. While DDI3.2 certainly seems to offer enough flexibility to handle most of the complex or tabular questions encountered in the legacy questionnaires, the finite scope of the editor and the skill level of its

intended users lead us to constrain the DDI3.2 profile we intend to use.

How we design the core structure of this editor and which DDI3.2 profile we implement in order to deal with complex questions has a wide range of repercussions for the rest of the project infrastructure. Our choices will affect issues such as how far we will be able to remain sufficiently true to the original questionnaires, how efficiently the searching of the USP will operate, how easy it will be to assign concepts, mappings and comparison schemes and, more generally, the consistency of the metadata across surveys. These repercussions need to be examined in order to weigh up our preferred options.

Data Service Infrastructure for the Social Science and the Humanities (DASISH): Improving Survey Quality in Cross-national Research

Eric Harrison (City University London, UK)

DASISH is an EU FP7 funded project that brings together the five ESFRI infrastructures within social science and the humanities to provide common solutions to common problems. The five infrastructures are CESSDA, CLARIN, DARIAH, ESS and SHARE. The work to improve survey quality in cross-national research includes three specific task sets, all connected to the development of tools for use in the DASISH survey projects, ESS and SHARE. An aim is that the tools developed under DASISH should also be able to serve other survey projects. The task sets comprise ...

1. a software for improved coding of occupation,
2. a multi-language questionnaire development tool suite consisting of three separate components that will communicate with each other: a questionnaire design documentation tool, a translation management tool and a question/variable data base, and finally ...
3. a fieldwork monitoring system.

This presentation gives an introduction to each of the DASISH task sets, with particular focus on the tasks to be presented in this section: a) the multi-language questionnaire development tool suite, where issues related to common requirements and a common metadata model for the three tools will be discussed, and b) the fieldwork monitoring, consisting of a transportable, standardised system of monitoring employing harmonised metadata files which can aid central field supervision, control and monitoring.

DDI on Rails

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

The metadata portal DDI on Rails is designed to accompany a researcher during his whole research project with cross-sectional or longitudinal data. It allows him to explore the data, personalize his datasets, and publish his results. The software is study-independent and open source, can document data with multiple versions/distributions and the specific characteristics of a longitudinal study, and is easy to use.

The system is currently in beta-testing and will be used for the next generation of our metadata portal SOEPinfo to document the data from the German Socio-economic Panel (SOEP) and its related studies.

The presentation gives an (mostly non-technical) introduction to the system, shows how the DDI standard helped with the implementation, and invites interested participants to join the beta test.

DataForge:SledgeHammer: Your Data/Metadata Management Tool
Pascal Heus (Metadata Technology North America)

At the IASSIST 2013 conference last June, we previewed "DataForge", a new toolkit aiming at addressing practical needs of statistical data managers and researchers, and facilitating the use and adoption of metadata standards such as DDI. At the core of DataForge is "SledgeHammer", a powerful utility for reading/writing data across packages, producing various flavors of metadata such as DDI, computing summary statistics, publishing into databases, and performing other useful operations around statistical data files.

SlegdeHammer was initially released as a freeware command line utility. The package has now come to maturity and will shortly be made available under both a freeware "community" and commercial "Pro" editions. The new version comes equipped with a graphical desktop user interface, making the package more easily accessible to the casual user.

Our EDDI presentation will showcase SledgeHammer core features, describe the differences between the Community and Pro editions, highlight the latest changes (including support for DDI 3.2), and demo various use cases.

For more information, please visit: http://www.openmetadata.org/dataforge/sledgehammer

Data Model and Data Standard - a Happy Marriage?
Oliver Hopt (GESIS - Leibniz Institute for the Social Sciences), Brigitte Mathiak (GESIS - Leibniz Institute for the Social Sciences)

When implementing an application, there are many requirements. Most of them are application-specific, but others arise from wanting to use standards. Implementing the model close to the standard allows for easy data exchange, but may hurt in terms of sub-optimal structure for the actual tasks of the application. Implementing it far from the standard reverses the situation. In this paper, we seek the middle ground. By using a light coupling, we implement a prototype that has both easy data exchange and a flexible structure.

Using Extended Attributes in Data Analysis Software- Controlled Vocabularies, Tools and DDI *[full paper]*
Larry Hoyle (Institute for Policy & Social Research, University of Kansas)

All of the major data analysis software packages now allow some form of user defined extended attributes on variables and most also allow these attributes for the datasets themselves. In each case these attributes can be seen as a pair of strings (attribute name, attribute value). They can also be seen as a subject, predicate, object triple (variable, "has" attribute name, attribute value). This paper explores potential uses of these attributes and suggests directions for developing best practice guidelines for their use.

Colectica 5: A New Generation of Open Metadata Tools
Jeremy Iverson (Colectica), Dan Smith (Colectica)

Colectica is a software suite for managing statistical data. It is based on the DDI metadata standard and provides tools for documenting data, importing metadata from existing sources, and publishing documentation on the Web and in other formats.

The new Colectica 5 integrates feedback from national statistics institutes, university archives, and commercial data collection organizations. This session will highlight the new functionality available in Colectica 5, including: DDI 3.2, support for Eurostat data quality reporting, deep integration with data collection systems, new ways to explore and discover data, and a translatable user interface.

Open Source DDI-L Search
Jannik Jensen (Danish Data Archive)

With a study collection fully documented in DDI-L a handful of advantages for dissemination services and support comes along. DDA has built the infrastructure to facilitate dissemination of DDI-L on the web.

This infrastructure incorporates:

- Multi-faceted search
- Landing page with micro format mark-up prepared for search engine harvesting
- Codebook rendering
- URN resolution

These elements are built into one single platform - on which DDI-L collection deployment is a breeze.

The infrastructure will be published open source and by giving the presentation we will like to invite the DDI-L community to make use of this infrastructure and take part in further development activities.

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

Statistics Denmark and the Danish Data Archive have in collaboration taken on the initiative to form a Danish user group to share and exchange knowledge on DDI. The goal is to promote, facilitate and align usage of DDI in Danish infrastructure using data for research and other kinds of use.

The overall idea behind creating this group is: If we use common terms and understanding of metadata regarding content and technical form - nationally as well as internationally - we can ensure:

- Better dissemination of metadata,
- Increased use of metadata,
- Better quality on products based on common metadata
- Better collaboration and lower costs on development and operation on projects using data across organizational boundaries.

The user group will reach out to Danish organizations using and producing and archiving statistical and research data e.g. social and health care topics.

The presentation will in detail lay out the goals and forthcoming activities of the Danish DDI user group.

Metadata Requirements to Document Data Analyses and Syntax Files in a Virtual Research Environment
Uwe Jensen (GESIS - Leibniz Institute for the Social Sciences)

Metadata supporting collaborative data analyses within a virtual research environment demands more elaboration within the DDI standard. The running 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.

Encoding Provenance of Social Science Data: Integrating PROV with DDI *[full paper]*

Carl Lagoze (University of Michigan School of Information), Jeremy Williams (Cornell Institute for Social and Economic Research - CISER), Lars Vilhuber (Cornell Institute for Social and Economic Research - CISER), William Block (Cornell Institute for Social and Economic Research - CISER)

Provenance is a key component of evaluating the integrity and reusability of data for scholarship. While recording and providing access provenance has always been important, it is even more critical in the web environment in which data from distributed sources and of varying integrity can be combined and derived. The PROV model, developed under the auspices of the W3C, is a foundation for semantically-rich, interoperable, and web-compatible provenance metadata. We report on the results of our experimentation with integrating the PROV model into the DDI metadata for a complex, but characteristic, example social science data. We also present some preliminary thinking on how to visualize those graphs in the user interface.

GSIM and DDI: Working Together

Therese Lalor (UNECE), Guillaume Duffes (INSEE)

The Generic Statistical Information Model (GSIM) is a reference framework of information objects, which enables generic descriptions of the definition, management, and use of data and metadata throughout the official statistical production process. GSIM version 1.0 was released in December 2012.

During 2013, a new project called Frameworks and Standards for Statistical Modernization was sponsored by the High Level Group for the Modernization of Statistical Production and Services (HLG). One of the priorities for this project was to focus on the implementation of GSIM. A task team was commissioned to map GSIM to DDI - initially at a conceptual level. The presentation will show the results of the mapping work, in particular where DDI supports GSIM well and where gaps have been found.

After completing the conceptual mapping, a task team is now working on creating DDI profiles for a number of GSIM objects. It is thought that these profiles will be used to standardise use of DDI in statistical organisations.

Common Statistical Production Architecture

Therese Lalor (UNECE), Arofan Gregory (Metadata Technology)

If the official statistical industry had greater alignment at the business, information and application levels, then sharing software would be easier. The Common Statistical Production Architecture (CSPA) will bring together these existing frameworks (for example, GSBPM and GSIM) in addition to

new frameworks about Statistical Services to create an agreed top level description of the 'system' of producing official statistics.

CSPA is a statistical industry reference architecture for statistical production. Its focus is to support the facilitation, sharing and reuse of Statistical Services both across and within statistical organizations. Statistical Services are defined and have invokable interfaces that are called to perform business processes.

Depending on what information is being represented in practice, DDI and SDMX are expected to provide the basis for CSPA implementation specification. The CSPA implementation specification for Statistical Services will specify what schema should be used for representing a particular GSIM information object, and exactly how the chosen schema will be applied for the particular purpose.

This work will inform the work being undertaken under the auspices of UNECE to improve GSIM and create DDI profiles.

Historization and Versioning of DDI-Lifecycle Metadata Objects - Findings in the STARDAT Project

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

The GESIS Data Archive for the Social Sciences provides high quality data and documentation of survey datasets. With the project STARDAT the data archive integrates existing DDI-Codebook software tools to one open source application suite supporting DDI-Lifecycle.

The talk reports about the experiences in extracting and implementing a Java-based extensible object model with object-relational persistence focused on historization and versioning of metadata in the sense of DDI 3.1. Using the example of the archival use case of importing metadata from statistical analysis systems like SPSS I show the possibilities of repository history, object history, object difference comparison, publishing metadata objects and staged reuse of metadata. Finally I discuss the implications of DDI schema changes on the implementation related to coming soon DDI 3.2.

Standardized Quality Declarations with DDI, SDMX, and Colectica

Mogens Nielsen (Statistics Denmark), Jeremy Iverson (Colectica), Dan Smith (Colectica)

Eurostat requires its members to report about quality using SDMX reference metadata standards. These standards align well with the new support for quality declarations in DDI 3.2. Statistics Denmark and Colectica worked together to build a comprehensive system for managing quality information and for reporting this information to Eurostat in the required formats.

The implementation of the SDMX reference metadata reporting standards ESMS and ESQRS at Statistics Denmark forms part of a general documentation system where the generalized quality concept inventory, SIMS (Single Integrated Metadata Structure), will form the basis for reporting ESMS, ESQRS and other reporting standards. The general documentation system is based on standards including GSBPM, DDI, SDMX and others, and uses standard software: Colectica. The general documentation system integrates variables, classifications, concepts and methodology. The Methodology element includes SIMS and other specific information used in a Danish context. The statistical concepts used in SIMS (e.g. Labor Status) are implemented in a way that is integrated with DDI and SDMX, reusing concepts across all GSBPM phases in Statistics Denmark. The use of standards is expected to be gradually harmonized with the use of standards in Eurostat and other organizations working together with Statistics Denmark in order to benefit from using common understanding and "common language".

Mapping DDI-RDF Discovery - a Hackathon Story

Olof Olsson (Swedish National Data Service), Jannik Jensen (Danish Data Archive)

Late in October 2013 a group of developers decided to meet in Copenhagen and hack on the DDI Lifecycle and DDI Codebook to transform this into RDF the DDI Discovery Vocabulary way.

The XSLT mappings are useful for organizations who want to make their data documentation (based on DDI XML) available in RDF.

The DDI-RDF Discovery Vocabulary provides a common easy to use vocabulary describing essential parts of both the Lifecycle and Codebook version of DDI.

During the hackathon we also tested some of the use cases of the vocabulary specification by running SPARQL queries against the transformed collections.

This presentation will report on findings and results from this hackathon as well as lessons learned from this way of facilitating work processes.

New Tools for Complex Surveys: the DASISH Questionnaire Design Documentation Tool and Question Variable Data Base

Hilde Orten (NSD - Norwegian Social Science Data Services), Håvard Bakkmoen (NSD - Norwegian Social Science Data Services), Yvette Prestage (City University London), Sally Widdop (City University London)

Two new tools to enhance the process of documenting questionnaire design and development and survey variables are currently being developed under the Data Service Infrastructure for the Social Sciences and Humanities (DASISH) project. The primary aim of the questionnaire design documentation tool (QDDT) and the question variable data base (QVDB) is to assist large-scale survey projects in the processes related to questionnaire design and development, as well as in data and metadata production, archiving and dissemination. Secondly, the tools aim to serve researchers and students in exploring metadata from existing projects, or in designing new research.

Initially planned for the European Social Survey (ESS), which is a complex international biennial cross-sectional survey, the tools will also be designed to serve other projects, to interoperate with other systems and tools, and to act as reusable models for other tools. Both the QDDT and QVDB tools will be modelled based on DDI-Lifecycle.

This presentation gives an introduction to each of the two tools, touching on purposes, usages, requirements and system models. The main focus will be on how we work with DDI at a conceptual level, as well as on the planned usage of DDI elements that are new in DDI 3.2 (public review version).

Building a Harmonized Data Market for Longitudinal Data with MIDUS and DDI

Barry Radler (University of Wisconsin - Madison), Jeremy Iverson (Colectica), Dan Smith (Colectica)

Researchers who wish to use data from longitudinal studies or to replicate other's research must currently navigate thousands of variables across multiple waves and datasets to answer simple analysis questions. A tool that allows researchers to create documented and citable data extracts that are directly related to their queries would allow more time to be spent on public health research questions instead of data management.

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 unique blend of social, health, and biomarker data collected over several decades.

In late 2013, the United States National Institutes of Health funded MIDUS to create a DDI-based, harmonized data extraction system. This tool 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.

REST API for Nesstar. Present Your Data and Metadata in New Ways - Fast.

Ørnulf Risnes (NSD - Norwegian Social Science Data Services)

NSD will continue to release Nesstar as an integrated solution.

However, there is an increasing demand from the user-community for building alternative applications for data publishing, dissemination and analysis, and to integrate Nesstar with in-house production systems and dissemination channels.

Responding to these changes in the data and metadata landscapes, NSD started developing a Java based Nesstar API in 2011. The Nesstar API has matured and stabilized since then, and is continuously being put to use in new applications, even in .Net environments.

However, since the birth of the Java-based API there has been a demand for a platform-agnostic version of the API. NSD has now addressed this demand, and released the Nesstar REST API.

The Nesstar REST API uses JSON as exchange format, and currently supports:

- Metadata navigation and –delivery
- Microdata analysis (cross-tabulation, correlation, regression)
- Cubes/aggregate data

This talk will demonstrate Nesstar as an increasingly programmable DDI-based dissemination platform, focusing on the new REST API. It will show examples of present websites and services that use the Nesstar REST API in combination with libraries like jQuery and HighCharts to enrich static and dynamic web pages with content and functionality from Nesstar Servers.

A New CESSDA Portal for European Research Data Discovery

Ørnulf Risnes (NSD - Norwegian Social Science Data Services), John Shepherdson (UK Data Archive), Pascal Heus (Metadata Technology)

Under the Data without Boundaries Project (http://www.dwbproject.org), work packages 8 and 12 are are focusing on the development of a one-stop shop data discovery portal for research data held by the CESSDA archives, National Statistical Institutes (NSIs), and other agencies across Europe. The portal implementation is a joint collaborative effort led by the Norwegian Social Science Data Services (NSD), the UK Data Archive (UKDA), and Metadata Technology (MT), with the assistance from other agencies.

By the end of the project, a prototype web based portal will be made available, with content harvested from an initial group of participating metadata providers. The primary objective is to empower researchers with the ability to discover data sets of interest through the exploration of study and variable level metadata.

The portal will further be equipped with a metadata provider section facilitating metadata quality assurance, usage analysis, and content management. Machine-actionable interfaces will be available to authorized external applications, enabling interaction with the portal´s search index and metadata repositories.

This presentation will provide an overview of the portal infrastructure and implementation strategy, including architecture, metadata model, technologies, harvesting mechanisms, use-cases for data discovery, metadata quality issues, and secondary use via portal APIs.

Harmonizing Between Different Agencies Using DDI Profiles

David Schiller (IAB - Institute for Employment Research, Germany), Ingo Barkow (German Institute for International Educational Research), Marcel Hebing (DIW Berlin - German Institute for Economic Research)

Developing software to support the DDI-L standard provides a challenge to agencies. The DDI-L standard is in most cases much too vast for most individual tool requirements. DDI-L 3.1 contains more than 900 main nodes in its schema while a survey software most likely only needs 50-60 of them. The idea is therefore to use DDI Profiles to specify a sub-set of requirements for the individual purpose. A software solution for surveys could therefore use two different DDI Profiles to express its compatibility to other similar software, e.g. the profiles "Survey Design" and "Data Collection". For other parts of the lifecycle which are relevant to other agencies similar DDI Profiles can be specified (e.g. "Administrative Data", "Processed Data" and "Transaction Data"). During the design of software tools like Rogatus (DIPF, TBA21, OPIT and IAB) and DDI on Rails (SOEP) these issues encountered as the software is supposed to be compatible with other tools like Colectica and Questasy. Therefore the process to create DDI Profiles for this harmonization has begun. Furthermore discussions with ABS to support their GSIM-based DDI Profiles are on the way as well.

Proposing a Metadata Solution over Multiple RDCs in the German Context

David Schiller (IAB - Institute for Employment Research, Germany), Ingo Barkow (German Institute for International Educational Research)

There is a wide ranch of research data available in Germany. Within the last decade a great number of new Research Data Centre (RDC) originated, offering a variety of different information for scientific research. Currently 25 RDCs are accredited under the umbrella of the German Data Council (RatSWD). Having does data available is a good thing for researchers; at the same time finding the best data for a given project is not easy at all. Currently they have to look at data documentations in different formats and spread over 25 homepages. Researchers need a single point of access and a structured way to search the available data. Information about datasets, research potential of variables and about how to access data are important in that regard. A reliable and machine-readable standard used by all RDCs would enable the use of software tools that allow researchers to effectively discover the richness of research data available in Germany. The case Germany is only an example for the need for a standard like DDI and it shows that goal of having an effective way to explore the landscape for research data is not yet reached. DDI still must be used by more data providers.

The Questionnaire Generation in the French National Statistical Institute

Eric Sigaud (INSEE - French NSI)

INSEE runs many surveys and needs to deliver them on different media: paper questionnaires, specialized tools like Blaise, or web forms. In an effort of rationalization, INSEE recently initiated the construction of a generic platform for survey data collection based on a central metadata repository.

One of the goals is to implement a questionnaire generator driven by active DDI metadata, offering different output formats based upon the same specifications, a "living documentation" of the survey instruments, and automation of the questionnaire building process.

The first development steps of this generator were based on two operational use cases.

First, we had to provide an electronic collection for the ESA - the main French structural business survey - which had only paper forms. We developed an XML process taking a DDI 3.1 questionnaire description and producing personalised OpenDocument forms ready to be downloaded and filled offline.

Second, we had to provide a web collection for CIS - a European business survey about innovation. With an upgrade of the generator, we were able to create web forms, using XForms semantics, based upon DDI 3.2 specifications of the questionnaire.

The presentation will outline the generator architecture and demonstrate the results obtained so far.

Generating Blaise Surveys from the Data Documentation Initiative's Metadata Standard using Colectica

Dan Smith (Colectica), Jeremy Iverson (Colectica)

The DDI-Lifecycle (DDI-L) metadata, described in XML, includes information about questions, code lists, variables, questionnaires, instrument flow, and conceptual components. These metadata items can be precisely identified and versioned to track changes made during development and over different waves of a survey. This paper details how the precise descriptions of metadata items found in DDI-L can be leveraged to generate complete Blaise questionnaires, a process that helps reduce programming errors, creates an audit trail, and saves valuable programmer time. A process for converting DDI-L to Blaise questionnaires will also be detailed using a DDI-L editing tool named Colectica Designer.

Documenting a survey is a time intensive process prone to errors and often done at the end of the survey design process where some knowledge may have already been lost. In addition to being used in the Blaise code generation, the DDI-L metadata also recorded the entire questionnaire design and implementation process from the beginning. This allows documentation and code books to be automatically generated based off the same DDI-L XML files. The standardized XML format also allows sharing with other organizations, enabling reuse of questionnaire sections across surveys.

Using DDI to Automate Blaise Instrument Generation

Simon Wall (Australian Bureau of Statistics)

The Australian Bureau of Statistics (ABS) is currently transforming the way it undertakes its business. A substantial component of this transformation is the development of the next generation of statistical infrastructure, for which DDI has been adopted as a key enabling standard for describing metadata.

Currently Blaise instruments are developed through hand coding by programmers, a simple but lengthy and expensive process. The ABS aims to give power back to subject matter experts by giving them a seamless instrument design interface to specify their collection instruments without the need for routine manual intervention by programmers. To this end, the ABS is developing a tool which allows subject matter experts to specify instruments in business terminology, and capture the specifications in DDI. We are also developing a service which automates Blaise instrument generation using that DDI.

This presentation will outline the development and experience at the ABS, so far, of automatically generating Blaise instruments and the challenges encountered.

Developing a Transportable, Standardised System of Monitoring Employing Harmonised Metadata Files Which Can Aid Central Field Supervision, Control and Monitoring

Sally Widdop (City University London, UK), Johanna Bristle (Max Planck Institute for Social Law and Social Policy), Yvette Prestage (City University London, UK)

As part of the DASISH project, researchers from the European Social Survey (ESS) and the Survey of health, ageing and retirement in Europe (SHARE) are working together to develop a computerised, mobile fieldwork management system that could be utilised on future rounds of the ESS and other face-to-face surveys. SHARE's existing bespoke sample management system provides researchers

with real time data on interviewer contact attempts and updated sample information. We are currently developing a similar tool, which can be used on any face-to-face survey. Installed on a portable device like a smart phone, tablet or personal digital assistant (PDA), the fieldwork management system ('slim sms') will employ harmonised metadata files and produce timely aggregate statistics on fieldwork progress and characteristics, which could be used to aid central field supervision, control and monitoring. This presentation will give an overview of the sample management system utilised by SHARE and summarise the key features that will be required to adapt the tool for other surveys. In addition, it will seek to explore how the new fieldwork management system application can be made compatible with DDI and/or DDI-L.

Metadata Management for Research Data with DBKfree
Wolfgang Zenk-Möltgen (GESIS - Leibniz Institute for the Social Sciences)

DBKfree is the open source version of the Data Catalogue software DBK that the GESIS Data Archive for the Social Sciences uses to manage the documentation of research data in the archive. It can be used to document internal workflows regarding long-term archiving as well as building a web site with a catalogue for users to search and find archived research datasets. With DBKfree, a broad study description may be generated that is compliant with and may be exported to DDI-Codebook and DDI-Lifecycle. Integrated into DBKfree is the possibility to register DOI Names with da|ra, the GESIS service for persistent identifiers. The metadata schema of DBKfree is compliant to da|ra as well as to DataCite. Another function of DBKfree is the download possibility or the shopping cart order for documents and datasets, which may be configured in multiple ways. The presentation will show a general overview about DBKfree from a user's and a provider's view, and highlight new developments in version 2.0. Special attendance will be given to supported metadata formats like DDI.

Posters / Software Demonstrations

Early Implementation of DDI Lifecycle on the Integrated Version of the Cornell National Social Survey

Florio Arguillas (Cornell Institute for Social and Economic Research - CISER), Jeremy Williams (Cornell Institute for Social and Economic Research - CISER), William Block (Cornell Institute for Social and Economic Research - CISER), Warren Brown (Cornell Institute for Social and Economic Research - CISER)

The annual Cornell National Social Survey (CNSS), conducted by the Survey Research Institute (SRI), polls adults aged 18 and over on a wide range of current public policy topics such as immigration, environment, employment, technology, politics, family, and economy, among many others.

As the designated repository of CNSS datasets, the Cornell Institute for Social and Economic Research (CISER) curated, created, and made available a public use version of each dataset. To further enhance the value of the surveys to researchers, CISER created and curated an integrated version of the public use files by harmonizing the annual datasets and ensuring consistent variable types, names and value-name pairs for variables that were asked in multiple years; and adding a critical variable for integration which was absent in each dataset--survey year.

This paper details our early implementation of the DDI Life Cycle on the integrated version of the CNSS noting, among others, its source datasets (the curated public use version), their recent versions (the further curated public use version of the datasets), its integrated version, metadata record creation, and the creation of documentation compliant with the DDI lifecycle specification.

The New Microdata Information System (MISSY): Integration of DDI-based Data Models, an Open-Source Software Architecture, and Independent Persistence Service Implementations

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

MISSY is a DDI-based system which provides structured metadata for microdata from official statistics online. Within the current project the content will be enlarged to cover integrated microdata from European statistics. This also implicates an enhancement of the technological basis including the development of a new metadata preparation tool (MISSY Editor). In addition to detailed variable level metadata the new MISSY Editor allows comprehensive study level documentation which includes country specific information on a yearly basis. The DDI-RDF Discovery Vocabulary is an ontology of the Data Documentation Initiative in the world of Linked Open Data. We reuse this vocabulary as the core data model, since the discovery use case is the most important one for our purposes. We extend this core data model by a MISSY specific data model which fulfills further requirements not met by the DDI-RDF Discovery Vocabulary. We demonstrate how we implement these data models and how we persist them using multiple formats such as DDI-XML, DDI-RDF, and relational databases. We also offer a live demonstration of the new MISSY Editor.

DataForge:SledgeHammer: Your Data/Metadata Management Tool

Pascal Heus (Metadata Technology North America)

At the IASSIST 2013 conference last June, we previewed "DataForge", a new toolkit aiming at addressing practical needs of statistical data managers and researchers, and facilitating the use and adoption of metadata standards such as DDI. At the core of DataForge is "SledgeHammer", a powerful utility for reading/writing data across packages, producing various flavors of metadata such

as DDI, computing summary statistics, publishing into databases, and performing other useful operations around statistical data files.

SlegdeHammer was initially released as a freeware command line utility. The package has now come to maturity and will shortly be made available under both a freeware "community" and commercial "Pro" editions. The new version comes equipped with a graphical desktop user interface, making the package more easily accessible to the casual user.

Our EDDI presentation will showcase SledgeHammer core features, describe the differences between the Community and Pro editions, highlight the latest changes (including support for DDI 3.2), and demo various use cases.

For more information, please visit: http://www.openmetadata.org/dataforge/sledgehammer

Colectica for Excel: Increasing Data Accessibility using Open Standards
Jeremy Iverson (Colectica), Dan Smith (Colectica)

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 add-in, 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 add-in 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 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.

SPSS and DDI-L
Jannik Jensen (Danish Data Archive)

In the move to DDI-Lifecycle as data documentation standard for DDA (Danish Data Archive) we have chosen SPSS statistical format as an intermediate data format. We will like to share findings and lessons learned in curating/ingesting statistical files originally in or transformed into SPSS. We will present the difficulties we have discovered and what we have been done to work around these in the curation process as well as on the ICT side of the DdiEditor.

From Curation to Publication of DDI-L Metadata
Jannik Jensen (Danish Data Archive), Anne Sofie Fink (Danish Data Archive)

The DDA DdiEditor was developed as a generic suite of tools to produce DDI-Lifecycle metadata for survey data sets. The DdiEditor supports curation of variables, questions, codes, categories, instrumentation, universes and concepts as well as linkage of these meta-elements.

In 2012 the DdiEditor was been supplemented by an indexing platform allowing the user to search (simple or advanced) in study description questions, variables, categories, universes and/or concepts. The study level description is publish as landing pages allowing the user interactively to move from

one study description to another by clicking on a metadata element e.g. a keyword. Similarly the traditional codebook presenting questions, variables, universes, concepts and instrumentation has be enriched by interactivity and graphics for descriptive statistics for all variables.

All metadata are published in a DDI-L XML format ready to import for external search engines and other services. Every study is given a persistent identifier in form of a DOI (digital object identifier).

In order to be able to offer the user the possibility of online analysis and download DDI-L metadata is converted to DDI-C and exported to Nesstar-based services.

We would like to invite more organisations (that is other than DDA) to try out the DdiEditor and indexing platform as a tool for producing and disseminating DDI-L metadata.

Modelling an Evolving, Geospatial SKOS Code List

Ørnulf Risnes (NSD - Norwegian Social Science Data Services), Ole Voldsæter (NSD - Norwegian Social Science Data Services)

The success of Linked Data and Metadata relies on canonical concepts and code lists that enable data from different sources to be connected and compared.

As DDI continues to build bridges to semantic web resources and technologies, generic models for representing complex and evolving structures become increasingly relevant to the DDI standard and related tools.

But how can we create code lists that accurately describe rapid changes in sets of administrative regions?

This talk describes a general solution to the problem of creating a SKOS code list that can represent and keep track of administrative regions that change, merge or split over time.

It will also demonstrate how SPARQL-queries can be used to extract knowledge about a region at given time-point, its lifespan and its relationships with other regions.

Managing and Sharing Data within the Collaborative Research Center SFB 882 in Practice: Data Documentation within a Virtual Research Environment for the Social Sciences.

Johanna Vompras (University of Bielefeld LibTec, Library Technology and Knowledge Management), Christoph Broschinski (Collaborative Research Center (SFB 882), Data and Information Infrastructure Project)

With the promotion of a Virtual Research Environment (VRE) by the INF project (integral part of the SFB 882) advisory and developmental services in the domain of information infrastructure are provided. The VRE combines both a work and a project-specific research platform. The work platform bundles various tools for administration, project management, and time- and location-independent collaboration in a single environment adapted to researcher's specific working processes. The research platform combines data management with further developments of social science methodologies - with a strong focus on scientific information handling (like archiving, re-use or sharing of data sets) and standardized data documentation. Our infrastructure is built upon on a web-based cloud framework (ownCloud) and provides researchers functionalities as integration of shared storage devices, calendar synchronization, and collaborative editing tools.

In this demonstration, we will practically 1) show up the technical challenges in building-up a VRE for the social sciences and 2) present the current state of our development of the DDI documentation component. It is implemented as an "ownCloud App" and supports an easy-going documentation across several stages of the research data life cycle - by including the collected, processed, and already annotated data within the VRE.

Easy DDI Organizer (EDO) - Open Source Survey Planning and Metadata Management Software

Yuki Yonekura (Center for Social Research and Data Archives, Institute of Social Science, The University of Tokyo)

Easy DDI Organizer (EDO) is software for survey planning and metadata management. It enables users to record survey metadata along with data lifecycle such as study purpose, sampling procedure, mode of data collection, questions, question sequence, variable descriptions, and bibliographic information. It also supports importing variable level metadata from SPSS files, importing metadata from DDI documents, exporting several documents such as codebook, questionnaire DDI documents. These features will be demonstrated in the poster session. Source code and documentations will be distributed through project homepage, although it is under preparation now. Information on project homepage will be announced in the session.

Concurrent Tutorials

Learn How to Use the Nesstar REST API. Use Javascript and HTML to Bring Your DDI Metadata and the Power of Nesstar to Your Website

Ørnulf Risnes (NSD - Norwegian Social Science Data Services)

The Nesstar REST API builds upon the Nesstar API, but it is platform agnostic and even easier and faster to use.

The REST API supports metadata navigation and delivery. It also supports tabulation, correlation, regression as well as cube functionality such as querying, slicing and dicing.

This hands-on tutorial will get you started using the Nesstar REST API, and show how you can put your Nesstar based data and metadata to re-use in minutes.

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 Nesstar REST API before the tutorial.

Bring your laptop. A Git client, a Java RE and Apache Maven will also be needed. No particular editor or other tools are required.

Data and Metadata Management Using DDI

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

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?