EDDI EUROPEAN DDI USER CONFERENCE

gesis
Leibniz Institute
for the Social Sciences

DDI®

IZA IDSC
INTERNATIONAL
DATA SERVICE CENTER

8th Annual European DDI User Conference (EDDI16)

December 6-7, 2016, Cologne, Germany

Hosted by GESIS – Leibniz Institute for the Social Sciences

http://www.eddi-conferences.eu/eddi16

Schedule and Program with Abstracts

Final version as of November 29, 2016

The conference will start on Tuesday, December 6 at 9:00 and will end on Wednesday, December 7 at 16:30.

Tutorials will take place on Monday, December 5 from 9:00 to 17:00.

Side meetings will take place on Monday December 5, Wednesday December 7, Thursday December 8, and in the week after the EDDI.

Contents

Venue

Stadthotel am Römerturm
Sankt-Apern-Straße 32, 50667 Cologne, Germany

Registration	Lobby
Main conference	Rooms Kolping and Babilon, Stadthotel am Römerturm
Tutorials	GESIS – Leibniz Institute for the Social Sciences Unter Sachsenhausen 6-8, 50667 Cologne Rooms West I, Ost/East, and Schulungsraum/Training Course
Side meetings	GESIS – Leibniz Institute for the Social Sciences Unter Sachsenhausen 6-8, 50667 Cologne Rooms West I, West II, and Ost/East
Conference Dinner December 6, 2016	Ludwig im Museum Heinrich-Böll-Platz, 50667 Cologne
Informal Get-Together	See detailed schedule below

Venue

Schedule

Sunday, December 4, 2016

19:00	**Informal Get-Together**
	Location: Bier-Esel, Breite Straße 114, 50667 Cologne

Monday, December 5, 2016

8:30	Starting Registration - Location: GESIS Lobby		
9:00 - 10:30	**Concurrent Tutorials**		**Side Meeting**
	Concurrent T1 (full day) **Data Documentation**	Concurrent T2 (half day) **Reusing and Sharing Metadata**	
	Location: GESIS West I	Location: GESIS Ost/East	
	Document Questionnaires and Datasets with DDI: A Hands-On Introduction with Colectica (abstract)	**Documenting Panel Data Using DDI** (abstract)	
	Jeremy Iverson, Dan Smith (both Colectica)	Knut Wenzig (German Institute for Economic Research (DIW Berlin) / German Socio-Economic Panel (SOEP))	
	Block 1 of T1	Block 1 of T2	
10:30 - 11:00	**Break** - Location: GESIS Cafeteria (basement)		
11:00 - 12:30	Block 2 of T1	Block 2 of T2	
12:30 - 13:30	**Lunch** - Location: GESIS Cafeteria (basement)		
13:30 - 15:00	Concurrent T1 (continued) Location: GESIS West I	Concurrent T3 (half day) **Questionnaire Capture** Location: GESIS Schulungsraum/Training Course	Meeting M1 (half day) **Big Data Europe** Location: GESIS Ost/East (closed meeting) Organizer: Ivana Versic 14:00-17:30
	Block 3 of T1	**Have "Questions"? We Have a Structure for Them** (abstract) Wendy Thomas (University of Minnesota Minnesota Population Center), Jon Johnson (UCL Institute of Education), Guillaume Duffes (Insee - National Institute of Statistics and Economic Studies) Block 1 of T3	
15:00 - 15:30	**Break** - Location: GESIS Cafeteria (basement)		
15:30 - 17:00	Block 4 of T1	Block 2 of T3	
18:00	**Informal Get-Together**		
	Location: Stadtgarten, Venloer Straße 40, 50672 Cologne		

Tuesday, December 6, 2016

8:15	**Starting Registration.** Location: Stadthotel Lobby, Registration Desk
9:00 - 10:15	Conference Plenary P1: Welcome and Keynote Location: Stadthotel Room Kolping Chairs: Jon Johnson and Joachim Wackerow (both EDDI Conference Co-Chairs)

	Welcome Christof Wolf (President of GESIS – Leibniz Institute of the Social Sciences)
	Keynote: Metadata: Foundation, Philosopher's and Rosetta Stones (abstract) Keith Jeffery (past Director IT at STFC Rutherford Appleton Laboratory)

10:15 - 10:45	**Break** - Location: Stadthotel Cafeteria
10:45 - 12:15	**Concurrent Conference Sessions**

Concurrent A1: **Questionnaire Capture** Location: Stadthotel Room Kolping Chair: Wendy Thomas	Concurrent A2: **CESSDA** Location: Stadthotel Room Babilon Chair: Mogens Grosen Nielsen
Archivist & Mapper: Simplifying and Modernising Questionnaire Entry (abstract) Will Poynter	**A Capability Development Model for Assessing and Improving Distributed Infrastructures and their Services.** (abstract) Mike Priddy, Trond Kvamme, Marion Wittenberg
Questionnaire Generator (abstract) Guillaume Duffes	**The Reference Model for Social Science and Humanities Data Infrastructures: towards a common architectural model for Research Infrastructure architectures.** (abstract) Mike Priddy, Maarten Hoogerwerf, John William Shepherdson, Johan Fihn
Rich Metadata from the Start (abstract) Oliver Hopt	**Is this Metadata Management Tool of Any Use? Extending CESSDA's Software Maturity Matrix to the DDI Domain** (abstract) John William Shepherdson

12:15 - 13:30	**Lunch** - Location: Stadthotel Restaurant

Tuesday, December 6, 2016 (continued)

13:30 - 14:45	**Concurrent Conference Sessions**	
	Concurrent B1: **Questionnaire Capture** Location: Stadthotel Room Kolping Chair: Kerrin Borschewski	Concurrent B2: **Harmonisation** Location: Stadthotel Room Babilon Chair: Oliver Hopt
	The DASISH Questionnaire Design and Documentation Tool – Functionalities and Examples from the Tool (abstract) Benjamin Beuster, Hilde Orten	**Enhancing the Quality of Metadata with DDI-L in a Transitioning DDI-C Environment** (abstract) Alina Danciu, Alexandre Mairot
	Question Banks, Reusability, and DDI 3.2 (abstract) Dan Smith	**Lessons from a Large-Scale Metadata Archiving Team Using DDI-Lifecycle** (abstract) Hayley Mills
	Steps towards a Single Point of Access for Survey Questions across Europe: The Euro Question Bank Project (abstract) Wolfgang Zenk-Möltgen, Azadeh Mahmoud Hashemi	**The e-Lab Metadata Registry: Supporting Harmonisation Using DDI, OWL and Reporting Guidelines** (abstract) Philip Couch, Chris Munro
14:45 - 15:15	**Break** - Location: Stadthotel Cafeteria	
15:15 - 16:15	**Concurrent Conference Sessions**	
	Concurrent C1: **Questionnaire Capture** Location: Stadthotel Room Kolping Chair: Jon Johnson	Concurrent C2: **Harmonisation** Location: Stadthotel Room Babilon Chair: Johan Fihn
	Using DDI in Harmonizing Longitudinal Data for Open Online Analyses (abstract) Juha Samuel Ranta, Markus Tuominen	**The Library of Online Harmonisations: The SERISS Data Harmonisation Platform** (abstract) Martin Friedrichs, Kristi Winters
	Assessment of the Swiss National Objectives in Education (abstract) Ingo Barkow, Catharina Wasner	**Building a Better Search and Discovery Platform Using Elasticsearch** (abstract) Stefan Jakobsson, Olof Olsson, Johan Fihn, Pablo Millet
	Discussion: Questionnaire Capture / Re-usable Questions (abstract)	**Towards an Open Infrastructure for Supporting Relationship Discovery Between Scholarly Assets** (abstract) Chris Munro, Philip Couch, Jon Johnson
16:15 - 16:30	**Break** - Location: Stadthotel Cafeteria	

Tuesday, December 6, 2016 (continued)

16:30 - 17:30	Concurrent Conference Sessions	
	Concurrent D1: **DDI-Related Publications** Location: Stadthotel Room Kolping Chair: Michelle Edwards	Concurrent D2: **Reusing and Sharing Metadata** Location: Stadthotel Room Babilon Chair: Iris Alfredsson
	Publishing DDI-Related Topics - Current Status and Related Working Group of the DDI Alliance (abstract) Knut Wenzig	**Controlled Vocabularies for CESSDA** (abstract) Anne Etheridge
	Publishing DDI-Related Topics - Advantages and Challenges of Creating Publications (abstract) Joachim Wackerow	**Creating a DDI-Lifecycle based Metadata Schema for the GESIS Data Archive** (abstract) Esra Akdeniz, Wolfgang Zenk-Möltgen
	Discussion: Publishing DDI-Related Topics (abstract)	**The Past, Present and Future of Geocoded Survey Data at the GESIS Data Archive** (abstract) Stefan Müller, Wolfgang Zenk-Möltgen, Stefan Schweers
18:30	**Conference Dinner**	
	Drinks from 18:30, dinner served at 19:00 Location: Ludwig im Museum, Heinrich-Böll-Platz, 50667 Cologne	

Wednesday, December 7, 2016

8:30	Starting Registration. Location: Stadthotel Lobby, Registration Desk

9:00 - 10:00	**Conference Plenary P2: Keynote**
	Location: Stadthotel Room Kolping
	Chair: Alexia Katsanidou (Head of GESIS Data Archive for the Social Sciences)

	Welcome
	Jared Lyle (Executive Director of the DDI Alliance)
	Keynote: Funding, Policies, Community Building - Data Sharing from a Funders Perspective (abstract)
	Stefan Winkler-Nees (DFG - German Research Foundation)

10:00 - 10:30	**Break** - Location: Stadthotel Cafeteria

10:30 - 12:00	**Concurrent Conference Sessions**

Concurrent E1:	Concurrent E2:
Issues in Building Infrastructure	**Official Statistics**
Location: Stadthotel Room Kolping	Location: Stadthotel Room Babilon
Chair: Barry Radler	Chair: William Block
Lessons Learnt From the Development of a DDI-Based Discovery Platform (abstract) Jon Johnson	**The Picasso Project** (abstract) Kathryn Stevenson, Flavio Rizzolo
Sharing Metadata Using the DDI "Standard" – 20 Years Later - How are we doing? (abstract) Timothy Parsons, Brandon Kowalski, Michelle Edwards	**Documenting Microdata from Official Statistics – Benefits of a DDI-Based Backend Infrastructure** (abstract) Jeanette Bohr, Matthäus Zloch
	Fulfilling User-Needs and Improving Quality and Efficiency through Statistical Standards and Reuse of Common Metadata (abstract) Mogens Grosen Nielsen

12:00 - 13:30	**Lunch** - Location: Stadthotel Restaurant

Poster and Software Demonstrations

While lunch time - Location: Stadthotel Room Kerpen - Chair: Kerrin Borschewski

The DASISH Questionnaire Design and Documentation Tool (QDDT) – A Business Tool for Capturing the Questionnaire Design Process (abstract)
Benjamin Beuster, Hilde Orten

CLOSER's Principles for Large-Scale Metadata Archiving Using DDI-Lifecycle (abstract)
Hayley Mills, Will Poynter

Developments and Perspectives in DDI Implementation Projects at SSJDA (abstract)
Akira Motegi

DDIR: An R Package for Handling DDI Files (abstract)
Yasuto Nakano

Building a Better Search and Discovery Platform Using Elasticsearch (abstract)
Stefan Jakobsson, Olof Olsson, Johan Fihn, Pablo Millet

Re-Purposing DDI-L in the Production Process (abstract)
Andrew Peters

Wednesday, December 7, 2016 (continued)

13:30 - 15:00	Concurrent Conference Sessions	
	Concurrent F1: Colectica and Blaise Location: Stadthotel Room Kolping Chair: Ingo Barkow	**Concurrent F2:** Research Infrastructure Location: Stadthotel Room Babilon Chair: Catharina Wasner
	Blaise, Colectica and DDI (abstract) Lon Hofman	**Documenting and Re-Using of DMP Information – The Distinct Role of DDI** (abstract) Uwe Jensen, Sebastian Netscher
	What's New in Colectica 5.3 (abstract) Jeremy Iverson	**An Infrastructure for Stimulating Researchers to Deposit Data** (abstract) Johan Fihn
	The MIDUS-Colectica Portal (abstract) Barry Radler	**"Beam us up Scotty" - The CESSDA Metadata Management Core Metadata Model Goes into Round Two** (abstract) Kerrin Borschewski, Wolfgang Zenk-Möltgen, Julia Hermann
15:00 - 15:30	**Break** - Location: Stadthotel Cafeteria	
15:30 - 16:30	Conference Plenary P3: Panel Discussion, Report on DDI Specifications, and Outlook Location: Stadthotel Room Kolping Chair: Mari Kleemola	
	Panel Session: Re-Use of Software and Administered Metadata (abstract) **Panelists:** William Block (CISER - Cornell Institute for Social and Economic Research), Barry Radler (University of Wisconsin), Mogens Grosen Nielsen (Statistics Denmark), John Shepherdson (UK Data Service, University of Essex)	
	DDI Specifications - Recent Developments Wendy Thomas, Jon Johnson, Joachim Wackerow (all Technical Committee of the DDI Alliance)	
	Announcement of 5th NADDI, Invitation to EDDI17 and Goodbye CISER/NADDI team, Jon Johnson and Joachim Wackerow (both EDDI co-conference chairs), and Next Year's Host (name will be disclosed in session)	
17:00 - 18:00	Side Meeting M2: **Active Data Management Plans** Public Meeting Location: Stadthotel Room Kolping Organizer: Uwe Jensen	
18:00	**Informal Get-Together** Location: Brauerei Päffgen, Friesenstraße 64-66, 50670 Cologne	

Thursday, December 8, 2016

9:00 - 17:00	Side Meetings		
	Concurrent M3: **International Colectica User Conference (ICUC) 2016** Public Meeting Location: GESIS Room Ost/East Organizers: Jeremy Iverson and Dan Smith (both Colectica)	Concurrent M4: **DDI Software Developers Meeting** Public Meeting Location: GESIS Room West I Chairs: Oliver Hopt and Matthäus Zloch	Concurrent: **CESSDA Meetings**
	See description	For more information see contact the email group	Concurrent M5: **CESSDA Metadata Management Project** Closed Meeting Location: GESIS Room West II Organizer: Wolfgang Zenk-Möltgen
			09:00-14:30
			Concurrent M6: **CESSDA Technical Work Group** Closed Meeting Location: GESIS Room A112 Organizer: John Shepherdson
			10:00-12:00
			Concurrent M7: **CESSDA Euro Question Bank Project** Closed Meeting Location: GESIS Room West II Organizer: Wolfgang Zenk-Möltgen
			15:00-18:00
18:00	**Informal Get-Together** Location: Schreckenskammer, Ursulagartenstraße 11-15, 50668 Cologne		

December 12-16, 2016

9:00 17:30	Side Meeting
	M8: DDI Specification Development
	DDI Lifecycle Moving Forward Project Closed Meeting Location: GESIS A112 Organizer: Joachim Wackerow

Keynotes

Metadata: Foundation, Philosopher's and Rosetta Stones

Keith Jeffery (Keith G Jeffery Consultants) (↑ schedule)

Research is now digital in execution and reporting/preservation. The end-to-end process of research – from idea to proposal to funded project to outputs – is supported by ICT (Information and Communication Technologies).

The various entities of the research domain require digital description to facilitate discovery, contextualization (for relevance and quality but including rights, costs, security, and privacy restrictions) and action. Metadata is the magic key for this; starting with datasets metadata now describes software, workflows, persons, organisations, equipment, computing resources and more.

Metadata must have formal syntax and declared (multilingual) semantics. Most existing metadata 'standards' do not meet these criteria, but many can be interconverted to a subset of a canonical form that does to permit interoperation. CERIF (Common European Research Information Format: a European Union Recommendation to Member States) is a widely-used canonical data model to meet these objectives. RDA (Research Data Alliance) is evolving a list of metadata elements to be recommended to support the operations described above; the set accords well with CERIF and – like CERIF - is a superset of other metadata 'standards'.

Keith Jeffery is an independent consultant and past Director IT at STFC Rutherford Appleton Laboratory with 360,000 users, 1100 servers and 140 staff. Keith holds 3 honorary visiting professorships, is a Fellow of the Geological Society of London and the British Computer Society, is a Chartered Engineer and Chartered IT Professional and an Honorary Fellow of the Irish Computer Society. Keith is past-President of ERCIM and past President of euroCRIS, and serves on international expert groups, conference boards and assessment panels. He has advised government on security and green computing. He chaired the EC Expert Groups on GRIDs and on CLOUD Computing. He is serving as co-chair in several working and interest groups of the Research Data Alliance (Metadata, Metadata Standards Directory, Data in Context, Metadata Standards Catalog).

Funding, Policies, Community Building - Data Sharing from a Funders Perspective

Stefan Winkler-Nees (DFG - German Research Foundation) (↑ schedule)

The German Research Foundation (DFG) funds projects for basic research in all disciplines. Since the late 90s, data produced in these projects were already in focus to ensure reproducibility and to verify the scientific results achieved. However, the aspect of re-use of research data in future projects, potentially targeting also interdisciplinary topics, reached greater importance over time. Therefore, DFG implemented funding opportunities for projects to develop and implement infrastructures and information systems for research data. One basic condition for these projects is to consider discipline specific requirements the intended information service needs to meet – in other words, to follow a "response mode". Consequently, infrastructure projects should involve professional information management, IT expertise and researchers, who help to define these requirements. Overall objective is to reach an infrastructure ecosystem, which allows "*…researchers in all disciplines [to] access all research data quickly and easily in a straightforward process in order to carry out research at the highest level and produce excellent results.*" (Alliance of Science Organisations in Germany, 2015).

Dr. Stefan Winkler-Nees is marine geologist. He studied in Munich and Kiel and worked as paleo-climate scientist in a number of research projects in Europe and Australia. In 2001, he joined a software company as consultant and account manager. Since 2007, he works for the German Research Foundation (DFG). First, he was responsible for the German contribution to the global Ocean Drilling Project. Today, he is programme director at the Department of Scientific Library Services and Information Systems. His task is

to develop and to run funding programmes to develop and to implement research data information systems. He is active in national and international networks of funding agencies, research organisations and information infrastructures initiatives.

Discussion Sessions

Plenary Panel Session: Re-Use of Software and Administered Metadata
(↑ schedule)

DDI was initially intended for preservation and exchange. Additionally, DDI Lifecycle provides the reuse of identifiable metadata items along the stages of one study, in multiple waves of a study, or across studies and organizations. The DDI specification provides only a structure for this purpose. A working infrastructure requires also software using this structure, and administered metadata. The concept of exchanging and reusing metadata can be extended to software as well.

- What is the vision of a distributed infrastructure for reusable DDI software components and reusable, administered metadata?
- What are the challenges and barriers in building this?

Panelists will make initial comments on the topic and discuss the issues. Questions and comments from the audience will broaden the discussion.

Discussion: Questionnaire Capture / Re-usable Questions
(↑ schedule)

This campfire session invites the audience to discuss issues around questionnaire capture and re-usable questions. This session is related to the previous three sessions on questionnaire capture.

Discussion: Publishing DDI-Related Topics
(↑ schedule)

This campfire session invites the audience to discuss issues around DDI-related publications. This session is related to the previous two presentations on DDI-related publications (current status, advantages, and challenges) which provide some food for thought for this discussion. The intention is to raise more awareness on writing publications in the DDI community and to identify possible barriers.

Presentations in Concurrent Sessions

(in alphabetical order by the last name of the first author)

Creating a DDI-Lifecycle based Metadata Schema for the GESIS Data Archive

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

Track: Reusing and Sharing – **Session Type:** Short Presentation - Project Report

The project DDI-LimDAS is the effort to create a metadata schema on the basis of DDI-Lifecycle for all metadata items, which are processed at the Data Archive. So far, DDI-Codebook has been used as the predominantly metadata format and only single applications were based on DDI-Lifecycle. Diverse applications are in use for the purposes of documenting the data holdings on study and variable level, for registering and discovering datasets, for supporting new questionnaire design, for enabling data cleaning, harmonization, and long-term archiving. To enable interoperability of metadata between all those applications, a common data model is absolutely necessary. This will reduce complexity and enable a centralized metadata store.

Preliminary results and experiences have already been presented at EDDI15. Since then, progress was made in collecting the base metadata schemata from more applications and mapping them to a common metadata model. Another significant new result is the mapping of each element of the common schema to a DDI-Lifecycle element. The mapping also contains constraints so that a valid DDI Instance and a DDI Profile can be generated. The presentation will show the overall state of the project and examples of the mappings to the common metadata schema.

Assessment of the Swiss National Objectives in Education

Ingo Barkow (HTW Chur), Catharina Wasner (HTW Chur) (↑ schedule)

Track: Questionnaire Capture – **Session Type:** Regular Presentation - Project Report

In Switzerland, the main responsibility for education and culture lies with the 26 cantons. In 2006 it was decided that some cornerstones of the Swiss education system has to be harmonized nationally. The Swiss national objectives in education are some of these cornerstones. They describe which competencies students in all cantons should have obtained after 2, 4 and 9 years of school. In 2016 first surveys about the national objectives in education were conducted. More are planned for 2017.

For the management of these surveys software tools will be developed which support documentation with DDI for the questionnaires. Further metadata standards are currently in discussion to be used because no current standard alone seems to meet all demands. In our presentation we would like to present where DDI can be supportive in the area of educational assessments and where gaps within DDI has to be filled for this special field.

The DASISH Questionnaire Design and Documentation Tool – Functionalities and Examples from the Tool

Benjamin Beuster (NSD - Norwegian Centre for Research Data), Hilde Orten (NSD - Norwegian Centre for Research Data) (↑ schedule)

Track: Questionnaire Capture – **Session Type:** Short Presentation - Project Report

The Questionnaire Design Documentation Tool (QDDT) is developed 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. The workflow of the European Social Survey is the use-case and DDI 3.2 serves as the basis for the conceptual model of the tool.

The QDDT has been presented at previous EDDI conferences. The current presentation provides an overview of the current developments, including functionalities and giving examples from the tool.

Documenting Microdata from Official Statistics – Benefits of a DDI-Based Backend Infrastructure

Jeanette Bohr (GESIS - Leibniz Institute for the Social Sciences), Matthäus Zloch (GESIS - Leibniz Institute for the Social Sciences) (↑ schedule)
Track: Official Statistics – **Session Type:** Regular Presentation - Project Report

Microdata from official statistics are a valuable data source for the social sciences. Providing a central source of structured and user-friendly data documentation encourages the use of microdata from official statistics for scientific purposes and reduces the individual effort in data exploration and preparation. For that purpose, GESIS has developed an online metadata system (MISSY) and a powerful documentation tool that fully satisfies the documentation process. MISSY addresses (social-) scientists working with data from official statistics (Eurostat microdata and the German Microcensus) and focuses on a very detailed, variable level documentation. The system consists of several technical components to allow for an efficient and standardized metadata management process:

- an import application to capture metadata (semi-)automatically from different sources,
- a metadata editor to edit and store the metadata and to produce several output formats (e.g. setup routines, pdf codebooks),
- an extensible and DDI-based backend data model that allows transformations into several export formats (e.g. DDI-RDF, DDI 3.2),
- and a public data documentation service (MISSY-Web) to facilitate the usage of official statistic microdata for researchers.

This presentation will introduce the MISSY metadata infrastructure and will demonstrate the benefits and the potentials of structured data documentation for both the metadata service provider and the potential users.

"Beam us up Scotty" - The CESSDA Metadata Management Core Metadata Model Goes into Round Two

Kerrin Borschewski (GESIS - Leibniz Institute for the Social Sciences), Wolfgang Zenk-Möltgen (GESIS - Leibniz Institute for the Social Sciences), Julia Hermann (GESIS - Leibniz Institute for the Social Sciences) (↑ schedule)
Track: Research Infrastructure – **Session Type:** Short Presentation - Project Report

At EDDI15 the CESSDA Metadata Management (CMM) group presented the project which had just been started. One year later, the creation of the Metadata Portfolio Version1 is under way and considerations for the CMM Phase 2, and therefore the "next, evolved and adapted generation" of the Metadata Portfolio, are being made. This presentation will focus on the current status of the CMM Core Metadata Model Version 1. The authors will present the relevant interim results and highlight the broad use of DDI-Lifecycle within the model. The model takes into account the needs of resource discovery, the CESSDA Euro Question Bank, preservation, and multi-lingual items. The extension and adaptation of the Core Metadata Model for Phase 2 of the project is intended to meet further user needs (identified in Phase 1) and the requirement to coordinate the model with the DDI Moving Forward data model. Additionally, we intend to portray the ideas, prospects and challenges for this next, extended and amended version of the Core Metadata Model.

Questionnaire Generator

Guillaume Duffes (Insee - National Institute of Statistics and Economic Studies) (↑ schedule)
Track: Questionnaire Capture – **Session Type:** Regular Presentation - Project Report

Insee has been developing for more than four years a statistical survey questionnaire generator which takes as input DDI 3.2 files and produces web forms (Xforms format).

Until 2015, the DDI files were mainly written manually which gave to the DDI modeller a wide flexibility to determine the right DDI representations. This is possible as long as the modeller knows the DDI model very well and his work is specifically dedicated to modelling DDI questionnaire.

However, with the view of generalising, streamlining and enhancing the metadata-driven questionnaire generation process, a generic tool is needed for subject-matter statisticians to design their questionnaire. "Generic" here means independent of the DDI terminology that statisticians ignore. That's why Insee has built a prototype Graphical User Interface (GUI) which can produce DDI files in an user-friendly manner and pass them to the generator.

Nevertheless the complexity of the DDI model willy-nilly makes the use of the GUI complex, even though the GUI is not strictly bound to the DDI model to the extent possible.

Hence precise guidelines for designing questionnaires are necessary for both subject-matter people who will use the GUI and IT developers who complement the questionnaire generator and need to understand the DDI output objects.

Those guidelines are gathered in a guidebook on the design of DDI questionnaire. The UNECE Modernisation Comittee on Standards accepted to review this guidebook a few months ago.

This presentation will describe the components of the infrastructure set up for generating survey questionnaire in a machine-actionable way: the GUI for questionnaire design, and the questionnaire generator. It will also explain why it is essential to produce detailed guidelines for using those tools and implementing the DDI standard.

Controlled Vocabularies for CESSDA

Anne Etheridge (UK Data Service, University of Essex) (↑ schedule)
Track: Reusing and Sharing – **Session Type:** Regular Presentation - Community Impact

The CESSDA Metadata Management (CMM) project, led by FSD, plans to develop, promote and implement a standardised metadata design, content and practice for all CESSDA data assets. The design will be based on the Data Documentation Initiative (DDI). The project will produce a portfolio of the metadata fields (described in a separate paper) and controlled vocabularies for some of those fields to enable cross-collection/cross-country data retrieval. The controlled vocabularies will, wherever possible, be those from the DDI suite of controlled vocabularies. If none of those are suitable then CVs will be suggested to the DDI CVG group for inclusion, else the CMM will create their own. Most members of the CMM CV team are also members of the DDI CVG. This paper describes the process of choosing those vocabularies and highlights particular ones. There are eight CESSDA archives directly involved in this project, out of the 15 CESSDA archives that will directly benefit from this model. The CVs have to be usable by all CESSDA archives so they must be multilingual. The project is also collaborating on a management system for the vocabularies, to manage not only the CVs from the project but those from DDI as well.

An Infrastructure for Stimulating Researchers to Deposit Data

Johan Fihn (SND - Swedish National Data Service) (↑ schedule)
Track: Research Infrastructure – **Session Type:** Regular Presentation - Project Report

Swedish National Data Service (SND) has during 2016 run a national pilot project with the purpose of looking at possibilities to establish a national infrastructure with the purpose of making research data more accessible. The project introduces the collaboration between SND, university archives and university

libraries, where SND works as a back office knowledge center and university librarians functioning as a front office towards researchers.

An online form for general descriptions of research data has been developed where researchers, with the help of university librarians, describe their data. The metadata produced is described in DDI and will be packaged with all data when it's archived as well as being the basis of a potential national search portal for research data produced at Swedish universities.

The Library of Online Harmonisations: The SERISS Data Harmonisation Platform

Martin Friedrichs (GESIS - Leibniz Institute for the Social Sciences), Kristi Winters (GESIS - Leibniz Institute for the Social Sciences) (↑ schedule)
Track: Harmonisation – **Session Type:** Regular Presentation - Project Report

Harmonising variables is an important part of social science research. It adds value to existing dataset through its reuse and extends compatibility to other datasets. However, these intellectual contributions are often lost because researchers can't publish their work.

For variable harmonisation work to be of scientific value the harmonisation process must be documented and published in a precise and transparent ways, and different types of harmonisations must be comparable (e.g. SPSS, CharmStats or DDI). The simplification and routinisation of data harmonisation workflows will create better data synergies.

Winters and Netscher published standards for documenting variable harmonisation work but a digital location for harmonization work is needed.

Our solution is an online library where digitally documented harmonisation routines will be archived and accessed. The Library of Online Harmonisations will advance data harmonisation by transforming it into a digital documentation process of outputs that are publishable, findable, citable and replicable using only online resources.

Such a platform requires metadata and data harmonisation workflow to be represented in a standardised and machine actionable way. We will present how DDI can be utilized in in the creation of the Library of Online Harmonisations.

Blaise, Colectica and DDI

Lon Hofman (Statistics Netherlands) (↑ schedule)
Track: Colectica & Blaise – **Session Type:** Regular Presentation - Community Impact

Colectica and Statistics Netherlands have entered into a long term partnership to build software linking Blaise, Colectica, and the DDI Lifecycle standard. The software allows survey researchers to build surveys faster, to leverage the DDI metadata standard, and to generate rich documentation and reports. The tools will improve transparency into the data capture process.

The first tool offers an intuitive survey design surface and questionnaire palette, allowing survey designers to build questionnaires without learning a domain specific language. Questions, blocks, and logic can be created within the program or reused from question bank powered by DDI. Reusing standardized questions will assist in creating more comparable data.

The software stores questionnaire specifications using the open DDI and GSIM standards, and can connect to metadata repositories and question banks powered by Colectica software. Data descriptions can be linked with source questions, creating harmonized data and showing data lineages.

Surveys designed with this tool can be fielded using Blaise 5 on the desktop, on the Web, and on mobile devices. The tool converts the DDI metadata into a Blaise project and source code. Changes to surveys made with the tool can be published and executed within the Blaise environment, allowing rapid iteration while developing surveys.

Two additional tools help convert information between the Blaise and DDI formats: the Blaise-Colectica Blaise to DDI Connector and the Blaise-Colectica DDI to Blaise Connector. Additionally, the teams will publish an official Blaise to DDI mapping paper.

Rich Metadata from the Start

Oliver Hopt (GESIS - Leibniz Institute for the Social Sciences), Claus-Peter Klas (GESIS - Leibniz Institute for the Social Sciences) (↑ schedule)

Track: Questionnaire Capture – **Session Type:** Regular Presentation - Project Report

In the GLES project, GESIS will prepare and run several surveys on the German national election in 2017. In EDDI2015 we provided an overview of the workflow necessary. The mostly manual process of creating questionnaires to running the study and documenting it for research and evaluation will be supported by several web-based tools. One major tool developed within the project is the questionnaire editor. On the frontend it is a collaborative editor to create, edit and structure questionnaires. It enables to create questions, question grids and logical blocks as main components.

Answer domains, interviewer instructions and filter statements are managed separately for reuse.

On the backend side it is fully DDI-L enabled based on the DDI-FlatDB architecture with full history on revisions and versions.

To support the creation process a question database will be developed to find questions from other studies or older waves. The information that the questions are re-used will be stored in the new questionnaire for methodological research. At any time, the questionnaire can be exported to DDI, PDF or office formats. In addition, statistical file prototypes can be created for handover to the survey institutions to predefine the structure and naming conventions of the later dataset.

What's New in Colectica 5.3

Jeremy Iverson (Colectica) (↑ schedule)

Track: Colectica & Blaise – **Session Type:** Regular Presentation - Community Impact

Colectica is software for managing statistical data. It is used by national statistical organizations, university research groups, and data collection agencies to provide well-documented data to researchers and the public. Colectica is built on open standards like DDI and GSIM, ensuring that information can be presented in numerous formats and shared among different organizations and tools.

In this session we will demonstrate new features added to Colectica in 2016, including:

- Colectica Datasets: a new application to add rich metadata directly to statistical data files, and to convert statistical data files to open formats
- Colectica Questionnaires: a new, streamlined survey design application including tighter integration with computer assisted interviewing systems like Blaise and Lime Survey
- Import formats: SAS and SAS catalog files
- Colectica Portal: data lineage visualizations, allowing researchers to trace statistical data back to its source
- Colectica Portal: Elasticsearch for fast, advanced queries
- Colectica Portal: maintenant en français

Documenting and Re-Using of DMP Information – The Distinct Role of DDI

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

Track: Research Infrastructure – **Session Type:** Short Presentation - Project Report

Metadata and standard compliant exchange and re-use of information of data management plans (DMP) along the data lifecycle concern a large range of different topics. Past conference discussions, e.g. at the EDDI 2015 or the IASSIST 2016, demonstrated that 'no simple' solutions exist to tackle related structural (what's missing in current DDI flavors?) and conceptual challenges (how to standardize diverse DMP structures?).

Of particular interest is the consideration to separate data management from project management issues looking forward to standards that simplify documentation, re-use and work with DMP information for relevant user groups.

To investigate such challenges, the current presentation will

- describe core differences between data and project management related DMP aspect
- exemplarily separate project management and data management issues, outlining standards of project management aspects (e.g. CASRAI, CERIF; PRINCE 2)
- focus on core data driven DMP elements to manage with DDI
- discuss conclusions about the distinct DDI role managing research data for re-use

Thereby, we aim to push forward DDI discussion on DMP's standards to support researchers and research projects.

Lessons Learnt From the Development of a DDI-Based Discovery Platform
Jon Johnson (UCL Institute of Education) (↑ schedule)
Track: Issues in Building Infrastructure – **Session Type:** Regular Presentation - Project Report

A major output of the CLOSER (Cohort & Longitudinal Studies Enhancement Resources) project is the creation of CLOSER Discovery (discovery.closer.ac.uk) to enable discovery of the data collected – and the related survey questions and measurements – across the 8 participating studies

Using a combination of commercially available software and in-house developed tools to manage the creation and ingest of questionnaires, mapping of variables to questions and tagging of searchable topics has been able to create an online resource that enables researchers to view and appraise data from eight leading UK longitudinal studies

The presentation will discuss lessons learnt managing DDI-L during the project associated with:

1. the use of a mix of technologies;
2. working across the social, behavioural and biomedical domains;
3. working with different implementations of the same version of DDI-L

The presentation will also describe some of the benefits that a platform adhering to DDI-L will bring to both the research and data collection communities.

Enhancing the Quality of Metadata with DDI-L in a Transitioning DDI-C Environment

Alina Danciu (Centre de données socio-politiques (UMS 828), CNRS/SciencesPo Paris), Alexandre Mairot (Centre de données socio-politiques (UMS 828), CNRS/SciencesPo Paris) (↑ schedule)
Track: Harmonisation – **Session Type:** Regular Presentation - Project Report

The French Center of Socio-Political Data has previously presented its reflection on the process of shifting from DDI-C to DDI-L and the creation of a DDI-L compliant metadata system.

At the time being, we are in the process of harmonising our metadata. To do this, we are using our DDI-L compliant metadata database presented at EDDI15. By taking the example of a twelve-wave political study, our presentation aims at showing how the use of DDI-L improves the capturing of metadata and of their lifecycle in a transitioning DDI-C environment.

In this paper, we are going to show how the dynamic process of harmonisation improved intrinsically the quality of our codebook metadata, allowing us to surpass artificially the DDI-C limits. Besides, we will discuss the use of this database as a documentation tool, which ensures metadata ingests and, at the same time, allows us to test its robustness and stability.

Lessons from a Large-Scale Metadata Archiving Team Using DDI-Lifecycle

Hayley Mills (UCL Institute of Education) (↑ schedule)
Track: Harmonisation – **Session Type:** Regular Presentation - Project Report

Ingesting metadata from legacy instrumentation into the DDI-Lifecycle is complex and resource demanding. One of CLOSER's aims is to document over 100,000 questions from over 380 instruments through a large centralised metadata team. This presentation uses data collected from over 2.5 years of manual metadata entry to describe the lessons learnt during this process. It will set out the considerations for undertaking a similar project including considerations for recruitment and cost estimations.

The Past, Present and Future of Geocoded Survey Data at the GESIS Data Archive

Stefan Müller (GESIS - Leibniz Institute for the Social Sciences), Wolfgang Zenk-Möltgen (GESIS - Leibniz Institute for the Social Sciences), Stefan Schweers (GESIS - Leibniz Institute for the Social Sciences) (↑ schedule)
Track: Reusing and Sharing – **Session Type:** Short Presentation - Project Report

Last year at EDDI15 we introduced our work on geocoded survey data that have been enriched with spatial data attributes at the GESIS Data Archive. This year, we suggest solutions and unveil our metadata scheme reflecting the challenges of documenting geocoded survey data at the archive. By splitting survey data attributes and spatial data attributes logically we were able to implement a documentation structure that also supports the dissemination of these data to the end-users. Moreover, we showcase present and future directions of our work regarding metadata of geocoded survey data with spatial data attributes. Since our present approach demands a lot of manual work, we intend to propose automatic solutions to the metadata processing. This includes metadata that are directly transferred from spatial datasets as, e.g., ISO19115 metadata fields to the corresponding DDI metadata fields. Furthermore, we aim to automatically collect metadata that are produced during the processes of geocoding, data linking, and disclosure analyses. These metadata will be included in a future metadata scheme that uses the wide and rich opportunities of DDI-Lifecycle. We complete our presentation with final thoughts on the future infrastructure at the archive: a spatial data infrastructure for social science survey data.

The e-Lab Metadata Registry: Supporting Harmonisation Using DDI, OWL and Reporting Guidelines

Philip Couch (University of Manchester), Chris Munro (University of Manchester) (↑ schedule)
Track: Harmonisation – **Session Type:** Short Presentation - Project Report

For some diseases, the symptoms associated with the disease arise from different pathophysiological mechanisms. In order to improve the way we treat such diseases, we could reclassify the disease into distinct sub-diseases that correspond to the different underlying mechanisms. This analysis often requires the pooling of data across multiple studies that might be based on different populations and that are collected and managed in different ways. Before analysis, the data need to be harmonised, a process that involves mapping similar variables across the studies onto 'standard variables'.

Here we present the e-Lab project, which is developing an approach and software to support the harmonisation process. We discuss the data model behind a new metadata registry that combines DDI with other standards including the Web Ontology Language. We also introduce the use of Variable Report Models, which specify the information that must be available when a primary measurement or outcome is made available to an analyst. For example, this additional information may include qualifying factors that provide the context of the data needed for the analyst to correctly interpret the results.

Towards an Open Infrastructure for Supporting Relationship Discovery Between Scholarly Assets

Chris Munro (University of Manchester), Philip Couch (University of Manchester), Jon Johnson (UCL Institute of Education) (↑ schedule)
Track: Harmonisation – **Session Type:** Short Presentation - Project Report

There has been significant investment in the development of infrastructure for the curation and discovery of scientific assets, as demonstrated by the increasing number of institutional repositories and catalogues. However, despite this investment, it still remains very difficult to explore the relationships between these assets. For example, there is often no standard mechanism for a researcher to make available relationships between the findings presented in a publication and a dataset that was used as part of the analysis. There are examples where studies have needed to use text mining techniques to explore the literature to determine these relationships. This information is important for funders to demonstrate the usefulness of a dataset. It also useful for researchers to find other work relating to a dataset that may be important to consider for future work.

Here we present our work to develop an approach and tooling to allow relationships to be expressed in a standard and open way. We discuss the use of 'Research Objects' to aggregate artefacts and to provide information about sources of metadata. We also discuss our plans for a distributed infrastructure that negates the need to support management by a single central organisation.

Fulfilling User-Needs and Improving Quality and Efficiency through Statistical Standards and Reuse of Common Metadata

Mogens Grosen Nielsen (Statistics Denmark) (↑ schedule)
Track: Official Statistics – **Session Type:** Regular Presentation - Project Report

Since 2011 Statistics Denmark has been working on building a common metadata system based on DDI. In the ongoing work, we are implementing integrated metadata for quality, concepts, classification and variables compliant with GSIM and other standards. We follow the UN-principles: reuse metadata where possible, integrate metadata into the business processes (GSBPM) and use metadata actively aiming for metadata-driven production. The presentation claims that the use of metadata to improve fulfilment of user-needs, improve quality and improve efficiency require …

a) improvement and precision in the terminology that is used when we talk about metadata,
b) better understanding of the role of metadata in relation to our users,
c) better understanding of the role of metadata in the production processes, and
d) using common overall architecture models to ensure successful implementation.

Building a Better Search and Discovery Platform Using Elasticsearch

Stefan Jakobsson (SND - Swedish National Data Service), Olof Olsson (SND - Swedish National Data Service), Johan Fihn (SND - Swedish National Data Service), Pablo Millet (SND - Swedish National Data Service) (↑ schedule)

Track: Harmonisation – **Session Type:** Short Presentation - Project Report

The Swedish National Data Service (SND) is in the process of rebuilding its search platform to support better discovery utilizing Elasticsearch. The metadata indexed requires the profile to support social science, humanities, environmental data and general GIS data. We will present and discuss the index mapping for Elasticsearch and how the fields are mapped from DDI-L. In the live demo you will be able to see the possibilities of exploring the metadata by queries. The first release of the discovery portal will be presented.

Sharing Metadata Using the DDI "Standard" – 20 Years Later - How are we doing?

Timothy Parsons (CISER - Cornell Institute for Social and Economic Research), Brandon Kowalski (CISER - Cornell Institute for Social and Economic Research), Michelle Edwards (CISER - Cornell Institute for Social and Economic Research) (↑ schedule)

Track: Issues in Building Infrastructure – **Session Type:** Regular Presentation - Project Report

DDI was developed as the standard for describing social science and statistical data 20 years ago. But what is a standard and how do we define one? Should the standard specify the model or should it specify the content? How do we know if we have it right? Advances in technology have propelled the rapid sharing of information. Systems are far more interconnected in enabling users to share information worldwide in a pace unimaginable back when DDI was first developed. Has DDI been able to keep up with these advances to serve the needs of users and systems? Where does DDI fit in this new world? We will demonstrate, how and where DDI may be falling behind and areas where we excel.

Archivist & Mapper: Simplifying and Modernising Questionnaire Entry

Will Poynter (UCL Institute of Education) (↑ schedule)

Track: Questionnaire Capture – **Session Type:** Regular Presentation - Project Report

Whether entering an historical questionnaire or designing a new data collection instrument, the combination of Archivist and Mapper will simplify the task. Archivist is a modern web-based questionnaire editor that is capable of importing and exporting the DDI3.2 Questionnaire Profile. (Gierl, 2014) Mapper is a unique web-app that provides a user interface to the complex task of connection questions and variables and applying a controlled vocabulary to items.

Both Archivist and Mapper have been designed with the aim of making large-scale ingests as efficient as possible, both in time, but also expertise; therefore their user interfaces are highly intuitive to people without a working knowledge of DDI. Archivist has also been developed with extendibility in mind, allowing groups to write plugins to augment Archivist's basic questionnaire editing functionality (e.g. adding further response domain types).

A Capability Development Model for Assessing and Improving Distributed Infrastructures and their Services.

Mike Priddy (DANS - Data Archiving and Networked Services), Trond Kvamme (NSD - Norwegian Centre for Research Data), Marion Wittenberg (DANS - Data Archiving and Networked Services) (↑ schedule)
Track: CESSDA – **Session Type:** Regular Presentation - Community Impact

As part of the CESSDA SaW (Strengthening and Widening) project a capability development model (CDM) was instigated to aid current and aspiring service providers to participate in CESSDA to their maximum ability: some partners have a long history and have set high ambitions and funding, whereas other partners are in the process of setting up their archives, sometimes with limited funding.

Rather than setting fixed requirements for each service and provider within an infrastructure, the CDM can be used to define both the organisation's desired state and provide effective guidance on how to improve services that form part of the infrastructure gradually, initially to the minimal and then, to the desired state for the service provider. The CDM can be adapted to other infrastructural contexts in the social sciences and humanities.

In this this presentation we will show the model, discuss how it can be used to identify gaps in service provision as well as managed improvements. Furthermore, we will consider how it could be adapted for use in other distributed infrastructures, or used in tandem with other maturity models (such as CESSDA's software maturity matrix), plus future developments of the model.

The Reference Model for Social Science and Humanities Data Infrastructures: towards a common architectural model for Research Infrastructure architectures.

Mike Priddy (DANS - Data Archiving and Networked Services), Maarten Hoogerwerf (DANS - Data Archiving and Networked Services), John William Shepherdson (UK Data Service, University of Essex), Johan Fihn (SND - Swedish National Data Service) (↑ schedule)
Track: CESSDA – **Session Type:** Regular Presentation - Community Impact

Research (data) infrastructures are primarily challenged by their own distributed nature. They consist of organisations who provide various types of services in different countries, with different business models & national environments, leading to different implementation of the services and standards. A prerequisite for addressing these challenges is a common understanding about the nature of the infrastructure, the diverse objectives of participants, functions and solutions. A common vision, model and architecture are instrumental in providing the foundation for this common understanding, as is a framework to identify and relate important aspects for central stakeholders. Furthermore it is an essential step for the alignment of the technical implementations with the high-level goals.

The Reference Model for Social Sciences and Humanities (RM-SSH) is an abstract model identifying and describing the concerns and related behaviour that infrastructures share, and translates these into common types of information and computation. Thus allowing for consistent internal development, maintenance, operation, and for the identification of opportunities for the development of interoperable and shared services and standards.

The MIDUS-Colectica Portal

Barry Radler (University of Wisconsin) (↑ schedule)
Track: Colectica & Blaise – **Session Type:** Regular Presentation - Project Report

DDI 3.2 is an apt metadata standard for describing related series of studies and datasets, such as those found in longitudinal research projects. MIDUS (Midlife in the U.S.), a national longitudinal study with over 12,000 cases and a broad blend of social, health, and biomarker data, has greatly benefitted from adopting DDI 3.2. In particular, this presentation describes how MIDUS has implemented DDI infrastructure to create a harmonized data extraction system that provides users of multi-disciplinary longitudinal MIDUS datasets the information needed to better analyze, interpret, and share them. Such a

system allows researchers to search across datasets for variables of interest, identify and harmonize related longitudinal versions of variables, and easily create customized data extracts and codebooks that are directly related to specific research questions. One area in which DDI has proven less satisfactory is in the documentation of non-survey data captures that do not necessarily require a Question construct, per se. How are such non-survey data--including those from bio-medical, administrative, and experimental paradigms—to be described by DDI? This presentation will demonstrate some of the non-survey data and instruments in MIDUS and discuss future directions for better documenting such data using DDI

Using DDI in Harmonizing Longitudinal Data for Open Online Analyses

Juha Samuel Ranta (Finnish Social Science Data Archive), Markus Tuominen (Finnish Social Science Data Archive) (↑ schedule)
Track: Questionnaire Capture – **Session Type:** Regular Presentation - Project Report

This presentation will introduce a new service Tietomilli, Finnish Social Science Data Archive's new easy-to-use interface for accessing key survey data on societal participation, values and attitudes of the Finns, and changes over time. The service enables online use of the data archived at FSD without requiring statistical skills. Target user groups are schoolchildren, students, journalists, decision-makers as well as ordinary citizens.

We use FSD's DDI Codebook variable level documentation that exists in XML format to combine and harmonize longitudinal datasets for Tietomilli. Access to study level documentation is also provided in the service. Reusability of FSD's DDI Codebook descriptions is enhanced by using in-house tools. We will describe the tools and how we use DDI variable level documentation to make harmonization process of survey data more affordable and less discouraging of a task. Our presentation will focus on the back-end processes and programs.

We will also discuss the challenges of the project and ideas for further developments. The project is funded by the Finnish Open Science and Research Initiative and is set to come to conclusion at the end of 2016.

Link to service http://www.fsd.uta.fi/suomi-monitori/.

Is this Metadata Management Tool of Any Use? Extending CESSDA's Software Maturity Matrix to the DDI Domain

John William Shepherdson (UK Data Service, University of Essex) (↑ schedule)
Track: CESSDA – **Session Type:** Regular Presentation - Community Impact

As part of its efforts to promote good practice for own-use software development and help meet its five common interoperability characteristics, CESSDA has defined a software maturity matrix. The work was initially based on NASA's Reuse Readiness Levels and revised in accordance with the "Capability Development Model" from the CESSDA SaW project. It now features ten criteria and five levels for each. It can be used when evaluating 3rd party software for potential adoption, or as acceptance criteria for components developed for use within CESSDA's Research Infrastructure.

Whilst it gives an indication of the general technical quality and likely maintainability and extensibility of a software product, it doesn't address its fitness for purpose for performing domain-specific tasks. This core can easily be extended to incorporate additional criteria and levels for a given domain, but what are they in the case of DDI and does the community want or need such a software maturity matrix?

The presentation covers the origins of CESSDA's software maturity matrix, looks at the core criteria and levels and goes on to suggest some DDI-specific extensions. A straw poll at the end will give an initial answer to the question about the community's enthusiasm or otherwise!

Question Banks, Reusability, and DDI 3.2

Dan Smith (Colectica) (↑ schedule)
Track: Questionnaire Capture – **Session Type:** Regular Presentation - Community Impact

Question banks provide researchers the opportunity to reuse well tested questions and promote harmonization of collected data. Creating questions, and sequences of questions, which are easily reusable, is now possible using DDI 3.2. This talk will discuss the technical aspects of reusability, and highlight Colectica's new questionnaire editor that can connect to DDI 3.2 enabled question banks.

Questions in DDI consist of more than just a question text and response descriptions. They also allow for dynamic text based on previous information in a questionnaire. The same is true for a series of questions that may include filtering logic. The logic may rely on data captured in a different location within the survey, leading to a challenge when reusing DDI described items directly. Luckily for DDI users, in DDI 3.2 a loose coupling between inputs and outputs was included in the standard to allow directly reusing questions which rely on additional information. This loose coupling allows items in a question bank to be directly reused.

This summer, the staff at Colectica have worked to implement this new DDI 3.2 loose coupling feature. This talk will explain the background, technical approach and challenges faced when implementing loose coupling, demonstrate creating and searching question banks for reusable item descriptions, and a demonstration of including reused questions into a new survey questionnaire, highlighting Colectica's new questionnaire editor. It will also show how reusing a question can create links between collected instance variables and their harmonized data elements.

The Picasso Project

Kathryn Stevenson (Statistics Canada), Flavio Rizzolo (Statistics Canada) (↑ schedule)
Track: Official Statistics – **Session Type:** Regular Presentation - Project Report

Building on Statistics Canada's *metadata-driven* architectural principle and metadata strategy themes: *drive, make available, structure* and *manage*, Picasso is an enterprise solution for statistical data and metadata management. Automated business rules will ensure metadata is gathered uniformly, adhering to common architecture, governance and policy instruments.

Picasso, a three-year project launched in 2015, replaces local solutions with a hub for managing metadata for all surveys, administrative files and record linkage projects; a data service centre function for all 'fit for use' data files; and enterprise search and discovery using metadata to facilitate reuse of information. New tools and components include a metadata designer with an entity lifecycle management and registration process.

The solution architecture is based on a hybrid relational/semantic graph (RDF) core registry and repository with a data model driven by standard vocabularies, e.g. SKOS/XKOS, PROV-O, and reference models, e.g. GSIM, DDI 4 and SDMX. Picasso component and external systems interact with the RDF core via a Data Access Layer and Entity Services to access metadata entities via Common Information Exchange Models. Standard vocabularies and models ensure efficient information exchange internally and to external users through the Agency's website and Research Data Centres.

Publishing DDI-Related Topics - Advantages and Challenges of Creating Publications

Joachim Wackerow (GESIS - Leibniz Institute for the Social Sciences) (↑ schedule)
Track: Publishing DDI-Related Topics – **Session Type:** Short Presentation - Community Impact

This presentation will outline the advantages of creating publications on DDI-related topic for both the DDI community and the authors. It will also highlight the challenges and barriers in creating written papers. The pros and cons will be discussed from the perspective of possible authors coming from multiple backgrounds.

Publishing DDI-Related Topics - Current Status and Related Working Group of the DDI Alliance

Knut Wenzig (German Institute for Economic Research (DIW Berlin) / German Socio-Economic Panel (SOEP)) (↑ schedule)

Track: Publishing DDI-Related Topics – **Session Type:** Short Presentation - Community Impact

While publishing articles in all fields of science and research is a common practice, it is less common in the area of metadata and data management, especially in the DDI community. This can be seen as problematic because gained insights tend to disappear after projects are offline and the next similar project has to start from the very beginning - which is a common (and perhaps preferred) practice for most software engineers.

But: If one would subsume metadata and data management under science, sustainable research would look different and last but not least the principle "publish or perish" is relevant for many involved experts, too.

This presentation will give an overview of the current status of DDI-related publications. Furthermore the ongoing work of a related working group of the DDI Alliance will be described.

Steps towards a Single Point of Access for Survey Questions across Europe: The Euro Question Bank Project

Wolfgang Zenk-Möltgen (GESIS - Leibniz Institute for the Social Sciences), Azadeh Mahmoud Hashemi (GESIS - Leibniz Institute for the Social Sciences) (↑ schedule)

Track: Questionnaire Capture – **Session Type:** Regular Presentation - Project Report

The Euro Question Bank (EQB) project will create a single point of access for survey questions across Europe. For social science researchers, the EQB will provide an integrated and easy to use portal which enables search queries over all survey questions from CESSDA archives in different languages. Functionalities of the EQB include faceted searching, cross national comparisons, and multilingual comparisons. If the available documentation exists, browsing by keywords, concepts, and collections will be possible.

Recently, the project was expanded by including the CESSDA members FORS, FSD, NSD, SND, and UKDA. The implementation of EQB is based on the DDI-Lifecycle metadata standard and provides both DDI-Lifecycle and DDI-Codebook import and export functionalities. The system architecture makes use of the Open Source Metadata Harvester Tool and the DDI Flat DB concept, which both have already been presented at previous EDDI conferences.

Posters and Software Demonstrations

(in alphabetical order by the last name of the first author)

The DASISH Questionnaire Design and Documentation Tool (QDDT) – A Business Tool for Capturing the Questionnaire Design Process

Benjamin Beuster (NSD - Norwegian Centre for Research Data), Hilde Orten (NSD - Norwegian Centre for Research Data) (↑ schedule)

Track: Questionnaire Capture – **Session Type:** Poster/Software Demonstration - Project Report

The primary aim of the Questionnaire Design Documentation Tool (QDDT) is to assist large-scale survey projects in the questionnaire design and development process. It's primarily developed to track the development of research concepts, questions, response domains and instruments for questionnaire modules with specific topics. DDI 3.2 serves as a basis for the conceptual model of the tool.

This poster provides a live demonstration of the tool as of December 2016.

CLOSER's Principles for Large-Scale Metadata Archiving Using DDI-Lifecycle

Hayley Mills (UCL Institute of Education), Will Poynter (UCL Institute of Education) (↑ schedule)

Track: Research Infrastructure – **Session Type:** Poster/Software Demonstration - Community Impact

A substantial part of CLOSER (Cohorts and Longitudinal Studies Enhancement Resources) is to electronically document over 100,000 questions from over 380 instruments for use in the CLOSER search platform, CLOSER Discovery (http://discovery.closer.ac.uk). Whilst carrying out the metadata ingest process, a set of 5 principles were developed and fine-tuned in order to outline and protect the overall standard by which the metadata is recorded. Due to the complexity and non-standardised format of the legacy instruments, there are situations where the entry principles conflict, and therefore one or more principle must be broken in order to follow the more significant principle(s). This poster outlines the updated principles in order of importance, and describes how the careful breaking of one or more principles can be applied in order to preserve the remaining principles and produce a standardised approach.

Developments and Perspectives in DDI Implementation Projects at SSJDA

Akira Motegi (Institute of Social Science, The University of Tokyo) (↑ schedule)

Track: Software Tools – **Session Type:** Poster/Software Demonstration - Project Report

We aim to discuss developments and perspectives in DDI implementation projects at Social Science Japan Data Archive (SSJDA). Our implementation is twofold. First, we have conducted Nesstar system operation since 2012. While the Nesstar system has been widely operated, there still remains enough room for its implementation in Japan. Our recent developments will be discussed with the detailed explanation focused on the published datasets. Our second implementation is the Easy DDI Organizer (EDO), a metadata editing and management and survey planning software development project. The main features will be explained such as metadata-editing based both of the DDI-L and DDI-C, the file import and the export functions, and its open source. Given these developments and experiences, the perspectives for further development will be discussed.

DDIR: An R Package for Handling DDI Files

Yasuto Nakano (Kwansei Gakuin University) (↑ schedule)

Track: Software Tools – **Session Type:** Poster/Software Demonstration - Project Report

'**DDIR**' is an R package which handles information in DDI format in the R environment.

Because a DDI format file collects/contains all information we need in research activities, it is efficient to use one DDI file as a source of information in any step of research activities even for small research project groups or for individual researchers.

In R environment, there is no standard data format for social research data . In many case, we have to prepare numerical data and label or factor information separately. If we use DDI file as a data file with **DDIR** in R, only one DDI file is needed to be prepared. DDI file could be a standard data format of social research data in R environment, just same as 'sav' file in SPSS. **DDIR** realizes integrated social research analysis environment with R, and ensures it as a reproducible research.

Building a Better Search and Discovery Platform Using Elasticsearch

Stefan Jakobsson (SND - Swedish National Data Service), Olof Olsson (SND - Swedish National Data Service), Johan Fihn (SND - Swedish National Data Service), Pablo Millet (SND - Swedish National Data Service) (↑ schedule)

Track: Harmonisation – **Session Type:** Poster/Software Demonstration - Project Report

The Swedish National Data Service (SND) is in the process of rebuilding its search platform to support better discovery of social science, humanities, environmental, and general GIS data utilizing Elasticsearch. We will present the platform we built on Elasticsearch and how DDI-L fields are used for search and filtration. In this live demo you will be able to see the possibilities of exploring the metadata by queries.

Re-Purposing DDI-L in the Production Process

Andrew Peters (UCL Institute of Education) (↑ schedule)

Track: Issues in Building Infrastructure – **Session Type:** Poster/Software Demonstration - Project Report

The Centre for Longitudinal Studies has worked with UK data collection agencies and software vendors to provide DDI-L compliant documentation of the data collection implementation in Unicom Intelligence.

The intention has been to achieve this without having to overhaul internal systems and infrastructure involved in the process.

Whilst this has provided a rich source of accurate information on the implementation (as opposed to the original specification), it has presented challenges as we try to re-purpose that across the data production and documentation pipeline.

The presentation will describe how we have re-used the DDI output from CAI systems to assist in data specification, data validation, and production of machine readable documentation as well as the provision of DDI compliant metadata to the archive to support the deposit of data. Finally the presentation will detail the pros and cons of re-purposing the CAI DDI output as well as the challenges that have arisen in trying to share that metadata within and between institutions.

Tutorials

Documenting Panel Data Using DDI

Knut Wenzig (German Institute for Economic Research (DIW Berlin) / German Socio-Economic Panel (SOEP)) (↑ schedule)

Track: Reusing and Sharing – **Session Type:** Tutorials or Workshop, half day

The key characteristics of panel studies include repeated measures for a more or less stable sample over time. The core challenge in documenting panel studies is the documentation of these repeated measures (usually questions) and the resulting variables because various reasons can require modifications of measures over time – resulting in comparable but not identical data structures.

The DDI standard provides not one but multiple options for the documentation of panel data. In this workshop we like to present various options and discuss their feasibility for common use cases. The German Socio-Economic Panel (SOEP) will provide the primary use case, but participants are also invited to introduce and discuss their own use cases.

Document Questionnaires and Datasets with DDI: A Hands-On Introduction with Colectica

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

Track: Reusing and Sharing – **Session Type:** Tutorials or Workshop, full day

This workshop offers a hands-on, practical approach to creating and documenting both surveys and datasets with DDI and Colectica.

Participants will build and field a DDI-driven survey using their own questions or samples provided in the workshop. They will then ingest, annotate, and publish DDI dataset descriptions using the collected survey data. The course will cover the following DDI content areas:

Questionnaire Design

- Survey Instruments
- Questions
- Concepts and Universes
- Question banks

Dataset Documentation

- Datasets and dataset layouts
- Summary Statistics
- Code Lists and Categories
- Data harmonization with RepresentedVariables and ConceptualVariables

Attendees may optionally bring their own Windows laptops to participate in the hands-on exercises.

Have "Questions"? We Have a Structure for Them

Wendy Thomas (University of Minnesota Minnesota Population Center), Jon Johnson (UCL Institute of Education), Guillaume Duffes (Insee - National Institute of Statistics and Economic Studies) (↑ schedule)
Track: Reusing and Sharing – **Session Type:** Tutorials or Workshop, half day

The most significant area of DDI adoption since 2008 has been around the development, management, and use of Questions. The creation of question banks, development of tools to organize and field questionnaires, and interest in new and specialized means of data capture has fueled development in the DDI. For those of you that have "Questions" and want to do something with them this half-day tutorial will address the following:

Question structures in DDI:

- What have these structures been designed to support
- How do they work - with what types of questions?
- How can they be organized - grouping and questionnaires

Creating Question Banks:

- What features are used
- Where are problems - and how to deal with them
- Trade-offs between semantics and representation: the special case of question grids

Creating Questionnaires:

- Organizing question flows and adding content - instructions, statements, and visual display
- Managing the flow of data from capture to variable

Question development and management:

- Creating new questions
- Managing old questions

Public Side Meetings

International Colectica User Conference (ICUC) 2016
Dan Smith (Colectica), Jeremy Iverson (Colectica) (↑ schedule)
Session Type: Side Meeting, public

Colectica will host a users meeting featuring a day of DDI and Colectica related discussion focusing on practical approaches to DDI metadata management, innovative research, successful implementations and lessons learned, and Colectica extensions and tools developed by and for the community.

The theme for 2016 is "**Surveys and Datasets**"

This is a public event. Individuals from statistical agencies, survey designers, data producers, managers, and users who are interested in either DDI, recent DDI implementations, or Colectica tools are invited to participate.

ICUC is co-located with the annual European Data Documentation (EDDI) conference. This provides **four days** of exciting content around the **DDI metadata standard**!

DDI Software Developers
Oliver Hopt and Matthäus Zloch (both GESIS - Leibniz Institute for the Social Sciences)
Session Type: Side Meeting, public

The meeting is an opportunity for software developers who work with the DDI standard to gather together and discuss topics like recent developments and techniques, upcoming changes of the standard, interoperability or just meet other developers.

Active Data Management Plans
Uwe Jensen (GESIS - Leibniz Institute for the Social Sciences)
Session Type: Side Meeting, public

Data Management Planning (DMP) is becoming highly relevant, since funding guidelines expect sharing and re-use of publicly funded data. The DDI working group on Active Data Management Planning (ADMP) aims incorporating DMP into current DDI versions. The meeting will discuss current status and future action items of this endeavour.

Closed Side Meetings

Big Data Europe

Ivana Versic (CESSDA - Consortium of European Social Science Data Archives)
Session Type: Side Meeting, closed

CESSDA Metadata Management Project

Wolfgang Zenk-Möltgen (GESIS - Leibniz Institute for the Social Sciences)
Session Type: Side Meeting, closed

The key objective of the CESSDA Metadata Management Project is to develop, promote and implement a standardised metadata design, content and practice for all CESSDA data assets. The result will be the CESSDA Metadata Standards Portfolio Version 1 that will become the standard for all Service Providers.

The CESSDA members are FSD (Lead), ADP, CASD, DDA, GESIS, NSD, SND, UKDS.

The side meeting will support the project work related to DDI (access only to project members).

CESSDA Technical Work Group

John Shepherdson (UK Data Service, University of Essex)
Session Type: Side Meeting, closed

CESSDA Euro Question Bank Project

Wolfgang Zenk-Möltgen (GESIS - Leibniz Institute for the Social Sciences)
Session Type: Side Meeting, closed

Aim of the project is to develop and implement the Euro Question Bank, a central search facility across all CESSDA's survey holdings in a way which provides as much coverage of questions used in surveys as possible. The Euro Question Bank will be based on the DDI-Lifecycle metadata standard.

The CESSDA project members are GESIS (Lead), FORS, FSD, NSD, SND, UKDA.

The side meeting will support the project work related to DDI (access only to project members).

DDI Lifecycle Moving Forward

Joachim Wackerow (GESIS - Leibniz Institute for the Social Sciences)
Session Type: Side Meeting, closed

This is a working meeting of the DDI Moving Forward Project focusing on the next generation of DDI. DDI 4 will provide more complete coverage of the data life cycle, and will be model-based and model-driven.

Authors Index

Organizations Index

About EDDI and EDDI16

EDDI Conference Description

EDDI (Annual European DDI User Conference) is the annual conference for users of DDI (Data Documentation Initiative), a metadata specification for the social, economic, and behavioral sciences. It is run by GESIS and the IDSC of IZA under the auspices of the DDI Alliance.

EDDI is designed to provide forum where DDI users from Europe and the world can gather to showcase their work and their progress toward DDI adoption, as well as discuss any questions or challenges they may have about the standard.

EDDI includes presentations, poster sessions, and discussion sessions. The conference closes with a "meet the experts" session in which users will have a chance to lobby for their point of view with representatives from the Technical Committee of the DDI Alliance. The philosophy of EDDI is to be an open, inclusive DDI community-building activity.

In a different European country each year, the conference (including related meetings) usually spans a week in early December.

EDDI16 Committees

Organization Committee

- Nikos Askitas, IDSC of IZA - International Data Service Center of the Institute for the Study of Labor, Germany
- Kerrin Borschewski, GESIS - Leibniz-Institute for the Social Sciences, Germany
- Uwe Jensen, GESIS - Leibniz Institute for the Social Sciences, Germany
- Eric van Um, GESIS - Leibniz Institute for the Social Sciences, Germany
- Joachim Wackerow, GESIS - Leibniz Institute for the Social Sciences, Germany
- Ulrike Weyerke, GESIS - Leibniz-Institute for the Social Sciences, Germany

Program Committee

- Jon Johnson, UCL Institute of Education, United Kingdom
- Mari Kleemola, FSD - Finnish Social Science Data Archive, Finland
- Mogens Grosen Nielsen, Statistics Denmark, Denmark
- Joachim Wackerow, GESIS - Leibniz Institute for the Social Sciences, Germany
- Knut Wenzig, German Institute for Economic Research (DIW Berlin) / German Socio-Economic Panel (SOEP), Germany
- Wolfgang Zenk-Möltgen, GESIS - Leibniz Institute for the Social Sciences, Germany

Conference Co-chairs

- Jon Johnson, UCL Institute of Education, United Kingdom
- Joachim Wackerow, GESIS - Leibniz Institute for the Social Sciences, Germany