EUROPEAN DDI USER CONFERENCE

7th Annual European DDI User Conference (EDDI15)

Hosted by Statistics Denmark (DST) and National Archive of Denmark

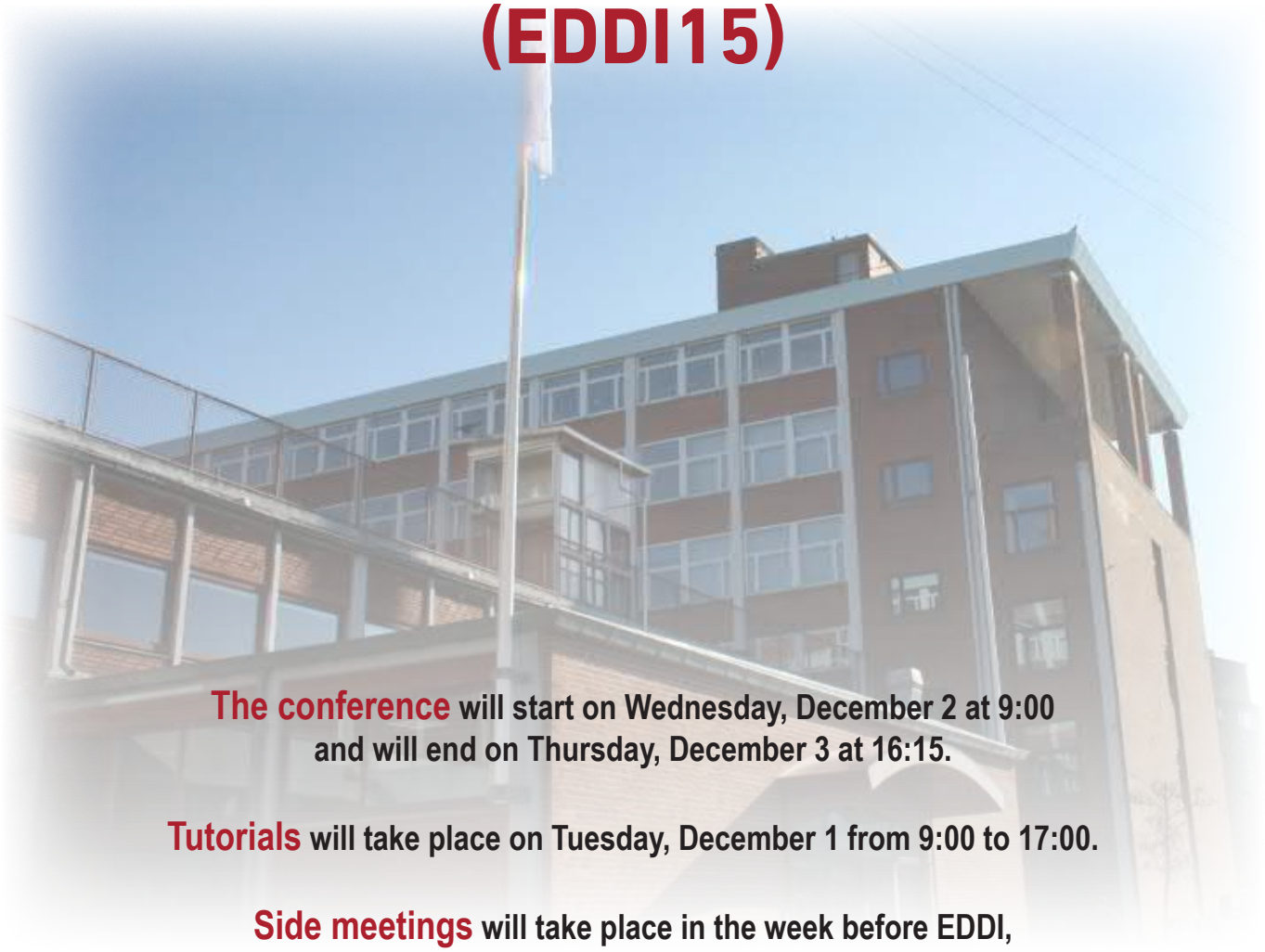
December 2-3, 2015

At: Royal School of Library and Information Science (RSLIS/IVA),
University of Copenhagen
Birketinget 6, DK 2300 Copenhagen S

STATISTICS DENMARK

RIGSARKIVET
THE NATIONAL ARCHIVES

Welcome
to the
7th
Annual European DDI User Conference (EDDI15)



The conference will start on Wednesday, December 2 at 9:00 and will end on Thursday, December 3 at 16:15.

Tutorials will take place on Tuesday, December 1 from 9:00 to 17:00.

Side meetings will take place in the week before EDDI, on Tuesday, December 1 and on Friday, December 4.

More information: www.eddi-conference.eu/eddi15

STATISTICS DENMARK

RIGSARKIVET
THE NATIONAL ARCHIVES

DDI

gesis
Leibniz-Institut für Sozialwissenschaften

IZA

IDSC

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, Leibniz Institute for the Social Sciences and the IDSC of IZA (Institute for the Study of Labor) 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.

The Data Documentation Initiative (DDI) is an international effort to create a standard to describe statistical and social science data. Documenting data with DDI facilitates interpretation and understanding - both by humans and computers. The freely available international DDI standard describes data that result from observational methods in the social, behavioral, economic, and health sciences. The DDI metadata specification supports the entire research data life cycle.

DDI metadata accompanies and enables data conceptualization, collection, processing, distribution, discovery, analysis, repurposing, and archiving. Use DDI to Document, Discover, and Interoperate!

Organizing Committee

Nikos Askitas, IDSC of IZA - International Data Service Center of the Institute for the Study of Labor, Germany
Anne Sofie Fink, Danish Data Archive, Denmark
Christian Lindgaard Olesen, Danish Data Archive, Denmark
Diana Ransgaard Sørensen, Statistics Denmark, Denmark
Joachim Wackerow, GESIS - Leibniz Institute for the Social Sciences, Germany
Catharina Wasner, GESIS - Leibniz Institute for the Social Sciences, Germany, Germany

Program Committee

Denis Grofils, Eurostat, Luxembourg
Jannik Vestergaard Jensen, Danish National Archives, Denmark
Jon Johnson, IOE - Institute of Education, Centre for Longitudinal Studies, United Kingdom
Mogens Grosen Nielsen, Statistics Denmark, Denmark
Joachim Wackerow, GESIS - Leibniz Institute for the Social Sciences, Germany
Knut Wenzig, DIW Berlin (SOEP), Germany
Wolfgang Zenk-Möltgen, GESIS - Leibniz Institute for the Social Sciences, Germany

Conference Chair

Joachim Wackerow, GESIS - Leibniz Institute for the Social Sciences, Germany

Venues and pre-conference schedule

VENUE INFORMATION

Royal School of Library and Information Science (IVA), University of Copenhagen
Birketinget 6, DK-2300 Copenhagen S., Denmark

Registration	Royal School of Library and Information Science (IVA), University of Copenhagen Birketinget 6, DK-2300 Copenhagen S.
Main conference	Lecture Hall and Auditorium **Royal School of Library and Information Science (RSLIS),** in Danish: Det Informationsvidenskabelige Akademi (IVA) University of Copenhagen, Birketinget 6, 2300 Copenhagen S
Tutorials	Rooms A1.08 and A.1.25, RSLIS/IVA
Side meetings	DDI Lifecycle Moving forward Location: Room Adolph Jensen Statistics Denmark Sejrøgade 11 2100 Copenhagen Ø

Developers meeting	Room A.1.25, RSLIS/IVA
CESSDA Metadata Harvesting tool and Technical Framework projects	Room A.1.20, RSLIS/IVA
CESSDA Metadata Management project	Room A.1.20, RSLIS/IVA
International Colectica User Conference (ICUC)	Room A.1.08, RSLIS/IVA

Conference Dinner December 2, 2015	The Danish National Museum (Restuarant Julian), Nye Vestergade 10, 1471 Copenhagen K
Informal get-together	See detailed schedule below

23-27 NOVEMBER 2015

9:00-18:00	**DDI Lifecycle Moving Forward** Location: Room Adolph Jensen, Statistics Denmark, Sejrøgade 11, 2100 Copenhagen Ø

MONDAY 30 NOVEMBER 2015

19:00	**Informal get-together** Location: Ørsted Ølbar Nørre Farimagsgade 13, 1364 Copenhagen K (close to Nørreport Station)

Conference schedule

TUESDAY, DECEMBER 1, 2015			
8:15	**Starting Registration** Location: Lobby adjacent to Lecture Hall, RSLIS/IVA		
9:00-10:30	**CONCURRENT TUTORIALS**		
	Concurrent T1 (full day) **Data Documentation** Location: Room A1.08, RSLIS/IVA **A hands-on, practical introduction to DDI using Colectica** Jeremy Iverson, Dan Smith (Colectica)	**Concurrent T2 (half day)** **Reusing and Sharing Metadata** Location: Room A1.25, RSLIS/IVA **Documenting panel data using DDI** Marcel Hebing, Knut Wenzig (DIW Berlin)	
10:30-11:00	**Break** – Location: Lounge outside room A1.08 and A1.25, RSLIS/IVA		
11:00-12:30	**Block 2 of Tutorial 1**	**Block 2 of Tutorial 2**	
12:30-13:30	**Lunch** – Location: Lounge outside room A1.08 and A1.25, RSLIS/IVA		
13:30-15:00	**Concurrent Tutorials 1 (continued)** Location: Room A1.08, RSLIS/IVA **Block 3 of Concurrent Tutorials 1**	**Concurrent Tutorial 3 (half day)** **Open Data and Linked Open Data** Location: Room A1.25, RSLIS/IVA **The interface between spatial and statistical data** Wendy Thomas (Minnesota Population Center)	**Side Meeting M1** **CESSDA Metadata Harvesting tool and Technical Framework projects** Location: Room A1.20, RSLIS/IVA **Starting time: 13:00** *(by invitation only)* Organizer: John Shepherdson
15:00-15:30	**Break** – Location: Lounge outside room A1.08 and A1.25, RSLIS/IVA		
15:30-17:00	**Block 4 of Concurrent Tutorial 1**	**Block 2 of Concurrent Tutorial 3**	**Block 2 of Metadata harvesting**
18:00	**Informal get-together** Location: Taphouse Lavendelstræde 15, 1462 Copenhagen K (close to Rådhuspladsen)		

Conference schedule

WEDNESDAY, DECEMBER 2, 2015

8:15

Starting Registration
Location: Lobby adjacent to Lecture Hall, RSLIS/IVA

9:00-10:15

CONFERENCE PLENARY P1: WELCOME AND KEYNOTE
Location: Lecture hall, RSLIS/IVA
Chair: Joachim Wackerow (EDDI Conference Chair)

Welcome
Asbjørn Hellum (General Director of National Archive of Denmark)

Keynote: Building a statistics lighthouse for all decision makers
Lars Thygesen (Director of Thygesen Statistics Consulting)
Carsten Zangenberg (Director for Communication and Sales at Statistics Denmark)

10:15-10:45

Break – Location: Lecture hall, RSLIS/IVA

10:45-12:15

CONCURRENT CONFERENCE SESSIONS

Concurrent A1	Concurrent A2:
Official Statistics Location: Lecture hall, RSLIS/IVA Chair: Wendy Thomas	**User needs..., Software / Tools: Confidential Data** Location:Auditorium, RSLIS/IVA Chair: Mari Kleemola
Classifications in Statistics Denmark – current situation and steps towards a DDI-based classification system Diana Ransgaard Sørensen (Statistics Denmark)	**Improving Access and Data Security to Confidential Labor Market Data** Warren A. Brown (CISER), Stephanie Jacobs (CISER), David Schiller (IAB), Germany), Joerg Heining (IAB), Germany
Towards common metadata at Statistics Denmark using GSIM and DDI Mogens Grosen Nielsen, Flemming Dannevang (Statistics Denmark)	**RAIRD - A privacy-preserving, browser-based statistical package** Ørnulf Risnes (NSD)
Management of Statistical Classifications with DDI 3.2 Dan Smith (Colectica)	**The Metadataportal (MDP) of the German Institute for Employment Research (IAB) – first live demonstrations** Dana Müller, David Schiller (IAB, Germany)

12:15-13:45

Lunch – Location: Lecture hall, RSLIS/IVA

13:45-15:15

CONCURRENT CONFERENCE SESSIONS

Concurrent B1	Concurrent B2:
Reusing and Sharing Metadata, User Needs..., Data Harmonization Location: Lecture hall, RSLIS/IVA Chair: Michelle Edwards	**Open Data and Linked Open Data, Reusing...** Location: Auditorium, RSLIS/IVA Chair: Marcel Hebing
Experimenting with DDI-L at the French Center of Socio-Political Data (CDSP): applying DDI-L to a political longitudinal study Simon Le Corgne, Alina Danciu, Alexandre Mairot (CDSP)	**Encapsulating DDI Metadata within a Spatio-Temporal Data Feed for British Historical Statistics** Humphrey Southall, Michael Stoner (University of Portsmouth)

Conference schedule

WEDNESDAY, DECEMBER 2, 2015 - *continued*

13:45-15:15	**CONCURRENT CONFERENCE SESSIONS - continued**	
	Concurrent B1 – continued **Use of DDI-Lifecycle at the GESIS Data Archive** Catharina Wasner (GESIS)	*Concurrent B2 – continued* **Georeferenced Survey Data at the GESIS Data Archive** Stefan Müller, Wolfgang Zenk-Möltgen, Stefan Schweers (GESIS)
	Best Practices for Documenting Repeated Studies in DDI Jeremy Iverson (Colectica)	**How can DDI make the most of RDF?** Franck Cotton, Guillaume Duffes (INSEE)
15:15-15:45	**Break**– Location: Lecture hall, RSLIS/IVA	
15:45-16:45	**CONCURRENT CONFERENCE SESSIONS**	
	Concurrent C1 **Software / Tools, Reusing...: Data Collection** Location: Lecture hall, RSLIS/IVA Chair: Knut Wenzig	**Concurrent C2:** **User Needs, Efficient Infrastructures and Improved Quality** Location:Auditorium, RSLIS/IVA Chair: Tito Castillo
	The DASISH Questionnaire Design Documentation Tool: Updates on the development of the tool Benjamin Beuster, Stig Norland, Hilde Orten (NSD)	**What comes first? Metadata or Data Access?** Michelle Edwards, William Block (CISER)
	Implementing DDI-Lifecycle for Data Collection within the German GLES project Manuela Blumenberg, Claus-Peter Klas, Wolfgang Zenk-Möltgen (GESIS)	**First Results from the Survey on Metadata Management in the Educational Sciences** Ingo Barkow (HTW Chur)
17:00-	**Before dinner** Exhibition 17-18: For those interested, an exhibit about the Danish prehistoric period can be seen at the National Museum of Denmark. **Conference Dinner** Location: National Museum of Denmark, Prinsens Palæ, Nye Vestergade 10, 1471 Copenhagen K Drinks 18.00-18:30 Dinner (buffet) 19:00-21.00 Coffee 21.00-21.30 **After dinner** Location: Mojo Blues Bar, Løngangstræde 21, 1468 Copenhagen (close to the National Museum)	

Conference schedule

THURSDAY, DECEMBER 3, 2015

9:15-10:15 — **CONFERENCE PLENARY P2: KEYNOTE**

Location: Lecture hall, RSLIS/IVA
Chair: Hans Jørgen Marker

Keynote: The Danish National Forum for Research data Management
Henrik Pedersen (Professor at University of Southern Denmark and president for the Danish National Forum for Research data Management).

10:15-10:45 — **Break**– Location: Lecture hall, RSLIS/IVA

10:45-12:15 — **CONCURRENT CONFERENCE SESSIONS**

Concurrent D1	Concurrent D2:
Metadata versus Data and Related Ethics, User Needs..., Reusing... Location: Lecture hall, RSLIS/IVA Chair: Mogens Nielsen	**Incentives to Document Data, User needs...** Location:Auditorium, RSLIS/IVA Chair: Iris Alfredsson
Documenting the lifecycle of every data point. A generic model for cell-level metadata in RAIRD Ørnulf Risnes (NSD)	**Exchanging Data Management Plans with DDI - Implementing Data Management in Social Science Funding - The German Case** Uwe Jensen, Sebastian Netscher, Kerrin Borschewski (GESIS)
Towards Metadata management with SDMX and DDI for the Deutsche Bundesbank Integrated Microdata Analysis System IMIDIAS Christopher Johannes Schild (Deutsche Bundesbank, Department Statistics)	**Introducing the CESSDA Metadata Management Project** Wolfgang Zenk-Möltgen (GESIS), Mari Kleemola (Finnish Social Science Data Archive), Anne Etheridge (UK Data Archive), Anne Sofie Fink Kjeldgaard (Danish National Archives)
A generic production environment - use of GSIM in Statistics Sweden Klas Blomqvist (Statistics Sweden)	

12:15-13:45 — **POSTERS AND SOFTWARE DEMONSTRATIONS**

Location: Lecture hall, RSLIS/IVA
Chair: Catharina Wasner

simultaneously with lunch

Automated Survey Enactment and DDI Integration
Stelios Alexandrakis (consulent)

DDI on Rails
Marcel Hebing (DIW Berlin)

Metka – the metadata database for FSD
Matti Heinonen, Katja Moilanen (Finnish Social Data Archive)

Implementing DDI in the SSJDA – Easy DDI Organizer project and Nesstar system operation
Akira Motegi (Institute of Social Science, The University of Tokyo)

The Epidemiology-France web portal: a metadata catalogue of French health databases
Simon Saint-Georges (Portail Epidemiologie-France / ITMO Santé Publique (Aviesan/INSERM)

Using DKAN – an Open Source Portal Solution for Publishing Social Science Data: Lessons learnt at DSZ-BO
Johanna Vompras (Bielefeld University Library)

Conference schedule

THURSDAY, DECEMBER 3, 2015 – *continued*		
12:15-13:45	**Lunch** – Location: Lecture hall, RSLIS/IVA	
13:45-14:45	**CONCURRENT CONFERENCE SESSIONS**	
	Concurrent E1 **Software / Tools** Location: Lecture hall, RSLIS/IVA Chair: Barry Radler	**Concurrent E2:** **Software / Tools** Location:Auditorium, RSLIS/IVA Chair: Guillaume Duffes
	Metadata in a Thousand Files: Bringing it all Together Claude Gierl, Jon Johnson (UCL, London)	**Open Source Metadata Accumulator for CESSDA (OS-MAC)** John William Shepherdson (UK Data Archice)
	A healthy portal is a portable portal Jannik Vestergaard Jensen (Danish National Archives), Stephanie Roth (Swedish National Data Service), Olof Olsson (Swedish National Data Service), Snorre Davoen (NSD), Ørnulf Risnes (NSD), Sami Granstedt (Finnish Social Science Data Archive)	**DDI-Flat-DB – a lightweight framework for heterogeneous DDI sources** Claus-Peter Klas, Oliver Hopt, Alexander Mühlbauer, Wolfgang Zenk-Möltgen (GESIS)
14:45-15:15	**Break** – Location: Lecture hall, RSLIS/IVA	
15:15-16:15	**CONFERENCE PLENARY P3: CURRENT STATUS AND OUTLOOK** Location: Lecture hall, RSLIS/IVA Chair: Bill Block	
	DDI Specifications (including Moving Forward/DDI4): Recent developments Joachim Wackerow and Wendy Thomas (both Technical Committee, DDI Alliance)	
	Questions and Grumbles? Answers from the Technical Committee Arofan Gregory, Wendy Thomas and Joachim Wackerow (all Technical Committee, DDI Alliance)	
	Announcement of 4th NADDI, Invitation to EDDI16 and Goodbye Barry Radler (NADDI team), Joachim Wackerow (EDDI core team), and Next Year's Host (name will be disclosed in session)	
18:00	**Informal get-together** Location: Ørsted Ølbar Nørre Farimagsgade 13, 1364 Copenhagen K (close to Nørreport Station)	

Conference schedule

FRIDAY, DECEMBER 4, 2015			
9:00-17:00	**SIDE MEETINGS**		
	Concurrent M2 **International Colectica User Conference (ICUC 2015)** Location: Room A1.08, RSLIS/IVA Organizers: Jeremy Iverson and Dan Smith (both Colectica) For registration and program, see webpage of ICUC 2015	**Concurrent M3** **DDI Developers Meeting** Location: Room A1.25, RSLIS/IVA Chair: Jannik V. Jensen (Danish National Archives)	**Concurrent M4** **CESSDA Metadata Management** Location: Room A1.20, RSLIS/IVA *by invitation only* Organizer: Anne Sofie Fink Kjeldgaard (Danish National Archives)

The Danish National Museum

Tutorials

A hands-on, practical introduction to DDI using Colectica

Jeremy Iverson (Colectica), Dan Smith (Colectica)

This workshop offers a hands-on, practical approach to getting started with data documentation standards like DDI. It starts with a high-level overview of statistical metadata the DDI content model, and then teaches how to create metadata in the DDI format. Participants will also learn how to publish rich metadata in multiple formats. Attendees may bring their laptops and data to participate in the hands-on exercises.

Introduction to Metadata
This session provides an introduction to using structured metadata and discusses the benefits of using standardized metadata.

- Non-technical session
- General introduction to the usefulness of standardized metadata
- Non-technical introduction to DDI
- Opportunities created by using standardized metadata
- Case studies of organizations using Colectica and DDI-Lifecycle
    - ◦Statistics New Zealand
    - ◦Statistics Denmark
    - ◦Midlife in the United States longitudinal study

The DDI Information Model
This session provides a detailed look at the content described by DDI, and explores how your agency's information relates to the DDI terminology.

- Series and Studies
- Quality Statements
- Data Collection, Survey Instruments, and Questions
- Data Files, Data Record Layouts, and Summary Statistics
- Code Lists and Categories
- Concepts, Universes (populations), and Organizations

Hands on with Colectica and DDI
These sessions focus on creating and documenting surveys and data using the Colectica software, and understanding how valid DDI XML is being generated automatically by the tools. Bring laptops to follow along and to document your own existing content.

- Technical training in Colectica
- Document datasets
- Create surveys
- Publish survey specifications, and generate source code for CAI and CATI systems
- Link survey descriptions with datasets
- Create rich data dictionaries

Documenting panel data using DDI

Marcel Hebing (German Socio-economic Panel Study (SOEP), DIW Berlin), Knut Wenzig (German Socio-economic Panel Study (SOEP), DIW Berlin)

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.

The workshop starts with a short introduction of both panel studies and the DDI standard. Therefore, no previous knowledge of the DDI standard is required to participate in the workshop. The goal for the workshop is to gain a deeper understanding of possible documentation strategies for panel studies.

The interface between spatial and statistical data

Wendy Thomas (University of Minnesota Minnesota Population Center)

DDI has always recognized the importance of spatial description for social science data and has increased the range of descriptive content in each of its major releases. As interest in spatial analysis and presentation increases the question of how to accurately use spatial information to link data sets and layer information from various domains has become a major concern.

Some discussions focus on standards and how they interact, others on semantic linkages. However, many of the difficulties in using the spatial dimension to traverse data sets lies in the lack of implicitly comparable content. Most social science data files do not have specifically associated spatial files for presentation. Most social science data does not provide explicit geographic information that can be easily processed by machines. Most published statistical tables rely on geographic place names rather than more specific coding systems.

This workshop focuses on spatial content rather than technical approaches to linkage and will cover:

- Information used to create accurate links between data sets at the spatial level
- Structures in commonly used standards to capture this information
- Content that should be provided by data producers in order to support the search for related data in an open-data environment
- Special issues involved in linking historical data and non-standard spatial types

The workshop is based on over 20 years of creating interfaces between spatial and statistical data including the work of National Historical Geographic Information System and Terra Populus at the Minnesota Population Center (University of Minnesota).

Plenary presentations abstract & keynotes

Wednesday, December 2, 2015

Building a statistics lighthouse for all decision makers

Lars Thygesen (Thygesen Statistics Consulting), Carsten Zangenberg (Statistics Denmark)

The purpose of official statistics is to provide a common basis for people's understanding of society, and thus help them making informed decisions. This makes it necessary to provide users with information of the available statistics; first regarding all statistics from each data provider (e.g. Statistics Denmark), and second between several data providers, in a country via a national portal and eventually across the world via portals at regional and global level.

In order to get this goal, metadata standards are a must. The use of the standards must be so intuitive that users, advanced as well as simple, can benefit. This requires introduction to terminology and information to users, aiming at making it easy to navigate in the information about statistics provided in the portals. In addition the standards must be understood and implementable by producers of statistics in the production processes.

Dissemination processes are especially important, since these processes ensure that external users are supported as best possible. Metadata standards used as part of an intuitive common terminology must be integrated into the dissemination systems so that the "distance" between the statistics and metadata will be as short as possible and increase users' understanding of official statistics. DDI together with other standards should be a key to solving all of these tasks.

Lars Thygesen

Lars has been working in official statistics for a lifetime, most of the time in Statistics Denmark, where he has been involved in developing the architecture of the Danish Statistical system.

He also worked in the OECD, developing their statistical system and databases.
He is now working as an independent consultant in statistics.

Throughout his career he has had a special interest in using and sharing excellent ideas and systems from best practice in other countries, and he has been especially focusing on how international standards can help us being clever, inventing the wheel just once and working together across borders.

Carsten Zangenberg

Carsten is director for Communication and Sales at Statistics Denmark.

He has been head of communication for several years and before that, he has been working with all aspects in the production of statistics and statistical analysis.

Through many years of experience with user needs and desires he has a special interest in dissemination of statistics, including how to integrate metadata in the dissemination.

Plenary presentations keynotes

Thursday, December 3, 2015

The Danish National Forum for Research data Management

Henrik Pedersen (Professor at University of Southern Denmark and president for the Dan¬ish National Forum for Research data Management)

Henrik Pedersen

Henrik is Professor of Mathematics since 1997, The Dean of Science and Engineering 2001-06, and Dean of Science 2006-15 at The University of Southern Denmark.

Henrik earned his M.Sc. in Mathematics and Physics from the University of Copenhagen, 1980, and a D.Phil. in Mathematics from University of Oxford, England, 1985, funded by The Royal Society, London.

Henrik has been member and Chairman of numerous boards of public and private companies, and Chairman for Science Ventures Denmark Ltd, spinning out companies based on IPR mainly from the university; He has been active establishing collaboration with Health Sciences, Humanities and Social Sciences and has been active in establishing collaboration with foreign companies and institutions including The Chinese Academy of Sciences.

Henrik was appointed president of The Danish National Forum for Research data Management in 2015.

Presentations in concurrent sessions

In order of concurrent session

Concurrent A1

Classifications in Statistics Denmark – current situation and steps towards a DDI-based classification system

Diana Ransgaard Sørensen (Statistics Denmark)

In the production of statistics at Statistics Denmark several classifications are in use. The classifications have not been subject to a centralised coordination. This has led to little reuse, duplication, and the endeavors to implement international standards has only been at a few domains.

In a pilot study, we introduced classifications as part of an integrated metadata system, which is based on the GSIM 1.1 Statistical Classification Model using the standard DDI 3.2. We used Business Process Management (BPM) as the overall methodology. First, we described existing work processes (as-is) and made descriptions of future work processes (to-be). The Generic Statistical Business Process Model (GSBPM) was used as the framework to have a common language and understanding. Secondly, we described use-cases and specifications in collaboration with the classification experts. Finally, a first prototype for a common classification system was developed.

The pilot project will be finalised and evaluated by the end of this year. Hereafter, we will continue the work on coordination and the use of a common classification system.

Towards common metadata at Statistics Denmark using GSIM and DDI

Mogens Grosen Nielsen (Statistics Denmark), Flemming Dannevang (Statistics Denmark)

Since the 1990's, Statistics Denmark has operated with separate classifications, quality declarations, variables and concepts systems. In 2011, we initiated projects aiming at a common integrated metadata-system in order to improve quality and facilitate dissemination of statistics. At the beginning of 2015, Statistics Denmark launched a DDI-based system handling concepts and quality information for 237 statistics. At the end of 2015, we expect to include variables, concepts and classifications in one common metadata system. GSIM and DDI 3.2 are used as standards for building models at conceptual, logical and implementation levels. Colectica is used as software.

The claim in this paper is that the road towards common metadata in a statistical context requires: a) improvement and precision in the terminology we use when we talk about metadata b) better understanding of role of metadata in relation to our users and c) better understanding of the role of metadata in production of statistics. The paper presents various perspectives on metadata including metadata in a helicopter perspective, metadata in a business processes and metadata implemented in Colectica using GSIM and DDI 3.2.

Management of Statistical Classifications with DDI 3.2

Dan Smith (Colectica)

Managing classifications in a major activity of statistical organizations. This presentation introduces best practices for managing Statistical Classifications using DDI 3.2 in conformance with the GSIM 1.1 Statistical Classifications model. The presentation also introduces the new statistical management functionality found in the recently released Colectica 5.2.

The Generic Statistical Information Model (GSIM) is an internationally endorsed reference framework for statistical information, and is based on the previous Neuchâtel model. In 2013, best practices for managing Statistical Classifications were described in the Copenhagen Mapping, a robust mapping between the classification model in DDI 3.2 and the GSIM classification model.

The Copenhagen Mapping also provides a controlled vocabulary for use with DDI 3.2. Statistical Classification support will be demonstrated in the new release of Colectica, showing DDI's ability to robustly manage Statistical Classification. Finally, this presentation will describe the simple upgrade path for Statistical Classification management from DDI 3.2 into future versions of DDI.

Concurrent A2

Improving Access and Data Security to Confidential Labor Market Data

Warren A. Brown (Cornell Institute for Social and Economic Research (CISER), Cornell University), Stephanie Jacobs (Cornell Institute for Social and Economic Research (CISER), Cornell University), David Schiller (Institute for Employment Research (IAB), Germany), Joerg Heining (Institute for Employment Research (IAB), Germany

The Cornell Institute for Social and Economic Research (CISER), Cornell University and the Institute for Employment Research (IAB), German Federal Employment Agency are collaborating to expand use of IAB's confidential Sample of Integrated Labour Market Biographies (SIAB). DDI 2.5 is used to enable researchers to discover the files by means of variable level searching in a repository of metadata on U.S. and German labor market related data files. The repository is the Comprehensive Extensible Data Documentation and Access Repository (CED2AR) being developed by researchers at Cornell University with funding from the U.S. National Science Foundation.

CED2AR provides researchers access to machine-readable codebooks with variable characteristics thus enabling researchers to develop detailed proposals for access to these data that are submitted to IAB. Researchers with approved projects are able to access and analyze the data using the Cornell Restricted Access Data Center (CRADC), a remote access virtual data enclave using remote desktop protocol. In the initial testing phase several researchers located in Europe and North America are successfully accessing and analyzing the Scientific Use Files of the SIAB. The project is well on its way to realizing the goal of wider access to researchers while improving

Presentations in concurrent sessions (continued)

secure management of confidential data.

RAIRD - A privacy-preserving, browser-based statistical package

Ørnulf Risnes (NSD - Norwegian Social Science Data Services)

RAIRD (Remote Access Infrastructure for Register Data) is a web-based system for confidential research on full population event data (spell-data) from a set of Norwegian administrative registers. RAIRD is currently under development and testing, and will move into production in 2017.

The RAIRD platform supports on-the-fly import and processing of event data into a disclosure-limiting web based statistical package for remote data processing and analysis.

The RAIRD technology stack is generic and may be used for other data sources and -types, e.g. survey data, panel data and of course event data.

This presentation will demonstrate RAIRD's end-user interface, and explain key aspects of the architecture and design principles, including:

- the use of immutable data structures and databases
- the combination of multiple programming languages and paradigms
- an overview of the supporting data and metadata models
- disclosure prevention models

RAIRD is a joint project between Norwegian Social Science Data Services (NSD) and Statistics Norway (SSB).
The RAIRD development project is funded by the Research Council of Norway.

The Metadataportal (MDP) of the German Institute for Employment Research (IAB) – first live demonstrations

Dana Müller (Institute for Employment Research (IAB), Germany), David Schiller (Institute for Employment Research (IAB), Germany)

The German Institute for Employment Research (IAB) is the research institution of the Federal Employment Agency. Via the Research Data Centre administrative and survey data is provided in a standardized way to researchers.

One of the most important issues to solve is providing a standardized, well organized, and user friendly data documentation. In the last years the IAB implemented a new Metadataportal (MDP), build by TBA21. The tool include numerous feature, for example, documentation of merged data sources, import of xml-files with information about variables or questionnaires, import of frequency tables, reporting, and offer a user-friendly search interface.

One of the most important features was implemented to enable the collaborative work of different IAB departments. While the Research Data Centre is responsible for the final release of research data, a number of departments are involved in the process of documenting the data. According to that some metadata information are restricted and for internal use only. Others are for public access. This is solved by a sophisticated rights management system working as backbone of the MDP. The public web portal is under construction.

Within this presentation the new IAB MDP will be introduced and presented as a live demo.

Concurrent B1

Experimenting with DDI-L at the French Center of Socio-Political Data (CDSP): applying DDI-L to a political longitudinal study

Simon Le Corgne (Sciences Po - CDSP), Alina Danciu (Sciences Po - CDSP), Alexandre Mairot (Sciences Po - CDSP)

The French Center of Socio-Political Data has presented its reflection on the process of shifting from DDI-C to DDI-L at EDDI14. This year, we will discuss the creation and storage of a DDI-L compliant XML record by capturing metadata of a nine-wave political study of the ELIPSS panel. Determining how best to recognise continuities between metadata collections within the same study, including question continuity and methodological continuities has been a primary challenge. To answer it, the starting point was the creation of a questions database.

As seen at the 2014 DDI workshop in Dagstuhl, the minimum requirements that a metadata system should meet before being able to import/export DDI-L are uniqueness of items, versioning and granularity. To conceive such a database, we had to start by using simple tools. We first identified metadata in CSV files that include variable-level information. We then performed a semi-manual import from these files to the database using importing scripts. Once we removed automatically the redundancy, with a further stage of human control, we generated the structure of the DDI-L compliant XML file. Our paper will present this process and discuss its replication to other DDI-C documented studies.

Use of DDI-Lifecycle at the GESIS Data Archive

Catharina Wasner (GESIS - Leibniz Institute for the Social Sciences)

A metadata schema for all use cases in the data archive will be developed in DDI-LimDAS to reduce complexity by harmonizing the preliminary work done in different projects so far. Prospective developments will be considered to facilitate a consistent and coordinated future DDI management which involves all DDI using projects at the data archive and beyond. This presentation introduces the GESIS project DDI-LimDAS and its goals, preliminary results and forthcoming activities.

Best Practices for Documenting Repeated Studies in DDI

Jeremy Iverson (Colectica)

Data are more interesting when they show how things change over time, instead of providing a single snapshot. For this reason, national statistical organizations routinely collect data at monthly, quarterly, and annual intervals. Longitudinal studies collect the same information about people on a repeated basis over long periods of time. To get the most value from these data sources, researchers need high quality metadata describing the data structures and how they change over time.

DDI Lifecycle provides nine core elements to model such repeated data. Although the DDI standard is often portrayed as having hundreds of overwhelming elements, these nine

Presentations in concurrent sessions (continued)

elements form a simple, elegant information model to represent data and to power rich, interactive visualizations and publication formats.

This presentation outlines a seven step process for documenting repeated data using the DDI standard, starting with basic data descriptions, followed by harmonization and mapping, and finally publication of the metadata in various formats.

This method of modeling repeated data is possible today with DDI Lifecycle, and aligns with the approaches used by the Generic Statistical Information Model (GSIM) and DDI Moving Forward (DDI 4).

Concurrent B2

Encapsulating DDI Metadata within a Spatio-Temporal Data Feed for British Historical Statistics

Humphrey Southall (University of Portsmouth), Michael Stoner (University of Portsmouth)

The GB Historical GIS holds 14m. diverse statistical data values in a uniform structure linked to a geospatial ontology of reporting units and a domain ontology of statistical concepts, the latter based on DDI concepts and especially the Aggregate Data Extension. This presentation describes the addition of a Linked Data API enabling access to this "big data" structure.

Our PastPlace DataCube API implements the World Wide Web Consortium's Datacube Vocabulary and is implemented as a Java application using Apache Jena, an open source Semantic Web framework for Java (https://jena.apache.org/). Jena serialises RDF graphs into different output formats including RDF/XML, Turtle, and Notation 3. To select data we draw on two distinct mechanisms already developed for the Vision of Britain download sub-system (www.VisionOfBritain.org.uk/data):

(1) The spatial strategy starts with the client specifying a point coordinate and a broad statistical theme, the system returning the set of reporting units whose boundary polygons cover the point, with thematically relevant nCubes from which the user can select.

(2) The thematic strategy allows the user to reach nCubes by moving down a topic hierarchy or through free-text searching of labels and descriptions within the metadata. Either strategy returns data values encapsulated within RDF variants.

Georeferenced 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

A current research project at the GESIS Data Archive will implement a geo-referencing service to merge survey data with small-scale spatial data (GeorefUm).

The project focuses on technical and conceptual challenges that are implied by realizing such a service, e.g. acquiring and harmonizing spatial data as well as archiving and documenting spatial and georeferenced survey data. As a first step to gather expertise in processing spatial data in combination with survey data, environmental noise and census data was added to the German General Social Survey 2014 which is curated at the archive.

The next step included the documentation of spatial data and georeferenced survey data on the study and the variable level by using and extending the DDI compliant metadata scheme of the archive. This was done with the help of the ISO 19115 metadata standard for spatial data that is already implemented in DDI-Lifecycle. Moreover, a compliance with the INSPIRE infrastructure is intended to exchange spatial data between data producers and users.

The presentation will discuss these operations with an emphasis on the development of a metadata scheme that describes sources, harmonization and the content of the georeferenced survey data.

How can DDI make the most of RDF?

Franck Cotton (INSEE), Guillaume Duffes (INSEE)

The development of DDI 4, including a RDF serialization, is under way.

The broad objective of RDF is to define a mechanism for describing resources that makes no assumptions about a particular application domain, nor defines a priori semantics of any application domain. Thus, existing RDF vocabularies can be straightforwardly reused by the future DDI implementation model.

Some vocabularies are already well-known and even used in the statistical community, for example SKOS (Simple Knowledge Organization System), others are developed under the auspices of the DDI Alliance (DISCO, XKOS…) or widespread in the Semantic Web community (the Dublin Core, also included in the previous versions of DDI).

Insee is going to manage its official concepts and classifications using SKOS and XKOS as core modeling standards. Some DDI controlled vocabularies and the recent UNECE glossary could also be transformed to RDF and linked to our metadata.

The presentation will demonstrate how Insee articulates the usage of RDF vocabularies with the DDI implementation to make its metadata semantically richer and their reuse more consistent and focused. A quick demo will show how the concepts can be:

- Managed in RDF
- Published as linked metadata
- Replicated as DDI in a Colectica Repository

Concurrent C1

The DASISH Questionnaire Design Documentation Tool: Updates on the development of the tool

Benjamin Beuster (NSD - Norwegian Social Science Data Services), Stig Norland (NSD - Norwegian Social Science Data Services), Hilde Orten (NSD - Norwegian Social Science Data Services)

Continuing from the DASISH project, two new tools to enhance the process of documenting questionnaire design and development and survey variables are currently being further developed and refined under the Synergies for Europe's Research Infrastructure in the Social Sciences (SERISS) project.

The primary aim of the questionnaire design

Presentations in concurrent sessions (continued)

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.

The revised version has a simplified data model and implements a RESTful API, which is used by the web frontend. All objects are transferred as JSON objects. This means that any modern system can interact with and use the data stored within with ease.

Although the database does not fully implement DDI, it is designed to be compatible with DDI 3.2, and DDI import and export will be implemented as add-ons to QDDT.

Implementing DDI-Lifecycle for Data Collection within the German GLES project

Manuela Blumenberg (GESIS - Leibniz Institute for the Social Sciences), Claus-Peter Klas (GESIS - Leibniz Institute for the Social Sciences), Wolfgang Zenk-Möltgen (GESIS - Leibniz Institute for the Social Sciences)

In the GLES project (http://www.gesis.org/en/elections-home/gles/), GESIS will prepare and run several surveys on the German national election in 2017.The overall process to prepare the surveys does consist of several steps beginning with preparing and evaluating previous questionnaires within a distributed team, hand over the questionnaires to the polling agency for running the surveys, and finally hand over the results for evaluation, documentation and archival processes. All these steps and hand-overs are currently done manually, in some cases semi-automatically. Word or PDF files are used for discussion, documentation and hand-over, resulting in repeated and error-prone documentation steps.

In order to harmonize the process and to raise efficiency between all participants, GESIS will drive a pilot project with GLES as main use case. The project will deliver all components and features to support surveys construction, hand-over and documentation. Starting from the first question and study description, DDI-Lifecycle will be the basis to automate creation and documentation of the new survey instruments. Within the research data lifecycle, this project represents a major step to use DDI within the phases of Study Planning and Data Collection.

Concurrent C2

What comes first? Metadata or Data Access?

Michelle Edwards (Cornell Institute for Social and Economic Research (CISER), Cornell University), William Block (Cornell Institute for Social and Economic Research (CISER), Cornell University)

The Data Archive at the Cornell Institute for Social and Economic Research (CISER) has been providing access to a variety of data sources for the Cornell community and beyond since 1981, and our services are in great demand.

Our goal over the next year is to evolve CISER's legacy archival environment to a flexible DDI-compliant data portal serving data to multiple audiences from a variety of social science data sources.

On campus clients include the Cornell-based National Data Archive on Child Abuse and Neglect (NDACAN) and another [soon to be named] major social science data archive; both seek to develop DDI-compliant metadata for their respective collections. Further afield, CISER is using its CED2AR platform to partner with the Institute for Employment Research (IAB), German Federal Employment Agency to expand use of IAB's confidential Sample of Integrated Labour Market Biographies (SIAB).

This presentation will highlight the process, challenges, questions, and possible answers of how CISER will evolve to ensure that Cornell based data archives continue to provide easy data access to our worldwide research community, via a DDI based infrastructure.

First Results from the Survey on Metadata Management in the Educational Sciences

Ingo Barkow (HTW Chur)

Representing educational content like cognitive items from computer-based assessment in data archives or repositories differs significantly from the archival of questionnaires or other means of data documentation in the social sciences. Though DDI 3.2 - and subsequently now DDI 3.3 - offer extended possibilities to represent simple stimuli like multiple choice tests more complex items like simulations cannot be represented.

To identify what the needs of users from the educational domain actually are two studies were conducted in the last six months as part of a doctoral thesis. While the results from the first survey were presented in April on the NADDI conference in Madison, Wisconsin this talk will show the final combined results of these two surveys as they were presented in the final PhD defense.

Concurrent D1

Documenting the lifecycle of every data point. A generic model for cell-level metadata in RAIRD.

Ørnulf Risnes (NSD - Norwegian Social Science Data Services)

RAIRD is a web-based system for confidential research on full population event data from a set of Norwegian administrative registers.

Event data (or spell-data) can be hard to understand, and hard to work with for researchers. The vision of the RAIRD-project is to build a user-friendly, web-based environment where researchers can query, process, manage and analyze event data through an intuitive interface. RAIRD must also ensure that outputs are anonymous, and that all analyses may be replicated.

Novel data and metadata-models were required to support the RAIRD-vision. It was important to develop a model with a clear reflection of the temporal aspect of events and associated state-changes in microdata. Moreover, new models needed to support attributes/annotations as well as fine-grained version control for individual data points.

Presentations in concurrent sessions (continued)

Such models have now been established through collaboration between the two RAIRD partners (Statistics Norway and NSD), and via substantial and frequent interaction with the GSIM and DDI modeling communities.

This presentation will describe the essentials of these new data and metadata-models, which should be useful in a wide range of scenarios, also outside RAIRD.

The RAIRD development project is funded by the Research Council of Norway.

Towards Metadata management with SDMX and DDI for the Deutsche Bundesbank Integrated Microdata Analysis System IMIDIAS

Christopher Johannes Schild (Deutsche Bundesbank, Department Statistics), Anja Treffs (Deutsche Bundesbank, Department Statistics)

Deutsche Bundesbank recognised an increasing need for microdata and is shifting away from the dominating macrodata paradigm of the financial sector towards an Integrated Microdata Analysis System (IMIDIAS).

Problems coming with the IMIDIAS initiative are that the microdata are disintegrated and dispersed and metadata are unstructured or stored in SDMX which is in its origin not made for microdata. At the same time, there is an increasing and heterogeneous demand from data users within Deutsche Bundesbank and outside coming through the new Research Data and Service Centre (RDSC).

One of the challenges that have to be addressed is defining a Metadata Model covering all demands. Discussions have shown that there are different possible organizational and technical approaches. While there is an existing SDMX infrastructure with some metadata there are additional needs verbalised by the RDSC that maybe can be addressed best by a DDI implementation. This presentation will focus on the IMIDIAS Metadata Model and the current ideas for bringing the existing SDMX infrastructure and the additional microdata-specific metadata needs together in one common IMIDIAS infrastructure.

A generic production environment - use of GSIM in Statistics Sweden

Klas Blomqvist (Statistics Sweden)

Statistics Sweden has launched a new program for creating a production environment for “enterprise wide, process based”-statistical services. One of necessary the preconditions for this initiative to provide metadata that can be used “actively” through the production process from design to dissemination.

These metadata have to be well-structured and of good quality in order to be used for building services in a new platform for processing and analyzing. Statistics Sweden is therefore creating and implementing a method for concept harmonization and also making use of the recently developed GSIM based model for a central metadata repository (presented at EDDI 2014) for handling Statistical programs, information management and data structure as well as for defining concepts and variables.

Other main areas of this development program are:

- Further developing the Triton platform
- Creating a platform for the dissemination process
- Creating statistical programs for products by implementing them into the platforms

One of the goals of the new production environment is to enable content harmonization. Focus is therefore on the role of metadata to provide reusability in order to increase efficiency and quality.

Concurrent D2

Exchanging Data Management Plans with DDI - Implementing Data Management in Social Science Funding - The German Case

Uwe Jensen (GESIS - Leibniz Institute for the Social Sciences), Sebastian Netscher (GESIS - Leibniz Institute for the Social Sciences), Kerrin Borschewski (GESIS - Leibniz Institute for the Social Sciences)

As an integral part of research, extensive data management is gaining in importance. It enables data sharing, enhances the quality of research, and increases transparency as well as replicability of scientific findings. However, implementing systematic data management plans (DMP) in social science research is challenging. In the cases of Germany, several initiatives are underway to promote data management as well as the development and implementation of guidelines for well-structured DMPs in research funding.

This presentation provides an overview of the current debate among German stakeholders on data management principles and focuses on approaches to implement DMPs in funding applications. In line with a recently established DDI working group, the presentation discusses the questions on how to implement and how to exchange DMPs. It investigates DDI facilities to meet such challenges. In particular, this implies the support of research projects by a DPM as a planning tool for systematic data management, enabling the creation of research data readily prepared for data sharing and reuse.

Introducing the CESSDA Metadata Management Project

Wolfgang Zenk-Möltgen (GESIS - Leibniz Institute for the Social Sciences), Mari Kleemola (Finnish Social Science Data Archive), Anne Etheridge (UK Data Archive), Anne Sofie Fink Kjeldgaard (Danish National Archives)

This presentation will give the first glimpse of the Consortium of European Social Science Data Archives (CESSDA) Metadata Management project. The project plans 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 within CESSDA. It will encompass support for resource discovery, question banks, preservation and data access and help all Service Providers achieve the Data Seal of Approval (DSA) certification requirements related to metadata issues.

Presentations in concurrent sessions (continued)

Concurrent E1

Metadata in a Thousand Files: Bringing it all Together

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

The CLOSER (Cohorts and Longitudinal Studies Enhancement Resources) Project at the Institute of Education, University College London, has now reached a more mature phase: metadata from the participant studies is being routinely produced and is starting to be ingested into the repository which will provide the backend of a search portal aimed at researchers. We concentrate in this paper on the Loader which implements the ingestion phase. The main files being ingested originate from a number of sources, but all have in common the fact that they are using the DDI-L (3.2) metadata standard as their format. Secondary mapping files are also being ingested for linking elements together.

The ingestion workflow is gradually being developed in free-standing mode in the first instance. The integration to a live website may bring up further issues. For now, the steps include building a concept hierarchy, ingesting the survey, questionnaire and variables metadata, linking variables to their source question, etc. The list is not exhaustive and the software is still in the middle of its active development phase. This piece of work is also very much a learning process where we are gradually gaining experience in the intricacies of DDI-L (3.2), about the most appropriate elements to use, about the ones to ignore, about the flexibility it offers and about the smooth integration such a standard makes possible.

A healthy portal is a portable portal

Jannik Vestergaard Jensen (Danish National Archives), Stephanie Roth (Swedish National Data Service), Olof Olsson (Swedish National Data Service), Snorre Davoen (NSD - Norwegian Social Science Data Services), Ørnulf Risnes (NSD - Norwegian Social Science Data Services), Sami Granstedt (Finnish Social Science Data Archive)

Under the wings of NordForsk, four Nordic countries gathered together in a series of meetings to make Nordic health care data more visible, discoverable, and searchable for researchers by means of merging and harmonizing metadata.

This resulted in a new prototype search portal system through using previously developed ideas and tools and by adding new concepts into the mix. The session will present the findings and future development plans and hacking tasks for upcoming events within the scope of the project. And, yes, a live demo will be available for the first 15 attending people as well as free source code handouts. We kindly invite you to join the collaboration: http://nordichealthdata.github.io/.

Concurrent E2

Open Source Metadata Accumulator for CESSDA (OS-MAC)

John William Shepherdson (UK Data Archive)

The CESSDA workplan for 2015 is part of a commissioning process that will produce a raft of policies and common infrastructure to benefit both its Users and Service Providers. One work item is for a metadata harvesting tool: this presentation outlines our proposal to create an extensible, Open Source tool that is capable of harvesting metadata in a variety of formats (including DDI Codebook) and making it available for consumption by search Portals and other downstream systems. The vision is that developers will continue to contribute adaptor modules long after the project funding has ended, so that many metadata input and output formats will be supported over time.

DDI-Flat-DB – a lightweight framework for heterogeneous DDI sources

Claus-Peter Klas (GESIS - Leibniz Institute for the Social Sciences), Oliver Hopt (GESIS - Leibniz Institute for the Social Sciences), Alexander Mühlbauer (GESIS - Leibniz Institute for the Social Sciences), Wolfgang Zenk-Möltgen (GESIS - Leibniz Institute for the Social Sciences)

The current usage of DDI is heterogeneous. It varies over different versions of DDI, different grouping, and unequal interpretation of elements. Therefore provider of services based on DDI implement complex database models for each developed application, resulting in high costs and application specific and non-reusable models. To overcome these problems we developed a prototype of an efficient REST-based database storage for DDI.

The main idea behind this architecture is to enable an abstract layer of entities, like StudyUnit, Question, etc. These entities are individually stored in the Flat-DB as XML snippets for access and search. The partitioning rules are given by a configuration file and are not implemented – thus, they can differ even for one specific DDI version. Each entity is reflected in Java as implemented entity, accessing information by configurable XPath expressions. As use cases for the DDI-Flat-DB we worked on three different web-based client applications: A variable search, a study level editor (based on Vaadin UI and Spring), and a mechanism generating printable variable reports (in the STARDAT project based on Freemarker). All share the same entity implementations and the same REST access classes. With our talk, we want to discuss the benefits and drawbacks of this approach.

Posters and Software Demonstrations

Automated Survey Enactment and DDI Integration

Stelios Alexandrakis (Consulent)

SURVANT (SURVey Authoring and eNactment Toolkit) is a suite of software tools that allows technical and non-technical users alike to author a data collection workflow ("survey") and enact it, using software they are already familiar with.

Users author their "survey" specification in Microsoft Word using a simple, structured textual format. SURVANT then behind the scenes parses the text, creating an intermediate language SurvantML (SML), which is used by the engine to enact the "survey", collect the data and perform any custom processing.

SURVANT comes in different flavours depending on where the "survey" is enacted. SURVANT-desktop runs on a desktop, laptop or tablet whereas SURVANT-online runs on a web server on-premises or in the cloud. SURVANT allows integration with external APIs such as REST services or hardware devices and databases. For example, it may be used to extract measurements from a medical device or database in a clinical setting.

SURVANT is designed to be flexible, extensible and interoperable with other systems. Towards this aim, DDI adoption is being considered in the form of a mapping between SurvantML and DDI. This allows SURVANT to import and enact DDI surveys but also export DDI metadata in order to comply with emerging standards.

DDI on Rails

Marcel Hebing (German Socio-economic Panel Study (SOEP), DIW Berlin)

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

Metka – the metadata database for FSD

Katja Moilanen (Finnish Social Science Data Archive), Matti Heinonen (Finnish Social Science Data Archive)

For many years FSD has used plain DDI XML files for marking up and storing both study level and variable level metadata. This has worked well even if the process is error-prone and laborious, sometimes requiring manually editing files that are over 140,000 lines of XML. In addition to DDI files, we have also used a relational database for administrative metadata and file management. Some of the metadata has been stored both in the files and in the database, resulting in the duplication of information and associated problems.

This poster will present FSD moving to a better managed solution, a metadata database called Metka. With Metka, all metadata is gathered to one single authoritative database. The system is based on modern web technologies. Data objects are stored in a relational database with a data model that conforms to DDI. Metka simplifies metadata entry and improves quality as structured fields are used extensively, for example, for controlled vocabularies. All committed changes are automatically versioned and can be audited. Metadata can also be exported to DDI. Extensive search functionality will help customer services when dealing with information requests.

Implementing DDI in the SSJDA – Easy DDI Organizer project and Nesstar system operation

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

We will introduce the two projects for DDI implementation in the SSJDA (Social Science Japan Data Archive). The first is Easy DDI Organizer (EDO), software for survey planning and metadata management. EDO was developed to help researchers and students to edit and manage metadata based on DDI 3.1. File-import/export function is another salient feature of EDO, which supports importing variable level metadata from SPSS files as well as exporting codebook and questionnaire. In this poster session, we will introduce the contexts and features of EDO with some demonstrations provided.

The second project is Nesstar system operation. While the Nesstar system has been widely operated across countries, its full implementation remains to be seen in Japan. The SSJDA has started the operation of Nesstar system since 2012. The number of the published data amounts to about 70. We will discuss the lesson from the past operation and perspectives for the future development.

The Epidemiology-France web portal: a metadata catalogue of French health databases

Simon Saint-Georges (Portail Epidemiologie-France / ITMO Santé Publique (Aviesan/INSERM)

The "Epidémiologie-France" web portal is a collaborative website which records individual French health databases, allows identifying the main health data sources available in France, and provides access to a synoptic description of their essential characteristics.

In order to better work with French and foreign counterparts, a metadata harmonization project was launched in 2014. Later that year, we started working towards a new DDI-compliant version of the portal, which is about to be released by the end of the year.

Today there are more than 800 databases identified and described, most of which are available in both French and English. In 2016 we will conduct an evaluation of our project, and we plan to further develop our partnerships with foreign counterparts, and eventually work on similar platforms on an international level, in order to better inscribe the French research into a global perspec-

tive.

We will discuss past and future developments, exchange on possible perspectives and make a short demonstration of the website.

Using DKAN – an Open Source Portal Solution for Publishing Social Science Data: Lessons learnt at DSZ-BO

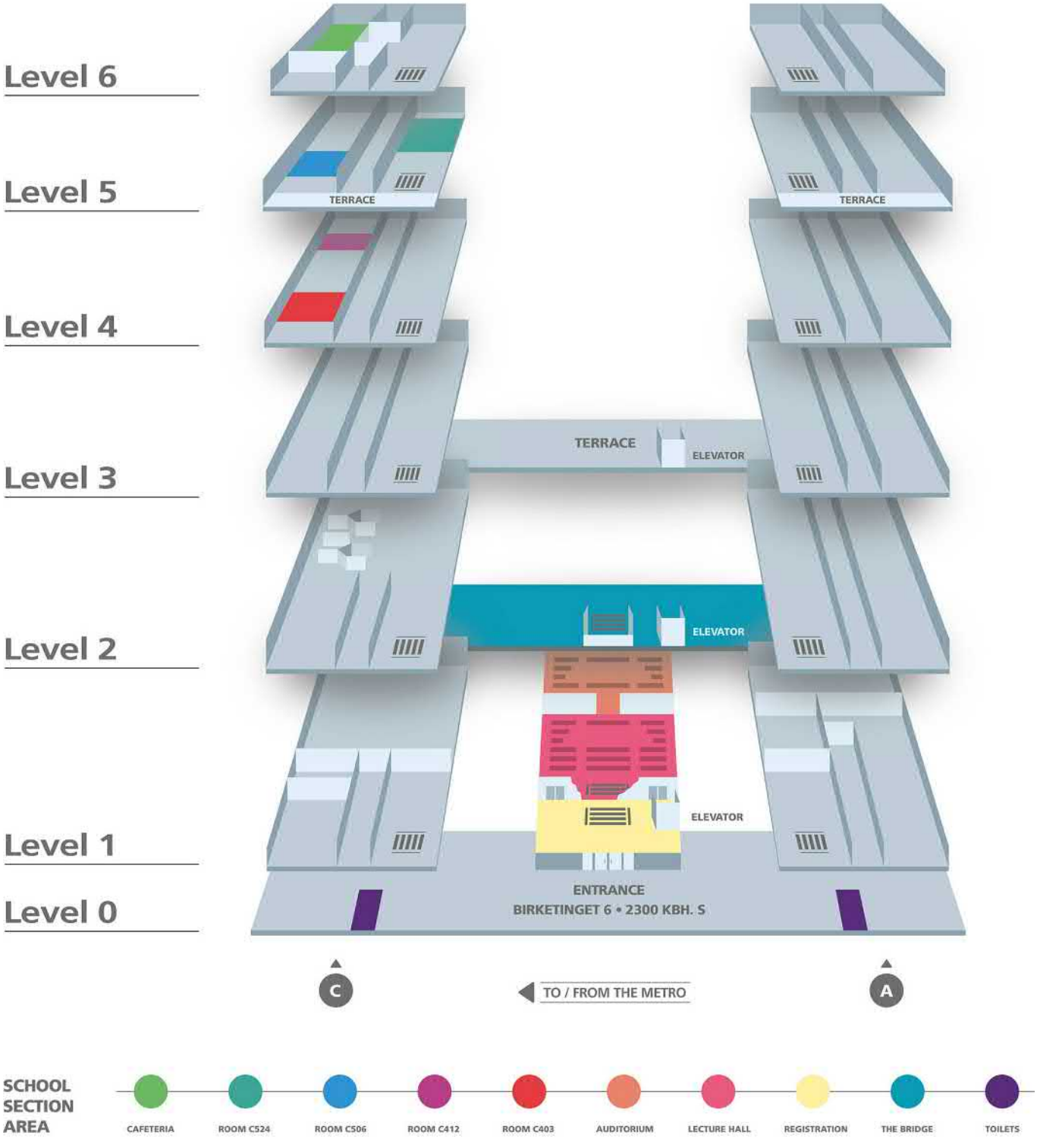
Johanna Vompras (Bielefeld University Library)

In the past few years, more and more institutions, universities, and single researchers aim to provide their data for reuse to the public. From the perspective of universities, there is a huge need for rolling-out data portals across all disciplines -- to enable researchers to document, collect, normalize data, and first of all: to give them simple tools for presenting their research output.

One promising open source solution is DKAN, a Drupal implementation of the popular CKAN data publishing platform, which bundles Drupal core, two content types for representing a dataset (in our case a "study"), resources, and efficient data retrieval with adjustable faceted search. Additionally, it provides features including customizable metadata schemas, metadata-APIs, and dataset harvesting capabilities. For extending the functionalities, all Drupal modules are available for it, like bibliography, backup management, data visualization etc.

This presentation will provide a schematic overview of steps needed to set up DKAN and to migrate existing DDI-3-1-based documentation data and resources of the DSZ-BO (Bielefeld, Germany) into the new portal solution. We will interactively demonstrate the publishing and data access workflows, the import and presentation of DDI XML, and the obtained benefits - both from a technical and economic point of view.

Map of RSLIS (IVA)



Authors

Alexandrakis, Stelios 7,19
Barkow, Ingo 6, 16
Beuster, Benjamin 6, 15
Block, William 6, 16
Blomqvist, Klas 7, 17
Blumenberg, Manuela 6, 16
Borschewski, Kerrin 7, 17
Brown, Warren A. 5, 13
Corgne, Simon Le 5, 14
Cotton, Franck 6, 15
Danciu, Alina 5,14
Dannevang, Flemming 5, 13
Davoen, Snorre 8, 18
Duffes, Guillaume 6, 15
Edwards, Michelle 6, 16
Etheridge, Anne 7, 17
Gierl, Claude 8, 18
Granstedt, Sami 8, 18
Hebing, Marcel 4, 7, 10, 19
Heining, Joerg 5, 13
Heinonen, Matti 7, 19
Hellum, Asbjørn, ... 5
Hopt, Oliver 8, 18
Iverson, Jeremy 4, 6, 10, 14
Jacobs, Stephanie 5, 13
Jensen, Jannik Vestergaard 8, 18
Jensen, Uwe 7, 17
Johnson, Jon 8, 18
Kjeldgaard, Anne Sofie Fink 7, 17
Klas, Claus-Peter 6, 8, 16, 18
Kleemola, Mari 7, 17
Mairot, Alexandre 5, 14
Moilanen, Katja 7, 19
Motegi, Akira 7, 19
Mühlbauer, Alexander 8, 18
Müller, Dana 5, 14
Müller, Stefan 6, 15
Netscher, Sebastian 7, 17
Nielsen, Mogens Grosen 5, 13
Norland, Stig 6, 15
Olsson, Olof 8, 18
Orten, Hilde 6, 15
Pedersen, Henrik 7, 12
Risnes, Ørnulf 5, 7, 8, 14, 16, 18
Roth, Stephanie 8, 18
Saint-Georges, Simon 7, 19
Schild, Christopher Johannes 7, 17
Schiller, David 5, 13, 14
Schweers, Stefan 6, 15
Shepherdson, John William 8, 18
Smith, Dan 4, 5, 10, 13
Southall, Humphrey 5, 15
Stoner, Michael 5, 15
Sørensen, Diana Ransgaard 5, 13
Thomas, Wendy 4, 10
Thygesen, Lars 5, 11
Treffs, Anja 17
Vompras, Johanna 7, 20
Wasner, Catharina 6, 14
Wenzig, Knut 4, 10
Zangenberg, Carsten 5, 11
Zenk-Möltgen, Wolfgang, ... 6, 7, 8, 15, 16, 17, 18

Organisations

Bielefeld University Library 7, 20
Centre for Longitudinal Studies, Institute of Education, UCL, London 8, 18
Colectica 4, 5, 6, 9, 10, 13, 14
Consultant 7, 19
Cornell Institute for Social and Economic Research (CISER), Cornell University 5, 6, 13, 16
Danish National Archives 5, 7, 8, 17, 18
Deutsche Bundesbank, Department Statistics 7, 17
Finnish Social Science Data Archive 7, 8, 17, 18, 19
German Socio-economic Panel Study (SOEP), DIW Berlin, ... 4, 7, 10, 19
GESIS - Leibniz Institute for the Social Sciences 6, 7, 8, 14, 15, 16, 17, 18
HTW Chur 6, 16
INSEE 6, 15
Institute for Employment Research (IAB), Germany 5, 13, 14
Institute of Social Science, The University of Tokyo 7, 19
NSD - Norwegian Social Science Data Services 5, 6, 7, 8, 14, 15, 16, 18
Portail Epidemiologie-France / ITMO Santé Publique (Aviesan/INSERM) 7, 19
Sciences Po – CDSP 5, 14
Statistics Denmark 5, 11, 13
Statistics Sweden 7, 17
Swedish National Data Service 8, 18
Thygesen Statistics Consulting 5, 11
UK Data Archive 7, 8, 17, 18
University of Minnesota Minnesota Population Center 4, 10
University of Portsmouth 5, 15
University of Southern Denmark 7, 12