EDDI EUROPEAN DDI USER CONFERENCE

SOEP DIW | **DDI®** | **gesis** Leibniz Institute for the Social Sciences | **IZA IDSC** INTERNATIONAL DATA SERVICE CENTER

10th Annual European DDI User Conference (EDDI18)

December 4-5, 2018, Berlin, Germany

Hosted by SOEP (The Socio-Economic Panel) at DIW (German Institute for Economic Research)

www.eddi-conferences.eu/eddi18

Schedule and Program with Abstracts

Final version as of December 3, 2018

The conference will start on Tuesday, December 4 at 9:00 and will end on Wednesday, December 5 at 15:15.

Tutorials will take place on Monday, December 3 from 9:00 to 16:15.

Three side meetings will take place in the margins of the conference.

Contents

Venue

Conference:	[Headquarter of Leibniz Association](#) Chausseestraße 111, 10115 Berlin
Side meeting and tutorials:	[DIW](#) - German Institute for Economic Research Mohrenstraße 58, 10117 Berlin

Registration for: **Tutorials on Monday** **Conference**	 DIW, 3rd floor Headquarter of Leibniz Association, 5th floor
Main conference	Headquarter of Leibniz Association (short in program: Leibniz) Rooms: • 5th floor, Hannover (plenary and concurrent 1) • 1st floor, Berlin (concurrent 2)
Side meetings	DIW • M1: 3rd floor, rooms 3.3.002A (Joan Robinson Room) and 3.3.002C • M2: 3rd floor, room 3.3.002A (Joan Robinson Room)
Tutorials on Monday	DIW • T1: 3rd floor, room 3.3.002C • T2: 3rd floor, room 3.3.002C • T3: 3rd floor, room 3.3.002A (Joan Robinson Room)
Conference Dinner on Tuesday	[Palais in der Kulturbrauerei](#) Schönhauser Allee 36, 10435 Berlin
Informal Get-Together	[Zosch](#), Tucholskystr. 30, 10117 Berlin

Venue

Schedule

November 26 – November 30, 2018

9:00 - 17:00	**Side Meeting**
	M1: DDI Specification Development
	DDI Technical Committee and DDI 4 Moving Forward Project – Combined Meeting (abstract) Closed Meeting Location: DIW, 3rd Floor, Rooms 3.3.002A (Joan Robinson Room) and 3.3.002C Organizer: Kelly Chatain, Wendy Thomas, and Joachim Wackerow

Sunday, December 2, 2018

19:00	**Informal Get-Together**
	Location: Zosch, Tucholskystr. 30, 10117 Berlin

Monday, December 3, 2018

9:00 - 17:00	**Side Meeting**
	M2: CESSDA
	CESSDA, Trust, Metadata and Roadmaps ([abstract](#))
	Closed Meeting Location: DIW, 5th Floor, Room 5.2.010 (Anna J. Schwartz Room) Organizer: Herve L'Hours

8:30	Starting Registration - Location: DIW, 3rd floor
	Tutorials

	T1 (half day)	
	Location: DIW, 3rd Floor, Room 3.3.002C	
	Same Same But Different - Documenting Panel Data Using DDI ([abstract](#))	
	Knut Wenzig (German Institute for Economic Research (DIW Berlin) / German Socio-Economic Panel (SOEP))	
9:00 - 10:30	Part 1 of T1	
10:30 - 10:45	**Break** - Location: DIW, 3rd floor	
10:45 - 12:15	Part 2 of T1	
12:15 - 13:15	**Lunch** - Location: DIW, 1st floor, Lounge	
	T2 (half day)	**T3** (half day)
	Location: DIW, 3rd Floor, Room 3.3.002C	Location: DIW, 3rd Floor, Room 3.3.002A (Joan Robinson Room)
	Document Questionnaires and Datasets with DDI: a Hands-on Introduction with Colectica ([abstract](#)) Jeremy Iverson (Colectica)	**Technical Workshop: Implementing DDI4 - Data Description Model** ([abstract](#)) **Cancelled** Darren Stephen Bell (UK Data Archive)
13:15 - 14:45	Part 1 of T2	Part 1 of T3
14:45 - 15:00	**Break** - Location: DIW, 3rd floor	
15:00 - 16:30	Part 2 of T2	Part 2 of T3
18:00	**Informal Get-Together** Location: [Zosch](#), Tucholskystr. 30, 10117 Berlin	

Tuesday, December 4, 2018

8:15	**Starting Registration** Location: Leibniz, 5th floor
9:00 - 10:15	**Conference Plenary P1: Welcome and Keynote** Location: Leibniz, 5th floor, Room Hannover Chairs: Joachim Wackerow (EDDI Conference Co-Chair)

Welcome

Stefan Liebig, Director of SOEP - The Socio-Economic Panel, Member of DIW's Executive Board (German Institute for Economic Research)

Keynote: Making FAIR Data a Reality... and the Challenges of Interoperability and Reusability (abstract)

Simon Hodson, Executive Director of CODATA - Committee on Data for Science and Technology of the International Science Council

10:15 - 10:45	**Break** - Location: Leibniz, 5th floor
10:45 - 12:00	**Concurrent Conference Sessions**

Concurrent A1: **User Needs** Location: Leibniz, 5th floor, Room Hannover Chair: Iris Alfredsson	Concurrent A2: **Questionnaire Design and Documentation** Location: Leibniz, 1st floor, Room Berlin Chair: Jon Johnson
Providing Data in a User-Friendly Manner at the Center for Socio-Political Data Paris (abstract) Alina Danciu, Alexandre Mairot (both Center for Socio-Political Data, Sciences Po Paris)	**The Questionnaire Design and Documentation Tool (QDDT) - a DDI based tool for assisting questionnaire design teams in their work** (abstract) Hilde Orten (NSD - Norwegian Centre for Research Data), Stig Norland (NSD - Norwegian Centre for Research Data), Sarah Butt (European Social Survey (ESS) ERIC Headquarters, City, University of London)
A New Data Publisher for the SSH Community (abstract) David Schiller (TBA21 Germany GmbH)	**The Question Variable Database (QVDB) - a Portal for the ESS** (abstract) Benjamin Beuster (NSD - Norwegian Centre for Research Data)
Linking Web Survey Tool and Data Archives (abstract) Irena Vipavc Brvar (Slovenian Social Science Data Archives (ADP)), Vasja Vehovar (University of Ljubljana), Nejc Berzelak (University of Ljubljana)	**How the CESSDA Euro Question Bank Integrates with Different Technologies** (abstract) Azadeh Mahmoud Hashemi, Katarina Boland, Alexander Mühlbauer, Esra Akdeniz, Wolfgang Zenk-Möltgen (all GESIS - Leibniz Institute for the Social Sciences)

12:00 - 13:15	**Lunch** - Location: Leibniz, ground floor, Atrium

13:15 - 14:45	**Concurrent Conference Sessions**	
	Concurrent B1: **Reusing and Sharing Metadata** Location: Leibniz, 5th floor, Room Hannover Chair: Alina Danciu	Concurrent B2: **Software / Tools** Location: Leibniz, 1st floor, Room Berlin Chair: Hilde Orten
	Finding Ways Between Metadata Models (abstract) Mari Kleemola, Katja Moilanen (both Finnish Social Science Data Archive (FSD)), Maja Dolinar, Irena Vipavc Brvar (both Slovenian Social Science Data Archives (ADP))	**The CESSDA Controlled Vocabular Manager Using DDI-FlatDB** (abstract) Sigit Nugraha, Claus-Peter Klas (both GESIS - Leibniz Institute for the Social Sciences)
	From Shadow to Light: Make Data Visible. DDI, a Collaborative Ground for the Library and the Center for Socio-Political Data (CDSP) of Sciences Po (abstract) Cynthia Pedroja (Library, Sciences Po Paris), Geneviève Michaud (Center for Socio-Political Data, Sciences Po Paris)	**Envisioned Architecture for Empirical Survey Documentation Based on DDI-FlatDB** (abstract) Claus-Peter Klas, Oliver Hopt (both GESIS - Leibniz Institute for the Social Sciences)
	Collect, Store and Re-Use Paradata in DDI – We Need a Working Group! (abstract) David Schiller (TBA21 Germany GmbH)	**Documenting and Publishing Statistical Data with Colectica and DDI** (abstract) Jeremy Iverson (Colectica)
14:45 - 15:15	**Break** - Location: Leibniz, 5th floor	

15:15 - 16:15	Concurrent Conference Sessions	
	Concurrent C1: **Infrastructure** Location: Leibniz, 5th floor, Room Hannover Chair: Jared Lyle	Concurrent C2: **DDI and R** Location: Leibniz, 1st floor, Room Berlin Chair: Knut Wenzig
	CESSDA DataverseEU Project (abstract) Vyacheslav Tykhonov (Data Archiving and Networked Services (DANS))	**DDI R-UI: a Tool to Describe Data in R** (abstract) Domingo Scisci (UniData - Bicocca Data Archive University of Milano-Bicocca)
	Dataverse as a Single Source of Truth in a DDI-Compliant Environment (abstract) Geneviève Michaud, Baptiste Rouxel, Romain Mougin (all Center for Socio-Political Data, Sciences Po Paris)	**Implementing the DDI4 Model in the R Language** (abstract) Larry Hoyle (Institute for Policy & Social Research, University of Kansas), Joachim Wackerow (GESIS - Leibniz Institute for the Social Sciences)
16:15 - 16:45	Birds-of-a-Feather Discussions	
	Concurrent BOF1: Location: Leibniz, 5th floor, Room Hannover	Concurrent BOF2: Location: Leibniz, 1st floor, Room Berlin
	Collaborations between Institutions/Libraries using DDI Wyne Nekesa	**Benefits and Challenges of Implementing the DDI Controlled Vocabularies (CV) at your Institution** Wolfgang Zenk-Möltgen
	Improving Metadata for Projects, Collections and Datasets through Workflow Automation Software Jonas Kahle	**Use of DDI: How to Gather and Improve Empirical Evidence for DDI Usage around the World** Barry Radler
	DataverseEU Vyacheslav Tykhonov	**How Does DDI Support the Researchers in Applying for Third Party Funded Projects, e.g. with a Data Management Plan** Jared Lyle
	Shared Libraries/Software Components for DDI Mari Kleemola	
18:30	**Conference Dinner**	
	Drinks from 18:30, dinner served at 19:00 Location: Palais in der Kulturbrauerei, Schönhauser Allee 36, 10435 Berlin	

Wednesday, December 5, 2018

8:30	Starting Registration Location: Leibniz, 5th floor
9:00 - 10:00	Conference Plenary P2: Panel Location: Leibniz, 5th floor, Room Hannover Chair: Knut Wenzig
	Metadata Sharing: Rights and Obligations (abstract) **Panelists:** Scott Hofer (University of Victoria), Jeremy Iverson (Colectica), Jon Johnson (CLOSER, IOE (UCL Institute of Education, University College London)), Mari Kleemola (FSD - Finnish Social Science Data Archive)
10:00 - 10:30	**Break** - Location: Leibniz, 5th floor
10:30 - 12:00	**Concurrent Conference Sessions**

Concurrent D1: **Documenting Statistical Data** Location: Leibniz, 5th floor, Room Hannover Chair: Mikko Saloila	Concurrent D2: **Applying DDI** Location: Leibniz, 1st floor, Room Berlin Chair: Irena Vipavc Brvar
New Improvements in DDI 3.3 (abstract) Dan Smith (Colectica)	**Not-so-big data: Using DDI to Promote Findable, Re-Usable Data across Small N Studies in the Traumatic Stress Field** (abstract) Nancy Kassam-Adams (Children's Hospital of Philadelphia & University of Pennsylvania)
Process and Rationale for the New DDI 3.3 Version, a User Point of View (abstract) Guillaume Duffes (Insee - Institut National de la Statistique et des Etudes Economiques)	**Challenges of Using DDI in Research Data Sharing in Uganda. A Case of Selected Research Institutions.** (abstract) Winny Nekesa Akullo (Public Procurement and Disposal of Public Assets Authority. IASSIST)
Development of Metadata, Conversion and Archiving of the Time Series Data of the Completed Censuses and Surveys of the BBS. (abstract) Chandra Shekhar Roy (Bangladesh Bureau of Statistics)	**Indexing DDI with ElasticSearch: The ExploreData Project** (abstract) Karam Abdulahhad, Claus-Peter Klas (both GESIS - Leibniz Institute for the Social Sciences)

12:00 - 13:30	**Lunch** - Location: Leibniz, ground floor, Atrium
	Posters and Software Demonstrations While lunch time - Location: Leibniz, ground floor, Atrium - Chair: Andreas Franken
	Planning Feasibility Studies on the Applicability of the DDI Standard for "One Health Surveillance" Data Harmonization in Europe (abstract) Taras Günther, Tasja Buschhardt (both Federal Institute for Risk Assessment), Fernanda Dorea (National Veterinary Institute), Matthias Filter (Federal Institute for Risk Assessment)
	Colectica Portals: Public Use Data on the Web (abstract) Jeremy Iverson, Dan Smith (both Colectica)
	DDI Implementation Projects at SSJDA: Ongoing Projects and a Future Task (abstract) Satoshi Miwa (Institute of Social Science, The University of Tokyo.), Kazuhiro Kezuka (Tokyo Institute of Technology), Tomoyuki Masuko (Institute of Social Science, The University of Tokyo.)
	CLOSER's Metadata Portal: Improving Discovery (abstract) Hayley Mills (CLOSER Project, Department of Social Science, UCL Institute of Education)
	An R Package for Putting DDI to Practical Use as a Personal Tool : DDIR (abstract) Yasuto Nakano (School of Sociology, Kwansei Gakuin University)
13:30 - 14:30	**Concurrent Conference Sessions**

13:30 - 14:30	Concurrent E1: **Rich Metadata** Location: Leibniz, 5th floor, Room Hannover Chair: Wendy Thomas	Concurrent E2: **Metadata Discovery and Harvesting** Location: Leibniz, 1st floor, Room Berlin Chair: Uwe Jensen
	Powering an Energy Research Portal with a DDI Property Graph (abstract) Deirdre Lungley, Darren Bell (both UK Data Archive)	**The New CESSDA Data Catalogue** (abstract) John Shepherdson (CESSDA ERIC), Carsten Thiel (CESSDA ERIC)
	Colectica Portals: Public Use Data on the Web (abstract) Jeremy Iverson (Colectica)	**Kuha2 - A New Open Source DDI Aware Metadata Server** (abstract) Toni Erik Sissala, Matti Heinonen (both Finnish Social Science Data Archive (FSD))

14:30 - 14:45	**Break** - Location: Leibniz, 5th floor
14:45 - 15:30	Conference Plenary P3: Reports on DDI Specifications, and Outlook Location: Leibniz, 5th floor, Room Hannover Chair: Jon Johnson (EDDI Conference Co-Chair)
	Report of the Birds-of-a-Feather Sessions
	News from the DDI Alliance Technical Committee
	Update on DDI 4
	Announcement of 7th NADDI, Invitation to EDDI19 and Goodbye Jared Lyle, Jon Johnson, and Next Year's Host (name will be disclosed in session)
18:00	**Informal Get-Together** Location: Zosch, Tucholskystr. 30, 10117 Berlin

Thursday, December 6, 2018

9:00 - 16:00	**Side Meeting**
	M3: CESSDA
	CESSDA Expert Seminar 2018 Follow-Up: practical session on Jenkins pipelines, Docker containers, Kubernetes clusters (for details see event page.)
	Closed Meeting
	Location: DIW, 3rd Floor, Room 3.3.002A (Joan Robinson Room)
	Organizer: John Shepherdson (CESSDA ERIC)

Keynote

Making FAIR Data a Reality… and the Challenges of Interoperability and Reusability

Simon Hodson (CODATA - Committee on Data for Science and Technology of the International Science Council) (↑ schedule)

Track: Reusing and Sharing Metadata – **Session Type:** Keynote

Building on previous summaries of the attributes that make research data most usable by other scientists (for replication, reanalysis and reuse), the framework provided by the FAIR data principles has gained support from policy makers, funders and researchers and has helped progress towards the vision of Open Science. Combined with the principle that research data should be 'as open as possible, as closed as necessary', FAIR data has allowed important principles to be communicated in a more accessible and persuasive way.

This talk will first explore how to implement the FAIR data principles and related and enabling practices, with reference to the work of the European Commission's Expert Group on FAIR data which is developing a report and Action Plan on how to make FAIR data a reality.

Second, the talk will consider the challenges of interoperability and reusability in relation to the aspirations of those interdisciplinary and transdisciplinary research areas that seek to address the most important global, environmental and societal challenges. Encouraging the development and growing maturity of vocabularies and standards required for interoperability and the integration of data from diverse disciplines will be a major challenge for 21st century research. An initiative of CODATA and the new International Science Council seeks to shed light on ongoing challenges of data availability and integration, taking research into infectious diseases, disaster risk and resilient cities as test cases.

Discussion Sessions

Metadata Sharing: Rights and Obligations

Jon Johnson (UK Data Service) (↑ schedule)
Track: Reusing and Sharing Metadata – **Session Type:** Panel Discussion

As metadata and data converge, particularly where evolving standards such as DDI4 can simultaneously be the data, the code that generated it and its metadata; it is apposite to revisit questions of intellectual copyright, licencing and sharing in a way that a metadata in the form of catalogue record has hitherto not.

Developing and maintaining such granularity of metadata can be an expensive proposition. There are costs associated with editing and publishing data and metadata.

Whilst most institutions have a data policy, very few have a metadata policy and those that do, seem to be aligning on a policy that advocates or requires a Creative Commons Zero (CC0) licence. The perceived advantages of having such a policy are that it provides a clear and easy-to-understand guidance for users especially where metadata is shared across borders.

There are obligations which could reasonably be requested of those to whom metadata is shared. If it is being re-shared, is it up-to-date with the original source, has it been altered and is that transparent to a user, have derived products attributed the original source?

This session will seek to identify the main challenges that a step change in the nature of metadata means for providers and recipients of shared metadata and the subsequent technical challenges this may engender.

Birds-of-a-Feather Discussions

TBA

Presentations in Concurrent Sessions

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

Indexing DDI with ElasticSearch: The ExploreData Project

Karam Abdulahhad (GESIS - Leibniz Institute for the Social Sciences), Claus-Peter Klas (GESIS - Leibniz Institute for the Social Sciences) (↑ schedule)
Track: Software / Tools – **Session Type:** Regular Presentation - Project Report

DDI 3.2 presents a solid meta-data model to encode social sciences surveys meta-data, such as studies and variables. However, making this data accessible is another story. In Explore-Data project, our goal is to present a dynamic, multi-facets, and multilingual search portal for both studies and variables. Accordingly, we transform the content of DDI, available via DDI-FlatDB, into ElasticSearch index.

We consider two main document types in our index, namely study and variable. Each document type consists of a set of fields in order to keep as much of study/variable structure as possible. To support the multilingual content, we build many copies of each field for each language.

During the indexing process, we were obliged to deal with migration of data, missing values, remove duplicated and redundant data, and sometimes to generate new data. Index content can be queried via classical Google-style queries. However, the same content can be also browsed via many pre-defined facets in an e-Commerce websites style. Both ways can be used simultaneously in order to give to users a tool to dynamically access data and filter result list.

We have indexed about 6000 studies, distributed on 13 study collections such as ALLBUS, European Values Study, etc. In addition we currently index about 150K variables. In our presentation we introduce the Explore Data project at GESIS and describe indexing challenges to transform DDI Lifecycle 3.2 to ElasticSearch on study and variable level.

Challenges of Using DDI in Research Data Sharing in Uganda. A Case of Selected Research Institutions.

Winny Nekesa Akullo (Public Procurement and Disposal of Public Assets Authority. IASSIST) (↑ schedule)
Track: Open Data and Linked Open Data – **Session Type:** Regular Presentation - Community Impact

Uganda has a number of institutions that carry out research both for decision making and evidence based planning. This data is described using various standards and is disseminated to various stakeholders. Some of these stakeholders who are researchers re-use the data for further analysis. The Data Documentation Initiative (DDI) on the other hand is an international standard for describing the data produced by surveys and other observational methods in the social, behavioral, economic, and health sciences. DDI is a free standard that can document and manage different stages in the research data lifecycle, such as conceptualization, collection, processing, distribution, discovery, and archiving. This paper shall examine the challenges the research institutions face when sharing research data. The objectives of the research are; to find out whether the research institutions use DDI for data sharing; to find out the challenges they face; and propose ways in which they can use DDI to improve data sharing.

The Question Variable Database (QVDB) - a Portal for the ESS

Benjamin Beuster (NSD - Norwegian Centre for Research Data) (↑ schedule)
Track: Reusing and Sharing Metadata – **Session Type:** Regular Presentation - Project Report

The Question Variable Database (QVDB) with the Colectica platform as technical backbone can be described as a system for storage and retrieval of questions and variables, and facilitating reuse of their metadata and metadata components. The overall aim of the QVDB is to serve the ESS, and potentially other survey programmes, in their business processes of specifying, documenting, versioning and disseminating survey data.

The European Social Survey (ESS) is an academically driven cross-national survey that has been conducted across Europe since its establishment in 2001. Every two years, face-to-face interviews are conducted with newly selected, cross-sectional samples.

In particular, this presentation describes how the QVDB was populated with 8 waves of ESS data including the transformation from DDI-C into DDI3.2 and it gives examples on how variables from different points in time or different datasets correspond to each other using the three levels structure of ConceptualVariable, RepresentedVariable and Variable (DDI3.2 variable harmonization)

Providing Data in a User-Friendly Manner at the Center for Socio-Political Data Paris

Alina Danciu (Center for Socio-Political Data, Sciences Po Paris), Alexandre Mairot (Center for Socio-Political Data, Sciences Po Paris) (↑ schedule)
Track: User Needs, Efficient Infrastructures and Improved Quality – **Session Type:** Regular Presentation - Project Report

The Center for Socio-Political Data (CDSP) of Sciences Po has been providing access to social sciences data resources from 2005. Our goal over the next year is to evolve our environment to a more user-friendly DDI-compliant structure.

In the context of the new French law "Loi du numérique" from 2017, we need to question our quantitative data access procedures. We need to evaluate which are the resources that can be downloaded in open access and which ones need to be restricted in terms of access, in order to preserve the confidentiality of the survey respondents.

Another element to take into account in order to provide data in a user-friendly manner is the improvement of the quality of our metadata. Our unequal interpretation of the DDI elements in our environment needs to be addressed in order to improve the user's data discovery and comparative analysis experience: a metadata harmonization process is currently ongoing (use of controlled vocabularies, metadata re-use…)

This presentation will highlight the processes, challenges and questions for providing easier data access to our users' community on a DDI-compliant platform.

Process and Rationale for the New DDI 3.3 Version, a User Point of View

Guillaume Duffes (Insee - Institut National de la Statistique et des Etudes Economiques) (↑ schedule)
Track: Official Statistics – **Session Type:** Regular Presentation - Community Impact

Insee has made a heavy use of DDI 3.2 for several years, especially the data collection and logical product parts.

Although DDI 3.2 is a major enhancement from DDI 3.1 and has proven to be quite complete, some specific features were considered as missing by Insee. They were then reported back to the DDI Technical Committee over the course of years.

This presentation will:

- Describe the interaction process between the DDI Technical Committee and Insee
- Give an overview of the different requests for change issued by Insee
- Depict which ones were integrated in DDI 3.3 and were not, and the reasons why
- Outline the process Insee intends to set up to integrate the official DDI 3.3 schemas as soon as they are published.

Implementing the DDI4 Model in the R Language

Larry Hoyle (Institute for Policy & Social Research, University of Kansas), Joachim Wackerow (GESIS - Leibniz Institute for the Social Sciences) (↑ schedule)
Track: Software / Tools – **Session Type:** Regular Presentation - Project Report

Model based DDI4 lends itself to a representation in a language with object oriented structures. This presentation describes an ongoing project implementing the complete DDI4 model in R. We currently have the complete structure represented as R6 objects. Future work will focus on a graphical user interface for entering metadata and on capturing metadata about transformations. Import and export of DDI and a human readable codebook are also planned.

Each DDI4 class is modeled as an (object oriented) R6 class. This approach yields several useful capabilities. Properties and relationships of objects are validated as to correct datatype and cardinality as they are instantiated. Existence of DDI URNs is validated against an internal registry and via an XML catalog. The latter allows for validating the existence of an external object. DDI metadata are attached to existing R objects as attributes having DDI URNs as their values.

The DDI4R package, including documentation, is generated programmatically directly from the DDI4 model XMI file available on the DDI4 website.

Having metadata objects in R may also offer interesting future possibilities for operations directly on metadata. Sum and difference operators on Collections, for example might be useful in harmonization.

Documenting and Publishing Statistical Data with Colectica and DDI

Jeremy Iverson (Colectica) (↑ schedule)
Track: Software / Tools – **Session Type:** Regular Presentation - Community Impact

Colectica is software used to document and publish statistical data using open standards. The software is used by national statistical organizations and major longitudinal studies worldwide.

Colectica provides several tools:

- Colectica Questionnaires for specifying surveys in a standard way
- Colectica Datasets for documenting SAS, SPSS, Stata, and other statistical datasets
- Colectica Designer for documenting the entire data lifecycle
- Colectica Portal for pushing richly-documented data on the Web, with full variable-level lineage and concordance over time

This presentation will provide an overview of the tools, show how they are used in production at various statistical agencies and research projects, and will highlight new functionality available in 2018.

Colectica Portals: Public Use Data on the Web

Jeremy Iverson (Colectica) (↑ schedule)
Track: Software / Tools – **Session Type:** Regular Presentation - Community Impact

Colectica Portal is built to search and display data documentation for longitudinal and repeated data collections. Variables can be harmonized and linked across study waves, and changes in measurements, coding, and descriptions are encoded using open data documentation standards such as DDI. Questionnaires from multiple waves can be documented, including sections, routing and filter questions, and linked to the data which was collected.

Colectica Portal uses Colectica Repository to store and index the graph of information used to describe these complex studies

Several national statistical organizations and international longitudinal studies publish rich metadata on the Web using Colectica Portals. This presentation will demonstrate several of these portals, discuss how information was ingested into the repositories for display online, and explore how off-the-shelf, general purpose software can provide a resource-efficient way to publish data with rich, interactive, searchable codebooks on the Web.

Not-so-big data: Using DDI to Promote Findable, Re-Usable Data across Small N Studies in the Traumatic Stress Field

Nancy Kassam-Adams (Children's Hospital of Philadelphia & University of Pennsylvania) (↑ schedule)
Track: Data Harmonization – **Session Type:** Regular Presentation - Project Report

Beyond big data or longitudinal studies, DDI may be useful with the 'long tail' of smaller research studies, where data re-use can provide valuable insights beyond the original study findings. The traumatic stress field presents an exemplar use case - while studies yield rich and potentially re-usable data, datasets are not easily discoverable and few have metadata that readily enables harmonization and re-use. For child acute trauma (e.g., disaster, injury, violence), understanding risk and protective factors for traumatic stress and health outcomes requires prospective studies that follow children identified soon after the event. These resource-intensive studies generally enroll small samples (i.e., N's of 50 to 500) and employ similar but non-identical measurement strategies; cross-study analyses of pooled and harmonized data can add great value. The Prospective studies of Acute Child Trauma and Recovery (PACT/R) Data Archive (www.childtraumadata.org) is an international collaborative effort to create a research resource for this field, with a searchable online portal and harmonization algorithms. PACT/R currently includes 26 datasets representing >3700 trauma-exposed children. This project report will (a) illustrate initial application of DDI metadata to the PACT/R Archive, and (b) describe progress and challenges in applying DDI to traumatic stress data to facilitate discoverability and harmonization.

Envisioned Architecture for Empirical Survey Documentation Based on DDI-FlatDB

Claus-Peter Klas (GESIS - Leibniz Institute for the Social Sciences), Oliver Hopt (GESIS - Leibniz Institute for the Social Sciences) (↑ schedule)
Track: Software / Tools – **Session Type:** Regular Presentation - Project Report

To document an empirical survey for understanding, findability and re-use several steps are necessary. In our case we start with an existing statistical file containing the survey data. The documentation starts with importing variable names and labels, answer codes and categories using R and storing these informations in corresponding DDI 3.2 Lifecycle elements using the DDI-FlatDB.

In addition variable statistics e.g. in form of frequency tables can be automatically created and stored. Next step is to newly enter or connect an existing questionnaire using the questionnaire editor. After these steps a deep documentation on variable level will be possible through a variable editor incl. connecting controlled vocabularies using the CESSDA CV manager. Furthermore we will integrate a table

generator, to create more sophisticated cross-reference, frequency or descriptive tables. For more specific tables, we integrate direct R script handling via RServe.

In this presentation we will describe the planed architecture, tools and editors. All planed services are implemented to be collaborative, multilingual web applications and directly support multilinguality on the documentation level. Search and browse of study- and variable-level is provided within the ExploreData project.

Finding Ways Between Metadata Models

Mari Kleemola (Finnish Social Science Data Archive (FSD)), Katja Moilanen (Finnish Social Science Data Archive (FSD)), Maja Dolinar (Slovenian Social Science Data Archives (ADP)), Irena Vipavc Brvar (Slovenian Social Science Data Archives (ADP)) (↑ schedule)
Track: User Needs, Efficient Infrastructures and Improved Quality – **Session Type:** Regular Presentation - Community Impact

In an era of open data and research data management movements, the research community is facing a growing number of metadata models and standards, recommendations and requirements for research data. It's enough to make your head spin and get lost.

This presentation provides guidance for surviving in the metadata jungle. We will present findings from a crosswalk of metadata models relevant to CESSDA ERIC, European data archives, and researchers in the social sciences. These include, for example, OpenAIRE, the DDI-based CESSDA Data Catalogue metadata and CESSDA Metadata Model, Dublin Core, DataCite, da|ra and Dataverse. We will discuss the advantages and disadvantages of the 'lowest common denominator' approach to resource metadata compared to a 'rich disciplinary metadata' approach like the DDI, and provide some examples of data services and their take on metadata.

The presentation builds on the work of CESSDA's Metadata Management and DataverseEU projects, and Data Without Boundaries project.

Powering an Energy Research Portal with a DDI Property Graph

Deirdre Lungley (UK Data Archive), Darren Bell (UK Data Archive) (↑ schedule)
Track: Other – **Session Type:** Regular Presentation - Project Report

The UK Data Service in collaboration with University College London is building an infrastructure to allow energy researchers to access near real-time domestic energy consumption and link this with relevant contextual data e.g., Energy Performance Certificates, weather data, social science data, and conceptual and geographical ontologies.

This data is stored and queried using an open source property graph – Janusgraph, backed with an open source NoSQL database – HBase. The UKDS has chosen DDI, in particular DDI4, as the schema to store this data, enabling powerful variable linkage.

This presentation aims to illustrate how large volumes of data stored in a distributed system using such a schema, can empower researchers to join, filter and aggregate multiple datasets to create a data product suitable to answer their research questions. DDI4's variable cascade is used to signpost linkage options and the datum level granularity of DDI4's Logical Data Description package is utilised to store both energy and weather readings.

How the CESSDA Euro Question Bank Integrates with Different Technologies

Azadeh Mahmoud Hashemi (GESIS - Leibniz Institute for the Social Sciences), Katarina Boland (GESIS - Leibniz Institute for the Social Sciences), Alexander Mühlbauer (GESIS - Leibniz Institute for the Social Sciences), Esra Akdeniz (GESIS - Leibniz Institute for the Social Sciences), Wolfgang Zenk-Möltgen (GESIS - Leibniz Institute for the Social Sciences) (↑ schedule)
Track: Reusing and Sharing Metadata – **Session Type:** Regular Presentation - Project Report

The CESSDA Euro Question Bank (EQB) project provides a central search facility across all CESSDA's survey questions in different languages, as shown at EDDI last year. It uses a metadata schema based on DDI-Lifecycle standard and also provides conversion mappings from DDI-Codebook. The latest work in EQB was to finalize this metadata schema in close collaboration with the CESSDA CMM2 project and is now one of the major project results. It combines an object structure that is not too complicated with properties that are compatible to the DDI schema. Using this metadata schema, CESSDA service providers can supply their contents from various formats to EQB.

The EQB system architecture consists of the EQB-Frontend, implemented with the Vaadin user interface technology, and the EQB-Backend, using Elasticsearch for the search index and the FlatDB relational database on MySQL to store DDI snippets. In addition, EQB uses a challenging technology called Docker. Docker is an operating system virtualization tool to manage, centralize, ease and deploy heterogeneous applications and services. The presentation will show the different components of the EQB application, focusing on new developments during the last year and implementation issues with the DDI metadata used.

Dataverse as a Single Source of Truth in a DDI-Compliant Environment

Geneviève Michaud (Center for Socio-Political Data, Sciences Po Paris), Baptiste Rouxel (Center for Socio-Political Data, Sciences Po Paris), Romain Mougin (Center for Socio-Political Data, Sciences Po Paris) (↑ schedule)
Track: Software / Tools – **Session Type:** Short Presentation - Project Report

At the Center for Socio-Political Data of Sciences Po, DDI-documented data is published on several repositories. The publication workflow is not user-friendly for the data managers, as it has several entry points and duplicated processes. Additionally, there is an important risk of having platform-dependent versions.

In order to preserve the system's consistency and integrity, a Single Source of Truth (SSOT) is needed.

To solve this problem, we designed a new infrastructure and workflow, coherent and interoperable by design, and compliant with both DDI and the FAIR principles. We focused on a flexible, upgradable infrastructure. After a thorough review of available solutions, we opted for a Dataverse repository as our SSOT. We are currently in the process of building a tailor-made data ecosystem, through the use of APIs querying our single repository.

This presentation will focus on the processes and technologies used to set up this viable DDI-compliant platform and on the reasons for choosing this solution.

The CESSDA Controlled Vocabular Manager Using DDI-FlatDB

Sigit Nugraha (GESIS - Leibniz Institute for the Social Sciences), Claus-Peter Klas (GESIS - Leibniz Institute for the Social Sciences) (↑ schedule)
Track: Software / Tools – **Session Type:** Regular Presentation - Project Report

CESSDA CV-Manager is a tool that provides comprehensive functionalities to manage hierarchical controlled vocabularies (CVs).

This tool manages the entire workflow process of a CV, from creation, review process until its publication, as well as providing versioning and mulitlinguality functionalities, which allows the collaborative CVs management in an easy and efficient way.

CESSDA CV-Manager is based on DDI Flat-DB, an efficient REST-based database storage for DDI. Each published CV is stored as XML inside DDI Flat-DB documented as SKOS RDF vocabulary.

The published CVs is indexed with ElasticSearch for user-friendly searching. In addition to the RDF format, the tool also allows CVs to be exported to other formats such as PDF and HTML. To incorporate the CVs into your own documentation tools, several REST endpoints are provided, such as selecting CVs, CV code suggestion or to fetch the whole CV for browsing purposes.

In the presentation the productive CESSDA CV manager will be described and we will discuss how to adopt and operationalize other complex XML based meta-data schema such as SKOS using DDI-FlatDB. As a next step we plan to extend the CESSDA CV manager to become a full flavored collaborative thesaurus manager.

The Questionnaire Design and Documentation Tool (QDDT) - a DDI based tool for assisting questionnaire design teams in their work

Hilde Orten (NSD - Norwegian Centre for Research Data), Stig Norland (NSD - Norwegian Centre for Research Data), Sarah Butt (European Social Survey (ESS) ERIC Headquarters, City, University of London) (↑ schedule)
Track: Software / Tools – **Session Type:** Regular Presentation - Project Report

Successful large-scale international survey projects often put lots of effort into the questionnaire design and development process. The QDDT has been developed to assist the questionnaire development process by allowing questionnaire design teams to reuse metadata elements at different levels, to maintain them, to publish them and to track the development of metadata elements over time. The tool is developed and tested as part of the EU funded DASISH and SERISS projects.

The first version of this DDI based tool has just been released, and will be used in the questionnaire design process of the European Social Survey in round 10. This presentation focuses on how the workflow of questionnaire designers is supported by the DDI based design of the tool.

From Shadow to Light: Make Data Visible. DDI, a Collaborative Ground for the Library and the Center for Socio-Political Data (CDSP) of Sciences Po

Cynthia Pedroja (Library, Sciences Po Paris), Geneviève Michaud (Center for Socio-Political Data, Sciences Po Paris) (↑ schedule)
Track: Reusing and Sharing Metadata – **Session Type:** Regular Presentation - Project Report

With a catalogue of nearly one million documents and a wide range of digital resources, Sciences Po's Library is recognized as a reference for the social sciences in Europe.

60% of its references are provided in foreign languages.

Numerous research teams produce actively data each year (wide thematic range: political science, economy, sociology, history, …).

Until recently, data was "hidden": no curation, preservation, documentation, management was formally planned by the research teams. From 2015, a new project was set up by Sciences Po, whose purpose is to curate data and add value to them by using metadata (DDI, MODS, METS, etc.) to make them visible and maintained over time.

The CDSP, which is specialized in the curation and dissemination of quantitative and qualitative SHS surveys and datasets, collaborates with the Library on the technical aspects of the set-up of Sciences Po's institutional repository.

This presentation will focus on three points: How to integrate DDI to an existing metadata schema at the Library? How about the interoperability between DDI and other existing standards and information systems at the library? How to collaborate with the other Sciences Po research and data centers? Is the "library's DDI" compatible with "CDSP's DDI"?

Development of Metadata, Conversion and Archiving of the Time Series Data of the Completed Censuses and Surveys of the BBS.

Chandra Shekhar Roy (Bangladesh Bureau of Statistics) (↑ schedule)
Track: Official Statistics – **Session Type:** Regular Presentation - Project Report

Bangladesh Bureau of Statistics (BBS) has vowed to convert Census/Surveys micro data for next generation technology format use processed in IBM proprietary technology. After Bangladesh's independence, there is a rich repository of statistical information in IBM 360 to ES/9000 mainframe tapes, dating back to late 1970s and early 2000s time.

The overarching objective is to strengthen the prevailing national statistical archiving system. BBS will be making available of this large volume of converted data to the citizens of Bangladesh (if needed world community) so that academic and scholarly debates can take place taking cognizance of historical data.

This is a big step towards dissemination of Big Data covering Bangladesh's economic developmental trend since its birth in 1971. By revisiting time series data, it is hoped that well-informed and meticulous policies can be designed and formulated in the future. Most fundamentally, availability and easy accessibility to such a large volume of Big Data will inspire reassessing economic theories and indicators of development informing Bangladesh's position in global rankings like SDG.

Key Words: mainframe, historical data, IBM, archiving, SDG.

Collect, Store and Re-Use Paradata in DDI – We Need a Working Group!

David Schiller (TBA21 Germany GmbH) (↑ schedule)
Track: User Needs, Efficient Infrastructures and Improved Quality – **Session Type:** Regular Presentation - Community Impact

Metadata is data about data? And what is paradata then? If metadata describes the already collected research data, describes paradata the process of data collection? If so, paradata is highly important to assess the quality of research data; a crucial topic for researchers. At the same time rich paradata can help to evaluate and improve the data collection process itself. Resulting in better data quality, better research and better knowledge. So fine, paradata seems to be important but how is paradata collected? As written notes? Standardized? If the interviewer had time to take notes? Is it collected at all? Why not improving research by collecting paradata using software (within a tool chain) and store it in a useful and common standard, let's say: DDI? Strange, a topic that seems to be of high importance and at the same time so many question marks. What should be done? Maybe a working group would be good!

This talk will give an introduction into the issue of using paradata to improve research. Best practice ways to proceed will be discussed. The goal is to establish a working group that finds solutions that allow to deal with paradata using DDI and related software products.

A New Data Publisher for the SSH Community

David Schiller (TBA21 Germany GmbH) (↑ schedule)
Track: User Needs, Efficient Infrastructures and Improved Quality – **Session Type:** Regular Presentation - Community Impact

A data publisher for Social Science and Humanities is under development. In 2018 a consortium of currently three partners started the development of a new tool to support researcher and data providers.

The software will not be developed for a special institution. The goal is to build a generalized tool that will cover the main needs of a wider community. For example by providing a main component coffering the general functionality needed. In addition special modules will be developed for special demands. Those modules need to be developed alongside with partners defining those special demands.

Some of the main features are: a modern responsive design, supporting modern standards and multilinguality. Frameworks and libraries used should be stable, well tested, open source and widely used. The software should support extensibility through plugins both in frontend and backend. Import formats

for the portal should support outputs from Nesstar publisher (DDI Codebook 1.2.2), CESSDA (CMM), SPSS, Stata, SAS, various CAPI/CATI software. Export functionalities will cover at least DDI 3.2 and 3.3. The requirements for running the portal should be kept as simple as possible and rely only on free software.

The presentation will give an overview on planned features, service and support; and may also open up the possibility to widen the current consortium.

DDI R-UI: a Tool to Describe Data in R
Domingo Scisci (UniData - Bicocca Data Archive University of Milano-Bicocca) (↑ schedule)
Track: Software / Tools – **Session Type:** Regular Presentation - Project Report

DDI R-UI (provisional name) is a graphical interface to document data inside R, based on the DDI (Codebook) standard. Such tool can take advantage of the power of R software and its packages in describing, managing and visualising statistical data. The Italian SSDA UniData has been managing its data archiving workflow with the R software for several years, so the DDI R-UI can meet the need of documenting data directly within the same environment. Why R? Because it's open, it's free, it's easy to expand (new tools can be packaged to ease the installation and update) and tailor to its own needs, no any additional software/server installation is required, and it can be used both in local and remote environments. At EDDI18 an early implementation of DDI R-UI is presented, that includes the following features: metadata in multiple languages, controlled vocabularies, export in XML format, and export in JSON format for publication in Dataverse. Although DDI R-UI has been created for the specific needs of UniData, it is thought that it can be easily adapted for its use in other small and/or low-resource archives. Lastly, some ideas for further development of the tool will be presented.

The New CESSDA Data Catalogue
John Shepherdson (CESSDA ERIC), Carsten Thiel (CESSDA ERIC) (↑ schedule)
Track: Software / Tools – **Session Type:** Regular Presentation - Project Report

The new CESSDA Data Catalogue provides a unified multilingual search/browse interface to Social Science and Humanities (SSH) data stored in decentralised and specialised data centres operated by CESSDA Service Providers. The metadata describing the Service Providers' holdings is harvested via Open APIs and indexed (by language) to provide an aggregated collection covering a wide range of institutions.

This work relates to two of CESSDA's four pillars, namely Technology and Tools & Services (the other two being Training and Trust). By building a modern cloud-based infrastructure, CESSDA is able to operate its services efficiently and effectively without dependency on any single national infrastructure. The adoption of state-of-the containerisation methods ensures the scalability needed to serve the European SSH research community. Secondly, by mapping existing DDI metadata formats to the CESSDA Metadata Model, existing well established (e.g. OAI-PMH) interfaces and exchange methods can be seamlessly harnessed to underpin this addition to CESSDA's suite of user friendly tools & services.

Kuha2 - A New Open Source DDI Aware Metadata Server
Toni Erik Sissala (Finnish Social Science Data Archive (FSD)), Matti Heinonen (Finnish Social Science Data Archive (FSD)) (↑ schedule)
Track: Software / Tools – **Session Type:** Regular Presentation - Project Report

Kuha2 is a metadata server that provides descriptive social science research metadata for harvesting via multiple protocols and a variety of metadata standards. Kuha is still evolving but it has been in production use at FSD from April 2018, serving metadata e.g. for the CESSDA Data Catalogue. The system is open source and released under EUPL license.

Kuha supports OAI-PMH and OSMH for harvesting. Metadata output is provided in DDI-C 2.5 and unqualified DC for OAI-PMH and in JSON for OSMH. Input is through a RESTful API in JSON or by utilizing

built-in mappings for DDI 1.2.2 and 2.5 XML. Kuha's internal metadata model comprises CESSDA Metadata Model v1 mandatory elements.

Kuha2 could provide archives with an easy-to-use entry point for integrating into CESSDA Data Catalogue. This presentation will discuss Kuha2's features, its usability for prospective user and our future plans with it.

New Improvements in DDI 3.3
Dan Smith (Colectica) (↑ schedule)
Track: User Needs, Efficient Infrastructures and Improved Quality – **Session Type:** Regular Presentation - Community Impact

The upcoming release of the DDI Lifecycle 3.3 standard includes many new content areas, and makes it easier for developers to build applications that use the standard. This talk will review the new content available in DDI 3.3 in the following areas: Classification management (based on GSIM/Neuchatel), Non-survey data collection (Measurements), Sampling and Weighting, Questionnaire Design, Support for working with DDI as a Property Graph (properties on items and references), and the Quality Statement improvements (useful for Eurostat reporting). In addition, the talk will present the now-complete formal model that is behind DDI 3.3 and how it was used to automatically generate the 3.3 documentation.

Linking Web Survey Tool and Data Archives
Irena Vipavc Brvar (Slovenian Social Science Data Archives (ADP)), Vasja Vehovar (University of Ljubljana), Nejc Berzelak (University of Ljubljana) (↑ schedule)
Track: User Needs, Efficient Infrastructures and Improved Quality – **Session Type:** Regular Presentation - Community Impact

I'm sure you already heard the saying, the less work researcher will have the more likely it is survey data will actually reach data archive. At the Slovenian Social Science Data Archives (ADP), we wanted to take one step closer to researchers. We will present the work done in cooperating with the Centre of Social informatics (CSI) at the Faculty of Social Sciences, University of Ljubljana. The centre develops an open source tool for online surveys, 1KA – One click survey (www.1ka.si). As of 2018, 1KA has over 44,000 users that annually collect more than one million questionnaires. It is used by a large number of Slovenian researchers from all sectors as well as PhD students conducting online surveys for their research.

The ambition of ADP and CSI is to enable export of question level metadata, paradata, and general survey information (name, language, survey creator, field period, questionnaire duration etc.) in DDI 2.5 format that could be directly deposited in the archive. The presentation will outline the approach to developing the export of desired survey characteristics from the survey software, demonstrate its implementation and expose the key challenges related to exporting sufficiently detailed and consistent survey description in the DDI format.

CESSDA DataverseEU Project
Vyacheslav Tykhonov (Data Archiving and Networked Services (DANS)) (↑ schedule)
Track: Software / Tools – **Session Type:** Regular Presentation - Project Report

One of the projects for 2018 within the Consortium of European Social Science Data Archives (CESSDA ERIC) is the DataverseEU project. The aim of the project is to develop a service for CESSDA Service Providers, based on the Dataverse software developed by the Harvard Institute for Quantitative Social Science (IQSS), that would be adjusted to the needs of the CESSDA community, especially small and/or emerging archives. To fulfill this, additional functionalities of the Dataverse software are planned to be developed within the project. Project activities in 2018 include the installation of the software using a Docker container on the CESSDA cloud platform, the enhancement of the system to enable connection with a chosen Persistent Identifier (PID) provider, the development of an API for the CESSDA Controlled

Vocabularies, Topic Classification, European Language Social Science Thesaurus (ELSST), and the development of a multilingual user interface. Activities in 2019 are aiming to expand the metadata fields towards the DDI lifecycle, to develop different interfaces for national long-term preservation solutions and more.

Project partners are: ADP (Slovenia), AUSSDA (Austria), DANS (the Netherlands), GESIS (Germany), SND (Sweden) and TARKI (Hungary).

The project will be partly continued in the SSHOC (Social Sciences & Humanities Open Cloud) project.

Posters / Software Demonstrations

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

Planning Feasibility Studies on the Applicability of the DDI Standard for "One Health Surveillance" Data Harmonization in Europe

Taras Günther (Federal Institute for Risk Assessment), Tasja Buschhardt (Federal Institute for Risk Assessment), Fernanda Dorea (National Veterinary Institute), Matthias Filter (Federal Institute for Risk Assessment) (↑ schedule)
Track: Data Harmonization – **Session Type:** Poster/Software Demonstration - Community Impact

Zoonotic diseases and microbial antibiotic resistance pose an increased worldwide health risk due to globalization. One Health Surveillance (OHS) describes the systematic collection, validation, analysis, interpretation and dissemination of data collected on humans, animals, food and the environment to implement more effective, evidence- and system-based health interventions. The lack of harmonized surveillance systems and architectures to support data interoperability across different domains (Public- and Animal Health as well as Food Safety) and stakeholders (e.g. institutions, countries, governmental agencies, and European authorities) is currently hampering a true One Health approach in risk assessment and management. In 2018 the European Commission funded the research project ORION (One health surRveillance Initiative on harmOnization of data collection and interpretation) under the European Joint Programming One Health framework (https://onehealthejp.eu/). Within this project we will explore whether the Data Documentation Initiative (DDI) can be used to support harmonization of One Health Surveillance data in Europe. We will present the results of the requirement and process analysis for several pilot applications.

Colectica Portals: Public Use Data on the Web

Jeremy Iverson (Colectica), Dan Smith (Colectica) (↑ schedule)
Track: Software / Tools – **Session Type:** Poster/Software Demonstration - Community Impact

Colectica Portal is built to search and display data documentation for longitudinal and repeated data collections. Variables can be harmonized and linked across study waves, and changes in measurements, coding, and descriptions are encoded using open data documentation standards such as DDI. Questionnaires from multiple waves can be documented, including sections, routing and filter questions, and linked to the data which was collected.

Colectica Portal uses Colectica Repository to store and index the graph of information used to describe these complex studies

Several national statistical organizations and international longitudinal studies publish rich metadata on the Web using Colectica Portals. At this poster, Colectica staff will be available to demonstrate the portals and discuss strategies for creating such portals.

DDI Implementation Projects at SSJDA: Ongoing Projects and a Future Task

Satoshi Miwa (Institute of Social Science, The University of Tokyo.), Kazuhiro Kezuka (Tokyo Institute of Technology), Tomoyuki Masuko (Institute of Social Science, The University of Tokyo.) (↑ schedule)
Track: Software / Tools – **Session Type:** Poster/Software Demonstration - Project Report

Social Science Japan Data Archive (SSJDA) has developed DDI implementation projects. We present two ongoing projects and a future task in SSJDA. The first implementation project is the Nesstar system operation since 2012. Some datasets which are provided from SSJDA are available in the Nesstar system and the number of access and users are increasing. Since mutual link with SSJDA's website is also started, cooperation between services is working more effectively. The second project is development of the

software Easy DDI Organizer (EDO). EDO is an open source tool for editing and managing a metadata along with data lifecycle. Our future task is to provide users with tools to encourage users to understand and use social research. Panel surveys usually have complex structures that impede analysis. At the same time, I would like to integrate these multiple projects and improve the process of data archiving work. Discussion of these projects and future issues.

CLOSER's Metadata Portal: Improving Discovery

Hayley Mills (CLOSER Project, Department of Social Science, UCL Institute of Education) (↑ schedule)
Track: User Needs, Efficient Infrastructures and Improved Quality – **Session Type:** Poster/Software Demonstration - Project Report

CLOSER Discovery (https://discovery.closer.ac.uk/) is a portal for researchers to search and browse social science and biomedical metadata from thousands of questions and variables, across several of the UKs longitudinal studies. The poster outlines; the searchable metadata, controlled vocabularies, the infrastructure, as well as the latest features and future developments. A demonstration of CLOSER Discovery will accompany the poster.

An R Package for Putting DDI to Practical Use as a Personal Tool : DDIR

Yasuto Nakano (School of Sociology, Kwansei Gakuin University) (↑ schedule)
Track: Software / Tools – **Session Type:** Poster/Software Demonstration - Project Report

Information of social research activities (e.g. research data, research questions, variable conceptualizations, questionnaire sentences, variable names, value labels, etc.) are in Babel. If we could have a standard format of information, social research would be more efficient and more reproducible. DDI is a hope in this chaos.

'DDIR' is an R package which handles information in DDI format on R environment. 'DDIR' contains functions not only to import information from metadata files in DDI format, but also to export information to a metadata file in DDI format. Importing functions retrieve variable information (e.g. variable labels, value labels, questionnaire sentences) and integrate them with data into a data frame. Exporting functions tidy up necessary information (e.g. document description, study description, file description, data description) in DDI xml format.

Especially with this exporting function, we can make DDI metadata files for our own. It means that DDI could be a personal tool even for small research projects. Furthermore, there are a lot of functions which import/export social research data in various format (e.g. sav, csv, stat?), 'DDIR' could be a hub of miscellaneous data formats.

Tutorials

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

Technical Workshop: Implementing DDI4 - Data Description Model
Darren Stephen Bell (UK Data Archive) (↑ schedule)
Track: Other – **Session Type:** Tutorials or Workshop

The workshop will be an opportunity for developers and other implementers to discuss possible options available for implementing the prototype DDI4 Data Description model.

Work at the UK Data Archive indicates that the current DDI4 RDF binding is not performant at scale and could leverage existing vocabularies to a greater extent.

The workshop will be an opportunity to explore additional strategies, namely using graph morphisms other than RDF triples, going beyond SPARQL to consider additional graph query languages like Gremlin and simplifying some of the proprietary structures of the DDI4 model like RelationSpecification, by re-using existing vocabularies like XKOS instead.

Document Questionnaires and Datasets with DDI: a Hands-on Introduction with Colectica
Jeremy Iverson (Colectica) (↑ schedule)
Track: Software / Tools – **Session Type:** Tutorials or Workshop

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 Concordance with Represented Variables and Conceptual Variables

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

Same Same But Different - Documenting Panel Data Using DDI
Knut Wenzig (German Institute for Economic Research (DIW Berlin) / German Socio-Economic Panel (SOEP)) (↑ schedule)
Track: Reusing and Sharing Metadata – **Session Type:** Tutorials or Workshop

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.

Closed Side Meetings

CESSDA Expert Seminar 2018 Follow-Up: practical session on Jenkins pipelines, Docker containers, Kubernetes clusters

John Shepherdson (CESSDA ERIC) (↑ schedule)
Track: Other – **Session Type:** Side Meeting

For details see event page.

CESSDA, Trust, Metadata and Roadmaps

Herve L'Hours (UK Data Archive) (↑ schedule)
Track: Other – **Session Type:** Side Meeting

This session will present the current activities of CESSDA and seek feedback on our initial Trust roadmap. We will consider the full data lifecycle, the archival role, the need for managed and aligned metadata and the role for the DDI standard and associated tools in reaching CoreTrustSeal compliance.

The CESSDA Trust Group has expanded its activities beyond an initial introduction to the CoreTrustSeal and support for those making applications. There are now a range of CESSDA audiences ranging across aspiring members, those actively seeking certification, and successfully certified Repositories working on maintenance and renewal of certification. (i.e. planning to renew). The Social Science and Humanities Open Cloud (SSHOC) project will see further Trust cooperation across SSH ERICs in Europe and all archives face a range of changing expectations from interdisciplinary data, to FAIR compliance and the emergence of the European Open Science Cloud (EOSC).

In order to be CoreTrustSeal compliant, all archives must provide evidence showing how they operate with a data and metadata management system that is suitable for ensuring integrity and authenticity during the process of ingest, archival storage, and data access. Enough information must be provided to ensure the Designated Communities can assess the substantive quality of the data and effectively reuse data and metadata over time despite changes in technology or changes to the community. Archives have different approaches to assure this, from adopting international standards that provide rich disciplinary metadata, such as the DDI (i.e. CESSDA Metadata Model), to using other perhaps less rich metadata standards.

This session will present the current activities of CESSDA and seek feedback on our initial Trust roadmap. We will consider the full data lifecycle, the archival role, the need for managed and aligned metadata and the role for the DDI standard and associated tools in reaching CoreTrustSeal compliance.

DDI Technical Committee and DDI 4 Moving Forward Project – Combined Meeting

Kelly Chatain, Wendy Thomas, and Joachim Wackerow (↑ schedule)
Track: Other – **Session Type:** Side Meeting

DDI Roadmap

- Finalization of 3.3 based on review
- Prepare COGS to accept 3.3 schema content and generate output
  - XML schema (this would be a future version)
  - UML model of 3.3 to reflect Moving Forward approach
  - Documentation infrastructure
  - OWL/RDF output
- Evaluate differences between 3.3 and output from COGS
- Transition paths between versions across DDI

DDI 4 Moving Forward Project

- High-level and conceptual discussion on the direction of DDI 4 regarding content and modeling
- Review of post-prototype issues and prototype public release feedback
  - Role of UML
    - Use of UML features
    - Refinement of modeling rules
  - Binding transformation rules
  - Role of Canonical XMI as an expression of the model
- Plans for new production framework (COGS)
  - Production process entering content and COGS production line
  - Clarification of validation checks for entered content
- Creation of an iteration cycle regarding modeling, representations, and prototype software implementation to ensure a robust specification on all of these levels.
- Creation of validation suites on all of these levels and integration into the production framework
- Structured documentation

Authors

Organizations Index

Funding and Sponsoring Organizations

EDDI18 is funded by the German Research Foundation DFG (Deutsche Forschungsgemeinschaft).

DFG Deutsche
Forschungsgemeinschaft
German Research Foundation

EDDI18 is sponsored by the German Data Forum RatSWD (Rat für Sozial- und Wirtschaftsdaten).

RatSWD.
German Data Forum
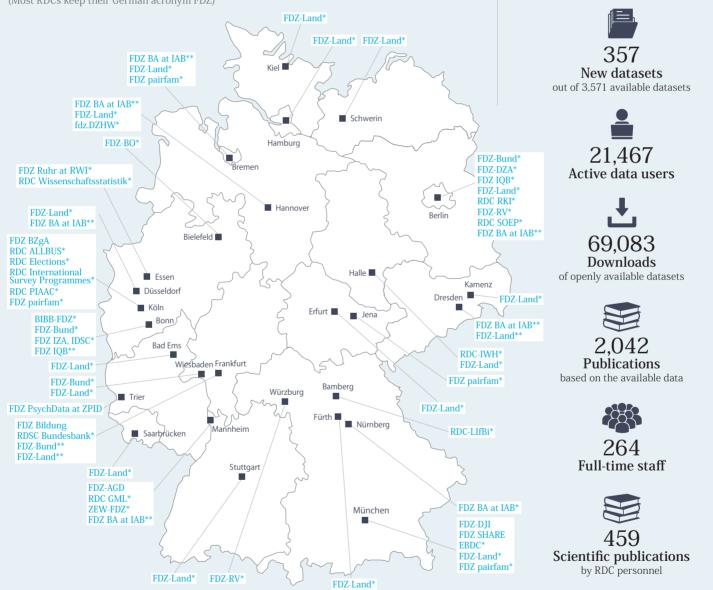
RatSWD.

German Data Forum

Research Data Infrastructure accredited by the RatSWD

Social, behavioural, and economic research often requires access to sensitive microdata on individuals, households, or businesses. To satisfy the demands of data producers, data users, and data protection, the German Data Forum (RatSWD) - as one of many tasks - accredits research data centres (RDCs). Consequently, a network of 31 German RDCs has developed since the early 2000s.

The RDCs provide researchers with user-oriented, cost-efficient access to a wide range of data. Since 2009, the RDCs have been organised in the FDI Committee to coordinate the research data infrastructure. This bottom-up coordination fosters efficient governance as well as securing the quality of the research data infrastructure through monitoring procedures.

Available data

RDCs provide data from the following areas of research:

- Social Data
- Behavioural Data
- Economic Data
- Qualitative Data
- Education Data
- Other Data
- Health Data

Research Data Centres in 2018

(Most RDCs keep their German acronym FDZ)

FDZ-Land*

FDZ-Land*   FDZ-Land*

Kiel

FDZ BA at IAB**
FDZ-Land*
FDZ pairfam*

Schwerin

FDZ BA at IAB**
FDZ-Land*
fdz.DZHW*

Hamburg

FDZ-BO*

Bremen

FDZ-Bund*
FDZ-DZA*
FDZ IQB*
FDZ-Land*
RDC RKI*
FDZ-RV*
RDC SOEP*
FDZ BA at IAB**

FDZ Ruhr at RWI*
RDC Wissenschaftsstatistik*

Hannover

Berlin

FDZ-Land*
FDZ BA at IAB**

Bielefeld

FDZ BZgA
RDC ALLBUS*
RDC Elections*
RDC International
Survey Programmes*
RDC PIAAC*
FDZ pairfam*

Essen

Düsseldorf

Halle

Kamenz

Dresden   FDZ-Land*

BIBB-FDZ*
FDZ-Bund*
FDZ IZA, IDSC*
FDZ IQB**

Köln

Bonn

Erfurt   Jena

FDZ BA at IAB**
FDZ-Land**

Bad Ems

RDC-IWH*
FDZ-Land*

FDZ-Land*

Wiesbaden Frankfurt

FDZ pairfam*

FDZ-Bund*
FDZ-Land*

Trier

Würzburg

Bamberg

FDZ PsychData at ZPID

Fürth   Nürnberg

FDZ-Land*

FDZ Bildung
RDSC Bundesbank*
FDZ-Bund**
FDZ-Land**

Saarbrücken

Mannheim

RDC-LIfBi*

FDZ-Land*

Stuttgart

FDZ-AGD
RDC GML*
ZEW-FDZ*
FDZ BA at IAB**

München

FDZ BA at IAB*

FDZ-DJI
FDZ SHARE
EBDC*
FDZ-Land*
FDZ pairfam*

FDZ-Land*   FDZ-RV*

FDZ-Land*

Key figures of the accredited RDCs in 2017

357
New datasets
out of 3,571 available datasets

21,467
Active data users

69,083
Downloads
of openly available datasets

2,042
Publications
based on the available data

264
Full-time staff

459
Scientific publications
by RDC personnel

*RDCs with guest researcher workstations / **External RDC guest researcher workstation   Source: RatSWD Activities Report 2017   © RatSWD 2018

www.ratswd.de/en

SPONSORED BY THE

Federal Ministry
of Education
and Research

About EDDI and EDDI18

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.

EDDI18 Committees

Organization Committee
- Nikos Askitas, IDSC of IZA - International Data Service Center of the Institute for the Study of Labor, Germany
- Andreas Franken, German Institute for Economic Research (DIW Berlin) / German Socio-Economic Panel (SOEP), Germany
- Uwe Jensen, GESIS - Leibniz Institute for the Social Sciences, Germany
- Jon Johnson, UK Data Service, United Kingdom
- Christine Kurka, German Institute for Economic Research (DIW Berlin) / German Socio-Economic Panel (SOEP), Germany
- 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

Program Committee
- Alina Danciu, Center of Socio-Political Data, Sciences Po Paris, France
- Jon Johnson, UK Data Service, United Kingdom
- Mari Kleemola, FSD - Finnish Social Science Data Archive, Finland
- Mikko Saloila, Statistics Finland, Finland
- 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, UK Data Service, United Kingdom
- Joachim Wackerow, GESIS - Leibniz Institute for the Social Sciences, Germany

Map