30 November - 2 December, 2020
Online (France)

12th
ANNUAL EUROPEAN DDI USER CONFERENCE
(EDDI20)

DDI®

SciencesPo
CENTRE DE DONNÉES SOCIO-POLITIQUES

gesis
Leibniz-Institut
für Sozialwissenschaften

IDSC
INTERNATIONAL
DATA SERVICE CENTER

I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

EUROPEAN DDI USER CONFERENCE

12th Annual European DDI User Conference (EDDI20)

Schedule and Program with Abstracts

Final version of November 26, 2020

The **conference** will start on Monday, November 30 at 10:00 and will end on Wednesday, December 2 at 16:15 (**all times are CET**).

Tutorial and workshops will take place on Monday, November 30 from 10:00 to 17:30 (**all times are CET**).

More information:

https://eddi20.sciencesconf.org

Organizers address

Welcome to EDDI 2020

We would like to extend a warm welcome to the 12th Annual meeting of the European DDI Users conference.

This time last year, the DDI community was looking forward to spending the first few days of December 2020, discussing with colleagues and enjoying the best that Paris can offer. But alas, it was not to be!

The Program Committee had to make (what seemed at the time) a difficult decision to postpone the face-to-face event. But, thanks to the generosity of our hosts at Science Po, who have provided access to their on-line platform, we have managed to provide a space for the community to come together to report on progress over the last 12 months.

The more so, as like many in the DDI community, it has been an exceptionally busy year, many of us have been at the heart of the response to the pandemic. Designing studies, collecting data and preparing it for research and analysis and releasing it for use.

We will no doubt examine the response to the pandemic over the next few years, but the tsunami of data generated during the pandemic has shown the need for well documented data is more important than ever if, data within a country from different sources is to be reliably combined, let alone comparing across countries.

If recent news reports are to be believed, vaccines are coming on stream, and over time the pandemic may start to come under control. We very much hope so, and it will enable us all to return to some form of normality.

Meanwhile, we have one tutorial, 2 workshops, and 27 presentations to enjoy in this year's EDDI. The conference covers a variety of topics from DDI standards and data management to DDI software. We would like to thank the Program Committee, presenters and session chairs for their willingness to share their work, time and knowledge. We are sure the presentations will stimulate meaningful and innovative discussion both during and after the conference.

We are hopeful that we will be able to meet in person next year, in Paris, where Science Po will be able to give us the real world welcome that we had all wished for.

Mari Kleemola & Jon Johnson
Co-Chairs EDDI 2020

Table of contents

Schedule (all times are CET)

Monday, November 30, 2020

10:00 - 11:00 : Tutorial

10:00 - 11:00 : What can DDI do for you? An introduction to the DDI

Benjamin Beuster (Norwegian Centre for Research Data (NSD))
Jane Fry (Carleton University)
Topics : Tutorial **- Session Type :** Tutorial or Workshop

Are you interested to learn about what DDI can do for your organization or institution? DDI is an international standard for describing data from the social, economic and behavioral sciences, currently moving into new fields. The standard contains metadata items that can be used to develop and document at different stages in the data lifecycle, from the first conceptualization through data collection, processing and dissemination and archiving. This tutorial provides an overview of the work products of the DDI Alliance. The conceptual basis of DDI will be described, introducing the participants to the main building blocks and items of the standard. Practical examples on how DDI can be used beneficially in the business processes of organizations and institutions that manage research data will also be shown. The overall approach of the tutorial is DDI-version agnostic. The examples shown will however be based on specific DDI versions (DDI-Codebook, DDI-Lifecycle).
Main focus will be put on the following areas:
- Data description and variable management
- Questionnaire design and implementation
- Question and variable banking
- Making your data and metadata FAIR (Findable, Accessible, Interoperable and Reusable) using DDI

15:00 - 17:30 : Workshops - Parallel Session

15:00 - 17:30 : Workshop 1 - Generating DDI – Cross-Domain Integration (DDI-CDI) in the Language of your Choice. Model Driven Architecture for Metadata Standards in Practice

Parallel session
Joachim Wackerow (GESIS - Leibniz-Institute for the Social Sciences)
Topics : Workshop 1 **- Session Type :** Tutorial or Workshop

This tutorial provides the knowledge and skills on how a syntax representation (or encoding) of an UML model can be generated in any language. The focus is on the DDI – Cross Domain Integration (DDI-CDI) model, which is in many important respects independent from the social science domain and therefore suitable for cross-domain purposes (the publication is planned for 2021).
The first part of the tutorial gives an overview of the model-driven approach, the model itself, used UML modeling techniques, model-to-model transformation, and the generation of syntax

representations in XML Schema (for XML instances) and in OWL (for RDF instances). It describes how the UML model can be used as basis to generate a syntax representation in a chosen language. The second part provides time for exercises. Eclipse modeling tools are used for this purpose. Participants are encouraged to choose a language in which they wish to use DDI-CDI. You should bring your own laptop and be ready to install the required software (a detailed list will be provided) prior to the tutorial.

Agenda :

Part I :

- Overview
- Model-driven approach
- DDI-CDI model – UML subset and model structure
- Model-to-model transformation
- Generation of XML Schema for XML instances and OWL Vocabulary for RDF instances

Approach for the generation of any syntax representations

Part II :

- Eclipse environment
- Exercises

15:00 - 17:30 : Workshop 2 - DDI Cross Domain Integration (DDI-CDI). An introduction to the draft specification using real life examples

Parallel session
Hilde Orten (Norwegian Centre for Research Data (NSD))
Arofan Gregory (Consultant)
Topics : Workshop 2 **- Session Type :** Tutorial or Workshop

The DDI Cross Domain Integration (DDI-CDI) is a new product within the DDI family of specifications to be released in 2021. Unlike earlier DDI metadata specifications, DDI-CDI is intended to be used in combination with DDI-Codebook and DDI-Lifecycle, as well as potentially stand-alone.

This tutorial shows you how DDI-CDI can be used to document a broader range of data structures than earlier specifications: traditional rectangular files, long/event data files, multi-dimensional data cubes, and key-value ("big") data. It also covers the description of transformations between these types and describing the details of data integration. Examples of provenance and process documentation will also be covered.

This tutorial introduces you to the main aspects of DDI-CDI using real life examples, covering topics related to data description, variable structures and harmonization, and process and provenance description.

Tuesday, December 1, 2020

10:00 - 11:15 : Questionnaires

Chair : Wolfgang Zenk-Möltgen

10:00 - 10:20 : Strategies for (re)creating Questionnaires in DDI-Lifecycle

Hayley Mills (CLOSER, UCL Institute of Education)
Topics : Questionnaires **- Session Type :** Regular presentation

Implementing questionnaires in DDI-Lifecycle data collection provides a rich source of metadata that can enhance the understanding of data for researchers, discovery and data and survey management, more so where data is widely reused.

Ideally one would capture the metadata for questionnaires as they are produced, however, for logistical, organisational and resource reasons, this is currently not widespread.

If we are to be able to provide this layer of rich metadata, finding ways to quickly and efficiently capture the questionnaires and link that to both the data and previously asked questions is crucial.

The presentation will discuss approaches that CLOSER has employed to lower the barriers for studies to provide high quality questionnaire metadata in DDI-Lifecycle for CLOSER Discovery, and for internal data management purposes at the studies themselves.

10:20 - 10:40 : Using Colectica to document socio-political surveys in DDI-Lifecycle: feedback and benefits

Lucie Marie (Sciences Po, Center for Socio-Political Data (CDSP), CNRS, Paris)
Topics : Questionnaires **- Session Type :** Regular presentation

With the Open Science movement, the patterns of data sharing are evolving. In this context, the French Center for socio-political data has launched in 2020 an experimental project that aims to build a question bank using DDI-Lifecycle and Colectica - Designer, Repository and Portal. Beyond the final goal to be compliant with the CESSDA repositories standards, so that data and metadata may be harvested in the European Question Bank, this new tool is implemented for two additional purposes. On the one hand, to make data more discoverable and findable at variable level. On the other hand, to implement an internal less time-consuming documentation protocol by defining a metadata model at the variable level – a reference document for data managers to check for paradigms of organized data.

This presentation outlines the main steps of the process of implementation of DDI-Lifecycle using the Colectica softwares package, including challenges met to build the question bank. More broadly, it gives feedback about the entry cost of the shift from DDI-Codebook to DDI-Lifecycle, as well as the process of enhancing metadata documentation quality, especially harmonization of items and work on granularity at the variable level.

10:40 - 11:00 : Specification and generation of questionnaires – a metadata-driven approach powered by DDI

Fabienne Perray-Gibert (Institut national de la statistique et des études économiques (INSEE))
Topics : Questionnaires **- Session Type :** Regular presentation

Insee has been developing during the last years a metadata-driven survey data collection process including a questionnaire generator code-named Eno. This tool operates on a formal description of the questionnaire in DDI and executes a completely automated chain of transformations to produce the actual collection instrument(s) (web and paper questionnaires). The questionnaires can be

personalized and, depending on the output format, a variable amount of flow logic can be implemented.

Questionnaires from 30 different business surveys have been produced by Eno, and this number will continue to increase, especially because Eno will soon support questionnaires for household surveys. Interfaced with Eno is a questionnaire design user interface, code-named Pogues, that connects with the generation process. More specifically, Pogues produces the DDI description of the questionnaire which is then submitted to an embedded instance of Eno. Using this tool, a survey manager or questionnaire designer can specify his web, paper and soon interviewer questionnaire in a friendly way and visualize the generated result in one click.

Pogues and Eno are open source, internationalized, and thus can be used directly by other statistical agencies. They are currently using DDI version 3.3. Some have already shown their interest.

10:00 - 11:30 : Issues in Access and Discovery

Chair : Maja Dolinar

10:00 - 10:20 : Access control for scientific use files under new regulations

Lorraine Adam (Adisp)
Constance Hemmer (Institut national d'études démographiques (INED))
Topics : Efficient Infrastructures **- Session Type :** Short presentation

Scientific use files (intermediary between public use files and secure use files), called production and research files (FPR) in France, have been available to the scientific community for several years via Quetelet-Progedo diffusion. Due to the evolution of regulations with the law "pour une république numérique", which promote the open data and continuing to respect the General Data Protection Regulation, we had to find a specific legal status to the product and research files.

In Quetelet Progedo Diffusion, the national archive of data from official statistics (ADISP) and the French Institute for Demographic Studies (INED) have these types of files in their repositories. Thus, those two data providers have established a collaboration with the Statistical Secret Committee, the organization who guarantees the application of the rules of statistical secret in France, in order to set up an harmonized procedure for the production and research files.

This presentation will focus on this specific access control we utilize in Quetelet Progedo Diffusion.

10:20 - 10:40 : Documentation of confidential data made available by the CASD

Raphaëlle Fleureux (CASD)
Halima Bakia (CASD)
Topics : Efficient Infrastructures **- Session Type :** Regular presentation

The online display of the data made available by the Secure Access Data Center (CASD) meets a need of data users. How can one request data from the various accreditation authorities without knowing whether their content is relevant to a project's issue? The confidential nature of these data, covered by the secrecy, requires that accredited users access them only in a secure environment.

This presentation will describe the documentation procedure set up by the CASD, based on the DDI3 standard and using the Colectica Designer software. Since the data made available by the CASD is confidential, the documentation process must necessarily be carried in the secure environment. The result of this documentation is then displayed on the CASD website in an open access. It describes all the databases made available, their variables and work is in progress to describe the codes. As the

number of data made available is significant, only the most recent available year of each series is currently documented.

10:40 - 11:00 : ExploreData to Search, Access, and Recommend Social Sciences Surveys Metadata

Karam Abdulahhad (GESIS - Leibniz-Institute for the Social Sciences)
Topics : Efficient Infrastructures **- Session Type :** Regular presentation

DDI (3.2 in our case) offers a solid framework to document social sciences surveys metadata. However, DDI isn't well adapted to make metadata searchable. GESIS isn't an exception, where GESIS's data archive has rich and deeply prepared and documented data collections. In the current settings, GESIS's researchers use a bunch of tools to access data, such as DBKSearch or ZACAT. Many of these tools are either outdated or not sufficiently user-friendly. ExploreData aims to make these data collections easily accessible, findable, and searchable. It also seeks to replace a bunch of tools by one easy-to-use tool.

ExploreData mainly deals with two entity types, namely study and variable. For each entity type, we maintain a bunch of fields in order to keep as much structure of the study/variable original DDI 3.2 as possible. ExploreData also supports multilingual content. During data processing, we deal with data migration, missing values, removing duplicated and redundant data, multilingual content, and sometimes generating new data. To end-users, the content can be searched via classical keyword queries, and also browsed via many pre-defined dynamic facets.

In addition to search and browse, ExploreData presents a content-based variables recommendation functionality, where the user is able to get a list of "similar" variables of a given variable.

Our repository contains more than 6100 studies, distributed over 193 collections such as ALLBUS and EVS. It also contains more than 215K variables. In our presentation we introduce ExploreData and describe data processing challenges.

15:00 - 16:30 : DDI-CDI / Open Data

Chair : Ingo Barkow

15:00 - 15:20 : DDI and Cross-Domain Data Sharing: Collaboration with CODATA's Decadal Programme

Arofan Gregory (Consultant)
Topics : DDI-CDI / Open Data **- Session Type :** Regular presentation

The development of DDI Cross Domain Integration (DDI-CDI) has allowed for the application of DDI standards to data not only within the traditional social, behavioural, and economic domains, but also to data from many other areas of research. Increasingly, data is being re-used across traditional domain boundaries, raising a number of complicated challenges for data documentation. CODATA - the data-focused arm of the International Science Council - has identified DDI-CDI as an important tool in establishing guidelines and a community of practice to facilitate cross-domain data sharing, and will highlight it in the soon-to-be-launched Decadal Programme: Making Data Work for Cross-Domain Grand Challenges (https://codata.org/initiatives/strategic-programme/decadal-programme/).

The use of DDI-CDI will be highlighted in the initial case studies around infectious disease, resilient cities, and the use of global policy monitoring data in research. This talk outlines the growing collaboration between this international initiative and the DDI Alliance around DDI-CDI.

15:20 - 15:40 : DDI-CDI Overview, Status, and Potential Implementation

Arofan Gregory (Consultant)
Hilde Orten (Norwegian Centre for Research Data (NSD))
Topics : DDI-CDI / Open Data **- Session Type :** Regular presentation

DDI Cross-Domain Integration (DDI-CDI) represents a new focus for DDI metadata specifications: it applies the DDI Moving Forward conceptual model to the problems of describing new forms of data and their provenance, and the integration of disparate types of data with those we have traditionally described with DDI-Codebook and DDI-Lifecycle. As a companion standard to these tools, DDI-CDI leverages the model-driven aspects of DDI Moving Forward to align with standards of many different types, and to provide a bridge between diverse data sources and across domain boundaries.
This presentation describes the overall feature set of DDI-CDI, covering description of rectangular, wide/event, multi-dimensional, and key-value ("big") data structures and the datum-oriented mechanism for describing integration across data sets. It also presents the two approaches for modeling process/provenance, and for aligning these with other standards and syntaxes.Typical expected implementations are summarized, and the current state of development will be described as the specification moves from the current public review phase to a production release.

15:40 - 16:00 : The ODISSEI Portal: Linking Survey and Administrative Data

Tom Emery (Erasmus University Rotterdam)
Topics : DDI-CDI / Open Data **- Session Type :** Full paper

The ODISSEI Portal will combine metadata from a wide variety of research data repositories into a single interface, allow advanced semantic queries to support findability, and facilitate data access. The project will include metadata of (a) all datasets of Statistics Netherlands, including the metadata of the microdata catalogue, (b) all datasets developed supported by ODISSEI in the Netherlands (EVS, GGP, SHARE, ESS, NTR, HSN, LISS) which predominantly apply DDI standards. The project will also develop a metadata ingestion pipeline to make sure that the Portal can be maintained and kept up to date. This pipeline will be used by the hosting party to add new datasets during and after the end of the project. Currently, existing tools for findability in the social sciences are limited in that they only identify specific terms or synonyms (e.g. United Kingdom question bank or the Question Variable Database). The ODISSEI Portal will extend and improve search functionality by using semantic queries which will enable broader probabilistic matching and link functions over an enriched knowledge graph representation of the FAIR metadata catalogue. This incorporates the context of specific terms which are crucial in social research. By using rich and extensible data structures developed within the linked data community, ODISSEI will evolve the relatively flat metadata catalogues in use today into the highly interlinked and graph-based structures needed to conduct advanced semantic searches.

16:00 - 16:20 : Using linked metadata and quality standards for the documentation of Insee's statistical operations

Franck Cotton (INSEE)
Topics : DDI-CDI / Open Data **- Session Type :** Regular presentation

Insee publishes on its web site documentation about the statistical operations that it conducts. Behind the scenes, an important effort has been made recently to break down these texts into precise, well-defined and harmonized items. For this, we leveraged a standard used in the European Statistical System for quality reporting: the SIMS model (Single Integrated Metadata Structure). We actually structured and refined the SIMS by translating it to OWL/RDF, adding a few attributes and specifying more rigorous data types (for example, a contact organization would be an organization object instead of just descriptive text).

Splitting operation descriptions into the SIMS layout required some initial work from the subject-matter statisticians, but it will greatly simplify future regulatory quality reporting. Also, publishing documentations as RDF will allow to integrate them into a global metadata information system and link them to concepts, codes, organization schemes, the product catalogue, etc. RDF maximizes usability and machine-actionability, so it is possible to offer a variety of dissemination formats: a dedicated API, exports as office documents or SDMX metadata reports ready to send to Eurostat, etc. This will greatly improve and accessibility of our metadata, discoverability of our products and comparability at the national and European levels.

15:00 - 16:15 : Data Harmonization

Chair : Mari Kleemola

15:00 - 15:20 : Cross study variable concordance with DDI Lifecycle

Jeremy Iverson (Colectica)
Topics : Data Harmonization **- Session Type :** Regular presentation

Since DDI Lifecycle 3.2, the variable cascade has allowed data producers to define the concordance among variables in datasets. Since at least 2014, several longitudinal studies and national statistics organizations have used this metadata structure to document and publish information on variables within single studies or organizations. Separating data definitions into conceptual variables, represented variables, and instance variables allows precise documentation of the data and how it changes over time.

The same basic metadata structure can be used to document concordance across studies. Since separate organizations often create conceptual variables with similar meanings as conceptual variables from other organizations, the challenge becomes declaring the comparability of conceptual variables. For the intellectual content work of deciding on comparability, this presents an easier task than individual concording a large number of instance variables. For the technical task of defining the concordance in a structured manner, DDI Lifecycle offers a standardized solution.

This presentation will provide an analysis of the DDI Lifecycle metadata structure used to perform the cross study concordance; discuss the workflow used to harmonize hundreds of variables from several large, longitudinal studies; and demonstrate the software tools used to create, publish, and visualize the data concordance.

15:20 - 15:40 : DDI, Dataverse and Colectica: our data management combo

Alina Danciu (Sciences Po)
Topics : Data Harmonization **- Session Type :** Regular presentation

Providing to our community a user-friendly interface and data visualisation tool, a powerful search engine and a fluid data download process have always been top priorities of the French Center for Socio-Political Data. In the past year, we worked intensively to implement our Dataverse repository, which answers at least partially to these questions.

During the process of importing our Nesstar metadata into Dataverse, we encountered a certain number of issues we had not anticipated, primarily linked to the compatibility between DDI-C and Dataverse. This presentation will mainly tackle this topic, as well as the process of implementing the DDI controlled vocabularies and the harmonisation process of our metadata (institution names, authors, keywords...).

In terms of variable management and visualisation, we decided to experiment with the Colectica tools (Designer, Repository and Portal), that we recently implemented at the CDSP and that we intend to use in order to build a DDI-L question bank. The project is in its experimental phase. We'll offer feedback on our data management choices at this stage, complementary with our Dataverse tool.

15:40 - 16:00 : Question driven harmonisation of data - the variable cascade in practice

Jon Johnson (CLOSER, UCL Institute of Education)
Topics : Data Harmonization **- Session Type :** Regular presentation

Harmonisation of data is largely a manual process. Although there are naming conventions that are embedded in many studies (especially panel studies), that can provide strong clues to the equivalence of variables, it is by no means perfect and requires a priori knowledge.
The capture of questions (in DDI-Lifecycle instruments) linked to the output variables, allows generic methods to be developed to reliably identify and organise variables which may be candidates for harmonisation.
The presentation will discuss the approach CLOSER has taken to develop user facing tools that allows researchers to determine the criteria they wish to apply to harmonising variables, these include, the question, response domain, mode, and study characteristics.

Wednesday, December 2, 2020

10:00 - 11:30 : Developing DDI

Chair : Iris Alfredsson

10:00 - 10:20 : Data management, statistics and reporting driven by DDI in the healthcare marketing area

Alexandre Mairot (Exastat)
Topics : Developing DDI **- Session Type :** Regular presentation

Exastat is a company specialized in the data management and the production of statistics for its partners. Currently, Exastat works essentially with its sister companies focused either on the healthcare marketing, Stethos, Stethos International, Stethos Social Lab, or on the data producting, Exafield.

The main challenges for Exastat are to propose and develop the partnerships with the protection of the confidential business information and the transparency for the work process. In May 2020, Exastat upgraded its workflow to increase the services quality. In order to consolidate the information system by the normalization of the data management process and the improvement of the data treatment reliability, its architecture is based now on a metadata model DDI compliant (standards and controlled vocabularies). At present, all statistics production or reporting are driven by the metadata structured in the DDI standard. Besides, Lifecycle is used for keeping a trace of the Exastat activities and Codebook for disseminating to the partners.

This submission will present the programs, interfaces and procedures that are been built in order to accomplish these purposes in the context of a competitive and commercial area.

10:20 - 10:40 : Euro Question Bank: Microservices and UI for fielded survey question search

Thomas Krämer (GESIS - Leibniz-Institute for the Social Sciences, Mannheim)
Topics : Developing DDI **- Session Type :** Regular presentation

With the Euro Question Bank EQB, researchers can find and reuse fielded survey questions.
The source metadata
In this session we present the workflow behind the Euro Question Bank UI including
- DDI Metadata Harvesting
- Metadata indexing
- Features and functionalities of the Euro Question Bank UI
- Faceted search
- Multilingual search
- Question comparison
- Reuse of questions for your own research

We will explain
- the organisational and technical preconditions for institutions that participate in the Euro Question Bank EQB as content providers
- functionalities of the Euro Question Bank EQB with a special focus on multilingual search
- challenges of sparse metadata
- possibilities to join the EQB
- future directions on how to leverage the potential of rich DDI metadata on variable and study level

Target audience: Social sciences institutions, documentaries, CIOs

10:40 - 11:00 : The DDI Alliance Training Committee – What's it all about

Jane Fry (Carleton University)
Topics : Developing DDI **- Session Type :** Short presentation

The DDI Alliance Training Committee has been thoroughly re-organized recently. It started with appointing two co-chairs who were given the initial task of updating the mandate. Subsequently, a call went out for new members, with the result that the overall membership of the committee was expanded to 15 participants.

Another new development in the Training Committee was the formation of four new working groups. They are focusing on: producing the Training Library; updating relevant portions of the DDI Alliance website; outreach to other organizations and audiences; and a gap analysis group to identify new areas of focus. These Working Groups work closely together as they have activities which are overlapping and complementary with each other.

This presentation will detail several exciting new activities that are taking place in the Working Groups. Chief among these is the publication of a DDI Training Library, based on the outputs from a workshop held at Schloss Dagstuhl in 2018. Further efforts of the groups include a focus on outreach to new audiences, and the development of new training materials to address topics including classification management and DDI-CDI. Collaborations with other organizations such as GO FAIR, RDA, and CODATA are also being explored.

11:00 - 11:20 : Usability of DDI for Addressing the Issues of Ethnographic Data

Dolinar Maja (University of Ljubljana, Faculty of Social Sciences, Slovenian Social Science Data Archives)
Topics : Developing DDI **- Session Type :** Full paper

The Statement on Data Governance in Ethnographic Projects by the European Association of Social Anthropologists exposes core methodological and ethical practices that have concrete implications for the norms and forms of data management in ethnology and anthropology. These include (1) issues of data ownership, as ethnographic materials are co-produced by researchers and research participants and embedded in particular social contexts, (2) challenges of archiving, as ethnographic research data are always part of social relations and thus not always easily archived, (3) issues of consents, since dynamics of ethnographic participation do not always enable obtaining prior informed consent from participants, (4) challenge of custodianship, as researchers bear responsibility for ethnographic materials that are usually negotiated with research participants, (5) possibilities of embargo in cases where materials cannot be anonymized and (6) issues of public access and sharing. The Empirical Humanities Working Group of the Research Data Alliance deals with finding the right metadata standard and use cases from a broad range of metadata standards to provide researchers from empirical humanities (history, folklore, cultural anthropology and other fields) the appropriate guidance in research data management. The paper aims to show that DDI Codebook 2.5 can be an appropriate metadata standard that addresses the various kinds of ethnographic data (such as recorded interviews, field notes, video and audio materials, and photographs, among others) and can resolve the challenges of ethnographic data as stated in the above-mentioned Statement.

10:00 - 11:30 : Reuse and Quality

Chair : Tuomas Alaterä

10:00 - 10:20 : Documenting is not enough : an example of collaborative work in a French Research Infrastructure

Elodie Petorin, (Archives de données issues de la statistique publique - ADISP)
Ada Chmilevschi (PUD-L)
Topics : Reuse and Quality **- Session Type :** Regular presentation

The question of data access has become an increasingly important topic in scientific research: the FAIR principle, the Research Data Alliance, the Open Science Cloud, etc... As a matter of fact, the challenge of data and metadata findable, accessible, interoperable and reusable (FAIR) is becoming more and more important in the scientific community these days.
But yet, how can we produce FAIR data? Also, how to disseminate them? The Large Research Infrastructure Progedo has set up a network of university data platform, each of these skills platforms aims to provide real supports which form the necessary link between digital platforms and users. Furthermore, the interest of working upstream the data production, in collaboration with the researchers, allows us to obtain detailed documentation consisting in document of context as well as metadata. Moreover, downstream the chain of data, the national archive of data from official statistics (ADISP), a service of Progedo, documents those metadata to the DDI standard and disseminate it to the European scientific community.
To conclude, the purpose of this presentation will be to highlight the betterment of this collaborative work between Adisp and the university data platform in the data life cycle and the interest of it.

10:20 - 10:40 : Making models into reality: the DDI Profile journey at CESSDA

Darren Bell (UKDS)
Kerrin Borschewski (GESIS)
Topics : Reuse and Quality **- Session Type :** Regular presentation

Several years ago, CESSDA had a dream. We would synthesize metadata from European partners and harmonize it into beautiful repositories, whether data catalogues or question banks, by making DDI less terrifying for normal human beings. Thus, the CESSDA Metadata Model (CMM) was born. And then we woke up.
This presentation introduces the CESSDA Metadata Office, the CMM, and how we developed the original vision to support the CDC (CESSDA Data Catalogue) and EQB (European Question Bank). We describe the challenges along the way, from capturing first requirements in spreadsheets to refining the model into an XML schema, then a UML model, and finally, the first real-world implementation of DDI Profiles, which will change the way we measure metadata quality. This has been not only a technical journey but a pan-European collaboration between end-users and metadata experts who communicate in English but are often still speaking different languages. We also offer suggestions (learnt the hard way) about how DDI Profile mechanisms could be improved in future versions. Lastly, we outline future developments in 2021 and how the CESSDA Metadata Validator can help repositories to quality assure their published metadata.

10:40 - 11:00 : Paradata - from by-product to standard documentation

David Schiller (University of applied science of the grisons)
Topics : Reuse and Quality **- Session Type :** Regular presentation

By looking at the current examples and use cases of paradata this paper develops a comprehensive description of what paradata is and how it is used in the preparation, capture, processing, and analysis of data. It explores a general model of paradata in terms of what should be captured and how it can be organized to better support the structured capture of paradata and its metadata to better understand the processes surrounding the creation of high-quality data for analysis. The paper proposes the development of a structured metadata model that can be used in relationship to other widely used standards such as DDI.

11:00 - 11:20 : Reusing and sharing metadata: A case study of quantitative survey research on the integration of ethnic and migrant minorities (EMMs) across Europe

Ami Saji (Sciences Po)
Laura Morales (Sciences Po)
Topics : Reuse and Quality **- Session Type :** Regular presentation

This presentation will showcase how the EMM Survey Registry - a new online searchable database developed jointly by COST Action 16111 Ethmigsurveydata and the H2020 project SSHOC - promotes the reuse and sharing of metadata for quantitative surveys conducted across Europe focusing on the integration of EMMs. The presentation will begin by describing the metadata schema that was developed for the EMM Survey Registry. It will then illustrate how the metadata schema was not only conceptualised and designed to provide rich and meaningful information to users of existing quantitative survey data on EMMs' integration, but also to conform with DDI Codebook (a task undertaken by Alina Danciu of Sciences Po, CDSP and Alexandre Mairot, formerly with Sciences Po, CDSP) to allow the EMM Survey Registry to be interoperable with relevant data archives. The presentation will conclude by discussing strategies for fostering buy-in from the user communities to use and contribute to the EMM Survey Registry, which is essential to making the EMM Survey Registry a sustainable and state-of-the-art tool for reusing and sharing metadata of EMM surveys in Europe and beyond.

15:00 - 16:15 : Managing DDI

Chair : Alina Danciu

15:00 - 15:20 : Customizing a DDI compliant repository: a detailed course with Dataverse

Geneviève Michaud (Sciences Po)
Baptiste Rouxel (Sciences Po)
Topics : Managing DDI **- Session Type :** Regular presentation

Among academic institutions data repositories are a trending addition to the ecosystem of open science, and more often than not, there is a push towards self hosting. Dataverse is one of the most popular software in such cases, as it offers out of the box much desirable features, focusing on explorability, interoperability and a promise for flexibility. From a software development point of view, the open model driven by community feedback is a great asset.

Since 2018 at the Center for Socio-Political Data, we opted for migrating our DDI metadata bank from Nesstar to Dataverse. We were eager to customise Dataverse metadata schema to follow as closely as possible the Cessda Metadata Model main elements defined by our metadata experts (controlled vocabularies). We had to compose with a quite relaxed implementation of DDI 2.5 standard by Dataverse.

We will share strategies, tools and workarounds we set up to approach our goal. We have also strived to make the most out of the tool, using templating mechanisms or enhancing the user interface. We will conclude with a selective list of features we want to promote in order to better align Dataverse with standards as DDI or OAI-PMH.

15:20 - 15:40 : Managing Social Science Metadata at the Data Archive

Wolfgang Zenk-Möltgen (GESIS – Leibniz Institute for the Social Sciences)
Topics : Managing DDI **- Session Type :** Regular presentation

Increasing requirements for standardized metadata documentation possibilities arise from several challenges that a data archive faces nowadays: There are higher amounts of data to be archived. Data stem from a wider variety of data collection methods. The context of data collection and data re-use is more and more inter-disciplinary. Finally, more institutions exchange data and metadata. These challenges require a well-defined set of metadata standards to be used, extending the perspective beyond the use of DDI as the most prominent Social Science metadata standard. Notably, management of controlled vocabularies, supplied by DDI, ISO, CESSDA or others is needed. Practical considerations also include the use of DDI-Codebook or DDI-Lifecylce, specific versions of these and conversion back and forth between them.

GESIS prepares several large-scale national and international study collections but also archives smaller studies from single-time research projects or dissertations. GESIS creates and maintains documentation on study and variable level according to DDI and other standards, and offers data and documentation for re-use. For this purpose the metadata is transferred to several portals, both at GESIS and outside, e.g. to da|ra, DataCite, CESSDA Data Catalogue, ExploreData, or special purpose portals like x-econ or VerbundFDB. To meet the demands of these portals, and at the same time fulfil the needs of complex survey series as well as of more basic research datasets, it is required to manage the use of documentation standards, workflow processes, and software tools at the data archive.

15:40 - 16:00 : The Advantages and Disadvantages of Three Software Solutions for the Development of a DDI-based Survey Catalogue

Arianna Caporali (Institut national d'études démographiques(INED))
Topics : Managing DDI **- Session Type :** Short presentation

To facilitate data discovery, data accessibility and data reusability, it is key to develop data catalogues that use DDI as a framework for harmonized metadata and for complete data documentation. It is also important to implement software capable of enhancing DDI potentials to locate and explore the data. What characteristics should these software solutions meet? What criteria should be used to evaluate different tools?

This paper presents a comparison between three different tools, i.e., NADA Microdata Cataloging Tool, Dataverse and Colectica Repository, conducted at the DataLab of the French Institute for Demographic Studies (INED). We need to replace Nesstar, the software we used so far, that is being discontinued. The new tool should allow to promote our socio-demographic surveys among the largest scientific community, targeting both expert and non-expert users. This tool should require limited IT support, be compliant with DDI-Codebook and compatible with PROGEDO-Quetelet Diffusion, the French social science data portal. Based on the specific needs of INED catalogue, we report on our evaluation and on the pros and cons of each tool considered. We also discuss the

benefits of software making use of DDI-Lifecycle for our longitudinal surveys. We conclude with the reasons that will make our choice.

15:00 - 16:15 : DDI Software

Chair : Carsten Thiel

15:00 - 15:20 : A Collaborative Questionnaire Editor based on DDI 3.2 LC

Claus-Peter Klas (GESIS - Leibniz-Institute for the Social Sciences, Cologne)
Topics : DDI Software **- Session Type :** Regular presentation

At GESIS we developed a collaborative questionnaire editor as a web portal initially for the use case of the German national election studies (GLES).
The features of the questionnaire editor are:
- Creation, editing and structuring of questionnaires with questions, question grids and free statements
- Translation of questions, answers, interviewer instruction, filter statements and free statements into (European) languages
- Workflow support through status on questions
- Sophisticate administration of users and roles and role management
- Simple discussion tool to discuss questions e.g. for question development
- Export to Word files with templates
- Plan for API to connect the questionnaire editor with CESSDA European Question Bank

We will demonstrate the questionnaire editor and also give access to our test version to try out creating your own questionnaire.

15:20 - 15:40 : A converter for DDI: From CodeBook to Lifecycle

Claus-Peter Klas (GESIS - Leibniz-Institute for the Social Sciences, Cologne)
Topics : DDI Software **- Session Type :** Regular presentation

On earlier conferences, we talked about the concepts of our approach for efficient management of DDI files for documentation applications called DDI-FlatDB. We also showed how to use this approach for bringing metadata into search and dissemination applications, e.g. the GESIS Questionnaire Editor. Now we want to share the experiences using DDI-FlatDB when it comes to convert instances between major DDI versions, e.g. DDI CodeBook 2.5 to DDI Lifecycle 3.2
This conversion of DDI is an absolute necessity to support infrastructures to convert their metadata for the use in new arising documentation tools or for the development of such tools and services.
A general point is, that we promised, that we could convert any DDI version to another, just by configuration files. But due to significant different models, e.g. CodeBook vs Lifecycle, we realized, that still programmatic work is needed. Making e.g. question texts, codes and categories from NESSTAR or DDI 2.x variables usable within applications that were implemented along DDI 3.x is challenging. Based on the splitting and entity parsing mechanisms of DDI-FlatDB, we managed to create a converting path from e.g. NESSTAR outputs from various sources into DDI LC 3.2 on variable and question level. This is the first step for indexing questions for the CESSDA European Question Bank search portal or the GESIS ExploreData Variable search portal. We will present the open source conversion tool and describe the process and the lessons learned.

Dan Smith (Colectica)
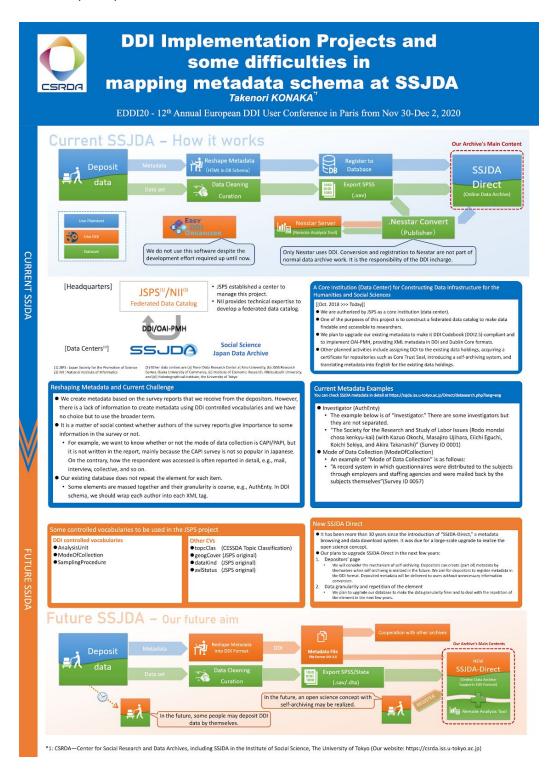**Topics :** DDI Software **- Session Type :** Regular presentation

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

In this session we will demonstrate new features of Colectica 6, including:
- Support for DDI Lifecycle 3.3
- Management of statistical classifications based on the GSIM and DDI 3.3 classification models
- Cross study harmonization and visualization of Conceptual Variables. When organizations each create their own Conceptual Variables, new techniques are needed for cross organization harmonization.
- Colectica Datasets: a new application to add rich metadata directly to statistical data files, and to convert statistical data files to open formats
- Colectica Question Bank: a new web based, streamlined question design application including approvals, publications, and workflows.

Poster

Social Science Japan Data Archive (SSJDA) has been developing DDI implementation projects. We present ongoing projects and some difficulties in the project. Now, users can access more than 1300 datasets and metadata, but those are not based on DDI format. We plan to make metadata in the DDI format (DDI 2.5). But we have some problems with DDI implementation projects. In particular, it is difficult to map between the controlled vocabularies in DDI and some survey methods commonly applied in social surveys in Japan. There are also differences in the granularity of metadata, such as the 'Topic Classification' element. We like to discuss these projects and these difficulties in the project with the EDDI20 participants.

Authors Index

Organizations Index

EDDI Code of Conduct

The European DDI (EDDI) holds its conferences 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.

To provide all participants - members and other attendees, speakers, exhibitors, staff and volunteers - the opportunity to benefit from the event, EDDI is dedicated to a positive, safe and harassment-free conference experience in which diverse participants may learn, network and enjoy the company of colleagues in an environment of mutual human respect and regardless of gender, sexual orientation, gender identity, gender expression, disability, physical appearance, ethnicity, religion or other group identity or political beliefs.

We recognize a shared responsibility to create and hold that environment for the benefit of all. Harassment, therefore, is specifically prohibited. Harassment is any verbal or non-verbal behaviour that threatens another person or group and that reinforces social structures of domination related to gender, gender identity and expression, sexual orientation, disability, physical appearance, body size, race, age and/or religion.

Examples of harassing behaviour include
- The gratuitous use of sexual images in public spaces,
- Deliberate intimidation,
- Stalking,
- Following,
- Harassing photography or recording,
- Sustained disruption of talks or other events,
- Inappropriate physical contact, and
- Unwelcome sexual attention.

Speakers are asked to frame discussions as openly and inclusively as possible and to be aware of how language or images may be perceived by others, especially given the international aspect of this organization.

All participants are expected to observe these rules and behaviours in all conference venues, including online venues, and conference social events. Conference participants seek to learn, network and have fun.

Please do so responsibly and with respect for the right of others to do likewise. If you wish to notify a breach of the Code of Conduct, please speak to or write to the conference Co-Chairs or with a member of the Program Committee.
You will receive an acknowledgement of your complaint as quickly as is practicable but at a maximum within 24 hours.

Participants asked to stop a hostile or harassing behaviour are expected to comply immediately. Conference participants violating these rules may be sanctioned or expelled from the conference without a refund at the discretion of the EDDI Program Committee.

On behalf of the Program Committee, eddi20prog@googlegroups.com
Mari Kleemola (Conference Co-chair), mari.kleemola@tuni.fi FSD - Finnish Social Science Data Archive, Finland
Jon Johnson (Conference Co-chair), jon.johnson@ucl.ac.uk CLOSER, UCL, United Kingdom

About EDDI and EDDI20

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.

EDDI20 Committees

Conference Co-Chairs

Jon Johnson, CLOSER, UCL, Institute of Education, United Kingdom
Mari Kleemola, FSD – Finnish Social Science Data Archive, Finland

Program Committee

Alina Danciu, Center of Socio-Political Data, Sciences Po Paris, France
Uwe Jensen, GESIS - Leibniz Institute for the Social Sciences, Germany
Jon Johnson, UK Data Service, United Kingdom
Mari Kleemola, FSD - Finnish Social Science Data Archive, Finland
Nicolas Sauger, Center of Socio-Political Data, Sciences Po Paris, France
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

Organisation Committee

Nikos Askitas, IDSC, IZA – Institute of Labor Economics, Germany
Alina Danciu, Center of Socio-Political Data, Sciences Po Paris, France
Marie D'arcimoles, Center of Socio-Political Data, Sciences Po Paris, France
Uwe Jensen, GESIS – Leibniz Institute for the Social Sciences, Germany
Jon Johnson, CLOSER, UCL, Institute of Education, United Kingdom
Mari Kleemola, FSD – Finnish Social Science Data Archive, Finland
Nicolas Sauger, Center of Socio-Political Data, Sciences Po Paris, France
Amélie Vairelles, Center of Socio-Political Data, Sciences Po Paris, France