‹ddi› DIRECTIONS
A newsletter of the Data Documentation Initiative

From the Director

This issue of *DDI Directions* provides a look at some of the events and developments taking place in the DDI community over the past few months. As always, thanks go to all of the hardworking people who support DDI and are helping to spread its adoption.

Mary Vardigan, Director, DDI Alliance, vardigan@umich.edu

In This Issue

Volume VII, Number 3, July 2014

New Officers Lead DDI Scientific Board

In May, the DDI Alliance chose officers to oversee the Scientific Board, the arm of the Alliance that has responsibility for shaping the DDI specifications. The new officers are Adam Brown, Statistics New Zealand, Chair, and Steve McEachern, Australian Data Archive, Vice Chair.

Adam Brown   Steve McEachern

Members Join the Alliance

The DDI Alliance recently welcomed two new member institutions into the membership:

- The ITMO Sante Publique, Aviesan -- Epidemiology France Portal. Simon Saint-Georges will serve as the Member Representative and Mari Lhosmot as the Scientific Board Representative.
- The University of Kansas, Institute for Policy & Social Research (IPSR) (Associate Member). Larry Hoyle will be the Member Representative.

New DDI Representatives in Place

Three new DDI representatives have recently taken over this position for their organizations: Nicole Kirgis is now the representative for University of Michigan, Survey Research Operations, and Anne Sofie Fink Kjeldgaard is now representing the Danish Data Archive. Vigdis Kvalheim is the new representative for the Norwegian Social Science Data Service.

Meeting of Members Held in Toronto

The DDI Alliance met in Toronto on June 2 with the morning devoted to administrative matters and the afternoon focused on substantive concerns. Chair Gillian Nicoll (Australian Bureau of Statistics) led the Meeting of Members and Scientific Board Vice Chair Steve McEachern (Australian Data Archive) chaired the first official meeting of the Scientific Board.

Photo by Sanda Ionescu

*Scene from DDI Meeting of Members: From left, Catharina Wasner (GESIS),
Dan Gillman (Bureau of Labor Statistics), Wolfgang Zenk-Möltgen (GESIS),
Kelly Chatain (Survey Research Operations, University of Michigan),
Jannik Jensen (Danish National Archive), Tom Piazza (University of California, Berkeley)*

Second Annual NADDI Conference Held in Vancouver

The [2nd Annual North American DDI Users Conference](#) took place in Vancouver on March 31-April 2, 2014. Organized by the University of British Columbia, Simon Fraser University, University of Alberta Libraries, and the Institute for Policy and Social Research at the University of Kansas, the meeting was held at Simon Fraser University, with the theme "Documenting Reproducible Research."

Ann Green discusses DDI and data quality at NADDI 2014.

The Keynote Speaker for the conference was Ann Green. Ann spoke about committing to data quality and the need for "intelligent openness" with respect to providing access to data. The conference program featured 20 sessions on a wide range of topics. Training workshops were held on Monday, March 31: Jane Fry, Carleton University, provided an introduction to DDI in the morning, showcasing the power of DDI metadata; in the afternoon Barry Radler, University of Wisconsin, and David Johnson, University of Kansas, led a session on the use of DDI to document health-related data.

The conference hosted around 40 participants from five Canadian provinces, eight U.S. states, and three countries.

NADDI presentations are now [online](#).

New Collaboration Site Established for DDI Moving Forward Project

DDI 4 Project Manager Thérèse Lalor of the Australian Bureau of Statistics has implemented a [communication platform for the DDI community](#) to encourage progress on DDI 4. There is a great deal of information on the site, including a slide deck for lay audiences called "Explain DDI 4 to Me," information on all of the sprints, and plans for future work.

NADDI Sprint Results in Important Outcomes

A "sprint" was held in Vancouver, BC, Canada, the week of March 24, just before the NADDI conference, with the goal of accelerating progress on DDI 4 content modeling. Led by Project Manager Thérèse Lalor, the sprint participants accomplished a lot. One important outcome of the meeting was that a new architecture for DDI 4 was specified. This innovative approach involves a Library of DDI elements and Functional Views on the Library. Functional Views are similar to DDI profiles in that they are subsets of the full Library. They focus on common use cases – for example, the set of DDI elements needed for data discovery, a simple codebook, a simple instrument, etc.

Participants wrote a [paper](#) that describes this approach and further develops the technical framework for generating XML and RDF representations of DDI 4. [More about the sprint](#) is available on the DDI collaboration site.

DDI Modeling Work Taking Place Online

Olof Olsson of the Swedish National Data Service (SND) has set up a Web site for DDI 4 modelers to enter information about DDI elements (see [lion.ddialliance.org](#)). Based on the properties and relationships entered, UML model graphs can be generated. The Web site, implemented using the Drupal content management system, makes the development of DDI more transparent and open.
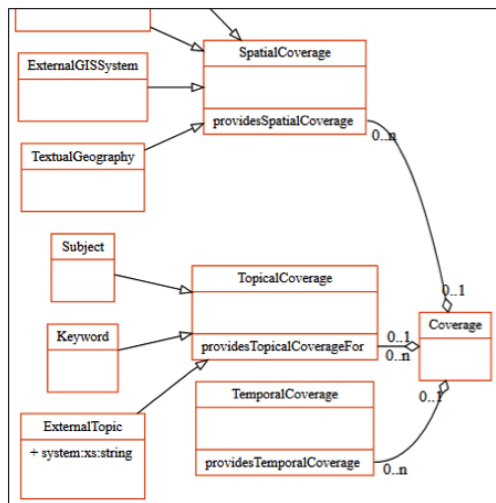
Section of a graph representing data discovery elements in DDI 4 on the lion.ddialliance.org site

IASSIST Session Focuses on DDI Tools

A session on "No Tools, No Standard – Software from the DDI Community" was held during the IASSIST conference, which took place in Toronto, Ontario, June 3-6. This session, which led off with a presentation on the DDI developers community, provided a summary of some of the tools and functionality now available for DDI.

- Samuel Spencer, Australian Institute of Health and Welfare, presented the Canard Questionnaire Suite, which is designed to produce and publish well-documented questionnaires using DDI metadata. The target audience for this tool, which has a drag-and-drop, point-and click interface, is researchers and statisticians. Built using XSLT plug-ins, the tool can export DDI as XForms, DDI 3.1, and 3.2, as well as other XML and text formats.
- Ingo Barkow (TBA21 and DIPF) provided a first look at the generic DDI-L version of the Rogatus Survey/Case Management System. Rogatus, which means "respondent" in Latin, has two parts: Rogatus Survey, and Rogatus Repository (for dissemination and preservation). Based on the Generic Longitudinal Business Process Model (GLBPM), Rogatus covers the entire lifecycle, including managing fieldwork and designing and translating questionnaires.
- Olof Olsson, Johan Fihn, Akira Olsbanning (Swedish National Data Service) and Jannik Jensen (Danish National Archive) provided updates on XSLT stylesheets for DDI that transform DDI XML into other formats. The stylesheets, which are free, support MARCXML (a schema based on MARC 21); DataCite version 3.0; DDI-to-XHTML interactive codebook; transformations from DDI Lifecycle to DDI Codebook, including HTML-formatted fields; and Nesstar and IHSN codebook formats. All transforms are available online for uploaded DDI Lifecycle files.
- Andrew DeCarlo (Metadata Technology North America) presented the OpenDataforge Tools Suite, which consists of several DDI-based tools. Sledgehammer reads statistical packages and creates DDI metadata; Caelum generates reports and codebooks, as well as other documentation; and Caster works with databases (MySQL, Oracle, SQL Server, DB2, HSQL, Postgres, and Sybase) to generate standard documentation (DDI, Triple-S) and statistical package set-up files.
- Dan Smith (Colectica) presented the Colectica 5 (now with DDI 3.2) software, which has several new features, including support for aggregate data (NCubes); quality reporting; a translatable user interface; and support for user-created add-ins. There are new survey features as well: new question types (grids, question blocks, rankings, distributions, scales); flowchart visualization of survey flow logic; and parameters and bindings for describing processing in instruments. Colectica 5 supports Blaise, RedCap, and paper questionnaires.
- Marcel Hebing (DIW Berlin – German Institute for Economic Research) discussed two

utilities he has developed: DDI on Rails for SOEP (German Socio-Economic Panel Study) and r2DDI, both of which are open source implementations. DDI on Rails supports the entire data lifecycle for researchers using SOEP data. It integrates the Solr search and metadata cross-tabulations to examine how variable labels, values, and questions have changed over time. The r2DDI program is a script generator that works on a basket of requested variables.

- Olof Olsson and Jannik Jensen presented an update on the DDI-RDF Tool (Disco), which resulted from DDI Hackathons held in 2013. The tool, which translates DDI to the RDF Discovery vocabulary (Disco), is based on a deployed SPARQL Web service endpoint (Jena).

New DDI Profiles Resources Available

Two new tools (XSLT stylesheets) developed by Joachim Wackerow are now available for DDI Profiles. One creates a DDI Profile template on the basis of an existing DDI instance. The other renders a DDI Profile in HTML, including links to the field-level documentation of the used elements. Versions for DDI 3.1 and 3.2 are available.

Also, a set of DDI Profiles has been developed by the community working with the Generic Statistical Information Model (GSIM). These profiles are intended to harmonize the use of DDI in statistical organizations and to help ensure that GSIM implementations using DDI will interoperate. As part of the GSIM work, DDI Profiles were developed for Basic Technical Objects, Variable, Represented Variable, Questionnaire, and Codelist.

New RDF Vocabularies Under Review

Three new RDF vocabularies – DDI-RDF Discovery (Disco), Physical Data Description (PHDD), and Extended Knowledge Organization System (XKOS) – have been created and are now out for public technical review. The vocabularies are scheduled to be finalized by the end of the year.

Copenhagen Mapping Out for Review

The Copenhagen Mapping is an effort to implement GSIM 1.1 Statistical Classifications using the new DDI 3.2. This work started at a Classifications workshop hosted by Statistics Denmark in December, and also included participants from Statistics Sweden, the Danish Data Archive, and Colectica. A first draft of the Copenhagen Mapping has been completed, and the authors are looking for feedback.

New DDI Controlled Vocabulary Published

The DDI Alliance Controlled Vocabularies Working Group (DDI-CVG) is pleased to announce the publication of a new controlled vocabulary for describing the Mode of (data) Collection. For more information regarding the use of the controlled vocabularies published by the DDI Alliance, and the work of the DDI-CVG please see the Controlled Vocabularies section on the DDI Alliance Web site.

DDI Handbook Project Begins

Organized by Joachim Wackerow and supported by GESIS-Leibniz Institute for the Social Sciences, a group met in Cologne during the week of November 11-15, 2013, to launch a new project with the aim of producing a collection of best practices for using DDI. These best practice descriptions will be modular with a homogeneous format, allowing reorganization in multiple ways. The primary structure for the collection will be organized in alignment with the DDI Lifecycle. A goal will be to involve the DDI community in producing a shared body of resources for all organizations and individuals using the DDI specification. Best practices will be reviewed by a team of editors and reviewers and published on a dedicated Web site.

The project will produce guidelines for institutions introducing DDI into their workflows and for

other institutions already using DDI Codebook and shifting some of their workflows to DDI Lifecycle, particularly in the area of archival processing activities. Others contributing to the activity include Peter Granda (ICPSR), Larry Hoyle (University of Kansas, Institute for Policy and Social Research), Catharina Wasner (GESIS), and Wolfgang Zenk-Möltgen (GESIS).

Joachim Wackerow gave a [presentation](#) on the Handbook at IASSIST in June.

Meeting on Documenting Data Transformations Held in Ann Arbor

On June 9-10, a group of international experts met in Ann Arbor, MI, to discuss requirements for tools to document data transformations in DDI. With support from NSF and the World Bank, the meeting focused on defining a standard language for describing data transformations. While a system-independent language currently under development was considered, the group decided to create a new standard appropriate for variable-level transformations. To that end, a list of the fundamental data transformation commands in each of the four main statistical packages (SAS, SPSS, Stata, and R) was compiled and consolidated. The group intends to continue on as a DDI working group and the group's output will be integrated into the DDI modeling efforts. Tools will be developed in a future phase of the project, and they will support both DDI Codebook and DDI Lifecycle.

Colectica To Develop Data Curation Software With Yale ISPS and Innovations for Poverty Action

Colectica is partnering with the [Institution for Social and Policy Studies (ISPS)](#) at [Yale University and Innovations for Poverty Action (IPA)](#) to develop a repository for research data from randomized controlled trials in the social sciences. The repository will be an expansion – and major upgrade – of the existing [ISPS Data Archive](#). The main objective of this project is the technical integration of the various curation processes with inventory management, metadata workflow, and Web access.

"DDI 3.2 Lifecycle allows precise description of curation and data archival operations over time," said Jeremy Iverson, a partner at Colectica. "We are pleased to be working with ISPS and IPA on this exciting project."

Colectica and the partners will develop a software platform that allows archives to leverage the DDI 3.2 standard for data documentation and to structure the curation workflow, which includes checking data for confidentiality and completeness, creating preservation formats, and reviewing and verifying code. The new integrated system will combine several open source and off-the-shelf components with a new, web-based ISPS-IPA Data Pipeline application, and will enable a seamless framework for collecting, processing, archiving, and publishing data.