

DDI Alliance Annual Meeting of the Scientific Board
Monday, June 1, 2015
Minnesota Population Center, University of Minnesota
Minneapolis, Minnesota, USA

Minutes

Present:

- Ingo Barkow, German Institute for International Educational Research
- Anja Burghardt, Research Data Centre of the German Federal Employment Agency, Institute for Employment Research (IAB)
- Michelle Edwards, Cornell Institute for Social and Economic Research (CISER)
- Johan Fihn, Swedish National Data Service (SND)
- Daniel Gillman, U.S. Bureau of Labor Statistics (Associate Member)
- Arofan Gregory, Metadata Technology (Consultant)
- Marcel Hebing, German Socio-Economic Panel Study (SOEP)
- Matti Heinonen, Finnish Social Science Data Archive
- Larry Hoyle, University of Kansas, Institute for Policy and Social Research (IPSR) (Associate Member)
- Chuck Humphrey, University of Alberta
- Sanda Ionescu, Inter-university Consortium for Political and Social Research (ICPSR)
- Anne Sofie Fink Kjeldgaard, Danish Data Archive
- Vigdis Kvalheim, Norwegian Social Science Data Service (NSD)
- Amber Leahey, University of Toronto
- Jared Lyle, Inter-university Consortium for Political and Social Research (ICPSR)
- Steven McEachern, Australian Data Archive (ADA)
- Katherine McNeill, Massachusetts Institute of Technology (MIT)
- Ron Nakao, Stanford University
- Gillian Nicoll, Australian Bureau of Statistics (ABS)
- Olof Olsson, Swedish National Data Service (SND)
- Carol Perry, University of Guelph
- Anita Rocha, University of Washington, Center for Studies in Demography & Ecology (CSDE)
- David Schiller, Research Data Centre of the German Federal Employment Agency, Institute for Employment Research (IAB)
- John Shepherdson, United Kingdom Data Archive
- Dan Smith, Colectica (Associate Member)
- Wendy Thomas, University of Minnesota, Minnesota Population Center
- Leanne Trimble, University of Toronto

- Heidi Tvedt, Norwegian Social Science Data Service (NSD)
- Mary Vardigan, Inter-university Consortium for Political and Social Research (ICPSR)
- Joachim Wackerow, GESIS - Leibniz Institute for the Social Sciences
- Marion Wittenberg, Data Archiving and Networked Services (DANS), The Netherlands

Observers:

- Shane McChesney, Nooro Online Research
- Barry Radler, University of Wisconsin

1. Model-based DDI – Update on Status

Steve McEachern, Vice Chair of the Scientific Board, opened the meeting with a discussion of the new model-based specification and its status. He noted that a draft release was available for review and the comment period had been extended to June 15, 2015. For this first release, there was an emphasis on base objects and packages of the Library such as individuals, organizations, data types, basic processes, and primitives – in essence, the building blocks. Functional Views would follow, which would serve as profiles or subsets of the overall model to enable users to document their data using only the elements needed.

It was reported that an adjusted timeline for the planned releases was in the process of being prepared. The DDI Moving Forward Advisory Group (AG) was guiding the direction of the development work along with the active working groups and modelers. In addition to the model, teams were also actively producing bindings to XML and RDF.

Enhancements

Several enhancements were being undertaken with respect to the new specification:

Data Citation – Larry Hoyle and Mary Vardigan had a small NSF grant to work on ensuring that DDI was extended to include comprehensive information on data citation. With the new model, there is now a mechanism to add creators and contributors as well as their roles and degrees of contribution. Additional work was taking place to allow other kinds of information to be structured. Almost any object in DDI can be cited.

Datum – The Data Description Group was looking at new ways to think about the “datum.” This work aligns with work going on at the NSD. The group was also looking at how to document new types of data like biomedical instruments.

Methodology – The new model-based DDI will have much richer information about methodology.

Integration – Work was taking place to bring DDI Codebook and Lifecycle into an integrated model, and there were discussions about calling the model-based DDI simply “DDI,” as it would be a single integrated framework.

Data Management Plans – Along with Study Inception, this is on the list for future work to be undertaken as part of the model-based specification.

Challenges

Resourcing

Resourcing was seen as one of the biggest issues. A lot of people, including many with in-depth expertise, were involved in the development process for the new specification, but there was a need for even more people to be involved. There were still some open and unscoped content areas to work on, which would require yet more resources. Another huge challenge mentioned was that all the production work was purely voluntary. There had been a lot of difficulty getting all the pieces to work together. Having more use cases to help expand and clarify structures and test against would be welcomed.

It was pointed out that the role of the Scientific Board from a technical standpoint is to make a case for the resources needed to develop the technical standard. The Scientific Board can articulate the case and then take this to the Executive Board, which can determine how to fund the request or allocate resources.

A view was expressed that the Alliance cannot continue to work as an altruistic organization without paying people. The alternative is to insist that if you use it, you pay for it. NSOs, which tend to have more funding, should pay for DDI membership if they are using it. However, the Bylaws state that “The DDI standards shall be publicly and internationally available free of charge to anyone.”

It was noted that what the organization had accomplished with in-kind contributions was impressive, but there were concrete things that we could possibly pay for to move things forward. Project management as well as development and maintenance of the production framework were possible candidates for funding. We are used to everyone performing heroics, but we may be able to pinpoint critical activities that are funding priorities.

Project management

Lack of a Project Coordinator or Manager to guide the model-based development was another challenge identified. There was no dedicated manager for the past year, although the AG had managed the work as well as it could. This is as an ambitious work program and as such really calls out for someone to manage it. In the budget going forward, we have a role for a Communications person since we received no response to the request for a Project Manager.

There is not enough money to fund these positions as they should be funded (\$60-70K to be successful).

Sustainability

As mentioned in the partnerships discussion, DDI needs to connect to other standards and domains. Also, we should not do the modeling and production in a new way every time we release a new version; rather, we need a stable structure and then we can just add content to that structure.

To ensure a solid foundation, we should solicit outside views on what we have accomplished and plan to do going forward, asking others if the structure we have built is good and how it aligns with other standards. Also, we have a good list of design rules. We should now review our work according to the design rules, and then review the rules themselves. Planning was under way to incorporate some of this review into the sprint to take place in October at Schloss Dagstuhl.

Opportunities

A suggestion was made that we should prioritize the Data Management Plans (DMPs) Functional View as we have a lot of expertise to draw on and previous work to build on. Also, from the marketing perspective this is a good direction to move in since most funding agencies now require DMPs. There is a new Research Data Alliance (RDA) Working Group on Active Data Management Plans, which CASRAI (Consortia Advancing Standards in Research Administration Information (CASRAI) is participating in, and we should join up with this work. We can also look at the DMP editor built at University College London. Members were advised that this is not a mapping exercise for DDI and Data Management Plans – rather, it is the exchange between them that we need to facilitate. How can we exchange DMPs so that they can be integrated into DDI?

A question was asked regarding when groups could start building customized Functional Views on their own. The response was that when variables, data descriptions, instruments, and codebooks were in place and accurately modeled, we would be closer to this goal.

A point was raised about integrating existing DDI Working Groups into the model-based development process, which was viewed as a good idea. The Scientific Board can take the lead in managing this.

Licensing and Copyright

Wendy Thomas, Chair of the Technical Committee, noted that in the past the specification and its documentation were released separately and had separate licenses. However, the new specification will be an integrated whole and thus, the Technical Committee (TC) has discussed

licensing in this new environment. They reviewed most of the existing licensing structures and wrote a proposal, which they have submitted to the AG for comment. The proposal is to shift from the current “copy left” license (GNU Lesser General Public License for the specification) to a more permissive license, Creative Commons International. This means that users would be free to share and adapt the work even commercially but they must attribute and provide appropriate credit for the parts that came from DDI. Any training materials should be copyrighted and allow commercial use.

The TC advised that the Executive Board should also review the proposal and it should be cleared by UM counsel. Licensing would be discussed further during the upcoming year and would probably be on the agenda for the next Scientific Board meeting. The policy needs to be approved by the Scientific Board. More permissive licenses have advantages but also carry risks, so this is not an easy decision. Members noted that they have connections to the Open Knowledge Foundation and could consult with them about these issues. Also raised was the point that the Alliance needs a Contributor’s Agreement.

2. DDI Lifecycle 3.3 Update

The TC noted that they had performed an internal review and were finalizing the update and fixing issues discovered. For the most part DDI 3.3 addresses bugs and incorporates the Survey Design and Implementation Group’s work on weighting, sampling, and design process.

3. Atlassian Products

The Alliance has acquired an open source license to use these Atlassian products:

- Confluence – This serves as the collaboration wiki for the DDI development work and the broader Moving Forward project. Working Groups use it to document their work and meetings.
- Jira
- Bitbucket

The Confluence implementation also includes a separate section for DDI committees and working groups. The TC has a dedicated page and it is planned that other groups will do the same. There is a basic template for committee and group pages so that the structure is consistent and standardized. Each group page has a purpose, a list of members, minutes, and a calendar as well as connection details for calls.

4. CSPA Update

Gillian Nicoll provided an update on the Common Statistical Production Architecture (CSPA), reference architecture intended to foster interoperability and to help NSOs in their modernization work, based on common standards:

- GSBPM (Generic Statistical Business Process Model)
- GSIM (Generic Statistical Information Model)
- DDI/SDMX

CSPA is currently developing a Logical Information Model, which has caused some concern in that DDI is also creating a logical model. Timing is the issue: DDI is on a slower timeframe, and the NSOs need a logical model sooner than DDI can deliver it. To prevent conflicts between the models, there is participation by a few key people across the initiatives.

5. Working Group Reports

A few of the current DDI working groups reported on their activities over the past year.

Technical Committee

Wendy Thomas summarized the Technical Committee's work over the past year. She noted that they were currently engaged, along with the Modeling Team, in the Moving Forward development work. The role of the TC is to prepare the releases for publication and review. After the new specification is completed and that project ends, the TC will go back to its traditional maintenance and development role.

RDF Vocabularies Group

Joachim Wackerow noted that there was a formal review of the RDF vocabularies during the year and comments were received. The group will now focus on final review as a final step before publishing the vocabularies.

Controlled Vocabularies Group

Sanda Ionescu reported that the CVG had published many new CVS and updated other during the year. Also, some of the archives in Europe were using many of the vocabularies created and there was wider interest in them. We should continue to raise awareness of these as they can be used in metadata catalogs as well as in DDI instances. Translations in German, Swedish, and Finnish are available for some of the vocabularies.

6. Celebration

The day's meetings adjourned with a celebration of the 20-year anniversary of the DDI.