

DDI Alliance Scientific Board Meeting
Sydney, NSW, Australia -- UNSW Sydney
June 1, 2019

Minutes

Participants:

Tuomas Alaterä (Finnish Social Science Data Archive)
Iris Alfredsson (Swedish National Data Service)
Ingo Barkow (HTW Chur)
Cathy Fitch (Minnesota Population Center)
Jane Fry (Carleton University)
Jared Lyle (ICPSR)
Steve McEachern (Australian Data Archive)
Marianne Myhren (Norwegian Centre for Research Data)
Ron Nakao (Stanford University)
Barry Radler (MIDUS, University of Wisconsin--Madison)
Wendy Thomas (Minnesota Population Center)

Virtual Participants:

Dan Smith (Colectica)
Jon Johnson (UCL - CLOSER)
Guillaume Duffes (INSEE)

Scientific Board Structure

Ingo Barkow, Vice Chair of the Scientific Board, began the meeting by noting the current limitations of the Scientific Board structure (see Appendix 3). He indicated that the Scientific Board has not been active and that the Alliance Bylaws are not clear enough about the Board's boundaries and processes. He noted that it is hard to represent a group when they don't know what the members want.

He outlined a proposal to create an Acting Committee of the Scientific Board whose members are elected by all members of the Scientific Board. The Acting Committee would be formed by five elected member representatives and two external advisory experts. The Acting Committee would meet monthly in a conference call, as well as one yearly face-to-face meeting in the margins IASSIST, EDDI, or NADDI.

The details of the proposed Acting Committee would be developed by a working group over the next couple of months. The working group would propose the new structure, including clarifying

the roles, number of people, and necessary changes to the Alliance Bylaws. Ingo would chair the working group.

It was asked what the composition of the working group would be. Ingo indicated it would be open to anyone. It was asked how the Scientific Board will ensure fair representation on the Acting Committee so it does not become an echo chamber. It was discussed how people should be elected to the Acting Committee. It was noted that the Scientific Board has never before asked for comments or input from its members, so why should it create a small group if they haven't asked the whole group in the first place. It was also asked what the role of the individual members will be.

It was noted that the working group will discuss and propose solutions for these questions. It was suggested to have two external representatives on the working group. It was noted that the new group needs to be organized to get community feedback rather than making decisions in closed chambers. It is hoped that the new body can ask member representatives about priorities; we only know about reported issues, not what the scientific board representatives think.

As a next step, the Executive Board will ask for volunteers and require external representation on the working group.

Scientific Board Leadership

Steve McEachern, Chair of the Executive Board, reviewed a proposal discussed by the Executive Board in their March 2019 meeting to postpone elections for the Chair and Vice-Chair of the Scientific Board until July 2020. The three year terms for the Chair and Vice-Chair positions of the Scientific Board end in June 2019. Given the proposed changes to the structure of the Scientific Board, the one year delay will allow the proposed new structure to be voted on and implemented before new leadership is elected.

The Scientific Board membership should review the proposal to delay Scientific Board leadership elections, with the option to reject. No one attending the annual meeting voiced objection. An email will be distributed to the membership informing them of the proposal. If any members reject the proposal to delay elections, the Alliance will hold an election.

Training

Ingo discussed a proposal to publish a modular training library for further reuse and self-guided training (see Appendix 4). The goal of the training library is to provide consistent and appealing teaching material, covering all relevant areas, following a version-agnostic approach, and providing version-specific details. The training library would form the basis of teaching tutorials to multiple audiences like tutorials in the margins of conferences or workshops in summer schools. It would also be the basis of online training material. It was noted that the current training library has gaps in terms of covering all relevant areas and has limitations regarding content consistency and look-and-feel of the slides. The proposal for the training library is to

fund a paid person to edit the library. Volunteers can help in providing material for missing areas.

It was requested to define the training library. It was asked who is the audience for the training library, where are the gaps, and whether this effort will produce new training material. Additionally, it was discussed whether the training document manager activities could be combined with the efforts of the proposed marketing manager.

Scientific Board Direction and Goals

Ingo presented slides discussing possible goals for the Scientific Board for the next year, including interoperability of metadata standards, improving infrastructure around all DDI specifications, and project management for all DDI Alliance activities (see Appendix 2).

For slide 7, it was requested to better define and update “Project Management for all DDI Alliance activities.”

Moving Forward

Specifications Roadmap

Wendy Thomas, Chair of the Technical Committee, discussed the DDI Roadmap document prepared and recently updated by the Technical Committee (see Appendix 9). The document provides a broad discussion about existing products and how to frame DDI 4.

It was noted that development of the DDI specifications has shifted focus to applied usage and the environments in which they're used. Also, development work on the different lines is not distinct; that is, the Alliance has added things to Codebook and parallel things in Lifecycle. This speaks to the issue of whether to cut off development and just fix bugs, or to continue to develop different lines.

Regarding DDI 4, concern was raised about being able to offer users a production model. It was recommended to keep discussion of DDI 4 to a minimum until the Alliance has a product to offer, as well as functionality that DDI 2 or DDI 3 cannot offer. It was suggested that functionality in DDI 4 could be added to DDI 3 to enhance DDI 3.

It was recommended to ask the community for their set of requirements and needs for different streams of development, including: Which version are you implementing now? What will you need in the future? What is the timeline? Where do you want to devote voluntary resources? Understanding these questions will help figure out timelines and priorities for the roadmap.

DDI Core

Steve McEachern described the DDI Core, which will use the core features of the DDI 4 model that are the most robust to date, conceptual, data description, and process, with a 'core' DDI 4 release that is implementable and the base on which to update the rest of the model. The goal is a DDI 4 Core release for review and publication in December 2019.

DDI Core will be fit for cross-domain integration of data and ready for data coming from new sources (e.g., “big data,” social media, sensor data, etc.). Features will include: the conceptual aspects of variables and classifications; the datum-oriented description of data; and the use of the process model to describe data lineage (the processing involved in the provenance of data). The architecture of DDI4 Core can be understood as a kernel on which additional features and user-oriented views can be arranged on layers. (See Appendixes 5-8, 10.)

Technical Committee FY2020 Workplan

Wendy Thomas reviewed the fiscal year 2020 work plan for the Technical Committee (see Appendix 12), including:

- Reviewing and publishing DISCO
 - Content is ready to be put out for a 30-day review in June
- Preparing DDI 3.3 for publication
 - Final entry and documentation
- Resolving 5 DDI 4 Prototype review issues
 - Overall best practices for XML structure
 - Shift from document base to multi-use instances (document, transfer, discovery systems, etc.)
 - Specification of default values that are inherited by contained objects
 - Broad feedback from ICPSR regarding interaction of DDI products, audiences, etc.
 - Broad feedback from NSD regarding role of DDI4
- Reviewing issues filed for DDI-Codebook and preparing a new version
 - Currently have a set of issues related to improving link with DataVerse
 - New issues are being filed in terms of issue arising from the new World Bank tool
- Shifting DDI Lifecycle and DDI-Codebook production work to COGS
 - We have tested out the documentation production using the DDI 3.3 public review
 - Test and finalize input-output scripts
 - Complete and test output to ensure coverage and output consistency
 - Finalize transfer and open for access
- Preparing for the shift of DDI 4 development work to COGS from Drupal
 - Finalize scripting for XML input and output

Standards Development document

Wendy Thomas discussed a proposal to update the [Standards Development and Review Process and Procedures](#) document (see Appendix 13). The current document is complicated, does not have a good way to approve minor changes, and requires a two-thirds majority for all votes.

It was suggested that voters should be able to indicate “I have no objection but I can’t evaluate.” It was also suggested that standards reviews and votes should be explained in lay people terms so the general membership can understand how the new changes will benefit them.

The Technical Committee will review and update the document. Thereafter, it will be sent to the Scientific Board and the Executive Board for review.

It was noted that the Alliance needs to add visible functionality to the web site about how to file an issue for the Technical Committee to address.

SDTL

Steve McEachern discussed a proposal by George Alter and the C²Metadata project for the DDI Alliance to monitor, maintain, and update the Structured Data Transformation Language (SDTL) as a standard (see Appendix 14). SDTL is an independent intermediate language for representing data transformation commands. Statistical analysis packages (e.g., SPSS, Stata, SAS, and R) provide similar functionality, but each one has its own proprietary language. SDTL consists of JSON schemas for common operations, such as RECODE, MERGE FILES, and VARIABLE LABELS. SDTL provides machine-actionable descriptions of variable-level data transformation histories derived from any data transformation language. Provenance metadata represented in SDTL can be added to documentation in DDI and other metadata standards.

It was noted that maintaining SDTL would be coordinated in the same way the Alliance handles XKOS or Controlled Vocabularies. If there is a group that wants to do maintenance, our role is to handle the organizational aspects and review for publication.

The next step will be to take this proposal to the Technical Committee. The Executive Board should also discuss the process for deciding how the Alliance formally reviews and accepts proposals like this, including what is the advantage to DDI and the individual product.

Appendix 1

Appendix 2

Appendix 3

Appendix 4

Appendix 5

Appendix 6

Appendix 7

Appendix 8

Appendix 9

Appendix 10

Appendix 11

Appendix 12

Appendix 13

Appendix 14