DDI Alliance Annual Meeting of the Scientific Board
Monday, June 2, 2014
University of Toronto
Toronto, Ontario, Canada

Minutes

Present:

- Ingo Barkow (Institute for International Education Research -- DIPF)
- Kelly Chatain (University of Michigan, Survey Research Operations)
- Franck Cotton (National Institute of Statistics and Economic Studies -- INSEE)
- Johan Fihn (Swedish National Data Service -- SND)
- Dan Gillman (U.S. Bureau of Labor Statistics)
- Arofan Gregory (Metadata Technology)
- Marcel Hebing (German Socio-Economic Panel Study (SOEP), DIW Berlin)
- Chuck Humphrey (University of Alberta)
- Sanda Ionescu (Inter-university Consortium for Political and Social Research -- ICPSR)
- Jannik Jensen (Danish Data Archive -- DDA)
- Mari Kleemola (Finnish Social Science Data Archive – FSD)
- Thérèse Lalor (Australian Bureau of Statistics -- ABS)
- Jared Lyle (Inter-university Consortium for Political and Social Research -- ICPSR)
- Steve McEachern (Australian Data Archive -- ADA)
- Katherine McNeill (Massachusetts Institute of Technology – MIT)
- Ron Nakao (Stanford University Libraries)
- Gillian Nicoll (Australian Bureau of Statistics -- ABS)
- Tom Piazza (University of California, Berkeley, Computer-Assisted Survey Methods -- CSM)
- 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 -- UKDA)
- Dan Smith (Colectica)
- Jon Stiles (University of California, Berkeley, UC DATA)
- Wendy Thomas (University of Minnesota, Minnesota Population Center)
- Leanne Trimble (University of Toronto, Scholars Portal)
- Mary Vardigan (Inter-university Consortium for Political and Social Research -- ICPSR)
- Joachim Wackerow (GESIS – Leibniz Institute for the Social Sciences)
- Catharina Wasner (GESIS – Leibniz Institute for the Social Sciences)
- Marion Wittenberg (Data Archive and Network Services – DANS)
- Wolfgang Zenk-Möltgen (GESIS – Leibniz Institute for the Social Sciences)

Virtual participant
- Adam Brown (Statistics New Zealand)

Observer

- Barry Radler (University of Wisconsin)

Future Procedures of the Scientific Board

The first official convened meeting of the Scientific Board took place during the afternoon, led by Vice Chair Steve McEachern with Chair Adam Brown participating remotely. Steve mentioned that the SB had oversight for the standard and needed to determine how it would operate and what its priorities would be. The Bylaws are relatively open in terms of the SB and its functions.

The rationale for creating the SB was that the Alliance had reached a level of maturity that it required an organizational restructuring for its existing members to bring in focused experts. Up until the new Bylaws were created, Alliance members had served in both administrative and substantive/scientific capacities.

It was pointed out that the Alliance needs to think about how to create a framework for more people to be involved during these transition years. A suggestion was made that the SB Chair report to the EB periodically to ensure good communication flow.

Moving Forward Process

Thérèse Lalor, DDI 4 Project Manager, reported on the status of the DDI Moving Forward Project. The project had begun at a Dagstuhl workshop devoted to this purpose in 2012. During that workshop, participants came up with project drivers and design principles for the new model-based specification and published their findings. After a series of sprints, the project has gained momentum and is developing an architecture and structure that align with the DDI 4 vision and drivers and design principles. However, the project is still determining timelines/operating procedures for achieving the goals in the timeframe that we have.

Architecture
Thérèse explained that the DDI 4 architecture will have a library of elements with functional views on the library. She likened this model to the use of an archive (library of objects), with an archivist only pulling out information relevant to a research purpose (functional view).

The DDI library of objects is organized according to packages – the functional view is the subset of objects a user needs for a particular purpose. Developers might want to look at the entirety of the library, however.

Management and Draft Release Plan
As pointed out in the earlier meeting, DDI 4 has an ambitious timeline with a deadline of March 2015. To achieve this accelerated schedule, the project is using agile development techniques – this is an iterative way to plan and guide a project based on a constant cycle of planning and iteration driven by user needs. A big benefit of this approach is that it can take new user needs into account as development proceeds. Agile is iterative and incremental, adaptive, user-based, and transparent. These characteristics tick the boxes on many of the DDI 4 design principles.

Agile works by defining a content backlog. Each item in the backlog has an associated user story. All of the user stories are available on the UNECE-hosted DDI Alliance collaboration wiki. We will be putting subsets of products from the backlog into beta releases. This idea of releasing smaller chunks of the specification means that there are no big-bang releases as with previous DDI specifications.

In terms of products to be released, Conceptual Core, Process/Agent, and Discovery were scheduled to be part of the first release. Next would be Simple Instrument, Simple Data Description, and Classification.

Process
The workflow for producing Functional Views consists of:

- Setting requirements
- Content and data modeling
- Documenting and producing outputs
- QA by the Technical Committee
- View released for comments

A question was raised about the estimation of how much effort would be required to produce each view as such estimations are a big part of agile development. The answer was that we have a good sense of the effort involved in some parts of the specification, but there are items that are not currently resourced at all. Estimation is in many ways implicit, and we always have more than we can achieve. We have a Project Manager and teams of people doing the work, which is a big step forward for the Alliance.

We have also tested the production framework and now have a good idea of how this will work. There is no overt link yet from the user stories wiki to the modeling, but they are the foundation for the work. The task teams have a link to the Drupal development site, which is linked from the wiki, so there is traceability.

This project represents a culture shift for the Alliance as we are attempting to work in an accelerated manner and not as deliberately as before. The agile principle is that there is an 80/20 rule for beta releases, with the teams attempting to cover 80% of the domain and not waiting for perfection.

Governance of the DDI 4 Project

Steve McEachern raised the issue that for proper governance of the DDI 4 project, which is not specified in the Bylaws, we need a Product Owner who can make decisions, particularly at sprints when time is of the essence. This is another aspect of agile development to ensure that progress continues.

The Technical Committee (TC) plays an important role in the DDI 4 project, participating in all teams with responsibility for carrying products forward; they are not a choke point. Out of this project participation in the TC has broadened, which is a positive development.

SB Chair Adam Brown had pulled together a proposal outlining governance of the DDI 4 project, which meeting participants reviewed. In the proposal, the SB Chair is the Product Owner and there is a Moving Forward Working Group to advise him or her. The working group is a small representative body including people from the user community, the data producer community, the archive community, technical people, and the Executive Director. The Project Manager reports to the Product Owner. They meet monthly and receive regular reports. The EB does not play a role other than financing it and setting boundaries within which the project operates.

This proposal generated a wide-ranging discussion and several points were raised:

- The Product Owner is not the same as the Project Leader but is the owner of the requirements and functions as the person authorized to make decisions on behalf of the advisory group. It was suggested that we come up with a term other than Product Owner for this role.
- This person has overall responsibility to resolve internal conflicts. In a sprint, the Alliance is investing money in the sprint, but others have long-term interest in the standard and decisions must be taken with care, so we need efficient means of making a decision and breaking logjams. The Product Owner also sets releases and what is in the backlog. The wiki is the place to propose user stories; the Product Owner can move this into the backlog.
- Other options suggested were that the SB appoint a Product Owner separate from the Chair, or the Chair and Vice Chair assign somebody to play that role.
- Regarding sprints, participants want to know the specific work they will be tasked to do at a sprint. We need to ensure that we have early consensus on the topics to be covered in each sprint. The SB Chair should be empowered to do this. It was pointed out that it could be dangerous for decisions to be made by a single person, and the full SB may not agree.
- In a sprint there is a need for the Product Owner to provide objective opinions with authorization from the Alliance to act in this capacity. At the same time we need to ensure that the Product Owner is not making decisions in isolation but is advised by a representative group of the SB.
- It was pointed out that a Sprint is short – if a wrong decision is made, we can go back and change it; this is the agile, iterative way of doing things.
- It was requested that any decisions made at a sprint be captured with pros and cons and reported to the full SB. The point was also made that the Chair of the SB has other tasks related to the chairing role, so the Product Owner position should perhaps be held by another person.

In general there was support for the idea of a Product Owner. The SB will send email inviting expressions of interest for the Product Owner and participants on the Advisory Committee.

The Chair and Vice Chair will reword the proposal and bring it back to the SB.

HLG, GSIM, CSPA Update

Thérèse reported on activities taking place in the world of official statistics. There is a High Level Group for Modernization of Statistical Products and Services (HLG) that consists of chief statisticians and heads of national and statistical offices working to modernize the process of producing official statistics.

The Moving Forward work is informed by the HLG process, the goal of which is the development of global standards. There is a need and willingness to work together, and the HLG activities are mostly resourced by volunteers with donations from some NSIs.

The HLG tends to work by selecting a project and working for a year on it. They created the Generic Statistical Business Process Model (GSBPM), which provides a common way to talk about the business process of producing data and is being used by 50 organizations around the world. They also recently finished a new version of the Generic Statistical Information Model (GSIM), a reference model for the information that is important to statistical offices. Like DDI, GSIM was produced in a series of sprints. It has four main parts: statistical needs, process, structure, concepts. Some parts of DDI 3.2 were influenced by GSIM, and DDI 4 is aligning with it. The Conceptual and Process groups looked at this alignment in the sprint.

In terms of GSIM products, "Clickable GSIM" permits one to produce object-related views. It would be great if DDI 4 could mimic this functionality. Also, GSIM has been mapped to DDI 3.1 with profiles available for variables, codelists, and questionnaires, among others.

The HLG is also working on a Common Statistical Production Architecture (CSPA) so that NSIs can share IT costs; this is designed as "plug and play" architecture with standard interfaces. The HLG built five services in 2013 based on DDI interfaces.

The HLG also sponsors a modernization committee on standards – Arofan Gregory is the representative for the DDI Alliance.

In general, the DDI and GSIM partnership has been positive and the GSIM and DDI specifications are converging.

New Collaboration Solutions

The Alliance is currently using a number of different tools and technologies in its work; it makes sense to combine technologies so that we can work on a common platform. The Atlassian product suite provides a wiki, issue tracker, and code repository. If we adopt this suite, Stash would be used for the source code repository, Jira for issue tracking, and Confluence as the wiki platform for collaboration. The Alliance should qualify to use these products for free because our specifications are open. The following data

there would be a meeting of interested people to discuss this further. The goal was to get this moving so that we have a functioning platform in September.

A question was raised about how this suite of products was chosen because Github, used by most DDI developers, would be an alternative. The response was that there was a requirements list that served as the foundation for selecting the Atlassian products.

Another question was raised about whether the application includes maintenance. This is still an open question. There are two options: have the products hosted by a member or use the on-demand option of the company. Either way configuration has to be done; software updates are a non-issue with the on-demand service. Backups are another open question.

ISO

Arofan Gregory reported that this project was now starting to move forward with support from the World Bank. Olivier Dupriez (World Bank), Arofan Gregory (Open Data Foundation and Metadata Technology), Dan Gillman (U.S. Bureau of Labor Statistics), and Stuart Feder, who helped the SDMX standard attain ISO certification, had been meeting about the best way to approach ISO.

Previously it was thought that the most expeditious route for DDI to become an ISO standard would be to follow the path of SDMX by applying to Technical Committee 154 (ISO/TC 154 Processes, data elements and documents in commerce, industry and administration). Now an alternative has been suggested -- ISO/TC 46 Information and documentation.

There is now a two-page summary of the DDI effort which is required to get the process started. A key next step is to create the standard ISO 25-page document that is in effect a rendering of the specification itself in document format. This document serves as the official representation of the specification and links off to the DDI standards. Technical details are not part of the process. This approach means that both DDI Codebook and DDI Lifecycle would be covered under the ISO imprimatur.

The point was made that there is a marketing aspect to having ISO certification also.

Upcoming Meeting on Documenting Data Transformations

A meeting was to be held the following week in Ann Arbor to work on documenting data transformations in DDI. This work was seen as preparatory to developing a tool to update DDI XML metadata when the associated data change.

A new nonproprietary Validation and Transformation Language has been developed by the SDMX technical working group. The proposed tool could use such a language, but there are other options and we should be able to sub in any non-proprietary language.

It was noted that SDMX is looking for feedback on VTL.

Training Status

During the past year the EB took a step back to determine what the training agenda of the Alliance should be, the kinds of principles we should adhere to, and the types of training we should support or encourage. There is still a need for high-level "What is DDI?" type materials suitable for managers.

GESIS continues to support the Dagstuhl DDI training workshops, and Joachim Wackerow had proposed a "train the trainer" opportunity to enable a new trainer to be part of these workshops. The EB approved the proposal and there had been quite a bit of interest in it.

The idea was that the training capacity should be improved. While a core group had been doing the training for several years, DDI was getting more popular and the demands for DDI training had increased. The goal was for an additional young trainer to meet with the other trainers in Mannheim and to work together with the trainers in the Dagstuhl workshop. After this intensive learning experience, the person should be able to provide at least a two-day training on his or her own.

Proposals were due June 22, 2014. The person selected would be supported in terms of travel, accommodation, and food, and would receive a small stipend.

There is also a need to form a Training Working Group. Some participants at the meeting volunteered, but more people are needed to help scope out the training area. It was pointed out that the 2011 Alliance minutes laid out the training issues and launched the discussion. There is still not enough introductory DDI training at low cost for universities.

DDI workshops generate a massive set of slides, which are not yet officially published but are available under a CC license. New training can leverage these slides, but the commentary going with each is also important.

We should look at doing some online Web-based training, perhaps a Webinar series.

A day-long DDI 3.2 training will be conducted in advance of the next EDDI conference, which will take place in December 2014 in London.

DNS Work Update

Joachim Wackerow reported that he had been working on the issue of resolving DDI URNs, which involves communication with the Domain Name System (DNS), a hierarchical distributed naming system for computers, services, or any resource connected to the Internet or a private network. The Alliance needs to interface with DNS to resolve DDI URNs now that 3.2 has been published and the identification mechanism is stable. Achim will be developing a Request for Comment document for the Internet Assigned Numbers Authority *(*IANA*)* and then will apply for the DDI URN and DNS entry, which will help to resolve DDI URNs in the future. URNs will be resolved in a two-stage process: the DDI agency ID gets resolved by DNS and the requesting software receives an answer as to what kinds of DDI services are available; then the local system is queried and URNs are resolved.

RDF Vocabularies

The Alliance has begun the public technical review for the three RDF vocabularies developed over the last few years -- Disco, Simple Data Description, and XKOS – and people are encouraged to comment. The idea is to capture feedback over the next month, make possible refinements during the summer, and then hold a final public review in the fall, after which there will be a vote to publish the vocabularies. This process should be complete by the end of the calendar year.

After this informational item, the Meeting of the Scientific Board was adjourned.