DDI Alliance Expert Committee Meeting
June 4, 2012
Washington, DC

Minutes

Member Participants

Nikos Askitas (Institute for the Study of Labor – IZA)
Sonia Barbosa (Harvard University, Institute for Quantitative Social Science)
Ingo Barkow (Institute for International Education Research -- DIPF)
Olivier Dupriez (World Bank Development Data Group)
Michelle Edwards (University of Guelph)
Dan Gillman (US Bureau of Labor Statistics)
Arofan Gregory (Metadata Technology)
Sue Ellen Hansen (University of Michigan, Survey Research Operations)
Jani Hautamaki (Finnish Social Science Data Archive – FSD)
Marcel Hebing (German Socio-Economic Panel Study -- SOEP)
Chuck Humphrey (University of Alberta), Chair
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), Vice Chair
Vigdis Kvalheim (Norwegian Social Science Data Service -- NSD)
Amber Leahey (University of Toronto)
Steve McEachern (Australian Data Archive -- ADA)
Katherine McNeill (Massachusetts Institute of Technology – MIT)
Ron Nakao (Stanford University Libraries)
Jonathan Palmer (Australian Bureau of Statistics)
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)
Sam Spencer (Australian Bureau of Statistics)
Jon Stiles (University of California, Berkeley -- UC DATA)
Wendy Thomas (University of Minnesota, Minnesota Population Center)
Mary Vardigan (Inter-university Consortium for Political and Social Research -- ICPSR)
Joachim Wackerow (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)

Observers

Thomas Bosch (GESIS – Leibniz Institute for the Social Sciences)
Susan Mowers (University of Ottawa)

Introductions

After introductions, Chair Chuck Humphrey began the meeting by drawing participants' attention to the two parts of the meeting: the morning would be devoted to administrative matters while the afternoon would focus on substantive and technical issues. This is in line with the new Alliance structure put forth in the revised Bylaws.

Revised Charter and Bylaws, Member Form

The report of the External Review conducted in 2011 recommended that the DDI Alliance revamp its Bylaws to set up a new organizational structure more suited to a mature organization. To that end, a Governance Task Force, led by Ron Nakao, was constituted last year to work on the new Bylaws. The Task Force issued a penultimate draft for the Committee's review.

An important feature of the revised Charter and Bylaws was a separation between the Member Representatives who meet annually to discuss Alliance business and administrative matters and the Board of Experts, whose responsibilities lie in shaping the substantive and technical aspects of the specifications. The new Bylaws also have an Executive Board, elected by the members, as a successor to the Steering Committee currently in place.

Questions were posed about whether the host institution is hard-wired into the Bylaws. The Charter sets up the condition for a Host Institution but does not name a specific organization. The Bylaws describe the current Host as the University of Michigan. It was pointed out that if the host organization were to change, that would constitute a big enough change that a change in the Bylaws would be appropriate.

Another question dealt with why the terms of the Chair and Vice Chair are term-limited. This was intentional, to ensure that opportunities for leadership were spread across the community.

It was suggested that an organizational diagram would be helpful since there are now more official Alliance bodies described (Executive Board, Member Representatives in Annual Meeting, Board of Experts, and the Technical Committee). This was seen as a good suggestion.

The Committee was in general agreement with the direction of the revised Charter and Bylaws. A calendar for approving them was outlined:

- June through July 1, 2012: Governance Task Force cleans up text, makes it all consistent, and creates a version for review
- July 1-October 1: Members review and vet with their institutions if desired
- October 1-Mid-December: Steering Committee reviews and produces final version for vote
- Mid-December: Vote

The ABS representative volunteered to canvass prospective DDI members in the NSI community to see what they want to get from membership.

This timetable was to go to the Steering Committee the next day as a recommendation.

Tiered Membership and Membership Incentives

The review recommended that the DDI Alliance investigate a tiered membership model to generate additional revenues. Two types of tiered membership models were discussed: a model based on commitment to DDI and a model based on number of employees in an organization. The committee voiced strong approval for the commitment model and for tiers that provided levels of influence through number of votes. This model would involve members self-selecting into categories based on their level of support for DDI and their desired influence in voting. Factors discussed included:

- It would be ideal to have smaller organizations and even individuals as members paying lower fees. This would make it more of a long tail organization. We should investigate OASIS, which does have the possibility for individual memberships.
- We need to clearly delineate the membership categories and what one gets for joining in each category.
- We need to guard against bloc voting.
- It was pointed out that as currently described, the commitment model has a North American orientation and it should be more inclusive.
- We should add a category for software vendors.
- We also need a system of sponsors so that organizations can pay more to move the organization forward.
- We don’t have to increase fees all at once but can do so gradually, and we should incorporate an inflation factor into fees.
- We should discuss whether organizations providing high levels of in-kind support and contributions may be compensated through reduced fees.
- We should investigate the Blaise model with different levels.

The Chair conducted an informal poll that suggested that most members could afford to pay more (e.g., 20 percent more) if fees were increased gradually. A show of hands also showed that nine members thought their organizations could pay double the current fees.

The Expert Committee asked that the commitment model be taken to the Steering Committee as a recommendation. The model will be written up and sent out for review with the revised Bylaws.

Role of NSIs in DDI

Jonathan Palmer discussed the current perspectives of national statistical institutes (NSIs), which are increasingly under pressure from a variety of factors, including declining budgets and the data deluge. To remain relevant, NSIs need to standardize data production to produce data more efficiently and to do so they need standards to support the work.

ABS is leading the effort to produce the GSIM reference model. This model has been embraced as an international standard with several NSIs lining up behind it. If NSIs build a common platform, they want it to be viable and stable, so standards play an important role. A number of NSIs are collaborating, including the Scandinavian NSIs, Statistics Canada, and ONS in the UK.

Timing of the 2013 DDI Meeting

It was decided that the 2013 DDI Alliance meeting should occur on Monday, May 27, just before the annual IASSIST meeting in Cologne, Germany (May 28-31).

If all goes according to plan, July 1, 2013, will mark the beginning of the newly structured Alliance (this will be the beginning of the 2014 fiscal year). We will be introducing the new format of the General Assembly for members and a separate meeting of the Board of Experts. Nominations for the Executive Board may also take place at the May 2013 meeting, with voting to follow.

The terms of the current Chair and Vice Chair of the Alliance Expert Committee technically end on January 1, 2013. However, the Alliance members present at the meeting expressed a desire that the Chair and Vice Chair continue in these roles through the transition period to shepherd in the newly formed Alliance. Thus, their terms will end July 1, 2013. There was a motion to this effect, which was seconded and passed unanimously.

Financial Position

The budget shows that the Alliance will most likely end the fiscal year (the fiscal year ends on June 30) with expenditures exceeding revenues for the year by about $20,000 to $25,000. This will mean dipping into reserves. The Alliance holds approximately $75,000 in reserves.

The forecasted 2013 budget reflects new priorities detailed in the meeting agenda, including costs to hire a UML modeler, ISO certification costs, Collective Mark costs to protect IP, and more foreign travel, especially to progress the alignment of DDI with SDMX and the development of GSIM. It was pointed out that we should add the cost of completing the DDI URN resolution work; this will cost approximately $2000. Revenues will increase slightly as the Alliance has two members joining on July 1, 2012, so revenues will increase to $85,000 (34 times $2500).

The FY2013 forecast also shows expenditures exceeding revenues by approximately $20,000, pointing up the importance of increasing Alliance revenues for financial health of the organization.

IPR Protection

Members questioned the wording in the agenda document describing the purpose of the Collective Mark (“The Collective Mark would indicate to the world that the Alliance has the exclusive right to develop, control, and distribute the various DDI products [and no one else has this right]”) and advised that the word “distribute” should be removed because software vendors may legitimately distribute the schema in their software (either in the form of the schema itself or in another technical form) and this is protected by the GNU license. Having the distribution of the standard be free is very important, but modifying it and claiming to be the owner of it is not permitted.

It may be possible to use the new trademark logo to indicate that software is using DDI and is roughly compliant. “Powered by DDI” might be a possibility for a tag line. This would be a self-assertion of compliance because the Alliance cannot get into the business of certifying conformity – there would be too much overhead involved. In ISO terms conformity means using the standard in exactly the way one is told to use it. A subgroup of the Technical Implementation Committee (TIC) should hammer out how we interact with software vendors around DDI compliance.

DDI Mission and Guiding Principles

The group thought that this was an important document that should tie in some way to the Charter and Bylaws. Some wording changes were suggested:

- Add DDI’s importance for preservation in addition to access
- Change the wording about the domains DDI covers to reflect the wording in the Bylaws regarding data on human activity and other observation data
- Emphasize the goal of interoperability by moving it up further in the document, but express the term itself in more general terms
- Mention technology with language such as “leveraging today’s technology”
- Remove the specific reference to XML schema since this is only one expression of DDI

Members should have a chance to review this document before it is published.

Status of DDI Lifecycle Version 3.2

Wendy Thomas, Chair of the Technical Implementation Committee (TIC), gave an update on the next version of DDI Lifecycle, Version 3.2. After a lot of feedback from the community about bugs, changes have been made to improve the specification, and it should be ready for a Technical Review by the Expert Committee soon, with a version for public review ready in July. The TIC will provide information to facilitate informed review of the changes.

The bug tracking system Mantis is open to the public so anyone can look at the status of the development version.

Identification Issue

Last year the TIC had discussions about the optimal way to approach identification in the DDI specification, and two approaches were put forward. The Alliance consulted outside expertise about the two options, but the feedback was mixed.

Consequently, the TIC decided to take a compromise approach. The new canonical format of the URN includes only the basic objects -- urn:ddi, agency, identifier, and version number. The deprecated version includes all of the objects of the URN in 3.1 except for the version date of the maintainable. In the canonical format, the uniqueness of the identifier can be set to the Maintainable level (as in earlier versions of DDI) or to the Agency level. This compromise will be built into Version 3.2 so that users have both options. It is advised, however, that new implementations should use the newer identification mechanism as the older one is deprecated. While Version 3.2 of DDI will support both systems, that will probably not be the case going forward. In addition, DDI 3.2 distinguishes between payload metadata (like variable labels) and administrative metadata (like UserID). Changes in the administrative metadata would not result in a need to change the version number of the object. A roundtrip transformation between both 3.2 identifier types (canonical and deprecated) will be possible.

Resolving the identification issue means that the application to the Internet Engineering Task Force (IETF) can proceed based on the new ID types. This work will define the DDI URN for resolution purposes.

Progressing the Standards: What Comes Next?

DDI 3.2 has reached a point where it is restricted by the use of XML schema as the canonical and development structure. The TIC advises that the Alliance begin a new model-driven approach, allowing for a smoother integration of new content areas. This will permit easier development and easier expression of the model in languages like XML, RDF, etc. Goals of the new version include:

- Complete data life cycle coverage
- Broadened focus for new research domains
- Robust and persistent data model (for the metadata), extension possibilities, implementation for different technical domains
- Simpler specification that is easier to understand and use including better documentation

It is proposed that a workshop be held at Schloss Dagstuhl, Wadern, Germany, October 22-26, 2012, to focus on gathering requirements for and modeling this new DDI version.

Some members raised the issue that migration to this new version may be painful, but there will be a migration path. It was also stressed that having a UML model will signal a strengthening of the standards.

The Dagstuhl workshop will be the process for developing the model but a full model will not come out of it. This is likely to be a multi-year project. We should aim for not only the model but at least one expression of it.

There are several new substantive and technical areas that we want to integrate:

- Abstraction of data capture/collection/source. The current data collection module is questionnaire-centric. We should also be able to describe register data and data in the natural and health sciences (i.e., from technical devices or from laboratory analyses). There would be an abstract layer for data collection with the possibilities for “plug-ins” to handle different types of data.
- New content on sampling, survey implementation, weighting, and paradata coming out of the Survey Design and Implementation Group
- New content developed by the Qualitative Working Group
- Framework for data and metadata quality

- Framework for access to data and metadata
- Process (work flow) description across the data life cycle, including support for automation and replication
- Integration with existing standards like GSBPM/GSIM, SDMX, CDISC, Triple-S
- Disclosure review and remediation -- documenting risk-level factors throughout the lifecycle related to confidentiality concerns and to specific outputs at various stages
- Data management planning
- Development of standard queries and/or interface specifications (such as REST) which are needed to allow for interoperable services based on the DDI standard information model

A question was raised regarding whether there will be a halt to new features during the process of building the model. The TIC said that they would not rule out a version 3.3 but it was probably unlikely.

The point was made that it is time for the DDI to take a quantum leap, but perhaps a third party should provide an assessment and market research to influence future directions. The response was that we will draw upon outside expertise for the model development, including experts in GSIM, environmental areas, medical data, etc. It would be difficult to find reviewers for the schema.

The TIC is optimistic that we can find good people for the workshop and that it will be successful, but it is not as clear what should happen after that. The TIC is small and cannot do everything it is charged with. We need to think about separating functions so that there is a UML modeler on the team as well as an implementer of the various renderings – XML, RDF, etc. This requires more people and probably more money. The point was made that this is not just a TIC problem but an Alliance problem. We will need to look outside the DDI community for help.

A question was raised regarding whether we are spreading ourselves too thin by taking on so much. The TIC responded that the current process of bug fixing on the Lifecycle specification as it exists now is not sustainable, and the model-driven approach is the best way to have a solid and reliable standard that is responsive to new audiences. We need to make DDI work well for data integration across domains, so we will need to change some of the very social science-specific language and terminology. It would also be a good goal to have the DDI be more accessible for individual researchers.

ISO Certification for DDI

A document summarizing conversations with Dan Gillman, Arofan Gregory, and Stuart Feder of the Bank of International Settlements about moving DDI into ISO was discussed.

The World Bank has successfully established DDI in many countries through its programs, and DDI is now used in over 100 statistical agencies in 67 countries. Some of the newer countries in which DDI is being adopted, e.g., Mexico and Russia, are interested in using standards with ISO backing, so this is a good time to move forward with this effort. The World Bank can support this process.

The advice of that group was that the best ISO committee for DDI to approach is the TC154, which deals with trade data and economic growth. This is the group that SDMX is working with and it will be very helpful if we can learn from the experience of SDMX and follow their lead once they have completed the ISO process. Currently SDMX is a Technical Specification but they are going for full international standard status. They have a Framework approach, which will permit them to continue to distribute successive versions of SDMX from the SDMX Web site. This would work well for DDI also.

To move forward with ISO, a model of the DDI specifications will be necessary as well as an alignment with ISO 11179. It was suggested that we might start with ISO certification for DDI Codebook but it was pointed out that the Framework approach will permit linking to all versions of the standard.

Getting ISO status can be a long process but if the World Bank could become an ISO Class A observer, we would be able to fast-track DDI to a Technical Specification. This would help to facilitate the process.

We need to have a better sense of what resources are involved and what the steps in the process are. Currently the Alliance has set aside $5000 for this activity, but we should gather more accurate cost estimates.

We need to determine whether the World Bank is already a Class A observer, and we also need a task group to take this forward. Several people volunteered to participate in a task force, including Olivier Dupriez, Arofan Gregory, Steve McEachern, Dan Gillman, and Amber Leahey.

Process for Changes to the Specifications

A document describing the process for changing the DDI specifications was provided. This document was based on text that had previously been part of the Bylaws but that is being removed from the revised Bylaws under the rationale that the change process document should exist separately so that the Bylaws will not need to be amended if the change process is altered. The new version of the change process shortens the time frames for approval of changes because this was seen as excessive in the past.

The document needs to be cleaned up with some additional detail added. Wendy Thomas, Mary Vardigan, Michelle Edwards, and Mari Kleemola volunteered to work on a new version and to put it out for review by the Committee.

Report on RDF/Linked Data Work

The Alliance has done a lot of work in this area, with a particular focus on data and metadata discovery. An RDF expression of a subset of DDI is in progress. Also, Dan Gillman and Franck Cotton are working on a vocabulary describing statistical metadata based on SKOS, the Simple Knowledge Organization System, which provides a model for expressing the basic structure and content of concept schemes such as thesauri, classification schemes, subject heading lists, and other similar types of controlled vocabulary. This work began at Dagstuhl in 2011, and there will be a second Dagstuhl workshop on this topic October 15-19.

Reports of Working Groups

Survey Design and Implementation (SDI). After this group completed its work on questionnaire development and sampling, they began to focus in two new areas, Paradata and Weighting, with subgroups devoted to each.

- Paradata – This group is focused on documenting paradata -- the process metadata generated in the course of data collection. Right now they are documenting paradata elements related to the data collection module; this type of information is often useful in assessing survey and data quality. The GSIM work is relevant here as well.
- Weighting -- The group will consider the various types of weights used in both demographic and economic surveys (e.g., the various types of probability/design and replicate weights) and the characteristics of their description and purpose that should be included in DDI-L. An initial task was to accumulate various case studies about weighting, discuss some of the algorithms used in producing weights during the various stages of the survey collection process (e.g., sampling, stratification, nonresponse), and define all of the required inputs, both informational and programmatic, to fully document these operations. The Household, Income and Labour Dynamics in Australia (HILDA) Survey has been particularly useful.

Developers Group. While not a formal Working Group, a group of developers has held two meetings and is about to hold a third in DC. The goal is to make sure that implementations of DDI are interoperable. In the Gothenburg meeting during EDDI, the group discussed issues such as:

- Demonstrations of tools and services using DDI
- DDI fragments
- DDI Tools Catalogue
- Availability of DDI examples
- DDI 3.2 and keeping up with TIC
- Agency Registry
- Single namespace
- Ideas for tools
- DDI and DataCite

It was pointed out that others might be interested in minutes from the meetings so these will be added to the DDI Web site.

Tools Catalog. The Tools Catalog facilitates the ability of developers to publish information about their tools and developing tools. There are tutorials and support infrastructure to do this.

Controlled Vocabularies Group. Several controlled vocabularies have been published and there will be several more made available for 3.2. There are tools for publishing vocabularies that anyone can use, but procedures are needed for when others want to submit new or revised vocabularies. The group is developing a versioning policy to cover this. Right now the copyright of the DDI Alliance is hard-coded into the tools, but others should be able to contribute new and revised vocabularies to the Alliance. The working group will send a draft of its versioning policy to the TIC.

Administrative Group. This group is chaired by David Schiller of the Institute for Employment Research (IAB) in Germany, which has data on everyone employed in Germany in terms of their wages, times of employment, etc. The data are not collected for scientific research but do have scientific value, which is typical of most administrative data. IAB would like to modify its administrative data and distribute them to the scientific community.

The group is aimed at determining whether DDI can describe administrative data effectively or whether it needs additional elements for this purpose. David is looking for people in the group who have good use cases for administrative data who would be willing to describe their data and documentation and compare them with DDI elements. He hopes to get the group together at EDDI in Bergen in December.

It was pointed out that the Banca d'Italia has some register data that has been described in SDMX and that will also be described in DDI so that the two standards and their relationships can be compared. This may be useful for the group. The group might also look at some of the secure data being presented at the IASSIST conference. Describing register data is a major focus of GSIM.

Documentation Working Group. This group has been improving the documentation of the DDI standards.

Qualitative Working Group. This group met at EDDI in Gothenburg with 25 people in attendance. The plan is for a smaller group to meet in Bergen at EDDI 2012.

Web Site Working Group. Sam Spencer added links to the DDI schema and documentation to the front page of the DDI site and refined the map showing where DDI is being used around the world. He will also be looking into making the DDI site content more dynamic, perhaps through a blog, and will provide links to the DDI examples code repository.

Disclosure Risk Working Group. The point was made that this group should have a wider focus on sensitive data and not just those social science data that pose risk of disclosure. This group has yet to get started.

Experimental Data. A small group has been working on documenting experimental data in DDI and a question was raised about whether they should form a working group. Kate McNeill will put a description together for others to review.

Report on EDDI 2011, Plans for EDDI 2012, and NADDI 2013

The 2011 European Users Group (EDDI 2011), held in Gothenburg, Sweden, last December, was a big success with over 80 people from 17 countries and 3 international organizations in attendance (100 people in all including side meetings).The fourth EDDI will take place December 3-4, 2012, in Bergen, Norway. There is a call out to host EDDI 2013.

The first ever North American Users Group meeting will take place April 2-3, 2013, at the University of Kansas. Larry Hoyle is organizing the conference and is looking for a keynote speaker as well as volunteers to work on the program.

Move to JIRA/Confluence for DDI Specification Development

The TIC is considering a move to this system for bug tracking and documentation. This system has been adopted widely and can add transparency to the development process. Developers can volunteer to provide examples as can other users.

Because DDI is freely available, we may be able to get a free license. There will still be costs involved in getting a server, configuring and installing the software, populating the system, and maintaining it. The UKDA has volunteered to maintain a JIRA instance so we should talk with them further about what is needed for the other tasks. The DDI Secretariat will coordinate this work.