DDI Alliance Expert Committee Meeting Minutes
Tampere, Finland
May 25, 2009

Participants

- Iris Alfredsson (Swedish National Data Service -- SND)
- Atle Alvheim (Norwegian Social Science Data Service -- NSD)
- Nikos Askitas (Institute for the Study of Labor, IZA, Germany)
- Arofan Gregory (Open Data Foundation)
- Sue Ellen Hansen (Survey Research Operations, University of Michigan)
- Sanda Ionescu (Inter-university Consortium for Political and Social Research -- ICPSR)
- Taina Jääskeläinen (Finnish Data Archive – FSD)
- Jannik Jensen (Danish Data Archive -- DDA)
- Uwe Jensen (GESIS)
- Stefan Kramer (Yale University)
- Fredy Kuhn (Swiss Foundation for Research in Social Sciences -- FORS)
- Hans Jørgen Marker (Chair, Danish Data Archive -- DDA)
- Ken Miller (United Kingdom Data Archive -- UKDA)
- Ron Nakao (Vice Chair, Stanford University)
- Tom Piazza (University of California, Berkeley)
- Pilar Rey del Castillo (Centro De Investigaciones Sociologicas--CIS, Spain)
- Jon Stiles (University of California, Berkeley)
- Wendy Thomas (University of Minnesota)
- Mary Vardigan (Inter-university Consortium for Political and Social Research -- ICPSR)
- Joachim Wackerow (GESIS)
- Wolfgang Zenk-Moeltgen (GESIS)
- Toby Burrows (member as of July 1) (Australian Social Science Data Archive -- ASSDA)
- Steve McEachern (member as of July 1) (Australian Social Science Data Archive -- ASSDA)
- Gavan McCarthy (member as of July 1) (Australian Social Science Data Archive -- ASSDA)
- Dan Gillman (observer) (US Bureau of Labor Statistics)

Introductions

Chair Hans Jørgen Marker called the meeting to order and introductions were made. Several new people were in attendance at the meeting.

Award for Ken Miller

The DDI Alliance awarded a plaque to Ken Miller of the UK Data Archive to honor his excellent contributions to the DDI project over the years. Ken participated in the TIC group and also led the effort to develop controlled vocabularies. He will be missed.

DDI Tools Status (Mary Vardigan)

An update was given on the status of the Foundation Tools Program and the status of tools building in general.

Foundation Tools Program

To recap, a group of stakeholders came together in 2007 at IASSIST in Montreal to build tools, and $30K was contributed by the group, with other in-kind contributions pledged. The following resources, which added to $20K, have been built:

- XML components for open source tools
- Validation and resolution tools
- Tools architecture
- Tools Web site (tools.ddialliance.org)

Last year at IASSIST in Stanford, more stakeholders joined the effort in order to build a DDI editing suite of tools, to which the Alliance proposed to contribute $40K. An RFP was subsequently issued through the University of Michigan, but no viable responses were received.

The Canadian RDC is building editing tools, so it is now thought that we do not need to build these tools ourselves. Instead we should reinvigorate the Foundation Tools project to spend out the $10K remaining from the initial amount. Three projects have been suggested:

- Agency registration service
- URN resolution service
- Coordination/communication of software development so that we know what is happening

The first two items were prioritized as important to move tool development forward.

Tools in General

Because there are many DDI 2 instances in existence, a converter to DDI 3 is needed in order to capitalize on the new features of DDI 3. We may want to create a Task Force with wide representation to develop requirements so that it works for all.

User needs for tools are also informed by the various DDI projects going on in different groups. Not all users will be ready to move to DDI 3 at the same time. We need to put forward a well-documented business case for adopting DDI 3, supported by solid use cases that will demonstrate its advantages. Data producers in particular need to become aware of its increased potential.

We also want to emphasize DDI's complementarity with other standards and the possibility of integrating it in a larger metadata framework.

NSF and NIH, which are funding most US research projects, also need to be convinced of the value of using DDI, but the efforts of promoting the standard have to be supported by tools and expert assistance that will facilitate the adoption of DDI.

On the other hand, we need to make it clear that we will continue to support DDI 2 as long as it has a significant number of users. We may also want to consider a phased approach for moving to DDI 3, in which parts of DDI 3 may be used to expand DDI 2 instances.

Formation of Alliance for Data Archive Technologies (Mary Vardigan)

This alliance met for the first time in October 2008 in Washington, DC, and will meet again in Australia in July. The Australian Data Archive (ASSDA), ICPSR, and the UK Data Archive are spearheading this initiative, with the objective of developing a modular, open-source, and community-developed technology system for data preservation and access. There is a focus on standards, including OAIS, Fedora, and DDI.

DDI XML Consultant Status (Mary Vardigan)

The Alliance issued an RFP for XML expertise in April and two responses were received. The Steering Committee selected Arofan Gregory of Metadata Technology to continue to provide XML consulting.

DDI User Support Specialist Position (Mary Vardigan)

A new position has been created on a trial basis to offer support to implementers of DDI. Wendy Thomas, University of Minnesota, had been doing this informally and now will be paid for six months. She will continue to help implementers, assist in documenting the data model, and write best practices.

Outreach to Statistical Agencies (Mary Vardigan)

During the year, the Alliance sent invitations to several national statistical agencies to join DDI as Associate members with voice but no vote, but only one accepted so far (Bureau of Labor Statistics). We need to make a bigger push to penetrate this community. Mary and Wendy will be attending the International Statistical Institute meeting in Durban, South Africa, in August with the hope of spreading the word about DDI and will be on the program with a presentation about exporting DDI from computer-assisted interview systems.

CESSDA Project (Ken Miller)

The CESSDA Preparatory Phase Project (PPP) is an on-going two-year project that explores ways to develop the organization's activity and capacity. One major goal is for CESSDA to become a legal entity; another goal is to create a unique authentication method (using Shibboleth) that would be applied at the portal and would allow users to access all of the member archives without having to sign in separately for each one.

The common portal is now built on Nesstar and DDI 2, but the PPP is also exploring possibilities of moving to DDI 3, or at least using parts of it that provide added functionality. Comparison is an area that appears very promising as comparative projects are heavily used among members, both across time and geographies.

CESSDA's controlled vocabularies will also map to DDI 3, as representatives from the two groups have worked together to create common lists. A controlled vocabulary will also be needed for the Shibboleth authentication.

CESSDA is also looking into the possibility of developing applications like metadata registries, primarily a question bank, which may make use of DDI 3.

DDI and METS (Arofan Gregory)

Mary Vardigan and Arofan Gregory attended a workshop in Raleigh in May on Developing Integrative Practices: METS (Metadata Exchange and Transmission Standard), DDI, TEI (Text Encoding Initiative), and EAD (Encoded Archival Description). This is an effort to determine how the standards can work together. An outcome of the meeting was a plan to seek funding to create a resource that will describe the range of metadata standards available, what their specific purposes are, and how they should be integrated.

Specification Changes -- DDI Version Differentiation, Maintenance, and Development

See report of the Technical Implementation Committee (TIC).

DDI Financial Position -- How Best to Allocate Resources (Mary Vardigan)

Currently, the Alliance has a favorable fund balance with about $100K in reserve. The Alliance currently pays for:

- Staff time and benefits
- Some travel (METS meeting, ISI meeting, Canadian RDC workshop)
- Training (e.g., Dagstuhl event)
- Meetings/hosting
- XML consultant
- User support specialist

Suggestions were elicited on how to use our money to best effect. Support for developing tools and more DDI presentations emerged as the best areas in which to invest more money. Presentations would increase awareness of DDI's potential, but we should also have a small application, like a pilot question bank, to show how it can be implemented.

As far as revenues are concerned, no significant changes in membership are currently anticipated. Australia is joining, and GESIS and Harvard have merged their previous two memberships into one. Whether dues are set at the right level is a question that we will have to revisit with the Steering Committee.

DDI Evaluation

Most effective organizations undergo periodic evaluations. The group talked briefly about how we might structure an evaluation for the Alliance and what we should evaluate – e.g., the specification itself, the

development process, the governance structure, the extent to which we have met our goals, or all of these things. The Steering Committee was to take up this issue the next day.

Upcoming DDI Training

There will be another week-long workshop at the Leibniz Institute – Schloss Dagstuhl in Wadern, Germany, in October, organized by GESIS, followed by a more advanced forum in which people who actually work on producing use cases of DDI applications (metadata creation and use) will meet to get advice, exchange ideas, and plan for papers that would report on their activity. Instructors for both workshops are Arofan Gregory (Open Data Foundation), Wendy Thomas (Minnesota Population Center, University of Minnesota) and Joachim Wackerow (GESIS).

The ICPSR Summer Program has organized a short course on DDI, to be held at Cornell University in Ithaca, New York, on July 13-16. The course is titled "Documenting Data Using DDI 3.0: Supporting Research, Collection Management, and Access," and instructors are Wendy Thomas (Minnesota Population Center, University of Minnesota) and Arofan Gregory (Open Data Foundation).

Arofan Gregory will hold a training session in June, targeting European statistical agencies.

European DDI Users Group (EDDI)

EDDI 2009 will take place December 4, 2009, at the IDSC of IZA in Bonn, Germany. The day before on December 3, the DDI Alliance will sponsor a half-day DDI workshop on "Putting DDI to Work for You!" This meeting is being arranged by Nikos Askitas (IZA) and Joachim Wackerow (GESIS). Instructor for the workshop is Wendy Thomas, from the University of Minnesota.

Best Practices Event

A summary of the Best Practices workshop that was held at Dagstuhl in November 2008 was presented. A diversity of topics was covered, and twelve draft Best Practices were produced, currently in the final editing stage. They are accessible on the Web, but still hidden from public view. The Expert Committee will review them in June and they will be published when approved.

DDI Conflict of Interest Policy

The Steering Committee has suggested adopting a conflict of interest policy because it is a common practice in standards development for transparency purposes. A draft policy was modeled on the W3C conflict of interest form. It attempts to address how the Alliance makes disclosures and where and how we make this information available. The draft will be reviewed and its wording carefully analyzed to make sure that it does not generate legal problems for Alliance members.

Legal Status of Alliance

The DDI Alliance currently holds the copyright on the specification. Acquiring legal status would protect its ownership rights. However, it is complicated and expensive to become a legal entity.

A viable alternative would be to amend the section in the Bylaws that refers to intellectual property and specify that the copyright is jointly owned by all current members, and the rights are lost at the time any one member withdraws from the Alliance.

Inclusion in a formal standards body (like ISO) would also protect DDI, but that is also a lengthy and relatively difficult process.

Working Groups

Technical Implementation Committee (Wendy Thomas)

Version 3.1 and Future Changes
DDI Version 3.1 was approved prior to the Expert Committee meeting. It was noted that there will be a slight delay in its publication due to the need for some corrections that arose during the review period. Changes to be made include: a consistent ordering of name, label, and description throughout the specification; correct ordering of CharacterSet and ArrayBase in BaseRecordStructure substitutions; fixing the URN structure (which was found to be illegal); introduction of xml:lang attribute in all maintainables; and the possibility of specifying the preferred language in r:Reference. When creating the new version, TIC has tried to keep backward incompatibility at a minimum.

A note was made about numbering versions: a change in the version number to the left of the decimal denotes new functionalities and a new structure; a change in the first decimal implies minor changes and backward incompatibility; finally, a change in the second decimal indicates minor changes and a backward compatible version.

In order to minimize the impact of changes that are backward incompatible, any new version containing such changes will be accompanied by documentation of all changes in spreadsheet format as well as instructions for upgrading instances to the new version.

For future versions, the production process will also include an internal TIC review period, which will allow producing cleaner versions. It would also be helpful if the specification upgrade and review processes were given more visibility on the DDI Alliance Web site.

TIC Agenda
Upcoming topics/issues on TIC's agenda are:

- A complete review of code/command specifications/formal language in order to ensure machine-actionability (the language needs to be software neutral)
- Introducing two new record layouts: xml-, and spreadsheet-type
- A review of xs:string and r:CodeValue usage -- Some fields that might qualify for controlled vocabularies are currently defined as xs:string and should be changes to CodeValueType
- A review of the links between the physical and logical descriptions of data for various data storage options (this is necessary for the creation of on-the-fly datasets)
- Enabling an assessment of the quality of the metadata (does it provide information about the research standards used, quality indicators and quality control of the data collection process,

are questions/concepts being provided or not, the producer's efforts to assess data quality, etc.?)

- Adding new features to the specification:
  - Better documentation of scales
  - Coverage of qualitative data
  - Features needed for creation and deployment of registries
  - Mapping to other standards
  - URN resolution
  - Support for the genericode approach – how to structure and name external genericode files in a standardized way so that applications know what to expect when parsing them (genericode does not allow creation of hierarchies, but a repetition of the upper level before each lower level can be used to mimic them)

DDI 2 Maintenance and Support and Migration from DDI 2 to DDI 3

Currently there are a large number of DDI 2 users and instances and no solid production tools for DDI 3. Many DDI 2 users would benefit from a phased transition to DDI 3, which would involve adding some features to DDI 2 to help migration. Any additions that we might agree to make would be present in DDI 3; however, in an enhanced DDI 2 they would be optional, and any new version of 2 would be backward compatible. User groups that would benefit from a transitional version of 2 should also contribute resources toward its enhancement.

Although concerns were voiced about the possibility that an enhanced DDI 2 would slow down the adoption of DDI 3 and the creation of DDI 3 tools, as well as the development of DDI 3 itself, it was generally agreed to give further consideration to the possibility of supporting a transitional DDI 2 within the specified limitations.

Usability and Outreach Working Group (Stefan Kramer)

The UOG has agreed on two-year rotating, staggering terms for its co-chairs, with Stefan ending his term this year and Katherine McNeill serving for another year. The UOG is therefore looking for a new co-chair, as well as recruiting new members.

Since the UOG has a wide range of goals, it might also be useful to consider creating several mini-groups so that each can focus on a particular issue.

In the past year the major UOG activity has been the Web site redesign, which has made good progress and will be finalized in the next few months. The new DDI Alliance Web site will use the Drupal content management system, with a new layout and the possibility for all members to add content.

In addition to launching the new Web site, other major goals for next year will be providing input to the tools project, primarily for the editing suite, as well as facilitating usability and outreach activities of other Alliance members.

Controlled Vocabularies Working Group (Ken Miller)

Throughout the past year, the group has held bi-weekly video-conferences and has finalized an impressive number of proposed controlled vocabularies for DDI 3 fields.

The list was presented to the meeting participants.

After review of the controlled vocabularies by the Expert Committee, the TIC will convert them to genericode, and they will be published as DDI code lists, each with its own version numbering. Before handing them over to TIC, the lists should be converted to a "clean" format, like a spreadsheet, so that their conversion to genericode can be automated.

A suggestion was made and approved that each entry should be represented by a value, which would be language-independent, to facilitate the use of multiple language versions.

Taina Jääskeläinen will be the new chair of this working group, as Ken Miller is preparing to retire.

Survey Design and Implementation Working Group (Dan Gillman)

In the past year, this group has identified two major topics that it needed attention – sampling and questionnaire design – and has split into two subgroups to better address the issues.

The sampling subgroup has been working on a draft identifying sampling frames and schemes, and will also address data quality issues.

The questionnaire design subgroup has started by identifying the specific metadata that needs to be included in DDI – development process for the questionnaire design, translation issues, including pre-testing, etc.

In the future, this group will need to cooperate more closely with the Controlled Vocabularies group, as it appears that some of the topics overlap (i.e., sampling, data collection methods, etc.).

As DDI 3 is no longer centered on documenting survey-type studies, but attempts to cover a variety of data types, it was suggested that Study Design and Implementation might be a more appropriate name for this group.

Working Group on Preservation (Mary Vardigan for Nancy McGovern)

This working group's charge is to investigate issues around DDI and long-term preservation. These issues arise because DDI 3 emphasizes metadata reuse through a system of references, meaning that it is possible to reference content outside the DDI instance.

Specifically, the group proposes to look at preservation of DDI content at three levels:

- Community level: Because DDI is based on metadata reuse, content may be shared among organizations and may be distributed around the world, so it is important to ensure that these shared resources are preserved in some way.

- Organization level: Each organization needs to implement strategies for preserving DDI that it uses, taking into account the referencing of external information that may occur. For example, an organization might develop a preservation template for DDI and define rules for when the template will be applied.
- Object level: At this level, ICPSR needs to create a set of elements essential for preservation metadata for the objects it curates. It is hoped that this set of elements may become part of the DDI standard for others to use. ICPSR will share results with the community. The strategy for this last level is to review the PREMIS data dictionary and identify a subset of elements to be used.

So far, discussions within this group have centered primarily around identifying the actual entity(ies) that an archive or other organization might want/need to preserve.

There is an increased interest in creating registries, and consensus is sought on the set of objects/elements to be included.

The relation with the OAIS model is also being examined within this working group.

New Working Groups

Specific areas that need development and the proposed scope of the new group activities need to be further examined.

It was proposed that smaller, investigative groups be created to work with TIC in identifying current needs and proposed coverage; alternatively, to avoid duplication of work, this investigative effort may be carried out by a single, larger group that might then split up according to particular interests.

One area that should be examined more closely is Archive Management Information. Secondary Analysis has been considered, but probably does not merit special attention, since the analysis process is similar to primary analysis (which is covered appropriately). Work on qualitative data coverage will be resumed following a temporary slowdown.

The TIC is also looking for new members willing to contribute to the development of the specification.