

# Committee Meeting Minutes

October 7, 2002  
Washington, DC

**Committee Members Present:** Grant Blank, Bill Bradley, Cavan Capps, Pat Doyle, Dan Gillman, Ann Green, Bjorn Henrichsen, Peter Joftis, Marc Maynard, Ken Miller, Jostein Ryssevick, Merrill Shanks, Wendy Thomas, Mary Vardigan, Ron Wilson

## 1 Discussion of revised Charter and DDI Alliance

The Committee reviewed a revised version of the DDI Alliance Charter that reflects changes made over the summer since the June meeting in Storrs, CT, with input from Richard Rockwell, Myron Gutmann, and Bjorn Henrichsen. The IFDO membership and the CESSDA institutions have seen this version and have committed to being host associations. The next task is for the Committee to request officially through the Chair that ICPSR and Roper become the initial host institutions for the Alliance.

Timing of the transition from the Committee structure and NSF funding to the self-supporting Alliance structure is an important consideration. NSF funding ends February 28, 2003, and we don't want to have an extended period with no funding. There is funding for one more committee meeting before the grant expires.

The Steering Committee needs to meet to start up the Alliance, preferably in January 2003. Most prospective academic members will need membership forms by March to secure money for the membership fee in their budgets for the upcoming fiscal year.

Length of the DDI Alliance membership term was also discussed. There are many U.S. institutions and organizations, as well as government agencies, that cannot commit to a three-year term of membership, although this is less of a problem in Europe. It was decided to make the term of membership one year but to encourage a longer commitment. It was suggested that the information about length of membership term be removed from the Charter and instead inserted into the membership form signed by an institution. The actual membership fee will be moved to the form as well since it may change over time and we don't want to have to amend the Charter for this purpose.

Other points raised regarding the Charter included:

- 1 We should change the language involving "employees" of

member organizations to read "employees or designated representatives," since some organizations like IASSIST do not have paid employees.

- 2 An individual may only represent a single member organization, but this shouldn't be a problem.
  - 3 Observers may be invited to attend Expert Committee meetings by the Director.
  - 4 A representative who cannot attend a meeting may send an alternate but must inform the Director. This can be handled informally and doesn't need to be written into the Charter.
  - 5 The Expert Committee Chair should be one of the two delegates from the Expert Committee to the Steering Committee.
  - 6 The section of the Charter on procedures for changing the specification actually should refer to major changes only. The Steering Committee needs to develop a procedure for handling minor changes.
  - 7 We need a short sentence on Intellectual Property. Any product developed by the Alliance would hopefully be open source and in the public domain.
- 2 Also raised was the issue of version control for the DDI specification. We need more certainty about the goals for a completed version and the long-term direction of the standard. Many people are basing software on it now and need to know where it is headed.
- In terms of what next steps need to be taken, ICPSR as one of the host institutions needs to take the necessary actions to begin the process of setting up the Alliance Secretariat. A Memorandum of Agreement between the Alliance and ICPSR, or some other formal arrangement, may be necessary. We also need a membership form and need to begin soliciting members very soon, with the goal of a meeting of the Expert Committee in the spring.
- The decennial and corporate metadata repository (CMR) branches of the U.S. Census Bureau need to become more familiar with the DDI. It would be helpful to have some materials and a strategy for promoting membership, particularly in government statistical agencies like Census. To that end, a Working Group was formed,

consisting of Pat Doyle, Bill Bradley, Dan Gillman, Ernie Boyko, and Cavan Capps. It was pointed out that representatives from Nesstar recently met with the International Division of the Census Bureau and found that branch very receptive to the DDI concept.

- 3 ICPSR report on changes to DTD approved at last meeting  
ICPSR submitted a report detailing the changes made to the DTD since the meeting in Storrs. Staff at ICPSR were able to make the necessary revisions, but the Committee needs to continue to look for an XML expert.
- 4 ICPSR report on Census 2000 markup using the aggregate nCubes extension; other experience with nCubes  
ICPSR submitted a report describing the markup done so far with Wendy Thomas's help.  
Concern was expressed about the status of the ncubes aggregate specification since proposed revisions have not been acted upon and momentum has been lost. The aggregate extension is close to being finished but cannot yet be frozen because we need to do further work. We need to try to address time series and the logic of hierarchies as well.
- 5 Report of Geography Working Group  
Ron Wilson, Chair of the Geography Working Group, reported that the group met with the Center for Spatially Integrated Social Science (CSISS) project in Santa Barbara in August and began to formulate some proposals for additions to the DDI specification. Specifically, they propose the addition of a bounding box element with coordinates to specify the geographic region covered by the dataset. This comes from Geography Markup Language (GML). We also need an attribute at the variable level to indicate that a particular variable is a geographic variable. We need to set some goals and determine what type of functionality we want to provide for with the geographic identifiers and also which resources will be incorporated into the metadata and which will simply be linked to. Wendy Thomas also provided proposals to handle markup of Census files in which the geographic code identifiers are embedded in the data.  
The Geography Working Group will have a final report before the next meeting. In the meantime, it was proposed that the Aggregate and Geography Working Groups merge, to be chaired by Peter Joftis. ICPSR may have some funds left over after this meeting, which could be put toward a meeting of the combined Working

Groups.

## 6 Additional changes to DTD

### 1 For relational files

A handout describing the logic of complex files was distributed, along with a summary of the information needed to link records. The current DTD already contains much of what is needed, and there are only two additions necessary to render the DTD robust enough to document complex file structures. This proposal would also cover complex relationships among separate files. The Working Group will put together some sample markup before the next meeting to demonstrate what is needed to make this approach work.

### 2 Funding agency/grant number linkage

This item was deferred until Version 2 since the scheme we currently use works, albeit less elegantly than we would like.

### 3 New File Group element

The issue was raised that the DTD does not handle file grouping in a way that is parallel to how it handles variable and category grouping. The Committee advised that we don't want to start documenting files other than data files in Section 3.0. Rather, we need to document files of other types (SAS/SPSS data definition statements, PDF versions, etc.) in 5.0 Other Material and add an IDREF attribute there to link these other files to the relevant data file in Section 3.

### 4 Modification of ExtLink element

Regarding external links embedded in text and how to render them, some members of the Committee advised that they inserted character entities (& etc.) for angle brackets embedded in text sections as a workaround.

## 7 DDI data model

Some work has already been done by Nesstar and Health Canada on a DDI data model with extensions to ISO 11179, and further work will be undertaken in the next two months. The model as it exists now is not complete and only contains some of the DDI elements, but it could serve as a basis for a full model.

The data model is an important longer-term goal for Version 2. It is tedious work to build a data model but something we need as the core of the standard. Having a data model will make it easy to

move to RDF and Schema and other formats. By January we should have a model to work with, and at that point we will form a Working Group to carry out this activity.

8 External funding

Bjorn reported on funding for social science infrastructure in Europe. Right now the social sciences are well funded under the 5th EU Framework. Calls for the 6th Framework will go out soon.

9 Next meeting

The next meeting, to be held in Washington and arranged by the Roper Center, was tentatively set for Friday, January 17, 2003. As NSF funding terminates February 28, 2003, we need to work to complete this version of the DTD by then, as promised in the grant application.