DDI Committee Meeting Minutes

June 29, 2001
Washington, DC

Committee Members Present: M. Shanks, Chair, M. Altman, G. Blank, E. Boyko, B. Bradley, C. Capps, C. Dippo, D. Gillman, P. Granda, A. Green, P. Joftis, T. Piazza, R. Rockwell, J. Ryssevik, T. Staples, W. Thomas, M. Vardigan

Other Meeting Participants: Emiel Kaper, Statistics Netherlands; Steve Kolodrubetz and James Summe, Office of Strategic Planning at the Centers for Medicare and Medicaid Services; Chris Oster, Health Canada

Introduction

After welcoming the visitors to the meeting, M. Shanks raised the issue of how the DDI effort might be supported in the long term. Can we find a structure not based on grant funding? Is a membership structure appropriate? How can we become an international standard and find a "home" for the DDI?

He also indicated that the Committee will be undergoing a transition period soon, with a new Chair and possibly a new Committee composition.

Changes to Version 1.0

The Committee agreed upon a set of changes to Version 1.0 of the DTD. These changes, which are all noninvalidating, will be written up and incorporated into a draft DTD that will be available on the DDI Web site.

Aggregate/Tabular Data Specification

The Committee reviewed the "compromise" specification developed principally by Wendy Thomas and Emiel Kaper. It was pointed out that documentation should theoretically carry the instructions on how to generate tables, but perhaps not the data themselves. However, many archives have documentation that does include frequency tables the Eurobarometers are examples of this.

The nCube group at Statistics Netherlands are talking with SPSS, so it may be possible to initiate contact with SPSS about the DDI effort. If we form an Implementers Group, such a group could interface with several of the commercial software firms.

Details of the aggregate specification were presented:

- The specification generally describes a result set an n-dimensional matrix, or nCube.
- Each cell has a relationship to others in the data matrix structure.
- The developers extended the var (4.2) specification to document aggregate data.
- Each cell is described by coordinates.
- Logical structure is separated from physical storage. This permits data to be stored and presented in different ways.

It was pointed out that to process data, a mapping from the logical to the physical is necessary.

The Committee agreed to adopt the recommended changes in 3.3 and 4.0 as well as making 3.1 repeatable. Additional testing of the specification needs to happen, especially with respect to rectangular files with marginals, like the Eurobarometer. Several Committee members volunteered to participate in the betatest of the aggregate data model: W. Thomas (University of Minnesota), B. Bradley (Health Canada), D. Gillman (BLS), C. Capps (Census Bureau), J. Ryssevik (NESSTAR), and the California Counts group at Berkeley. This group will report back at the next meeting. It was suggested that ICPSR set up a betatest site for this testing similar to the site used for the formal betatest preceding publication of Version 1.0. Testers should also make use of the threaded codebook list to communicate on this topic.

Report on DDI/ISO 11179 Meeting

A meeting was held the previous day to discuss harmonization of the DDI with ISO 11179. In attendance were A. Green, J. Ryssevik, C. Oster, D. Gillman, B. Bradley, and P. Joftis. The group reviewed a mapping between 11179 and the Corporate Metadata Repository (CMR) extensions to the DDI.

D. Gillman suggested that there are a variety of paths and organizations that could facilitate the DDI s becoming an accredited standard:

- World Wide Web Consortium (W3C)
- Object Management Group (OMG)

- ISO-IEC
- IEEE
- ANSI

One option would be to pursue accreditation under the sponsorship of IASSIST/IFDO. Also, it is possible to seek rapid approval as a Publicly Available Standard. Peter Joftis will pursue this further.

DDI Proposal to NSF

The Committee was asked to communicate with R. Rockwell about any errors discovered or any significant items of high priority that should be added. The Committee should determine whether the proposal adequately covers what it seeks to do.

Letters of support are helpful and should be solicited. Further, additional citations could be incorporated to enhance the proposal.

Next Meeting

The Committee tentatively agreed to meet December 3 and 4, 2001 (Monday and Tuesday), in Washington, DC. Tentative items for discussion include:

- results of beta testing of aggregate data
- hierarchical files
- Version 2.0 structure
- ISO 11179 mapping review