DDI Codebook

Draft: 2021-02-15

# History of Codebook

Codebook is the original DDI product which was intended to be a single unified specification for describing social science research data. While initially published to cover a simple study, the intent was to continue development work to cover repeated studies, studies with multiple data files or storage types, and complexity. The work done by the DDI Alliance was started as a means of addressing the next steps in DDI development, and as such, simply a new version of DDI which covered additional data types and coverage information. However, accomplishing these goals required a reworking of the means of expressing DDI (at that time an XML DTD) and a revision of the basic design rule of backwards compatibility (see Design Rules).

With the publication of DDI Version 3.0, the first version breaking this rule and shifting to XML Schema as a means of expression, it was anticipated that users of Version 2.1 would shift to Version 3.0. It shortly became clear that features of the earlier versions -- use of XML ID and IDRef for internal identification, minimal infrastructure requirements, and availability of software for both DDI creation and management – meant that earlier versions met some needs that were not being addressed with the new model and that continuation, maintenance, and even development of the earlier model was required by this community of users.

Around 2019/2020 the decision was made to declare that DDI consisted of a suite of products including DDI Codebook and DDI Lifecycle. These two products need to be closely aligned to support ease of content transfer between the two products but that they address the needs of different user communities.

# Design Goals

* Continue to meet the need for primarily descriptive metadata for single data sets (with one or more related data files)
* Keep infrastructure requirements low particularly in terms of required content including identification
* Improve coverage of descriptive metadata to reflect new content coverage in Lifecycle and relationships with other DDI products (for example use of Controlled Vocabularies and support for SDTL content)
* Respond to user requests for maintenance and improvements to support the use of Codebook as content for data discovery and access systems in use within libraries and archives

# Design Rules

* All versions must be backward compatible, i.e. an instance created in an earlier version will parse against the current version when the namespace is changed to the newer version
* DDI-C addresses simple studies and does not cover grouping, comparison, or other broad areas of expansion found in the DDI-L product line. Additions may be made which enhance detail in its current coverage area
* Additions to DDI-C must be reflected in the content of DDI-L (either as a reflection of current DDI-L content or through filing a bug to add content to DDI-L in a future revision)
* DDI-C wants to continue the looser rules around identification, making the use of ID content optional except to support reference functions within an instance

## Appendix

Excerpts from Minutes of the DDI Committee, June 15, 2002– Storrs, VT

(Other agenda items: DDI Alliance, charter, membership issues, etc., Funding opportunities, including the October NIH competition, Working Group reports, Resolution of Weights Issue, Status of Tag Library, and Versioning Plan for DDI)

**Research for DDI 2: How to begin**

The Committee developed a list of goals for DDI 2, including:

* Modularity
* Time series
* Aggregate data
* Object -oriented model
* Collection level, “families of datasets”
* Interoperability
* Instructions for applications
* Best practices
* Support for multiple languages, if XML:lang doesn’t cover this adequately

**Plan for moving forward, Fall meeting date**

We need to move ahead with another version of the draft charter, which can be circulated. Richard is heading this effort. Also, the following changes to the specification were approved and need to be implemented:

* Create stub schema and readme, with default stylesheet
* Incorporate Aggregate Working Group Report items 1-4
* Incorporate new content model for nested files
* Incorporate TEI equivalents for formatting
* Add new content models for weights
* Add on/off switch for CALS table model