

Technical Committee Meeting 21-25 October 2019: Summary

Attendees: Wendy Thomas, Jon Johnson, Larry Hoyle, Jeremy Iverson, Dan Smith, Johan Fihn Marberg, Oliver Hopt, Dan Gillman (as available), Barry Radler (Monday-Tuesday), Flavio Rizzolo (remote, as available), Jay Greenfield (remote, as available)

Meeting Page:

<https://ddi-alliance.atlassian.net/wiki/spaces/DDI4/pages/723255303/In-person+Meeting+21-25+October+2019>

The Monday/Tuesday agenda was adjusted to fit the availability of remote members. Entry and review of DDI 3.3 changes were added to the agenda as well as review of outstanding TC issues.

Results of Meeting:

DDI Roadmap:

This discussion took place within the context of current and future user groups of DDI products, the mix of DDI products, marketing, production frameworks, and integration. The initial point of discussion was the meaning of a single “DDI”. Given the expansion of DDI products outside of the “DDI specification” as well as the technical constraints and content needs expressed by various user groups the goal of a single integrated DDI specification seemed difficult to attain. Instead, the idea of DDI as a suite of products which covered a specified topical and application area unified by a common conceptual model emerged. In this scenario each product would specify coverage of the conceptual model in terms of content and application support. Development efforts would be towards improving consistency across the products to support transfer of content from one product to another, clear mapping content between products, and a shift in the definition of the DDI products to focus on their content coverage and applied uses. From a marketing perspective it allows for promoting DDI as a whole while directing potential users to specific products in the suite that support specific activities and technical environments.

Notes from this discussion are found in the TC Minutes under 20191021 – 20191025

TASK: Complete draft of proposal for updated DDI Roadmap for member review through the Scientific Board

Production Framework

New TC Production Framework page listing information for DDI-Lifecycle and Controlled Vocabularies Pipeline

<https://ddi-alliance.atlassian.net/wiki/spaces/DDI4/pages/729284679/TC+Production+Framework>

COGS

COGS is a development and production environment that takes a “registry” approach to the management of DDI content. Identified-Versioned objects are intended to be managed by registries for the purpose of reuse. The content of DDI may be expressed in XML, RDF, XMI, JSON, as well as other commonly used expressions. Documentation and graphs of objects are also produced. By managing the content of these objects using CSV files in a GIT repository, a broad range of implementers and users can develop content, test it in an iterative manner, and submit it for inclusion in DDI. The use of COGS supports:

- Automated verification of content against design rules
- Automated production of multiple expressions of the DDI content to support various uses and applications
- Options for iterative testing of new content
- Options for increased involvement in DDI development by lowering the technical bar for participation

The applicability of COGS to the main DDI products was discussed. Initial usage was envisioned for DDI-Lifecycle. Discussions of the utility of COGS for DDI4 was focused on the production of documentation. There is a difference in perspective between the UML modeling approach and the COGS Registry approach. Specific issues were identified and these should be the basis for further discussion on the maintenance of DDI4 over time. One point affecting the XMI output of COGS (currently EA flavor of normative OMG UML 2.4.2 and 2.5 with diagrams) is that of the appearance of diagrams and their content. XMI output of COGS should meet the expectations and needs of the UML user community. The possibilities for the application of COGS to DDI-Codebook were discussed and will be examined over the coming year. The move of either the DDI-Lifecycle or DDI-Codebook specifications would result in a structurally different XML output. Some design rule changes such as dropping the use of global attributes or unions, and the substitution group approach for Physical Structure would drive these changes. The proposal is to take DDI-Lifecycle 3.3, move it to COGS, and produce a DDI-Lifecycle 3.4 for technical review so that developers can focus on the structural and design changes independent of content additions. Dan Smith will do a test load of DDI-L 3.3 when last few items are entered. A similar approach would be used to examine a change in the DDI-Codebook schema.

Documentation Production

Work was completed on resolving publication and production issues for the DocFlex production of HTML documentation for DDI-Lifecycle and DDI-Codebook object level metadata.

Production of the high-level documentation through COGS based on restructured text files is now functional and only requires additional content coverage prior to publication. This is an on-going project and was not part of the October meeting agenda. Object level documentation from COGS will be not be provided as part of the publication package for DDI-Lifecycle 3.3 as it reflects the design rules that will be reviewed by DDI-Lifecycle 3.4.

There was discussion of what documentation should be included in the official publication package and what should be supplemental. Separating high-level documentation as well as alternate distribution structures allows some flexibility for improving documentation between official publications.

Controlled Vocabulary Production – post CESSDA

Clarified the output content of the CESSDA CV management system. All outputs (HTML, SKOS, PDF) have multi-lingual content as required by DDI. The addition of a language selection bar on the HTML output allows for filtering by language. DDI would also like to provide a multi-lingual output of each CV in a common DDI-Lifecycle CodeList structure allowing users to point to a content structure that was already familiar to them. Oliver wrote the needed transformations and tools pipeline to produce a package of output for updating the ICPSR hosted site for Controlled Vocabularies. Use of this pipeline reduces the workload for updating based on new publications to updating of the page listing all current CV's and their download options and the page listing version histories. Currently CESSDA does not provide a

“push” option for initiating this process (it is planned for the future). A command line tool has been created to grab the CESSDA output, run it through a DDI production pipeline, and provide ICPSR staff with a zipped package for publication. TC is currently coordinating this work with the CV working group and Michael Iannaccona of ICPSR. See document on Controlled Vocabularies on the Meeting Page.

DDI Resolution

URL format for API resolution (resolving a DDI URN to an API format) – did not get to

TC-4 Informed that this is now progressing again – keep tracking

High Level Documentation Activities

Integration

If the approach of a DDI Suite of products is supported by the DDI Community this implies the need for a number of documents:

- Creation of a conceptual model underlying DDI – Flavio will explore this and we will add to future TC agenda
 - Conceptual level mapping of products to the overall model
 - Use to identify commonalities, points of similarity/dissimilarity
 - Use of this information for Marketing and Training to explain DDI overall and applications for specific products
- Explore implications for the design and development rules for each product in terms of coverage and technical implementation
 - How does this effect the addition of new content across products

Mapping

A first pass at mapping DDI4 Core content objects to DDI-Lifecycle 3.3 complex elements was started. A mapping page will be created to bring together this and other product-to-product mapping information. Progress was delayed until DDI4 specific (property and relationship level content) was solidified and the DDI-Lifecycle 3.3 updates were completed. Many of the mapping issues are found in the property and relationship details so that having stable content for mapping is a must. The approach for publication will start with a spreadsheet as an easy means for both visualization and processing of information. While many areas have relatively clear object to object mapping, changes regarding the logical and physical description of data are extensive and resulted in a many-to-many mapping at the object level. This section must be done property to property. This work must be continued post-meeting.

Best Practices Document

The Best Practices Document for DDI 3.2 and future versions was reviewed and updated. Dan Smith will enter updates and provide the PDF for version 1.1 of this document. This document advises on best practices to support future planned developments in DDI design rules. Additional items not added to the Best Practices Document will be placed in High-Level Documentation and/or be used in defining the product level support for various applications.

DDI 3.3 entry and review

DDLIFE issues which were filed as part of the DDI Lifecycle 3.3 review were processed.

Of the 53 issues identified during the review process:

No.	Status	Description
3	Open	creation of the change log, determination of package content and readme.txt update
3	In Process	creation of a validation test (not required for 3.3 publication); documentation of methodology section – post meeting
3	Resolved	Resolved during meeting requiring entry
3	Resolved	Corrections in examples requiring finalization of 3.3 content to complete – post meeting
19	Closed	Closed prior to October meeting
32	Closed	Entered into the 3.3 schema, reviewed and closed during meeting

TC Issues Update

At the beginning of the meeting there were 19 TC issues unresolved. These were reviewed and discussed as follows:

Issue No.	Topic	Status Action [notes]
TC-201	CV Production: decisions and implementation regarding versioning, production and usage	No change [See notes above for progress during meeting]
TC-157	This is what the issue on primitives is about...the appropriate UML primitives and the appropriate translation to bindings. I'll link the issues and transfer the information to make sure it is not lost.	No change
TC-4	ARPA registration of DDI URN	No change
TC-1	Define process for migration from Drupal to COGS - this has been pretty well determined by the EA-> Canonical XMI -> csv file -> COGS however this only addresses what was put into the Core - do we lose the rest or do we capture it in COGS in some way	Changed to In Progress [discussed where issues were in terms of content and approaches – not critical to DDI4 production for review]
TC-208	Generic agency value - add information to Best Practices document	Change to Closed
TC-207	Review use of common properties across versions	Changed to In Progress [generate from finalized 3.3, 2.5 and DDI4 to identify during mapping and integration process]
TC-205	Overall alignment across DDI versions	Changed to In Progress [Flavio is exploring higher level conceptual model of DDI to aid in this work as part of integration and mapping]
TC-204	Alignment of classification models across DDI versions and XKOS	Changed to In Progress [Priority item in mapping]
TC-203	High level feedback DDI4 (NDS) - level of interoperability between standards (line of DDI	Added cross-version label [interoperability work]

	development); how do the lines hold together?; how does DDI4 fold into continued development	
TC-189	Feedback from ICPSR staff committee on DDI4 prototype - interoperability across models; migration is a major issue; possible use of underlying conceptual model	No change [Need to tie into integration draft]
TC-167	Specification of default layout properties at the file level - how should these be recorded and relayed to developers/implementers?	Added labels [This will be part of developers information for DDILifecycle 3.3]
TC-78	Ensure that source target cardinalities are flipped when content is transferred to COGS	No change [per Oliver to ensure this is correct in both input and output]
TC-3	Recommended validity check for DDI4 - Valid realization of pattern (in relation to COGS requirements)	Changed to In Progress
TC-209	COGS interpretation of content and resulting XMI	Changed to In Progress [role of COGS created XMI see discussion under COGS]
TC-163	Document information - primarily the idea of serialized information, expressions that are transitory, expressions used for preservation, transfer, etc.	No change
TC-80	Incorporation of UML class relationship information to COGS	Changed to In Progress [role of COGS created XMI see discussion under COGS]
TC-55	Error: PSM for XSD has identical xmi:id - is this still a valid issue	Changed to Closed prior to meeting
TC-210	Face-to-Face meeting	No change [will close when draft documents from meeting are completed]