

# DDI Alliance Executive Board Meeting

## 20 February 2019

*Present: Bill Block, Cathy Fitch, Jared Lyle, Barry Radler, Joachim Wackerow, Wendy Thomas (invited guest)*

### DDI 4 Prototype Review – Status Report

Wendy Thomas, chair of the DDI Technical Committee, discussed the “DDI 4 Prototype Review: Status Report” (see Appendix 1). The prototype review period has been extended to the end of March. The review has been sent to DDI and RDF mailing lists. Targeted follow-up requests for review have been made to a handful of institutions.

Board members noted the limited number of people and institutions who have responded so far and noted that earlier DDI 4 reviews had a broader set of respondents, although they were much more focused and not so broad in scope.

The Board discussed the overall non-response. Some community members may not have an immediate need for DDI 4 and may not feel the need to provide feedback. The Board discussed how to get feedback from the overall community about what is needed in DDI 4.

A Board member asked whether it would be worth using Alliance funds to pay organizations for feedback, although other Board members indicated the Alliance does not want to start paying for feedback.

### MRT Sprint Proposal

The Board discussed a proposal for a new DDI 4 working group that will produce a core DDI 4 release available for use by the end of 2019 (see Appendix 2). The new working group is requesting approximately \$7,000 USD in funding to support a face-to-face planning meeting to be held in connection with NADDI in Ottawa in April (see Appendix 3).

It was noted that the Alliance has underspent so far this fiscal year on planned meetings and there are no plans for a sprint around IASSIST this year. It was also noted that a representative from the NSD - Norwegian Centre for Research Data emailed the Executive Board in support of the proposed NADDI sprint. The Board noted that NSD has not yet provided feedback for the DDI 4 prototype review. The Board requested Jared contact the NSD and request that they provide DDI 4 prototype feedback.

Concern was raised about the working group meeting prior to the completion of the DDI 4 prototype review, although it was countered that not approving the meeting could stop or slow the initiative. It was noted that while Alliance funding for travel and hosting is significant,

individuals' in-kind contributions are highly valuable and the Alliance should treat these resources carefully. It was also noted that the Executive Board has the responsibility of representing the entire community and carefully stewarding the use of limited resources, and that just because the Board does not fund one meeting does not mean it does not support DDI 4.

The similarities between the existing DDI 4 modeling group and the new MRT working group were noted, including questions about why a new group is needed. It was noted that the new MRT working group provides a more comprehensive perspective -- including modeling, representations (XML and RDF), and test instances. The other change is the focus on the core of DDI 4 and not on the many, many views. The MRT group is seeking to build a stable and consistent core onto which later (2020+) things can be added on without changing the DDI 4 core. The core will include conceptual (starting with a concept, but can also be a code list, etc.; different forms of variable cascade), data description (also variable cascade, but also physical description), and process (workflow process, data transformation description).

The deliverables of the MRT sprint proposal were discussed. Board members indicated a desire to see a stronger statement of deliverables.

The Board agreed to fund the proposed NADDI sprint but would like the MRT working group to incorporate feedback of the prototype review into the MRT NADDI sprint. The Board also requests the MRT working group submit back to the Executive Board concrete deliverables that will be produced by the end of the sprint.

## Action Items

- Jared will contact NSD to request DDI 4 feedback.
- Jared will notify the MRT working group that the NADDI sprint request was funded and that they need to: (1) incorporate prototype review feedback into the sprint, and (2) submit back to the Executive Board concrete deliverables that will be produced by the end of the sprint.

## Appendix 1

# DDI4 Prototype Review: Status Report

Wendy Thomas – 2019-02-19

## Actions:

Review period began: 2018-10-25

Initial period was to end 2019-02-04 (reminding sent out on 2019-01-16)

Review period was extended to 2019-03-31 as the request for review was not distributed to the RDF community as planned

## Results:

As of 2019-02-18 we received:

33 comments via the JIRA tracking system:

- 27 from Larry Hoyle,
- 4 from Flavio Rizzolo,
- 1 from Joachim Wackerow
- 1 from ICPSR

Prior to the public review 37 Issues were identified by the Technical Committee that were determined should be dealt with post-prototype review (either the issue was complex or required an amount of time which would have delayed the release of the Prototype for review). These issues were noted in the Technical Committee comments on the Prototype.

The TC has performed only a single triage on the filed issues just prior to the EDDI Sprint. This covered the first 22 issues filed and resulted in 7 review issues plus 20 post-prototype issues to the Modeling Team. These dealt primarily with broader modeling issues.

In mid-January the Advisory Project Management Group discussed the response rate and we were informed that the announcement to the semantic web lists was delayed. The decision was made to extend the review period to the end of March and to follow-up with individuals and organizations who had previously provided feed-back on earlier reviews.

Of the filed issues approximately 70% deal with specific minor bugs or improvements, one issue was filed on the overall prototype, one regarding the Statistical Classification and its relationship to the DDI product XKOS, and 10 issues on broader modeling issues (use of UML, Functional Views, Patterns, etc.). A summary of the issues identified during the TC review is found at

<https://ddi-alliance.atlassian.net/wiki/spaces/DDI4/pages/589824023/DDI+Prototype+TC+Review+comments>

## Appendix 2

# MRT - Modeling, Representation, and Testing Lifecycle: A Proposed Working Group for Building DDI 4 Core

*Draft 1.0, 28 January 2019*

*(Contributors to this document include: Guillaume Duffes, Dan Gillman, Jay Greenfield, Arofan Gregory, Oliver Hopt, Larry Hoyle, Jon Johnson, Hilde Orten, Flavio Rizzolo, Wendy Thomas, and Achim Wackerow)*

## Overview

This document proposes a re-organization of the DDI 4 Modeling Team to incorporate suggestions coming from various quarters that a shorter development and testing cycle be adopted, and that the scope of the overall modeling effort become more focused. The structure and focus of each of a number of sub-teams is outlined, to be coordinated by an overall working group.

In order to guarantee the practical utility of the model, people and projects – both internal and external to the DDI Community - will be enlisted to act as testers for the draft models as appropriate. This re-organization provides for a liaison function so that the findings of groups testing and implementing drafts of the model are tracked and incorporated in a systematic fashion. It will be the function of the MRT Working Group to manage these relationships.

The scope of the effort is also defined, as are efforts to identify and formalize the business and technical requirements to which the group is building. The timelines are intended to be short – rapid development cycles will help to produce useful standards output in the shortest possible timeframes. This document addresses work over the course of the calendar year 2019.

## MRT Working Group

The MRT working group focuses on the iterative lifecycle of modeling, representation, and testing. This group replaces the previous modeling group. The new group should have experts on each stage of the MRT lifecycle, i.e. modeling, representations, testing of metadata instances, and software for transformation and testing. These experts can be involved in weekly phone calls on an on-demand basis: not all participants need to be involved in every task of the group.

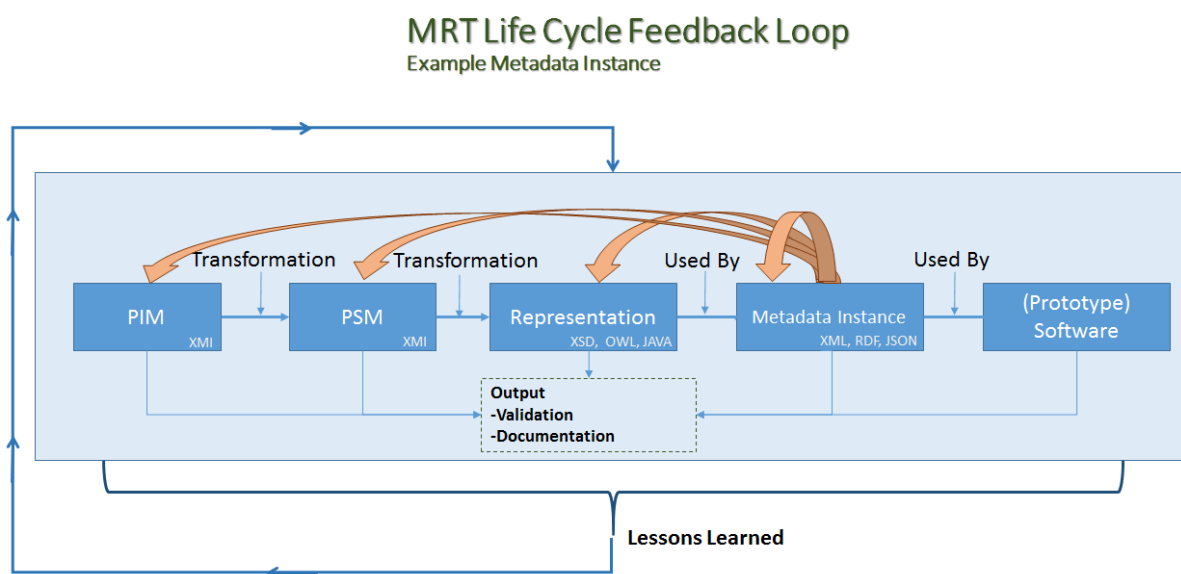
The intention of combining modeling, representation, and testing in an iterative lifecycle is to achieve more efficient production of the specification. Any major change on the model level can be immediately evaluated on other levels and its effects seen in practical application. The outcome of this testing can then be fed back into the modeling work, driving improvement on all the different levels. This approach avoids long isolated work on just one stage of the MRT

lifecycle. A suitable iteration frequency needs to be determined for the work process, but is intended to be measured in weeks, not months.

A combined approach for modeling, representation, and testing would enable a balanced perspective on the different needs of different aspects of implementation and use. It would enable a robust model with robust representations and documentation. All of these, taken together, are the DDI 4 specification - not just the model itself.

The result of the work should be a first release of DDI 4. It would consist of the model, the major representations, and the documentation for these. The overall framework for modeling, generating the representations, generating documentation, and testing will be established as a necessary part of producing the release.

The figure below shows the envisioned MRT life cycle:



(Key: PIM = platform independent model; PSM = platform-specific model)

The model needs to be improved according following requirements:

- Creating the model
  - Consistency and stability of the model
  - Persistent expression of the model in canonical form
  - UML conformance / usage of UML
- Using the model
  - With UML tools, connecting to other models
  - Subsetting the model, useful views or other subsetting mechanism
  - Supporting the main representations efficiently
  - Lossless roundtrip of metadata in different representations

The generation of representations needs transparent and straightforward transformation rules.

The work of the group should focus on DDI 4 core with following goals:

- Specification as UML model and representations as XML Schema and OWL

- Library from which subsets for major use cases and audiences can be built
- Core areas, i.e. conceptual, data description, and process
- Transparent model which is suitable as a basis for interoperation/alignment with other standards
- Clear identification of conformance to or divergence from previous versions of DDI Interoperable UML library which can be used in major UML tools.
- Applicable to data from other domains as well as social science, with domain-specific features implemented as extensions and clearly identified (if any). The data description should support logical and physical description of rectangular, key-value pair (e.g. Hadoop), and event data, as well as data cubes and individual datum cells. It should also address the modelling of collections of metadata items related to data description.

It is recognized that a determination of the production platform and process will be an immediate and critical topic to be addressed by the MRT to enable this cyclical development. This topic will be undertaken as a matter of priority, working with other relevant groups within the Alliance.

## DDI 4 Core

The focus should be on the core features of the DDI 4 prototype which comprises the conceptual, data description, and process areas. These three areas are useful for a large variety of audiences and use cases independent of specific needs. They represent a common ground for different requirements. The focus on these three core areas enables the specification of a DDI 4 core which is suitable for cross-domain use. The scope of each of the three areas needs to be determined.

A robust approach (regarding model and representations) for these three features would enable two things:

- Providing a core DDI 4 release which would be available for use
- Providing a robust “engine” for adding additional content features like data capture and other generic and specific components. The “engine” represents the model, representations, related transformations, test instances, prototype software, and the production framework.

The release of DDI 4 Core is roughly planned for the end of 2019. The overall attempt is to prove that DDI 4 is a usable and promising specification.

DDI 4 Core needs to be done in a way that additional areas can be added later without any change to the core. In UML terms, addition of further features can be achieved by adding new packages and classes or by specialization of existing classes.

## Alignment with other Metadata Standards

DDI 4 Core should be aligned with or interoperate effectively with *at least* the following other standards:

- GSIM – The Generic Statistical Information Model. Reference standard in official statistics

- DCAT – The Data Catalog Vocabulary is a W3C RDF vocabulary designed to facilitate interoperability between data catalogs published on the Web.
- SSN – The Semantic Sensor Network (SSN) ontology is a W3C ontology for describing sensors and their observations, the involved procedures, the studied features of interest, the samples used to do so, and the observed properties, as well as actuators.
- CSV on the Web – A W3C standard to express useful metadata about CSV files and other kinds of tabular data.
- PROV-O – A W3C standard to represent and interchange provenance information generated in different systems and under different contexts.
- BPMN – The Business Process Model and Notation is a graphical representation for specifying business processes in a business process model.

The alignment should be done in such a way that DDI 4 supports combined use with each of these standards. Ideally, metadata can be migrated from DDI 4 to each of these standards and vice versa, as appropriate to identified use cases.

## Additional DDI 4 Areas

For each additional area, a clear distinction should be made if the area is suitable for cross-domain use or is specific for the social science domain (like data capture with questionnaires). Work on additional areas can be done at the same time as the MRT group is working. But any additional work should have second priority and can only be done if the goals of a DDI 4 Core release are not affected. Example areas are data capture with questionnaires, coverage of DDI Codebook, methodology, and qualitative data.

## Requirements

### Technical Requirements

The MRT group should improve the technical requirements regarding each stage of the MRT lifecycle and create related guiding documents, based on input from the business requirements (see below), review activities of the existing efforts, and input from implementers. Furthermore, the model needs to be improved in several identified areas:

- Simplification of the model (i.e. less inheritance and less specialized classes)
- Review of collections (use of appropriate UML properties, use of collections throughout the model)
- Review of design patterns (relationship to acknowledged software design patterns, relevance of design patterns for users of the model and of the representations)
- Review of views (definition and effective use of subsets of the model)
- Model as portable UML library which can be imported into major UML tools (validation and reuse (in other context) of the model)

## Business Requirements

Work on business requirements can be done as a parallel activity to the MRT work on the core and fed into identified technical requirements. This work on business requirements is independent because areas of DDI 4 Core can be useful for any use case and audience. Providing a robust “engine” will be necessary for any features of specific business requirements. It is expected that implementers of the core will help refine business requirements vis-a-vis other standards and more generally for the core in their specific applications. The business requirements will be important in the future (after the release of DDI 4 Core) to determine additional areas which should be added to the DDI 4.

A review of the documentation is necessary with the goal that a descriptive language is used which is accessible to a wide user community. The class-level documentation should include mappings to earlier versions of DDI (and potentially to other useful targets) at with a level of detail sufficient to support consistent implementation.

## Organization and Structure

It is anticipated that several different tasks will need to be completed. These would include the following:

1. **MRT Working Group Coordination** – a coordination team will perform this task, holding regular meetings to manage the overall work of the group. Attendance would include specifically invited members of the other task teams on an as-needed basis, but would also consist of a core of people responsible for managing the work. Meetings would be open to all members of the effort.
2. **Modeling Tasks** – modeling work will be conducted by task teams responsible for further developing and maintaining the model in each core area, and for identifying and tracking the technical requirements of the work. The outputs of these teams would be representations of the model in an agreed XMI format (ideally Canonical XMI) which would form the input to the generation of representations. Each team would focus on one of the DDI Core areas – other modeling work might be conducted in other groups as agreed, but these would not be the top priority of the overall effort.
3. **Representation Tasks** – each representation being developed as part of the standard would have a task team responsible for taking each iteration of the model and producing bindings into their target syntax, and to provide assistance to those working on the documentation of the representation/model. These teams are likely to be small, and it is anticipated that representations will be generated from the model, so that the job of repeatedly generating representations is not onerous.
4. **Documentation Task** – the overall documentation for the model and representations is handled by another task team working in coordination with the other teams. To the extent possible, documentation should be generated from the model and representations, according to whatever system is used for producing and packaging it (presumably using the existing tools and making any needed adjustments).



5. **Testing and Liaison** – members of the group will be identified as liaisons to different projects acting as implementers and testers of the ongoing drafts. These individuals would feed back issues and technical and business requirements to other groups as needed.

It is intended that the structure of these task teams be fairly fluid, so that communication is as easy as possible. A high degree of cross-membership in the task teams is anticipated, given the size and expertise of the group and its members. Each task would have a lead assigned as the person responsible for achieving that task and communicating any status and issues to the coordination task team. It is assumed that the leads would be regular attendees at weekly coordination meetings. Face-to-face meetings would support the overall work most efficiently. Possible options seem to be in the margins of NADDI in April and at Dagstuhl in October. It is anticipated that the MRT group would work closely in conjunction with the Advisory Group and other parts of the DDI Alliance (e.g., the Technical Committee) to ensure that results are in line with the overall workings of the organization.

## Appendix 3

# Planning for Possible MRT Sprint in NADDI Margins

Version 1.0, 30 January 2019

The group of people mentioned in the document on “MRT - Modeling, Representation, and Testing Lifecycle: A Proposed Working Group for Building DDI 4 Core” proposes a face-to-face working meeting in Ottawa prior to NADDI on April 22-24. The overall goal is to work on a DDI 4 Core described in the mentioned document. The proposed meeting would need funding support from the DDI Alliance.

The meeting will focus on two main areas:

- Documentation of datum-based model application to examples of following data structures: rectangular data, event data (wide and narrow data), single datum points, multidimensional data like data cubes and aggregate data.
- Discussion and possible resolution of structural model issues:
  - Simplification of the model (i.e. less inheritance and less specialized classes)
  - Review of collections (use of appropriate UML properties, use of collections throughout the model)
  - Review of design patterns (relationship to acknowledged software design patterns, relevance of design patterns for users of the model and of the representations)
  - Review of views (definition and effective use of subsets of the model)

Nine people would be available in person for this meeting. Most of people would need funding from the DDI Alliance. In summary:

- All people would need support for food
- Five people would need support for additional accommodation (they are anyway at NADDI)
- One US-based person would need support for travel and accommodation
- One Europe-based person would need support for travel and accommodation
- Clarification needed for one Europe-based person

Another person would be available by phone.

The meeting could take place on ...

- Monday, April 22 (StatCan is closed on Easter Monday) in the NADDI conference hotel (\$200 CD\$ + 13% HST (including projector, screen, flip charts and coffee/tea for breaks))
- Tuesday and Wednesday, April 23/24 in rooms of StatCan

The hotel costs would be on ...

- Sunday and Monday, April 21/22, CAD 139.00 +17.52% tax (MAT+HST) per night per person
- Tuesday and Wednesday, April 23/24, CAD 126.00 +17.52% tax (MAT+HST) per night per person

Estimated total cost: \$9,408.88 CAD.

Name	Travel	Hotel	Food	Meeting Room	Total
Person 1	1800	622.86	200		
Person 2	600	622.86	200		
Person 3		622.86	200		
Person 4			200		
Person 5		622.86	200		
Person 6		622.86	200		
Person 7		622.86	200		
Person 8		622.86	200		
Person 9		622.86	200		
Hotel Room on 22 April (StatsCan closed)				226	
<b>Total</b>	2400	4982.88	1800	226	9408.88
*Currency in CAD					
	Accommodation:				
	\$326.71/SunMon,				
	296.15/TueWed				
	Food = \$50/day				
	for 4 days				
	Travel costs are				
	rough estimates.				