MRT DDI 4 Core Sprint, Dagstuhl 2019: Agenda and Deliverables

*Version 1.0 – 6 September 2019*

# Overview

This document outlines the work agenda and deliverables for the week-long Sprint to be held at Schloss Dagstuhl in Wadern, Germany from September 30 to October 4. The goal of the workshop is to finalize the modelling work on the MRT DDI 4 Core, covering the data description and (provenance) process portions of the DDI 4 Prototype model, as outlined in other project documents. These would form the basis of an initial release to be published for review at the start of 2020.

While it is recognized that the final review version of this package will not be ready by the end of the Dagstuhl Sprint, the substantive work should be completed. Further efforts through the end of the calendar year can thus be focused on packaging and finalization of documentation and presentation.

This document describes the work agenda for the workshop to achieve this goal and specifies the deliverables which are anticipated. As is often the case, some deliverables are seen as “nice to have,” rather than as firm requirements – these will be identified. If possible, such deliverables will be produced (as occurred in the Ottawa MRT Sprint earlier this year), but they are not seen as critical path for delivering by end of year (the schedule to which the group has committed).

# Participants

A significant portion of the working group will be attending the Sprint in person. Those who are unable to attend will be available virtually for those topics which require their input. Virtual attendees are noted below.

All participants are regular members of the Working Group, and the process for the workshop will mirror the normal working process of the committee.

**Organizers:**

Arofan Gregory, acting MRT chair, Consultant

Hilde Orten, workgroup Infrastructure and Records, NSD

Joachim Wackerow, institutional host, GESIS A Leibnitz Institute for the Social Sciences

**In-Person Attendees:**

Flavio Rizzolo, Statistics Canada

Daniel Gillman, US Bureau of Labor Statistics

Larry Hoyle, University of Kansas

Jay Greenfield, Consultant

**Virtual Attendees:**

Wendy Thomas, University of Minnesota Population Center

Oliver Hopt, GESIS A Leibnitz Institute for the Social Sciences

# Work Agenda

The work will be organized around the areas of the specification being produced, with the intention of having drafts of the specifications and documentation in hand by the end of the workshop. Some additional activities will also be required, especially in the case of the creation of automatically generated bindings and documentation.

1. Develop the overview and Upper Model
2. Update and present the foundational portions of the Prototype needed for DDI 4 Core (concepts, categories, classifications, variables, etc.)
3. Update and present the patterns used in MRT DDI 4 Core
4. Document the Process model
5. Document the Data Description model
   1. Core model
   2. Unit record data
   3. “Tall/Skinny” data
   4. Cube Structure model
   5. Key-Value/No SQL model
6. Document exemplary Functional Views
   1. Data management view (process/provenance and data description)
   2. Data description view(s)
   3. External integration view – DCAT (possibly part of Standards Alignment document)
7. Document Approach to Standards Alignment using the DDI 4 Core (GSIM, DCAT, GSBPM, PROV-O, other DDI versions, etc.)
8. Finalize the Canonical XMI version of the model, and integrate with other relevant production systems (binding production, modeling tools)
9. Develop and document the XML Schema bindings for all portion of the model covered by DDI 4 Core
10. Develop and document an experimental OWL/RDF syntax binding, pending resolution of the approach to including other relevant vocabularies

This list indicates the work areas for the week. These will be organized into a presentation oriented toward users of the specification for actual release and will consist of a series of documents/packages, each addressing one or more of the agenda items listed above. However, the work in each area should result in a draft version of the text needed for the creation of the deliverables in a direct fashion – each area should produce specification text and similar materials, rather than working notes or internal documents.

# Deliverables

This list describes the anticipated deliverables of the workshop, forming the initial draft of the specification to be available for review at the start of 2020. All substantive work will hopefully be completed for this set of deliverables, for finalization during November and December of 2019. Organization of the content presented here may ultimately be organized into a different set of actual documents based on emerging requirements. The content and coverage, however, will not change as a result of this re-packaging.

Note that it has been agreed that each portion of the model will be presented in its canonical form without the process patterns shown, in order to simplify presentation of the model and make it easier for users to understand. Versions including the patterns will also be documented as appropriate for more sophisticated users (eg, maintainers and developers). This distinction will be reflected also in appropriate machine-readable products (eg, the Canonical XMI).

The OWL/RDF syntax representation will be offered as a prototype in this first release. Funding for an external expert to integrate relevant RDF vocabularies in the OWL/RDF syntax representation was not available. This was planned as second part of the RDF work after the mapping of the UML model to an OWL/RDF expression. We hope that this work can be continued in the next budget year as the integration of DDI 4 as domain agnostic metadata standard into the web of Linked Data seems to be crucial. Offering an RDF syntax representation without the integration with existing relevant RDF vocabularies may allow a broader community to give us feedback on where that integration should occur.

1. Overview and Purpose (document)
2. The Upper Model for DDI 4 Core (diagram and document)
3. Foundational Classes - concepts, variables, categories, classifications, etc. as presented in the Prototype, but updated to agree with the current modeling guidelines (diagrams and document)
4. Design Patterns in DDI 4 Core – an introduction to the design patterns used (diagrams and document)
5. The DDI 4 Core Process Model (diagrams and document)
6. The DDI 4 Core Data Description Model (diagrams and document)
   1. Core classes
   2. Unit-record data
   3. Tall/skinny data
   4. Cube structures for dimensional data
   5. Key-value/No SQL data
7. DDI 4 Core and External Standards (GSIM, GSBPM, DCAT, DDI-C, DDI-L, etc.) ***[NICE TO HAVE]***
   1. Use of external standards for DDI users
   2. Use of DDI 4 Core for external users
8. Functional Views in DDI 4 Core (exemplary views) (diagrams and document) ***[NICE TO HAVE]***
   1. Data Management View, encompassing process/provenance and data description
   2. Data Description View, covering the straight data descriptions case(s)
   3. External Standards View, showing how DDI can be used in conjunction with an external domain standard [specific example TBD]
9. Canonical XMI for DDI 4 Core (valid XML XMI describing the DDI 4 Core Model with documentation)
10. XML Schema Definition (XSD) for the DDI 4 Core Model (XSD schemas with documentation)
11. OWL/RDF Syntax Binding for the DDI 4 Core Model (RDF vocabulary definition and documentation) ***[NICE TO HAVE]***