Summary Report: DDI 4 Core Sprint, Ottawa, 22-24 April 2019

# Overview

This document provides a brief summary of the work which took place at the DDI 4 Core Sprint in the margins of the North American Data Documentation Initiative Conference, from 22 to 24 April 2019. It describes the agenda, participants, work-streams, and deliverables. These are placed in the context of the overall work plan of the group, with an emphasis on the year-end deadline for the work.

The Sprint was a meeting of the Modeling, Representation, Testing (MRT) working group. This group has been formed as part of the DDI Moving Forward project, and largely replaces the Modeling Team, but has a focus which includes not only modeling but also testing of the syntax representations and other work products.

In general terms, the Sprint was very successful – all of the anticipated deliverables were completed, and some topics for future work were discussed or explored, such that future progress will be more easily realized. In terms of the overall workplan, target milestones have been met, and in some cases exceeded. Significant work remains, but with the progress made at the Sprint, delivery of the DDI 4 Core at the end of December 2019, as a product ready for review and publication is still a realistic goal.

# Background: Goals and Timelines

The DDI 4 Core was identified by MRT as a subset of the DDI 4 released in the Prototype Review package (see the *DDI 4 Core Summary and Overview* document). It is intended to be a production release of some of the most useful functionality supported by that model and associated products, narrowed in scope to make resource issues more tractable. Emphasis is on the foundational metadata, data description, and some applications of the process model.

The MRT Working Group has adopted a working process somewhat different from earlier DDI 4 projects: a more limited scope has been identified, and short-term timelines established. The core features of the existing DDI 4 model are to be finalized and the entire standards product (the model, documentation, and syntax representations/bindings) is to be ready for review as a production release by the end of 2019. The working process is an iterative one, more fully embracing the Agile methodology which has to a limited extent informed all of the DDI 4 work up to this point.

Central to the work is the existence of a production system which will allow modeling to become part of a cycle which also includes the production of documentation and bindings. This system did not exist in a useful form at the start of the work, and prior to this Sprint half of the group’s efforts have been focused on developing this critical infrastructure from the existing one (the TC production framework and the Lion Repository). The initial move off of the previous infrastructure was achieved at this Sprint, which is an important milestone in the overall working of the group, even if one which is not as visible in terms of the eventual standards product to be delivered.

One change from earlier production processes is the use of Canonical XMI as a format for describing the model. This format was agreed in discussions with the TC as offering several benefits. It serves as an exchange format for the UML model between the MRT and the TC, being designed as a portable format for such models. Further, it can also be used directly as a deliverable by users across a wide range of UML tools, a feature which is of increasing as implementers use DDI in new ways (e.g., not as XML or RDF, but as a model for analysis packages, repositories, and other systems).

In terms of the model content, the existing scope has been narrowed, but the substantial work of the past years forms the basis for the group’s current efforts. It is in essence a finalization and productization of the model and derived products, informed by the recent implementation and review of the DDI 4 Prototype. This input has indicated that changes are needed in both the style and content of the model and related products. Further, the work will need to be passed on to the TC at the point where it is ready for public review and distribution – the TC is ultimately the part of the DDI Alliance which will maintain it. Thus, alignment and integration with the TC production and management systems has been given a high priority in the work of the MRT.

# Agenda and Participants

The agenda for the work was:

* Definition of the DDI 4 Core/Scope and Extraction from the Lion Repository (added just prior to the Sprint)
* Resolution of Open Modeling Issues/Production of UML Modeling Guidelines
* Production of Examples for Datum-Oriented Data Description/Gap Analysis
* The Role of Functional Views/Subsets
* Standards Alignment/”Round-Trippability” between Syntax Representations/Bindings
* Annotations, Access, and Citation Information (added during the Sprint)

The first agenda item was added after discussions subsequent to approval for Sprint funding by the Executive Board and was not tied to a specific deliverable at the time of approval. The agenda item regarding Standards Alignment was seen as perhaps too ambitious in a 3-day sprint and was presented as low-priority in terms of the work programme. The work on Annotations, Access, and Citation Information was not initially a part of the agenda but emerged during the Sprint as a result of work in other areas (notably responses to issues in the Prototype review and the work on Modeling Guidelines).

Participants at the Sprint were:

Dan Gillman, US Bureau of Labor Statistics

Arofan Gregory, Invited Expert

Larry Hoyle, University of Kansas

Hilde Orten, NSD - Norwegian Centre for Research Data

Flavio Rizzolo, Statistics Canada

Wendy Thomas, Minnesota Population Center, University of Minnesota

Joachim Wackerow, GESIS - Leibniz Institute for the Social Sciences

Jay Greenfield (virtual), Invited Expert

Oliver Hopt (virtual), GESIS - Leibnitz Institute for the Social Sciences

# Work Streams and Process

Work during the Sprint was prioritized according to the deliverables, and these served as the focus for the work. The production of each deliverable was broken down into discrete tasks taking one or more 1.5-hour sessions, performed by 1 – 3 people working in small breakout groups. The working process was Agile, with periodic re-organization and re-ordering of the groups based on reviews by the whole team after each work session. In the past, this Agile-based process has proven to be very efficient for groups with a limited number of participants, as at this Sprint. During the work, there was an emphasis on documentation of all discussions and decisions, so that it would directly contribute to deliverables and serve as a record for those not participating directly. All documentation is available on the Confluence site used at the Sprint as a collaboration platform.

* DDI 4 Core Scope Definition and Extraction from the Lion Repository
* UML Modeling Issues and Guidelines
* Datum-Based Application of the Model: Examples and Gap Analysis
* Analysis of Functional Views
* Alignment with External Standards and “Round-Trippability”
* Annotations, Access, and Citation Information

See the list of the deliverables, below, for a description of what each work stream addressed at the Sprint.

The work concluded with a planning session to organize future work, based on progress in each area at the Sprint, and in relation to the plan for producing the December 2019 package. This included identifying future activities and milestones, taking into account both holiday schedules during the summer and at years’ end, and the increased productivity anticipated at the planned Dagstuhl Sprint in the fall.

# Deliverables

The following is a brief description of each of the deliverables produced during the Sprint. Note that some work streams produced more than one deliverable. In some cases, this exceeded the initial expectations of the Sprint – additional deliverables are generally in the form of position papers or proposals which will be addressed by the MRT Working Group in their normal weekly calls and future Sprints. Links to the latest version of all documents on the Confluence wiki are provided for each work stream.

## Work Stream: DDI 4 Core Scope Definition and Extraction from the Lion Repository

* **DDI 4 Core Definition of Scope** – document outlining supported functionality, with the details of what was included in the extraction from the Lion Repository to support the DDI 4 Core work. Possible mid- and future-term functionality is identified, based on the portions of the model included in the Prototype Review package, and the extent of the work already done in different functional areas. [[LINK TO CONFLUENCE]](https://ddi-alliance.atlassian.net/wiki/spaces/DDI4/pages/689274975/1.%2BDDI%2BCore%2BScope?preview=/689274975/691634177/2_MRTScope.docx)
* **DDI 4 Core Model Extraction** – the model subset to be used for the purposes of finalizing the DDI 4 Core was extracted in the form of (proprietary) XMI. This was then modified to conform to the Canonical XMI form agreed as the production format between the MRT and the TC. Limited testing was performed in various tools and is ongoing.
* **DDI 4 Complete Model** – to guarantee that none of the work thus far conducted in the DDI 4 project is lost, a second extraction was performed including all of the model minus that portion which has been discarded over the course of DDI 4 modeling work in the Lion Repository. This extraction was for archival purposes and remains in the proprietary XMI format produced by the Lion Repository.

## Work Stream: Modeling Issues and Guidelines

* **UML Modeling Guidelines –** this document reflects changes made to the modeling style of the DDI 4 Core in response to open issues following the Berlin Sprint, those raised during the Prototype Review, and consideration of UML best practice and tools support. Decisions reflected here will be implanted in the DDI 4 Core model by MRT moving forward.

The modeling guidelines are a key tool to resolve filed issues on the model in a consistent way. They provide rules which ensure that the model is conformant to UML, and that it uses the chosen UML subset efficiently. The new UML guidelines complement existing business modeling guidelines on the formal UML level and ensure a clear and consistent implementation model which can be transformed efficiently to multiple syntax representations. [[LINK TO CONFLUENCE]](https://ddi-alliance.atlassian.net/wiki/spaces/DDI4/pages/689438803/2.%2BCore%2BModelling%2BIssues?preview=/689438803/691732487/3_Modelling%20recommendations.docx)

## Work Stream: Datum-Based Application of the Model: Examples and Gap Analysis

* **Examples of Datum-Based Applications –** This document provides three illustrative examples for a general audience, showing how the DDI 4 Core model can be used to describe unit-record data, multi-dimensional data/aggregates, and event data. This will in future form a key input to documentation and training activities – extension with an introduction to the model itself, bridging the gap between the conceptual level and the formal (UML), will be added, and the excellent introduction to the Variable Cascade produced at the Dagstuhl workshops in 2018 will be incorporated. [[LINK TO CONFLUENCE]](https://ddi-alliance.atlassian.net/wiki/spaces/DDI4/pages/689242207/3.%2BExample%2BDocumentation%2BApplication%2Bof%2BDatum-Based%2BApproach%2Bto%2BDifferent%2BStructures?preview=/689242207/691929092/4a_Documenting%20Data%20Structures.docx)
* **Proposals for Additions to the Model –** Having identified the existing gaps in the production of the first deliverable, proposals were created for addressing them. These take the form of a proposal addressing the classes and properties needed to better describe Event data, and a proposal for describing multi-dimensional data (aggregates, time-series, etc.). These do not represent major extensions of the existing model, but refinements to make it both more powerful and easier to use for these specific applications. Reaching agreement within the MRT group to these proposals is the next step, following which appropriate changes will be made to the model for testing. [[LINK TO CONFLUENCE - Tall Layouts]](https://ddi-alliance.atlassian.net/wiki/spaces/DDI4/pages/689242207/3.%2BExample%2BDocumentation%2BApplication%2Bof%2BDatum-Based%2BApproach%2Bto%2BDifferent%2BStructures?preview=/689242207/691601415/4b_ClassesForTall%20Layouts.docx) [[LINK TO CONFLUENCE - Multi-dimensional/Cube Layouts]](https://ddi-alliance.atlassian.net/wiki/spaces/DDI4/pages/689242207/3.%2BExample%2BDocumentation%2BApplication%2Bof%2BDatum-Based%2BApproach%2Bto%2BDifferent%2BStructures?preview=/689242207/691666994/4c_Describing%20a%20Data%20Cube%20in%20DDI4.docx)

*Note:* The group will be assessing No-SQL, key-value, and some other types of data stores to determine how best to address these within the current scope. This was not a work-stream during the Sprint but is an open issue for near-term work.

## Work Stream: Analysis of Functional Views

* **Functional Views Position Paper –** Over the course of DDI 4 development, the role played by Functional Views has evolved. This aspect of DDI 4 was included in the Prototype Review package but has never been fully defined or documented. This position paper considers the history, purpose, and means of determining how the DDI Model can be made more approachable and easier to use on the basis of this mechanism. Methodical analysis of the model served as the basis for this discussion, which promises to be a fruitful technique for identifying the most useful organization of the DDI 4 Core products from a functional perspective. [[LINK TO CONFLUENCE]](https://ddi-alliance.atlassian.net/wiki/spaces/DDI4/pages/689340516/4.%2BFunctional%2BViews%2BSub-Setting%2Bof%2Bthe%2BModel?preview=/689340516/691568659/5_FunctionalViews.docx)

## Work Stream: Alignment with External Standards and “Round-Trippability”

* **Position Paper on Standards Alignment** – The importance of aligning with external standards, to optimize the value of the DDI 4 standard, has long been recognized. Over time, the ideas about what this means in terms of specific uses of the standard has evolved. This paper summarizes the standards which may need to be addressed, and the issues raised when alignment with these is considered. It is anticipated that this topic will be a near-term one for discussion within MRT and will result in refinements of the existing model to better align with specific external standards. Further, it is expected to impact the technical expression of the model in syntax representations, to support transfer across these expressions and technology platforms. It is hoped that this work will produce guidelines for this aspect of DDI 4, as the basis for future work. [[LINK TO CONFLUENCE]](https://ddi-alliance.atlassian.net/wiki/spaces/DDI4/pages/689274988/5.%2BAlignment%2Band%2BUse%2Bof%2BOther%2BStandards%2BRound-Trippability%2Bof%2BBindings?preview=/689274988/691732493/6_RelationshipToOtherStandards.docx)

## Work Stream: Annotations, Access, and Citation Information

* **Position Paper on Annotations, Access, and Citation Information –** The modeling of this information in the Prototype Review package has been identified in some of the issues raised during the review and is also impacted by considerations of the modeling style for the DDI 4 Core work. This paper considers possible resolution of these issues and changes needed in the modeling guidelines as the basis for further discussion within MRT. [[LINK TO CONFLUENCE]](https://ddi-alliance.atlassian.net/wiki/spaces/DDI4/pages/691273767/6.%2BAnnotations%2BAccess%2Band%2BCitation%2BInformation?preview=/691273767/691667003/7_Annotation-Access-Provenance.docx)

# Summary and Assessment

Overall, the Sprint was extremely productive. All of the anticipated milestones were met, and the set of deliverables was somewhat larger than anticipated. In terms of the overall workplan, the MRT is now positioned to begin the iterative production and testing cycles identified as their intended work process.

While a significant amount of work remains, at this point the project seems to be firmly on-track, with a delivery of a production-ready DDI 4 Core package to the TC for review and publication at the end of 2019 expected. The results of the Prototype Review have been addressed as appropriate, including the integration of issue resolutions into the Jira issue-tracking system being used by both the TC and the MRT (this should ease transition to the TC at time of delivery).

All issues considered during the Sprint which were raised during the Prototype Review are been updated in Jira, working in coordination with the TC to make sure that all changes are known to both groups.

The package to be delivered covers only a subset of the overall DDI 4 model but will provide essential functionality to users and lay the foundation for expanded functionality to be supported moving forward.

Overall, the Sprint exceeded expectations in terms of deliverables, and helped to point the way toward final delivery in December and a relatively painless hand-off to the TC at that time.

Thanks go to all participants for their hard work and excellent focus during the Sprint. Thanks also to the DDI Alliance for supporting the Sprint and making this focused and intensive work possible.