DDI Alliance Scientific Board Meeting
Sydney, NSW, Australia -- UNSW Sydney
June 1, 2019

Minutes

Participants:

Tuomas Alaterä (Finnish Social Science Data Archive)
Iris Alfredsson (Swedish National Data Service)
Ingo Barkow (HTW Chur)
Cathy Fitch (Minnesota Population Center)
Jane Fry (Carleton University)
Jared Lyle (ICPSR)
Steve McEachern (Australian Data Archive)
Marianne Myhren (Norwegian Centre for Research Data)
Ron Nakao (Stanford University)
Barry Radler (MIDUS, University of Wisconsin--Madison)
Wendy Thomas (Minnesota Population Center)

Virtual Participants:

Dan Smith (Colectica)
Jon Johnson (UCL - CLOSER)
Guillaume Duffes (INSEE)

Scientific Board Structure

Ingo Barkow, Vice Chair of the Scientific Board, began the meeting by noting the current limitations of the Scientific Board structure (see Appendix 3). He indicated that the Scientific Board has not been active and that the Alliance Bylaws are not clear enough about the Board's boundaries and processes. He noted that it is hard to represent a group when they don't know what the members want.

He outlined a proposal to create an Acting Committee of the Scientific Board whose members are elected by all members of the Scientific Board. The Acting Committee would be formed by five elected member representatives and two external advisory experts. The Acting Committee would meet monthly in a conference call, as well as one yearly face-to-face meeting in the margins IASSIST, EDDI, or NADDI.

The details of the proposed Acting Committee would be developed by a working group over the next couple of months. The working group would propose the new structure, including clarifying

the roles, number of people, and necessary changes to the Alliance Bylaws. Ingo would chair the working group.

It was asked what the composition of the working group would be. Ingo indicated it would be open to anyone. It was asked how the Scientific Board will ensure fair representation on the Acting Committee so it does not become an echo chamber. It was discussed how people should be elected to the Acting Committee. It was noted that the Scientific Board has never before asked for comments or input from its members, so why should it create a small group if they haven't asked the whole group in the first place. It was also asked what the role of the individual members will be.

It was noted that the working group will discuss and propose solutions for these questions. It was suggested to have two external representatives on the working group. It was noted that the new group needs to be organized to get community feedback rather than making decisions in closed chambers. It is hoped that the new body can ask member representatives about priorities; we only know about reported issues, not what the scientific board representatives think.

As a next step, the Executive Board will ask for volunteers and require external representation on the working group.

Scientific Board Leadership

Steve McEachern, Chair of the Executive Board, reviewed a proposal discussed by the Executive Board in their March 2019 meeting to postpone elections for the Chair and Vice-Chair of the Scientific Board until July 2020. The three year terms for the Chair and Vice-Chair positions of the Scientific Board end in June 2019. Given the proposed changes to the structure of the Scientific Board, the one year delay will allow the proposed new structure to be voted on and implemented before new leadership is elected.

The Scientific Board membership should review the proposal to delay Scientific Board leadership elections, with the option to reject. No one attending the annual meeting voiced objection. An email will be distributed to the membership informing them of the proposal. If any members reject the proposal to delay elections, the Alliance will hold an election.

Training

Ingo discussed a proposal to publish a modular training library for further reuse and self-guided training (see Appendix 4). The goal of the training library is to provide consistent and appealing teaching material, covering all relevant areas, following a version-agnostic approach, and providing version-specific details. The training library would form the basis of teaching tutorials to multiple audiences like tutorials in the margins of conferences or workshops in summer schools. It would also be the basis of online training material. It was noted that the current training library has gaps in terms of covering all relevant areas and has limitations regarding content consistency and look-and-feel of the slides. The proposal for the training library is to

fund a paid person to edit the library. Volunteers can help in providing material for missing areas.

It was requested to define the training library. It was asked who is the audience for the training library, where are the gaps, and whether this effort will produce new training material. Additionally, it was discussed whether the training document manager activities could be combined with the efforts of the proposed marketing manager.

Scientific Board Direction and Goals

Ingo presented slides discussing possible goals for the Scientific Board for the next year, including interoperability of metadata standards, improving infrastructure around all DDI specifications, and project management for all DDI Alliance activities (see Appendix 2).

For slide 7, it was requested to better define and update “Project Management for all DDI Alliance activities.”

Moving Forward

Specifications Roadmap
Wendy Thomas, Chair of the Technical Committee, discussed the DDI Roadmap document prepared and recently updated by the Technical Committee (see Appendix 9). The document provides a broad discussion about existing products and how to frame DDI 4.

It was noted that development of the DDI specifications has shifted focus to applied usage and the environments in which they’re used. Also, development work on the different lines is not distinct; that is, the Alliance has added things to Codebook and parallel things in Lifecycle. This speaks to the issue of whether to cut off development and just fix bugs, or to continue to develop different lines.

Regarding DDI 4, concern was raised about being able to offer users a production model. It was recommended to keep discussion of DDI 4 to a minimum until the Alliance has a product to offer, as well as functionality that DDI 2 or DDI 3 cannot offer. It was suggested that functionality in DDI 4 could be added to DDI 3 to enhance DDI 3.

It was recommended to ask the community for their set of requirements and needs for different streams of development, including: Which version are you implementing now? What will you need in the future? What is the timeline? Where do you want to devote voluntary resources? Understanding these questions will help figure out timelines and priorities for the roadmap.

DDI Core
Steve McEachern described the DDI Core, which will use the core features of the DDI 4 model that are the most robust to date, conceptual, data description, and process, with a ‘core’ DDI 4 release that is implementable and the base on which to update the rest of the model. The goal is a DDI 4 Core release for review and publication in December 2019.

DDI Core will be fit for cross-domain integration of data and ready for data coming from new sources (e.g., “big data,” social media, sensor data, etc.). Features will include: the conceptual aspects of variables and classifications; the datum-oriented description of data; and the use of the process model to describe data lineage (the processing involved in the provenance of data). The architecture of DDI4 Core can be understood as a kernel on which additional features and user-oriented views can be arranged on layers. (See Appendixes 5-8, 10.)

Technical Committee FY2020 Workplan

Wendy Thomas reviewed the fiscal year 2020 work plan for the Technical Committee (see Appendix 12), including:

- Reviewing and publishing DISCO
    - Content is ready to be put out for a 30-day review in June
- Preparing DDI 3.3 for publication
    - Final entry and documentation
- Resolving 5 DDI 4 Prototype review issues
    - Overall best practices for XML structure
    - Shift from document base to multi-use instances (document, transfer, discovery systems, etc.)
    - Specification of default values that are inherited by contained objects
    - Broad feedback from ICPSR regarding interaction of DDI products, audiences, etc.
    - Broad feedback from NSD regarding role of DDI4
- Reviewing issues filed for DDI-Codebook and preparing a new version
    - Currently have a set of issues related to improving link with DataVerse
    - New issues are being filed in terms of issue arising from the new World Bank tool
- Shifting DDI Lifecycle and DDI-Codebook production work to COGS
    - We have tested out the documentation production using the DDI 3.3 public review
    - Test and finalize input-output scripts
    - Complete and test output to ensure coverage and output consistency
    - Finalize transfer and open for access
- Preparing for the shift of DDI 4 development work to COGS from Drupal
    - Finalize scripting for XMI input and output

Standards Development document

Wendy Thomas discussed a proposal to update the Standards Development and Review Process and Procedures document (see Appendix 13). The current document is complicated, does not have a good way to approve minor changes, and requires a two-thirds majority for all votes.

It was suggested that voters should be able to indicate “I have no objection but I can’t evaluate.” It was also suggested that standards reviews and votes should be explained in lay people terms so the general membership can understand how the new changes will benefit them.

The Technical Committee will review and update the document. Thereafter, it will be sent to the Scientific Board and the Executive Board for review.

It was noted that the Alliance needs to add visible functionality to the web site about how to file an issue for the Technical Committee to address.

SDTL

Steve McEachern discussed a proposal by George Alter and the C^{2}Metadata project for the DDI Alliance to monitor, maintain, and update the Structured Data Transformation Language (SDTL) as a standard (see Appendix 14). SDTL is an independent intermediate language for representing data transformation commands. Statistical analysis packages (e.g., SPSS, Stata, SAS, and R) provide similar functionality, but each one has its own proprietary language. SDTL consists of JSON schemas for common operations, such as RECODE, MERGE FILES, and VARIABLE LABELS. SDTL provides machine-actionable descriptions of variable-level data transformation histories derived from any data transformation language. Provenance metadata represented in SDTL can be added to documentation in DDI and other metadata standards.

It was noted that maintaining SDTL would be coordinated in the same way the Alliance handles XKOS or Controlled Vocabularies. If there is a group that wants to do maintenance, our role is to handle the organizational aspects and review for publication.

The next step will be to take this proposal to the Technical Committee. The Executive Board should also discuss the process for deciding how the Alliance formally reviews and accepts proposals like this, including what is the advantage to DDI and the individual product.

Appendix 1

DDI Alliance Meeting
Saturday, June 1, 2019, 08:30-16:30
Hugh Dixson Theatre, UNSW Sydney
Gate 11, AGSM Building, Kensington Campus -- Directions

Virtual link

Agenda -- Meeting of Members				
Time	**Subject**	**Detail**	**Lead**	**Purpose**
08:30-09:00	Coffee			
09:00-09:05	Welcome		Steve	Introductions
09:05-09:20	State of the Alliance 2019		Steve	Update on last year's work
09:20-09:30	Alliance Budget	Financial Report	Jared	
09:30-10:00	Working Group Reports	-Marketing & Partnerships -Training -Technical Committee -DDI 4 Berlin sprint MRT group Prototype public review Ottawa sprint	Barry Jane Wendy	
10:00-10:15	Coffee break			
10:15-12:25	Strategic Planning	-Strategic Plan 2018-2022 (draft) -Open Letter: "The Case for Continued Support of a Model-Driven, Platform-Independent DDI" -Executive Board draft response to Open Letter	Steve	Get input and feedback
12:25-12:30	Proposed Date for Next Meeting		Steve	Agree on best day to meet
12:30-13:30	Lunch			

Agenda -- Meeting of Scientific Board				
Time	**Subject**	**Detail**	**Lead**	**Purpose**
13:30-13:35	Welcome		Ingo	Introductions
13:35-14:05	Scientific Board structure	-Improvements of Scientific Board structure -Decision-making -Postpone elections (chair, vice-chair) for one year -Discussion	Ingo Steve	
14:05-14:20	Training	-Training library -Training at conferences for new users -Discussion	Ingo	
14:20-15:05	Moving Forward program	-DDI 4 Core (see also scope and MRT group) -Technical Committee's DDI Specifications Roadmap -Future direction on DDI 4, and additional extensions views like data capture and codebook -Discussion	Steve Ingo Wendy	In-depth discussion of DDI4 development
15:05-15:20	Coffee break			
15:20-15:50	Scientific Board direction and goals for the year	-What are the goals? -What is the work plan? -Discussion	Ingo	Set goals for what to accomplish
15:50-16:20	Technical Committee	-What is the work plan? -Updating the Standards Development and Review Process and Procedures document -Discussion	Wendy	
16:20-16:30	SDTL proposal		Jared	

18:00 - Informal DDI group dinner

Appendix 2

Agenda Overview

- Scientific Board structure
  - Improvements of Scientific Board structure
  - Postpone elections for one year
  - Discussion
- Training
  - Training library
  - Training at conferences for new users
  - Discussion
- Moving Forward program
  - DDI 4 Core (working name)
  - Future direction on DDI 4, and additional extensions views like data capture and codebook
  - Discussion
- Scientific Board direction and goals for the year

Scientific Board Structure

- Improvements of Scientific Board structure
  - Acting Committee of the Scientific Board
    - Sub-committee of Scientific Board representatives
    - Activating the Scientific Board as level between the Executive Board and the working groups
  - Roles of Member Representatives
    - Clarification of roles of designated member representatives vs. scientific board representatives
  - Voting in Committees and Working Groups
    - A growing organization needs clear voting rules
  - Temporary working group for creating proposals
- Postpone elections (chair/vice-chair) for one year

DDI Training I

- Training library
  - Goals is consistent and appealing teaching material
  - For all relevant areas
    - Following a version-agnostic approach
    - Providing version-specific details
  - Basis of teaching tutorials to multiple audiences like:
    - Tutorials in the margins of conferences
    - Workshops in summer schools
  - Basis of online training material

Editor required for creating and coordinating

DDI Training II

- Training at conferences for new users
  - Continuous offer on DDI tutorials for …
    - maintaining a good user base
    - increasing the use of DDI
  - Tutorials in the margins of conferences where the audience might be interested in DDI
    - Focus on target groups that DDI is not already familiar with
    - Workshops at survey methodology summer schools

Moving Forward Program I

- DDI 4 Core (working name)
  - Fit for cross-domain integration of data
  - Ready for data coming from new sources, some of it in unfamiliar forms, like …
    - “big data,” social media, sensor data, etc.
  - Features include …
    - the conceptual aspects of variables and classifications
    - the datum-oriented description of data, and
    - the use of the process model to describe data lineage (the processing involved in the provenance of data)
  - DDI 4 Core is just a working name
    - Name should make sense in relationship to DDI-Codebook and DDI-Lifecycle
    - Some suggestions: DDI-Integration, DDI-Cross-Domain
  - Release of DDI 4 Core for review is roughly planned for the end of 2019

Moving Forward Program II

- Future direction on DDI 4, and additional extensions views like …
  - data capture, codebook, and qualitative
  - The architecture of DDI4 Core can be understood as a kernel on which additional features and user-oriented views can be arranged on layers
  - Kind of additions needs to be determined
  - A related sprint could be organized in the margins of the IASSSIST conference 2020 in Gothenburg

Scientific Board Direction and Goals for the year

- Interoperability of Metadata Standards
  - 2nd workshop on “Interoperability of Metadata Standards in Cross-Domain Science, Health, and Social Science Applications” in Dagstuhl, Oct 7-11, 2019
  - The workshop is organized again together with CODATA (Committee on Data of the International Council for Science)
- Improving infrastructure around all DDI specifications
  - Best practices should be created on when to use which specification (and which part). This could be guided along use cases.
  - Improving portability of DDI metadata, interoperability of DDI systems
    - Criteria of portability and interoperability, validation tools
    - DDI Profiles / Views for specific use cases / specific perspectives
    - Description of the workflow in integrating above items
  - Supporting reusable DDI metadata in organizations and across organizations. Prerequisites for using efficiently DDI metadata repositories and registries
    - DDI URN resolution enables web-based reuse of metadata, i.e. persistent identifier to URL of DDI item.
    - Standardized query/exchange protocol enables reuse of DDI items stored in local and remote repositories, common data element registries
- Project Management for all DDI Alliance activities
  - Integrated coordination would make sure that activities are made in a consistent way following a plan which the Scientific Board and Executive Board agreed on
  - An open question are the required resources for this task

Appendix 3

DDI Alliance – Scientific Board
Goals 2019/2020

Joachim Wackerow, Chair of the Scientific Board, 2019-05-21

This is part of a larger plan. The proposals comprehend the areas of DDI training, a new acting committee of the scientific board, interoperability, DDI 4, and project management. Each item has a relationship to a topic in the strategic plan.

Scientific Board

This item is related to the strategic plan topic "**Business Structure**" in the section "The DDI Alliance as an Organisation".

Acting Committee of the Scientific Board

It is proposed to create an Acting Committee of the Scientific Board as a standing committee.

According to the bylaws, the Scientific Board is the scientific and technical body of the Alliance. It is composed of Member and Associate Member Organization Designated Representatives. It represents the level between the working level of the Sub-Committees (like the Technical Committee as standing committee) and Working Groups on specific topics, and the steering level of the Executive Board which is responsible for setting the overall policy and budget for the Alliance among other tasks.

The experience shows that the current set up - a Scientific Board of approx. 40 member representatives with the two elected chairs – has limited impact regarding the described role of the Scientific Board.

The proposal is to create an Acting Committee of the Scientific Board whose members are elected by the all members of the Scientific Board. Chair and vice-chair could be the elected members with most of the votes. Additionally, external assigned members should be part of this Acting Committee to assure additional perspectives from a larger community. The role of these external experts is inspired by the idea of a Scientific Advisory Board (SAB). This approach avoids inventing an additional committee like a SAB.

The Acting Committee should be formed by five elected member representatives and two external advisory experts.

The Acting Committee should meet monthly in a conference call. The Acting Committee should meet once the year in a face-to-face meeting (like one day). The face-to-face meeting should be in the margins of one of the conferences which are relevant for DDI like IASSIST, EDDI, and NADDI.

The details need to be discussed and agreed in the next couple of months. A related change of the bylaws might be required.

Roles of Member Representatives

The experience shows that a clarification regarding the roles of member representatives and scientific board representatives would be helpful. A related document should be created.

Voting in Committees and Working Groups

The DDI Alliance became larger and more heterogeneous. With this background, it doesn't seem to be possible to rely always on a decision mechanism by agreement (which was very successful in the past). Therefore clear voting mechanisms should be described for all committees and working groups. This can affect the decision on items where no agreement is possible and the process of assigning chairs.

Next Task

A related temporary working group should be established which creates detailed proposals for all three items.

Appendix 4

DDI Alliance – Scientific Board
Goals 2019/2020

Joachim Wackerow, Chair of the Scientific Board, 2019-05-21

This is part of a larger plan. The proposals comprehend the areas of DDI training, a new acting committee of the scientific board, interoperability, DDI 4, and project management. Each item has a relationship to a topic in the strategic plan.

DDI Training

This item is related to the strategic plan topic "**Training**" in the section "The DDI Alliance as an Organisation".

DDI Training Library

The work on a general purpose DDI Training Library has started in the Train-the-Trainers workshop in Dagstuhl 2018. The goal of the training library is to provide consistent and appealing teaching material. It should cover all relevant areas, follow a version-agnostic approach, and provide version-specific details.

The training library should form the basis of teaching tutorials to multiple audiences like tutorials in the margins of conferences or workshops in summer schools. It should be also the basis of online training material.

The current training library has still gaps in terms of covering all relevant areas and has limitations regarding content consistency and look-and-feel of the slides.

Organizing and editing work is crucial to achieve the described goal. This work should be done by a paid person. Volunteers can help in providing material for missing areas.

Details are available in appendix 1.

DDI Tutorials

In terms of maintaining a good user base and increasing the use of DDI, it is important to provide a continuous offer on DDI tutorials. The tutorials should be taught in the margins of conferences where the audience might be interested in DDI. It is even more important in terms of outreach to provide this at conferences where the audience is not really aware of DDI but might be interested. Examples are survey methodology conferences like ESRA and RC33 but there are much more opportunities. Furthermore, tutorials should be taught at summer schools of empirical social sciences. Examples are the ICPSR and GESIS summer schools. It seems to be important to support travel and accommodation of interested instructors. The details on how to offer these possibilities (yearly tutorial plan, tender for instructors) need to be determined.

Some conferences were sponsored in the past by the DDI Alliance in a general way as a form of outreach and marketing. This money could be used in a more focused approach for DDI tutorials. This way, both goals – outreach and training – could be achieved.

Appendix

1 Plan for DDI Training Library

1. A modular training material body (the library) will be published for further reuse and self-guided training. This will include the content presently on the DDI Alliance website.
2. This training library is an official publication of the DDI Alliance. The workshop participants are contributors to this material. They should get the appropriate attribution.
3. A smaller group of the participants will contribute further to this effort.
4. The material should be reviewed and supported by multiple DDI Alliance working groups to achieve a consistent from different perspectives. This will include:
    a. Training Group. The appropriate group for this subject
    b. Marketing Group. Announcing and promoting the library.
    c. Website Group. Adding self-guiding material to the website.
    d. Technical Committee. Making sure that there are no factual errors in the technical content.
5. A person in the role of an editor will be assigned to achieve consistency on multiple levels. This work includes tasks which are often not attractive for volunteers (in contrast to the work on the content). Ideally, this person will be paid by the DDI Alliance for a limited period of time.
6. Further communication between the workshop participants will be supported. The goal is to foster exchange between participants and possibly mutual review of future training material.
    a. The DDI Alliance Confluence wiki can be used as cooperation platform.
    b. An email list will be created.
    c. A bi-monthly phone call will be scheduled.
7. The material will be published with the Creative Commons Attribute license.
8. The material will be published in a way that it is citable. This is important for the material but also for the contributors. Some possible approaches: the DDI Alliance Working Paper Series or the Zenodo repository. Both would provide persistent identifiers, i.e. DOIs.
9. Additional multiple training material can be provided in a “contributed” section on the DDI Alliance website. This could include for example future slide decks of the participants for their own tutorials, and the larger slide deck on DDI Lifecycle of older GESIS training workshops at Dagstuhl.
10. A person will be assigned who is responsible for the overall organization. Some options seem to be the DDI 4 project manager (with a future task), somebody of the Training Group, or the suggested editor.

Appendix 5

DDI Alliance – Scientific Board
Goals 2019/2020

Joachim Wackerow, Chair of the Scientific Board, 2019-05-21

This is part of a larger plan. The proposals comprehend the areas of DDI training, a new acting committee of the scientific board, interoperability, DDI 4, and project management. Each item has a relationship to a topic in the strategic plan.

DDI 4

This item is related to the strategic plan topic "Enable DDI specs to adapt to changes in information technologies and bindings" in the section "Standards and Work Products".

DDI 4 Core (working name)

The work on DDI 4 Core (conceptual, data description, process) is ongoing in the MRT group (Modeling, Representation, and Testing). The release of DDI 4 Core is roughly planned for the end of 2019. A second sprint (after the NADDI sprint) is important to support this work. A sprint is planned in Dagstuhl, Sep 30 – Oct 4, 2019.

After the release of DDI 4 Core, it is planned to add extension for additional areas like data capture, provenance, and qualitative. The kind of additions needs to be determined. A related sprint could be organized in the margins of the IASSSIST conference 2020 in Gothenburg.

Longer descriptions on DDI 4 Core and the MRT group are available on the related wiki page.

OWL/RDF Representation

The OWL/RDF representation of the DDI 4 model is an important goal of DDI 4. This work was planned in two steps, first step are the rules on the transformation and the creation of a related tool, second step is the integration of established RDF vocabularies in the DDI 4 OWL/RDF representation. This is common practice in the Semantic Web and will assure that DDI 4 can be well received in the Linked Data community. The first step was achieved by the work of a paid consultant in 2018.

The second step – integration of established RDF vocabularies – is the next planned task in collaboration with experts and possibly a consultant.

Appendix 6

DDI 4 Core Overview and Scope[1]

Purpose

Note on Process

The work on producing the DDI 4 Core was launched so that, following the Prototype review, some of the core features of the DDI 4 work could be made ready for production release, recognizing that with available resources a more narrow scope was desirable. Emphasis was placed on short-term delivery: the Modeling, Representation, Testing (MRT) working group has allowed itself a year to complete the work on the initial core release, with delivery of a final deliverable ready for review and release at the end of December 2019.

Goals

DDI has always faced the requirement of dealing with a large range of data, both for archival purposes and to provide support for the entire production lifecycle to large studies and statistical agencies. The result of this work is a model which in many important respects is domain-independent. Recent developments in the research world are placing a greater emphasis on cross-domain integration of data, and data coming from new sources, some of it in unfamiliar forms (e.g., "big data," social media, sensor data, etc.). Social Science is no different from other domains – the DDI community is faced with a requirement for a more flexible ability to describe and manage, now available in a wider variety of forms.

The DDI 4 Core is intended to provide useful functionality in response to this requirement. In presentations, review comments, and discussions it has become apparent that some aspects of the DDI 4 model included in the DDI 4 Prototype are of especially high value. Identified features include the conceptual aspects of variables and classifications, the datum-oriented description of data, and the use of the process model to describe data lineage (the processing involved in the provenance of data). These same features have been identified as of interest by participants at recent Dagstuhl workshops on the subject of cross-domain data integration, with a further emphasis on alignment with external standards and the use of a UML model as a primary deliverable.

The DDI 4 Core will include not only the XML and RDF syntax representations, but will also deliver the UML from which they are derived in the form of Canonical XMI, a portable, tool-independent expression of the model. This not only makes DDI available for representation in other syntaxes and systems, but provides a stable basis for the maintenance of the model into the future.

User-oriented subsets of the DDI 4 model are provided by the inclusion of Functional Views, organized to support the application of DDI to specific tasks. This approach was employed in the Prototype, and will be carried forward into DDI 4 Core in a refined form, having both a technical and documentary

[1] The term "DDI 4 Core" is used throughout as the current working name. The issue of a re-naming to bring the work in-line with other DDI products is discussed in the text.

expression. In addition, high-level documentation aimed at introducing the model to adopters has been added. Together, these should make the DDI 4 Core more adoptable and easier to approach.

Because of the use of DDI 4 Core for cross-domain integration, and for other purposes, some key external standards have been selected as candidates for specific alignment (PROV-O for provenance, GSIM as a higher-level model of core metadata, GSBPM as a framework for process description, and DCAT for data discovery). Documentation of alignment with this small set of selected standards will be part of the deliverable package. The use of existing RDF vocabularies in the RDF syntax representation of the model is anticipated, as a needed feature of alignment with standards/best practice in the Linked Data domain.

The idea that DDI 4 Core be re-branded to reflect its intended use has been discussed: DDI-Codebook and DDI-Lifecycle have brands which reflect their intended use, while DDI 4 Core does not. A re-branding would communicate to users the purpose of the new release, and minimize confusion as to which version of the standard is best suited for their applications - currently, the use of the version number indicates an erroneous relationship between versions 3 and 4 which is causing some confusion among the potential users of the new standard release, as it did when earlier releases were referred to as “DDI 2” and “DDI 3”. (Suggestions have been along the lines of "DDI - Integration," "DDI - Cross-Domain," etc. Identifying a better name for the DDI 4 Core will need further exploration and conversation with the Marketing group and others.)

Please see the following page for a list of features.

Features

The DDI Core will have the following specific features:

- Conceptual and data description
    - Ability to describe variables as they are used and re-used at different process stages (the "variable cascade") with a strong connection to a conceptual layer
    - Ability to describe traditional and non-traditional data formats (unit-record/rectangular data files, multi-dimensional/aggregate data, event-history ("tall-skinny") data, and other data formats (analysis of no-SQL is on-going)
    - Ability to describe classifications as used in description of data
- Process description
    - The ability to describe in detail the processes to which data have been subject, as an aspect of data provenance ("data lineage"); data capture and methodology are not described in detail
- User-oriented subsets
    - Definition of Functional Views/subsets intended for specific applications, in the form of relevant technical specification and documentation
- Technical Deliverables
    - A UML model expressed in Canonical XMI, a portable flavor of that standard which is supported by a large number of tools which process UML. (This will support various validation and further processing like the use of DDI in program syntaxes as automatically generated by some UML tools.)
    - XML Schema for use in XML implementations for preservation and exchange
    - OWL specification with supporting validation resources (i.e., ShEx) for use in RDF implementations for discovery in the Web of Linked Data
- Documentation Deliverables
    - Complete documentation of the syntax representations for XML Schema and OWL
    - Complete documentation of the classes and relationships in the UML model
    - High-level documentation, including an explanation of design, features, and scope, and introductory material for the major features of the model (the variable cascade, the datum-based data description, and the process model for describing provenance/lineage)
- Documentation for implementation, alignment, and/or integration with external standards (PROV-O, DCAT, GSBPM, GSIM)

Appendix 7

MRT: DDI 4 Core Definition of Scope

Ottawa Sprint, 22-24 April 2019

I. Overview

This document provides a description of the scope of the work in the MRT Working Group on the DDI 4 Core model. This will be the content of a first production release, following the prototype review of 2018.

The intention is to take a subset of the overall model, and incorporate feedback and comments from the prototype review, so that a useful set of functionality can be brought to market in a short timeframe.

The release date for the DDI Core model and bindings (XML and RDF) is the end of December, 2019. At this time, it is intended that a final version of the DDI 4 Core specification be available for review, and that a production release would follow after issues raised during a subsequent review have been addressed.

The portions of the model which have not been included in the core release will be addressed in subsequent work. The core is meant to provide support for some of the most common functions for which DDI is intended to be used. Significant areas of functionality are not addressed (Data Capture is a good example of this – there are others). While also extremely important, a more incremental approach to providing support for these functions has been chosen.

The working style of the MRT will be to create an initial core release, and to work iteratively to support additional functionality in subsequent releases. Testing and implementation will form part of this cycle. Thus, each delivered part of the model will need to support a meaningful set of functionality from the user's perspective.

For the DDI 4 Core, the chosen functionality includes data description. It is the nature of the DDI model that this central functionality requires a substantial amount of foundational material, addressing data structures at the logical and physical level, conceptual material, representations of variables, and the relationships of variables at several levels.

The DDI 4 Core will also include support for some applications of the process model. Specific cases regarding data lineage and production processes will be specified. Because neither Data Capture nor Methodology are to be undertaken at this time, however, the application of the process model will be somewhat limited. The fuller use of this model will be included in future releases, but initial implementation of the Data Management View from the prototype release have shown that there is immediate utility in including at least some applications of the process model in the initial DDI 4 Core release.

II. The DDI 4 Lion Repository and DDI 4 Core

The DDI 4 work has been using the Lion Repository as a collaborative working platform for the work leading up to the prototype release. The MRT work will focus on a narrower scope and will be using a different modelling platform.

The modelling environment and production platform used for the MRT work has been integrated with that used by the Technical Committee, so that a single production flow will exist at the time that the MRT work goes into production. In most cases, the same tools are being used for producing bindings and documentation, and relevant artefacts can be passed back and forth between these teams without any difficulty.

As a first step, the complete DDI 4 model as contained in the Lion Repository will be extracted and then further refined to contain only those portions of the model identified in this scope document.

III. Supported Functionality

Pending a more complete specification of user requirements, the functionality to be supported in the DDI 4 Core model has been decided through discussion among those concerned in the work. This is seen as acceptable because up to this point formal requirements have not been a major feature of DDI development. It is seen as a potentially desirable activity in future, as it would help to inform the further development of DDI work products generally, the DDI 4 Core included, but this effort is not part of the scope of the current MRT work at this time.

There are some features which are seen as easy to include and of sufficient utility to users to be selected, even though they may not represent the more complete coverage of these topics in the full DDI 4 model. In this sense, the scope of the work has been determined opportunistically. The goal is to maximize the utility of the DDI 4 Core release to the user community.

An example of this is seen with methodology. The full DDI 4 model contains both a Methodology Pattern and an implementation of it for Sampling. These have not been included in the DDI 4 Core. Because of the strong relationship between process description and methodology, however, some aspects of methodology are required for a useful implementation of the process model. These have been restricted to summary descriptions and external references for the immediate term.

It is anticipated that those interested in cross-domain data integration will be potential users of the new DDI 4 Core, as has been suggested by discussions at the Dagstuhl workshops on this subject. For this audience, the description of data, foundational metadata, and process description are all of interest. Alignment with some other popular standards is also important in engaging these users.

The following list provides a summary of the functionality which will be supported as forming the most useful core functionality for DDI users:

- **Describing data** – users will be able to describe data sets (variables, representations, associated definitions/concepts, identification). This covers the data dictionary portion of a codebook.
- **Describing concepts and their relationships (in non-data-description) roles** – concepts and concept systems are used in many specific ways related to data, but not specifically in the roles

- of categories or variables. Examples are their use in specialized geography descriptions, describing coverage, and supporting data discovery.
- **Describe logical organization of data** – the logical organization/structuring of data in several common forms (rectangular/unit-record data, cube/aggregate/time series data, event data) with reference to the logical content, but independent of how it might be physically formatted/stored.
- **Describe physical organization** – the physical organization/structure of data.
- **Description of transformation/relationships between data organization styles, to accommodate different uses of the data** – the use of the datum as a pivot point between different styles of data organization/structure, informing how the same datums can perform different roles in different uses of the data (flexible viewpoints used to assign roles to different variables for different purposes).
- **Describing data lineage processes** – the actions performed on data set to produce related data, with summary information regarding methodology and data capture, but without a full description of these in a detailed way (this is an extension point for further work).
- **Describing collections of items and their internal structures/relations** – A general pattern of collections will be presented, along with some specific implementations of the pattern as relevant to support other stated functionality.
- **Alignment with selected common standards** – [Add ISO 19115 family] Some specific standards have been identified which will be used in combination with DDI. The alignment with these standards needs to be supported, as appropriate to their intended use. These include the DCAT vocabulary, the PROV-O vocabulary, and the GSBPM. In some cases (PROV-O, DCAT) there are hand-off point between the standards which should be identified to enable their combined use. In the case of GSBPM, the description of data lineage might consist of a process which navigates the GSBPM framework, in which case they are complementary descriptions of the same basic information, used for different purposes (DDI for processing/documentation, GSBPM for communication).

IV. DDI 4 Core Packages and Classes

This section details the contents of the initial package pulled from the Lion Repository, to include *at least* those needed to support the functionality described above. This section is intended to rercord what was done, and to facilitate further work on the DDI 4 Core model, which will start from the set of classes listed, but will very likely exclude many of them if they are unused. Open issues where new classes will likely be needed are indicated in the first section – these are the subject of analysis proposals found in other working MRT documents.

A. Open Issues for December Release

From Datum-based Examples work:

- For Tall format Value columns, we need a generic datatype to describe datums from different instance variables

- For Tall format VariableRef columns, we need to be able to reference Value Mappings (and by association to Instance Variables)
- For Viewpoints applied to Data Cubes, we need to add a **dimension** role.
- Should a variable be allowed to have more than one role in the same Viewpoint?

From other places:

- How to handle Annotations and Citation Information if Identifiers change as a result of modelling work?

As an initial activity, the contents of the Prototype were reviewed and an initial pull was made from the Lion Repository to include *at least* the full set of classes to be included in the DDI 4 Core. (Because changes are being made to resolve modeling issues this includes some classes on which there are currently dependencies which may or may not be included in the published DDI 4 Core.)

At the same time, a more complete extraction was made from the Lion Repository of all of the Prototype Review classes as well as other parts of the model which were not published for review.

B. Initial DDI 4 Core Extraction

The following table is a list of all classes extracted for inclusion in the initial pull, organized as an alphabetical inventory. Note that many of these classes are included as a result of dependencies on other objects in their packages, and that the contents of the production DDI 4 Core model are likely to be different.

Class Name	Definition (Partial)	Package
Access	Describes access to the annotated object. This...	Discovery
Act	An Act is an indivisible, atomic step, i.e....	Workflows
Agent	An actor that performs a role in relation to a...	Agents
AgentListing	A listing of Agents of any type. The AgentList...	Agents
AgentRelationStructure	Defines the relationships between Agents in a...	Agents
Algorithm	An algorithm is an effective method that can...	MethodologyPattern
AlgorithmOverview	High level, descriptive, human informative,...	SimpleMethodologyOverview
AnnotatedIdentifiable	Used to identify objects for purposes of...	Identification
AppliedUse	Links the guidance instructions to specific...	Methodologies
AttributeRole	An AttributeRole identifies one or more...	LogicalDataDescription
AuthorizationSource	Identifies the authorizing agency and allows...	Representations

BoundingBox	A type of Spatial coverage describing a...	Discovery
BusinessAlgorithm	A Business Algorithm is used to express the...	BusinessWorkflow
BusinessFunction	Something an enterprise does, or needs to do,...	Methodologies
BusinessProcess	BusinessProcesses could be Generic...	BusinessWorkflow
Category	A Concept whose role is to define and measure...	Conceptual
CategoryRelationStructure	Relation structure of categories within a...	Conceptual
CategorySet	Specialization of a Concept System focusing on...	Conceptual
ClassificationFamily	A Classification Family is a collection of...	Representations
ClassificationIndex	A Classification Index is an ordered list (...	Representations
ClassificationIndexEntry	A Classification Index Entry is a word or a...	Representations
ClassificationItem	A Classification Item represents a Category at...	Representations
ClassificationItemRelationStructure	A complex RelationStructure for use with...	Representations
ClassificationSeries	A Classification Series is an ensemble of one...	Representations
ClassificationSeriesRelationStructure	Describes the complex relation structure of a...	Representations
Code	A type of Designation that relates a...	Representations
CodeList	A list of Codes and associated Categories. May...	Representations
CodeRelationStructure	Relation structure of codes within a codelist...	Representations
CollectionMember	Generic class representing members of a...	CollectionsPattern
Comparison	The minimal pattern for a comparison including...	CollectionsPattern
ComparisonMap	Provides a basic pattern for a comparison map...	CollectionsPattern
ComputationAction	Provides an extensible framework for specific...	Workflows

Concept	Unit of thought differentiated by...	Conceptual
ConceptRelationStructure	Relation structure of concepts within a...	Conceptual
ConceptSystem	A set of Concepts structured by the relations...	Conceptual
ConceptSystemCorrespondence	Relationship between 2 or more ConceptSystems...	Conceptual
ConceptualDomain	Set of Concepts, both sentinel and substantive...	Conceptual
ConceptualVariable	The use of a Concept as a characteristic of a...	Conceptual
ConditionalControlStep	Type of WorkflowControlStep in which the...	Workflows
ControlledVocabulary	The specification of a controlled vocabulary...	CustomMetadata
CorrespondenceTable	A Correspondence Table expresses relationships...	Representations
Coverage	Coverage information for an annotated object....	Discovery
CustomInstance	A set of CustomValues to be attached to some...	CustomMetadata
CustomItem	A custom item description. This allows the...	CustomMetadata
CustomItemRelationStructure	Contains a set of CustomItemRelations which...	CustomMetadata
CustomStructure	A Collection containing a set of item...	CustomMetadata
CustomValue	An instance of a key, value pair for a...	CustomMetadata
DataPipeline	A DataPipeline is a single traversal of the...	BusinessWorkflow
DataPoint	A DataPoint is a container for a Datum.	LogicalDataDescription
DataPointRelationStructure	A means for describing the complex relational...	FormatDescription
DataStore	A DataStore is either a SimpleCollection or a...	LogicalDataDescription
DataStoreLibrary	A DataStoreLibrary is a collection or, again,...	LogicalDataDescription
DataStoreRelationStructure	A structure for describing a complex relation...	LogicalDataDescription

Datum	A Datum is a designation (a representation of...	LogicalDataDescription
Design	The design pattern class may be used to...	MethodologyPattern
Designation	A sign denoting a concept.	Representations
DesignOverview	High level, descriptive, human informative,...	SimpleMethodologyOverview
EnumerationDomain	An abstract base to allow all codifications (...	Representations
ExternalMaterial	ExternalMaterial describes the location,...	Utility
GeographicExtent	Defines the extent of a geographic unit for a...	GeographicClassification
GeographicUnit	A specific geographic unit of a defined Unit...	GeographicClassification
GeographicUnitClassification	Describes the classification of specific...	GeographicClassification
GeographicUnitRelationStructure	Defines the relationships between Geographic...	GeographicClassification
GeographicUnitTypeClassification	A structured collection of Unit Types defining...	GeographicClassification
GeographicUnitTypeRelationStructure	Defines the relationships between Geographic...	GeographicClassification
Goal	Goals are the "things" a method aims to...	Methodologies
Guide	Provides a guide for the usage of a result...	Methodologies
Identifiable	Used to identify objects for purposes of...	Identification
IdentifierRole	An IdentifierRole identifies one or more...	LogicalDataDescription
IfThenElse	IfThenElse describes an if-then-else decision...	Workflows
IndexEntryRelationStructure	Structures relationship of Classification...	Representations
Individual	A person who may have a relationship to...	Agents
InstanceVariable	The use of a Represented Variable within a...	Conceptual
InstanceVariableRelationStructure	A realization of RelationStructure that is...	LogicalDataDescription
LevelStructure	The LevelStructure describes the nesting...	Representations
LogicalRecord	The LogicalRecord is a record definition. It...	LogicalDataDescription

LogicalRecordRelationStructure	Allows for the complex structuring of...	LogicalDataDescription
Loop	Iterative control structure to be repeated a...	Workflows
Machine	Mechanism or computer program used to...	Agents
MeasureRole	A MeasureRole identifies one or more...	LogicalDataDescription
MemberIndicator	Provides ability to declare an optional...	CollectionsPattern
MemberRelation	Defines one kind of relationship between one...	CollectionsPattern
MetadataDrivenAction	MetadataDrivenActions are Acts in which their...	Workflows
Methodology	Methodology brings together the design,...	MethodologyPattern
MethodologyOverview	High level, descriptive, human informative...	SimpleMethodologyOverview
Organization	A framework of authority designated to act...	Agents
Parameter	An Input or Output to a Process Step defined...	Workflows
PhysicalDataSet	The information needed for understanding the...	FormatDescription
PhysicalLayoutRelationStructure	A realization of RelationStructure that is...	FormatDescription
PhysicalOrderRelationStructure	PhysicalStructureOrder orders...	FormatDescription
PhysicalRecordSegment	A description of each physical storage segment...	FormatDescription
PhysicalSegmentLayout	The PhysicalSegmentLayout is an abstract class...	FormatDescription
PhysicalSegmentLocation	Among other things defines the location of a...	FormatDescription
Population	Set of specific units (people, entities,...	Conceptual
Precondition	A precondition is a state. The state includes...	Methodologies
Process	Process is an implementation of an algorithm....	ProcessPattern
ProcessControlStep	A Process Step that controls the ordering of...	ProcessPattern

ProcessOverview	Process is an implementation of an algorithm....	SimpleMethodologyOverview
ProcessStep	One of the constituents of a Process. It can...	ProcessPattern
RecordRelation	The RecordRelation object is used to indicate...	LogicalDataDescription
RelationStructure	The set of MemberRelations used to structure a...	CollectionsPattern
RepeatUntil	Iterative control structure to be repeated...	Workflows
RepeatWhile	Iterative control structure to be repeated...	Workflows
RepresentedVariable	A combination of a characteristic of a...	Conceptual
Result	Describes the results of a process for the...	Methodologies
SegmentByText	Defines the location of a segment of text...	FormatDescription
SentinelConceptualDomain	Description or list of possible sentinel...	Conceptual
SentinelValueDomain	The Value Domain for a sentinel conceptual...	Representations
Service	A means of performing a Process Step as part...	ProcessPattern
Sign	Something that suggests the presence or...	SignificationPattern
Signified	Concept or object denoted by the signifier...	SignificationPattern
Signifier	Concept whose extension includes perceivable...	SignificationPattern
SimpleCollection	Simple Collection container (set or bag) that...	CollectionsPattern
SpatialCoverage	A description of spatial coverage (geographic...	Discovery
Split	The components of a Split consists of a number...	Workflows
SplitJoin	SplitJoin consists of process steps that are...	Workflows
StandardWeight	Provides an identified value for a standard...	SimpleCodebook
StatisticalClassification	A Statistical Classification is a set of...	Representations

StatisticalClassificationRelationStructure	A structure for describing the complex...	Representations
StructuredCollection	Structured Collection container extends a...	CollectionsPattern
SubstantiveConceptualDomain	Set of valid Concepts. The Concepts can be...	Conceptual
SubstantiveValueDomain	The Value Domain for a substantive conceptual...	Representations
TemporalCoverage	Describes the temporal coverage of the...	Discovery
TemporalRelationControlStep	Defines complex synchronous or asynchronous...	Workflows
TopicalCoverage	Describes the topical coverage of the module...	Discovery
Unit	The object of interest in a process step...	Conceptual
UnitDataRecord	Gives a UnitDataRecord structure to a Logical...	LogicalDataDescription
UnitDataViewpoint	The assignment of measure, identifier and...	LogicalDataDescription
UnitSegmentLayout	UnitSegmentLayout supports the description of...	FormatDescription
UnitType	A Unit Type is a type or class of objects of...	Conceptual
Universe	A defined set or class of people, entities,...	Conceptual
ValueAndConceptDescription	A formal description of a set of values.	Representations
ValueDomain	The permitted range of values for a...	Representations
ValueMapping	Provides physical characteristics for an...	FormatDescription
VariableCollection	A collection (group) of any type of Variable...	Conceptual
VariableRelationStructure	RelationStructure for use with any mixture of...	Conceptual
VariableStatistics	Contains summary and category level statistics...	SimpleCodebook
ViewpointRole	A ViewpointRole designates the function an...	LogicalDataDescription
VocabularyEntry	One entry term and its definition in an...	CustomMetadata
VocabularyRelationStructure	Contains the Vocabulary Relations used to...	CustomMetadata

WorkflowControlStep	A subtype of WorkflowStep which controls the...	Workflows
WorkflowProcess	A Workflow Process is a realization of Process...	Workflows
WorkflowService	A means of performing a Workflow Step as part...	Workflows
WorkflowStep	One of the constituents of a Workflow. It can...	Workflows
WorkflowStepSequence	A WorkflowStepSequence controls the order of...	Workflows

C. Prototype Review Classes Excluded from the DDI 4 Core Initial Extraction

This section describes which packages, classes, and relationships, included in the Prototype Review, were not included in the initial pull from the Lion Repository. These were manually trimmed out of the initial pull as they would result in the inclusion of large numbers of unwanted classes.

Relationships:

From InstanceVariable:

Removed for Core Build as no capture information in included in core:

Relationship: sourceCapture (target Capture) - The source of capture for the values that populate this InstanceVariable. This may be any class using Capture as an extension base, currently RepresentedQuestion and RepresentedMeasurement. This is the direct source of capture for the content of the data. If a transformation is required the source is not multiple captures but the transformation.

From DataStore:

Removed the following content in preparation for Core Build:

Relationship: isInStudy (target Study) - A Study has at most one DataStore. Many studies can have the same set of record types.

From DataStoreLibrary:

Removed following from Core Build:

Relationship: isInStudySeries (target StudySeries) - Whereas a DataStore is associated with a Study, a DataStoreLibrary is associated with a StudySeries. Each StudySeries has at most one DataStoreLibrary.

From DataPipeline:

Removed the following for Core Build:

Relationship: isInStudy (target Study) - A study has at most one DataPipeline. The same DataPipeline can be used in many studies.

Packages:

The following packages found in the Prototype have been removed from Core Build

DataCapture

SamplingMethodology

StudyRelated

Classes:

The following classes have been removed:

From DataCapture:

BooleanResponseDomain

Capture

CodeListResponseDomain

ConceptualInstrument

ExternalAid

ImplementedInstrument

InstanceMeasurement

InstanceQuestion

Instruction

InstrumentCode

InstrumentComponent

NumericResponseDomain

RankingResponseDomain

RepresentedMeasurement

RepresentedQuestion

ResponseDomain

ScaleResponseDomain

Statement

TextResponseDomain

From SamplingMethodology:

SampleFrame

SamplePopulationResult

SamplingAlgorithm

SamplingDesign

SamplingGoal

SamplingProcedure

SamplingProcess

From StudyRelated:

Budget

ComplianceStatement

Embargo

ExPostEvaluation

QualityStatement

Standard

Study

StudyControl

StudyRelationStructure

StudySeries

From Utility:

DocumentInformation

FundingInformation

The following StructuredDataTypes have been moved to the unpublished package ClassesOnHOLD:

ConditionalText

ContentDateOffset

DynamicText

DynamicTextContent

FixedText

LiteralText

StudyIndicator

TargetSample

V. Post-December Work: Next Steps

Some important parts of the DDI model have been excluded from the DDI 4 Core for the reasons given above. It is recognized that integrating these parts of the model is important as a next step, after the December 2019 delivery, so that the corresponding user functionality is supported. In discussion, these functional areas included the following:

- **Data Capture –** The collection of data from questionnaires, registers, and other sources is a major use-case for DDI. As modelled in the prototype, this is an important implementation of the process-related portions of the model. Although process is included, as implemented in some other applications, data capture is an important and high-priority addition. Some work has already been done on the integration of the data capture model with the data description model – this needs to be finalized and agreed and is estimated to be a significant amount of work.
- **Methodology –** The prototype contained both a pattern for describing methodology, and a set of concrete classes which are related the pattern. From these, a model exists specifically for sampling. More complete support for methodology is contemplated, with specifics and priority yet to be determined.
- **Study Description (Purpose, Data Dictionary, Related Information) -** The prototype contained a model of the Study which is the basis for many of the real-world "codebook" implementations of DDI seen in previous versions of the standard. While the DDI 4 Core contains the heart of the information needed (the data dictionary portion) it does not cover some other types of information needed to represent codebooks. For migration purposes, this is an important model to finalize. Note, however, that it is a non-trivial effort, as the range of codebook metadata expressed in earlier versions of DDI is extensive.

Other areas could also be usefully included in a next release – identifying these is not within the scope of current MRT efforts.

VI. Post-December Work: Longer-Term Functionality

There are many areas which have been considered for inclusion in DDI 4 over the course of its development. Of these, some have been the subject of considerable work, and others have not gone much beyond the planning stages.

Identifying these is not within the scope for current MRT efforts but should be considered at this point in general terms, to make sure that future directions are not in some way unintentionally limited. One activity has been to make sure that any of the content within the Lion Repository which may in future become part of a production DDI 4 model has been preserved, in the case that the Lion Repository itself is taken offline.

Functionality mentioned here may also become the focus of more immediate work – it was not included in the Prototype Review package, and so is assumed to be of lower priority for release. It has been the focus of much past work, however, and represents a significant resource should it become a priority for inclusion.

Qualitative Data Description – a great deal of work has gone into describing the different types of metadata associated with non-quantitative data, including the addressing of segments in various types of files such as images, interview transcripts, videos, etc. Linking various types of information to these, and further documenting the process by which they are quantified and further analyzed is also supported. The interest in standardization in this area has recently become a topic of discussion within the Social Sciences and related domains, reflecting renewed interest. The REFI standard (an effort driven by QDA) is the most prominent example of this: https://www.qdasoftware.org/products-project-exchange/ .

Appendix 8

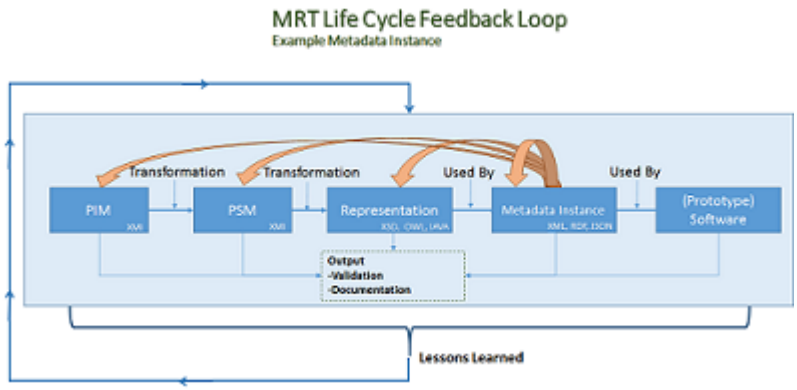
MRT - Modeling, Representation, and Testing Lifecycle Working Group

Created by Hilde Orten
Last updated May 29, 2019 by Jon Johnson

Purpose and goal of the MRT Working Group

The MRT working group focuses on the iterative lifecycle of modeling, representation, and testing. Further information is available in the document "MRT - Modeling, Representation, and Testing Lifecycle: A Proposed Working Group for Building DDI 4 Core".

The overview and scope of DDI 4 Core (working name) is described in a separate document.



MRT MINUTES

MRT TASKS

MRT MILESTONES

Team members and contact

Arofan Gregory	ilg21@yahoo.com
Dan Gillman	Gillman.Daniel@bls.gov
Flavio Rizzolo	flavio.rizzolo@statcan.gc.ca or flavio.rizzolo@gmail.com
Hilde Orten	hilde.orten@nsd.no

Meeting time and details

Regular meeting time for M
The meeting will take place
Wednesday at 0700 Califor
Minneapolis, 1000 Ottawa,
0200 Canberra.

Meeting dates: Weekly

Instructions to join the mee

1. Please join my meeting.
https://www3.gotomeeting

2. Use your microphone an
headset is recommended. (
telephone.

United States: +1(646)749-

Australia: +61283551035

Belgium: +32(0)42680154

Canada: + 1(647)497-9373

Denmark: +45 (0) 89 88 04

France: +33 (0) 182 880 93

Germany: +49 (0) 811 8899

New Zealand: +64 (0) 9 80

Norway: +47 21 01 89 06

Sweden: +46 (0) 852 500 2

United Kingdom: +44 (0) 2

Access Code: 731-015-190
Audio PIN: Shown after joi
Meeting ID: 731-015-190

Jay Greenfireld	greenfield_jay@bah.com
Joachim Wackerow	Joachim.Wackerow@gesis.org
Jon Johnson	jon.johnson@ucl.ac.uk
Larry Hoyle	larryhoyle@ku.edu
Oliver Hopt	oliver.hopt@gesis.org
Wendy Thomas	wlt@umn.edu

No labels

1 Comment

Hilde Orten

Under construction

Reply • Feb 08, 2019

Appendix 9

DDI Specifications Roadmap (Rationale and Considerations)

DDI Technical Committee (2 April 2019)

Background

This document updates the DDI Roadmap document prepared by the Technical Committee in mid-2017 and expands its coverage to also include DDI-C in light of the changes that have taken place in the intervening 18 months. The original DDI Roadmap was approved by the Executive Committee in October 2017.

DDI CodeBook

A substantial proportion of the content of metadata in DDI is in DDI-C. For the organisations which rely upon it and for many potential users who wish to document and discover data, DDI-C is more than sufficient to meet their needs, its simplicity, its ease of implementation and its wide-scale adoption are strong reasons to still advocate its usage for many data management, discovery and documentation purposes.

Surprisingly, for a standard that has been in existence for such a long period of time it is still attracting new users and existing users are (re)developing software to support the functionality that it enables. This ranges from organisations whose content does not easily support transformation to a more complex standard such as DDI-L to those such as Cornell who have developed software to decouple their data ingest pipeline from commercial vendors. For many scenarios a structured Codebook is a major step forward in ensuring the longevity of data over proprietary formats. In particular it is excellent for independent or student researchers who lack the infrastructure needed to support the use of more complex standards.

The major platform (Nesstar) upon which DDI-C (using a restricted version of the standard) has gained significant traction in the archives and for Low Middle Income Countries (LMIC) data collection is likely to be at the end of its useful life. Dataverse is replacing Nesstar in many organisations (including member organisations) in order to support the business functionality. For those organisations whose business needs are met by Dataverse, this makes sense, for those whom it does not, they have a decision point on whether to move to a new iteration of the standard if there is suitable software or stay with DDI-C and redevelop their systems.

DDI-C is still being actively being used and relied upon by a significant user base. DataPASS recommends a profile of DDI-C for its users. In addition, The World Bank has developed a full capability editor (currently awaiting release), and is also pushing forward further development of NADA[1] and supporting CAPI collection software[2]. This will almost inevitably lead to further demands to support enhancements to DDI-C.

Planned Development of DDI-C

DDI-C will need some minor updates, in particular to support changes to related standards such as Dublin Core for existing users and the provision of better support for use in Dataverse. Specific issues regarding

[1] https://github.com/ihsn/nada
[2] https://mysurvey.solutions/

improved compatibility with DataVerse and interaction with statistical software have been filed in the DDI-C issue tracker.

DDI-Lifecycle

DDI-L has gained significant traction, especially since the release of 3.2. The capability to do questions and a number of major bug fixes has attracted new users especially amongst longitudinal data resources and software from both commercial and open source communities.

Alongside these new users there has been substantial investment from individual studies (e.g. ESS, NHATS, MIDUS), consortia (e.g. CLOSER, CESSDA) and archives (e.g. GESIS) reflecting the better support for data management and opening up new possibilities such as the creation of genuinely usable question banks, that starts to solve major infrastructural issues which have bedevilled survey research for decades.

The uptake in use of DDI-L by NSO's (e.g. Insee, CSO, New Zealand and Danish Stats) has been possible because the standard has evolved to meet their specific use cases and that it can be delivered on existing technology stacks and integrated into backend systems, support for Neuchatel, better alignment with GSIM and enhanced questionnaire support for grids in the upcoming release 3.3, is a reflection of that engagement.

A best practices DDI-L document has been issued and continues to be updated.

The public review of DDI 3.3 has concluded and the Technical Committee are currently assessing the feedback and focused on getting that out as a final version in late-Spring 2019.

At the EDDI 2018 meeting the Technical Committee worked on the detailed planning, identified gaps and dependencies on establishing an automated production process based on a model using DDI 3.3 as the test bed.

Further Development of DDI-L

DDI-L 3.4

Pending community feedback it is envisaged that a further version DDI-L 3.4 which will contain:

- Bug fixes from DDI-L 3.3
- Structural revisions to the DDI-L 3.3 model to eradicate inconsistencies etc.
- A UML model that takes the payload of 3.3 and revises the structure of the model to reflect approach of Moving Forward Project
- Flattened XML schema
- Documentation as a PDF generated from source controlled restructured text files.
- OWL / RDF output

It is not envisaged that DDI-L 3.4 will be available as XML schemas in the same flavour as previous versions of DDI-L.

DDI-L 3.4+

Versions after 3.4 would start to incorporate work from DDI4 in a more formal manner, although as previously noted many aspects will already have been included.

In preparation for this, we would suggest that future sprints and virtual meetings:

- Determine strategy for a single production model from Drupal and COGS experience

- Perform gap analysis between 3.3 and Moving Forward content (focus on points of difference and resolve those differences in both naming and structure)
- Perform gap analysis between combined content of 3.3 and Moving Forward against GSIM
- Develop the work on mapping DDI OWL and other ontologies
- Technical Committee consult and implement a strategic work plan

Allied Standards (DISCO and XKOS)

Recent activity on resolving the final issues in DISCO may result in the publication of an official version in the coming year. XKOS has completed its final review and is being prepared for publication of a final, official version.

DDI-4

The EDDI 2018 meeting which combined the Moving Forward project and the Technical Committee sought to take stock of the issues raised thus far identified in the public review of DDI4.

Discussions at the EDDI 2018 meeting included a suggestion for using the core features of the DDI 4 model that are the most robust to date, conceptual, data description, and process, as the focus of this approach for a period of one year. The result being a 'core' DDI 4 release that is implementable and the base on which to update the rest of the model.

Other discussions have cast doubt on whether other parts of the coverage of DDI should indeed be extended as 'other standards' do it better. This discussion needs a resolution, as it has a significant impact on the relationship between all versions of DDI.

The direction of DDI-4 is subject to the outcome of the Public Review and documents that have emerged from the EDDI 2018 meeting.

Production considerations

One of the objectives of the DDI4 project was a production pipeline which moved away from the hand crafting that both DDI-C and DDI-L has relied upon through the auto-generation of the serialisation of the standard from a model. The TC has at its 2017 meeting and with the assistance of Colectica a pipeline which achieves this for DDI-L 3.3. This has enables the production of prototype versions of serialisations other than XML which may be of utility to the community. Further work was completed on this at the 2018 EDDI Sprint and it is hoped that this will enable a wider audience for the standard. It is envisaged that the development of such a pipeline can be extended to DDI-C to manage this also.

A robust model for all iterations of the standard opens up the possibility of providing better documentation and processes to enable users to have a more seamless migration path between versions than we are currently able to provide.

Appendix 10

Future Direction on DDI 4 Background

- DDI 4 was started years ago with the ideas …
  - … to comprehend …
    - all features of all existing DDI specifications (DDI-C, DDI-L, Disco)
    - description of new data structures
    - description of new data sources beyond surveys
  - … to be driven by an UML model
  - … to have generated syntax representations
    - XML Schema for preservation and exchange
    - OWL /RDF for discovery and interoperability

The realization turned out to be ambitious.

Future Direction on DDI 4
Where are we now?

- Work on DDI 4 Core since December 2018
  - Goal is a stable and consistent core
    - Public review Jan 2020
    - Onto DDI 4 core later (2020+) things can be added on without changing the core.
  - MRT working group provides a more comprehensive perspective -- including modeling, representations (XML and RDF), and test instances.
  - Including conceptual, data description, process
  - Agnostic to the specific domain, i.e. the social sciences
    - Reflects recent more interdisciplinary approaches in research
  - The Core features have the potential for cross-domain use
- DDI 4 Prototype review up to March 2019

- In parallel: improvement of DDI-Lifecycle resulting in DDI-L 3.3

Future of DDI Developments
Resources of the DDI Alliance

- Financial resources coming from membership fees, resulting in approx. 100,000 USD per year
  - Major increase of membership is not realistic
- Volunteering work
  - In-kind contributions of representatives of member organizations
  - Contributions of interested experts
- Other in-kind contributions
  - Organization of training workshops
- Cooperations with other organizations?

The work in working groups and sprints rely mostly on **volunteering work**. This is the most **valuable resource**.

Future of DDI Developments Questions I

- Can the DDI Alliance afford to support development of multiple specification tracks in parallel?
  - DDI-Codebook, DDI-Lifecycle, DDI 4, Disco
  - Other products: Controlled Vocabulary, XKOS
- Are extensions to DDI 4 Core a good investment into the future?
- Is the architecture of DDI-Lifecycle (XML Schema) a good basis for future development?

Future of DDI Developments Questions II

- Which role play the investments in DDI-Lifecycle (software in organizations, Colectica)?
- Which role play the requests of organizations (i.e. NSIs) for DDI 4 features like description of new data types and data sources beyond surveys?
- Can the DDI Alliance afford to loose the resource of volunteering contributions?
  - i.e. 8 people are working on DDI 4 Core. Most of them wouldn‘t be available for other work.

Future of DDI Developments Possible Scenario

- Would a parallel development for a limited time make sense to combine efforts, to avoid loosing volunteering work, and to ease conflicts in order to serve multiple user needs?
  - DDI 4 Core plus future extensions
    - As DDI Integration or Interoperability
  - DDI-Lifecycle - for surveys, ease of transition to DDI 4, no major other developments
  - DDI-Codebook - maintenance
  - DDI Controlled Vocabularies - development and maintenance
  - XKOS - maintenance
  - Disco (publication soon) – no maintenance, integration of Disco features in DDI 4

Appendix 11

DDI Alliance – Scientific Board
Goals 2019/2020

Joachim Wackerow, Chair of the Scientific Board, 2019-05-21

The proposals comprehend the areas of DDI training, a new acting committee of the scientific board, interoperability, DDI 4, and project management. Each item has a relationship to a topic in the strategic plan.

DDI Training

This item is related to the strategic plan topic "**Training**" in the section "The DDI Alliance as an Organisation".

DDI Training Library

The work on a general purpose DDI Training Library has started in the Train-the-Trainers workshop in Dagstuhl 2018. The goal of the training library is to provide consistent and appealing teaching material. It should cover all relevant areas, follow a version-agnostic approach, and provide version-specific details.

The training library should form the basis of teaching tutorials to multiple audiences like tutorials in the margins of conferences or workshops in summer schools. It should be also the basis of online training material.

The current training library has still gaps in terms of covering all relevant areas and has limitations regarding content consistency and look-and-feel of the slides.

Organizing and editing work is crucial to achieve the described goal. This work should be done by a paid person. Volunteers can help in providing material for missing areas.

Details are available in appendix 1.

DDI Tutorials

In terms of maintaining a good user base and increasing the use of DDI, it is important to provide a continuous offer on DDI tutorials. The tutorials should be taught in the margins of conferences where the audience might be interested in DDI. It is even more important in terms of outreach to provide this at conferences where the audience is not really aware of DDI but might be interested. Examples are survey methodology conferences like ESRA and RC33 but there are much more opportunities. Furthermore, tutorials should be taught at summer schools of empirical social sciences. Examples are the ICPSR and GESIS summer schools. It seems to be important to support travel and accommodation of interested instructors. The details on how to offer these possibilities (yearly tutorial plan, tender for instructors) need to be determined.

Some conferences were sponsored in the past by the DDI Alliance in a general way as a form of outreach and marketing. This money could be used in a more focused approach for DDI tutorials. This way, both goals – outreach and training – could be achieved.

Scientific Board

This item is related to the strategic plan topic "**Business Structure**" in the section "The DDI Alliance as an Organisation".

Acting Committee of the Scientific Board

It is proposed to create an Acting Committee of the Scientific Board as a standing committee.

According to the bylaws, the Scientific Board is the scientific and technical body of the Alliance. It is composed of Member and Associate Member Organization Designated Representatives. It represents the level between the working level of the Sub-Committees (like the Technical Committee as standing committee) and Working Groups on specific topics, and the steering level of the Executive Board which is responsible for setting the overall policy and budget for the Alliance among other tasks.

The experience shows that the current set up - a Scientific Board of approx. 40 member representatives with the two elected chairs – has limited impact regarding the described role of the Scientific Board.

The proposal is to create an Acting Committee of the Scientific Board whose members are elected by the all members of the Scientific Board. Chair and vice-chair could be the elected members with most of the votes. Additionally, external assigned members should be part of this Acting Committee to assure additional perspectives from a larger community. The role of these external experts is inspired by the idea of a Scientific Advisory Board (SAB). This approach avoids inventing an additional committee like a SAB.

The Acting Committee should be formed by five elected member representatives and two external advisory experts.

The Acting Committee should meet monthly in a conference call. The Acting Committee should meet once the year in a face-to-face meeting (like one day). The face-to-face meeting should be in the margins of one of the conferences which are relevant for DDI like IASSIST, EDDI, and NADDI.

The details need to be discussed and agreed in the next couple of months. A related change of the bylaws might be required.

Roles of Member Representatives

The experience shows that a clarification regarding the roles of member representatives and scientific board representatives would be helpful. A related document should be created.

Voting in Committees and Working Groups

The DDI Alliance became larger and more heterogeneous. With this background, it doesn't seem to be possible to rely always on a decision mechanism by agreement (which was very successful in the past). Therefore clear voting mechanisms should be described for all committees and working groups. This can affect the decision on items where no agreement is possible and the process of assigning chairs.

Next Task

A related temporary working group should be established which creates detailed proposals for all three items.

Interoperability of Metadata Standards

This item is related to the strategic plan topic "**Engagement with Global Digital Research Infrastructure**" in the section "The DDI User Community".

A second workshop on "Interoperability of Metadata Standards in Cross-Domain Science, Health, and Social Science Applications" is planned in Dagstuhl, Oct 7-11, 2019. The workshop is organized again together with CODATA (Committee on Data of the International Council for Science). This workshop builds on the outcomes of the first Dagstuhl meeting in 2018, further exploring how metadata standards can best support interdisciplinary research projects. To achieve the intended objectives and to build on last year's work, there will be a shift in emphasis. The 2018 event identified commonalities between several relevant standards, charting a direction for future work. The focus in 2019 will be more technical, looking at detailed aspects of identified approaches, and with specific examples provided in the outputs.

This activity has the goal to establish DDI 4 in a cross-domain context. It makes sure that DDI can have a larger community in the ongoing development of data science. It assures also that DDI is not developed in an isolated environment. Furthermore, it can be understood as an outreach activity in terms of getting in touch with a wider metadata community beyond the traditional DDI archive and survey community.

More information is available on the related wiki page.

DDI 4

This item is related to the strategic plan topic "Enable DDI specs to adapt to changes in information technologies and bindings" in the section "Standards and Work Products".

DDI 4 Core (working name)

The work on DDI 4 Core (conceptual, data description, process) is ongoing in the MRT group (Modeling, Representation, and Testing). The release of DDI 4 Core is roughly planned for the end of 2019. A second sprint (after the NADDI sprint) is important to support this work. A sprint is planned in Dagstuhl, Sep 30 – Oct 4, 2019.

After the release of DDI 4 Core, it is planned to add extension for additional areas like data capture, provenance, and qualitative. The kind of additions needs to be determined. A related sprint could be organized in the margins of the IASSSIST conference 2020 in Gothenburg.

Longer descriptions on DDI 4 Core and the MRT group are available on the related wiki page.

OWL/RDF Representation

The OWL/RDF representation of the DDI 4 model is an important goal of DDI 4. This work was planned in two steps, first step are the rules on the transformation and the creation of a related tool, second step is the integration of established RDF vocabularies in the DDI 4 OWL/RDF representation. This is common practice in the Semantic Web and will assure that DDI 4 can be well received in the Linked Data community. The first step was achieved by the work of a paid consultant in 2018.

The second step – integration of established RDF vocabularies – is the next planned task in collaboration with experts and possibly a consultant.

Improving Infrastructure around all DDI Specifications

The item is related to the strategic plan topic “Improvement of interoperable and distributed DDI infrastructure for use and reuse of DDI resources” in the section “Standards and Work Products”.

- Best practices should be created on when to use which specification (and which part). This could be guided along use cases.
- Improving portability of DDI metadata, interoperability of DDI systems
    - Criteria of portability and interoperability, validation tools
    - DDI Profiles / Views for specific use cases / specific perspectives
    - Description of the workflow in integrating above items
- DDI URN resolution enables web-based reuse of metadata, i.e. persistent identifier to URL of DDI item.
- Standardized query/exchange protocol enables reuse of DDI items stored in local and remote repositories, common data element registries

The two latter items support the idea of reusable DDI metadata in organizations and across organizations. They are prerequisites for using efficiently DDI metadata repositories and registries.

The recently announced plan of the Technical Committee on web-based agency id resolution and service discovery system fits well into this.

Project Management

This item is related to the strategic plan section “The DDI Alliance as an Organisation”.

Project management seems to be important when looking on the various not always well coordinated activities of the DDI Alliance. This would assure that the activities are made in a consistent way following a plan which the Scientific Board and Executive Board agreed on. A paid projector manager or an in-kind contribution could help here. The realization doesn’t look too realistic because of the limited budget and the fact that this doesn’t seem to be a typical in-kind contribution.

Appendix

1 Plan for DDI Training Library

1. A modular training material body (the library) will be published for further reuse and self-guided training. This will include the content presently on the DDI Alliance website.
2. This training library is an official publication of the DDI Alliance. The workshop participants are contributors to this material. They should get the appropriate attribution.
3. A smaller group of the participants will contribute further to this effort.
4. The material should be reviewed and supported by multiple DDI Alliance working groups to achieve a consistent from different perspectives. This will include:
    a. Training Group. The appropriate group for this subject
    b. Marketing Group. Announcing and promoting the library.
    c. Website Group. Adding self-guiding material to the website.
    d. Technical Committee. Making sure that there are no factual errors in the technical content.
5. A person in the role of an editor will be assigned to achieve consistency on multiple levels. This work includes tasks which are often not attractive for volunteers (in contrast to the work on the content). Ideally, this person will be paid by the DDI Alliance for a limited period of time.
6. Further communication between the workshop participants will be supported. The goal is to foster exchange between participants and possibly mutual review of future training material.
    a. The DDI Alliance Confluence wiki can be used as cooperation platform.
    b. An email list will be created.
    c. A bi-monthly phone call will be scheduled.
7. The material will be published with the Creative Commons Attribute license.
8. The material will be published in a way that it is citable. This is important for the material but also for the contributors. Some possible approaches: the DDI Alliance Working Paper Series or the Zenodo repository. Both would provide persistent identifiers, i.e. DOIs.
9. Additional multiple training material can be provided in a “contributed” section on the DDI Alliance website. This could include for example future slide decks of the participants for their own tutorials, and the larger slide deck on DDI Lifecycle of older GESIS training workshops at Dagstuhl.
10. A person will be assigned who is responsible for the overall organization. Some options seem to be the DDI 4 project manager (with a future task), somebody of the Training Group, or the suggested editor.

Appendix 12

TC Workplan 2019/2020:

- Reviewing and publishing DISCO
    - Content is ready to be put out for a 30-day review in June
- Preparing DDI 3.3 for publication
    - Final entry and documentation
- Resolving 5 DDI 4 Prototype review issues
    - Overall best practices for XML structure
    - Shift from document base to multi-use instances (document, transfer, discovery systems, etc.)
    - Specification of default values that are inherited by contained objects
    - Broad feedback from ICPSR regarding interaction of DDI products, audiences, etc.
    - Broad feedback from NSD regarding role of DDI4
- Reviewing issues filed for DDI-Codebook and preparing a new version
    - Currently have a set of issues related to improving link with DataVerse
    - New issues are being filed in terms of issue arising from the new World Bank tool
- Shifting DDI Lifecycle and DDI-Codebook production work to COGS
    - We have tested out the documentation production using the DDI 3.3 public review
    - Test and finalize input-output scripts
    - Complete and test output to ensure coverage and output consistency
    - Finalize transfer and open for access
- Preparing for the shift of DDI 4 development work to COGS from Drupal
    - Finalize scripting for XMI input and output

Appendix 13

DDI Alliance
Standards Development and Review Process and Procedures

The Data Documentation Initiative (DDI) Alliance develops standards and semantic products supporting the documentation, management, and integration of social science data and other data about human activity.

Definitions

DDI Codebook: The development line aimed at providing the comprehensive documentation needed to effectively use and analyze social science data focusing on simple surveys and data sets.

DDI Lifecycle: The development line aimed at documenting and managing data across the entire life cycle of research data from conceptualization to publication and beyond.

DDI Controlled Vocabularies: The lists of standardized terms that provide semantics (definitions of the meaning of metadata elements) and content (declaration of instructions for what and how values should be assigned to elements) rendered in a DDI specified structure using an XML format called Genericode (an OASIS specification).

Expedited Review: An abbreviated review process for Sub-Minor Version Changes. Expedited review must be approved by the Director in consultation with the Director's Advisory Group. At minimum, such a review includes notification of the Scientific Board describing the proposed changes and a period for comments.

Invalidating Change: A change to a specification such that instances created using the previous version of the specification may be invalid when parsed against the new specification.

Major Version Change: A revision resulting in a new version of the DDI Codebook or Lifecycle specification that changes the coverage, scope, or functionality. A major structural remodeling of a specification would also result in a Major Version Change. These changes are generally invalidating. The change results in a new product namespace. In the three-digit DDI version numbering scheme, e.g., V. 3.2.1, a major change is indicated by the first digit.

Minor Version Change: A revision resulting in a new version of a specification that contains corrections for bugs or minor changes to improve functionality of current features. Instances created using the previous version may be invalid when parsed against the new specification. The change results in a new product namespace. In the three-digit DDI version numbering scheme, e.g., V. 3.2.1, a minor change is indicated by the second digit.

Specification: In the DDI context, a specification is a set of requirements for metadata describing social science and other data about human activity. The specification consists of a model, its rendering(s) in specific technical languages (like XML), and its documentation.

Sponsor Request: A request with accompanying funding from an Alliance Member for specific changes to a DDI specification to advance the standard.

Sub-Minor Version Change: A revision resulting in a new version of a specification that contains corrections for bugs or minor changes such as relaxing cardinality constraints or field-level documentation improvements. All DDI instances created using the previous version of the specification will still be valid when parsed against the new specification. The change does not result in a new product namespace. In the three-digit DDI version numbering scheme, e.g., V. 3.2.1, a sub-minor change is indicated by the third digit.

Technical Committee (TC): The group that models, renders, maintains, and updates the specifications.

Validating Change: A change to a specification such that instances created using the previous version will be valid when parsed against the new specification.

Development Lines

The DDI standards and products are progressed through development lines that meet the needs of specific audiences, solve common issues in the relevant domain, and adapt to changes in the domain. Current development lines are DDI Codebook and DDI Lifecycle. A line of Controlled Vocabularies is also made available and maintained.

Proposals for Changes to the Standards

Every proposal for a modification to an existing DDI specification goes through one or more of the following processes, depending upon the extent of change involved, unless alternative processes are later approved by the Director, the Executive Board, and the Scientific Board. All proposals must be (a) accompanied by a revised model and technical implementation developed by the Technical Committee, (b) publicly posted for an open review period enabling comment from the DDI community, and (c) voted upon by the Scientific Board or receive expedited approval by the Director in consultation with the Director's Advisory Group.

The change procedure for the Controlled Vocabularies is outlined at http://www.ddialliance.org/controlled-vocabularies.

Scheduled Publication of New Versions

Major and Minor Version Changes can occur a maximum of twice a year with publication for review in January or July. It is anticipated that revisions resulting in a change to the namespace will occur less frequently than this. Sub-Minor Version Changes (those not affecting the namespace) may occur as required to address critical bugs or other non-invalidating minor changes.

Controlled Vocabularies are versioned to meet the needs of the community.

Development and Review Process

1. Identifying Objects of Change

User Requests. Any DDI Alliance member or user of DDI specifications may note and report DDI items that require change, either correction or expansion, through the DDI tracking system. The Technical Committee routinely reviews these reports and determines if an immediate sub-minor change is warranted or if the change should be noted and corrected in the next planned version change. In addition, the Technical Committee can note that an issue is broader than a specific correction to an existing element and can recommend to the Director that the broader issue be brought before the Scientific Board for consideration as a topic for a Working Group tasked with preparing a proposal for change to the standard.

Working Groups. The Scientific Board may authorize the creation of a Working Group to explore changes to a specification in terms of the depth of current DDI coverage or expansion of that coverage into new areas whenever recommended by the Director, when proposed by a Member Organization with the support of two additional Member Organizations, or upon its own initiative, such as focused discussions among members. The creation of a Working Group by the Scientific Board implies that resources should be directed towards support of the Working Group and the integration of the change into the standard once approved.

Sponsor Requests. In the event that a Working Group is not formed due to lack of resources or the competing need for limited resources, a proposal for change may also be developed as a Sponsored Request for Change. The proposal may be submitted by a Member Organization with the support of at least three other Member Organizations. The proposal must meet the same specifications as that developed by a Working Group and proceeds through the same review process. The Member Organization(s) agree to sponsor the proposal and provide the financial resources required to review, process, and implement the change.

2. Proposal Development and Content

A proposal for change may be prepared by an established Working Group, a sponsored Working Group, or the Technical Committee. A proposal for change should reflect the input of the DDI community concerned with the topic area. To help the community assess the proposal during review, the proposal must include a complete draft statement of content and functionality as well as information about the business case for the proposed change and the objectives that the change will achieve. When relevant, the proposal should also document solutions suggested but not followed up on to ensure that the proposal for change was thoroughly considered before submission.

A member of the group preparing the proposal should be designated the Architect and is responsible for ensuring that the proposal is complete, for coordinating with the Technical Committee in preparing a technical implementation for testing, and for answering questions arising during review. Working Groups are encouraged to involve a member of the Technical Committee as

a liaison to facilitate development of a well-integrated proposal. The Director must be kept apprised of all proposal development.

3. Technical Review

A proposal may be submitted to the Director by the Technical Committee or a Working Group. The proposal must include a model as well as a technical implementation for the purpose of testing. The provision of multiple Use Cases is strongly recommended.

The Director will then ensure that the proposal is circulated to all Alliance members for a Technical Review before it is made available for public comment. The Technical Review period should span one to two months. The proposal will be made available electronically and a means of tracking discussion and feedback will be established. The electronic discussion will be open to all Member Organizations, without constraint. As issues are raised during the Technical Review, it is expected that participants will provide feedback (formally or informally) on how to resolve the issues. The Architect of the specification should post responses to these issues. The Architect will consider the comments on each issue in consultation with the Technical Committee and make an informed decision on whether to change the proposal accordingly.

At the end of the Technical Review period, each member of the Scientific Board will be asked to give the Director a "Yes", "No," or "Not-Yet" vote. A "No" vote, to discontinue the Review and reject the specification, must be accompanied with comments to explain the vote. A "Not-Yet" vote indicates that the specification either needs substantially more design work or that the complexity of the specification requires a longer Technical Review cycle. A "Not-Yet" vote also must be accompanied with comments to explain the vote. The Director will consider the votes and make an informed decision as to whether to proceed to the next stage of the process, to extend the Technical Review, to dismiss the specification without prejudice and ask the sponsor to resubmit after redesign, or to reject the specification. In order to proceed to Public Review, at least 50 percent of those voting must have voted "Yes." The Director must explain his or her decision in writing to the Scientific Board.

At the end of a Technical Review the proposal may be made available in a beta version for immediate use and testing prior to incorporation into a new version of the specification. This is done at the discretion of the Director with input from the Technical Committee. This makes the content available for use prior to official publication and allows for further refinement of the proposal. This is particularly useful for proposals that add increased depth or new content to the specification.

4. Public Review with Proof of Concept and Change Log

If the proposal is accepted after Technical Review, then the proposal is made available for a Public Review of at least one month. The Director will make a reasonable effort to distribute the revised specification to the public, primarily by posting on the public Web site. All public comments will be published and openly available. Any issues raised during this Public Review must be publicly

answered by the Technical Committee. The period of Public Review may be extended at the discretion of the Architect, the Technical Committee, or the Director.

In parallel with the Public Review, a "proof of concept" implementation of the specification should be undertaken. Proof of concept typically requires a complete, public, portable implementation, but a public implementation is not a requirement. Alternatives to proof of concept may be proposed on a case-by-case basis and are subject to the discretion of the Scientific Board. It is not required that the lead sponsor of the proposal be the organization that undertakes the proof of concept implementation.

When relevant, a Change Log should also be made available, showing which elements in the specification were changed, how they were changed, and the rationale for the changes. If acceptable to the Scientific Board, the Change Log may be an alternative to the Proof of Concept.

5. Vote

When Public Review has ended and the Director believes proof of concept (or its alternative) has been established, each member of the Scientific Board will be asked to give the Director a "Yes" or "No" vote. A "Yes" vote indicates that the validity and usefulness of the proposed modification have been demonstrated and that the revised specification should now be accepted as a part of the DDI standard. A "No" vote indicates the case has not yet been made for the proposed modification. A "No" vote must be accompanied with comments to explain the vote. The Director will consider the votes and make an informed decision as to whether to accept the specification or to restart the process at some earlier stage. In order for the specification to be approved, at least two-thirds of the Scientific Board must vote "Yes," but the Director is not required to approve the specification even if a higher proportion of the Board recommends its approval. The Director must explain his or her decision in writing to the Scientific Board. Ordinarily, the Director will override a substantial vote of the Scientific Board only when he or she perceives that the proposal would compromise the DDI as an international standard. The decision of the Director may be vetoed by a two-thirds vote of the Executive Board.

6. Publication of the Approved Version Change

The Technical Committee will incorporate corrections noted in the Public Review and prepare the new version of the specification for publication. This will include a revised model, a technical implementation of the specification, field-level documentation as a usable guide, and revised high-level technical documentation. The Technical Committee will review corrections with the Director to determine if an additional review period is needed. Timing of publication is dependent upon resources and will be determined by the Director in consultation with the Technical Committee.

Process Requirements Based on Level of Change

1. Major Version Change

Major Version Changes are the result of adding new areas of coverage or making significant structural changes to the specification. These changes are generally invalidating but this is not a

requirement for a Major Version Change. Given the scope of change, Major Version Changes generally involve the development of one or more Working Groups or extensive consultation between the Scientific Board and the Technical Committee.

Major Version Changes involve a Technical Review addressing each specific area of change, a Public Review of the complete content of the new version, a vote to approve, and publication. At the discretion of the Director with advice from the Technical Committee, multiple areas of change may be rolled into a single Technical Review.

2. Minor Version Change

Minor Version Changes may involve a wide range of corrections and added depth to current coverage. They may be focused or wide reaching depending on the bugs and issues filed against the specification. Minor Version Changes may be initiated by the Technical Committee with the approval of the Director. They may or may not incorporate the proposals of Working Groups submitted while the Minor Version Change is in process. The Director may decide to place such a proposal on hold and not include it in the Minor Version Change if there is a need for further review or if the proposed change would greatly impact the larger specification (incorporation resulting in a Major Version Change).

A Minor Version Change involves a Technical Review, Public Review, a vote for approval, and publication. The Technical Review may be limited or bypassed if the Director determines that the process of developing the proposed changes involved sufficient input and review by Member Organizations concerned with area of coverage.

3. Sub-Minor Version Change

All Sub-Minor Version Changes must contain only validating changes. Common reasons for a Sub-Minor Version Change include a major or blocking bug found post-publication, expanded or corrected field-level documentation, or relaxing cardinality rules. The preparation of a Sub-Minor Version Change is proposed by the Technical Committee and must be approved in advance by the Director in consultation with the Director's Advisory Group. This is to ensure that resources are available for the work involved. Sub-Minor Version Changes may be expedited when the Director in consultation with the Director's Advisory Group determines that sufficient review has taken place with the concerned Member Organizations. At minimum an Expedited Review will include notification of the Scientific Board describing the proposed changes and a period for comments.

[2013-11-11]

Appendix 14

What is SDTL?

Structured Data Transformation Language (SDTL) is an independent intermediate language for representing data transformation commands. Statistical analysis packages (e.g., SPSS, Stata, SAS, and R) provide similar functionality, but each one has its own proprietary language. SDTL consists of JSON schemas for common operations, such as RECODE, MERGE FILES, and VARIABLE LABELS. SDTL provides machine-actionable descriptions of variable-level data transformation histories derived from any data transformation language. Provenance metadata represented in SDTL can be added to documentation in DDI and other metadata standards.

There are currently three parts of SDTL:

1. SDTL JSON schemas for data transformation commands
2. Function Library: Each statistical package has hundreds of functions for common mathematical, statistical, and text operations, such as LOG, SINE, AVERAGE, and LENGTH. The Function Library is a crosswalk between a standard SDTL representation of each function and the implementations of that function in various statistical languages. The Function Library minimizes program code for SDTL applications, because all functions can be handled in the same way, and additions to the Function Library do not require changes to program code.
3. Pseudocode Library: The Pseudocode Library provides human readable translations of SDTL commands. Like the Function Library, it is external to the SDTL JSON schemas, so it minimizes program code and can easily be updated.

The C^{2}Metadata Project has developed software applications based on SDTL:

1. Parsers for SPSS, Stata, SAS, and R translate command scripts in these languages into SDTL
2. Updaters for DDI and Ecological Markup Language (EML) incorporate SDTL into existing metadata files
3. A Pseudocode Generator translates SDTL commands into human-readable text
4. A DDI Formatter creates an HTML codebook that displays variable-level provenance metadata in pseudocode with hyperlinks to antecedent versions of variables.

These applications are prototypes to demonstrate the potential of SDTL, which will be available as open source code. We expect software developers to incorporate these functions into future products. We also anticipate future parsers for Python, SQL, and the SDMX Validation and Transformation Language (VTL). An updater for JSON LD metadata is under discussion.

SDTL also has the potential to provide translations between statistical packages. We believe that it will be possible for a software application to read a script written for SPSS and translate it into a script for Stata, SAS, or possibly R. The goal of the C^{2}Metadata Project is the creation of documentation, but we have found sufficient similarity among these languages for SDTL to serve as an intermediate step in the translation process. Please note that SDTL is limited to

describing data transformations. A similar approach to statistical analysis is possible, but beyond the scope of our current project.

Benefits of the DDI Alliance maintaining SDTL?

SDTL greatly enhances the value of DDI, because it is a key component of an automated metadata production process. Currently, DDI metadata is almost always created by data repositories not by data producers. Even when data are born digital, data producers discard provenance information that could be transported into DDI, because they do data management and variable transformations in statistical packages with minimal metadata capabilities. SDTL and the tools created by the C^2Metadata Project are designed to create a metadata life cycle that parallels the data life cycle. The same scripts that are used to transform and manage data files can be used to update metadata files. As a result, data producers can create more accurate and complete DDI metadata with less time and effort for them and for data repositories.

Like DDI, SDTL is a standard that should remain open and non-proprietary. However, SDTL will evolve and require periodic updates. The best way to assure that SDTL remains open and current is for the SDTL standard to be owned by a self-governing community, like the DDI Alliance.

What is required of the DDI Alliance?

If the DDI Alliance adopts SDTL as a standard, it would do three things.

First, the DDI Alliance would create a structure and process for maintaining SDTL as a standard. This would be similar to the process that maintains other DDI standards, but SDTL should be much easier to maintain than DDI. The SDTL standard is much smaller than the DDI standard, and it will have a smaller and more focused user community.

Second, the DDI Alliance would be the sole publisher of the SDTL standard, and the DDI Alliance would defend SDTL from attempts to claim or modify the standard by anyone else. Again, the DDI Alliance already plays this role for the DDI standard. Several years ago the DDI Alliance decided to use the trademark system to defend DDI with legal advice and assistance from the University of Michigan.

Third, the DDI Alliance would need to communicate with other communities that use SDTL. For example, the ecological research community has been very interested in SDTL, and the DataONE network of data repositories is discussing using SDTL to represent variable provenance metadata in its data catalog.

Who would do the work?

The DDI Alliance would appoint a working group to monitor and update the SDTL standard in response to the needs of the SDTL user community. As noted above, the SDTL community is likely to be smaller than the DDI community, but it will extend beyond the social sciences. If SDTL is successful, there will be volunteers willing to do this work.