EDDI Sprint Final Report

December 5, 2018
(updated links on January 14, 2019)

Joint Meeting of Technical Committee and DDI 4 Development Group, November 26-30, 2018

DDI Lifecycle Work

DDI-L work focused on the preparation of the COGS tool for support of both the DDI-L and DDI4 production work. Input flows, validation options, outputs, and post-output production pipelines were considered. Note that publication of DDI-L 3.3 is not dependent upon the completion of this work. DocFlex has been completed (with the exception of the final high level document information) and it is now being produced from the current schema structure for field level documentation and we are able to produce high level documentation through the current configuration of COGS.

Requirements for DDI-L COGS:
- Support of currently used XSD features
- Validate that all of 3.3 is imported correctly
- Ability to output DDI-L 3.3 in current XSD with supporting documentation
- Verify that the UML definitions are accurate and stable
- Output of XMI from CSV files

The overall production flow is still to be decided, through COGS output options or following the DDI4 generation from the XMI.

Production validation work has been specified, charted and begun for DDI-L in COGS. Some points regarding DDI4 in COGS have been added and work has been completed on a test transformation

for the canonical XMI to CSV which should allow for identification of transformation issues going into and coming out of COGS for this product. Several other steps have been completed or are in progress.

Related Documents

Current XSD Objects Used

GANTT Chart for COGS

DDI 3.3 Production Validation

DDI 4 Development Work

The DDI 4 work fell into two general areas: Production framework/work cycle and modeling.

Production Framework

Overview

An iterative approach for moving the DDI 4 development work forward was proposed in the form of the Modeling Representation Testing Lifecycle (MRT). The goal of this approach is to provide a ‘mostly’ automatic five stage development process from modeling through implementation (software testing) on a platform that is sustainable at the technical level and that can provide feedback input on a stage-by-stage basis. It would also allow for various UML modeling tools to be employed at the user’s discretion.

Figure 1. Modeling Representation Testing LifeCycle (MRT)

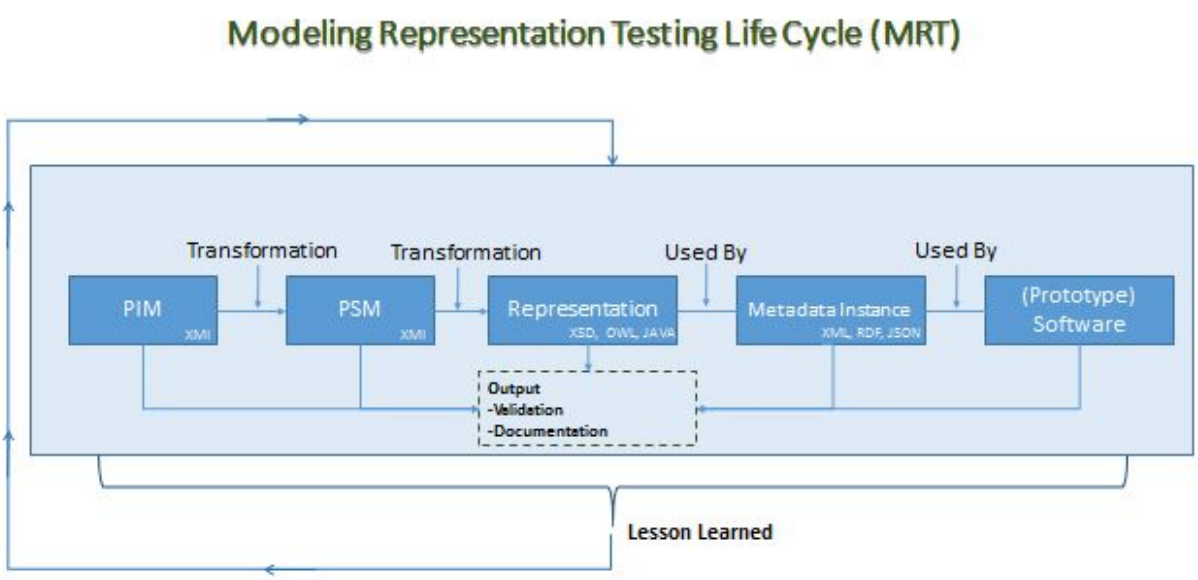


Figure 2. MRT Feedback Loop

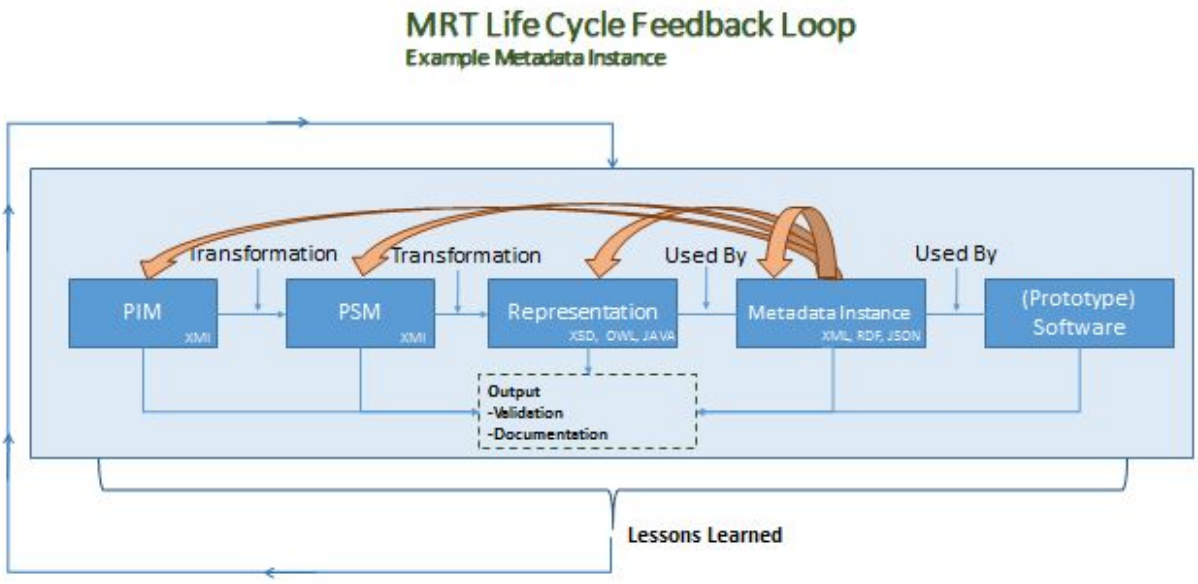
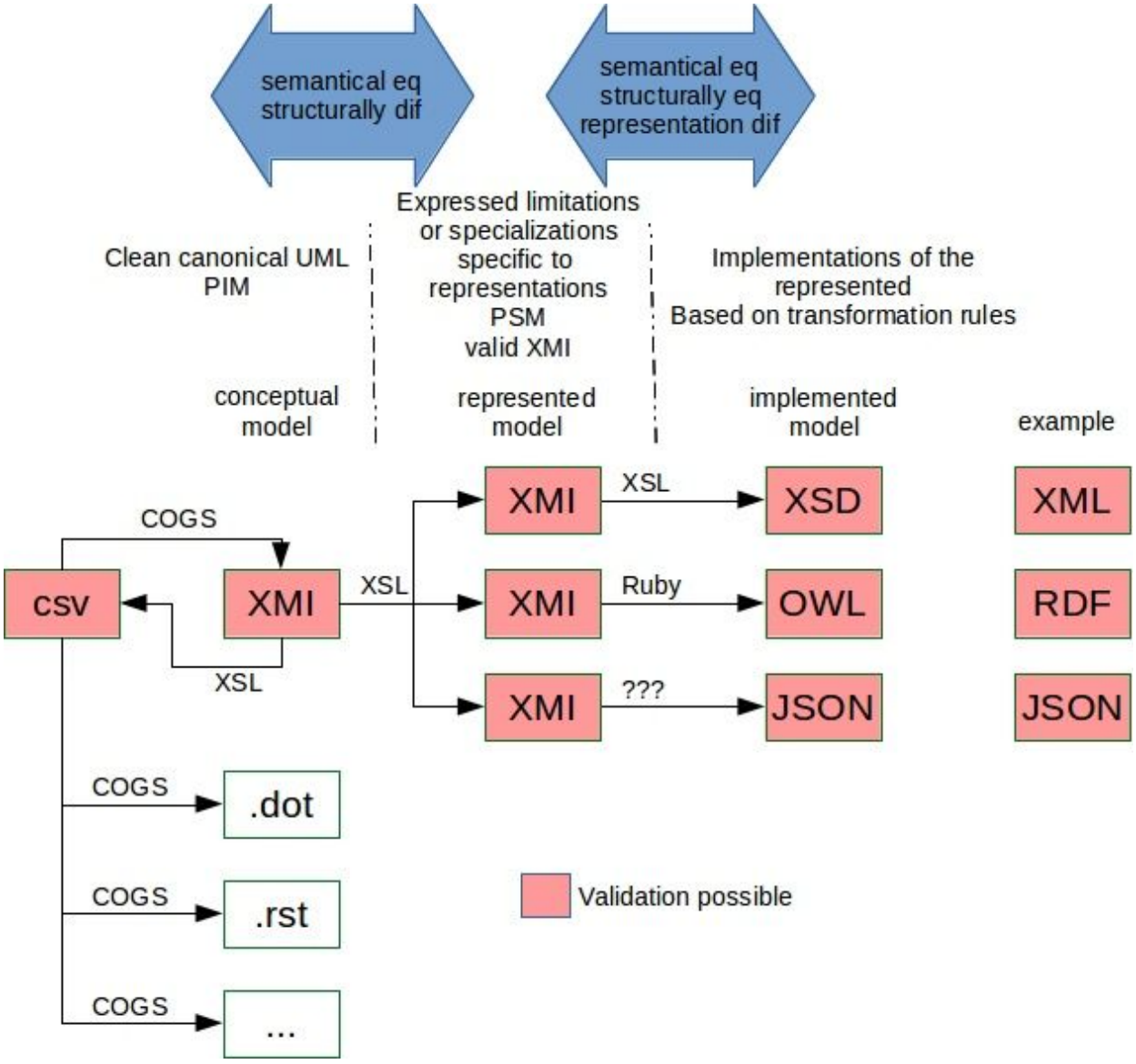


Figure 3. Production Workflow with Outputs and Validation



Requirements Breakdown

A list of requirements that covers function, output, and best practices was produced. Each item is associated with one or more stage(s) of the MRT and includes a note about whether it exists in the current production framework and in COGS. Comments on each requirement are also included.

DDI Production Framework Requirements Breakdown (link)

Preliminary work on a set of validation rules was begun, broken down into UML-specific rules, and a set of rules for Platform Specific Model (PSM), RDF, and XML. These documents are in the very early stages of work and are based on the validation documents produced during the Copenhagen Sprint in 2015.

UML Validation
PSM Validation
XML Validation
RDF Validation

General Discussion

The following points were made in the discussion of the MRT and workflow in general which require more input for resolution:

1. At the prototype software stage, the implementation, it is important to reach out to users who are actually doing this type of implementation. Converting from instance to instance is good, but not the entire perspective. This stage should also be used to import existing codebook and lifecycle instances for testing. Also need to test the 'uncommon' problems in real life, the edge cases, not just the mainstream.
2. There is a question about the canonical 'source of truth' in this pipeline. Should it be the csv or the XMI (PIM stage). In COGS the csv and the XMI are two guaranteed representations of the model, but can you have two 'sources of truth'?
3. Version control in the process needs clarification
4. Resources are required in order to bring this proposal to fruition.

Modeling

Overview

JIRA contains a number of issues pertaining to modeling which are labeled as post-prototype (known prior to the release of the prototype but not yet resolved) and prototype-review (filed as part of the prototype review process). To gain a broader understanding of the issues, they were grouped into categories and then linked to a 'master' issue for each category. A report of each other these categories contains a description of the requirements or a statement of the problem(s), a discussion of proposed solutions with questions and limitations in each area, and links to the issues in JIRA that are related to that area. The report as it stands can be found here in its entirety. The separate areas will be broken out and placed into the corresponding master issue in JIRA.

UML To Be Decided - Consolidated

Issues relating to data types, cardinality types, class content, class independence, and class membership, were broken down in a supporting document that covers the general purpose and importance to the model (UML) and representations (XML, RDF) along with comments.

Object Types Overview

Issues relating to Views and the Library - supporting documentation

Library and Functional Subsetting

MRT Group Proposal

A proposal was put forth for a next generation modeling working group with the MRT in mind, focused on the iterative lifecycle of modeling, representation, and testing. The proposal includes 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.

MRT Group (Modeling, Representation, Testing Lifecycle)

General Discussion

A number of topics require additional input and follow-up:

1. While addressing the modeling issues, it was agreed that it is necessary to revisit the goals of DDI 4 and create a list of requirements, both business and technical, to be evaluated and approved by the Scientific Board for moving forward. A distinction can be made between the more 'technical' requirements, say of UML features that are business-independent that could serve, for instance, the goals of finding the best way to work with other standards, and the more business-specific requirements that must include feedback and input from the DDI community.
2. The consolidated report addresses the JIRA issues and the direct input from the sprint participants, but additional input can and should be added to fill in any gaps. Each area/category should be assigned to a modeling group member.
3. The MRT proposal requires additional content as well as a consensus on the 'core features' for moving forward:
    - Include a statement about how MRT would interact with the 'content' creators and implementers, (not about new content, but focusing on making the core more robust).
    - Specific timeframe for 1 year to make the prototype 'fly'

- How and where the work is done - JIRA? Google Drive documents? What platform?
- Linking issues between BitBucket and JIRA or other higher level goals
- Resolving the pipeline issues with existing and accessible resources
- Outreach to tool implementers (developers)

Conference Call Notes Supporting Discussion:

EDDI Remote Call - Tues - November 27th
EDDI Remote Call - Wed -November 28th
EDDI Remote Call - Friday - November 30th

Packaging the Next Steps: Presenting to Scientific Board/ Funding from Executive Board

1. What's in it?
    a. Goals - Audience and Purpose for DDI 4 (w/reference to 3 and 2)
        i. What are the inputs? TC Review of 3, etc. Group discussions of 4, feedback from DDI User community…
        ii. Capture the discussion to date for consideration of moving forward and buy-in

    b. Modeling Requirements - Business and Technical (input from various sources)
        i. UML To Be Decided - Consolidated (prototype review and post-prototype)
        ii. Flavio (StatCan and other statistical agencies)
        iii. Jay and Arofan and the ALPHA project
        iv. Cross-Domain DDI 4 Dagstuhl 2018 Report
        v. UKDA implementation of Data Description (Jon and Darren Bell)
        vi. Other additions? Future ideas…

    c. Production Framework Internal requirements for sustainability of DDI products

2. Where is it?
    a. Which platform to use for generating and managing the documents?
    b. Mechanism for delivery and review

3. When is it?
    a. Must be ready in April 2019 for the Scientific Board to review prior to the DDI Annual Meeting at IASSIST.