DDI Roadmap

Edited: Wendy Thomas, Jon Johnson

Background

The DDI Moving Forward project was conceived in 2012 to address many of the limitations of DDI-C and DDI-L. These limitations were:

– Designed around the expectation that metadata is managed in XML resulting in excessive packaging
– Multiple namespaces (originally intended to allow use of subsets of metadata, approach proved to be inflexible)
– Hand crafting of schema
– Hand crafting of high level documentation
– Rise in use of RDF by community was not met by DDI standard (DISCO was an initial attempt to deal with this for discovery metadata, but is still unpublished)

The following year the Moving Forward Project was formally established with the following aims:

To build the next-generation DDI specification based on a data model with the following goals for delivery in 2015:

– Model based development with multiple bindings (XML and RDF at a minimum)
– An automated production process for the standard and documentation
– Coverage of DDI 3.2/3.3 payload
– Clear relationships with GSIM
– Develop with an understanding and intent of supporting good relationships with relevant standards
– Repackage the specification
– Enhance the documentation

Alongside the development undertaken in this project, there was a clear expectation from the existing community of users of DDI-C and DDI-L that there should be continued development to meet immediate requirements.

Assessment of progress on Moving Forward

The engagement with the user community has been substantial. Over 70 people have participated in the various activities since the inception of the project. There has also been significant input from other metadata standards or in specific areas where the standard is not being used but where there are overlaps and possible opportunities for alignment.

Areas where the project has been successful are:

– Engagement with other metadata standards
– Refinement of what a model driven approach means in practice
– Development of a more robust and sustainable documentation approach
– Lowered the barriers to involvement and contribution to the standard
– Transparency of development processes and management through Confluence and JIRA
– Datum and viewpoint have been insights that will have major utility
– Simplification of the serialisation in XML and the removal of the packaging which in turn makes tooling more straight-forward
– The scoping work on RDF
– Modelling of patterns e.g. collections, process, methodology and signification

Areas where the project has not been successful are:

- A clear alignment and migration path from either DDI-C or DDI-L
- Change management and control of versions and releases
- Production of serialisations other than XML e.g. RDF/OWL
- Stable production pipeline and development of the collaboration tools to meet evolving requirements
- Reproducing all of the functionality of DDI-L 3.2 in the model
- Ensuring adherence to the modelling rules and conventions.

It is our judgement that the prospects of delivering a stable release of DDI4 (as currently named) in the next 12-18 months is very low. There are many contributing factors why this is likely, but they are primarily encapsulated by the areas where the project has been less successful indicated above.

A side effect of the effort involved in the project is that the DDI-L line has faltered, as well as delaying associated initiatives such as DISCO and XKOS, which leverage the standard into new areas.

Proposal for a new direction

The concerns illustrated above have been the driver for the TC face-to-face meeting in June 2017, which had as its aim the long-delayed finalising of DDI-L documentation, release of a new point version and placing the production of DDI-L on a firmer footing.

This followed the issuing of two key documents from the TC, on best practices for DDI-L and the proposed roadmap for DDI-L at the all members meeting in May 2017.

Advantages of the new approach

This draws on lessons learned from the Moving Forward project, specifically:

- A more collaborative framework for development of the standard
- Lowering the barriers to the submission of contributions to the standard
- Model driven approach to create required outputs
- The integration of the production of the model and its associated documentation
- Addressing the deficiencies in the management of versioning and change control
- Reducing the dependencies on specific technologies and skills for the production process
- A clearer migration path for new features and structural changes for existing users of the standard
- Bring in all the development practices that have worked well (Confluence, JIRA, transparency, robust and sustainable documentation, engagement with other standards)

The intent is that this will be able to create a more sustainable development process to continue to improve the standard to meet community requirements.

Work on a new development environment is well advanced, based on both GIT and COGS (github/Colectica/DDItoCogs, and github/Colectica/cogs) which respectively, transform the existing schemas and creates an editable model based on CSV files, and transforms the model into the desired outputs. GIT will provide item and property level version tracking, the ability to create branches for specific work and experimentation, and the ability to manage merging updates and content via pull requests. All forks, branches, and pull requests will be immediately tested by the automated build infrastructure.

Additional work is required to add some of the functionality available in the current development platform, to capture some elements such as mapping to GSIM, explanatory notes etc. This work is planned to be completed in the near future, alongside adding in continuous build capability.

Proposed Outputs and Timeline

DDI-L 3.2

Revised documentation utilising the new production framework for DDI-L 3.2

DDI-L 3.3

DDI-L 3.3 will be issued for review as soon as is practicable, pending the production of accompanying documentation. This release will contain:

- Bug fixes from DDI-L 3.2
- Some enhancements to the previous version, including integration of several pieces of DDI4:
  - Non-survey data capture
  - Classifications
  - Unit description
  - Improvement to the variable cascade
- XML schema's in the same flavour as previous versions of DDI-L
- Documentation will comprise a PDF generated from source controlled restructured text files.

Not part of the official release, but for community comment will be

- Flattened XML schema
- A UML model
- Web based documentation

DDI-L 3.4

Pending community feedback it is envisaged that a further version DDI-L 3.4 will be released in spring 2018 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 schema's in the same flavour as previous versions of DDI-L.

Future versions

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 - for 2017/2018, 2018/2019

A visual of the proposed timeline to achieve a merged production process is found Appendix A.

Managing work over the coming 12-18 months
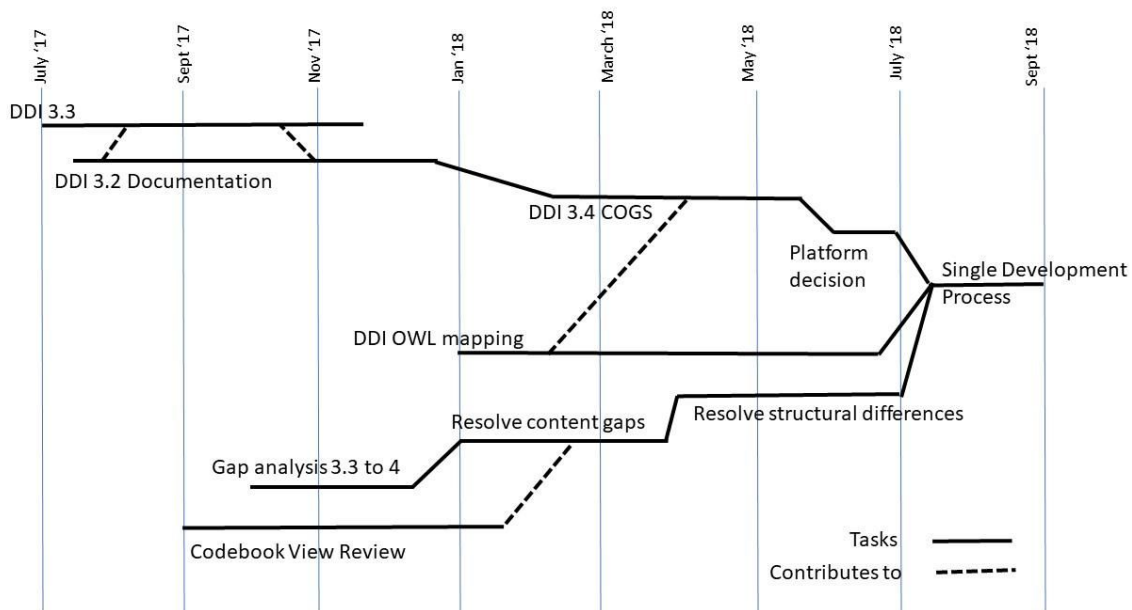
Activities would merge over the year. Initially the Modelling Team should merge into the Technical Committee. As the majority of these individuals are already on the Technical Committee this should be a minor adjustment other than integrating their scheduled activities and determining how to merge or route content from the Modelling Team confluence site to the Technical Committee confluence site.

The assumption is that a number of the activities such as the gap analysis and resulting workplan, as well as the DDI OWL mapping would begin in Dagstuhl. A clear strategic work plan covering finalization of various parts of DDI4 such as Process realizations, data capture, and data description needs to break out specific goals within the work and integrate it into the overall gap analysis and the work to resolve the gaps. It is anticipated that the work patterns of existing content groups will need to shift over the coming year to support on-going development within the Alliance as opposed to being related to a special project.