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.