# Lawrence Sprint Report

June 13, 2017
Kelly Chatain & Wendy Thomas

**Codebook Working Group**

The primary task of this group was to evaluate the DDI 4 Codebook View from a user’s perspective, to identify gaps, and include recommendations on documentation. Sanda Ionescu of ICPSR has worked with DDI Codebook for 18 years. Her approach to this task was to take an existing ICPSR DDI 2.0 codebook and replicate it using only the DDI Codebook View as expressed in an xml schema. Concurrently, Mehmood Asghar, a software developer from the World Bank, worked on reproducing the exact metadata that is captured via Nesstar for the IHSN by using the Data Description View. Nesstar has reached end-of-life and the World Bank has decided to create software (using electron.atom.iso) that mimics Nesstar’s functionality, with a few usability improvements. Nesstar has worked well for them and they’ve invested heavily in training countries to use the existing IHSN DDI profile consistently. Neither Sanda nor Mehmood had any significant understanding of DDI 4 prior to this week’s activity.

After working with DDI 2 and 3 so long, Sanda has a ‘map’ of the DDI in her mind. Using Codebook View required throwing away this map and starting over with a different structure, looking for familiar elements, and seeing what the schema will ‘allow’. At first this made DDI 4 appear much more complicated, due to the increased cross-referencing and IDs required, but ultimately showed its appeal by keeping much of the content only a few levels in (with some exceptions). Mehmood also had to shift his thinking in order to document the variables ‘outside’ of the data file and linking instead of from ‘within’.

What follows are specific issues and questions that arose while working through the examples. Some problems were fixed during the sprint while others are still outstanding. The outstanding issues were formally filed using the JIRA issue tracker (tracking ID appears next to item in the list). The resulting xml documents can be found on the sprint’s [codebook working page](https://ddi-alliance.atlassian.net/wiki/display/DDI4/Codebook%2BWorking%2BPage).

Issues

*General*

* Clean up use of content vs. languagespecificstring - need to identify all points where attributes point to complexdatatypes DMT-132
* xs:language should be different datatype in annotation. It should be a complex thing that can be xs:language and string. DMT-133
* ConceptualInstrument is not in Codebook View. See item in Study (below) (ADDED)

*Study*

* Design linked via hasDesign, but uses DesignOverview on the view list CWG-9
* hasAFunder indicates that should be an agent, could not insert agent, type can only be organization or individual or machine - ADDED documentation
* hasInstrument only allows for ConceptualInstrument in the attributes, not ImplementedInstrument. ADDED ConceptualInstrument to FV link is from ImplementedInstrument to ConceptualCodebook

*Variables*

* Need value for code DMT-128
* How to group variables? DMT-127
* Range and valid range FOUND

*Frequency*

* How to input frequencies? DDI4DATA-18

*Question*

* InstanceQuestion requires RepresentedQuestion, which is not in the codebook view. DMT-126
* Able to create a represented question when using the whole DDI schema DMT-126

*Layout*

* We cannot instantiate a physical layout in the codebook view. FOUND
* No location attributes in DD4 DMT-128
* The name of data file should be in dataStore (logical), but is missing one for the physical IMPORTANT DDI4DATA-19

*Coverage*

* in SpatialObject, we need a place to put human-readable lowest level of coverage. e.g. street, county, state, etc. DMT-134

[Note that all DMT assigned issues have been resolved except for DMT-127 which will be resolved as soon as the Collection Pattern is finalized.]

Recommendations for Documentation

* Visual representation arranged in a tree structure, similar to the [CDISC Operational Data Model (ODM)](https://ddi-alliance.atlassian.net/wiki/download/attachments/71908575/ODM1-3-0-Final.htm?version=1&modificationDate=1496335596327&cacheVersion=1&api=v2), for the Views.
* Clear guidance on where to start: In the xml, DDI is the root and from there you can proceed to Study and so on, etc.
* Provide clear (and many) examples of what content should go in each element/object. Limit technical details and references.
* Clearly define what abstract classes mean.
* Define what attributes, properties, and relationships mean in the context of the model. This should be spelled out for the layperson with specific examples. Larry wrote an example of this during the sprint. See [Using Relationships in DDI4](https://ddi-alliance.atlassian.net/wiki/download/attachments/71874537/UsingRelationshipsInDDI4.docx?api=v2) on the sprint wiki.
* Annotation will also require guidance on when and where to use, particularly for new audiences.
* Clear explanation of how to record what used to be part of the methodology section in DDI 2, such as sampling procedures, weighting, etc. The added flexibility of doing this in 4 means it needs more guidance.
* Specific guidance on creating IDs by hand. Oliver wrote an example during the sprint. See [Identifier Recommendation](https://ddi-alliance.atlassian.net/wiki/download/attachments/71874537/Identifier-Recommendation.doc?api=v2) on the sprint wiki.
* Clarification of what CollectionTypes mean, bags and sets.
* The relationships between instance, represented, and conceptual for both variables and questions needs to be made VERY clear.
* Also, clear language around DataStore and the PhyscialLayout options.
* Locators are buried quite deep. Clear language needed.
* Please use the actual element name in the xml instance to avoid confusion.

This information has been passed on to the documentation group in order to guide their continued work.

**Modelling Working Group**

The modelling team worked on clarification of what DDI 4 would encompass when ready for publication, revising and reviewing patterns (how they work and how they fit together), rules for realizations of patterns, laying out the plan for a gap analysis, and other documentation and modelling fixes as needed. The goal was to produce content drafts of specific topics that could be integrated into the overall documentation and finalized post-sprint.

The description of DDI 4 reviewed original documents regarding technical and process goals, coverage, and product content. This is a critical document for informing gap analysis work, transition documentation between DDI versions, and planning DDI 4 work for the next year.

Oliver Hopt wrote a [first draft of an overview](https://ddi-alliance.atlassian.net/wiki/download/attachments/71874537/WhatisDDI4andwhatisitsstate.docx?api=v2) describing the current state of DDI 4 from the technical perspective, with the purpose of knowing when a ‘ready’ state has been achieved. It includes sections on rules, bindings, xml, rdf, production flow, conceptual work, Lion, abstract classes, and patterns. Still to be described is content.

Wendy Thomas wrote an [environmental assessment](https://ddi-alliance.atlassian.net/wiki/download/attachments/71874537/environment-coverage.docx?api=v2) of DDI 4 in relation to past versions of DDI and other standards as a first step in a more thorough gap analysis. The Technical Committee has already begun work with DDI 3 and GSIM will be done during the summer by the Modelling Team.

The modeling team focused on reducing the complexity of the collection pattern and making it clear and easier to implement. Given the centrality of this pattern in DDI 4, simplification of this pattern is intended to address problems of complexity in the the process pattern and several realizations of the collection pattern throughout the DDI 4 model. Specific goals included collapsing description of a collection with its internal structure and making sure that it could support the current set of realizations within the model. Prior to the sprint the status of collections was reviewed and initial proposal developed. [Collection Pattern and Realizations](https://ddi-alliance.atlassian.net/wiki/display/DDI4/Lawrence%2BSprint%2C%2BMay%2B2017?preview=/70326219/71875614/Collections-116-114.pptx), [The Relativity of Collections](https://ddi-alliance.atlassian.net/wiki/display/DDI4/Lawrence%2BSprint%2C%2BMay%2B2017?preview=/70326219/71875657/Collections%20Relativity.pptx), and an initial [proposal for revision](https://ddi-alliance.atlassian.net/wiki/display/DDI4/Lawrence%2BSprint%2C%2BMay%2B2017?preview=/70326219/71875574/collectionrevisions.pptx) for the purpose of discussion. These were the basis for discussion on Tuesday and the development of a [revised proposal](https://ddi-alliance.atlassian.net/wiki/download/attachments/71874537/ProposedCollections.pdf?api=v2) on Wednesday. This was further revised with images captured during the discussion posted at [Modelling Team Work Page](https://ddi-alliance.atlassian.net/wiki/display/DDI4/Modeling%2BTeam%2BWork%2BPage) on the sprint wiki along with a number of related documents covering specific aspects of the collection and signification patterns.

Dan Gillman produced a draft of the [Rationale for the Signification Pattern](https://ddi-alliance.atlassian.net/wiki/download/attachments/71874537/Rationale%20for%20Signification%20Pattern%20%281%29.docx?api=v2), written specifically for those with limited technical expertise.

A number of proposed pattern diagrams were produced by Larry Hoyle and work is continuing.

Work to be done post-sprint:

* Describe the relationship between the Collection Pattern and GSIM Node/Nodeset (Dan)
* Relating the collection algorithm to EBNF (Extended Backus-Naur form)
* Collection Pattern June-July (need to complete this by end of July in order to update content for Codebook FV)
* Revise the new/proposed collection pattern based on Friday discussion (Wendy)
* Create realizations of basic types and create a View of these (Wendy)
* Generate a binding of the above (Oliver)
* Make examples using the XML generated from the realizations (Dan, Larry)