Codebook

Knutholmen Sprint | 23–27 May 2016

# Open issues

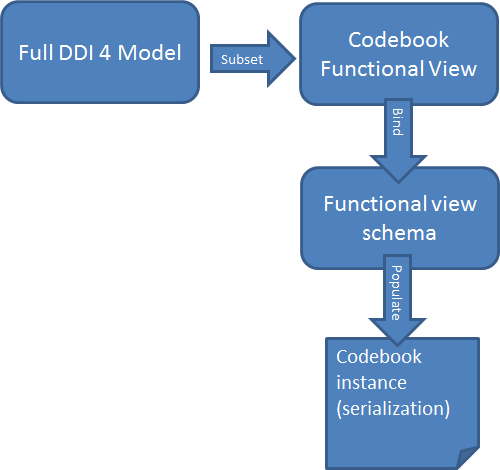
* Codebook Functional View
  + Mapping to DDI Codebook Specification
* Model (PIM)
  + What is missing in packages from which classes are used in the Codebook Functional View (CFV)
* Production Framework
  + PIM to PSM
  + PSM
  + PSM to XSD
  + Structure of XSD
  + RDF is probably parked for the meeting

# Clarifications re: “Functional view” vs “codebook as report” (Ornulf and Marcel)

During the sprint, Oliver Hopt and Ørnulf Risnes went on to make instances of DDI 4 Codebooks in XML and YAML, respectively. The instances are serializations of content that can be represented in the Codebook Functional View model (and by extension the XML-schemas and ontologies produced from the model). Jay called a codebook “a metadata report”, which seems to be a decent metaphor in many cases.

The following questions arose: Are the instances simply populations of the schemas/ontologies produced from the Functional View? And can these schemas/ontologies be used to produce “templated” for populating the Codebook by hand or for easy uptake by developers?

The answer seemed to be “yes”.



*Figure XX - Overview of the “path” (bad term) to a codebook instance*

The figure XX above indicates the relationship between the codebook instances, the model - and the layers/steps between.

# Discussions day 1

Dan Gillman was appointed as **decider** for the technical subgroup.

The first break-out started with a discussion about what a “functional view” is, and how it relates to the complete set of classes in DDI 4.

**The conclusion was:**

*At this point we only care about what properties we need in the Codebook, as a requirement, and leave it to the modelers to make that happen. As such, the discussion at this point should be content-based.*

**Scope discussion:**

Ørnulf: If we support DDI 2.5-use-cases only, there is no real value-add of DDI 4-codebook? Are we really limiting ourselves to cross-sectional survey data (the use-case for DDI 2.5).

Dan: By expressing Codebook in DDI 4-model, we are automatically able to tap into the reusability-aspect of DDI 4, something that was missing in DDI 2.5. So there is a value-add simply by expressing a Functional Codebook View in DDI 4 terms.

Jay: But is that sufficiently tempting to move people over from DDI 2.5 to DDI 4.

**Reviewing the attributes in the spreadsheet:**

This afternoon.

# Walking through the spreadsheet

## Document Description-section

Issues were identified regarding modeling of authors etc, before the Document Description was parked. It was parked because Document Description is not the domain of expertise of DDI nor the participants

## Study Description-section

We went through the spreadsheet point by point, and Oliver added issues/notes to column F in the spreadsheet in the “NewStartingPoint” workbook within the [spreadsheet](https://ddi-alliance.atlassian.net/wiki/display/DDI4/Documents+from+Sprint?preview=/38240266/38240277/CESSDAandLiteLH.xlsx)

Conclusion regarding methodology:

We are going to map from the Methodology-attributes in the spreadsheet onto the Design Overview in DDI 4.0. Design is an extension point of Methodology.

Example: Methodology for sampling corresponds to the name “Sampling” associated with “Design overview”.

# Creating an XML Instance of SimpleCodebookView

Going through the sheet named Knutholmen\_Sprint, all rows that contain precise mapping will result in a filled XML node in the sample instance.

The main issue was that within package Methodology not all things that are needed for SimpleCodebook are modeled in LION.

Currently left out but with mapping are Questions. The reason is, that the connection from captures to any other class within SimpleCodebook is not obvious. I would suggest to have a direct relationship from InstanceVariable to Capture called iscupturedBy. This would make it possible to leave the entire flow out but still have the questions in.

Also left out because some sort of obvious mapping but no obvious connection are Codes/Categories and Concepts.

# Codebook Instance in YAML

The following YAML-based example was produced with the XML-based example (done by Oliver Hopt) as background. It is not complete, and not necessarily well-formed or fully correct. But it should give an idea of what the codebook really could look like:

---

codebook:

DocumentInformation:

Id: "urn:ddi:di1"

Agency: "DDI Alliance"

Version: "1.0.0"

HasAnnotation:

Annotation:

Id: "urn:ddi:andi1"

Title:

String:

Language: "en-US"

Content: "Simple Codebook first test instance"

String

Language: "de-DE"

Content: "Simple Codebook erste Testinstanz"

Abstract:

String:

Language: "en-US"

Content: "some text"

String:

Language: "de-DE"

Content: "ein bisschen Text"

Creator:

Agent:

String:

Language: "en-US"

Content: "Oliver Hopt"

Contributor:

Agent:

String:

Language: "en-US"

Content: "Ørnulf Risnes"

Study:

Id: "urn:ddi:std1"

Agency: "DDI Alliance"

Version: "1.0.0"

HasAnnotation:

Annotation:

Title:

String:

Language: "en-US"

Content: "survey of the DDI users"

Language: "de-DE"

Content: "Befragung der DDI Nutzer"

SubTitle:

String:

Language: "en-US"

Content: "Knutholmen edition"

Language: "de-DE"

Content: "Knutholmen Ausgabe"

Identifier:

IdentifierContent: "http://doi.org/10.1234/lalala"

isURI: "true"

ManagingAgency:

Content: "DOI"

# This can be a controlled vocabulary entry as well

Creator:

Agent:

String:

Language: "en-US"

Content: "Oliver Hopt"

Contributor:

Agent:

String:

Language: "en-US"

Content: "Ørnulf Risnes"

Role:

Language: "en-US"

Content: "Data collector"

ControlledVocabularyName: "List of roles"

Publisher:

Agent:

String:

Language: "en-US"

Content: "drive.google.com"

Role:

Language: "en-US"

Content: "Distributor"

Copyright:

String:

Language: "en-US"

Content: "DDI Alliance"

HasTopcialCoverage:

TopicalCoverage:

Id: "urn:ddi:topcov1"

Agency: "DDI Alliance"

Version: "1.0.0"

Keyword:

Language: "en-US"

Content: "user study"

Keyword:

Language: "en-US"

Content: "coneheadded"

Subject:

Language: "en-US"

Content: "DDI internal review"

HasTemporalCoverage:

TemporalCoverage:

Id: "urn:ddi:tempcov1"

Agency: "DDI Alliance"

Version: "1.0.0"

CoverageDate:

NonIsoDate:

DateContent: "Mai 2016"

TypeOfDate:

Language: "en-US"

Content: "time period"

HasSpatialCoverage:

SpatialCoverage:

Id: "urn:ddi:spacov1"

Agency: "DDI Alliance"

Version: "1.0.0"

SpatialAreaCode:

Language: "en-US"

ControlledVocabularyName: "ISO 3166-1"

Content: "NOR"

Description:

Content:

Language: "en-us"

IsPlainText: "false"

Content: " <p xmlns=\"http://www.w3.org/1999/xhtml\">Somewhere in the nowhere.</p><p xmlns=\"http://www.w3.org/1999/xhtml\">Surrounded by norwegian nature</p>"

UnitType:

Id: "urn:ddi:ut1"

Agency: "DDI Alliance"

Version: "1.0.0"

Name:

Content: "INDIVIDUALS"

Context:

ControlledVocabularyName: "THE\_OFFICIAL\_UNIT\_TYPE\_CV"

HasAnnotation:

Annotation:

String:

Language: "en-US"

Content: "Individuals"

Language: "de-DE"

Content: "Personen"

Universe:

Id: "urn:ddi:uni1"

Agency: "DDI Aliance"

Version: "1.0.0"

Name:

Content: "TEENAGERS"

Context:

ControlledVocabularyName: "THE\_OFFICIAL\_UNIVERSE\_CV"

HasAnnotation:

Annotation:

String:

Language: "en-US"

Content: "Teenagers"

Language: "de-DE"

Content: "Teenagers"

HasInstanceVariable:

InstanceVariable:

Id: "urn:ddi:var1"

Agency: "DDI Aliance"

Version: "1.0.0"

HasAnnotation:

Annotation:

Id: "urn:ddi:anivar1"

Title:

String:

Language: "en-US"

Content: "sam1"

Abstract:

String:

Language: "en-US"

Content: "It's in the article!"

BasedOnRepresentedVariable:

RepresentedVariable:

Name:

Content: "SAM1"

DisplayLabel:

DescriptiveText:

UnitType:

Id: "urn:ddi:ut1"

Agency: "DDI Aliance"

Version: "1.0.0"

Name:

Content: "INDIVIDUALS"

Context:

ControlledVocabularyName: "THE\_OFFICIAL\_UNIT\_TYPE\_CV"

HasAnnotation:

Annotation:

String:

Language: "en-US"

Content: "Individuals"

Language: "de-DE"

Content: "Personen”