DDI Moving Forward, Sprint #1
Friday, November 1, 2013

Plenary

New use case

An additional use case was suggested: the library use case (DataCite data catalogs).
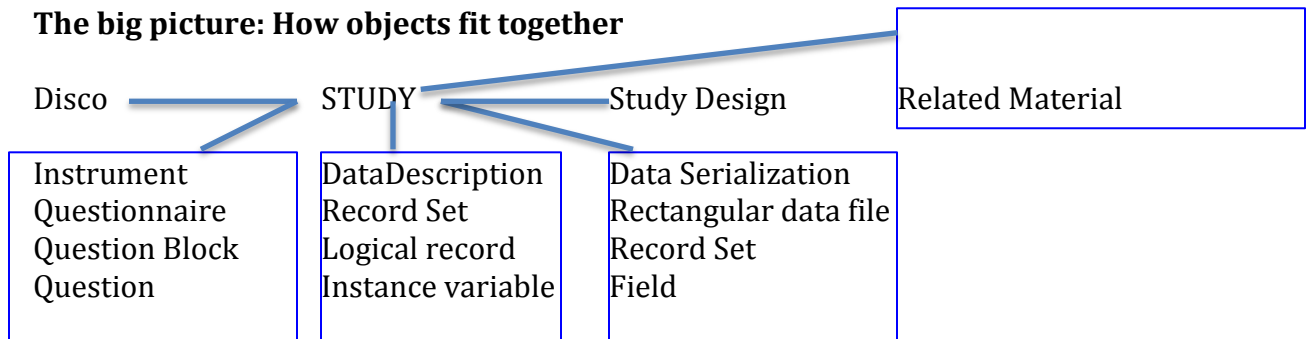
Content Group

Instrument Group

In the revised diagram, we capture links to response domain and to some kind of variable.

Cardinalities were reviewed and assigned.

The group developed an example questionnaire and applied the model to it.

The big picture: How objects fit together

Disco ⟶ STUDY ⟶ Study Design   Related Material

Instrument	DataDescription	Data Serialization
Questionnaire	Record Set	Rectangular data file
Question Block	Logical record	Record Set
Question	Instance variable	Field

Open Issues/Parking Lot Issues

Identification, maintenance, version

DDI 3.1 was not optimal in regard to identification but was ameliorated in 3.2. We can do better: One idea is to take the id mechanism in 3.2 and restrict to only URNs since we now have RDF representation and we can map to RDF. We have a hard constraint because of RDF – complex models have to have a URN.

- All metadata has to travel without loss of information.
- The id of an object through different versions should be recognizable.

Currently agency is the namespace of the identification system and is also used to maintain metadata. We need to decouple identification system from agency. Do we need to have separate maintainable and identifiable objects?

Web services

We need a proposal for restful syntax and this should be considered the 10th deliverable is this document.

Entending beyond rectangular data files

We need to extend our rectangular data file example and also cover groups of studies.

Minimum set of administrative metadata

Do we have a minimum set of administrative metadata on every object? Yes, because of RDF.

Grouping

Schemes got overloaded in DDI 3. We need to cover ad hoc packages of stuff (not maintained), maintainable packages of stuff, and topical (content of the same type) groups (last 2 are supported by schemes). We need an object for topical and ad hoc packages. All groups should contain just references; there should be no notion of containership

Process after the sprint

- Modeling people will pick up and associate pictures for objects; make changes and iterate through cycle
- Don't know base objects yet
- We will iterate around the model and then request feedback from the Content group around the three use cases
- Ideally we will put the model through a production run
- Jon Johnson will look into Docbook for field-level documentation
- We don't want modelers to lock us into certain structures
- The Technical Committee is where the modelers live going forward
- The TC and the content groups will work collaboratively
- Virtual meetings minimum of every two weeks
- When we have diagrams, bring others in to the process – those who signed on for specific areas at the Cologne meeting
- We need to think about how to involve others. To do this, we will need a document aimed at content modelers, including how to use Drupal.

Object templates

- We have right set of objects but there are gaps in the object templates – particularly in properties and relationships

- We need a mechanism for highlighting things that still need to be filled in -- Use comments field
- Systematically get good examples
- We need validation in the Drupal forms to ensure that everything is filled in -- automate the validation that things are complete.
- What's in Drupal is missing a lot of info like cardinality
- What is in in Word template and Drupal are not the same
- We don't have the right model for describing coded categories and code lists
- Attributes on study – are we sure we have all of them?

File sharing

We need a collaborative file sharing space – we can use DropBox

Open issues with XMI export

- XMI export doesn't show direction/type (composition/association) of relationship
- To dos:
    - Adding datatype field to content capture form
    - Transfer all items to production system

Workflow issue in EA

Importing XMI from Drupal will currently overwrite existing version in EA. Issue with consistent IDs

- Open issues with modeling
    - Explore how package versioning works in EA
    - Exporting XMI to EA is not efficient

XML binding

- Have gaps in the binding
- For object level mapping – have properties of literal - see Arofan's document
- Need to work on Web services and DDI 4 (10th DDI deliverable)
- All issues with modeling have been documented in the modeling documents

XSLT

- Finished with data types and enumeration
- Open issue:
    - Defining DDI core and DDI U2

RDF

- Have taken the model and created initial RDF representation
- Some comments on the xslt for the rdf binding:
  - axioms are camelized
  - underscore characters are deleted
  - human readable rdfs:label is implemented
  - axioms are named according to naming conventions

Additional Parking Lot Items

- Data dictionaries for qualitative Data
  - o Rectangular data file → Data File A
  - o Non-rectangular data file → DataFile A
  - o 1…N data dictionaries
  - o Object dictionary
  - o Codebook
- Find better language for "determinable role" etc.
- Semantics of collections
  - o Schemes insufficient.
  - o Need richer expressive capability (Jay Greenfield)
- Simple, single structure for CV's role lists, universe, internal/external
- Default Delimiter
  - o Inheritance or not?
  - o Should we try to merge all InstanceX and X?
- Dublin Core for Provenance, etc. for data resource
- For Identifiable/Versionable/Maintainable objects…do all have a name, label, description that they inherit from the IVM extension base?
- Does data serialization need a connection to LogicalRecord?
- E.g., when it is not detailed as in RetFile, but generic as in a link to an interactive Data ? website