DDI Moving Forward, Sprint #1
Tuesday, October 29, 2013

Plenaries

Modeling Design Principles

In the plenary meeting to start the day, the whole group reviewed two design principles:

- Gratuitous remodeling is forbidden
- Objects should represent actual things

Deliverables

In the plenary at the end of the day the group reviewed the deliverables again and recommended releasing the minutes from each day to give others not at the sprint a good sense of the discussions.

It was pointed out that flattening XML will happen through RDF so it is not clear that we need all nine of the deliverables. We can inject packaging hierarchy into the business rules at the binding stage. In response to the suggestion that we just release RDF, it was pointed out that we need to optimize both XML and RDF for processing in each syntax.

Drupal as input for modeling

We now have a Drupal input mechanism and an XSLT stylesheet in progress. We can use this on instances to see what we will get rather than the Word templates, which are static and not interactive. We want to quickly publish pieces of the model for targeted review.

Questions to address

At a small group meeting after the plenary, some questions were posed.

Do we need a tool available off-line? It was decided that no, we should concentrate on Drupal.

Documentation

What form should documentation take? We can use DocBook or DITA for documentation.

In the proposed production workflow, how should documentation be handled for syntax-specific? We will quickly need to decide what the outputs are for these documentation items:

- XML, RDF and OWL documentation
- General documentation of the model
- High-level documentation plus examples and use cases

Should we produce packages by function module? Should we separate technical documentation from general user documentation? What is the internal structure for documentation? We want to be able to create small general chunks of documentation that can be reused. We should look at SDMX documentation for a good example.

Business process and approvals

In terms of business flows, there is a strong relationship to the production flow. We need to decide at what point we do Alliance approvals and on what timetable as we are going to be iterating around the content and model. We will need both a Scientific Committee and a technical review.

We will have captured some good documentation – definitions of objects, properties, relationships, examples -- and can release things for review on a rolling basis. It was pointed out that the HTML community has an active, living standard. They do snapshots of it periodically and this is successful. DDI needs to be stable for longer periods, however.

Our current timetable calls for a first draft of the four functions we are working on to be delivered in January with the complete set in May and a first draft of three new functions. This kind of periodicity seems a good model.

We may find we cannot deliver all that we have promised. In that case we can put more resources toward completion or drop off some of the deliverables. In the latter case, we need a good set of priorities to determine what to forgo. We have a prioritized list of functions and should publish it.

It was also pointed out that the work needs to be shared and so skills and expertise need to be "swappable."

Content Group

The group split into two groups to focus on (1) the object descriptions for Concept, and possibly Designation and Sign, and (2) the object descriptions for Universe, Characteristic, Category, and Category Set, along with examples. The goal was to turn to the simple data description after that.

Concept

The group started by asking whether we should indeed add Designation with Sign and Vocabulary as foundational and decided in the affirmative.

The suggestion was made to use SKOS since that already exists, but it was decided that SKOS describes a vocabulary and has limited value for DDI. SKOS should be compatible with the model but shouldn't drive the model. Not all labels we have are from controlled vocabularies.

DDI has created an extension of SKOS called XKOS, which we may use when we get to modeling classifications.

Category set

The group discussing this was uncertain about whether we actually need it. We can add packaging and grouping at the binding stage but this will be different for different syntaxes.

The group considered the relationship between the category set "American States" and a characteristic. American States is a concept itself. The relationship between characteristic and category set was a problem for ISO 11179 also. A characteristic is kind of like a question and the category set is the potential answers. Sets of categories are managed metadata in real systems. Category set is a specific type of grouping. Do we need a grouping object of which category set is one type?

GSIM did a thorough job of this and SDMX and DDI have schemes. Not all items should live in schemes; is category set a subclass of an abstract grouping?

Some type of grouping object should be foundational. One needs a collection level organized around a set of designations. It was decided to use Concept System as modeled in GSIM.

Simple data description

A new data file description drawing on the DISCO vocabulary and the DCAT vocabulary was described and modeled. The use of DCAT is seen as a plus because we should use existing vocabularies when they make sense. The main connection point is the data file (a "distribution" in DCAT terms). The simple description models a CSV file, which is very basic. There are physical and logical components to this just as currently in DDI. Note that this is a simple description and not intended for nonrectangular files.

If the dataset changes, the version of the identifier of the physical instance will change.

A question was asked regarding whether the description can capture qualitative data. It appears to be flexible enough. Another question was how this description relates to the foundational metadata.

In reconsidering the modeling, the group created objects for Data File, File Structure, Field, and Record.

Technical Group

The group discussed the modeling template.

1. Aggregation and composition for example can also be represented using simple associations.
   Thomas Bosch recommends keeping the models as simple as possible, so aggregation and composition are not necessary.

   Decision: go with a plain "contains"

The semantic of associations should be seen in the names of the associations.
One problem would also be how aggregation and composition should be represented in the different types of representations.
The intended semantics of the relationship between variables and questions for examples should be seen when reading the name of the association and not of the association type.

2. Shouldn't we stay with official UML terminology?

- Relationships are associations
- Properties are attributes

   Decision: We will come up with a UML glossary for developers.

Thomas recommends an underscore character between the object ID and the property ID in order to see where the property ID begins.

3. 'The object ID should be unique' --> why not 'must'? It would be too confusing to have IDs in the model which are the same referencing different objects.

   Decision: We will get rid of all IDs in the template.

4. Do we really need roles?

   Decision: No, we will get rid of "target" and "source" roles in the template.

Thomas thinks Roles are also “syntactic sugar” making our model complex.
We can use the association name both for the source and the target object.
It would be also too confusing if there is a relationship child in the target object and a relationship parent in the source object.
Now we cannot see that the same association is meant.