DDI Moving Forward, Sprint #1
Thursday, October 31, 2013

The plenary started out with a "Soap Box" presentation by Ornulf Risnes. He suggested that under the principle of no gratuitous remodeling, we should stick with the elements of the simple codebook from DDI, which has a strong relationship with the DISCO vocabulary, which we are using for attributes at the study level.

Content Group

Instrument Group

Separating logical and physical instrument

The group started by asking how Instrument interfaces with the simple data description. The answer is that an instance variable links to Instrument.

We still have not fleshed out the response domain. Represented variable links to the value domain (uses the category responses). Response domain is a category set and not a coding. There is a variable and question connection.

The group discussed the need for something that represents the questionnaire as a generic form and not as a specific instrument.

- Example
    - Employment survey with a small set of questions
    - Done through different modes, each of which creates an instrument
    - We need to document the abstract thing that is the set of questions that exists independent of modes

GSIM has an Instrument Implementation. It is a class that separates questions and flow from the instrument itself. In GSIM Instrument is the abstract questionnaire. This involves the logical and the physical and is analogous to the variable and the instance variable. An important design principle espoused by GSIM is to separate the logical and the physical.

The group solved this by creating logical and physical instruments. The logical instrument is the instrument in theory (design that exists) and the physical instrument is the practice (includes mode). The relationship between physical and logical should be an association (instantiates or implements).

Statement

The group also split Statement into two different objects. For example, one might

use Statement to store a legal notice at the beginning of a questionnaire (this statement would not be maintained in a proper scheme because it is not intended to be reused, but other things would be reusable).

The next task on the instrument side is to fully realize a simple questionnaire using the new model.

Response domain

There were additional questions about Response domain. We aren't adequately representing skip and flow and the logic contingent on a response (can this live in instrument control?). Where are response choices recorded and how do they interact with control?

Codebook Group

Grouping

In terms of the simple codebook, the group went back to review DDI Lite. Is there a need for variable groups? We do need a grouping mechanism that can group flat lists dynamically. Current variable groups in DDI are static and we need to move away from that.

Elements of a codebook

What do we need in a codebook?

- Study description
- File description
- Data description

Some terminology in DDI is older and should be revised. File, for example, does not apply to databases. The group went forward using the term Data Serialization, equivalent to a physical file. A Study was considered a Data Resource.
What attributes do we import at the highest level? In terms of study-level (data resource level) attributes, the group agreed to use DISCO attributes – e.g., subject coverage, temporal coverage, geographic coverage.

Methodology, data production, and interoperability

It would be good to be more explicit about the scientific process in play for production of data in a Data Resource for interoperability purposes and to assess comparability. That is, there is a semantic interoperability issue because there is no way to compare at this high-level process level, e.g., the same study at different time periods. There is no framework to allow a researcher to tie to what the last researcher did in terms of study design. This may be a larger problem than DDI can

solve. There is also no way to compare experiments. Experimental data is another function to be addressed.

The Survey Design and Implementation working group has done a lot of work around sampling and weighting. Methodology is on the list of functions to be covered by DDI during the Moving Forward process.

Plenary

The larger group came together for further discussions.

Process information

There is a need for process and provenance information in DDI. Dublin Core is working on a provenance description with the W3C's PROV RDF vocabulary. This work can be pulled into DDI. We will need Organizations and People and an external process framework like GSBPM or GSLPM to plug in. We can reference out to the process framework to connect inputs and outputs with people. Process description would be an overlay. Organizations and People may need to be a separate function addressed during the modeling project.

The drawback is that the links go only in one direction and we haven't solved the interoperability problem. If you have generic way of describing process, how do people know what's going to be there to compare?

Identification

We need a mechanism for doing identification and will be discussing this in the technical group, outlining possible options.

Revised Design Principles

1. Interoperability and Standards
2. Simplicity
3. User Driven
4. Terminology
5. Iterative Development
6. Documentation
7. Lifecycle Orientation
8. Reuse and Exchange
9. Modularity
10. Stability
11. Extensibility
12. Tool Independence
13. Innovation
14. Actionable Metadata

15. Remodelling Discouraged
16. Objects Represent Actual Things
17. Separate logical from physical
18. Names are mutable
19. The model will only have features which reflect the common expressive capabilities of supported syntaxes/technologies (e.g., no multiple inheritances – RDF can do this, but XML cannot)

Grouping

We need a way to create a stack of instance variables. The grouping mechanism decided upon can be used for this. The grouping class should be general and a fundamental class capable of grouping any collection of types. There is an element called ItemCollection currently in DDI.

Drupal server

SND will host the Drupal server containing the object templates. We need to create a subdomain for this.

Technical Group Status

RDF

Thomas Bosch has been working on implementing the RDF mapping and the modeling guidelines, which include a list of the primitives that would go into utility packages and the core. This is not implemented yet in the Drupal drop-down list.

XML

Arofan Gregory and Oliver Hopt have been working on the XML mapping and XML binding rules. They are almost finished. The issue is that the simple types have not yet been done. They have just figured out how this looks in XMI.

Production process

The work on the production process is as complete as possible right now. We have to work on how the injection of XML from Drupal will work.

Restful query approach

Achim Wackerow made a proposal in 3.1 on a restful query approach so that all DDI services could become standard. In terms of statistical packages, SAS can call a web service but other packages are a mixed bag. We need to go through an exercise to determine how web services are used with DDI. The group did take on board the requirement that DDI become more web services friendly.

Content Group Status

Modeling

Larry Hoyle has been creating a UML diagram for the content group. This will be updated according to the object descriptions being created in Drupal.

Grouping

There will be future discussions on Individuals, Sampling and Methodology, and a Grouping Class.

In terms of grouping, there may be an abstract class created for this. We should take a stab at it here as it is important. It begins to answer questions about name, label, description, etc.

Response domains

This has not been fully developed. We have decided to link it under question (a capture type). Under response domain we have decided to use enumerated and described as types. Response domain maps to value domain. Question maps directly to variable. We have decided to pull that information in from variable and value domain. We have pulled in category under value domain which linked back nicely to foundational metadata.