DDI Moving Forward, Sprint #1
Wednesday, October 30, 2013

In the plenary, there were two "soap box" presentations (3-minute presentations to put forward a point of view):

Soap Box #1 – Single structure for codelist, etc.

Achim Wackerow proposed a single structure for code list (code and label) (codes and categories), internal controlled vocabulary, plus external additions, and external controlled vocabularies.

In the content group we have been considering a generic way of talking about code sets and controlled vocabularies. Achim's proposal is not as simple as the similar ones laid out, as there are other kinds of things we need to describe. An important area in statistics is classification schemes with structure and levels with precise meanings; this has additional constraints that a typical code list doesn't have.

Other points about this proposal and its coverage and rationale:

- Thesauri and other kinds of language organization systems that also fall into general framework of a code list and controlled vocabulary
- Response choices to a question or allowed values for a variable also
- Not so complex that we can't handle it
- 80% of use in DDI are simple and similar
- Motivation to have a simple structure to address this and then extend to the more complicated
- GSIM does this – abstract high-level way of describing these things. Defines three subtypes of this general class: not exhaustive but these three are critical to statistics.

Soap Box #2 – Using Smalltalk "collection"

Jay Greenfield also submitted a proposal regarding some of his ideas relating to the DDI base. His point is that DDI schemes are not sufficiently rich semantically. He suggests that we should grow itemset to conform to the Smalltalk abstract class "Collection" and all the imitators who have followed including HL7.

A Smalltalk Collection is a hierarchy of objects. It's a scale and the scale has an item pool traversed using CAT. The current DDI variable scheme including groups is not sufficient.

It was pointed out that this proposal merges the logical and physical. Jay will be at the next sprint for a fuller discussion.

Content Group

The group broke up into an Instrument group and a Codebook group.

Codebook Group

The Codebook group started on extending the simple data description to cover the variable, which will link into the foundational metadata and to the common codebook. It was decided to build on the work of the DISCO vocabulary in doing codebook modeling. DISCO has both Study and Study Group and a class that is a union of study and study group.

DISCO has universe, abstract, topical coverage (dcterm pointing to SKOS concept), spatial coverage (dcterms: Location), temporal coverage (start and end), kind of data (SKOS), and analysis unit.

The study has an instrument, a connection to data file, to variable, and from study to logical dataset. Study and groups connect to agents (person or organization), creator, contributor, publisher.

Instrument Group 1, Participants: Guillaume, Brigitte, Jenny

This small group discussed the instrument modelling and made these points:

- Objects first, then Essential properties
- Variable and Value Domain are there, but we don't discuss these
- Value Domain or Response Domain? (GSIM only has Value Domain as an object)
- Question Group is there but we don't discuss it
- For simplification we do not use Instance
- We don't need Survey Instrument as a sub-type of Instrument implementation
- Mode is a property of Instrument Implementation and not Data Channel
- Data Channel is there but we don't discuss it – it's the way to sampling, etc.
- Observation is an abstract class
- If we drop Survey Instrument, then Instrument is not an abstract class
- Instrument has no relationship to Instrument Control, because we need a mode for the Instrument Control
- We don't need Multiple Questions/Sub Question in addition to Question Block
- Measurement is there (not in GSIM), but we don't discuss it
- Observation as a new abstract class

Collapsed model

- We collapse Instrument and Instrument Implementation

- We allow there to be no Instrument Control
- We collapse Question Block and Instrument Control
- Flow and Control Transition are synonyms
- Interview Instructions are collapsed with Statement
- Instrument has Mode and Objective
- Question has Text
- Statement has Text and Type – e.g., interviewer instruction
- Instrument Control has Type – e.g., flow or block or random or sequence or matrix and Label

Instrument Group 2, Participants: Guillaume, Brigitte, Jenny, Wolfram, Sophie, Barry

This small group also discussed the instrument modelling and made these points: made these points:

- Inside the Instrument Control you can have everything except the instrument.
- Question can exist without Instrument Control.
- Instrument Control is not abstract.
- Instrument Control can capture these three things (Question, Statement, Observation) and itself.

- Observation -> Capture which is abstract
- Question is a type of Capture. Question has Text, others don't.
- Take matrix and block out of Instrument Control

- Capture can contain a task, etc.
- Leave flow control in Instrument Control, also sequence and random.
- Block is too complicated and needs its own Object.

- Instrument - * Instrument Control
- Instrument Control - * Instrument Control

In terms of Instrument, there is a need for something like a data capture protocol. The Instrument Group as a whole reported back having solved this problem (see above). Their group had two approaches – bottom up and top down – and they were able to meet in the middle with a solution that satisfied everyone. Their shared principle was parsimony – keeping things simple but applicable to any situation.

In their model, Instrument is a tool to capture data. There is a Flow control sequence documenting how to proceed through the instrument. There are links to Capture and then to Question. Capture is connected to variable also and there is a Question block in an extension. Statement holds interviewer instructions or layout instructions. There is an additional object called study design (need new term for

this) that defines how and why instruments are administered. Questionnaire has disappeared but is an instrument control object. This appears to hold for many types of data, including qualitative and experimental.

Plenary

The group concentrated on the update for the next day.