Data description and Data Capture Integration

Jon, Dan S., Johan, Wendy, Jon, Jay

Major classes involved:

Variable Cascade (Pkg: Conceptual): Conceptual Variable, Represented Variable, Instance Variable

Means of capturing (Pkg: Data Capture): Represented Question, Represented Measure, Instance Question, Instrument Component, Capture, Response Domains

Proposed touch points and questions around the

|  |  |  |  |
| --- | --- | --- | --- |
| SOURCE | TARGET |  |  |
| Response Domain | Represented Variable 0..n |  |  |
| Instance Variable | Capture (abstract)“sourceCapture”: | Think of this as pointing to the represented question | May not have a defined flow therefore require this to capture in these location |
| Instance Variable | Instrument Component (abstract)“wasCapturedUsing” | Think of this as pointing to the instance question |  |

|  |  |
| --- | --- |
| Object | Definition / Use |
| Instance Question | The usage of a Represented Question in a Question Flow (3.2 question construct); an instance question may be used multiple times in a question flow |
| Represented Variable | A Represented Variable is reusable and therefore can never really know all of the sources of data that are expressed in a way that meets its definition. |
| Instance Variable that is derived from multiple pieces of data | Points to the individual Instance Variables that were used to derive the instance variable in question |
| Capture point #1 | The relationship between the Response Domain and the Represented Variable that defines it (source is ResponseDomain, target is RepresentedVariable) |
| Capture point #2 | From InstanceVariable to the InstrumentComponent and Capture |

Documentation:

* Has to be clear that a Response Domain has a specific single response which can go specifically to a known single source (that is where the final content comes from)
* Need to document approach
* Relation is through the Response Domain which means this is a more refined than 3.2 and MUCH more refined than 2.5
* It is how we are defining the response options of the respondent and they are different that a representation. Response domains RESULT in representations at the point it gets to the Instance Variable

Record and Datum

* Capture is completely reusable
* You can say this was the flow for this individual case
* The way datum identifies the source of its content points to the record to the variable or how ever that relationship works. They are linked through the instance variable to the source.
* Datum (what is it?) there is some logical set of instance variables defined by someone and there is case (some individual of population) and the piece information captured is the datum
* A collection of DataPoints is in a record and a particular case of this is the Datum
* A lot of the datum specific detail tracking comes in at the execution of the instrument (instrument components) - this can vary by the system used for execution
* Blaise writes out instance variables to some record format and controls how that process occurs
* The structure we have captures the output of that process
* There is an execution layer between the capture and Record with the Datum

For next week provide some definitions:

* Basically we need the clear definitions of what the Response Domain is in DDI4 with some examples, what the touchpoints now are and how the act, and the relationship of the Datum with the execution layer.
* There is still a misunderstanding between groups of what an “instance” is which is the usage of a represented variable. While a Represented Variable has changed exacerbating this misunderstanding. The Instance Variable is vary tied up with an execution.
* How can one use Represented Variables as CDE’s if it is missing the sentinel values?
* Is there work that has to be done to sort out sentinel from substantive at the point of the Represented Variable?

Data Description:

* Draw the line here and think about datum in terms of execution layer rather than what “data capture” is covering.
* So then the topic for next week in Data Description needs to look at issues raised by realization of collection.

Data Capture:

* Data Capture can focus on their list of topics. This means these groups can/should meet separately.
* Implications of this model’s approach to specification of the response domain and applied usage of current managed response domain. For example the use of a numeric response domain across a number of Represented Variable.
* There is a reusable definition of value (such as percentage) as a component of represented variable.