On QuestionGrids and QuestionBlocks: Towards a Robust Data Capture Model for DDI 4

Barry Radler, 2/9/16

The current DDI DataCapture model in Lion is an attempt at creating a parsimonious and robust model that can document a variety of data capture contexts and instruments. In particular, its invocation of the Process model has streamlined the model so that many specific instrument functions, processes, and controls can be represented via the Process model instead of Data Capture description. This approach is sympathetic to the modelling principle followed during the Moving Forward sprints that “there should not be more than one way to describe one thing.” Indeed, the current DataCapture model appears capable of describing a variety of data capture scenarios with a circumscribed set of objects.

The Process model should be able to replicate any type of instrument control, including elements currently found in DDI 3.2 such as QuestionGrid and QuestionBlock. As such, I argue that these elements are unnecessary in the DDI 4 DataCapture model.

1. A QuestionGrid “structures the QuestionGrid as an NCube-like structure providing dimension information, labeling options, and response domains attached to one or more cells within the grid.”
	1. This definition suggests that the grid is a way of describing and locating cells (question items) in a table (grid), such that in its simplest form, one response to one question is the cell in a 1x1 grid. This distinction seems semantical; one can also identify a cell by its question number or identifier or label, and can define its “location” via the Process model (its location in a sequence). If a question grid can represent or locate a cell as a specific sequence of if-then-else steps, then the Process model can adequately represent a question grid.
		1. To elaborate, a question grid is *more appropriately considered a modal artifact* (a characteristic of the administration mode) than an object in its own right.
			1. Consider a household roster that enumerates and characterizes the members of a respondent’s residence. In a self-administered mode, it makes sense to present the series of questions in a HH roster to the respondent in grid form. This mode of administration eases the data entry (by the respondent), as well as data capture, transcription, and cleaning (by the data “capturer”). But the question grid is indigenous to this mode of administration, and is not a necessary individual element of a comprehensive data capture model. The same series of HH roster questions could be administered via a phone survey by an interviewer using CAI software. In such an instance, a grid is not only unnecessary but could prove inefficient, both to administer via CATI, and to code using the control language of the CAI program.
		2. The mode-related issues associated with QuestionGrid dovetail somewhat with display-related characteristics that DDI has traditionally avoided describing, i.e., typography, fonts, size, graphic design, layout, etc. Yet this information is legitimate metadata that is critical to accurately describing some instruments and measures.
			1. A legitimate question is whether the Process model, along with the ExternalAid object in DataCapture, should be able to adequately document all of the elements in this data capture context, i.e., a set of related questions administered in sequence in a specific mode with specific display characteristics.
		3. Taken to an extreme, if DDI accepts the QuestionGrid object as a mode-specific series of related itemized questions like those found in a HH roster, then what is to keep DDI from introducing other mode-specific elements in DDI 4 that describe evermore idiosyncratic data capture situations? Extend the QuestionGrid situation to other domains (e.g. clinical measures, cognitive assessments, laboratory experiments, multi-media product development research, etc.) and what is to keep DataCapture from becoming a top-heavy model overpopulated with objects that describe measures administered in specific modes?
2. Another example of potentially unnecessary objects in DDI 4 is QuestionBlock. “A QuestionBlock is a specific structure used in educational and other types of testing where an object (Stimulus Material) is provided and a set of questions are asked regarding the object.”
	1. QuestionBlock does not seem to have the same modal characteristics that QuestionGrid does, but it is another idiosyncratic data capture situation employing an object designed to describe one specific type of data capture. The features of a QuestionBlock situation are not so different from other survey-type data captures. The characteristics that make it unique (such as use of Stimulus Material, a characteristic shared by many other instrument types like intelligence tests and cognitive assessments) are more appropriately described by invoking ExternalAid within DataCapture, or by describing QuestionBlock situations in the Methodology model.
3. Ultimately, the inclusion of Grid and Block in the model (as individual objects) should be determined by whether use cases can or cannot be implemented by using the current DataCapture model. If the Grid or Block use cases break the current model, then the model needs to be appended. But then it needs to be determined where/how these objects will be incorporated.
	1. If Grid and Block are unnecessary as individual objects in DataCapture, then an alternative scheme needs to be offered that clearly describes where/how such data captures are described in the new DDI 4 model.
	2. There is some resistance to removing Grid and Block because some organizations have invested much effort introducing them to their DDI workflows. This is a legitimate concern, but it is not a sufficient reason for their inclusion in DataCapture going forward.

The resolution of the above issues will determine which of two directions the development of the DataCapture model proceeds: (1) a model able to describe many data capture processes using a minimum of robust elements or descriptors; or (2) a complicated model able to describe many data capture processes using an ever-expanding set (as new data capture situations are invented or described) of highly specific elements.