Metadata Management
Questions, Measures and Questionnaires

Hayley Mills
12 September 2019
CLOSER, UCL Institute of Education
Overview

• Questions and measures
• Questionnaires
• Parameters and bindings
Focus on DDI-Lifecycle

• Codebook is variable centric – it only provides a partial description of question as the source of data for a variable

• Lifecycle supports stand-alone question specification and management

• Lifecycle can describe the use of a question within a questionnaire flow-logic
Focus on DDI-Lifecycle

Codebook- Example
IHSN - 5th Census of Population 1992 - IPUMS Subset
El Salvador
Variables, questions and measurements

Variable
• Description of data
• A variable can come from a question or measurement

Question
• Describes a means of capturing data
• Specifies a text and the form of the expected response
• Questions can be organised in an instrument

Measurement
• Describes a means of capturing data
• Specifies the measurement and the form of the expected response
• Measurements can be organised in an instrument

<table>
<thead>
<tr>
<th>person_est</th>
<th>140</th>
<th>180</th>
<th>370</th>
</tr>
</thead>
</table>

Q1 How tall are you in inches?

Measured height

[ ] inches
Questions in Questionnaires

A1. Do you ever have a headache?

- yes, quite often: 1
- yes, sometimes: 2
- yes, I had one once: 3
- no, never: 4

Code list: 1, 2, 3, 4

Category: Health
Questions in CLOSER Discovery
discovery.closer.ac.uk

**Question label**: q1_A1

**Question text**: Do you ever have a headache?

**Code value**: cs_yqo_yq_yhoo_nn

**Category**: Multiple Choice Response Options

**Code list**: 
1. yes, quite often
2. yes, sometimes
3. yes, I had one once
4. no, never
# Questions and Measures

<table>
<thead>
<tr>
<th>Question Item</th>
<th>Measurement Item (DDI-L 3.3)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question Item Name</td>
<td>Question Item Name</td>
<td>Name</td>
</tr>
<tr>
<td>Label</td>
<td>Label</td>
<td>Label</td>
</tr>
<tr>
<td>Question Text</td>
<td>Type Of Measurement Item</td>
<td>Question, or description of measure</td>
</tr>
<tr>
<td>Question Intent</td>
<td>Measurement Intent</td>
<td>What are you trying to measure</td>
</tr>
<tr>
<td>Response Domain</td>
<td>Response Domain</td>
<td>How are you classifying the response</td>
</tr>
<tr>
<td>Response Cardinality</td>
<td>Response Cardinality</td>
<td>Number of allowed responses</td>
</tr>
<tr>
<td>Concept Reference</td>
<td>Concept Reference</td>
<td>Concept being captured</td>
</tr>
<tr>
<td>External Aid</td>
<td>External Aid</td>
<td>What might assist in response</td>
</tr>
<tr>
<td>Interviewer Instruction</td>
<td>Interviewer Instruction</td>
<td>Clarification, information to obtain consistent responses</td>
</tr>
<tr>
<td>Parameters / Bindings</td>
<td>Parameters / Bindings</td>
<td>Regulate the flow of information within an instrument</td>
</tr>
<tr>
<td>Represented Variable</td>
<td>Represented Variable</td>
<td>‘Template’ for the data produced</td>
</tr>
</tbody>
</table>
Response Domains

- A response domain defines response options to a question
- There may be more than one response domain for a question
- Response domains clarify the bounds of accepted valid and invalid values for the question
- Response domains capture a response that may then be recoded for entry in a data file
- They are not the same as variable value representations

<table>
<thead>
<tr>
<th>Style of Response</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Code</td>
<td>Pick list (codes and associated categories)</td>
</tr>
<tr>
<td>Category</td>
<td>Pick list set of categories (no associated code)</td>
</tr>
<tr>
<td>Numeric</td>
<td>Number field can be defined by type, range, and precision</td>
</tr>
<tr>
<td>Text</td>
<td>Text field can be defined by length and regular expression</td>
</tr>
<tr>
<td>DateTime</td>
<td>Date and/or Time field can be defined by format, range, and regular expression</td>
</tr>
<tr>
<td>Geographic</td>
<td>Set of fields to capture GPS position</td>
</tr>
<tr>
<td>Geographic Structure</td>
<td>Pick list for geographic levels (State)</td>
</tr>
<tr>
<td>Geographic Location</td>
<td>Pick list for individual geographic locations (Alabama)</td>
</tr>
<tr>
<td>Scale</td>
<td>Layout of a scale (Likert or similar)</td>
</tr>
<tr>
<td>Distribution</td>
<td>Layout of a distribution (assigning %)</td>
</tr>
<tr>
<td>Ranking</td>
<td>Ordering items</td>
</tr>
<tr>
<td>Location</td>
<td>Specifying the location on an image, sound, video, etc</td>
</tr>
<tr>
<td>Nominal</td>
<td>Marked or unmarked</td>
</tr>
</tbody>
</table>

### ccaf100 - A1: Child ever had a headache. CCAF file

**Dataset**

<table>
<thead>
<tr>
<th>Value</th>
<th>Label</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>-11</td>
<td>Triplet / quadruplet</td>
<td>13</td>
</tr>
<tr>
<td>-10</td>
<td>Not completed</td>
<td>8087</td>
</tr>
<tr>
<td>-1</td>
<td>Missing</td>
<td>22</td>
</tr>
<tr>
<td>1</td>
<td>Yes, quite often</td>
<td>580</td>
</tr>
<tr>
<td>2</td>
<td>Yes, sometimes</td>
<td>4365</td>
</tr>
<tr>
<td>3</td>
<td>Yes, I had one once</td>
<td>1325</td>
</tr>
<tr>
<td>4</td>
<td>No, never</td>
<td>1066</td>
</tr>
</tbody>
</table>

- **Valid**: 7336
- **Invalid**: 8122
- **Min**: 1
- **Max**: 4
- **Mean**: 2.39
Use of a Question

Managed within a Question Bank
- Made available for later use
- Managed in terms of version changes

Assembled into a Questionnaire
- Ordered
- Adding statements and information specific to the questionnaire
- Routing based on responses
Questionnaires

• A questionnaire mainly uses questions as a means of capturing data
• It describes the sequence of questions and measures, intervening text, and conditional routing
• DDI Lifecycle describes questionnaire starting with the Instrument item type
Things For You to Do

Thank you for filling this in. Children of the 90s loves to look at the things you draw!

A1. Do you ever have a headache?

- yes, quite often
- yes, sometimes
- yes, I had one once
- no, never

If no, go to question A3 below

A2. For your last headache, please shade in where the pain was in these two pictures of a head.
Questionnaires in CLOSER Discovery

Statement: Things For You to Do
Thank you for filling this in. Children of the 90s loves to look at the things you draw!

Question text: Do you ever have a headache?

Code list:
1. yes, quite often
2. yes, sometimes
3. yes, I had one once
4. no, never

Condition Description: If no, go to question A3 below
Condition Command Code: `qc_A1 == 4`
## Types of Control Constructs

<table>
<thead>
<tr>
<th>Control Construct</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence</td>
<td>Structural</td>
<td>A section of a questionnaire which contains Control Constructs. All Control Constructs fall within a single Master Sequence</td>
</tr>
<tr>
<td>StatementItem</td>
<td>Structural</td>
<td>Allows for the insertion of a statement (text or object)</td>
</tr>
<tr>
<td>ComputationItem</td>
<td>Structural</td>
<td>Supports Conditional Constructs and can insert validation or other checks</td>
</tr>
<tr>
<td>QuestionConstruct</td>
<td>Structural</td>
<td>Inserts a Question</td>
</tr>
<tr>
<td>IfThenElse</td>
<td>Conditional</td>
<td>Supports splitting the path through a questionnaire based generally on the response to one or more questions</td>
</tr>
<tr>
<td>Loop</td>
<td>Conditional</td>
<td>Repeat the sequence of the loop until the condition is satisfied. Sets up initial value, step, and condition</td>
</tr>
<tr>
<td>RepeatWhile and RepeatUntil</td>
<td>Conditional</td>
<td>Repeat “while” a condition is true or “until” a condition is met</td>
</tr>
</tbody>
</table>

Creating questionnaire flows

• If you want to reuse a specific order of questions create a Sequence and reuse it (e.g. block of questions in a panel study)
• It is easier to manage bundles of sequences than a long string of questions; organise your questionnaire flow
• If you need to manage the flow of data through your questionnaire use Input / Output Parameters and Binding
Thank you for filling this in. Children of the 90s loves to look at the things you draw!
Questions live in `<QuestionItem>`

- URN
- Display Language
- Question Name
- Question Label
- Question Text
- Response Domain (Code, Numeric, Text)

A1. **Do you ever have a headache?**

- yes, quite often
- yes, sometimes

**Question text**
Codelist - codes that reference categories

- URN
- Display Language
- Label
- Value
- Category reference
Categories

- URN
- Display Language
- Category Name
- Label

yes, quite often
yes, sometimes
yes, I had one once
no, never

```
<1:Category>
  <r:URN urn:ddi:uk.alspac:alspac_99_ttd-ca-156517:1.0.0 />
  <1:CategoryName>
    <r:String xml:lang="en-GB">156517</r:String>
  </1:CategoryName>
  <r:Label>
    <r:Content xml:lang="en-GB">yes, sometimes</r:Content>
  </r:Label>
</1:Category>
```
Conditions - IfThenElse

- URN
- Display Language
- Construct Name
- Description
- Command

```xml
<d:IfThenElse>
  <r:URN>urn:ddi:uk.alspac:alspac_99_ttd-if-018736:1.0.0</r:URN>
  <d:ConstructName>
    <r:String xml:lang="en-GB">c_qA1</r:String>
  </d:ConstructName>
  <d:IfCondition>
    <r:Description>
      <r:Content xml:lang="en-GB">If no, go to question A3 below</r:Content>
    </r:Description>
    <r:Command>
      <r:ProgramLanguage>pseudo-code</r:ProgramLanguage>
      <r:CommandContent> qc_A1 == 4 </r:CommandContent>
    </r:Command>
  </d:IfCondition>
</d:IfThenElse>
```

**Condition**

yes, I had one once

| 3 |

no, never

| 4 |

If no, go to question A3 below
Constructs combine to make a questionnaire.
Relationships in a questionnaire - Example
Relationships in a questionnaire - Example
Using parameters and binding

• If you need to manage the flow of data through your questionnaire use Input / Output Parameters and Binding

• In and out parameters assign an ID to the object entering a particular ‘box’ or exiting that ‘box’.

• Regulate the flow of information within an instrument

• It is recommended to track the parameter value at each step of the DDI questionnaire flow.
Parameters and binding - Schematic

Summary

• Documenting questionnaires and other data collection instruments provides rich information about both what was asked, the responses, and of whom and by who the questions were asked.

• This can be linked to variables, concepts, universes and other information that makes the data collection process more transparent, and provides a level of provenance for reproducibility.

• Structures in DDI-Lifecycle can assist better management of both question development and questionnaire specification.

• Reuse of questions both where they have been reused, and as a resource for future reuse creates a rich resource for harmonisation.