



Data Capture using Metadata

- Credits: input from meeeting at Eurostat July 2014. and part of presentation by Bryan Fitzpatrick
- im is to demonstrate designing and running a survey questionnaire based entirely on metadata
- Aim is to use DDI metadata
 - design the questions
 - organise the questions into a questionnaire
 - present the questionnaire
 - capture and save the responses
 - all based entirely on the metadata



Using DDI Metadata for Questionnaires

- DDI has metadata for Questions
 - a simple question goes in a Question Item
 - What is your age in years?
 - a complex question goes in a Multiple Question Item
 - Did you do paid work last week?
 - Full Time or Part Time?
 - How many hours?
 - A Multiple Question Item can contain Question Items or other Multiple Question Items

Using DDI Metadata for Questionnaires

- Questions can link to one or more Concepts
 - to indicate what the question is seeking to cover
 - Age, Sex, Country, Income, Occupation, ...
 - perhaps to qualify what is being covered
 - eg Non-farm income, Tertiary qualifications

Using DDI Metadata for Questionnaires

- Questions have:
 - Name
 - just a multi-lingual name, not used in questionnaires
 - Text
 - the question that is asked
 - can be conditional, multi-lingual, formatted
 - can even have mixed language
 - Question Intent
 - some elaboration about what is being sought
 - multi-lingual, formatted

Using DDI Metadata for Questionnaires

- Questions have Response Domains
 - what sort of answer is expected or valid
 - Numeric domain
 - can specify integer or decimal, valid formats and ranges, etc
 - Text domain
 - can specify format, length
 - Category Domain
 - valid list of multi-lingual values
 - not really very much use
 - Code Domain
 - valid list of multi-lingual values with codes
 - a classification

Using DDI Metadata for Questionnaires

- Questions do not go directly into a questionnaire
 - DDI calls a questionnaire an **Instrument**
- questions constitute a library available for use
 - a "Question Bank"
- questions are selected and assembled into an Instrument
- the assembling of questions is done with Control Constructs
- an Instrument identifies a single Control Construct that builds the questionnaire

Control Constructs

- Control Constructs are the critical component in building a questionnaire
 - they select the questions
 - they control the flow of the questions
 - branching and looping
 - they insert non-question text
 - "Now I want to ask you about other people in the household"
 - they can compute values
 - they link to Interviewer Instructions
 - structured DDI Interviewer Instructions
 - unstructured external interviewer instructions material

Control Constructs

- Several types of Control Constructs
 - Question Construct
 - selects a Question Item or Multiple Question Item
 - Sequence
 - selects a sequence of other control constructs of any type
 - If-Then-Else
 - defines an If condition with optional Elself clauses (multiple) and optional Else clause
 - each condition selects a single Control Construct to include

Control Constructs

- Several types of Control Constructs
 - Loop, Repeat-Until, Repeat-While
 - eg to loop over people in a household
 - Statement Item
 - inserts non-question multi-lingual text (conditional, formatted)
 - Computation Item
 - a calculation in some language that is assigned to a Variable

Instrument

- Identifies a single Control Construct to assemble the questionnaire
 - probably a Sequence construct
- Instruments can have multiple Software specifications
 - basically just identifying "software" used with instrument
 - Colectica: generate code for Blaise, Redcap etc

Instrument

- Instruments do not have any place for useful layout metadata
 - just the type of the layout
- We need quite a lot of information to do the layout
 - how to represent lists
 - tick boxes, list boxes, combo boxes, radio buttons
 - how to show flow logic
 - which questions to show at once, which to separate
 - can the respondent backtrack
- Colectica made improvements – flow diagram etc

Interviewer Instructions

- A formal DDI metadata type
- Organised, structured instructions
 - formatted multi-lingual text
 - may be conditional
- May link to external, non-DDI material
 - eg, PDF, Word documents
- Not used in this Proof of Concept

Classifications

- DDI holds Classifications as linked Code Schemes and Category Schemes
 - a Category Scheme is a list of Categories
 - flat list of multi-lingual names and descriptions
 - eg, Country names, Occupation names, etc
 - a Code Schemes selects Categories from Category Schemes, assigns a Code (not multi-lingual), and may specify a hierarchy
 - a Code Scheme may select Categories from multiple Category Schemes
 - multiple Code Schemes may select the same Categories

Code Schemes and Category Schemes

- Used for
 - Classifications
 - a Classification is a Code Scheme
 - Controlled Vocabularies
 - lists of standardised terms
 - defined by DDI, an organisation, a local area
- Used for
 - Response Domains for Questions
 - Representations for Variables
 - Code Representation

Variables

- A Variable is a container that will hold a data value
 - has a Name and Description (both multi-lingual)
 - can be linked to a single Concept
 - to indicate what the data represents
 - can be linked to multiple Questions
 - to indicate where the data comes from
 - can have a Representation
 - Code, Date/Time, Numeric, Text
 - with constraints on values
 - can identify a Response Unit and an Analysis Unit
 - a population that it can apply to

Logical Record

- A Logical record consists of a sequence of Variables
 - groups data values for a purpose
 - data from a questionnaire goes into one or more Logical Records
 - Logical Records can be linked
 - eg, Households and Persons
 - Logical Records are independent of any storage or stored format

Record Layouts and Physical Structures

- Map a Logical record to a physical record and an actual stored file format
- Can support a very wide range of structures and storage formats
 - CSV, Binary file, XML, database
 - multiple record types, linkages of many kinds

Physical Instance

- Holds information about actual data sets produced
 - links to Physical Structures, Record Layouts, and Logical records
- provides a central management of data from a collection



Use case 1 Demo




