Metadata Review

Introduction

The Metadata Review process is where the metadata owners (i.e. the studies) to review, enhance and approve their metadata ready for use in the search platform. This process includes:

- Expert review of documented questionnaire metadata
- In some cases, mapping of variables/derived variables to questions
- In some cases, linking of the 'topic' controlled vocabulary
- Expert review of any mapping or linking already completed
- Listing of any items for redaction
- Signing off of approved metadata

To facilitate this, a metadata bundle will be delivered for each instrument. A metadata bundle is a collection of all the files that will be fed into Colectica Repository in order to populate the Metadata Search Platform (MSP). A detailed breakdown of the contents of the bundle can be found in **Error! Reference source not found.**.

How to Review

Once you have downloaded the files in the bundle:

1. Unzip the file "caddies.zip".
2. Open "index.htm" in Mozilla Firefox[i].
3. Verify that the questions, answers, statements and questionnaire flow is an accurate rendering of the original questionnaire.
4. If you are conducting the mapping process, read the 'How to map variables' section below. If not, skip to step 5.
5. Verify that the variables attributed to each question are correct. Variables are displayed in blue. Derived variables are show in red and green.
6. Verify that the topics linked to each sequence are correct. This refers only to top level topics. A separate process will address question specific topics and lower level topics.
7. Check the two lists at the bottom of the page for the questions and variables that could not be mapped to ensure these are valid exceptions.
8. List all questions/statements you would like redacted from public display.

The questionnaire has been documented within the DDI structure as closely as possible. There will be cases where this is difficult and we have applied the following principles (in order of priority):

1. Maintain and do not alter the semantic meaning of the questionnaire

2. Do not correct the questionnaire

3. Only record what is contained within the questionnaire

4. Do not allow the data recorded to inform the metadata archiving

In Appendix B below is a screenshot of the index.htm file identifying the different components for your reference.

How to Map Variables

1. Open the folder "resources"
2. Open the file "mapping.xlsx"
3. The blue tab 'qlist' is populated with the questions to be mapped.
4. Populate the blue tab 'variables' with your variable data. You must use the prescribed format and headings.
5. In the red tab 'mapping' drag/copy the formulae in columns A to F down the sheet until all of your variables appear in Column A.
6. In the ID Column, type the question ID that relates to that variable. You should see the question text appear in Column F. The question IDs are listed on index.htm.
7. If the question is a grid (visible on index.htm) and the variable links to a specific cell within that grid, you will need to populate Columns D and E with the cell reference. See Appendix C for grid coordinates.
8. If a variable relates to more than one question, you will need to duplicate the variable in Column A of 'mapping' for each question and list each question in Column C.
9. If a variable is derived, leave Column C blank and **copy** the variable ID into Column A of 'DV mapping'. Do not remove it from 'mapping'.
10. In the red tab 'derived variables' you will now have a list of all derived variables in Column A. Drag/copy the formulae in Columns B to H down to cover all variables; Column B will be populated.
11. For each derived variable, create an additional entry in Column A for every variable it maps to. List that variable in Column C. Columns D to H will then populate.
12. Where there are multiple levels of derivation, Columns D to H will not populate until you reach the bottom level.
13. If the derived variable does not map to another variable, but to a question, add the question ID in Column E. You will need to add grid co-ordinates as per step 7 if relevant.
14. The two green tabs (mapping.txt and dv.txt) are used produce the output (.txt) files in the correct format for both variable (mapping) and derived variable mapping (dv) files. In both cases, drag/copy the formula down until all variables are displayed.
15. Copy the contents of mapping.txt into a Notepad file and save this as mapping.txt. Repeat this for dv.txt, saving it as dv.txt. It must be in the main bundle folder.
16. Close and re-open index.htm to see your changes displayed.

How to Link Topics

1. Open the folder "resources"
2. Open the file "linking.xlsx"
3. The blue tab slist is populated with all the sequences that appear in your metadata. Sequences are sections into which the questionnaire is divided. They can be children of other sequences, but this is not denoted in slist.
4. The blue tab CV is populated with a copy of the top level topic controlled vocabulary. This is fixed and must not be changed. Each topic has an ID number.
5. If required, in the red tab 'linking' drag/copy the formula in columns A and B down until all the sequences appear.
6. You should then select the relevant topic from the drop down list in column B in 'linking'.
7. If required, in the green tab 'linking.txt' drag/copy the formulae in columns A and B down until all the sequences appear.

8. Copy the contents of linking.txt into a Notepad file and save this as linking.txt. It must be in the main bundle folder.
9. Close and re-open index.htm to see your changes displayed.

Reporting Issues/Redactions

If you find an issue in the documented metadata, please list it in the relevant section below.

Issues with questionnaire metadata

```



```

Items for redaction

```



```

If you find issues in variable/topic mappings, you can either list them below or change them in the relevant spreadsheets:

- mapping.xlsx contains variable/derived variable mappings

- linking.xlsx contains topic mappings

```



```

Appendix A - Documents in the Metadata Bundle

Directory Structure

- index_files
- resources
    - mapping.xlsx
    - linking.xlsx
    - qlist.txt
    - slist.txt
    - CV.txt
    - [Optional useful files]
- ddi.xml
- dv.txt
- labels.txt
- linking.txt
- mapping.txt
- notes.txt
- redactions.txt

index_files

Please ignore this folder, it contains the various supporting files for index.htm (see below)

resources

This folder contains files that may be useful during the reviewing period (e.g. the Excel file use to map the variables or a scan of the questionnaire)

mapping.xlsx

This is the Excel file used to map the variables (both derived and normal) to the questions.

linking.xlsx

This is the Excel file used to link a topic from the controlled vocabulary to a sequence.

qlist.txt

This text file contains a tab delimited list of all the question literals and their related question ID. This file is used as an input for the variable mapping process (mapping.xlsx).

slist.txt

This text file contains a tab delimited list of all the sequences within the questionnaire. This file is used as an input for the topic linking process (linking.xlsx).

CV.txt

This file is a copy of the list of major topics from the agreed controlled vocabulary and their permanent IDs. This file is used to generate the topic linking (linking.xlsx).

ddi.xml

This is a DDI export of the questionnaire from Caddies. It contains all of the questionnaire metadata documented during the metadata ingest.

dv.txt

This is one of the six configuration files. Using a tab delimited format it describes the mapping from the derived variables to either other derived variables, normal variables or questions. This files is generated by mapping.xlsx.

labels.txt

This is one of the six configuration files. It contains a tab delimited list of labels that require replacing during import into Colectica Repository. The first column is the label as it is currently in ddi.xml and the second column is what the label will become.

linking.txt

This is one of the six configuration files. Using a tab delimited format it describes which major topics are linked to each of the sequences within a questionnaire. This file is generated by linking.xlsx.

mapping.txt

This is one of the six configuration files. Using a tab delimited format it describes the mapping from the variables to the questions. This files is generated by mapping.xlsx.
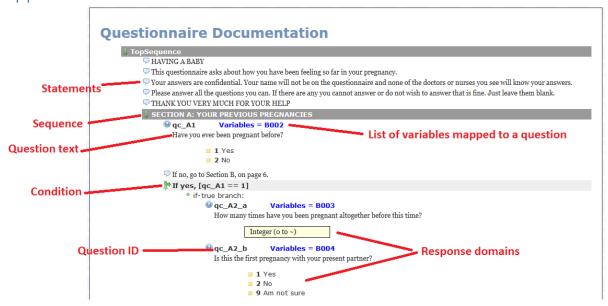
notes.txt

This is one of the six configuration files. It is used to attach a note to a question.

redaction.txt

This is one of the six configuration files. This contains a list of all questions and statements that are not to be made public for a given questionnaire.

Appendix B - Screenshot of index.htm

Appendix C - Grid Coordinates

	In the year before this pregnancy	In the first 3 months of this pregnancy	From 4 months of this pregnancy until now
	1-Yes 2-No	1-Yes 2-No	1-Yes 2-No
Did/do you use a VDU? (television type screen)	1;1	2;1	3;1
Are/were you mostly sitting?	1;2	2;2	3;2
Are/were you bending a lot?	1;3	2;3	3;3
Are/were you standing much of the time?	1;4	2;4	3;4
Are/were you doing repetitive, boring tasks?			
Did/does your job involve challenging and mentally demanding tasks?			
Are/were you using a lot of physical energy?			
In your job are/were you in contact with fumes or chemicals? (please describe)			

---

i Caddies has only be tested to work in Mozilla Firefox. If you cannot use Firefox the next best alternative would be Google Chrome.