Paper #DM01

The Metadata Survey Method, Phase II: Tools For General Use

Derek Morgan, Washington University Medical School, St. Louis, MO

ABSTRACT

The metadata method was developed to bypass obstacles related to staff resources for end-user data capture. It uses SAS® datasets to implement program control as well as provide questions and responses. A fixed series of SAS/AF® frames interact with the user, yielding a consistent user interface. The frame code does not change between surveys—the datasets do. Phase 1 consisted of the development and deployment of a prototype application to administer one specific survey.

Since the initial deployment of version 1.0 of the metadata software, the method has been refined, and design and development improvements have been made towards the ultimate goal of turning the metadata method into a system for end-user data capture. In order to utilize version 2 of the metadata software, a series of tools have been created so that semi-technical users may create and modify the control datasets necessary to implement the metadata system. These tools allow, with the exception of any custom frames necessary, rapid survey deployment from a paper form template without direct intervention from programming staff. This paper will briefly talk about the metadata method and its development, but the focus will be on showing and explaining the tools by showing how to create a sample survey form from a paper template to illustrate how these tools are used by staff.

WHAT IS THE METADATA METHOD?

It is a series of SAS/AF frames that administer single-response (select one) questions, multiple-response (select all that apply) questions, or free-text (character or numeric) entry. The frames do not change throughout the survey, but the objects are filled and/or activated according to the metadata contained in two SAS data sets. The first is the question data set, or QDS. The QDS is responsible for the flow-of control for the questionnaire, the presentation of the question text, the definition of the response group for each question, and the name of the variable where the data for each question are to be stored. The second data set is the response data set, or RDS. The RDS contains all of the standardized responses used in a questionnaire, grouped into display categories. These categories correspond to each unique set of responses for all the questions in a questionnaire. The method was originally developed to administer a long (407 questions) dietary survey instrument, because neither the FSEDIT procedure nor any of the standard SAS/AF data entry tools was an attractive option for the task.

HOW DOES IT WORK?

The QDS contains five variables related to flow-of-control, and one variable to indicate the order of the questions on the paper form. This ordering variable is only used to sort the data set for production of hardcopies of the metadata in question order, but it is not responsible for the survey's flow of control. By not using this ordering variable for the survey control, we allow for the insertion/deletion of questions without regard to their physical location in the QDS.

The first of the five variables uniquely identifies each record in the QDS without regard to its order. Two of the four remaining variables define the previous and next questions in the sequence. Normal survey flow of control requires a variable to dictate the next question in the survey sequence. In addition, since this is to be a computer-assisted interview, allowing a respondent to go back and change an answer would be a good thing, so we created a variable to represent the question previously asked. Both of these variables assume that there is no skip logic, or that the skip pattern was not triggered. If there is skip logic involved, the application needs to know where it’s going and what makes it jump, so this requires two variables as well: one to define the response that triggers the skip pattern, and another to define the next question in the survey sequence when the skip pattern has been invoked. Together, these four variables define the entire flow of control within the survey from the QDS.

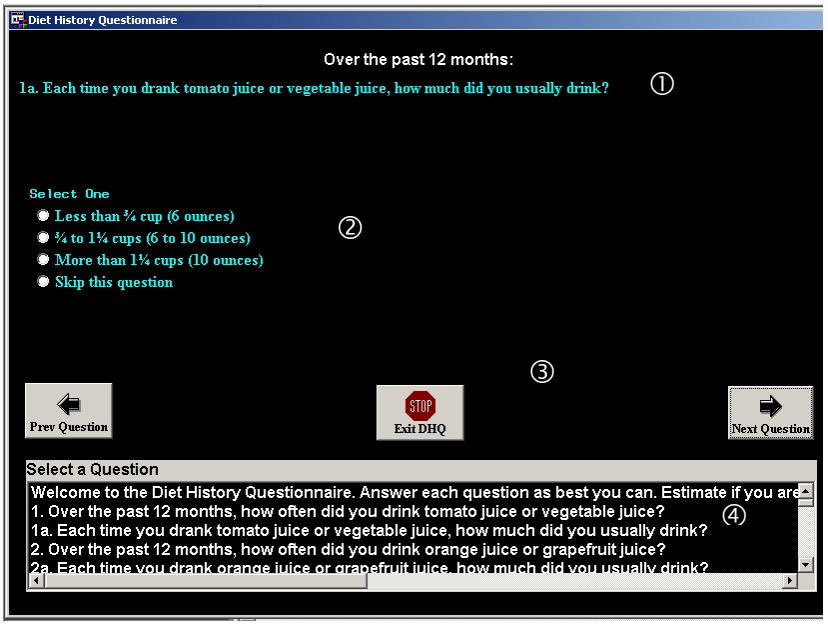
One column in the QDS contains the complete question text to be displayed, while another holds the name of the response category to be used. There is also a column that represents the name of the variable where the response is to be stored. Multiple response questions are stored as yes/no variables, one per possible response. This allows us to use a variable prefix in the QDS and the program sequentially numbers each response. Another column indicates the single response in a multiple-response answer that will cancel any positive responses, e.g., “none of the above.” There is a column to indicate whether the value entered in a text frame is to be stored as a character or numeric variable in the data capture table. Finally, there is a column in the QDS that tells the application whether to display a hard-coded prefix phrase with the question text. This is the complete list of variables in the QDS under version 1 of the metadata method:

TABLE OF VARIABLES IN THE QUESTION DATASET (QDS)

order	Num	4	Order that questions should appear (used for sorting ONLY)
q_no	Char	32	Question number (used by program)
q_text	Char	300	Question text as it is to be displayed
rcat	Char	32	Response category (R=single response, M=multiple-response, T=text field)
skip_to	Char	32	Question number (q_no) to skip to
skip_trigger	Num	4	Answer to trigger skip pattern
prev	Char	32	Previous question in sequence (q_no)
next	Char	32	Next question in sequence (q_no)
vname	Char	32	Dataset variable name associated with question
exc	Num	3	Exclusive answer in multiple-response category (ONLY used with M)
type	Char	1	Char or Num variable (for text (T) answers)
use_common_text	Num	3	Turn pre-text box on or off?

The main control frame incorporates the administration of single-response questions.

SINGLE RESPONSE FRAME

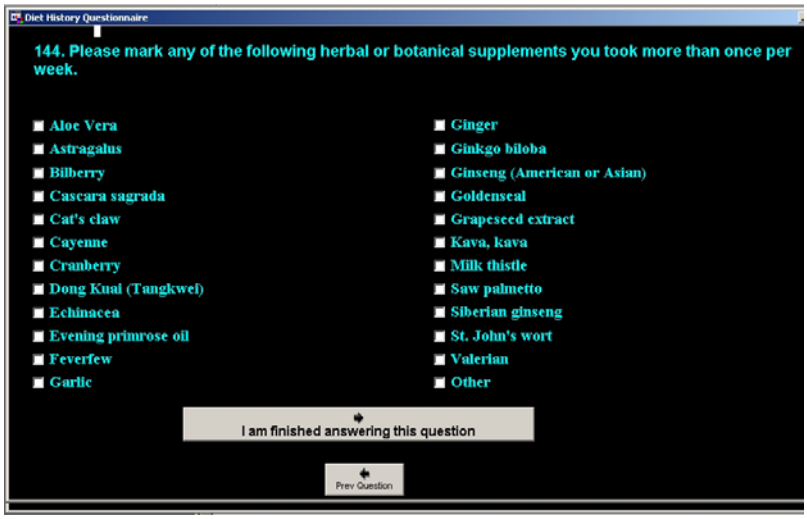


This is what the single response frame looks like. The question text is at the top (❶). The question prefix is centered, and in white. It uses a different font than the rest of the frame to make it stand out visually. A variable (USE_COMMON_TEXT) in the QDS controls whether it is visible or not. The blue text in ❶ is the question text (Q_TEXT) as specified in the QDS. The question area runs the entire width of the frame, but this particular question does not use the entire space. Item ❷ is the response box area. This area also encompasses the width of the frame, since there are several questions with more than one column of possible single-choice responses. The list that fills the radio box is generated dynamically from the RDS. The value of response category (RCAT) in each QDS record defines which records from the RDS are used to create the SCL list to be used. Area ❸ is the navigation area. It contains push buttons that cause the application to go to the next or previous question, and a stop button. The forward and backward buttons are only displayed if the survey context allows, that is, the last question shows no next button, while the first question shows no previous button. The combo box at the bottom (❹) is intended for administrative use only, and allows direct access to any question in the survey. It can be made invisible based on a toggle. In production for the general population, the toggle is a check box in the opening frame, but in our study, the toggle is based on the user's authorization level.

The elements required for multiple-response questions are the same as for single response questions, except that instead of getting the response from a radio box of mutually exclusive options, multiple answers are obtained from a list of checkboxes that can be independently checked. The respondent also must explicitly close the question. It is implicit that selecting one

answer in a single-choice will cause the survey to advance; the question has been completely answered. In a multiple-response situation, there is no way of knowing how many checkboxes the respondent wants to select. Therefore, the respondent must take some sort of action to signal that all the desired choices have been checked.

MULTIPLE RESPONSE FRAME



This is the two-column, 24-checkbox version of the multiple response frame. There are twenty-four choices shown in the above screenshot, which also happens to be the maximum number of choices for multiple-response questions in the DHQ. The two columns of 12 items are populated with a list from the RDS. The left column will be filled first, from top to bottom, and then the right. The question text from the QDS is at the top. Note that of the two push buttons at the bottom; "I am finished answering this question," is the preferred option as indicated by its size. The 12-choice, single-column version is identical in appearance except that there are only 12 checkboxes that fill the width of the screen.

Text responses are questions that require the user to type in text; and there are two identical, overlaid text box objects in the frame. Which one is active depends on whether the variable to be saved from this frame is character or numeric, which is stored as a part of the QDS in the variable TYPE.

Another feature of the metadata system is the ability to stop in the middle of a survey and have it resume where you stopped. Survey data are written to the dataset as each question is answered; therefore, in case of a system problem, only the data for the most recent question will be lost.

This allows simple survey administration where the bulk of control is defined via SAS data sets, and avoids the construction of custom FSEDIT screens or SAS/AF table and/or form viewers.

CHANGES AND IMPROVEMENTS TO RELEASE 1

The metadata system was originally designed for a specific survey with only three possible types of data entry. In order to expand the concept beyond this narrow scope, we had to think about what would be necessary for the use of this method as a generalized data entry tool. After all, what good does it do to have a standard method where you still have to custom code fifty percent of the pieces? A few more standard FRAMES were created and added to the package to administer the following types of data entry:

- Date (using a calendar.)
- Date as month, day, and year separately.
- Date as day, month, and year separately.
- Data where the units can be specified as either/or (e.g., pounds or kilograms).
- The metadata method was also enhanced to capture additional text if "Other (please specify)" was listed as an answer.

Another improvement centers around the common text; in version 1, the text was hard-coded, and the object that displayed the text was simply toggled on or off by the USE_COMMON_TEXT variable. Instead of a simple toggle, version 2 of the metadata system utilizes another dataset, one that contains header text. This allows the use of many different headers throughout the system.

The biggest enhancements, however, involve the skip logic. When you skip a question, the questions that are bypassed are only missing if they are not answered. What happens when someone goes back to a prior question after having followed the standard flow of control and then invokes a skip pattern? In release 1, the prior answers would remain. Now the metadata system will go through all of the skipped questions and explicitly set the responses to missing.

The second enhancement removed the skip logic control from the QDS, and placed it into a metadata data set of its own. This was done to address the situation where you can go to more than two possible destinations from a given question based on the responses to that question. In the old system, you could go to the next question or a different question based on one specific response. This caused problems when you had a structure such as:

If the answer is:	
Yes	Go to question 9b
No	Go to question 10
Maybe	Go to question 9d

Now the metadata system can handle an infinite number of destinations from single-response and numeric text response questions. From the example above, you could specify the default destination (next question) as question 10, and instruct the system to go to question 9b if the answer was "yes," or question 9d if the answer was "maybe."

All these enhancements lead us to one question: now that the metadata system is more robust than ever, how do we get people to use it?

And that brings us to phase two.

THE METADATA TOOLS

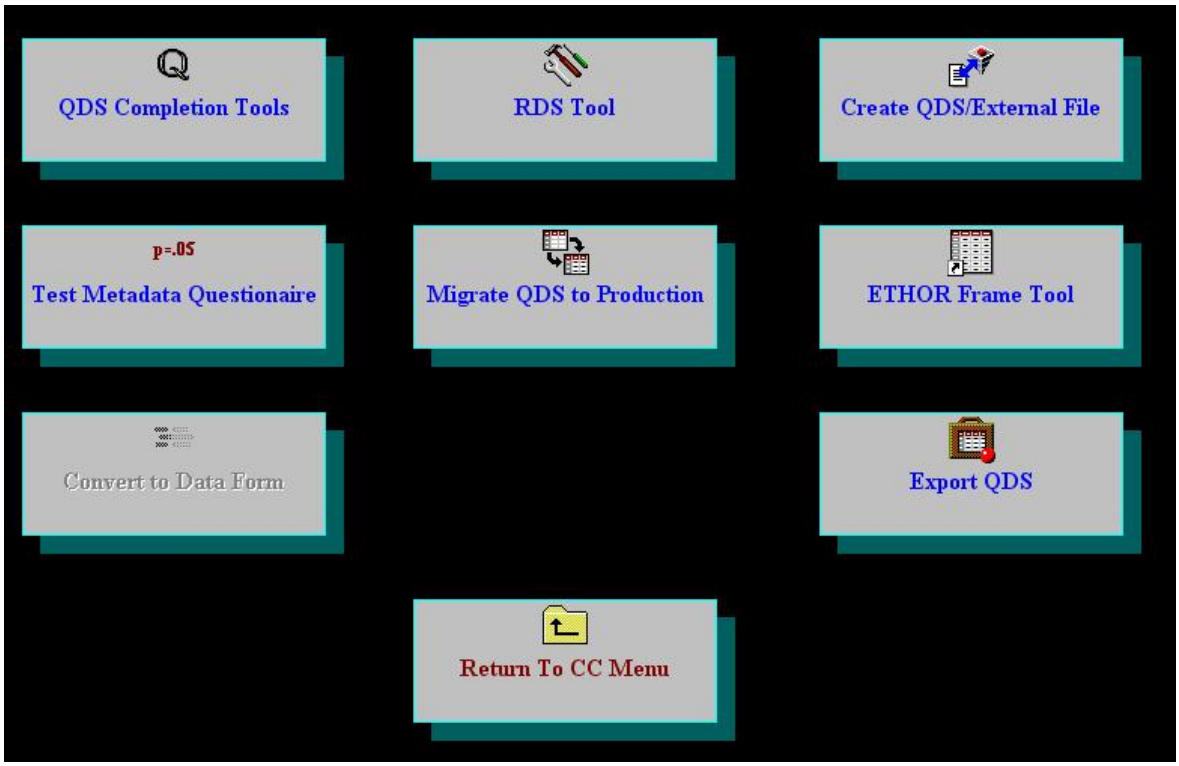
In order to allow for general usage of the metadata system, we must make it more accessible to non-technical users. Although the SAS data sets used are very simple in structure and can be created via an INPUT statement, or the IMPORT procedure from an Excel spreadsheet (since everybody seems to like to put things into Excel these days), or a VIEWTABLE window, or even Enterprise Guide, that still implies opening SAS and working with it to yield the data sets in the format that the metadata system requires.

We have created a series of tools to allow a non-technical user to start with an electronic version of a questionnaire, and by using these tools, accessed through a SAS/AF application, create a metadata system survey. The tools are:

- QDS2TXT: Creates the basic QDS from a raw text file, or will write out a finished QDS to a text file.
- RDS_TOOL: Allows manipulation of the RDS; create/delete/modify response groups from a text file or by direct data entry.
- QDS_TOOL: "Finishing" tool for QDS. Once the QDS is created from the text file, this allows interactive manipulation of the QDS. Define response categories, skip logic, variable names, and create header information. Also allows insertion/deletion of questions.

A second set of tools are used to create menus to access several surveys, as would be used in medical studies, add categories to the either/or frames, and to transfer the QDS to another operating system using Cross-Environment Data Access.

Metadata Tool Menu



The best way to show the user tools and to demonstrate their usage would be to follow the process of creating a metadata questionnaire from start to finish, so let's start with a sample paper form, and turn it into a metadata survey.

SAMPLE QUESTIONNAIRE

Here is a simple 2-page questionnaire for the Men In Black immigration desk. All aliens and returning earth space travelers must be registered with this questionnaire.

MIB Sol-3 Visitor Questionnaire

Interviewer Script: Welcome to Sol-3 (Earth) Intergalactic Customs. Before you may enter the Sol-3 zone, I need to ask you a few questions. You will be issued a zone visitor's visa at the completion of this brief interview.

A. **Arrival Date:** ☐☐-☐☐☐-☐☐☐☐
day month year

B. **MIB Staff Person Identifier:** ☐☐☐☐☐☐☐☐

1. Are you animal, mineral or vegetable?

- ☐ Animal
- ☐ Vegetable
- ☐ Mineral
- ☐ Other (please specify) _____

2. Are you a resident of the planet Earth?

- ☐ Yes **Go to Question 4.**
- ☐ No

3. What is your home galaxy? _____

4. What is the primary purpose of your visit?

- ☐ Business
- ☐ Pleasure
- ☐ Domination ***Please go to the Dept. of Megalomania in bldg. 5, rm 30300903A.***
- ☐ Complete Destruction ***PRESS PRE-EMPTIVE ANNIHILATION BUTTON IMMEDIATELY!***
- ☐ Re-settlement ***Please go to the department of residency transfer on level 6.***
- ☐ Returning Native

5. How long are you planning to stay? ☐☐☐ **days** ***(if 0, If you are only using the planet as a jump point, you do not have to register.)***

6. What is your date of birth or age (Sol-3 equivalent)?

DOB: ☐☐-☐☐☐-☐☐☐☐ **OR** **Age:** ☐☐☐☐☐☐
day month year

7. What vital organs are you missing (check all that apply)?

- ☐ Brain
- ☐ Eyes
- ☐ Ears
- ☐ Organs for Verbal Communication
- ☐ Organs for Telepathic Communication
- ☐ Organs for Visual Communication
- ☐ Heart
- ☐ Lungs/Gills
- ☐ Organs to process non-oxygenated environments
- ☐ Reproductive Organs
- ☐ Other (please specify) _____

8. How many limbs do you have? ☐☐☐☐

9. What is your familiar designation? _____

10. In case of emergency, how should we contact you?

- ☐ Telephone
- ☐ Subspace
- ☐ Non-visible light transmission
- ☐ Telepathic network transmission
- ☐ Other (please specify) _____

11. Please provide your contact frequency. _____

12. Please confirm the purpose of your visit.

- ☐ Business
- ☐ Pleasure
- ☐ Domination ***Please go to the Dept. of Megalomania in bldg. 5, rm 30300903A.***
- ☐ Complete Destruction ***PRESS PRE-EMPTIVE ANNIHILATION BUTTON IMMEDIATELY!***
- ☐ Re-settlement ***Please go to the department of residency transfer on level 6.***
- ☐ Returning Native

Thank you. Please enjoy your stay on the third planet of the Sol system.

BUILDING THE METADATA

The first step in building the metadata is to create the template data set for the data being collected, or the data capture table (DCT). This should contain all the variables in the survey, as well an ID variable(s) to identify the record, and three internal record keeping variables for the metadata system. Two of these variables track the entry person and a record modification datetime value. The third variable is used by the metadata system, and it keeps track of the last question answered in the survey, which is what allows a user to return to the same survey at the point where he or she left it.

Code to Create the Template for the SAMPLE data set

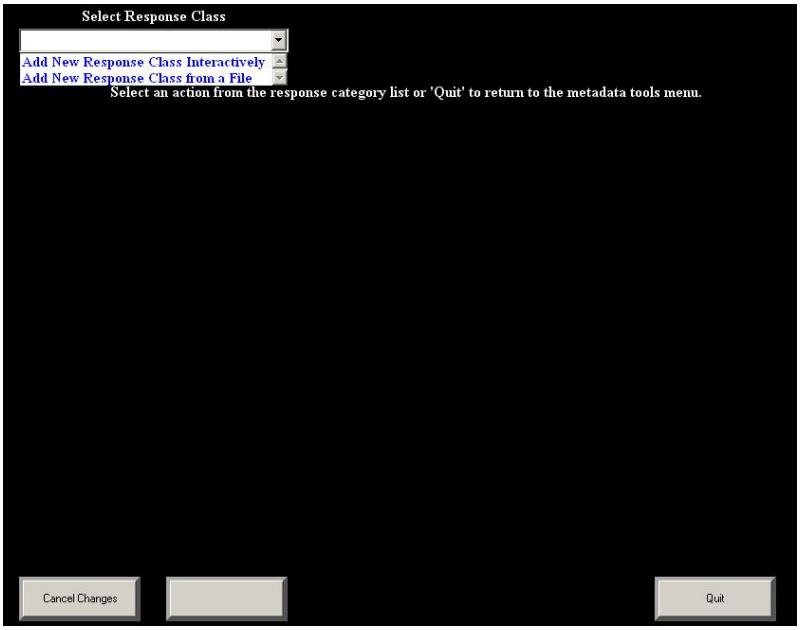
```
1   DATA clinic.sample (ENCRYPT=YES READ=&rdpswd WRITE=&wrpswd ALTER=&altpswd)
2   /* For information on how the macro variables are used in line 1, see the second item in
    the "References" section */
3   LENGTH id $ 5 date 4 staffno $ 8 amv 3 amv_othr $ 80 resident 3 home $ 80 purpose 3
    duration_of_stay 8 lifespandate 4 lifespanage 4 vitalorg1-vitalorg11 3 vitalorg_othr $ 256
    limbs 4 name $ 80 contact_via 3 contact_via_othr $ 80 contact 8 purpose2 3 entryid $ 8
    lastfld $ 32 touchdat 6;
4   FORMAT amv amv. resident vitalorg1-vitalorg11 yn. purpose purpose2 bpd. contact_via cmeth.
    version date lifespandate date9. touchdat touchdat18.;
5   INFORMAT amv amv. resident vitalorg1-vitalorg11 yn. purpose purpose2 bpd. contact_via
    cmeth. version date lifespandate date. touchdat touchdat.;
6   LABEL
7   date = 'Arrival Date'
8   staffno = 'MIB Staff Person'
9   amv = 'Are you animal, mineral or vegetable?'
10  resident = 'Are you a resident of the planet earth?'
11  home = 'What is your home galaxy?'
12  purpose = 'What is the purpose of your visit?'
13  duration_of_stay = 'How long are you planning to stay?'
14  dob = 'What is your date of birth?'
15  vitalorg1 = 'Missing vital organs Brain'
16  vitalorg2 = 'Missing vital organs Eyes'
17  vitalorg3 = 'Missing vital organs Ears'
18  vitalorg4 = 'Missing vital organs Verbal Communication'
19  vitalorg5 = 'Missing vital organs Telepathic Communication'
20  vitalorg6 = 'Missing vital organs Visual Communication'
21  vitalorg7 = 'Missing vital organs Heart'
22  vitalorg8 = 'Missing vital organs Lungs/Gills'
23  vitalorg9 = 'Missing vital organs non-oxygenated environments'
24  vitalorg10 = 'Missing vital organs Reproductive'
25  vitalorg11 = 'Missing vital organs Other'
26  vitalorg_othr = 'Missing vital organ description'
27  limbs = 'How many limbs do you have?'
28  name = 'What is your familiar designation?'
29  contact_via = 'In case of emergency, how should we contact you?'
30  contact = 'Contact frequency'
31  purpose2 = 'Confirm the purpose of your visit.'
32  ;;;;
33  STOP;
34  RUN;
```

Once the template dataset is in place, it's time to create the question and response data sets for the survey. It starts by using cut-and-paste to put the question text and responses into the information into plain text files that will be used by the metadata tools.

CREATING THE RDS

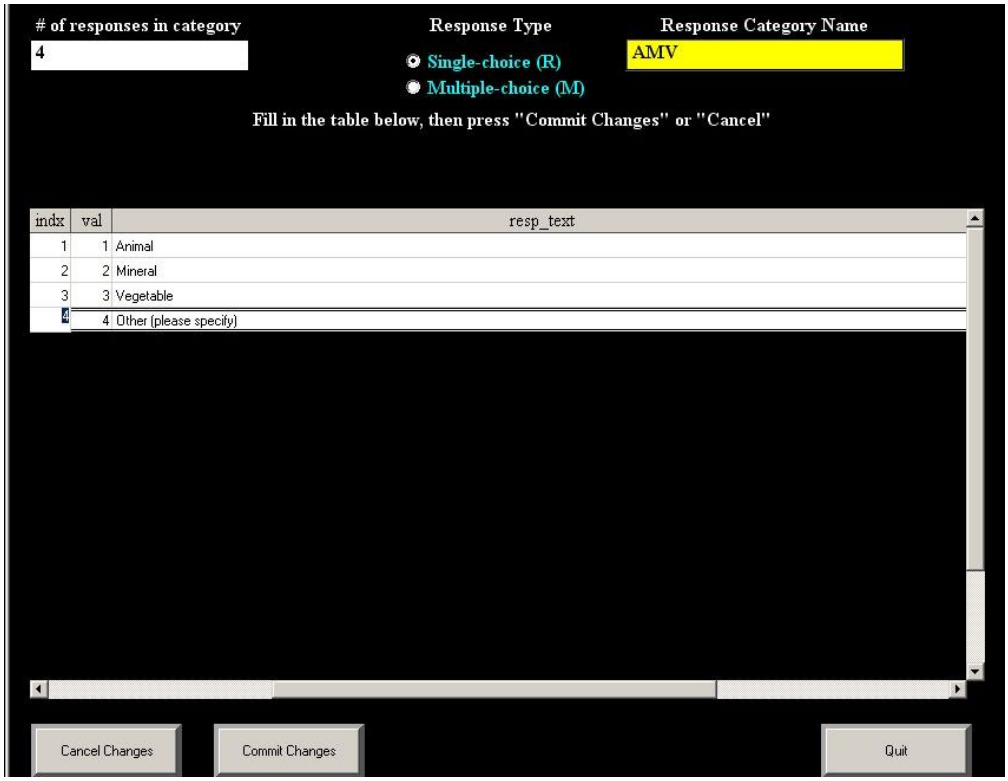
One of the principles of the metadata system is re-use of existing components, and this is especially true for the RDS. In practice, the response data set (RDS) grows over time, because many surveys use the same response groups (e.g., Yes/No). Any given set of responses only needs to be put into the RDS once, and it will then be available for use by any metadata survey. After the user clicks on the RDS Tool icon, the pull-down menu is used to determine which task to perform.

Starting the RDS Tool



Since this is going to be the first use of the metadata system (and therefore, a brand new RDS,) we will need to add the responses for all the questions from the survey. Let's do the responses for question 1 ("Animal"; "Mineral"; "Vegetable"; and "Other (please specify)") interactively. This is a single-response question with four possible responses. We have also given the response category a name, "AMV". If we left that blank, the response category would be assigned a sequential number, and would still be available for re-use. Simply type in the value that is to be stored in the database with that selection, and the text for each answer where indicated.

Entering a Response Category Interactively



We will use the external file method to add the remaining response categories. The file that the metadata system uses must be formatted in a very specific way.

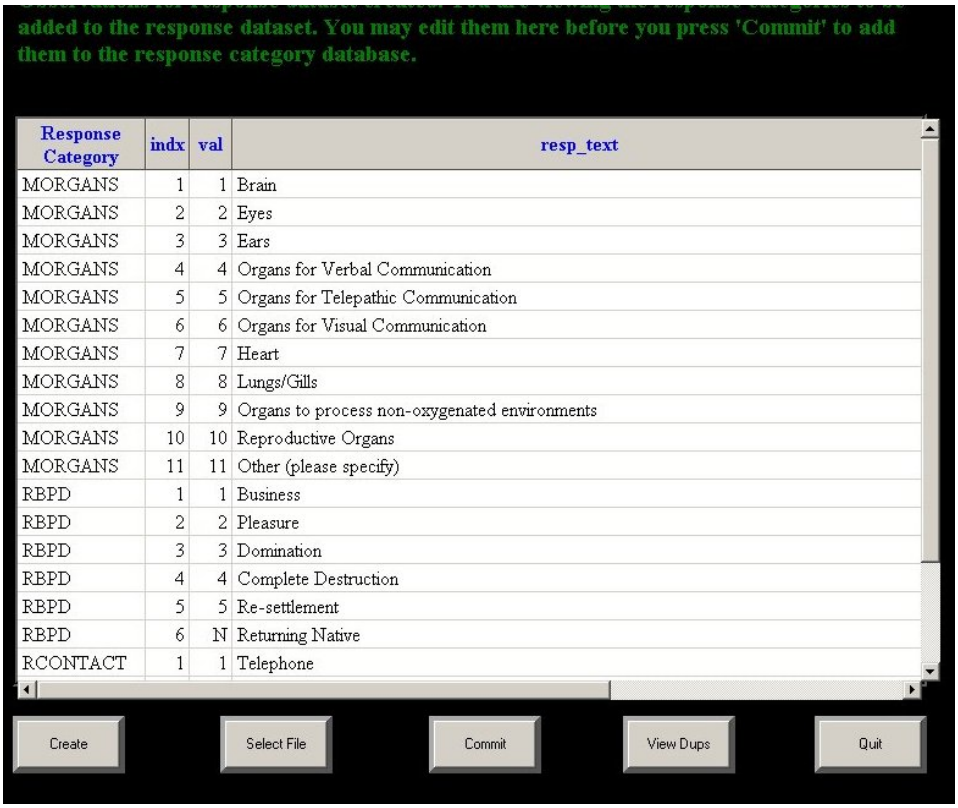
Metadata Response Input File

```
^RBPD¶
1 → Business¶
2 → Pleasure¶
3 → Domination¶
4 → Complete·Destruction¶
5 → Re-settlement¶
.n → Returning·Native¶
^MORGANS¶
1 → Brain¶
2 → Eyes¶
3 → Ears¶
4 → Organs·for·Verbal·Communication¶
5 → Organs·for·Telepathic·Communication¶
6 → Organs·for·Visual·Communication¶
7 → Heart¶
8 → Lungs/Gills¶
9 → Organs·to·process·non-oxygenated·environments¶
10 → Reproductive·Organs¶
11 → Other·(please·specify)¶
^RCONTACT¶
1 → Telephone¶
2 → Subspace¶
3 → Non-visible·light·transmission¶
4 → Telepathic·network·transmission¶
5 → Other·(please·specify)¶
```

The first line of each response group should begin with a caret character (^), followed immediately by the type of response group (R=single-response, M=multiple-response), and an optional name. If the name is omitted, a sequential number will be used to name the response group. The response levels must be listed in display order. Each line of the raw file contains the value that is to be stored in the data set for the given response, separated from the text of the response level by a tab character.

Notice also that the AMV category we just added is now in the drop down list. Clicking on an existing category will allow you to edit it interactively. Once you've selected the file name from a dialog box and pressed the "Create" button, you will see the response categories that you are about to create. You may edit them before you actually add them to the response metadata.

Adding Response Metadata from a File



Response Category	indx	val	resp_text
MORGANS	1	1	Brain
MORGANS	2	2	Eyes
MORGANS	3	3	Ears
MORGANS	4	4	Organs for Verbal Communication
MORGANS	5	5	Organs for Telepathic Communication
MORGANS	6	6	Organs for Visual Communication
MORGANS	7	7	Heart
MORGANS	8	8	Lungs/Gills
MORGANS	9	9	Organs to process non-oxygenated environments
MORGANS	10	10	Reproductive Organs
MORGANS	11	11	Other (please specify)
RBPD	1	1	Business
RBPD	2	2	Pleasure
RBPD	3	3	Domination
RBPD	4	4	Complete Destruction
RBPD	5	5	Re-settlement
RBPD	6	N	Returning Native
RCONTACT	1	1	Telephone

Once you've verified the response category name, the order (indx), the value to be stored (val), and the text to be displayed for each answer, you add it to the response metadata by pressing "Commit". Before a category is added, the existing metadata are checked to determine if you are trying to add a duplicate category. A duplicate category is one that has exactly the same order, and the same values and text for each response level. If this is the case, any duplicate categories will not be added, and the system will tell you what the existing category name is for each duplicate, so you can use that in the metadata questionnaire. This prevents redundancies in the response metadata.

CONSTRUCTING THE QDS

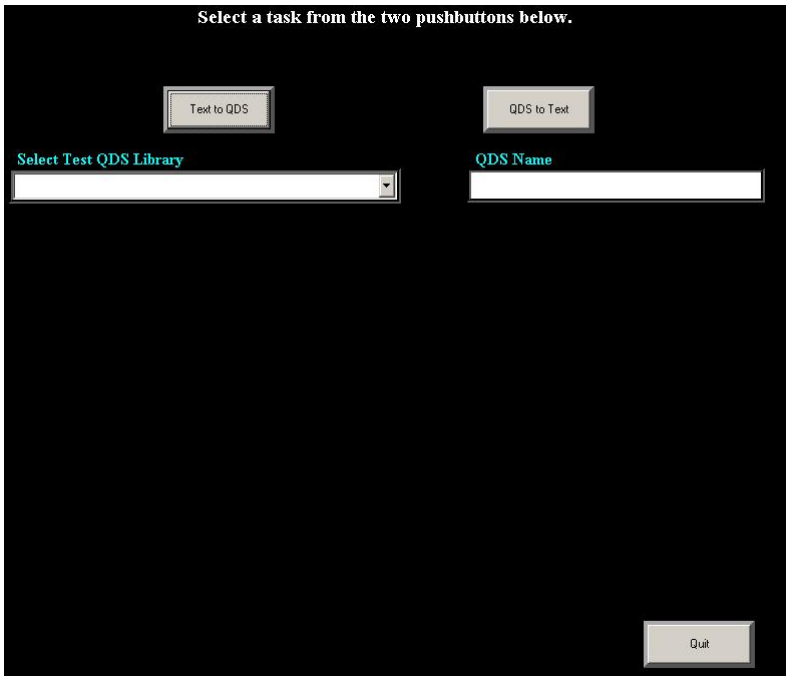
Now that we have the standardized responses ready, it's time to work on the questionnaire itself. This is a two-step process; first, we need to put the questions into a text file (cut-and-paste from your survey document file works great.) As with the response data, the system expects the questionnaire text file in a specific format:

The Questionnaire Text File

```
strt→ Welcome·to·Sol-3·(Earth)·Intergalactic·Customs.·Before·you·may·enter·the·
Sol-3·zone,·I·need·to·ask·you·a·few·questions.·You·will·be·issued·a·zone·pass·at·
the·completion·of·this·brief·interview.¶
A  →  Arrival·Date¶
B  →  MIB·Staff·Person¶
1  →  Are·you·animal,·mineral·or·vegetable?¶
2  →  Are·you·a·resident·of·the·planet·earth?¶
3  →  What·is·your·home·galaxy?¶
4  →  What·is·the·primary·purpose·of·your·visit?¶
X1 →  Please·go·to·the·department·of·residency·transfer·on·level·6.¶
5  →  How·long·are·you·planning·to·stay?·(Enter·.N·if·returnee)¶
6  →  What·is·your·date·of·birth·or·age·(Sol-3·equivalent)?¶
7  →  What·vital·organs·are·you·missing·(check·all·that·apply)?¶
8  →  How·many·limbs·do·you·have?¶
9  →  What·is·your·familiar·designation?¶
10 →  In·case·of·emergency,·how·should·we·contact·you?¶
11 →  Please·provide·your·contact·frequency.¶
12 →  Please·confirm·the·purpose·of·your·visit.¶
x2 →  Please·go·to·the·Department·of·Megalomania·in·building·5,·room·30300903A.¶
x3 →  PRESS·PRE-EMPTIVE·ANNIHILATION·BUTTON·IMMEDIATELY!¶
OK →  Thank·you.·Please·enjoy·your·stay·on·the·third·planet·of·the·Sol·system.¶
OKTRANS  →  Thank·you,·and·have·a·nice·day.¶
fini→ Press·"Exit"·to·end·survey.¶
```

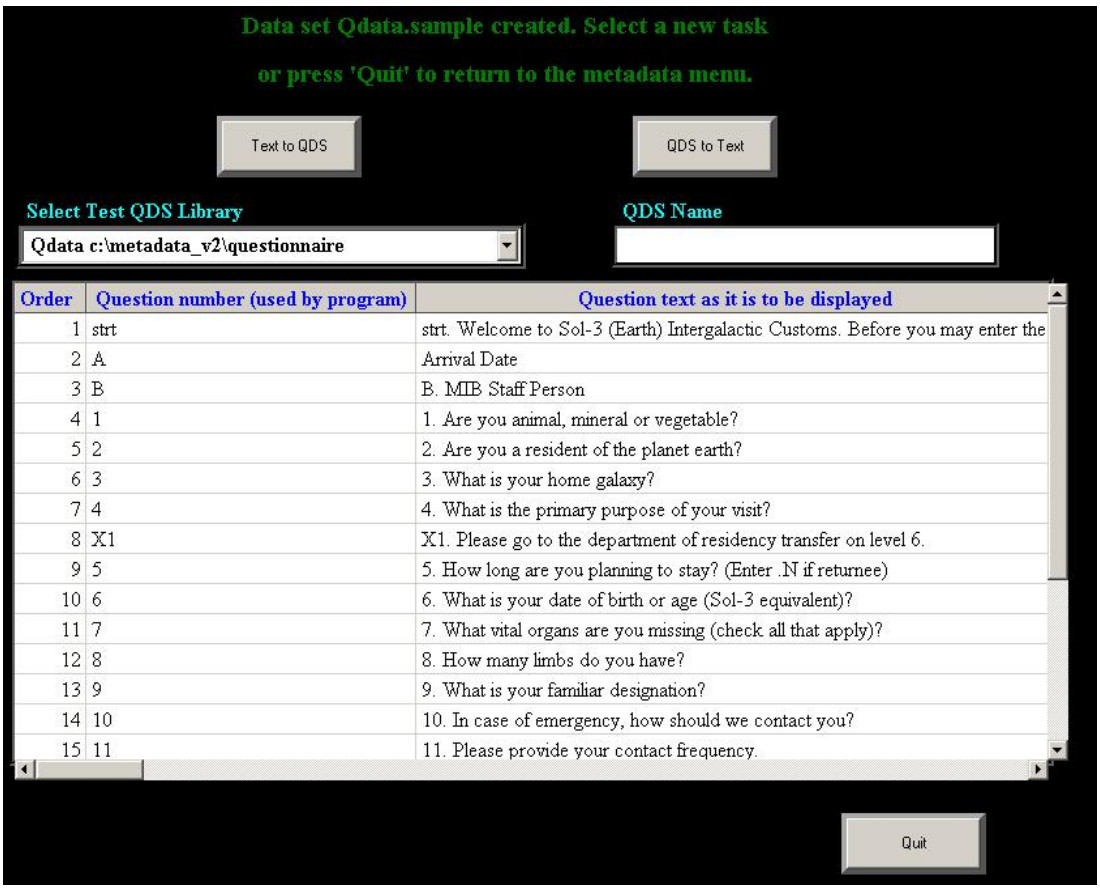
You can see that the file is again tab-delimited. The first field in each record is the internal question number identifier that we talked about earlier. The only fixed rules for naming the question records is that the internal numers are case-sensitive, and that the final question in the questionnaire must have the name 'fini'. Any prompts that do not require responses are also included in this file as well (items X1, x2, x3, OK, OKTRANS, and fini.) Once the file is ready, the next tool we use will be the "Create QDS/External File" tool. This is a two-way tool: it can create a QDS from an external text file, and it can also create an external file from a QDS (at any stage).

Create Basic QDS from Text File



You begin by selecting which task you want: creating a QDS from an external file, or creating an external file from an existing QDS. In this case, we want the first option, and we are prompted for the name of the external file. We select the library name from the pull-down list on the left, and enter a name for the QDS on the right. When these are filled in, and the external file has been selected, we press a button, and the basic QDS is built and then displayed.

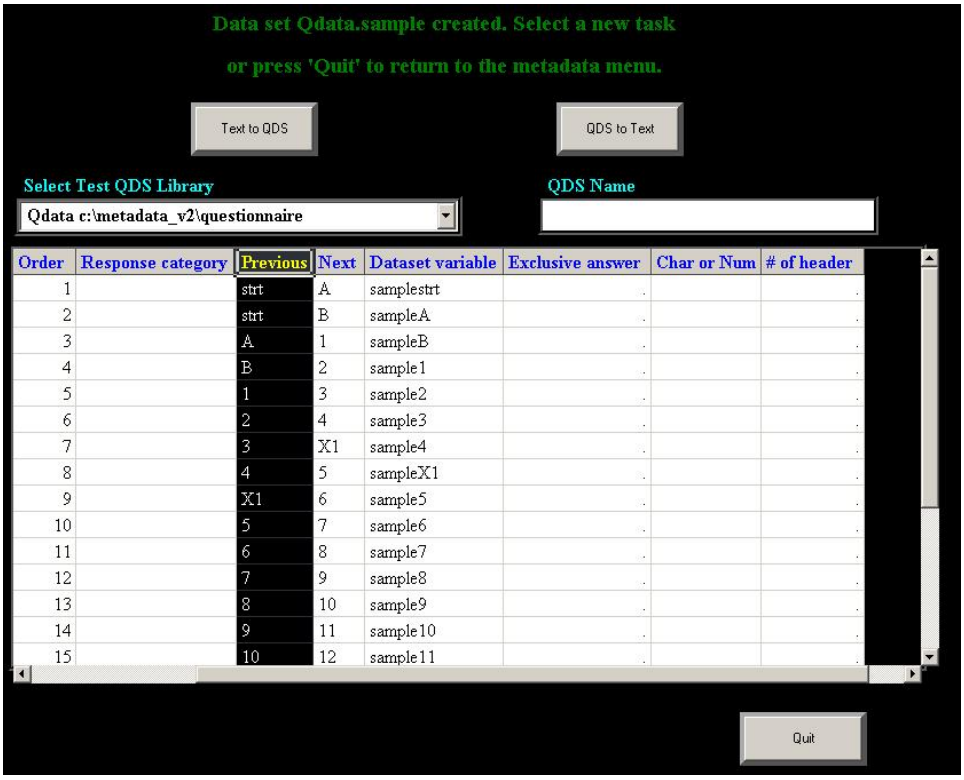
Base QDS from the External File



Order	Question number (used by program)	Question text as it is to be displayed
1	strt	strt. Welcome to Sol-3 (Earth) Intergalactic Customs. Before you may enter the
2	A	Arrival Date
3	B	B. MIB Staff Person
4	1	1. Are you animal, mineral or vegetable?
5	2	2. Are you a resident of the planet earth?
6	3	3. What is your home galaxy?
7	4	4. What is the primary purpose of your visit?
8	X1	X1. Please go to the department of residency transfer on level 6.
9	5	5. How long are you planning to stay? (Enter .N if returnee)
10	6	6. What is your date of birth or age (Sol-3 equivalent)?
11	7	7. What vital organs are you missing (check all that apply)?
12	8	8. How many limbs do you have?
13	9	9. What is your familiar designation?
14	10	10. In case of emergency, how should we contact you?
15	11	11. Please provide your contact frequency.

This is called the “base QDS” because only the question number and question text are truly ready to use. Default variable names are created, and default values of the QDS variables **prev** and **next** are filled in, providing the most basic flow-of-control for the metadata survey.

Default Settings for the QDS



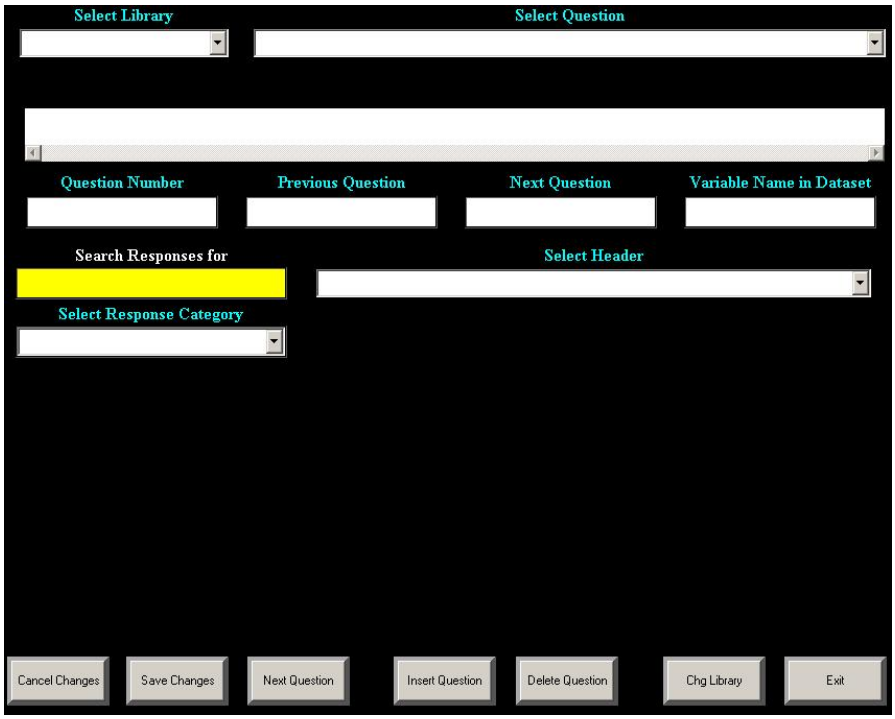
Order	Response category	Previous	Next	Dataset variable	Exclusive answer	Char or Num	# of header
1		strt	A	samplestrt	.		.
2		strt	B	sampleA	.		.
3		A	1	sampleB	.		.
4		B	2	sample1	.		.
5		1	3	sample2	.		.
6		2	4	sample3	.		.
7		3	X1	sample4	.		.
8		4	5	sampleX1	.		.
9		X1	6	sample5	.		.
10		5	7	sample6	.		.
11		6	8	sample7	.		.
12		7	9	sample8	.		.
13		8	10	sample9	.		.
14		9	11	sample10	.		.
15		10	12	sample11	.		.

This version of the QDS cannot be used right away. Response categories have not yet been defined for any of the questions, the variable names probably aren't correct, and the flow-of-control does not account for any skip patterns. The second step of the process of creating question metadata is using the QDS Completion Tool to correct all of these issues.

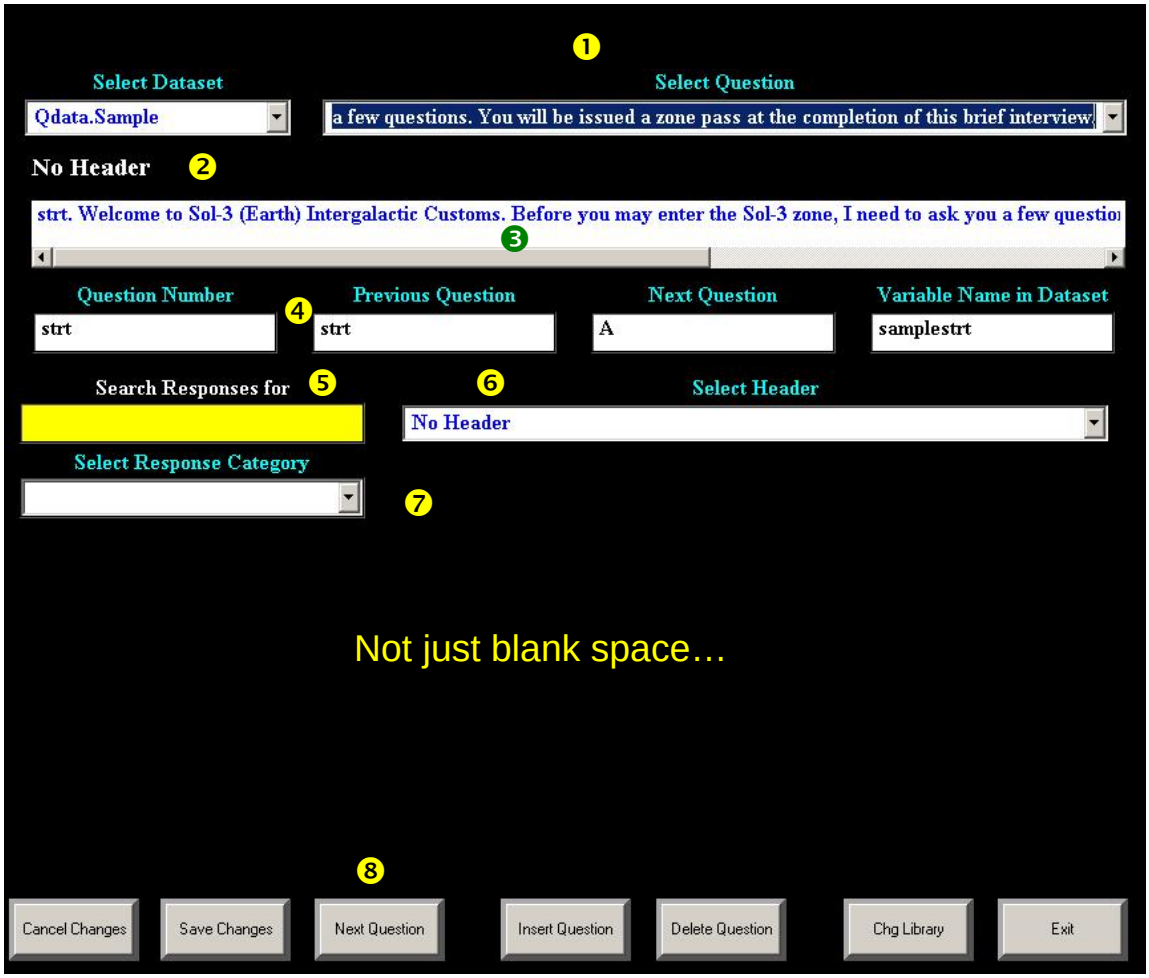
THE QDS COMPLETION TOOL (AKA THE QDS EDITOR)

It is used to edit the question metadata interactively, and is what really allows the metadata system to be utilized by non-technical users. The ("Select Library") pulldown allows you to select the name of the library where the QDS file is located. After you select that, it changes to display a list of all the SAS data sets in that library, so that you can select the QDS file you want to edit.

Starting the QDS Editor



Editing the QDS



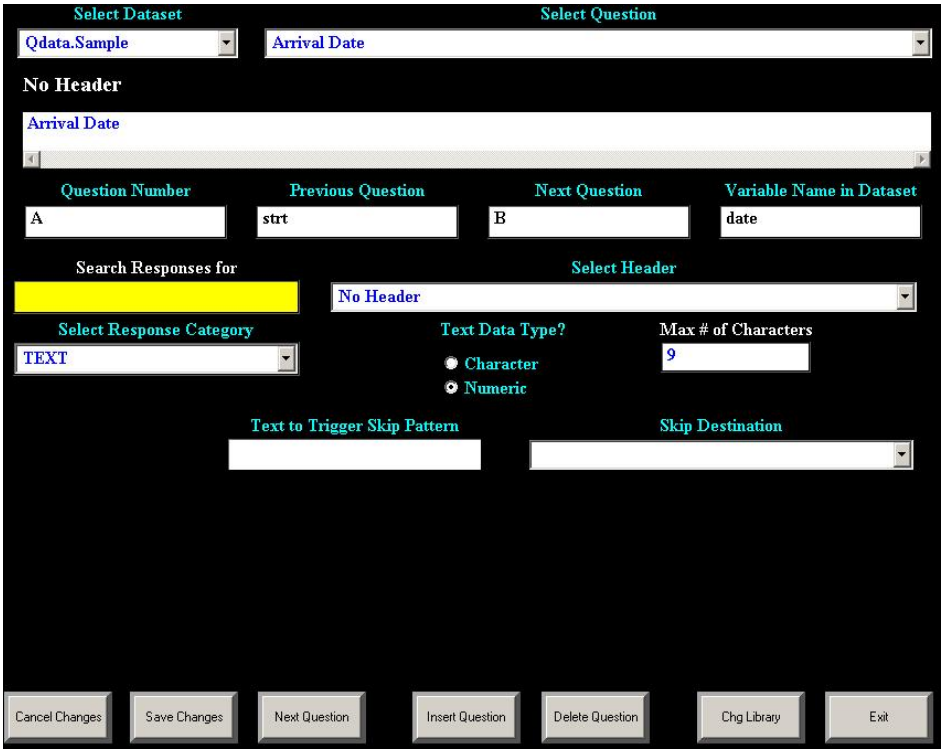
The QDS Editor gives access to all of the fields and features of the metadata system from a point-and-click environment. The questions can be selected from the pull-down menu at the top right (❶). All question records for this data set are in this pull-down, and selecting one will display all the information in that record, although you can navigate through the questionnaire from the beginning to the end via the "Next Question" button (❽) at the bottom of the tool. Any header associated with the question is displayed in (❷), and is associated with the question by selecting a header from (❻). If you need to create a new header, one of the choices in the header pull-down (❻) will open a window to allow you to type in the text of this new header, and it will automatically be added to the header metadata data set.

Item ❸ is where you can edit the text of the question itself. This usually needs to be done to add/remove the question numbers that were added to the question text in the previous step of creating the base QDS from an external file. However, if the text of a question changes, you can make the change here without having to change the external file, re-create the QDS and edit the entire metadata survey from scratch. Basic flow-of-control fields are accessed in ❹. The internal question number, previous and next questions, and the variable name in the data capture table (DCT) associated with the question are all here. A default variable name was created in the previous step by concatenating the name of the QDS file (in this case, "sample") with the internal question number. You can change that, or you can indicate that the text is for display only and does not require a response by removing any value in this field. For example, the first record in the question data set is a standardized script to be read. If we remove "samplestrt" from the "Variable Name in Dataset" field, the metadata system will display the text without asking for any response.

A pull-down list (❼) gives you the list of all the response categories available from the response metadata, as well as those special response categories built into the metadata system. To help you find the category that you are looking for, a search tool is available (❺). This will search the response metadata for the text in the search box, and will display all categories with that text. Where does it do this? In the area marked, "Not just blank space…".

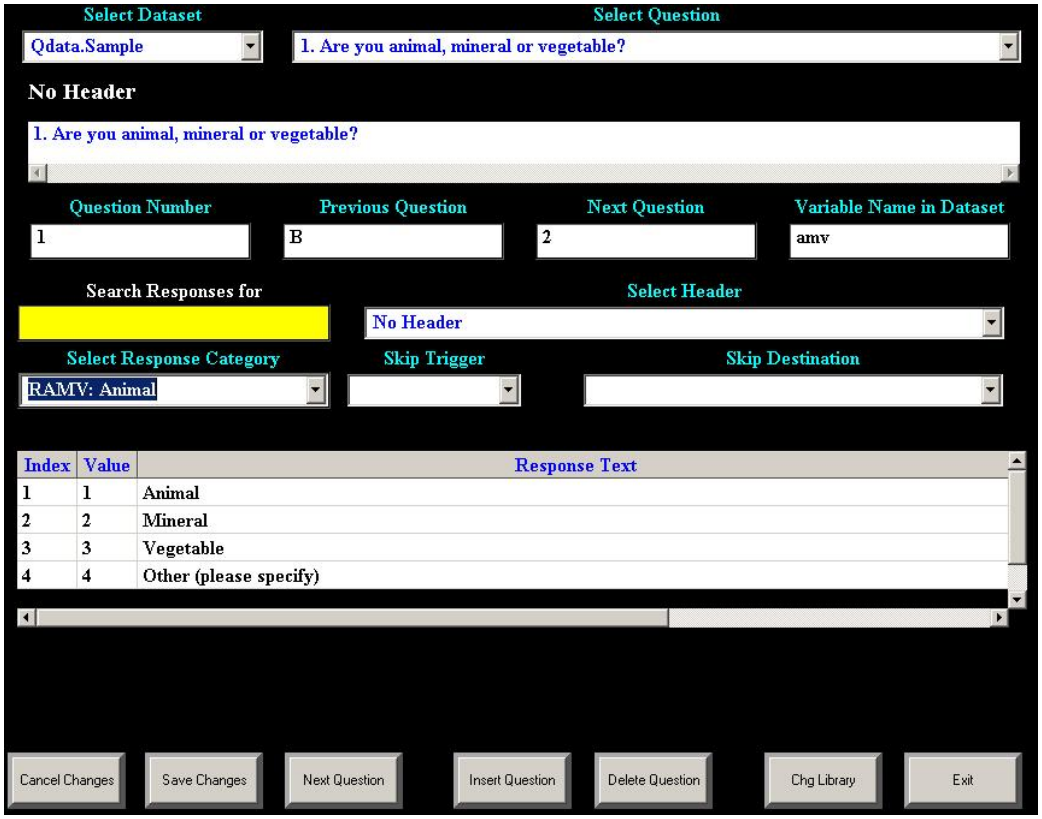
Let's show how the some of the various types of questions are handled by the editor. Starting at the top of the survey, and working our way down, we'll do the "Arrival Date" question first, which is a text response.

Text Response Question



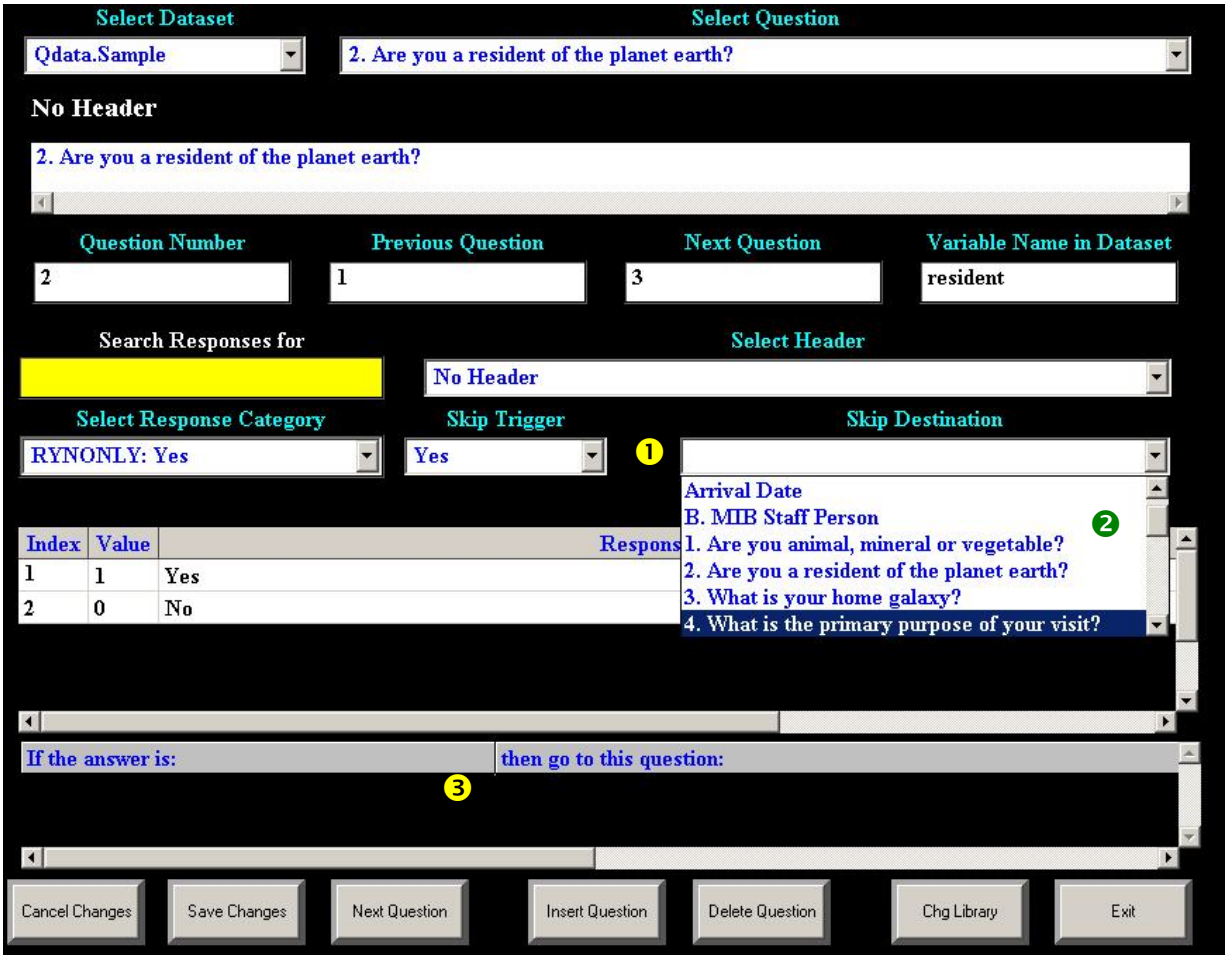
Now that the editor knows what type of response is expected, it will ask the questions that it needs to in order to complete the metadata for this record. When you select "TEXT", the metadata system needs to know whether the text is going to be stored as character or numeric, and the maximum number of characters. Any formats or informats are obtained from the DCT at run-time. Because it's numeric text, it also asks for text that will trigger a skip pattern. If it's not needed, just leave these blank. You can click on "Save Changes" or "Next Question" to commit the changes to the QDS. Next, we'll define a single-response question using one of the response categories we created.

Single Response Question Without Skip Pattern



As you can see, we've changed the variable name and selected the RAMV category from the pulldown. The QDS editor responds by displaying all of the levels for the category and has opened up boxes for skip logic. Since this question has no skip logic, we leave those blank, and move to the next question.

Defining Skip Logic

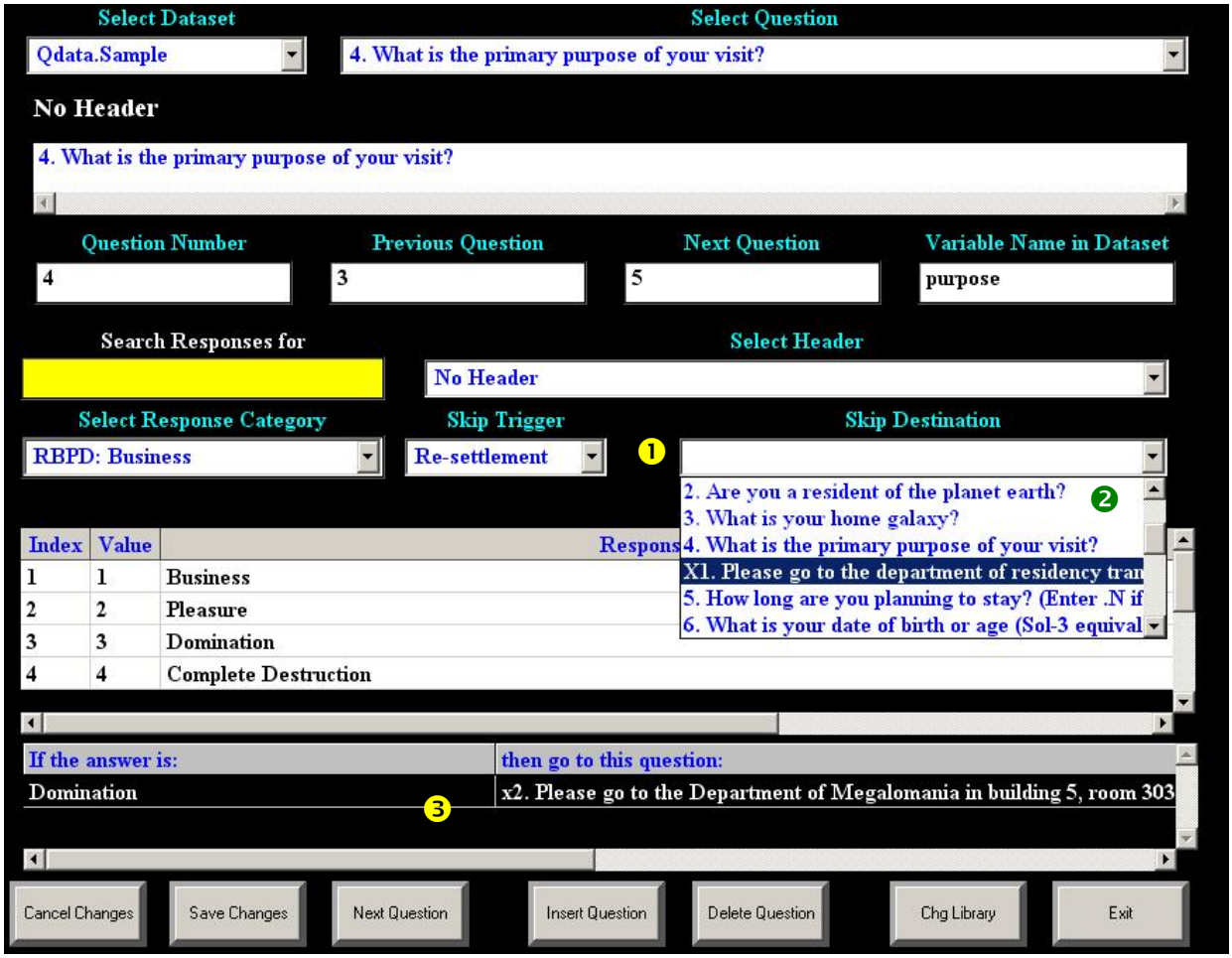


This question involves skip logic. The QDS editor allows you to select the trigger based on the response text (❶), and to define the skip destination based on the question text (❷). The result(s) of your selection(s) are then displayed in ❸, which is also interactive. Double-clicking on a record in ❸ will erase it from the skip logic metadata, allowing you to redefine or correct the skip logic for this question. There is no limit on the number of skip triggers or destinations for a given question. You may even define the same destination for multiple triggers if that is what the survey calls for. Let's look at question 4 from the questionnaire, which has three destinations based on skip logic, as well as a default flow of control.

4. What is the primary purpose of your visit?

- ☐ Business
- ☐ Pleasure
- ☐ Domination ***Please go to the Dept. of Megalomania in bldg. 5, rm 30300903A.***
- ☐ Complete Destruction ***PRESS PRE-EMPTIVE ANNIHILATION BUTTON IMMEDIATELY!***
- ☐ Re-settlement ***Please go to the department of residency transfer on level 6.***
- ☐ Returning Native

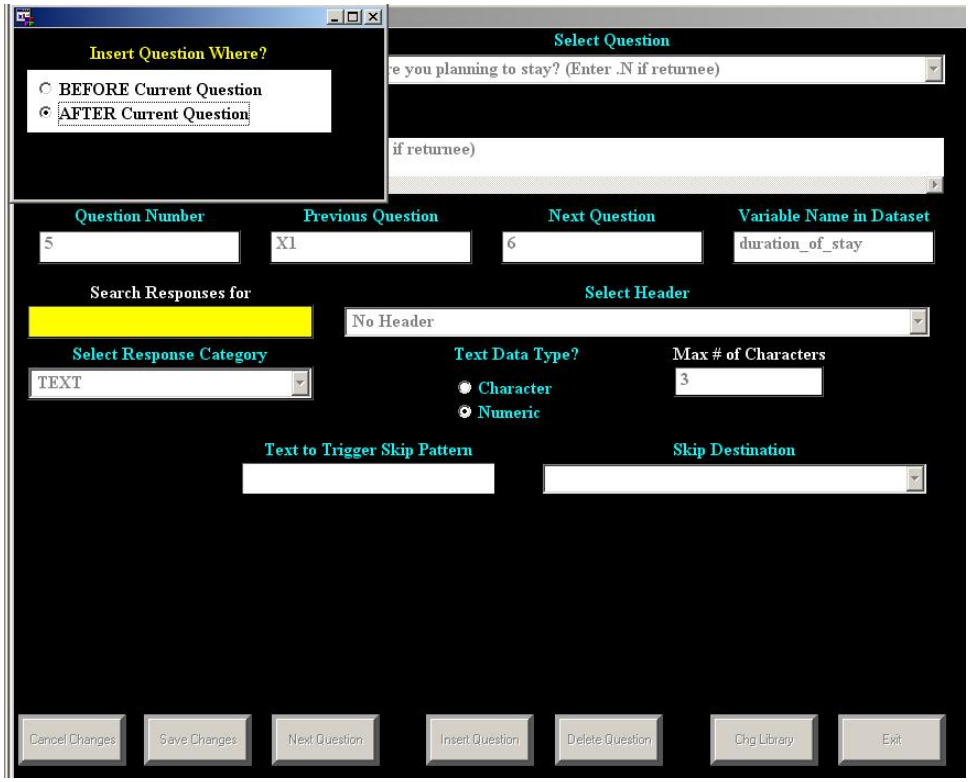
Defining Multiple Skip Destinations Based on Different Answers to the Same Question



Everything looks pretty much the same as the previous example, except that ❸ now has something in it. If the answer is "Domination", then the metadata system should go to the question with the text "Please go to the Department…". The tool communicates with users via the text of questions and answers, so it doesn't matter what the internal question number (which is what is stored in the metadata) is. Of course, if you change the internal question number, then your skip logic will be at the least, incorrect, or at the worst, it will cause the metadata system to fail. In the above example, we've already added the skip pattern for the "domination" answer, and are about to add one for the "Re-settlement" answer, which will show up in ❸ when we select the skip destination. If you make a mistake, it's easy to remove existing skip logic by double-clicking on the record you want to remove in ❸.

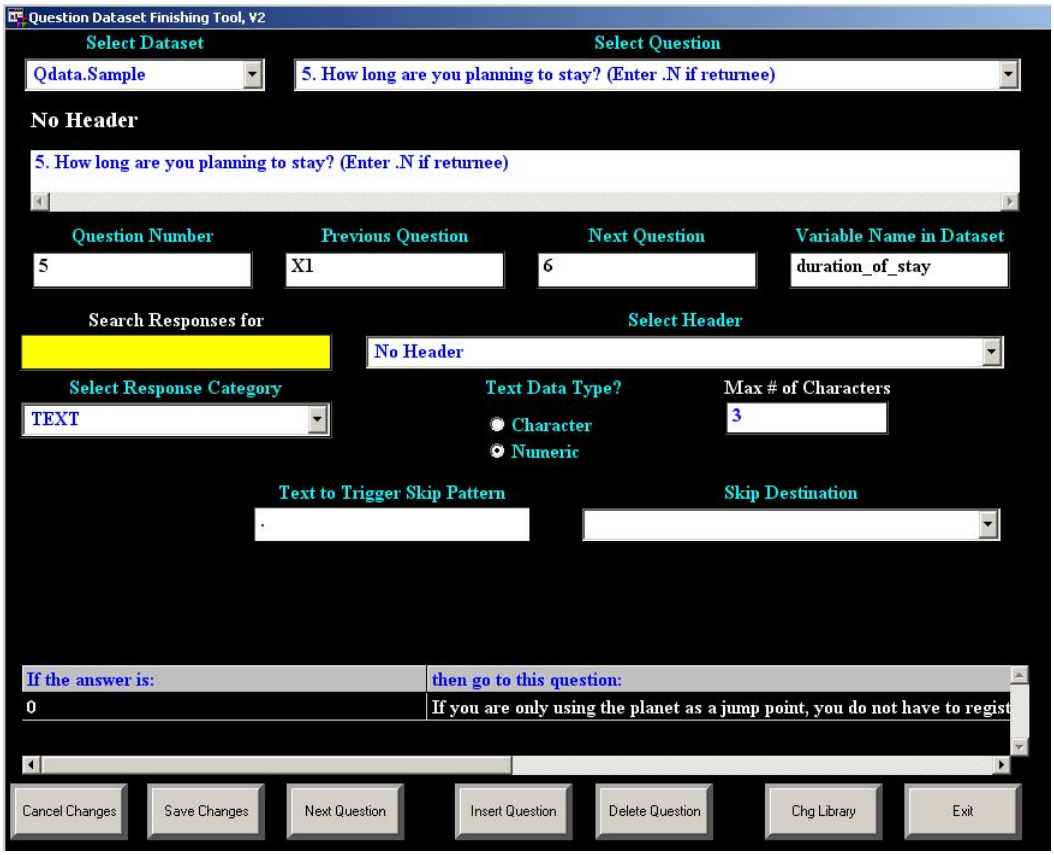
Another feature is the ability to add or remove questions from the metadata. Even though you were careful about setting up the question file that you read in, it appears that you forgot something. Question 6 has a skip pattern and the text of that skip pattern isn't in the question file, unlike all the other skip patterns. You can just add the new question record from the tool.

Adding A Question



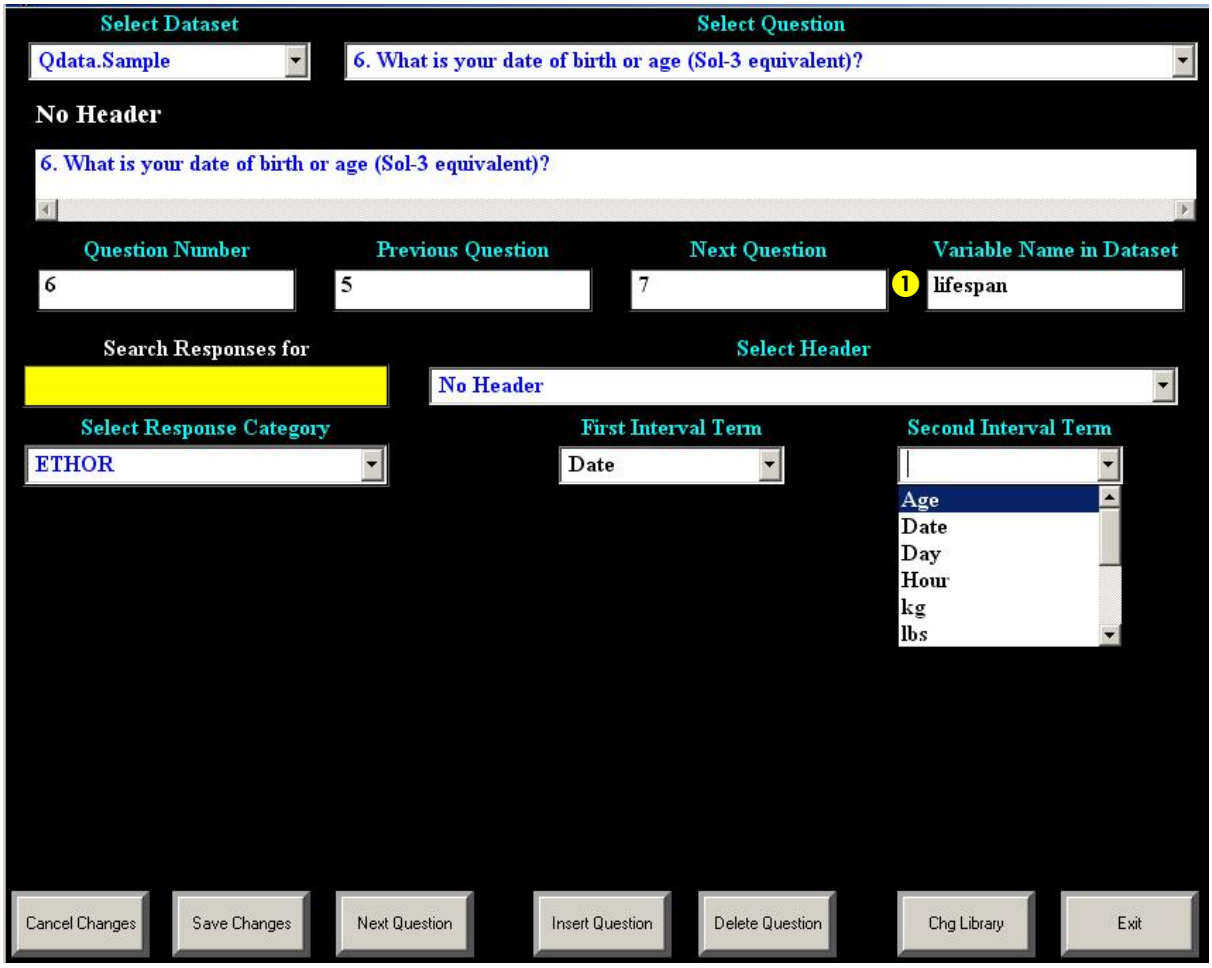
You can add a question anywhere in the file. The metadata system doesn't care where you put the actual record because the order of the records in the QDS does not determine flow of control. However, in practice, most users prefer inserting questions where they belong with respect to the paper form. Now we can go back to question 5 and insert the skip logic.

Text Skip Pattern



The next question in the survey gives two possible units for the response. Here's how the metadata handles it:

Either/Or Questions

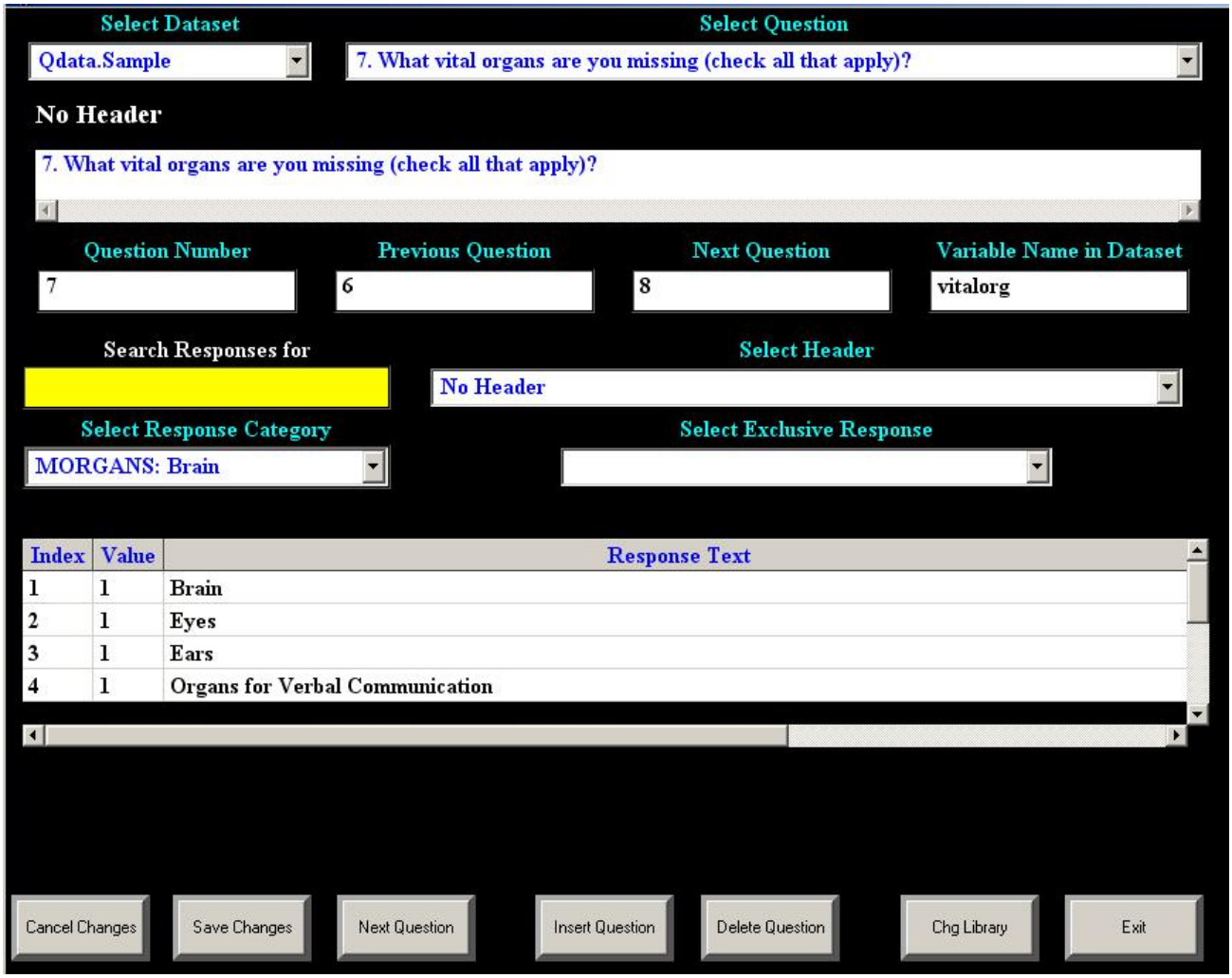


There's a special pre-defined response category called "ETHOR". The variable name you provide the metadata system is used as a root for the actual dataset variables ("lifespan" above (❶). You select the terms from the two pull-down boxes. They are filled with metadata of a list of terms, and there is a separate tool to add terms that aren't already in the term metadata. Since we are dealing with date and age in question 6, the variable names in the dataset are a concatenation of the root variable name you specified, and the term abbreviation. Therefore, the variable names are "lifespandate" and "lifespanage", only one of which will be non-missing for any given record. The term abbreviations are fixed, and can be obtained from the tool that modifies the term metadata.

```
DATA clinic.sample (ENCRYPT=yes READ=&rdpswd WRITE=&wrpswd ALTER=&altpswd)
LENGTH id $ 5 version 4 fc $ 2 visit $ 2 date 4 staffno $ 8 amv 3 amv_othr $ 80 resident 3
       home $ 80 purpose 3 duration_of_stay 8 lifespandate 4 lifespanage 4
```

The last type of question we'll cover here is the multiple-choice question, where you can select any one of a number of answers.

Multiple-Choice Response Question



The variable naming works the same way as the either/or questions, in that the variable name you supply in the QDS Editor is the root. Multiple choice questions are generally scored as yes/no, one variable per possible response, and they are numbered sequentially. Therefore, the response for "Brain" would be stored in the variable vitalorg1; the response for "Eyes" would be stored in vitalorg2, and so on. You also have the ability to define an exclusive response, which is the response that will reset all other choices to unchecked. If question 7 had a response of "None", then that would be an exclusive response, since it's contradictory to the question.

SUMMARY

The metadata system has evolved from a method to perform a single, specific data entry task into a generalized tool for rapid development of data entry forms from paper templates. This logical step in its life cycle required the construction of tools so that non-SAS programmers could use this to create computerized data entry forms, freeing programming staff for other projects. The metadata system and its tools are not perfect, and they don't cover every possible data entry situation (nor can they.) While the system isn't quite to the point where naïve users can just go in and do this, it does make it easy for people who have a small amount of experience with SAS to create fairly sophisticated data entry forms.

The metadata system is still in its relative infancy, and has not yet been used for heavy-duty data entry of multiple forms, although the tools are in place. Future plans call for the creation of a survey question data warehouse, and the development of tools to manage and utilize it, so that the manual step of creating a QDS from a text file will no longer be necessary. It will become a matter of dragging and dropping question records from the warehouse into the QDS being created. While the QDS Editor will still be necessary because skip patterns may change, or answers may use slightly different categories, the warehouse will allow for standardization of some variable names, and the presence of relevant defaults for variable name and response category will speed up QDS development.

Development of standard frames will continue when the situation warrants. However, the next major programming task for the metadata system is the migration from the thick-client methodology of SAS/AF to the thin-client presence on the web. This process has not been scheduled as of right now, but it is most assuredly in the pipeline for the future.

ACKNOWLEDGEMENTS

This work has been partially funded by NIH grants UY1 HL72524-01 and U01 AG023746.

REFERENCES

1. Morgan D. **Building a Data-Driven Computer Assisted Interview Using SAS/AF®**. SAS Institute Inc., Cary, N.C. Proceedings of the Thirtieth Annual SAS® Users Group International Conference**.** 2005 <http://www2.sas.com/proceedings/sugi30/018-30.pdf> (February 17, 2006)

2. Morgan D, Province, M. **Simplifying SAS® Security**. SAS Institute Inc., Cary, N.C . Proceedings of the Twenty-Seventh Annual SAS® Users Group International Conference. April 2002 <http://www2.sas.com/proceedings/sugi27/p103-27.pdf> (February 17, 2006)

CONTACT INFORMATION:

Further inquiries are welcome to:

Derek Morgan
Division of Statistical Genomics
Washington University Medical School
Box 8506, 4444 Forest Park Blvd.
St. Louis, MO 63108
Phone: (314) 362-3685 FAX: (314) 362-4227
E-mail: derek@wustl.edu

This and other SAS System examples and papers can be found on the World Wide Web at:

http://www.biostat.wustl.edu/~derek/sasindex.html