**Populating and Updating the MIDUS Portal**

This document describes how to document MIDUS project datasets in the MIDUS-Colectica online Portal. The Portal is a repository based on the Data Documentation Initiative (DDI) metadata standard which facilitates comprehensive and dynamic online documentation of research datasets. There are actually two versions of the MIDUS Portal:

* Public (<https://midus.colectica.org/>): as the name implies, this is a public-facing portal that contains the same versions of the data and documentation that are publicly available through ICPSR.
* DTD (<https://daytoday.midus.colectica.org/>): the Day-To-Day portal is a private backend version that is automatically updated each week and is only viewable by MIDUS staff. The DTD allows posting the most current data and documentation in a preliminary “sandbox” environment before being eventually made publicly available at ICPSR and the Public Portal. Because MIDUS publishes public data updates relatively infrequently, there usually are some differences between what is available in the Public and DTD Portals.

Both versions of the Portal use the same three basic sources to populate content and functionality:

* SPSS Datasets
* Extra Metadata Spreadsheets
* Concordance Tables

**SPSS Datasets**

The Portal is designed to describe the variables contained in MIDUS datasets so the most important variable-level metadata are derived from a well-documented SPSS dataset. The Portal obtains information about variable names, variable labels, codes, values, value labels, missing value designations, and generates frequency tables and summary statistics from a SPSS dataset. Details about creating a well-documented MIDUS SPSS dataset are available in the [MIDUS Naming and Coding Conventions document](https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/36346/versions/V7/datasets/1/files/1279607/downloadDoc/doc?path=/pcms/studies/0/3/6/3/36346/V7/files/1279607). Contact the Admin Core for the latest version or with any questions about formatting a dataset.

**Extra Metadata Spreadsheets**

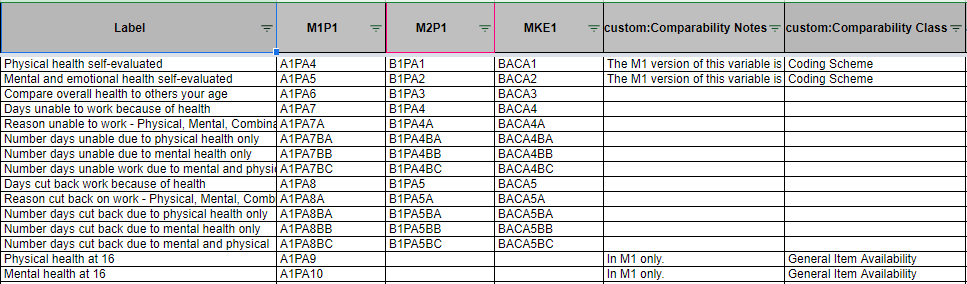
There is other important metadata about each variable that isn’t captured in the SPSS dataset. This extra information describes the questionnaire, instrument, or measurement protocol in more detail and is assembled in an Extra Metadata Spreadsheet. Each individual MIDUS dataset has an Extra Metadata Spreadsheet that includes the following fields (columns):

* Primary key: uniquely identifies each row in the table and maintains sequential order of variables.
* Name: variable name.
* PreQuestionText: any introductory or explanatory text.
* QuestionText: literal question text as asked in the instrument.
* PostQuestionText: any transitional or explanatory text.
* BackwardSkip: describes any skip patterns that filter responses to the current variable.
* ForwardSkip: describes any skip patterns or filters to subsequent variables.
* Notes: any extra information that is helpful in describing or interpreting the variable; this can include references to external documentation or noting that the variable is constructed or derived from other variables.
* InterviewerInstruction: any information about prompts, followups, allowable responses, etc. that is provided to the interviewer or data collector; can also include instructions printed in self-administered instruments.
* Topic: describes the hierarchical measurement organization of the dataset or instrument with different levels separated by colon – Instrument:Section:Topic:SubTopic. See the Project 1 example below.

Each metadata spreadsheet has an Updates tab where any changes should be documented, initialed, and date-stamped to provide an audit trail.

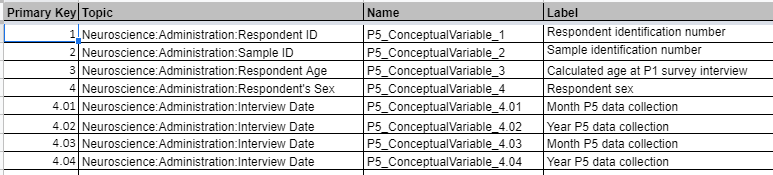
**Concordance Tables**

A primary reason for using longitudinal research designs is the ability to analyze change by observing the same measures over time while holding the conditions of measurement constant. Many longitudinal studies like MIDUS use concordance tables so equivalent measures/variables across waves/datasets are findable, linkable, and interpretable. The MIDUS concordance tables are essentially upgraded versions of the variable cross-walk spreadsheets MIDUS employed at MIDUS 2. These tables link or associate equivalent cross-wave variables to help researchers find the same variable in different datasets. In the concordance tables equivalent variables are listed in the same row while each longitudinal dataset is indicated by the column header. The concordance tables help researchers understand when variables are dropped or added to a dataset, or whether there are any differences in how they were measured over time. In the example below you can see that M1 variable A1PA4 was included in the M2P1 and MKE1 datasets but the Comparability Notes indicate that there were differences in the coding scheme used in the M1 version of this variable. You can also see that variables A1PA9 and A1PA10 were baseline measures only included in M1 (i.e., there are no equivalent variables in the other datasets so those cells in that row are empty).

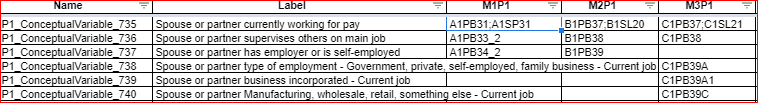


The current MIDUS concordance tables have evolved so that in addition to serving as a variable crosswalk for each MIDUS Project, they drive the interactive functionality of the online Colectica Portal. Because there are both a Public and DTD Portal, most projects have both a Public and DTD version of the concordance tables.

* ALL variables from all datasets within a Project should be represented in the concordance tables, even if there are no other equivalent variables in other waves.
* Both the Public and DTD concordance tables contain a Primary Key field (column). This field is critical for organizing and versioning the tables. Primary Keys are also used in each dataset’s metadata spreadsheets for the same reasons.
  + *New Primary Key values can be added to a concordance table (e.g., when a new measure introduced to an instrument) but Primary Key values are never deleted.* If a project is adding variables to a dataset, they can either:
    - Add new rows *to the end* of the concordance table and populate new primary keys beginning with the last value in the table. If the last primary key in the table is 168, then the new added rows would begin with 169. If a project is adding a new type or new Section of variables to their dataset, this may be the most appropriate route.
    - Insert new rows *within* the concordance table and populate new primary keys by inserting decimals after the previous key value. For example, if Project 5 is adding 4 new “Administration” variables it is probably most clear to insert them in the “Administration” section of the concordance table.

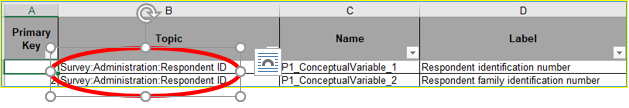


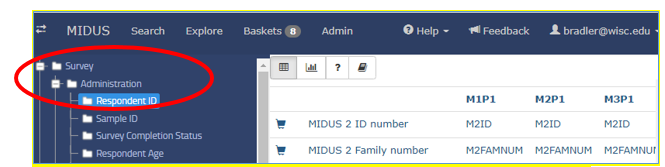
Four new variables inserted after Primary Key value “4” would have values of 4.01 through 4.04 as in the example below. New primary keys need to include leading 0's to maintain variable order sequence and facilitate automatic sorting. Note that the number of leading 0’s is determined by how many new variables are being inserted at any one point (if inserting 100+ variables, the decimal values would start with .001).

* + - Note that leading zeros are needed because spreadsheet software does not necessarily recognize human-readable trailing zeros and therefore cannot distinguish between “4.1” and “4.10.”
  + *Primary Keys cannot be deleted.* If a variable is removed from a dataset in a subsequent update, in the concordance table everything in the row of that variable can be deleted *except the primary key value*.
  + A similar process must be followed to add numbered Conceptual Variable Names ("Name" column), inserting decimals after the previous conceptual variable name number.
  + For situations in which a single Conceptual Variable contains multiple instance variables (called “one-to-many” variable relationships), simply include all of the instance variable names in a single cell separated by a semi-colon (it also helps to include all of the instance variable *labels* in a single cell separated by a semi-colon too). In the example below, two questions asked about R’s spouse or partner working status, one in the CATI and another in the SAQ. Because these two instance variables are conceptually the same, they are both referenced in the same row for the same Conceptual Variable. Each instance variable is only associated with a single primary key and no variables names are repeated.

Each concordance table includes the following fields (columns) that are used by the Portal to organize and display variable-level metadata:

* Primary key: uniquely identifies each row in the table and maintains sequential order of variables.
* Topic: describes the hierarchical measurement organization of the project with different levels separated by colon – **Project:Section:Topic:SubTopic** (SubTopics are not required but can expand a measurement hierarchy if needed). This determines how the Explore view (which uses the Concordance Table) displays the hierarchy. See the following example of the Project 1 Concordance Table and accompanying Portal Explore view:





Note that the Portal uses the information in the Topic column to determine which Project and which Section a conceptual variable belongs to. Regardless of the value of its primary key or its location in the concordance table, any “Survey” variable that includes “Administration” in its Topic column (e.g., “Survey:Administration:RespondentWeight”) will be displayed in the “Administration” Section of the Portal.

* Name: conceptual variable name, unique but non-descriptive. DDI uses conceptual variable names to link or relate all the variables in a single row.
* Label: conceptual variable label, descriptive.
* Variable names (dataset specific): these are the actual variable names from each dataset.
* Comparability Notes: notes that describe any differences among variables or threats to equivalence.
* Comparability Class: categories that broadly describe the type of differences.
* Display Values: determines the values that are displayed in the Portal for that row’s variables - some variables (such as IDs) do not need every value displayed. Each concordance table includes a “DisplayStat Directions” tab which details the different display options available.
* Stats: determines the summary stats that are displayed in the Portal for that row’s variables – some variables do not need means or standard deviations displayed. Each concordance table includes a “DisplayStat Directions” tab which details the different summary statistics available.
* Variable labels: The Portal populates variable label information from the datasets themselves, *not from the concordance table*. The variable labels in the concordance table help humans manage and understand the variables.

Each Concordance Table has an Updates tab where any changes should be documented, initialed, and date-stamped to provide an audit trail.