



Data Dictionary



THE UNIVERSITY OF
CHICAGO MEDICINE &
BIOLOGICAL SCIENCES



CENTER FOR
RESEARCH
INFORMATICS

© 2015 Vanderbilt University

Creating a Data Dictionary in REDCap

You can create your database fields either using the Online Designer or a Microsoft Excel spreadsheet. In either case, what you're creating is referred to as the data dictionary, a description of all the field (variable) names, field types, possible values, etc. We recommend using the Online Designer when you first begin using REDCap to get a feel for how it works, and also anytime you are making minor changes. In time, you might find it is easier to create a data dictionary in Excel, and upload this into REDCap. You don't have to create similar variables individually as in the Online Designer – you can take advantage of using Excel's copy/paste functionality. To learn how to create a data dictionary in Excel, a Data Dictionary demonstration file with coding examples can be downloaded from the Data Dictionary tab of any REDCap project.

Data Dictionary Excel Spreadsheet Columns

This section describes the function of each column in the data dictionary spreadsheet, and whether or not it is required or optional.

Note: The term "Column" here refers to Excel spreadsheet data dictionary columns. Also, the terms "field" and "variable", as used here, are essentially interchangeable. Both terms refer to a unique item of data to be collected and analyzed. "Field" is a database term, while "variable" is a data analysis term.

Column A – Variable/Field Name

Required

1. Variable/Field names specify the variable name that will be used in reporting, data export, and data analysis. They are not displayed on the data entry form.
2. **The first variable on the first form must be the record identifier** (e.g. Participant ID) because it will be used by REDCap as a key variable linking forms for a particular record. The default field name is "Study_ID". Demographics is normally the first form, but this is not required.
3. Field names may contain lowercase letters, numbers, and underscores, but no spaces or special characters.
4. Field names should always start with a letter.
5. Field names must be unique, and cannot be repeated within a database, even in different forms.
6. In general, variable names should be as short in length as possible while maintaining meaning to researchers. Remember they do not need to be too descriptive because the description will be included in the Field Label. A common example is the use of "dob" as a variable name, with the corresponding field label of "Date of Birth".
7. The recommended character length for a field name is 26 and the maximum is 100.
8. When using the Online Designer, REDCap automatically assigns a default field name based on the field label. It is recommended that you double check the default field name and modify if necessary.
9. **Remember that if you change a variable name in one place, you must change it everywhere it is used, e.g. calculations, branching logic, automated survey logic.**
10. **Remember that if you change a variable name, you will lose all data entered under the old variable name.**

Column B – Form Name

Required

1. Form Names correspond to individual data entry forms in the project.
2. Form names must be lowercase letters, numbers and underscores in the Excel spreadsheet, but will be displayed in REDCap with initial capitals. If your form name contains more than one word, connect the words with an underscore, such as "form_name". The underscore will appear as a space in REDCap.

3. All variables in a form must be in adjacent rows in the data dictionary. For example, you cannot have a variable in row 6 be in the “demographics_form”, a variable in row 7 be in the “first_visit” form, and then a variable in row 8 back in the “demographics_form”.
4. Forms will appear in the data collection section of REDCap in the order they appear in the data dictionary.
5. Variables within a form will appear in the order they appear in the data dictionary.

Column C – Section Header

Optional

1. Use a Section Header to break up the data entry form into sections. When rendered, any text included in the header information will show up as a preceding line of text – with a yellow background.
2. Section Headers are used to visually separate items within a form, primarily to aid data entry. If you are entering data directly into REDCap while interviewing a study participant, you may also want to use Section Headers to display interview script between questions, e.g. to introduce a new topic.

Column D – Field Type

Required

1. Specifying the field type determines what types of responses are allowed, and how the data entry area for that field will be displayed.
2. Field types include
 - a. *text box* – for capturing free text. The *text* field type should have validation (Column H) whenever possible. If the validation is “integer” or “numeric”, you should also include the allowable minimum and maximum values (Columns I & J) when possible. The text field type can also have suggested choices listed in Column G.
 - b. *notes box* – for capturing free text. Although there is no limit to the number of characters you can store in the either *text* or *notes* field types, there is a limit to the number of characters displayed on the data entry form for each field type. Text boxes display about 30 characters and notes boxes display about 300 characters.
 - c. *calculated field* - perform real-time calculations. Calculated variables display the result of a calculation based on responses to other variables. Data cannot be entered in calculated fields. Any variable in the database can be used in the calculated field regardless of data entry form. For longitudinal projects, you may need to specify the event name as well.
 - d. *drop-down list* - dropdown list of choices. Use if there are a large number of choices and only one choice is allowed. Must also have response options (choices) defined in Column F.
 - e. *radio buttons* - radio buttons for each choice. Use if there are a small number of choices and only one choice is allowed. Must also have response options (choices) defined in Column F.
 - f. *checkboxes* – multi-selectable checkboxes. Use if multiple choices are allowed. Must also have response options (choices) defined in Column F.
 - g. *yes-no* – radio buttons with yes and no options coded as No = 0 and Yes = 1.
 - h. *true-false* - radio buttons with true and false options coded as False = 0 and True = 1.
 - i. *slider* – visual analogue scale. Slider labels can be specified in Column F and slider numbers can be turned on in Column H.
 - j. *file upload* - upload a document. This field type allows you to attach a document (e.g. signed consent form) to the record. Note, that the file repository of REDCap is used to upload a project document, not a record/subject document to the project (e.g. blank consent form or IRB protocol).
 - k. *descriptive text* - text displayed with no data entry. Optional image/file attachment must be specified through the Online Designer.

- l. *begin new section* – section headers appear in yellow and can include optional descriptive text.

Column E – Field Label

Required

A Field Label (or variable label) is a word or phrase which should be more descriptive than the variable/field name. This is the text that will be on the form (instead of the variable name). For example the field name might be “dr_address” while the field label would be “Please enter the primary care physician’s address.”

Column F – Choices, Calculations, or Slider Labels

Required if field type is dropdown, radio, or checkbox, or calculation

1. All categorical field types (*dropdown, radio, checkbox*) must specify response options associating numerical values with labels. All calculated fields must specify the equation to be calculated here.
2. Choices using *dropdown, radio, or checkbox* field types are stored as numerical values, but choices can be displayed as text in the data entry form. For *dropdown, radio, checkbox* field types, this column is used to specify both what is displayed on the data entry form and what is stored in the database. For example, if “0, No | 1, Yes” is entered in Column F, “No” and “Yes” will displayed on the data entry form and “0” and “1” will be stored in the database. The | is critical to separate options in the list. Numerical codes do not need to be consecutive and do not affect the order the text is displayed.
3. All calculated fields must specify the calculation here. Examples of calculation syntax can be found at the Help & FAQ link at the bottom of the left hand panel of any REDCap project screen.
4. Optional slider field labels can be specified here. The slider field allows you to label three anchor points: left, middle, and right. An example might be “Strongly Disagree”, “Neutral”, and “Strongly Agree.” These label would be defined as follows in Column F: Strongly Disagree | Neutral | Strongly Agree

Column G – Field Note

Optional

1. Field notes are used to provide information to assist in data entry. They are mainly used to supplement the Field Label information. Examples are specifying the expected format of a validated field (e.g. phone number), or units (e.g. kg vs. lb).
2. The Field Note is rendered just below the field input element and shows up in small blue text.

Column H – Text Validation Type or Show Slider Number

Optional

1. Specifies data validation type for *text box* field type.
2. An error message is displayed if an entry does not match the expected format, and the entry is not saved to the database. In other words, the validation type is a hard validation.
3. Data validation types for *text box* fields are email, integer, letters only, MRN (10 digits), number, phone, ssn, and zip code. In addition, there are several date and time validation formats.

DATE_DMY	Example: 16-02-2011
DATE_MDY	Example: 02-16-2011
DATE_MY	Example: 02-2011
DATE_YMD	Example: 2011-02-16
DATETIME_DMY	Example: 16-02-2011 17:45
DATETIME_MDY	Example: 02-16-2011 17:45
DATETIME_YMD	Example: 2011-02-16 17:45

DATETIME_SECONDS_DMY	Example: 16-02-2011 17:45:23
DATETIME_SECONDS_MDY	Example: 02-16-2011 17:45:23
DATETIME_SECONDS_YMD	Example: 2011-02-16 17:45:23
TIME_HH:MM	Example: 17:45
TIME_MM:SS	Example: 45:23

4. For *slider* field type, specifies whether to display or hide the value (1-100) selected on the slider. An entry of 'number' in this column will display the value.

Columns I & J – Text Validation Min/Max

Optional

1. For text validation types of number, integer, or date, minimum and maximum acceptable values may be specified. You may use a min, max, neither or both for each individual field.
2. An error message, including the acceptable range, is displayed if the entry is out of range. However, the entry is stored to the database regardless of whether or not the entry is out of range. In other words, the Min/Max validation is a soft validation.

Column K – Identifier?

Optional

1. If any of the fields are a Subject Identifier, then place 'y' in the cell. If not, simply leave the cell blank.
2. Tagging fields as Subject Identifiers will prevent lower-level users from exporting data for these fields from the database, depending on user rights.
3. There are 18 pieces of information that should be marked as identifiers in a REDCap data dictionary:
 - a. Name
 - b. Fax number
 - c. Phone number
 - d. E-mail address
 - e. Account numbers
 - f. Social Security number
 - g. Medical Record number
 - h. Health Plan number
 - i. Certificate/license numbers
 - j. URL
 - k. IP address
 - l. Vehicle identifiers
 - m. Device ID
 - n. Biometric ID
 - o. Full face/identifying photo
 - p. Other unique identifying number, characteristic, or code
 - q. Postal address (geographic subdivisions smaller than state)
 - r. Date precision beyond year

Column L – Branching Logic

Optional

Branching logic can be applied to a field to specify whether or not it will be displayed, depending on values in other fields. For example, a question about pregnancy can be designated to be displayed only if the subject is female. Syntax for branching logic can be found at the Help & FAQ link at the bottom of the left hand panel of any REDCap project screen.

Column M – Required Field?

Optional

1. A field can be designated as “required” so that it must be completed before moving on to the next field. An error message is displayed if the field is left blank. However, the user can choose to ignore the error message and continue with data entry. In other words, this is a soft requirement.
2. If any of the fields require a value, then place ‘y’ in the cell. If not, simply leave the cell blank.

Column N – Custom Alignment

Optional

1. The location of text boxes or categorical responses (dropdown, radio, checkbox) can be specified as Right/Vertical, Left/Vertical, Right/Horizontal, Left/Horizontal.
2. The default setting, if not specified, is Right/Vertical. Right/Horizontal is specified as RH, Left/Vertical is specified as LV, and Left/Horizontal is specified as LH.

Column O – Question Number

Optional for surveys

REDCap can be set to auto-number questions on a survey. However, if you want a custom numbering scheme, you can specify each question number here. Any text in this column will be displayed as the question number.

Column P – Matrix Group Name

Required for matrix type fields

1. In order for different fields to be displayed in the same matrix/grid together, they must be given the same matrix group name. Otherwise they will get rendered as separate matrix groups.
2. Only letters, numbers, and underscores can be used in the name.

Column Q – Matrix Ranking

Optional for matrix type fields

A ranked matrix of fields is a matrix that can be used to 'rank' the choices within it so that no two fields in the matrix can have the same selected value for a given column. For example, if ranking is enabled for a matrix and a user then selects a choice for one field that has already been selected for another field, it will automatically deselect the value for that other field and display a quick message saying 'value removed!' to denote that the value was removed for the other field. This prevents any two fields from having the same selected value.

Column R – Field Annotation

Optional

Field annotations can be used to add explanatory notes or commentary about a given field. This can be used for several purposes, such as for the bookkeeping of a project's field structure (as metadata about the given field) for reference purposes regarding what the field represents or how it should be used (during data entry, analysis, etc.). Field annotations are not displayed on any page but are merely for reference. Summarily, field annotations do not have a specifically defined purpose, so users may use them in whatever way they so choose.