

How to Use a REDCap Data Dictionary

The Data Dictionary is a CSV file that holds the instruments of a project. In addition to building an instrument using Online Designer, instruments can be built or edited in a Data Dictionary, then uploaded into REDCap. Once a user is familiar with building projects in Online Designer and comfortable with Data Dictionary syntax, this can be an efficient way to build larger projects or make large scale changes. The Data Dictionary will not contain any information about survey settings, events in a longitudinal study, or other project-level settings.

To download the Data Dictionary, navigate to the “Data Dictionary” page for a new project and select “Download the current Data Dictionary.” This will download a CSV file in Excel (or a spreadsheet software of your choice) where the necessary information to build the project can be entered, saved and then uploaded into REDCap on the Data Dictionary page. In the worksheet, the **rows** will contain the fields, or variables; the **columns** will contain the required or optional information about each field.

Note: If your project is in production and in draft mode, you will see two download options: the current data dictionary being used, and the data dictionary with the draft changes. This can be an effective way to compare project changes.

Some things to remember when building/editing a data dictionary:

- The following columns **must** be completed for each field/variable: (A) Variable/Field Name, (B) Form Name, (D) Field Type, (E) Field Label.
- The Data Dictionary must be saved as a .csv file.
- The first variable on the first form should always be a record identifier. The REDCap system defaults to “record_id.”
- If using piping, branching logic, or calculations, the variable/field name utilized in that text must be in square brackets, e.g. [variablename].
- If using field embedding, the variable/field name utilized must be in curly brackets, e.g. {variablename}.

Column A – Variable/Field Name *(Required)*

The Variable/Field Name column specifies the variable names that will appear when data is reported, exported, and analyzed. It will also appear in the Codebook.

- Variable/field names cannot contain spaces or special characters – only letters, numbers, or underscores.
- Variables/field names cannot start with a number.
- Keep variable/field names brief and intuitive. For example, for a question asking for birthdate, a suggested variable/field name is “dob” or “birthdate.” “variable3” is not particularly descriptive, and “participants_date_of_birth” is unnecessarily long.
- **Variable/field names must be unique and cannot be repeated within a project.**

Column B – Form Name *(Required)*

Forms are a way to group variables within a survey or database.

- In the Data Dictionary, form names must be all lowercase and have no spaces. REDCap will display the names with initial capitals and will turn underscores into spaces after the Data Dictionary is uploaded.
- All variables that will be on the same form must be in consecutive rows.

Column C – Section Header

Section Headers are a way to separate questions on a form to either aid in data entry or provide instructions to survey takers. If you have multiple pages enabled as a survey setting, then your headers will be where the page breaks.

Column D – Field Type (*Required*)

Field type tells REDCap how to structure the answer format.

- Options include:
 - text – provides a box to enter text appropriate for short text, numbers, dates/times, etc.; there is no default text limit, but it is difficult to read more than a few words in short text boxes.
 - notes – provides a box to enter text appropriate for longer text, such as paragraphs or narratives; there is no default text limit
 - radio – provides radio buttons for multiple choice options
 - dropdown – provides a dropdown menu for multiple choice options
 - checkboxes – provides checkboxes for multiple choice options where more than one answer can be chosen
 - calc – calculates an equation by gathering information from other completed fields/variables
 - file – allows a participant or data entry personnel to upload a file; when paired with the “signature” validation, it provides a signature field
 - yesno – provides default yes/no answers (yes=1, no=0) as radio buttons; this is similar to the previous multiple choice questions, but is pre-coded
 - truefalse – provides default true/false answers (true=1, false=0) as radio buttons; this is similar to the previous multiple choice questions, but is pre-coded
 - descriptive – provides the user an opportunity to provide descriptive text, such as instructions, a vignette, etc., or display an image, file attachment, and other media; these fields are not included in data exports or reports
 - slider – provides a scale where participants or data entry personnel can select an answer by using a slider to set the response
- Double-check that there are no unnecessary spaces in this column, or REDCap will read the field type as invalid.

Column E – Field Label (*Required*)

The Field Label is where question text or directions are entered.

Column F – Choices, Calculations, OR Slider Labels (*Required for Radio, Dropdown, Checkboxes, Calculations, or Slider Fields*)

- Choices – Categorical or checkbox variables must specify response options that associate numerical values with textual labels. Syntax should be formatted as “1, Answer 1 | 2, Answer 2 | 3, Answer 3 | ... N, Answer N.”
 - If you use you the “yesno” or “truefalse” field types, you do not need to enter response options.
- Calculations – Any calculated fields must specify the calculation here.
- Slider Labels – Slider fields should have three anchor points: the two end points and the middle. These can be numerical (e.g. “1 | 50 | 100”) or textual (e.g. “Strongly Disagree | Neutral | Strongly Agree”). When exporting the data, the numerical value will be extracted.

Column G – Field Notes

Fields Notes are a place to provide extra information for survey-takers or data entry personnel. This can be a note about a proper format, an example, or anything else that may be considered helpful.

Column H – Text Validation Type OR Show Slider Number

- For text fields, it is recommended to use text validation when it’s appropriate. Options for text validation are:
 - date_dmy, date_mdy, date_ymd, datetime_dmy, datetime_mdy, datetime_ymd, datetime_second_dmy, datetime_seconds_mdy, datetime_seconds_ymd, email, integer, number, phone, time, and zipcode.
 - The “integer” validation only allows whole numbers. The “number” validation allows numbers with decimals.
- For slider fields, this column gives the user the option to display the value selected on the slider by entering “number” in this column.
- If you would like to use the signature feature, set the “Field Type” to “file” and enter “signature” in this column.

Column I – Text Validation Min

If the Text Validation Type is set to a number, integer, or date range, setting the minimum aids in valid data entry. If a number below the accepted range is entered, an error message is displayed. When collecting data, this validation can be overridden by the user or survey participant.

Column J – Text Validation Max

If the Text Validation Type is set to a number, integer, or date range setting the maximum aids in valid data entry. If a number above the accepted range is entered, an error message is displayed. When collecting data, this validation can be overridden by the user or survey participant.

Column K – Identifier?

In this column, anything that is an identifier should be marked with a “y.” To ensure compliance with HIPAA, there are 18 HIPAA identifiers that should be marked. These are:

- Name
 - If you are using a “Signature” field for participants to sign their name, that should also be marked as an identifier.
- All geographical identifiers smaller than a state

- Dates (other than year) directly related to an individual
- Phone numbers
- Fax numbers
- Email addresses
- Social Security Numbers (SSNs)
- Medical record numbers (MRNs)
- Health insurance beneficiary numbers
- Account numbers
- Certificate/license numbers
- Vehicle identifiers and serial numbers, including license plate numbers
- Device identifiers and serial numbers
- Web Uniform Resource Locations (URLs)
- Internet Protocol (IP) address numbers
- Biometric identifiers, including finger, retinal, and voice prints
- Full face photographic images and any comparable images
- Any other unique identifying number, characteristic, or code except the unique code assigned by the investigator to code the data

Column L – Branching Logic (Show field only if...)

Branching logic is a way to control what fields appear to which participants or data entry personnel, depending on answers to previous questions.

- For radio button questions, branching logic is coded as: [variable] = 'code#', where '[variable]' is the name of the field that must be answered first and 'code#' is the numerical code of the answer choice that must be selected in order for the field to appear.
- For checkbox questions, branching logic is coded as: [variable(code#)] = '1' for checked or '0' for unchecked.

Column M – Required Field?

Fields can be designated as required if this column is marked with a "y." These fields must be completed before moving to the next form, and an error message is displayed if the field is left blank. This is strictly enforced on surveys (meaning a participant cannot submit the survey or turn the page), however this can be overridden by a user in data entry mode.

Column N – Custom Alignment

- REDCap defaults to aligning most text boxes or response options to Right/Vertical. This can be changed to Left/Vertical (LV), Right Horizontal (RH), or Left Horizontal (LH).
- Slider fields default to Right/Horizontal.

Column O – Question Number (surveys only)

REDCap can auto-number questions on a survey, but a custom numbering scheme can also be specified in the Data Dictionary. If any branching logic is used, custom-numbering is required. Be sure to check that you did not number the questions in the field label if you turn on the Question Number survey setting.

Column P – Matrix Group Name

When multiple questions will use the same response options, such as several questions that utilize the same Likert scale, they can be organized into a matrix. To do this, enter the desired matrix name in this column for all questions that will be included in the matrix.

- Matrix names cannot contain spaces or special characters – only letters, numbers, or underscores.
- Matrix names must be unique and cannot be repeated within a project.
- All variables that will be in the same matrix must be in consecutive rows.
- Only radio buttons or checkboxes can be grouped into a matrix.

Column Q – Matrix Ranking?

To have a ranked matrix – or a matrix where choices are “ranked” so that no two fields in the matrix can have the same selected value – mark this column with a “y” for all fields included in the matrix.

Column R – Field Annotation

The Field Annotation column can be used for Field Annotations or Action Tags.

- Field Annotations are additional explanatory notes or commentary for a field. These do not appear on the form or survey.
- Action Tags are a way to customize data entry for individual fields in a survey or form to make sure the data is exactly what you need. An action tag is a command, starting with an “@,” that is entered into this column. To prevent Excel from attempting to do a calculation, enter a space before the @ symbol. Common Action Tags include:
 - @CHARLIMIT or @WORDLIMIT – sets the maximum number of characters or words that can be entered into a text or notes field; will also display the number of characters or words remaining
 - @MAXCHECKED – allows a checkbox field to have a maximum number of checkboxes selected (e.g., @MAXCHECKED = 3 means a maximum of 3 answer options can be checked)
 - @NONEOFTHEABOVE – allows for the designation of a checkbox choice to be a ‘none of the above’ option, thus ensuring no other choices can be selected if this choice is selected