

Creating a Data Dictionary

You can create your database fields either using the Online Form Editor or in an 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 Form Editor when you first begin using REDCap to get a feel for how it works, and also anytime you are making minor changes. In general, it is easier to create a data dictionary in Excel, and upload this into REDCap – since you don't have to create similar variables individually, but can take advantage of using copy/paste functionality. To assist you in learning how to create a data dictionary in Excel, we have provided a Data Dictionary demonstration file, which can be found in, and downloaded from, the Online Form Editor page. This file shows entries for each required column, as well examples of calculated fields and branching logic.

	A	B	C	D	E	F
	Variable / Field Name	Form Name	Section Header	Field Type	Field Label	Choices, Calculations, OR Slider Labels
2	study_id	demographics		text	ID	
3	enroll	demographics		text	Date subject signed consent	
4	fname	demographics		text	First Name	
5	lname	demographics		text	Last Name	
6	city	demographics		text	City	
7	state	demographics		text	State	
8	zip	demographics		text	Zipcode	
9	sex	demographics		dropdown	Gender	0, Female 1, Male
10	given_birth	demographics		radio	Has the subject given birth before?	0, No 1, Yes
11	num_children	demographics		text	How many times has the subject given birth?	
12	race	demographics		checkbox	Race	0, Caucasian 1, African American 2, Hispanic 3, Asian 12, Other
13	dob	demographics		text	Date of birth	
14	age	demographics		calc	Age	round(datediff([enroll],[dob],"y"),1)
15	height	demographics		text	Height (cm)	
16	weight	demographics		text	Weight (kilograms)	
17	bmi	demographics		calc	BMI	round([weight]*10000/([height]*[height]),1)
18	pcp	demographics		dropdown	Does patient have a primary care physician?	1, Yes 2, No
19	upload	demographics		file	Upload record documents	
20	notes	demographics		notes	Notes about record	
21	date_visit_b	baseline_data	Baseline Measurements	text	Date of baseline visit	
22	date_blood_b	baseline_data		text	Date blood was drawn	
23	alb_b	baseline_data		text	Serum Albumin (g/dL)	
24	prealb_b	baseline_data		text	Serum Prealbumin (mg/dL)	
25	creat_b	baseline_data		text	Creatinine (mg/dL)	
26	npcr_b	baseline_data		text	Normalized Protein Catabolic Rate (g/kg/d)	
27	chol_b	baseline_data		text	Cholesterol (mg/dL)	

Note: Although you will be working in Excel, the file type you work with is actually a ".csv" (comma separated variables) file. When saving your data dictionary, be sure to select the .csv format to upload to REDCap.

General Best Practices

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.

- a. The first variable on the first form should 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 variable name is “Study_ID”. Demographics is normally the first form, but this is not required. All new projects are provided with a sample Demographics form, but you are free to modify/replace this.
- b. Use categorical response (dropdown, radio button, checkbox) field types when possible to reduce risk of data entry error. If these fields are not feasible, use text fields with validation (date, phone, email, integer, number) whenever possible to reduce the use of free-text fields.
- c. When using a text field with validation types of number or integer, define range minimum/maximum as much as possible to allow REDCap to perform basic data validation/quality control.
- d. Put variables collected together on the same form to improve data entry workflow. Putting demographics together and labs together on separate forms makes data entry more reliable.
- e. Include Field Notes describing units, formats, etc. whenever appropriate. Do not assume the data entry person knows the expected units or formats.

Data Dictionary Spreadsheet Columns

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

1. Column A - Variable/Field Name

Required

- a. 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.
- b. Variable names may contain letters, numbers, and underscores, but no spaces or special characters.
- c. Variable names cannot start with a number.
- d. Variable names must be unique, and cannot be repeated within a database, even in different forms.
- e. Variable names should be brief; they do not need to be descriptive as the description will be included in the Variable Label. A common example is the use of “dob” as a variable name, with the corresponding variable label of “Date of birth.”
- f. Remember that if you change a variable name in one place, you must change it everywhere it is used, e.g. calculations, branching logic.

2. Column B - Form Name

Required

- a. Forms are groupings of variables within the database. It’s a good idea to divide your variables into several fairly short forms for ease of data entry, and more opportunities to save data at the end of each form.
- b. Form names must be all lowercase 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.
- c. 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”. Variables will appear on the form in the order they appear in the data dictionary.

3. Column C – Section Header

Optional

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.

4. Column D - Field Type

Required

- a. Specifying the field type determines what types of responses are allowed, and how they will be displayed. Field types include: dropdown, radio button, checkboxes, text box, note box, calculated field, file upload, and section header.
- b. Categorical field types (dropdown, radio buttons, checkboxes) must also have response options (choices) defined in Column G. Terms used in Column E to define these field types are: *dropdown, radio, checkboxes*.
- c. Text field types (text box or note box) should have validation (Column I) whenever possible. If the validation is “integer” or “numeric”, you should also include the allowable minimum and maximum values (Columns J & K). Text variable cannot also have choices listed in Column G. Terms used in Column E to define these field types are: *text, notes*.
- d. Calculated variables display the result of a calculation based on responses to previous variables. Data cannot be entered in calculated fields. The term used to define this field type in Column E is: *calc*.
- e. The file upload field type allows you to attach a document (e.g. consent form) to the record. The maximum file size for any document is 50Mb. The term used to define this field type in Column E is: *file*.
- f. A section header is a field type used to visually separate fields in a form. Text to be included in a section header should be entered in Column D. The term used to define this field type in Column E is: *text*.

5. Column E – Field Label

Required

- a. A Field Label (or variable label) is a word or phrase that is more descriptive than the variable/field name. It is displayed on the form because (instead of the variable name) because it provides more information to the reader.

6. Column F – Choices, Calculations OR Slider Labels

Required

- a. All categorical field types (yes/no, dropdown, radio buttons, checkboxes) must specify response options associating numerical values with labels. For example, Yes=1, No=0.
- b. All calculated fields must specify the calculation here. Examples of calculation syntax can be found in the REDCap Help Section and in the demonstration data dictionary.
- c. The slider field allows you to label three anchor points: left, middle, and right. An example might be “Strongly Disagree”, “Neutral”, and “Strongly Agree.”

7. Column G – Field Notes

Optional

Field notes are used to provide information to assist in data entry. Examples are specifying the expected format of a validated field (e.g. phone number), or units (e.g. kg vs. lb).

8. Column H – Text Validation Type OR Show Slider Number

Optional

- a. Format validation types for text fields are: date, time, integer, number, zipcode, phone, and email. An error message is displayed if an entry does not match expected format.
 - b. For slider fields, specify whether to display or hide the value (1-100) selected on the slider.
9. Columns I & J – Text Validation Min/Max
Optional
For text validation types of number or integer, minimum and maximum acceptable values may also be specified. An error message, including the acceptable range, is displayed if the entry is out of range.
10. Column K – Identifiers
Optional
To be HIPAA-compliant there are 18 pieces of information that must be marked as “identifiers” in a REDCap data dictionary.
1. Name
 2. Fax number
 3. Phone number
 4. E-mail address
 5. Account numbers
 6. Social Security number
 7. Medical Record number
 8. Health Plan number
 9. Certificate/license numbers
 10. URL
 11. IP address
 12. Vehicle identifiers
 13. Device ID
 14. Biometric ID
 15. Full face/identifying photo
 16. Other unique identifying number, characteristic, or code
 17. Postal address (geographic subdivisions smaller than state)
 18. Date precision beyond year
11. 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 previous fields. For example, a question about pregnancy can be designated to be displayed only if the subject is female. Syntax for branching logic is described in the REDCap Help Section and in the demonstration data dictionary.
12. Column M – Required Field
Optional
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.
13. Column N – Custom Alignment
Optional
The location of text boxes or categorical responses (dropdown, radio, checkbox) can be specified as Right/Vertical, Left/Vertical, Right/Horizontal, Left/Horizontal. The default setting, if not specified, is Right/Vertical.
14. Column O – Question Number (surveys only)
Optional
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.