

Introduction to REDCap for Clinical Data Collection

Shannon M. Morrison, M.S.

Quantitative Health Sciences – Cleveland Clinic Foundation, Cleveland, OH

Abstract

REDCap (Research Electronic Data Capture) is a free web-based tool developed by Vanderbilt University that aids in the collection of clinical data. The application is used primarily for creating databases and/or surveys by using the “Online Designer” in a web browser or constructing a “Data Dictionary” in Excel, which can be uploaded to REDCap. Multiple users can be added to a project and changes can be made quickly and easily online. Users have the ability to add validations and/or branching logic to variables to help with data quality, and once data collection is complete data can be exported to common statistical packages, including SAS®!

Introduction

The Department of Quantitative Health Sciences (QHS) at Cleveland Clinic offers research support in the form of statistical analysis, data management services, and statistical programming. The statistical programming team is charged with cleaning up a dataset from an investigator and getting it ready for analysis. This can often be an incredibly time-consuming process, as most datasheets received by the biostatistics team are full of text fields, misspellings, colors, special characters, etc. It is estimated that 20-80% of the total time the biostatistics team spends on a project is dedicated to data cleaning. This is obviously a waste of research time and money and is a frequent topic of discussion among team members – is there a way to cut down on data cleaning time?

Employees clinic-wide are also strongly focused on maintaining security of data. Everyone almost always work with sensitive data, and preventing violations of patient privacy is of the utmost importance. Any breaches of security can result in workplace and legal repercussions.

REDCap helps with both of these issues. Projects are neatly organized and there are built-in features that a user can add that will help with catching errors at data entry. When creating variables, one first chooses the field type – whether it be a multiple choice question, a text box, etc. Choosing correctly here can eliminate most issues – for example, it’s better to create a patient race variable as a multiple choice question rather than a text entry field (where one could make spelling mistakes). After choosing a field type, optional validations can be used – for example, patient age can be specified as having to be a number...so if a user enters an age of “purple,” the system will alert the user that this is not a valid entry. Minimum and maximum ranges can also be set here, and codes for multiple choice questions can be entered, similar to using the FORMAT PROCEDURE in SAS®. Lastly, branching logic can be used so that a question will only show up if a certain condition (or conditions) is/are met. For example, “Have you ever been pregnant?” – Yes/No” does not need to be asked of males, so branching logic can be added to indicate that the question should only show up if gender = Female for a previous question. In addition to these options in creating the variables, REDCap also has a “Data Quality” section that can highlight data discrepancies as well as a “Graphical Data View & Stats” feature that shows bar charts, box-and-whisker plots and summary statistics for variables as appropriate. These plots and values are not meant to be used for publication, but they are easy ways to get a visual range check of the data or to check for missing values.

Variables can also be designated as an Identifier, that is, a field that contains identifying information (name, Social Security Number, address, etc.). This is especially helpful for data export, as there is an option to export only de-identified data to ensure privacy. In addition to this feature, at Cleveland Clinic REDCap is backed up on secure servers, so data can never be lost. The REDCap website can be accessed from any computer that has internet access, so there is no need to email data or carry data on a flash drive that can be stolen or misplaced. This also ensures that whatever dataset a user is viewing or working on in REDCap is the most up-to-date data available.

Another security feature is the User Rights section. Here, the project leader can designate who has access to the project, what kind of access they have (read-only access? De-identified data only? Ability to create and delete records? etc.), and how long their access will last (with an option to set an expiration date). Users can obtain access to a project only by being invited by the project leader.

REDCap also features a Logging section that is automatically created for each project. In this section, each and every change that is made to a project is recorded, whether it’s updating a record, creating a new variable, etc. Changes can be filtered by user or by type of action, which can be a very handy tool for manuscript review or in the event of an audit.

The goal of QHS is to teach investigators to set up projects on their own before data collection begins. It is recommended that investigators meet with their biostatistics team beforehand to map out the REDCap project and

figure out what questions need to be asked. After collection data can be export to a .csv file. A SAS® program that assigns formats, labels, etc. that were set up in REDCap is also generated at data export. The hope is that this work ahead of time will result in a clean dataset full of numbers only that is ready for analysis.

Creating Forms

REDCap organizes data into what it calls “Forms.” Here, users can create the variables that will be used for data collection. There are two ways to create forms: pointing and clicking through pop-up boxes in REDCap’s Online Designer, or creating a “Data Dictionary” in Excel, which is then uploaded to REDCap.

Online Designer

Users can point-and-click through the Online Designer to create forms. This method is very user-friendly and is the one most often recommended to investigators.

My Projects
Project Home
Project Setup
Project status: Development

Data Collection [Edit instruments](#)

Record Status Dashboard
- View data collection status of all records

Add / Edit Records
- Create new records or edit/view existing ones

Data Collection Instruments:
SASHELP.CLASS

Applications

- Calendar
- Data Export Tool
- Data Import Tool
- Data Comparison Tool
- Logging
- File Repository
- User Rights
- Graphical Data View & Stats

Project Home Project Setup Other Functionality Project Revision History

Project status: Development Completed steps 0 of 6

Main project settings

Not started

Enable Use longitudinal data collection with repeating forms? ?

Enable Use surveys in this project? ? [VIDEO: How to create and manage a survey](#)

I'm done! Modify project title, purpose, etc.

Design your data collection instruments

Add or edit fields in your data collection instruments. This may be done by either using the Online Designer (online method) or by uploading a Data Dictionary (offline method), in which you may use either method or both. Quick links: [Download PDF of all data collection instruments](#) OR [Download the current Data Dictionary](#)

In progress

I'm done! Go **Online Designer** Upload Data Dictionary

Have you checked the [Check For Identifiers](#) page to ensure all identifier fields have been tagged?

Variables can be organized into forms, which can be traditional data collection instruments or surveys. The example project below has only one traditional data collection form called SASHELP.CLASS.

Project Setup Online Designer Upload Data Dictionary

[VIDEO: How to use this page](#)

The Online Designer will allow you to make project modifications to fields and data collection instruments very easily using only your web browser. **NOTE: While in development status, all field changes will take effect immediately in real time.**

Data Collection Instruments

Add new instrument: [Create](#) a new instrument from scratch

[Download](#) a new instrument from the REDCap Shared Library

Instrument name	Fields	View PDF	Instrument actions
SASHELP.CLASS	6		Rename Delete

Clicking on a form name takes us to an area of the Online Designer where we can edit that form. Variables are listed here, and this is where new variables can be added or existing variables can be edited. Clicking “Add Field” or “Add Matrix of Fields” creates a new variable (note that a matrix of fields is a group of questions that all have the same response: i.e. a matrix of “Describe your pain” questions may have fields for areas of the body – Neck, Head, etc. – and they will all have the same responses of Mild, Moderate or Severe).

Project Setup
Online Designer
Upload Data Dictionary

[VIDEO: How to use this page](#)

This page allows you to build and customize your data collection instruments one field at a time. You may add new fields or edit existing ones. New fields may be added by clicking the **Add Field** buttons. You can begin editing an existing field by clicking on the **Edit** icon. If you decide that you do not want to keep a field, you can simply delete it by clicking on the **Delete** icon. To reorder the fields, simply **drag and drop** a field to a different position within the form below. NOTE: While in development status, all field changes will take effect immediately in real time.

[Return to list of Data Collection Instruments](#)

Current instrument: **SASHELP.CLASS** [Preview instrument](#)

Variable: study_id

Study ID

NOTE: The field above is the record ID field and thus cannot be deleted or moved. It can only be edited.

[Add Field](#) [Add Matrix of Fields](#)

Demographics Information

[Add Field](#) [Add Matrix of Fields](#)

Variable: name

Name

[Add Field](#) [Add Matrix of Fields](#)

Variable: gender

Gender

☐ Male
☐ Female

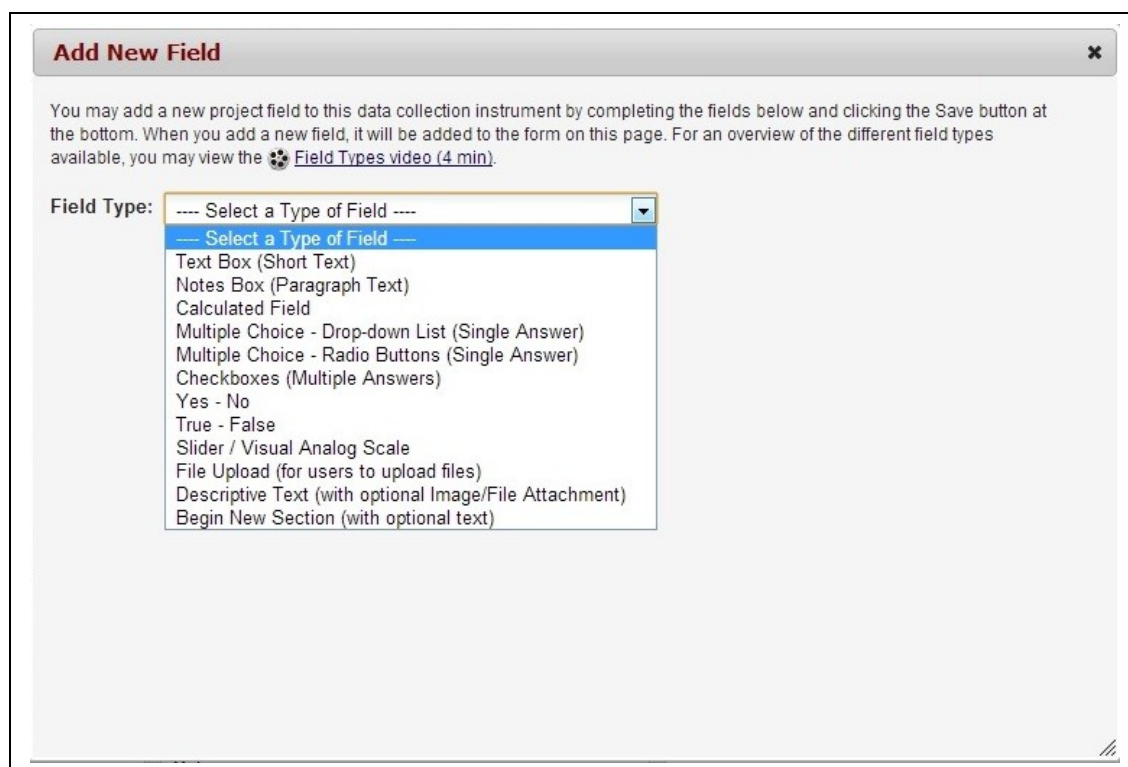
[reset](#)

[Add Field](#) [Add Matrix of Fields](#)

Describe your pain			
	Mild	Moderate	Severe
Neck	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Head	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Chest	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

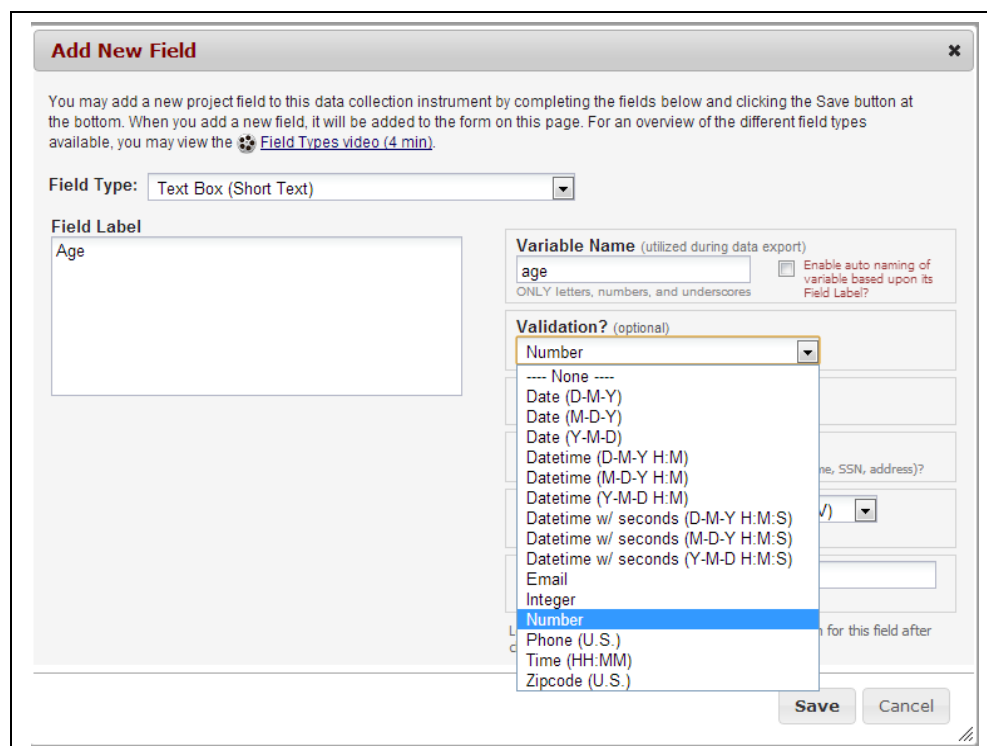
[reset](#)

Clicking on “Add Field” brings up the field type box. Here the user will choose the variable type, name and label the variable, assign formats (if applicable), designate the variable as “Required” or an “Identifier” (if applicable) and add notes if desired.



The screenshot shows the 'Add New Field' dialog box. At the top, there is a title bar with the text 'Add New Field' and a close button. Below the title bar, there is a paragraph of text: 'You may add a new project field to this data collection instrument by completing the fields below and clicking the Save button at the bottom. When you add a new field, it will be added to the form on this page. For an overview of the different field types available, you may view the [Field Types video \(4 min\)](#).' Below this text, there is a 'Field Type:' label followed by a dropdown menu. The dropdown menu is open, showing a list of field types: '--- Select a Type of Field ---', 'Text Box (Short Text)', 'Notes Box (Paragraph Text)', 'Calculated Field', 'Multiple Choice - Drop-down List (Single Answer)', 'Multiple Choice - Radio Buttons (Single Answer)', 'Checkboxes (Multiple Answers)', 'Yes - No', 'True - False', 'Slider / Visual Analog Scale', 'File Upload (for users to upload files)', 'Descriptive Text (with optional Image/File Attachment)', and 'Begin New Section (with optional text)'.

Validations can be set here. For example, if we only want adults in our study, we may want to indicate that 1.) Age should be a numeric value, and 2.) The minimum acceptable age is 18.



The screenshot shows the 'Add New Field' dialog box with the 'Field Type' set to 'Text Box (Short Text)'. The 'Field Label' is 'Age'. The 'Variable Name' is 'age'. There is a checkbox for 'Enable auto naming of variable based upon its Field Label?'. The 'Validation?' dropdown menu is open, showing a list of validation options: '--- None ---', 'Date (D-M-Y)', 'Date (M-D-Y)', 'Date (Y-M-D)', 'Datetime (D-M-Y H:M)', 'Datetime (M-D-Y H:M)', 'Datetime (Y-M-D H:M)', 'Datetime w/ seconds (D-M-Y H:M:S)', 'Datetime w/ seconds (M-D-Y H:M:S)', 'Datetime w/ seconds (Y-M-D H:M:S)', 'Email', 'Integer', 'Number', 'Phone (U.S.)', 'Time (HH:MM)', and 'Zipcode (U.S.)'. The 'Number' option is selected. At the bottom right, there are 'Save' and 'Cancel' buttons.

A field can also be designated as an Identifier if a de-identified dataset is needed for analysis:

Edit Field

You may add a new project field to this data collection instrument by completing the fields below and clicking the Save button at the bottom. When you add a new field, it will be added to the form on this page. For an overview of the different field types available, you may view the [Field Types video \(4 min\)](#).

Field Type: Text Box (Short Text)

Field Label
Name

Variable Name (utilized during data export)
name ☐ Enable auto naming of variable based upon its Field Label?

Validation? (optional)
---- None ----

Required?* ☒ No ☐ Yes
* Prompt if field is blank

Identifier? ☐ No ☒ Yes
Does the field contain identifying information (e.g. name, SSN, address)?

Custom Alignment Right / Vertical (RV)
Align the position of the field on the page

Field Note (optional)
Small reminder text displayed underneath field

Looking for Branching Logic? Try clicking the icon for this field after clicking the Save or Cancel button below.

Save **Cancel**

The Branching Logic Feature hides questions unless a certain condition/conditions is/are met. For example, if the project had a field titled “Pregnant? Yes/No,” we would not need to ask this question if the respondent is a Male. We can set Branching Logic by coding the syntax or using the Drag-N-Drop builder to indicate that the pregnancy question should only show up if the respondent is Female based on the earlier Gender question.

Add/Edit Branching Logic

Branching Logic may be employed when fields/questions need to be hidden under certain conditions. If branching logic is defined, the field will only be visible if the conditions provided are true (i.e. show the field only if...). You may specify those conditions in the text box below for the Advanced Branching Logic Syntax or by choosing the Drag-N-Drop Logic Builder method, which allows you to build your logic in a much easier fashion by simply dragging over the options you want. You may switch back and forth between each method if you wish, but please be aware that since the advanced logic allows for greater complexity, it may not be able to be switched over to the Drag-N-Drop method if it becomes too complex.

Choose method below for the following field: **pregnant - Pregnant?**

☒ **Advanced Branching Logic Syntax** [\(How do I use the advanced syntax?\)](#)

Show the field ONLY if...
[gender] = 1

— OR —

☐ **Drag-N-Drop Logic Builder**

Field choices from other fields
(drag a choice below to box on right)

- study_id = (define criteria)
- name = (define criteria)
- gender = Male (0)
- gender = Female (1)
- age = (define criteria)
- height = (define criteria)
- weight = (define criteria)
- sashelpclass_complete = Incomplete (0)
- sashelpclass_complete = Unverified (1)

Show the field ONLY if...

- ☒ ALL below are true
- ☐ ANY below are true

gender = Female (1)

Save **Cancel**

Data Dictionary

Data Dictionaries are created as Excel files that are uploaded to REDCap. Instead of pointing and clicking through the Online Designer, the Data Dictionary creates the REDCap forms. This method is not as user friendly as the Online Designer, but can be useful for projects with many similar variables.

A portion of the Data Dictionary for the above example is shown below (note that the dictionary was split into 2 screen shots to fit on the page and still be readable...in a true Data Dictionary there is one record per field):

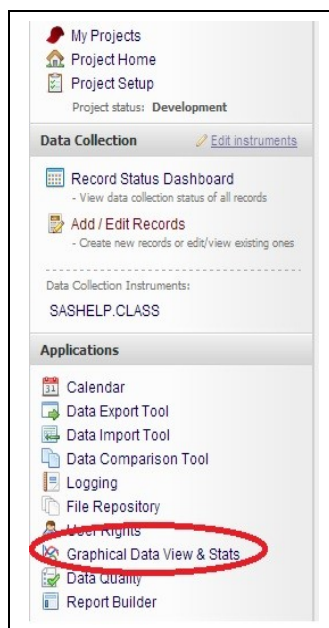
	A	B	C	D	E	F
1	Variable / Field Name	Form Name	Section Header	Field Type	Field Label	Choices, Calculations, OR Slider Labels
2	study_id	sashelpclass		text	Study ID	
3	name	sashelpclass	Demographics Information	text	Name	
4	gender	sashelpclass		radio	Gender	0, Male 1, Female
5	age	sashelpclass		text	Age	
6	height	sashelpclass		text	Height	
7	weight	sashelpclass		text	Weight	
8	pregnant	sashelpclass		yesno	Pregnant?	

	H	I	J	K	L
1	Text Validation Type OR Show Slider Number	Text Validation Min	Text Validation Max	Identifier?	Branching Logic (Show field only if...)
2					
3					
4					
5	number	18			
6	number				
7	number				
8					[gender] = '1'

Note that everything in the above boxes is shown here: Field Names, Form Names, Multiple Choice Codes, Field Type, Validations, Branching Logic, etc. Importing this Excel sheet into REDCap would create the same project as the point-and-click example creates above in the Online Designer.

Graphical Data View & Stats

This section of REDCap provides a way to get a quick look at the data to note any missing values or outliers.



Graphs provided include box-and-whisker plots for continuous variables and bar charts for categorical variables:



Descriptive statistics are also provided for continuous variables:

Plots Descriptive Stats


Total records: 19

Field	Missing	Min	Max	Mean	Median	StDev	CV
Study ID	0						
Name	0						
Gender	0						
Age	0	11.00	16.00	13.32	13.00	1.49	0.11
Height	0	51.30	72.00	62.34	62.80	5.13	0.08
Weight	0	50.50	150.00	100.03	99.50	22.77	0.23
Pregnant?	19						
Complete?	13						

It is important to note that 1.) Continuous variables are only shown here IF they are indicated as a Number or Integer in the Online Designer/Data Dictionary, and 2.) This section should be used for visual checks ONLY (i.e. looking for outliers or missing values). REDCap is meant to be used as a data storage tool, not as something to produce summary statistics or graphics for publication.

Data Quality

The Data Quality section is another tool for users to quickly assess missing values, outliers or any other potential data issues. REDCap provides common pre-defined “rules” to execute, and the user can code in custom rules as well to check the data.


Data Quality

This module will allow you to execute data quality rules upon your project data to check for discrepancies in your data. Listed below are some pre-defined data rules that you may utilize and run. You may also create your own rules or edit, delete, or reorder the rules you have already created. To find discrepancies for a given rule, simply click the Execute button next to it, or click the Execute All Rules button to fire all the rules at once. It will provide you with a total number of discrepancies found for each rule and will allow you to view the details of those discrepancies by clicking the View link next to each. [Read more detailed instructions.](#)

Data Quality Rules					Processing Complete!	Execute All Rules	Clear
	Rule #	Rule Name	Rule Logic (Show discrepancy only if...)	Total Discrepancies	Delete rule?		
	a	Missing values*	-	76 view			
	b	Missing values* (required fields only)	-	0 view			
	c	Field validation errors (incorrect data type)	-	0 view			
	d	Field validation errors (out of range)	-	0 view			
	e	Outliers for numerical fields (numbers, integers, sliders, calc fields)	-	4 view			
	f	Hidden fields that contain values**	-	0 view			
	g	Multiple choice fields with invalid values	-	0 view			
		<input type="text"/> <input type="button" value="Add"/>	<input type="text"/> Enter logic for new rule (e.g. [age] < 18) How do I use special functions?				

* The Missing Values rules above automatically exclude fields hidden by branching logic when searching for missing values. If a field is hidden by branching logic on a data entry form or survey, then it is expected that such a field would not have a value. Thus for these cases, the values for those hidden fields will not be classified as missing. Additionally, checkbox fields are also excluded since an unchecked checkbox is itself often considered to be a real value.

** The term 'hidden fields' refers to any fields on a survey or data entry form that are not being displayed because branching logic is hiding them, which assumes that the field's value should be blank/null.

Upon viewing discrepancies, users have the option of excluding certain results from appearing the next time the rule is run. Clicking on the hyperlinked value will take the user directly to that record's form, where the value can be changed.

Rule: Outliers for numerical fields(numbers, integers, sliders, calc fields)				
Discrepancies found: 4				
Record	Discrepant fields with their values	Status	Exclude 	
15	age = <u>16</u> (median: 13, stdev: 1.49)	Outlier	exclude	
11	height = <u>51.3</u> (median: 62.8, stdev: 5.13)	Outlier	exclude	
11	weight = <u>50.5</u> (median: 99.5, stdev: 22.77)	Outlier	exclude	
15	weight = <u>150</u> (median: 99.5, stdev: 22.77)	Outlier	exclude	

User Rights

The User Rights section is a nice security feature of REDCap. Here, the project leader can invite other users to work on the project. A REDCap user will only be able to view a project if they are the ones who created the project or if they are designated as a valid user by the project leader.

Project Setup

User Rights

Data Access Groups

This page may be used for granting new users access to the project and for editing the rights of current project users. You may edit the rights of a current user by selecting them from the dropdown list below or add a new user by entering their user name in the text box and hitting the Tab key.

Choose existing project user

-- select user --

OR type a new user and hit the TAB key

New user name

Project leaders can use this section to manage user permissions and (if applicable) usage expiration dates.

Adding new User "new_user"

Basic User Rights

Calendar

☒

Data Export Tool

☐ No Access
☒ De-Identified
☐ Full Data Set

Data Import Tool

☐

Data Comparison Tool

☐

Logging

☐

File Repository

☒

User Rights

☐

Data Access Groups

☐

Graphical Data View & Stats

☒

Data Quality

☐ Create & edit rules
☐ Execute rules

[What is Data Quality?](#)

Reports & Report Builder

☒

Project Design and Setup

☐

API

☐ API Export
☐ API Import

[What is the REDCap API?](#)

Settings pertaining to record locking and E-signatures:

Record Locking Customization

☐

Lock/Unlock Records

☒ Disabled
☐ Locking / Unlocking
☐ Locking / Unlocking with E-signature authority

Users with locking privileges also have access to the E-signature and Locking Mgmt page on the left-hand Applications menu.

[Watch video about locking](#)

[What is an E-signature?](#)

Allow locking of all forms at once for a given record?

☐

Settings pertaining to project records: [Explain these settings](#)

Create Records

☒

Rename Records

☐

Delete Records

☐

Expiration Date (if applicable)

Add User

- Cancel -

Data Entry Rights

NOTE: The data entry rights "only" pertain to a user's ability to view or edit data on the web page. It has no effect on what data is included in data exports.

	No Access	Read Only	View & Edit
SASHELP.CLASS	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>

Logging

All REDCap projects have an automatically created Logging section that keeps track of every single change made to a project. For each change, the log shows the date and time of the change, the name of the user who made the change, the “action” (what kind of change it was), and a brief description of the action.

 **Logging**

 [Download entire logging record to Microsoft Excel \(CSV\)](#)

This module lists all changes made to this project, including data exports, data changes, and the creation or deletion of users.

Filter by event: All event types (excluding page views) ▼

Filter by user name: All users ▼

Filter by record: All records ▼

Displaying events (by most recent): 1 - 54 ▼

Time / Date	Username	Action	List of Data Changes OR Fields Exported
12:46pm 08/08/2013	morriss2	Manage/Design	Download data dictionary
12:46pm 08/08/2013	morriss2	Manage/Design	Edit project field
12:46pm 08/08/2013	morriss2	Manage/Design	Reorder project fields
12:46pm 08/08/2013	morriss2	Manage/Design	Add/edit branching logic
12:45pm 08/08/2013	morriss2	Manage/Design	Download data dictionary
12:39pm 08/08/2013	morriss2	Manage/Design	Create project field
12:32pm 08/08/2013	morriss2	Manage/Design	Create project field
12:22pm 08/08/2013	morriss2	Manage/Design	Delete project field
2:33pm 04/24/2013	morriss2	Updated Record 9	
2:30pm 04/24/2013	morriss2	Updated Record 10	sashelpclass_complete = '2'
2:03pm 04/24/2013	morriss2	Updated Record 17	sashelpclass_complete = '1'
2:03pm 04/24/2013	morriss2	Updated Record 6	sashelpclass_complete = '2'
2:02pm 04/24/2013	morriss2	Updated Record 14	sashelpclass_complete = '2'
2:02pm 04/24/2013	morriss2	Updated Record 9	sashelpclass_complete = '1'
2:02pm 04/24/2013	morriss2	Updated Record 1	sashelpclass_complete = '2'
11:32am 02/07/2013	morriss2	Manage/Design	Create report
11:27am 02/07/2013	morriss2	Manage/Design	Download exported data file (CSV raw)
11:27am 02/07/2013	morriss2	Data Export	study_id, name, gender, age, height, weight, sashelpclass_complete
4:05pm 02/06/2013	morriss2	Manage/Design	Edit project field
4:05pm 02/06/2013	morriss2	Manage/Design	Edit project field
4:05pm 02/06/2013	morriss2	Manage/Design	Edit project field
4:05pm 02/06/2013	morriss2	Manage/Design	Modify project settings
4:04pm 02/06/2013	morriss2	Manage/Design	Rename data collection instrument
4:04pm 02/06/2013	morriss2	Created Record (import) 19	study_id = '19', name = 'William', gender = '0', age = '15',

Changes can be filtered by event, user name or record using the drop-down lists. The log itself can be exported to a .csv file and viewed in Excel.

In addition to the main log file, REDCap creates a log for each variable within a record. Clicking on the **H** button next to a field in the Online Designer brings up the history for that field.

SASHELP.CLASS [Modify this instrument](#) [VIDEO: Basic data entry \(16 min\)](#)

Download PDF of - select PDF download option -

Editing existing Study ID 9

Study ID 9

Demographics Information

Name **H** Jeffrey

Gender **H** ☒ Male ☐ Female [reset](#)

Age **H** 20 [View data history](#)

Height **H**

Weight **H** 84

Form Status

Complete? **H** Unverified [Save Record](#) [Save and Continue](#) [-- Cancel --](#)

Data History for variable "age" for record "9" [x](#)

Listed below is the history of all data entered for the variable "age" for Study ID "9".

Date/Time of Change	User	Data Changes Made
2:31pm 08/08/2013	morris2	20
4:04pm 02/06/2013	morris2	13

[Close](#)

Data Export

Once collection is complete, data can be exported to statistical packages for analysis.

Form: **SASHELP.CLASS**

Study ID (*study_id*)

Demographics Information

Name (*name*)

Gender (*gender*)

Age (*age*)

Height (*height*)

Weight (*weight*)

Pregnant? (*pregnant*)

Form Status

Complete? (*sashelpclass_complete*)

De-Identification Options (optional)

The options below allow you to limit the amount of sensitive information that you are exporting out of the project. Check all that apply.

Known Identifiers:

☐ Remove all known Identifier fields (*tagged in Data Dictionary*)

☐ Hash the Study ID (*converts record name to an unrecognizable value*)

Free-form text:

☐ Remove unvalidated Text fields (*i.e. Text fields other than dates, numbers, etc.*)

☐ Remove Notes/Essay box fields

Date and datetime fields:

☐ Remove all date and datetime fields

— OR —

☐ Shift all dates by value between 0 and 364 days (*shifted amount determined by algorithm for each record*)
[What is date shifting?](#)






[Deselect all options](#)

Submit

Here a user can select variables to export. The user can also choose to export a de-identified dataset, which would not include variables indicated as Identifiers in the Online Designer/Data Dictionary. Other de-identification options are listed here as well.

13

REDCap will generate certain files depending on which statistical package is being used for analysis.

		Download Syntax & Data
	Microsoft Excel You may download the survey results in CSV (comma-separated) format, which can be opened in Excel. You have the choice of downloading the data either with the full headers and answer labels or just with the answer codes (i.e. raw data). <i>NOTE: If you are using a version of Microsoft Excel prior to Excel 2007, due to limitations the data will only be read to 255 columns when opened.</i>	<div> <div>EXCEL CSV</div> <div>Labels</div> </div> <div> <div>EXCEL CSV</div> <div>Raw</div> </div> <div>Send file?</div>
	SPSS Statistical Analysis Software Instructions: Download and save all 3 files on the right to a common location. First, double-click on the Pathway Mapper (.bat) file, which will run quickly and invisibly. (If you are not using a Windows operating system, such as Mac or Linux, please see the Additional Instructions .) Now double-click on the *.sps file, which will open SPSS. When the file is loaded and displayed, choose Run-->All from the top menu options. This action will launch the script that will automatically read in all data and manipulate data fields with labels, option values, etc. Additional instructions	<div> <div>SPSS</div> <div>DATA CSV</div> </div> <div>Pathway Mapper</div> <div>Send file?</div>
	SAS Statistical Software Instructions: Download and save all 3 files on the right to a common location. First, double-click on the Pathway Mapper (.bat) file, which will run quickly and invisibly. (If you are not using a Windows operating system, such as Mac or Linux, please see the Additional Instructions .) Now double-click on the *.sas file, which will open SAS. When the file is loaded and displayed, choose Run (or Run-->Submit) from the top menu options. This action will launch the script that will automatically read in all data and manipulate data fields with labels, option values, etc. Additional instructions	<div> <div>SAS</div> <div>DATA CSV</div> </div> <div>Pathway Mapper</div> <div>Send file?</div>
	R Statistical Software Instructions: Use command read.csv('filename') to read in data file.	<div> <div>R</div> <div>DATA CSV</div> </div> <div>Send file?</div>
	STATA Analysis and Statistical Software Instructions: Download both files to common location and double-click on *.do file. This action will launch the script that will automatically read in all data and manipulate data fields with labels, option values, etc.	<div> <div>STATA</div> <div>DATA CSV</div> </div> <div>Send file?</div>

When using SAS® for analysis, three files are generated: a .csv file (which has the data), a SAS® file (which assigns labels and formats to the variables – these are directly from the Online Designer/Data Dictionary), and a Pathway Mapper (a Windows batch file which will tell the SAS® file where to locate the data if using SAS® on Windows).

A portion of the Excel sheet and SAS® program are shown below for the SASHELP.CLASS example:

	A	B	C	D	E	F	G	H
1	1	Alfred	0	14	69	112.5		2
2	2	Alice	1	13	56.5	84		0
3	3	Barbara	1	13	65.3	98		0
4	4	Carol	1	14	62.8	102.5		0
5	5	Henry	0	14	63.5	102.5		0
6	6	James	0	12	57.3	83		2
7	7	Jane	1	12	59.8	84.5		0
8	8	Janet	1	15	62.5	112.5		0
9	9	Jeffrey	0	13	62.5	84		1
10	10	John	0	12	59	99.5		2
11	11	Joyce	1	11	51.3	50.5		0
12	12	Judy	1	14	64.3	90		0
13	13	Louise	1	12	56.3	77		0
14	14	Mary	1	15	66.5	112		2
15	15	Philip	0	16	72	150		0
16	16	Robert	0	12	64.8	128		0
17	17	Ronald	0	15	67	133		1
18	18	Thomas	0	11	57.5	85		0
19	19	William	0	15	66.5	112		0


```

data redcap;
  set redcap;
  label study_id='Study ID';
  label name='Name';
  label gender='Gender';
  label age='Age';
  label height='Height';
  label weight='Weight';
  label pregnant='Pregnant?';
  label sashelpclass_complete='Complete?';
run;

proc format;
  value gender_ 0='Male' 1='Female';
  value pregnant_ 1='Yes' 0='No';
  value sashelpclass_complete_ 0='Incomplete' 1='Unverified'
    2='Complete';
run;

data redcap;
  set redcap;

  format gender gender_.;
  format pregnant pregnant_.;
  format sashelpclass_complete sashelpclass_complete_.;
run;

```

During the analysis phase, the biostatistics team can simply put a %INCLUDE statement at the beginning of the SAS® analysis program that points to the location of the above SAS® file generated by REDCap. The dataset WORK.REDCAP will be created using labels and formats assigned in REDCap, and will be ready for analysis.

Conclusion

Data quality and security are of the utmost importance, especially when working with sensitive data. Too much time and money is wasted cleaning the data to get it ready for analysis, and data can easily be lost or stolen if it is sent over email, saved on a flash drive, etc.

REDCap is a web-based tool that can help ensure data quality. Data fields are created using the Online Designer or Data Dictionary. Here, users can create variables with labels and formats, similar to how one would code variables in SAS®. In addition to choosing the data field type, a user can also indicate validations that can help in catching data entry errors. Branching Logic can be used to hide questions unless certain circumstances, decreasing the chance of unnecessary data entry. The Data Quality and Graphical Data View & Stats sections can be used to visually assess the data; here, it's easy to see out-of-range values and get a quick count of missing responses for each variable.

REDCap also boasts attractive security features. At Cleveland Clinic, REDCap data is backed up on secure servers, meaning that data will never be lost. In the User Rights section, the project leader can add users to a project and modify those users' permissions and usage expiration dates.

Every REDCap project automatically generates a log file. This file keeps track of every single change made to the project, and these changes can be filtered or exported to a .csv file. It is also possible to see changes at the variable level within each record.

Once data collection is complete and the team is ready for analysis, data can quickly be exported from REDCap and is ready to use in SAS®. The data itself is in a .csv file – REDCap also generates a SAS® file with labels and formats. Including this file in the analysis program will generate the analysis dataset.

REDCap is a great tool for any analysis team. If used correctly and to its full advantage, it can produce high-quality data that is stored securely, resulting in high-quality results.

References

- (1) Paul A. Harris, Robert Taylor, Robert Thielke, Jonathon Payne, Nathaniel Gonzalez, Jose G. Conde.
"Research electronic data capture (REDCap) – A metadata-driven methodology and workflow process for providing translational research informatics support," *J Biomed Inform.* 2009 Apr; 42(2): 377-81.
- (2) Worley, Sarah. "Overview of Databases for Research." Cleveland Clinic Presentation.
- (3) Butler, Robert S. "Data is a Dish Best Served Raw." Cleveland Clinic Presentation.
- (4) Sharp, John. "Introduction to REDCap." Cleveland Clinic Presentation.

Acknowledgements

The author wishes to acknowledge the QHS REDCap and Alpha Beta teams for their continued help and input when it comes to using – and talking about using – REDCap!

Contact Information

Your comments and questions are valued and encouraged. Contact the author at:

Name: Shannon Morrison, M.S.
Enterprise: Cleveland Clinic Foundation
Address: 9500 Euclid Avenue / JJN3-01
City, State ZIP: Cleveland, OH 44195
Phone: (216) 636-5413
Fax: (216) 444-8021
E-mail: morris2@ccf.org
Web: <http://lerner.ccf.org/qhs>

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.