

# REDCap General Guidelines

This document provides general guidelines for REDCap database design. When creating your database, keep in mind optimizing for data entry and data analysis.

## Preparation

We encourage research PIs to collect the least amount of Personal Identifiable Information (PII) or Personal Health Information (PHI) as possible. *Recommended reading before collecting PHI:* [A beginner's guide to avoiding Protected Health Information \(PHI\) issues in clinical research – With how-to's in REDCap Data Management Software](#)

- Plan your data collection form before you start creating the forms in REDCap.
- Document your data collection needs in a study protocol or similar document.
- Group your data fields/questions based on how they are being collected. For example:
  - Data source (chart review, patient assessment, lab report, etc.)
  - Time point (baseline clinic visit, follow-up visit, etc).
  - Method of collection (the question flow should mirror the actual data collection process).
- You may find it helpful to draft a paper form in Word before you start in REDCap
- Always keep a copy of your Data Dictionary (Downloaded to Excel)
- Do not change variable names or values of categorical field types once you begin collecting real data
- Group variables together that follow the data entry work flow
- Keep forms fairly short to minimize risk of data loss
- Use categorical field types (yes/no, multiple choice, etc.) whenever possible
- When using text fields, add field validation to minimize data entry errors
- Involve a statistician early during the development of the database
- Using the multi-page surveys feature will allow the participant to complete the survey in sections with each completed page uploading to REDCap
- Enable the Save and Return later feature for large surveys
  
- **Online Survey and Forms** – As a rule of thumb, online forms (especially in the case of surveys) are generally kept to a page or two. While you could make them much longer, users tend to get overwhelmed when they have to scroll across page after page of questions. Very long forms also take longer to complete. If you are distracted mid-form and do not save your work, your internet session could time-out (after 30 minutes of inactivity) and all unsaved data would be lost. So, if you think you're starting to get carried away with a long series of questions, it may be preferable to break your questions up into multiple forms.

- **Data Entry** – Group variables together that follow the data entry work flow, and use field types that minimize changing from keyboard to mouse. For example, you can enter a drop-down field option by typing the first character of the label, allowing you to “tab and type” through the data entry fields, while radio buttons require using the mouse to select an option). Keep forms fairly short to minimize risk of data loss (by saving more often when completing a form) and make it easier to identify data entry errors.
- **Data Analysis and Reporting** – Focus on how you want your data to look at the end of the study so you address all important data fields and design elements at the beginning. Consider creating the tables you hope to have for your final data analyses in draft form to check that you are collecting the data you will need. Use categorical response (drop-down, 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. 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.
- **Export only when necessary** – One of REDCap’s greatest strength is the security of your data. Take precaution when exporting data and only export data if you need to run reports or analyses outside of REDCap. Limit user privileges to only allow export rights to those who really need it.
- **Use REDCap’s Send-It Feature to send data** – Send-It is a secure data transfer application that allows you to upload a file (up to 20 MB in size) and then allow multiple recipients to download the file in a secure manner. Each recipient will receive an email containing a unique download URL, along with a second follow-up email with the password (for greater security) for downloading the file. The file will be stored securely and then later removed from the server after the specified expiration date. Send-It is the perfect solution for anyone wanting to send files that are too large for email attachments or that contain sensitive data.
- **Emailing Surveys** – You should not use the REDCap ‘Participant Email Contact list’ with group email addresses or distribution lists. The emailed invitations send only 1 unique survey link per email address; therefore, only the first person in the distribution group who clicks on the email link would be able to complete the survey.
  - For group distribution lists, you can email the general survey link provided at the top of the “Invite Participants” page directly from your email account. Or, you can add each individual email address from a distribution list to the Participant Contact list. You can copy/paste the emails from a list (word or excel) into REDCap.
  - The advantages of using REDCap’s Participant Contact list and the individual emails is that REDCap will track respondents and non-respondents for you. You’ll be able to email only non-respondents if you want to send a reminder. With the general distribution email, you won’t be able to track responses and participants will have the potential to complete the survey more than once.
- **Identifiers** – Understand the different types of identifiers:
  - The first variable listed in your project is the **unique identifier** which links all your data. **Do not use Protected Health Information (PHI) identifiers** such as medical record number or date of birth or initials as the unique identifier, as it could be accidentally displayed if a URL is created.

- In Data Entry projects, you must define the unique identifier field. For projects where a survey is the first data collection instrument, it is automatically defined as the Participant ID. The Participant ID value is numeric and auto-increments starting with the highest value in the project. If no records exist, it will begin with '1'. Users can define the unique ID for projects with surveys instead of using the participant\_id by having the first data collection instrument as a data entry form (do NOT enable it as a survey).
- The optional **secondary unique field** may be defined as any field on the data collection instruments. The value for the field you specify will be displayed next to the Participant ID (for surveys) or next to your unique identifier when choosing an existing record/response. It will also appear at the top of the data entry page when viewing a record/response. Unlike the value of the primary unique identifier field, it will not be potentially visible in a URL.
  - The data values entered into the secondary unique field must also be unique. The system will not allow for duplicate entries and checks values entered in real time. If a duplicate value is entered, an error message will appear and the value must be changed to save/submit data entered on the data entry instrument.
  - Common secondary unique identifiers are medical record numbers, subject name, and subject birth date.
- The **redcap\_survey\_identifier** is the identifier defined for surveys when utilizing the Participant Email Contact List and sending survey invitations from the system. The “Participant Identifier” is an optional field you can use to identify individual survey responses so that the participant doesn’t have to enter any identifying information into the actual survey. This field is exported in the data set; the email address of the participant is not.
- **Test and Retest the Project; repeat: TEST and RETEST!!** – Test the project prior to requesting the project be moved to Production Mode. The testing should include data entry, review of project unique identifier, data export formats, etc., to ensure the project design is suitable and appropriate. For surveys, pilot the instrument and test multiple iterations of potential answers.
  - Save a copy of the forms or data dictionary before changing any items in the test phase.
  - Test the project with fake data, in all instruments and events to validate instruments and event definitions, branching logic, calculated fields, and minimum/maximum ranges. Entering and saving test data is the only way to test that the branching logic and calculated fields are properly working. Do not enter real data into the development phase of the project.
  - Review test data: open data entry forms, create reports, export data and send to the Secondary Owner, co-investigators, and any data analyst or statistician to review.
    - It is important to think through the planned statistical analysis before collecting any data. A statistician can make sure you are collecting the fields you need, in the format you need them, in order to perform the needed statistical analyses. For projects with a large amount of data and many forms, consider having the data dictionary of the database reviewed by the Statistician. The data dictionary is easy to download. This will clearly communicate any defined branching logic that is not communicated in the raw data file or meta-data formatting accessible through the “Data Export” application. This will also clearly communicate the formulas from calculated fields. Also, send the blank case report form or other data collection tools. Have the Statistician perform a data export to ensure it does not extract identifiers. The statistician can give you feedback regarding the overall design of your database, as well as the definition of each field.

- After adequate testing, and before submitting, download a codebook and PDFs of forms.
- After adequate testing and downloading the codebook and PDFs of forms, submit the project to the REDCap Administrators to approve a move to Production Phase. Ideally, this should be done after the IRB has approved the data entry forms when IRB review is applicable.

- **Other Design Considerations-**

- The automatic survey feature requires respondents to answer every question (forced choice radio button or drop-down, etc.) in order to move to the next question. Consider having a choice be “I prefer not to answer” or “Other”. If desirable, a branched logic free text response can be linked to collect more specific information from the respondent on the reason for the answer, “Other”.
- Note that data dictionary uploads are only available in the Development Phase.
- Be sure your variable names and field types are correct before requesting to move to production phase – later changes in a variable name, field type, or field label can cause data loss. Similarly adding or modifying branching logic later can cause data loss.