# A black and red logo Description automatically generated

# REDCap Data Quality Process

**Introduction**

Implementing a project in REDCap automatically imposes a degree of structure on data collection. However, without a formal data quality process, even the best-designed project can be bogged down over time as a result of poor data quality. This document outlines a data quality process that offers a starting point for most REDCap projects. It is not intended as a detailed discussion of data quality practices.

**Critical Variables**

Identify your critical variables. For a research study database, these are normally the variables that represent study outcomes or any variables that have the capacity to influence outcomes. They may also include variables that represent dates and times. (No one, least of all a statistician, likes to see study events that are out of sequence.)

Supporting data, that is not critical, should also be collected as accurately as possible and you may wish to apply checks to this data. However, in a large project it is often most practical to focus resources on the most critical data first.

## Instrument Design

**Enforce Data Types**

Your REDCap project should be designed to enforce data types wherever possible. Use REDCap’s variable types and validations to ensure that data is formatted correctly during data entry.

* Avoid free text if the data is known to be categorical. Use radio buttons, drop-downs and checkboxes instead.
* Apply an appropriate validation to numeric variables. How many decimal places do you require? Is the variable collecting only integer data, etc?
* If you are collecting email addresses, phone numbers, etc. choose an appropriate validation.
* Apply data types and validations wisely. They are “hard” checks that can prevent data entry if applied incorrectly.

**Range Checks**

For numerical data add range checks to any variables where values are expected to fall within a specific range.

* Range checks are an aide to data entry, alerting the user to values that may be incorrect. They are “soft” checks. When range checks are violated REDCap displays a warning but it does not prevent data entry from continuing.
* If applied too rigorously range checks can be annoying or an impediment to data entry.
* For clinical values where results are expected to be within a “normal range” consider your study population. For example, in an oncology study where participants are often quite sick, range checks based on “normal range” may be too narrow and result in too many interruptions to data entry. In this scenario it is better to use normal range +/- a percentage.

**Inform Data Entry**

Where data is expected in a particular format or have known units it is helpful for data entry staff to understand what is expected. Use REDCap’s Field Note feature to display hints such as*:*

Systolic Blood Pressure: *mmHg. (Int)*

Weight: *Kg (nnn.nn)*

Name: *First, Last*

**Branching**

Use branching logic to display only variables that are relevant based on current context. For example, do not display the pregnancy test question if your participant is male or is not of childbearing potential.

**Data Quality Rules**

REDCap provides the Data Quality module where you can write rules to identify possible issues with your data. Building data quality rules in REDCap may look complex, but it is really not much harder than writing branching logic. REDCap does provide a few standard, built-in rules. These built-in rules can be useful, but may also create too much noise. It is often better to build custom rules to specifically check your data. Such rules might be…

* Missing data - For each critical variable, check that the data has been entered.
* If your study has inclusion/exclusion data, check that these questions are answered correctly to allow inclusion in the study.
* Has the study participant consented? If so, does the date of consent fall within an appropriate range of dates for the study?

**Data Listings**

Some checks are complex to program into data quality rules, or may be too time consuming to be practical. Under these circumstances it may be more efficient to configure reports that display the data to be checked. It is often easier to review a listing than to inspect individual pieces of data. For example, in a clinical trial where clinic visits and procedures are expected to follow sequentially, it is often best to simply list the dates and review them manually.

**Discrepancy Resolution**

Once you discover an issue or discrepancy in your data it will need to be resolved. For smaller, less complex projects this might simply involve correcting the data or emailing someone to establish a correct value. For a multi-site study you may wish to use REDCap’s built-in query process wherein queries can be entered into the system and assigned to users for resolution.

Consider writing and maintaining a data handling manual. This is an important tool that can help you and the study team handle issues consistently over time. Every time a situation arises that requires a decision you can document that decision in the data handling manual. Next time a similar situation arises the solution will already be documented and can be applied consistently.

**Process**

Once you have an appropriate study design and data quality checks implemented in REDCap it is helpful to establish and document a consistent and regular process for reviewing the data and applying the checks. The process could include:

* Manual review of newly entered data
* Running data quality rules
* Reviewing reports

How frequently you review your data using this process will depend on the study. Some studies will require a weekly review, others may only require a monthly or six-monthly review. Data quality should be reviewed on an ongoing basis throughout the project. Don’t leave it to the end. It is more efficient to handle data issues as they occur rather than after a significant time has elapsed. At the very least, conduct a quality review before each key project milestone, such as an interim analysis or independent data review.

Assign a member of the study team to take responsibility for data quality and to manage this process. In a multi-site study, it is helpful to make each site accountable for their data issues and REDCap keeps metrics relating to queries and query resolution which may be helpful.