

Protecting High Risk Data in REDCap

Background

While people may only think of clinical data that is available from electronic medical records as high risk data, many non-clinical projects contain data that includes high risk data, such as individually identifiable health information and protected health information (PHI). There is a wealth of information about an individual that could be considered high risk data, such as information about health status and provision of health care. There is also non-health related data that should be protected as high risk data, such as criminal activity or financial information.

This document outlines requirements and recommendations pertaining to protecting high risk data.

The Health Insurance Portability and Accountability Act of 1996 (HIPAA), as well as the Health Information Technology for Economic and Clinical Health Act of 2009 (HITECH), established rules protecting the privacy and security of individually identifiable health information. For example, the HIPAA Privacy Rule and Security Rule set national standards requiring organizations and individuals to implement certain administrative, physical, and technical safeguards to maintain the confidentiality, integrity, and availability of PHI.

REDCap is *HIPAA-capable*. It contains the necessary components for HIPAA compliancy, but it is the *environment* into which the software is installed that makes it compliant.

Therefore, HIPAA compliancy requires that certain practices, such as limiting access to high risk data and restricting export of high risk data, are thoroughly documented and communicated to users. As such, part of what makes REDCap *HIPAA-compliant* is its users.

What is High Risk Data?

ALL the following are designated *high risk data* and must be stored and transmitted in accordance with HIPAA standards:

- Health Information
- Individually Identifiable Health Information (PII)
- Protected Health Information (PHI)

This effectively means that REDCap users are expected to treat all health-related data (unless de-identified) as covered by HIPAA requirements—with the highest levels of privacy and security possible—regardless of the source of the data.

Further, the HIPAA Privacy Rule requires that investigators take reasonable steps to limit the use or disclosure of, and requests for, high risk data to the “minimum necessary” to accomplish the intended purpose. REDCap users are expected to always operate according to the “minimum necessary” standard (e.g., limit data access to necessary team members; do not export, share, or transfer data unless absolute necessary; etc.).

Health Information, Individually Identifiable Health Information, and Protected Health Information (PHI) are defined as follows:

1. Health Information
 - Any information, including genetic information, whether oral or recorded in any form or medium, that: (1) is created or received by a Health Care Provider, Health Plan, public health authority, employer, life insurer, school or university, or Health Care Clearinghouse; and (2) relates to the past, present, or future physical or mental health or condition of an Individual; the provision of health care to an Individual; or the past, present, or future payment for the provision of health care to an Individual. (45 CFR § 160.103)
2. Individually Identifiable Health Information
 - Information that is a subset of Health Information, including demographic information collected from an Individual, and that: (1) is created or received by a Health Care Provider, Health Plan, employer, or Health Care Clearinghouse; (2) relates to the past, present, or future physical or mental health or condition of an Individual; the provision of health care to an Individual; or the past present or future payment for the provision of health care to an Individual; and (3) a. identifies the Individual, or b. with respect to which there is a reasonable basis to believe the information can be used to identify the Individual. (45 CFR § 160.103)
3. Protected Health Information (PHI)
 - A subset of Individually Identifiable Health Information that is (a) transmitted by Electronic Media; (b) maintained in any medium constituting Electronic Media; or (c) transmitted or maintained in any other form or medium. (45 CFR §160.103)
 - Note: Information pertaining to a patient who has been deceased for more than 50 years is no longer Protected Health Information. Protected Health Information does not include Individually Identifiable Health Information in education records under FERPA or employment records held by a Covered Entity as an employer.

Again, REDCap users are expected to treat ALL health-related data (unless de-identified) as covered by HIPAA requirements—with the highest levels of privacy and security possible—*regardless of the source of the data*.

HIPAA Training

All REDCap users should complete HIPAA training, regardless of whether they reside in a “covered component,” and regardless of whether they will be working with high risk data.

HIPAA training is administered by individual institutions. Training is coordinated so that most people are only required to complete HIPAA training once per year. However, some users may be required to complete HIPAA training more than once in a year if their access requirements change.

Compliance Documentation

- If an UCH REDCap project is collecting research data involving human or animal subjects, the Principal Investigator or designated Project Administrator (e.g., lab manager, research coordinator) must acquire the appropriate compliance documents *prior* to collecting data.
- If the project is conducted at more than one institution, the project owner attests that appropriate regulatory approvals have been obtained prior to data collection.

De-Identification

There are two methods to de-identify Personal Health Information: the Safe Harbor Method (§164.514(b)(2)) and Expert Determination (§164.514(b)(1)).

If using the Safe Harbor Method, there are 18 pieces of information, or “identifiers,” linked to data that must be removed to consider data to be de-identified:

1. Names
2. All geographical identifiers smaller than a state, except for the initial three digits of a zip code if, according to the current publicly available data from the Bureau of the Census: the geographic unit formed by combining all zip codes with the same three initial digits contains more than 20,000 people; and the initial three digits of a zip code for all such geographic units containing 20,000 or fewer people is changed to 000
3. Dates (other than year) directly related to an individual
4. Phone numbers
5. Fax numbers
6. Email addresses
7. Social Security numbers (SSNs)
8. Medical record numbers (MRNs)
9. Health insurance beneficiary numbers
10. Account numbers
11. Certificate/license numbers
12. Vehicle identifiers and serial numbers, including license plate numbers
13. Device identifiers and serial numbers
14. Web Uniform Resource Locators (URLs)
15. Internet Protocol (IP) address numbers
16. Biometric identifiers, including finger, retinal, and voice prints
17. 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

However, UCH REDCap does NOT allow users to use or store any of the following sensitive data in the UCH REDCap system:

- Account numbers
- Social Security number
- Medical Record number
- Mother’s Maiden names
- Health Plan number
- Certificate/license numbers

- IP address
- Financial/PCI data
- Vehicle identifiers
- Biometric ID
- Full face/identifying photo
- Audio & Video Files

Data is considered de-identified according to this method once these 18 specific identifiers linked to an individual have been removed. In essence, de-identifying the data removes all information that could reasonably be used to re-identify an individual. Attention must be given to check that all 18 specific identifiers are removed, wherever those identifiers appear.

If using Expert Determination Method, you may use an expert with "appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable" to determine that there is a "very small" risk that the information, alone or in combination with other reasonably available information, could be used by the researcher to identify the Individual who is the subject of the information. The person certifying statistical de-identification must document the methods used as well as the result of the analysis that justifies the determination. The Institution must keep such certification, in written or electronic format, for at least six years from the date of its creation or the date when it was last in effect, whichever is later.

Anonymous Code Systems and Re-Identification

After de-identification of high risk data, data managers are permitted to use an anonymous code system, which assigns a code or other means of record identification to allow that information to be re-identified or linked to the dataset.

The mechanism for assigning codes and re-identifying records (the "key") **must NOT be:**

- Derived from any identifiers (e.g., using a participant's initials in lieu of their name)
- Used for any other purpose other than for re-identification
- Disclosed to others outside your group
- Stored on any machines, including those used for data collection/analysis

Limited Data Sets

A "limited data set" contains identifiers. A limited data set pertaining to health information is therefore always high risk data.

Identifiers that may remain in the information disclosed in a limited data set include:

- Dates such as admission, discharge, service, DOB, DOD
- City, state, five digit or more zip code
- Ages in years, months, days, or hours

Required Strategies to Maintain Security of Data in REDCap:

Minimum Necessary User Rights

- Learn more about user rights and roles on our User Rights and Privilege Access document

Flag identifiers

- When creating new fields, if your field label calls for identifying information (i.e., one of the 18 HIPAA identifiers), you must choose “Yes” next to Identifier - this is the ONLY mechanism the system has to ascertain if a field contains an identifier

Add New 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, Number, Date/Time, ...)

Field Label

Action Tags / Field Annotation (optional)

Learn about [@ Action Tags](#) or [using Field Annotation](#)

Variable Name (utilized in logic, calcs, and exports)

ONLY letters, numbers, and underscores

Enable auto naming of variable based upon its Field Label?

How to use [Smart Variables](#) [Piping](#)

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

Required?* No Yes
* Prompt if field is blank

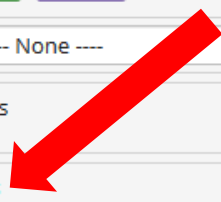
Identifier? No Yes
Does the field contain identifying information (e.g., name, SSN, address)?

Custom Alignment Right / Vertical (RV)

Align the position of the field on the page

Field Note (optional)

Small reminder text displayed underneath field



Restrict data export rights and ensure safe data export

- Restrict/limit Data Export user rights according to “minimum necessary” standard
 - For example, non-Illinois (i.e., external) collaborators **should not have** Data Export rights

Creating new role "External Collaborator"

Basic Rights

Highest level privileges:

- Project Design and Setup
- User Rights
- Data Access Groups

Privileges for data exports (including PDFs and API exports, reports, and stats):

- Data Exports No Access
 - * De-identified means that all free-form text fields will be removed, as well as any date/time fields and Identifier fields.
 - De-Identified*
 - Remove all tagged Identifier fields
 - Full Data Set
- Add / Edit Reports
 - Also allows user to view ALL reports (but not necessarily all data in the reports)

User rights table when “No Access” to Data Export is selected:

Project Design and Setup	User Rights	Data Access Groups	Data Export Tool	Reports & Report Builder	Calendar	Data Import Tool	Data Comparison Tool	Logging	File Repository	Record Locking Customization	Lock/Unlock Records	Data Quality (create/edit rules)	Data Quality (execute rules)	Create Records	Rename Records	Delete Records
✗	✗	✗	✗	✓	✓	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗	✗

- Export and/or transfer de-identified data according to “minimum necessary” standard
 - Ensure that “Remove all tagged identifier fields” is checked prior to exporting and/or transferring data

Utilizing Surveys in REDCap

REDCap has two online survey options, a private survey and a public survey.

- The **private survey** utilizes a participant's email address and REDCap sends a unique survey URL to each individual participant. Participants may only take the private survey one time.
- The **public survey** option involves a REDCap survey URL that can be posted on a website, emailed to a mailing list, etc.