University of Reading

Data management in REDCap

A good practice guide

Contents

Introduction

This guide to good practice in managing data collected from research participants is provided for users of REDCap, but the general guidelines and pseudonymous data processing principles can be applied to any research that involves the use of electronic tools to collect and process data.

REDCap is used to collect data from participants. Some research may collect sensitive information, for example about a person's health and medical history, or their sex life. Such information, where linked to an identifiable living person, constitutes special category personal data in data protection laws, and is Highly Restricted Information under the University's Classification Policy. Personal data that are not in themselves sensitive, such as a person's name and contact details, are subject to data protection laws and are classified as Restricted by the University. Restricted and Highly Restricted data must be processed in accordance with University policy.

Where personal data are collected by or on behalf of the University, both the University and those involved in processing the data have a legal responsibility to safeguard the information. The REDCap project owner must ensure that all members of a project team working on the direction or under the auspices of the University, including UoR staff, students, and non-UoR team members, are suitably trained in data protection and will act in compliance with all data protection laws while working on the project. These requirements are part of the Terms of Use that must be agreed to by all users of REDCap.

The risks involved in processing personal data and the potential for causing harm or distress by the inappropriate use or disclosure of such information must be understood by everyone involved in a project. REDCap provides the capability for multiple users, who may be University members or external to the University, to access personal data via a web browser from any device in any location, and to export the data to any other device or location. Sensible controls must therefore be applied.

General guidelines

Data management planning

Good practice in data management begins with planning. Every research project that involves the collection and use of research data should have a data management plan (DMP). A DMP is a structured document describing:

- what data will be collected or used in the course of a research project;
- how the data will be collected and managed on a day-to-day basis during the project;

- how relevant data will be managed on completion of the project, including plans for the retention and disposal of personal and confidential information, and for the preservation and sharing of de-identified research data in support of project findings.

The University provides information about data management planning tools and a REC DMP template, for use as part of a Research Ethics Committee submission. It is accompanied by guidance to help researchers plan for effective data management in compliance with legal and ethical obligations, throughout the lifetime of the research project and beyond.

Data management in REDCap

The following guidelines should be observed in all REDCap projects:

- REDCap should wherever possible be accessed only on the University campus or at an external user's place of work, via institutionally-maintained computing devices connected to the institutional network;
- If it is necessary to access REDCap from personal devices, these devices must have an appropriate level of security as directed by Digital Technology Services. Devices used to access Highly Restricted information must be encrypted;
- Mobile devices used to access REDCap must require password, code or biometric (fingerprint) authentication, and where possible be configured to enable remote wipe;
- REDCap should not be accessed in locations where unauthorised individuals may be able to view or intercept the information, and access via unsecured public wifi is not permitted;
- Local storage and caching of the information should be minimised as far as is practicable. Where it is necessary to export data to another location, this should be in a de-identified format wherever possible, and to a location with appropriate security;
- If identifiable information is collected in REDCap, all fields that will or are likely to contain direct identifiers must be tagged when they are created (see guidance below). The REDCap project owner should assign user rights allowing project users to access and export identifiable information only as needed for project purposes. For example, a statistician will not need to know the identities of individuals in a study or to export direct identifiers. User rights can be set to prevent access to specified data instruments (e.g. those containing identifiers) and to allow export of de-identified data only. Data Access Groups can be created to restrict access to records to subsets of the REDCap project team (e.g. where data collection takes place at multiple institutions, each institution might have its own Data Access Group, able to view only its own records).
- When a project is completed, and a final dataset is exported from REDCap for long-term preservation and sharing, relevant de-identification options should be used. These including removing tagged identifier fields, hashing record IDs (to de-

link them from participant information), removing free text fields, and removing or modifying date and datetime fields. The exported dataset must be carefully reviewed to ensure that it is fully de-identified.

Pseudonymous data processing in REDCap

While REDCap can be used to process identifiable participant information, data should be processed either anonymously (collecting no identifiable information, for example in an anonymous survey) or pseudonymously (using a non-identifying participant ID or key code) wherever this can be done without detriment to the project.

In most cases, it is likely that data in REDCap will be processed pseudonymously, using a linked key code to designate participants. Pseudonymous data that can be re-linked to identifiable participants are still personal data under data protection laws, so the guidelines outlined above apply. But separating direct identifiers from the key-coded data is a way of reducing the risk of causing adverse consequences should there be breach of security in REDCap.

Creating a pseudonymous key-coded dataset in REDCap involves the following:

- A separate spreadsheet or database, stored in a secure location and accessible only by individuals authorised on a need-to-know basis (such as the project PI and Study Manager), contains details of participants and their linked key codes;
- Key codes are non-identifying, typically being randomly-assigned numbers or character strings, such as REDCap record IDs;
- Participant records in REDCap are designated only by their key codes;
- No other direct identifiers are stored in REDCap (see below);
- The likelihood of being able to identify an individual by combining indirect identifiers stored in REDCap is very low;
- Use of free-text fields where users might enter identifiers is minimised.

Identifiers in REDCap

Identifiers in REDCap may be **direct identifiers**, **indirect identifiers** and **potential identifiers** (e.g. free text fields, especially when used in surveys, that might elicit identifiable information).

All direct identifier fields in REDCap instruments must be tagged as such when instruments are created. It is the responsibility of the PI to ensure this is done. In Add New Field and Edit Field there is an Identifier tag that can be selected.

Direct identifiers

Direct Identifiers include any of the following:

- Names (including first names, middle names and surnames, separately or in combination);

- Address;
- Full postcode;
- Phone number;
- Fax number;
- Mobile number;
- Email address;
- NHS number;
- National Insurance number;
- Other identifying reference numbers or codes, e.g. student ID, staff ID, passport number, driving licence number;
- Device attributes or serial numbers;
- Location data, including GPS data;
- Digital identifiers, such as IP addresses, and website URLs;
- Biometric elements, including finger, retinal, and voiceprints;
- Names of relatives, friends or colleagues;
- Photographs/video showing face or other distinguishing features.

Email addresses stored in REDCap for the purpose of emailing questionnaires to participants are direct identifiers. It is recommended that this method of sending surveys be used for low-risk data collection only. For projects that collect Highly Restricted information, for example relating to an individual's health or sexual activity, you are advised not to store email addresses in REDCap, and instead to send public survey links using a University email account.

Using the e-consent function to collect participant consent will also involve the storage in REDCap of direct identifiers. It is recommended that this function be used for low-risk data collection only. If used for studies involving the collection of Highly Restricted information, the ability of project members to access and export this information should be strictly controlled via user rights.

Indirect Identifiers

There are typically four types of indirect identifiers:

- Demographic data: e.g. marital status, gender, ethnicity, occupation, salary;
- Geographic data: most usually postcode, but may be any specific geographic location, e.g. GP surgery, school;
- Dates: e.g. date of birth, admission date;
- Medical identifiers: e.g. height, weight, unusual medical condition.

Two approaches can be taken to reduce the likelihood of re-identification using indirect identifiers:

- Decrease the granularity of the data you request: use ranges instead of exact values for items like salary, height or weight; record age at baseline instead of

date of birth; use month and year or year for other dates; record only the first three characters of a postcode;

- Store the indirect identifiers with the direct identifiers in a separate spreadsheet or database. Much of this information does not change over time and can be collected at the point of consent. Indirect identifiers that are required for the analysis can be imported into a statistical package when required. (Bear in mind that the resulting dataset would be considered identifiable and would need to be secured accordingly.)

Potential Identifiers

The use of fields that allow free text should be minimised. If REDCap is used to store information such as medical notes or adverse event reports, training and guidelines for the people entering these data must be provided, so that risks are minimised.

If free text fields are used to collect data directly from participants in surveys, each field should be risk assessed. If it is possible the user might enter identifiable data then a suitable warning, using the REDCap Field Note, must be added, and the field may need to be tagged as an Identifier. Content in free text fields not tagged as Identifiers should be reviewed, so that a field can be tagged as an Identifier if necessary.

User rights

When a new project is created by a REDCap Administrator, it is set up using a standard project template, with a set of user roles having pre-defined user rights. These roles are:

- **PI/Study Manager**: high-level user rights, including to copy projects, to design and set up projects, to modify project user rights, to view and edit all data, to create and rename records, and to export all data including identifiers;
- **Data Collector**: rights to view and edit data (these may vary between the different instruments created for a study, e.g. to restrict access to identifiable information), to create and rename records, and to export only de-identified data;
- **Data Analyst/Statistician**: rights to read only data (these may vary between study instruments, e.g. to restrict access to identifiable information), and to export only de-identified data.

The PI/Study Manager is able to modify the user rights for any users added to the project, and can modify the rights specifications for the default user roles, and create new user roles as required. The standard user roles described above have been designed to control data processing risks.

Where rights are granted to users, this must be done with caution, as permissions to view and export identifiable information carry risks relating to participant confidentiality and data protection, while permissions to edit and delete records may affect data integrity.

User rights for both individual users and user roles can be modified from the Project Setup tab on the Project Home page. Clicking on either the role or the user allows the associated rights to be edited. There are two sets of rights:

- Data entry rights can be set to **No Access**, **Read Only**, **View & Edit**, and **Edit Survey Responses**. These can be used to control access to instruments containing identifiable information, and can be specified for each instrument in a project. **No Access** should be used for users who do not require access to instruments containing identifiable information, e.g. demographics and consent instruments. Data entry rights only pertain to a user's ability to view or edit data on the web page; data export rights are set separately in Data Exports.
- Basic rights include highest level privileges (e.g. to modify other users' rights) that should be reserved for the PI/Study Manager only, data export rights (which should be set to de-identified by default), and various other privileges, some of which may enable users to access identifiable data, and should be granted with caution.

A detailed description of user rights is provided in the UoR REDCap Service Guide, which can be downloaded from the UoR REDCap web page.

Data Access Groups

The Data Access Groups feature provides another means of controlling access to identifiable information, and may be implemented to prevent disclosure of personal data outside the University where REDcap is used as part of a multi-site collaboration. This feature allows only users within a given Data Access Group to access records created by users within that group.

Data Access Groups can be created and users assigned to groups from the Project Setup tab on the Project Home page. Users do not have to be assigned to a Data Access Group. Unassigned users will be able to view all project records, so this privilege should be granted with caution.

Useful resources

Data protection for researchers. University of Reading IMPS office. https://www.reading.ac.uk/imps/data-protection/data-protection-and-research

Anonymisation. UK Data Service. https://ukdataservice.ac.uk/learning-hub/research-data-management/

Anonymisation: managing data protection risk. Code of practice. Information Commissioner's Office. https://ico.org.uk/media/1061/anonymisation-code.pdf

Good practice principles for sharing individual participant data from publicly funded clinical trials. Cancer Research UK, MRC, UKCRC Registered Clinical Trials Units, Wellcome Trust.

https://www.methodologyhubs.mrc.ac.uk/files/7114/3682/3831/Datasharingguidance2015.pdf

- Appendix 2: List of 28 potential patient identifiers in datasets;
- Appendix 3: Example Anonymisation Standard

Preparing raw clinical data for publication: guidance for journal editors, authors, and peer reviewers. Hrynaszkiewicz I et al. (2010). British Medical Journal 340:c181. https://doi.org/10.1136/bmj.c181