

# Queen's REB Guidelines on Research Data Retention and Data Repositories



Queen's  
UNIVERSITY

RE

# Queen's REB Guidelines on Research Data Retention and Data Repositories

**Version 1.0**

**07-Jun-2024**

© 2024, Queen's University

This publication is protected by copyright and can be used in accordance with the Creative Commons [CC BY-NC-SA](#) license. This license permits distributing, remixing, adapting, and building upon the material in any medium in any format for non-commercial purposes only as long as attribution is given to Queen's University. If the material is remixed, adapted, or built upon, the resulting material must be licensed in accordance with the Creative Commons CC By-NC-SA license or under identical terms.

Inquiries and permission requests for commercial use may be directed to:

Queen's University  
Vice-Principal Research  
Research, Compliance, Training, and Ethics  
355 King Street West  
Kingston, Ontario  
K7L 2X3  
[chair.greb@queensu.ca](mailto:chair.greb@queensu.ca), [hsreb@queensu.ca](mailto:hsreb@queensu.ca)  
[Research Compliance, Training and Ethics](#)

Queen's University is situated on [traditional Anishinaabe and Haudenosaunee Territory](#).

# Contents

Purpose .....	1
Background .....	1
Classification of data.....	1
Personally identifiable data .....	1
De-identified data.....	1
Anonymized data.....	2
Anonymous data .....	2
Indirectly identifying data.....	2
Research Data Storage and Retention .....	2
Queen’s University Senate Policy .....	3
Tri-agency Research Data Management (RDM) Policy .....	3
Health Canada Regulated Studies .....	3
US Regulated Studies.....	4
Research Data Repositories.....	4
Can all data be deposited in a research data repository? .....	4
How to find a research data repository .....	5

## Purpose

The purpose of this guideline is to:

- Provide clear guidance on the research data storage, retention policies and regulations that apply to human participant research and the use of research data repositories.

## Background

Queen's University (and affiliated hospital) researchers must comply with the research data retention policies that apply to their research data. Some of these policies are listed below. As part of the ethics application process, researchers must outline their data storage and retention plans; this will be reviewed and approved in conjunction with the study proposal itself by the appropriate research ethics board (REB). Research data repositories are digital repositories usually accessible via the internet that are specifically designed to store de-identified datasets.

## Classification of data

Study data can be classified through various methods, such as personally identifiable, de-identified, anonymized, and anonymous. These terms are not interchangeable.

### Personally identifiable data

Personally identifiable data is any information that allows for the identity of an individual (i.e., directly identifies an individual). For example: full name, date of birth, health card number, full postal code.



### De-identified data

De-identified data involves the removal of direct identifiers, but it may still be possible to re-identify individuals using indirect information or via a master linking log. For example, deleting or masking personal identifiers, such as participant name, with the use of a participant specific code.



## Anonymized data

Anonymization removes or alters identifiable information from datasets, ensuring that individuals cannot be directly identified or re-identified. For example, data that has phone number and name would be removed and it becomes impossible to re-identify the individuals.



## Anonymous data

Anonymous data refers to information that has never been associated with any identifiers. For example, filling out a survey online without providing any identifiers like email, phone number or name. Thus, there is absolutely no way to determine which response is associated with a particular individual.



## Indirectly identifying data

Indirectly identifying information is not enough to confirm someone's identity. However, when combined with other information, it could indirectly identify someone. For example, gender, age, geographical area, or additional information.



When designing a study, researchers need to determine what kind of data they are collecting from participants. Researchers should be aware that even if no directly identifiable information is being collected from participants it, may still be possible to determine a participant's identity if the potential participant pool is small and/or the information when collated can lead to identification.

## Research Data Storage and Retention

The length of time research data must be retained is determined by the type of research and the various policies and regulations that may apply. Many of these data retention policies and regulations outline the required minimums, as opposed to any maximums. As part of the ethics application process, researchers must outline their plans on data storage and retention, which will be reviewed and approved by the Research Ethics Board.

When determining how long research study data should be kept, researchers must determine which policies and regulations apply to their research. The sections below provide information about the data retention policies that most commonly apply to research data conducted at Queen's University and its affiliated hospitals. NOTE: There may be other policies that apply (e.g., school board, Tri-Agency, or other government agency policies) and the onus is on the researcher to determine if they are applicable to the research study. When multiple data retention policies apply to a research study, research study data should be retained per the policy that has the longest retention period.

## Queen's University Senate Policy

The Queen's Senate Policy on Integrity in Research (<https://www.queensu.ca/secretariat/policies/senate/integrity-research>) outlines the requirements and responsibilities of the Queen's community with respect to the conduct of research and scholarly activities in a manner consistent with the highest standards of ethical and scientific practice. Per this policy, researchers are required to retain complete and accurate research records, in a manner that will allow verification or replication of the work by others, within their personal control for a minimum of five (5) years from the date of publication or other form of presentation.

**Of note:** this policy does not specify where the data needs to be kept for the 5 years, however, it should be kept within the researcher's personal control. At a minimum, the data should be kept within a secure and encrypted Queen's OneDrive account to which the researcher has access. However, if the researcher should leave Queen's, they are responsible for ensuring the continued storage and access of this data (e.g., by establishing an appropriate data custodian to take over).

**For non-clinical trials:** Depositing the dataset into a research data repository would be a substitute for storing the dataset in a OneDrive account. If you can deposit your research data in a repository, you do **not** also have to keep that data in a separate location like OneDrive.

## Tri-agency Research Data Management (RDM) Policy

All research that is funded by Tri-agency funds must comply with the Tri-Agency Research Data Management Policy (2021) - <https://science.gc.ca/site/science/en/interagency-research-funding/policies-and-guidelines/research-data-management/tri-agency-research-data-management-policy>).

## Health Canada Regulated Studies

Health Canada Regulated Clinical trials must comply with Health Canada regulations which, since February 11, 2022, have required that clinical trials records for drugs and natural health products be retained for 15 years. The period for keeping records starts on the date the record is created. To simplify the process, sponsors may choose to "start the clock" for keeping all study records when the trial is completed or terminated.

## US Regulated Studies

US regulated studies, for any research that involved drugs, devices, or biologics being tested in humans, must retain records for a period of 2 years following the date a marketing application is approved for the drug for the indication for which it is being investigated; or, if no application is to be filed or if the application is not approved for such indication, until 2 years after the investigation is discontinued and the FDA is notified. Sponsors/researchers should receive written confirmation from the sponsor and/or FDA granting permission to destroy the records. (21CFR312.62.c).

According to FDA regulations, when a participant withdraws from a study, the data collected on the subject to the point of withdrawal remains part of the study database and may not be removed.

Research studies that are funded by US government agencies may also have specific data retention requirements. Researchers should check with the applicable agency to determine the data retention policy.

## Research Data Repositories

A research data repository is a digital repository, usually accessible via the internet, that is specifically designed to store research datasets (i.e., research data). Queen's typically recommends that researchers use Borealis, the Canadian Dataverse Repository (<https://borealisdata.ca/>), which is run by Scholars Portal and within which Queen's has a collection which it curates and administers (<https://borealisdata.ca/dataverse/queens>). In addition, all Canadian researchers have access to the Federated Research Data Repository (FRDR) (<https://www.frdr-dfdr.ca/repo/?locale=en>), which is run and maintained by the Digital Research Alliance of Canada. FRDR is curated by staff at the Alliance and is specifically designed to hold large datasets.

Research data repositories are designed such that researchers can share their data with others. This helps make verification and replication of research results easier and can help future researchers continue studies initially conducted by other researchers. They can also reduce the cost of research as data collection can be expensive and time consuming. Typically, research data repositories are not used as storage locations for data where only the researcher has access to the data.

Many journal publications now require that research data used to produce the results outlined in the publication be shared, with at least the journal editors or peer-reviewers and some require that the research data be available publicly to anyone who reads the publication.

## Can all data be deposited in a research data repository?

No. Some research data (e.g., confidential or sensitive data, copyrighted or trademarked data, licensed data, etc.) should not be deposited in a research data repository unless explicit permission has been provided. Non-health-related research data repositories do not have the level of security

and encryption required to store confidential or sensitive data. Data protected by intellectual property rights and/or legal agreements (i.e., licenses) may not be able to deposited or shared, even if the researcher had permission to view and use the data in the study.

Most de-identified or anonymized data collected by the researcher themselves, or data for which the identity of the participants cannot be determined, can be deposited into a research data repository. Identifiable data can potentially be stored in a research data repository, if the consent process explicitly asked for and received permission from participants to do so.

## How to find a research data repository

There are a number of ways a researcher can find a research data repository that will work for them.

1. They can contact the Research Data Management team at the Queen's Library for assistance (<https://library.queensu.ca/help-services/research-data-management>).
2. They can ask colleagues within their discipline what repositories they may have used in the past.
3. They can search for a repository on the Registry of Research Data Repositories (<https://www.re3data.org/>).