



University of
Southern
Queensland

Frequently Asked Questions: **Existing data sets**

Human research is research conducted with or about people, or their data or biospecimens. While some research involves the collection of new data, other research will use data that already exists in some form. Here are some frequently asked questions associated with the use of existing data sets:

1. What is an existing data set?

An existing data set is information, records, data or biospecimens that already exist prior to your proposed research project. The data may have been originally collected for a previous research project, collected for another purpose such as administrative, educational, clinical, service delivery, or organisational activities, or consist of user-generated content that already exists online, such as social media posts, comments, videos, blogs, discussion forum posts, or other publicly accessible digital content.

For the purposes of human research ethics review, an existing data set includes any existing human-related information that a researcher intends to access, collate, analyse, and reuse for a new research purpose. Examples of existing data sets include, but not limited to:

- student records, assessment items, or assessment results;
- administrative or organisational records, including quality assurance and evaluation activity data;
- data collected by a government body or organisation that is available for use by researchers upon access (e.g. the Australian Data Archive);
- data collected by other researchers held on open repositories (e.g. FigShare);
- data collect by you as part of a previous research project;
- clinical records or existing biospecimens; and
- social media posts or other publicly available online content.

2. Do I need to apply for ethical approval when using existing data sets?

Yes. The use of an existing data set for a new research purpose generally requires human research ethics review and approval. This is because the proposed use, analysis, publication, or dissemination of the data may differ from the purpose for which it was originally collected, regardless of whether the original collection underwent ethics review.

The ethics application allows the Human Research Ethics Committee (HREC) or non-HREC review body to assess the proposed use of the existing data set in the context of your proposed research project. This is particularly important where the data will be used, analysed, published, or disseminated in ways that differ from the purpose or conditions under which it was originally collected. The ethics application enables you to provide the context and information needed to determine whether your proposed use of the data set presents any new risk to the original participants and/or communities represented within the data set, or to the organisation that has provided you access to it.

In the ethics application, you will need to provide details about:

- a) how many records are in the data set;
- b) what information, records, data, or biospecimens the data set contains;
- c) whether the data was originally collected in an identifiable, re-identifiable or non-identifiable format, and any processes that will be used to remove identifiers prior to publication or dissemination;
- d) the original owner, source or custodian of the data set;
- e) whether ethical approval was obtained for the original collection and use of the data;
- f) who may have access to any identifiable information, who will be responsible for removing any identifiers from the data, and whether de-identification will occur before the data is provided to the research team or be undertaken by the research team as part of the project;
- g) evidence of permission to use the data set for the proposed research purpose (where required);
- h) how you intend to use the data within the proposed research project;
- i) whether the data or findings derived from the data will be published, disseminated, or otherwise shared;
- j) any explicit licensing agreements, terms and conditions, or restrictions governing the use of the data; and

- k) how consent was originally obtained, including whether participants agreed to future research use of the data, and any evidence of this.

3. Do I need permission from the original source to use data sets?

This will vary depending on the source of the data set and any terms and conditions that apply to its use. For example, some data may be publicly available online and freely available to download directly, although you might still need to agree to specific terms of use which must be recorded in your ethics application. Other data sets may require you to request permission from the data owner or custodian first, before they can be accessed or used.

Researchers are responsible for ensuring they have the appropriate authority to access and use the data for their proposed research purpose. Where permission is required, evidence of that permission must be included in the ethics application.

4. If data is open access, why do I need ethical approval to use it?

Open access does not necessarily mean unrestricted ethical use. As explained at Q2, human research ethics review considers how an existing data set will be used within the context of your proposed research project. This includes whether participants could reasonably expect their information to be used for research purposes, whether the proposed use creates any risks associated with analysis, publication, dissemination, or re-identification, and whether it is consistent with any permissions, conditions, or restrictions attached to the data. The ethics application provides a clear record of how, why and under what conditions the data will be accessed, managed and use throughout your project.

5. Why is social media considered an existing data set?

Social media is data that already exists online but is yet to be compiled into a usable data set. However, when people post online, they are not posting with the thought or intent of their posts being used for research purposes. Furthermore, some social media platforms may have their own requirements about how social media data is collected and reused. By treating social media as an existing data set, you will be able to document how the existing data is being collected and used, the terms and conditions of the social media platform, and your processes of deidentification. You may also wish to implement additional consent processes, such as contacting individual social media users to request permission to directly quote their posts, where this is practicable. For more information about social media and human research ethics, please see our [FAQ: Social Media Research and Ethics Approval](#).

6. What if the data doesn't yet exist but it will soon?

This is a common conundrum for projects intending to use future data sources, such as student assessment items or course evaluation surveys. Although these are not the only examples, they will be used here to help illustrate the key issues and considerations. When designing your research, it is important to consider whether consent can reasonably be obtained before data is collected. For example, if the project relates to a future trimester, could students be informed about the research component of the assessment or evaluation survey and be asked to consent to that data to be used for research purposes, or at least be given an opportunity to opt-out?

Where consent can reasonably be obtained, the project should be designed as direct participant-based research. In this case, students should be included as a participant group within the ethics application, rather than treating the future assessment or evaluation survey data as an existing data set.

However, there may be circumstances where direct participant recruitment is impracticable, making it unreasonable or impractical to obtain informed consent. In these cases, researchers should explain the circumstances within their ethics application and justify why the proposed data should be treated as an existing data set rather than collected through direct participant recruitment. As a general rule, where it is practical and reasonable to do so, researchers should seek to obtain consent and design their project in a way that maximises participants' opportunity to make an informed decision about the use of their data in research.

7. Where can I get help with my ethics application?

You can contact the [UniSQ Ethics Office](#) directly or drop into one of our weekly [HRE drop-in sessions](#). Additional resources and templates are available on the [Human Research Ethics webpage](#).