



# Ethics Guide to AusReddit

Version: 1.0

**Digital Observatory**

✉ [digitalobservatory@qut.edu.au](mailto:digitalobservatory@qut.edu.au)

**2 June 2026**

## 1 Ethics requirements for accessing AusReddit

Use of data in the AusReddit databank are subject to the National Statement on Ethical Conduct in Human Research (National Health and Medical Research Council (NHMRC), [2023](#)).

The Digital Observatory has applied for ethical clearance (granted by QUT's University Human Research Ethics Committee) to collect and maintain the AusReddit (Approval Number #8860, granted 30/08/2024 and effective until 30/08/2029).

Research projects with an Ethics exemption can only obtain aggregate data from the AusReddit. If researchers require access to raw data from the AusReddit for their research, **ethical clearance in accordance with the study's risk level is a requirement**. For most projects, low risk ethical clearance is sufficient. Importantly, the ethics application must cover usage of Reddit data. The Digital Observatory will require a copy of your ethical clearance before granting access to the databank.

If you are applying for low-risk ethical clearance, this guide provides information which might aid your ethics application.

## 2 Ethics issues

### 2.1 Data properties

---

Property	Details
Source	Existing datasets
Identifiability	Re-identifiable or potentially re-identifiable

---

### 2.2 Research participants and their involvement

Research participants are authors of Reddit posts and comments which are selected and included in the project's analysis. By the public nature of Reddit data, participants will not be directly contacted for recruitment and will not be asked to complete any study activity. Reddit data may contain the displayed usernames of the post authors - which are usually aliases, but no other identifying metadata are collected. Participants are not obliged nor coerced to be involved in the research as they will not need to be approached directly. Any relationship between the research team and any participant is purely coincidental.

### 2.2.1 Tips

- Be more specific about the selection criteria, e.g. only posts containing certain keywords, from a selection of subreddits, within a certain time period are included.
- In most cases and research projects, participants from vulnerable groups mentioned in the National Statement on Ethical Conduct in Human Research (Section 4) (National Health and Medical Research Council (NHMRC), 2023) are not specifically targeted. If this is true of your project, make an explicit statement.
- You may want to provide a rough estimate of the number of Reddit posts expected. The Digital Observatory can assist by conducting a preliminary volume check using your initial search queries.

## 2.3 Potential risks to research participants

There is the risk of identification, whereby some users have posted content (e.g. sexual orientation, membership of a minority, political beliefs) that explicitly self-identifies them as part of a potentially vulnerable population. This risk is, in general, low, based on the following grounds:

1. Users on open public platforms such as Reddit tend to use aliases rather than their real names.
2. Research has produced some evidence that people find use of their social media for academic research generally more acceptable than by journalists or social media companies (Hemphill, Schöpke-Gonzalez, & Panda, 2022).
3. It is reasonable to assume users understand the implications that posts that are visible to the public were intended to be so, especially in the case of Reddit, which makes clear to users the platform's use of their data<sup>1</sup>.

However, further measures will be taken to mitigate risks, namely:

- not to quote individual quotes,
- to report findings in aggregate forms,
- to avoid providing usernames of post authors,
- if quotes are required, to paraphrase and reword them to obscure/obfuscate identifiable details, and
- to explicitly mention and discuss ethical implications of using Reddit data in their studies.

---

<sup>1</sup><https://redditinc.com/policies/user-agreement>

### 2.3.1 Tips

- It is recommended that you adopt all the risk mitigating measures listed above. If this is not feasible for your research, only include in your ethics application those measures that apply.
- We recommend not publishing individual quotes or posts at all, as this may pose ethical challenges that require extra effort to address.

## 2.4 Waiver of Consent

A waiver of consent is requested for Reddit data for the following reasons:

1. Involvement in the research carries no more than low risk to participants. It is reasonable that the public nature of the users'/participants' contributions is clear to the users contributing to the platforms, since they can see contributions from all other participants publicly posting content to the platform. Specifically, in posting a thread or comment, users have agreed to Reddit's User Agreement, which states that the platform allows third parties to access public data on Reddit.
2. The benefits from such research justify any risks of harm associated with not seeking consent.
3. It is impracticable to obtain consent, due to the quantity of participants involved and the impracticability of contacting them via other means.
4. There is no known or likely reason for thinking that participants would not have consented if they had been asked, as discussed in Section 2.3.
5. There is an adequate plan to protect the confidentiality of data and the privacy of participants, as discussed in Section 2.3.

## References

- Hemphill, L, Schöpke-Gonzalez, A, & Panda, A. (2022). Comparative sensitivity of social media data and their acceptable use in research. *Scientific Data*, **9**, 643.
- National Health and Medical Research Council (NHMRC). (2023). National statement on ethical conduct in human research [Accessed: 10 September 2024]. *National Health Medical Research Council (NHMRC)*. <https://www.nhmrc.gov.au/about-us/publications/national-statement-ethical-conduct-human-research-2023>.