



# AusReddit: A guide for researchers

Version: 1.0.0

**Australia Digital  
Observatory**

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

**28 August 2024**

# 1 What is the AusReddit?

AusReddit is a research databank of Reddit posts and comments from Australian-related subreddits. As of 28 August 2024, the databank contains more than 4.9 million submissions and nearly 100 million comments from 593 subreddits.

## 1.1 How are subreddits identified to be “Australia-related”?

The list of Australian subreddits was created and iterated using various methods and sources. We began with known geography-specific subreddits such as r/brisbane, r/queensland, r/australia. More subreddits were added as they were identified, either from previous research or during the harvesting process of NewsTalk, Digital Observatory’s ongoing collection of user commentary on Australian news stories.

A subreddit and all its content are included in the databank if they meet the following criteria:

1. The subreddit’s topic explicitly mentions Australia, e.g. r/brisbane, r/queensland, r/australia, r/AusFinance.
2. The subreddit is about Australian entities, issues or organisations, e.g. r/AFL, r/NRL, r/InstantHotel.
3. Any subreddit known to have a substantial Australian community interacting, i.e. where many discussions about known Australian issues occur.

Please note that collecting data from these Australia-related subreddits is the only way we confine data to an Australian context. It does *not* mean the posts were posted in an Australian region or by an Australian user.

# 2 Who can use AusReddit?

AusReddit is available to academic researchers across Australia.

# 3 How do I use AusReddit?

There are three ways you can access and use AusReddit:

1. Via [ausreddit.digitalobservatory.net.au](https://ausreddit.digitalobservatory.net.au)
2. Via AusReddit public API
3. Via the Digital Observatory team, who will discuss your data needs and prepare datasets for you

### 3.1 AusReddit platform and API

The quickest way is to access AusReddit via the AusReddit website ([ausreddit.digitalobservatory.net.au](https://ausreddit.digitalobservatory.net.au)) and API. You will have to create an account and request databank access.

### 3.2 Working with the Digital Observatory team

If you need a guiding hand and a more tailored approach, you can contact the Digital Observatory. The DO team will discuss and provide advice on your data needs and other considerations. We also prepare datasets according to your specifications for you.

## 4 What do I need to get access?

Regardless of access pathways, you will need to provide sufficient information on your research project as well as an appropriate ethics approval and data management plan.

### 4.1 Ethics

Use of data in the AusReddit databank are subject to the National Statement on Ethical Conduct in Human Research.

The Digital Observatory has applied for ethical clearance (granted by QUT's University Human Research Ethics Committee) to collect and maintain the AusReddit. Research projects with an Ethics exemption can only access 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, but this depends on your institution's faculty or ethics committee.

In addition, researchers are advised to:

1. Treat all AusReddit data as sensitive and use mitigating strategies where appropriate,
2. Avoid quoting individual posts/comments,
3. Publish findings in aggregate forms,
4. If quotes are required, paraphrase and reword them to obscure and obfuscate identifiable details, and
5. Explicit mention and discuss limitations, potential bias, and ethical implications of using Reddit data, particularly AusReddit.

QUT researchers can find out more here: <https://www.qut.edu.au/research/why-qut/ethics-and-integrity> (QUT login required). The Digital Observatory will require a copy of your ethical clearance before granting access to the databank.

## **4.2 Data management plan**

A data management plan is recommended if you work with potentially personally identifiable data (e.g. raw submissions and comments) from the AusReddit. It might be a requirement in some cases, including HDR and several funded research projects, including ARC (Australian Research Council) and NHMRC (National Health and Medical Research Council).