



Australian Twittersphere: Factsheet

Version: 1.0

**Australia Digital
Observatory**

✉ digitalobservatory@qut.edu.au

12 September 2024

1 About the databank

The Australian Twittersphere is a longitudinal collection of tweets from a periodically updated list of Twitter accounts that were identified as Australian (i.e., have a stated connection to Australia in the free text fields of the account profile). Since its establishment six years ago, over **1.9 billion** tweets have been collected from more than 1 million Twitter accounts considered to be Australian. The earliest tweet captured was posted on 9 January 2007.

Following significant changes at Twitter/X and to its APIs, Digital Observatory's access to the API was revoked in early July 2023. The last tweet that was collected as part of the Australian Twittersphere was posted on 13 July 2023. However, some API instabilities were observed in early July. Therefore, **for research purposes, the Australian Twittersphere collection ends on 30 June 2023.**

1.1 Historical background

The Australian Twittersphere began in 2016 as the [Tracking Infrastructure for Social Media Analysis \(TrISMA\) project](#). Tweets were collected from approximately 530,000 Twitter accounts from early 2016 to May 2017.

In early 2018, the Digital Observatory was established and continued collecting tweets from the same accounts from March 2018 to December 2018 using Twitter's streaming API.

In late 2018, the Digital Observatory developed an additional collector which began collecting tweets from the same accounts using a number of Twitter's API endpoints.

In December 2020 and October 2021 the list of accounts was updated/refreshed to incorporate new accounts into the Australian Twittersphere. These two population refreshes ensured that the Australian Twittersphere remained up-to-date.

In October 2022, Twitter was sold to Elon Musk and API instability was observed throughout the remainder of 2022 and early 2023. However, the Digital Observatory was able to continue collection of the Australian Twittersphere. Throughout the first half of 2023, Twitter/X announced significant changes to their API. Specifically, the academic access track (which many researchers relied on) was removed, and Twitter/X began charging for API access.

Although the *Pro* level API access allows the collection of 1,000,000 million tweets per month at a cost of US\$5,000 per month, this would not have been sufficient to maintain the Aus-

tralian Twittersphere in its current form. In early July 2023, Twitter/X revoked the Digital Observatory's access to the API (as expected), and collection ceased.

The last tweet that was collected as part of the Australian Twittersphere was posted on 13 July 2023, however, some API instabilities were observed in early July and therefore, **for research purposes, the Australian Twittersphere collection ends on 30 June 2023.**

2 Collaboration process/How to use

Getting data from the Australian Twittersphere usually involves four steps:

1. Initial project discussion (1 hour)
 - Interested researchers can contact the Digital Observatory via email or submitting an enquiry form.
 - The Digital Observatory team will get in touch to schedule an initial meeting to discuss your project requirements.
2. Feasibility analysis (up to 1 week)
 - Preliminary investigations are done to determine whether there are enough relevant data in the Australian Twittersphere to address the research question.
 - This may consist of a volume check (basic counts of tweets matching specific search terms) and/or a feasibility analysis (descriptive analysis of tweets matching specific search terms).
3. Terms of Engagement (up to 1 week, longer for non-QUT researchers)
 - Researcher and the Digital Observatory agree on the project scope, deliverable(s), due date(s), and cost recovery (if applicable) by signing a non-contractual Terms of Engagement.
 - For external (non-QUT) researchers, a contractual QUT Services Agreement is to be signed by the researcher and the Digital Observatory.
4. Project delivery (basic data provision is typically up to one week, longer for aggregate data and analytical outputs)
 - Agreed deliverables are provided to the researcher in accordance with the Terms of Engagement.
 - This may include raw data from the Australian Twittersphere (typically in .csv format), aggregate data (e.g., descriptive statistics such as hashtag/keyword counts

over time), and/or analytical outputs (e.g., sentiment analysis, visualisations).

3 Ethics and Data management

Projects that use data from the Australian Twittersphere may require ethical clearance and/or a data management plan.

3.1 Ethical clearance

Data in the Australian Twittersphere are subject to the National Statement on Ethical Conduct in Human Research.

The Digital Observatory has ethical clearance (granted by QUT's University Human Research Ethics Committee) to collect and maintain the Australian Twittersphere. If researchers require access to raw tweet data from the Australian Twittersphere for their research, ethical clearance is a requirement. For most projects, low risk ethical clearance is sufficient but this depends on your institution's ethics committee.

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 (or exemption, if applicable) before providing access to raw tweet data.

3.2 Data management plan

The Digital Observatory recommends that a data management plan is written for the project. A data management plan is a live document that outlines how you will handle your data during and after your research. A data management plan is a requirement if your research is funded by the ARC or NHMRC, or you are a HDR student. QUT researchers can use the QUT Data Management Planning Tool and can contact their liaison librarian (QUT login required) for further assistance. External researchers should consult their own institutions to find out how to approach data management.

4 Frequently Asked Questions

4.1 Do I have to work at QUT to collaborate with the Digital Observatory and access the Australian Twittersphere?

Absolutely not! We work with researchers from many universities, as long as appropriate ethical clearance is in place (if applicable). Having a QUT collaborator does help, but it is not required. External researchers will be charged on a cost- recovery basis.

4.2 What geolocation data do you have?

In the Australian Twittersphere, there is a small amount of geolocation data (< 1% of tweets). This is because the majority of Twitter users opt against including geolocation information when posting tweets.

If your research question requires geolocation information, it might be possible to infer location information from the tweet text (e.g., the tweet mentions a specific location) or user profile information (e.g., the user indicates their location in their profile). However, these methods require additional work and are not suitable for all research questions, especially if high accuracy is important or fine-grained geolocation information is required.

Discuss your needs with us to see if there is a solution for your research question(s).

4.3 How long does it take to extract data from the Australian Twittersphere?

Extracting data from the Australian Twittersphere does not take long at all. However, the work of deciding what to extract requires some discussion and iteration, to ensure that the data will meet the needs of the research question(s).

4.4 How far back does the Australian Twittersphere capture?

The most reliable data goes back to March 2018, but some data are available before this time. There are other options for obtaining historical Twitter data (e.g., using Twitter/X's API at a cost), and we are happy to discuss this with you.

4.5 How much will it cost me?

The Digital Observatory operates to keep costs low for researchers by only charging on a cost-recovery basis. If you are a researcher within QUT, we also offer in-kind services for: - HDR (Higher-degree research) projects - Small projects - Projects in support of grant applications - ECR (Early-career researchers) projects

4.6 Do I need ethics approval?

It depends. If you are intending on publishing your findings or need access to raw tweet data (e.g., tweet text, account information) then you will require ethics. There are also other situations in which you will require ethics, such as research questions that require data about sensitive topics. Refer to the Ethics section above for more information.

If you are unsure, we can discuss your research question with you and can refer you to the Office of Research Ethics and Integrity or an ethics advisor for further advice if necessary.

4.7 Do I need a Data Management Plan?

A data management plan is recommended if you work with potentially personally identifiable data (e.g. raw tweets) from the Australian Twittersphere. 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).

4.8 Do I need any technical skills to collaborate with the Digital Observatory?

Absolutely not! We collaborate with researchers from a wide range of technical backgrounds, with different levels of technical know-how. We can tailor our output and deliverables to your specific needs and skills.