The Journal of Open Source Software

bskyr: An R Package to Interact with Bluesky Social

Christopher T. Kenny [1]

1 Princeton University

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor:

Submitted: 01 August 2025
Published: unpublished

License
Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

Summary

Bluesky Social is an offshoot of Twitter, designed to provide a decentralized alternative to traditional social media platforms. `bskyr` is an R package (R Core Team, 2025) that provides programmatic access to Bluesky Social. The package wraps the official Bluesky Social API, enabling users to retrieve posts, threads, user profiles, social graphs (e.g., follows, mutes, blocks), curated lists, and starter packs. It also supports content creation and interaction so that users can post, reply, like, repost, and follow directly from R. All data is returned in tidy data frames, making it compatible with standard R workflows for analysis and visualization. This allows users of `bskyr` to not only collect data for observational studies but also to engage with the platform programmatically, such as posting updates or interacting with other users.

Statement of Need

Following the acquisition of Twitter by Elon Musk in 2022, Bluesky Social launched a public beta in February 2023. Bluesky Social is built on the Authenticated Transfer (AT) Protocol, a decentralized social networking protocol that emphasizes user control, data portability, and interoperability. Bluesky has over 30 million users (The Bluesky Team, 2025). A similar package, Rtoot (Schoch & Chan, 2023), has been designed for an alternative decentralized platform Mastodon.

Bluesky's decentralized design makes it an appealing platform for researchers interested in social media data, as it allows for near complete access to public posts, user interactions, and social graphs without the restrictions often imposed by centralized platforms. In contrast, existing packages for Twitter, such as `academictwitteR` (Barrie & Ho, 2021) or `rtweet` (Kearney, 2019), have unfortunately become less reliable due to changes in Twitter's API access policies and the platform's evolving nature under new ownership.

Due to the decentralized nature of Bluesky Social, users can access a wide range of public data without the restrictions often imposed by centralized platforms. This is unlikely to be threatened by future changes to the platform, as Bluesky Social is open-source and designed to be accessible, with final control held by a public benefit corporation.

`bskyr` offers a consistent and user-friendly interface to nearly all public endpoints of Bluesky Social. It handles authentication, pagination, and API structure internally, allowing users to focus on data retrieval and interaction logic rather than protocol details. By supporting both reading data from and writing data to Bluesky, `bskyr` enables reproducible workflows that cover data collection, analysis, and programmatic engagement with the platform. This makes it a valuable tool for researchers, developers, and analysts working with social data in R.

Finally, `bskyr` offers flexibility in inputs and outputs. All core operations are supported with automatic tidying of data into familiar tidy formats from the `tidyverse` (Wickham et al., 2019). Users can opt out of cleaning returned data, allowing them to read in raw JSON as native R lists. Similarly, users can use `bskyr` to handle the uploads, but pass their own list objects.

DRAFT

The Journal of Open Source Software

Software design

The package attempts to balance tidiness with a light touch to avoid blocking future features as they become supported by the upstream API. The core goals of Bluesky and downstream clients differ greatly. Bluesky's overall service and underlying data are designed to support a complete social media service, while clients primarily download data or create data at a much smaller scale.

R packages on CRAN are limited to monthly updates, whereas upstream APIs can change frequently, so the package is designed to be robust to changes in the Bluesky Social API. Most, if not all, processing of data is optional, with an explicit opt-out via the `clean` argument. For the remainder, columns are processed based on their length and type. This avoids hard-coding any specific names or types that may change in the future. Of course, these can still break if a column's meaning is changed upstream, though such changes have been rare. This choice requires careful, manual updates any time the upstream API changes meaningfully, but will automatically work for any minor updates.

State of the field

When development started, `bskyr` was the only R package available for interacting with Bluesky Social. It has grown to be the most downloaded R package for using Bluesky Social. `atrrr` (Gruber et al., 2025), another R package, provides similar functionality for the primary posting and data collection features and is built with automatic parsing of Bluesky's lexicons. However, it only exports functions for a smaller set of endpoints than those supported by `bskyr`.

Programmatic access to the Bluesky Social API is available in several other languages. Notably, Bluesky Social provides its own official Typescript package and python has a popular AT Protocol SDK (Siamionau, 2025). Given the absence of any R alternatives at the beginning of development, `bskyr` was started to provide R users direct access to Bluesky Social. Additionally, studies of social media platforms are frequent in social science research, where the R language is most common. Further, developing an R package, rather than using packages from other languages, allows the use of `tidyverse`-style data frames with nested columns, which simplify this type of data analysis workflow in R.

Research impact statement

Bluesky Social is a relatively new social media platform, but has already seen significant adoption. Due to the decentralized design, researchers can use `bskyr` in support of a wide range of social media research, from running experiments to collecting data for observational studies.

`bskyr` is available on CRAN and has been downloaded at least 20,000 times since its release. Every function is fully documented and contains examples of function-by-function use. The package contains 3 vignettes which provide long-form examples of how to use the package.

The package itself is available under the MIT license. All development occurs on GitHub, where users can find latest releases, report issues, and contribute to the codebase. Tests are run after every commit on GitHub Actions and nightly on CRAN via testthat (Wickham, 2011).

Examples

At its simplest, `bskyr` formats data inputted by users into lists that can be passed to the Bluesky Social API. It also automates conversion from JSON to an R tibble when users want to read content from Bluesky. Below demonstrates the simplest workflow for authenticating, making a simple post, and then gathering data from a single endpoint.

DRAFT

Prior to making or collecting content, users must authenticate with Bluesky. To authenticate, all users need an account with a handle and an app password.

```
library(bskyr)
bs_auth(user = 'chriskenny.bsky.social', app_password = 'XXXX-XXXX-XXXX-XXX')
```

From there, users can utilize many of the functions in `bskyr` to interact with Bluesky Social. For example, users can post content to Bluesky Social using the `bs_post` function:

```
bs_post('Hello, Bluesky!')
```

Users can retrieve nearly all records from Bluesky Social. For example, users can get information on a specific user using the `bs_get_user` function:

```
bs_get_user('chriskenny.bsky.social')
```

AI usage disclosure

No generative AI tools were used in the development of the software or writing of this manuscript. ChatGPT 4o was used in a limited capacity to improve the coverage of the R package vignettes and all suggestions were manually reviewed.

References

Barrie, C., & Ho, J. C. (2021). academictwitteR: An r package to access the twitter academic research product track v2 API endpoint. *Journal of Open Source Software*, *6*(62), 3272. https://doi.org/10.21105/joss.03272

Gruber, J. B., Guinaudeau, B., & Votta, F. (2025). *Atrrr: Wrapper for the 'AT' protocol behind 'bluesky'*. https://doi.org/10.32614/CRAN.package.atrrr

Kearney, M. W. (2019). Rtweet: Collecting and analyzing twitter data. *Journal of Open Source Software*, *4*(42), 1829. https://doi.org/10.21105/joss.01829

R Core Team. (2025). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Schoch, D., & Chan, C. (2023). Rtoot: Collecting and analyzing mastodon data. *Mobile Media & Communication*. https://doi.org/10.1177/20501579231176678

Siamionau, I. (2025). *Atproto: The AT protocol SDK for python*. https://github.com/MarshalX/atproto

The Bluesky Team. (2025). *Bluesky user FAQ*. https://bsky.social/about/faq.

Wickham, H. (2011). Testthat: Get started with testing. *The R Journal*, *3*, 5–10. https://journal.r-project.org/archive/2011-1/RJournal_2011-1_Wickham.pdf

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., & others. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

DRAFT