

Bluesky Data: Collection, Cleaning, and Analysis

Tutor	Resul Umit (resul.umat@durham.ac.uk)
Organization	Digital Skills, University of Lucerne
Language	English
ECTS-Points	2
Contact	samuel.schmid@unilu.ch
Dates and time	17 October 2025 (10am – 4:30pm) 18 October 2025 (10am – 4:30pm) 24 October 2025 (10am – 4:30pm) 25 October 2025 (10am – 4:30pm)
Content	<p>The emergence of new social media platforms presents opportunities to explore uncharted data frontiers. Bluesky in particular is rapidly gaining a large and active user base while offering free access to its data, making it an invaluable resource for academics and professionals alike.</p> <p>Yet, leveraging social media data requires a specific set of skills. This course helps participants develop those skills through targeted knowledge, practical tools, and individual support for working effectively with Bluesky data.</p> <p>Through a hands-on introduction to collecting, preparing, and analysing Bluesky data in R, participants will build skills step-by-step, supported by guided exercises and one-to-one feedback on their own use cases. By the end of the course, they will be able to access and collect data from the Bluesky API, clean and prepare datasets for analysis, and apply specialised methods for studying social media content and networks.</p>
Preliminary Programme	Data Collection <ul style="list-style-type: none">• Setting up and authenticating access to the Bluesky API in R• Collecting posts and profiles

	<ul style="list-style-type: none"> Collecting other publicly available data <p>Data Cleaning</p> <ul style="list-style-type: none"> Working with lists and strings in R to process nested Bluesky data Cleaning and structuring posts and profiles for analysis Handling tags, mentions, emojis, links, and media <p>Data Analysis</p> <ul style="list-style-type: none"> Approaches to analysing Bluesky data at the profile, post, and combined levels Classification with dictionaries Classification with large language models Topic modelling Network analysis <p>Project Support</p> <ul style="list-style-type: none"> Peer and instructor feedback sessions on individual projects
Prerequisites/ Materials	<p>Participants should have basic knowledge of the R programming language, such as working with data frames and basic functions.</p> <p>Pre-course guidance and materials will be circulated beforehand to help participants get ready for the workshop. These will include instructions for installing and configuring required software.</p> <p>Assistance with technical setup will be available before and during the course to ensure that all participants can work smoothly with the Bluesky API and R on their own devices.</p>
Teaching Method	<p>The course is delivered through practical, hands-on activities. Short instructor-led demonstrations introduce concepts, functions, and methods, followed by guided exercises that allow participants to apply techniques immediately to real Bluesky data.</p>