

A data storage crisis: how UC Berkeley utilized Dataverse to respond to faculty data needs

maría a. matienzo

Head of Application Development Services
UC Berkeley Library IT

Storage Context

In April 2025, the Academic Senate Committee on Information Technology (CIT) wrote a letter to the Executive Vice Chancellor and Provost with the following proposals:

- **Increased storage:** baseline personal storage allocation of 5TB or greater.
- **Expandable storage:** simple way to pay to increase their allocations of storage
- **Longer-Term Preservation of Data Sets:** longer-term data storage and public availability based on funder and publisher mandates
- **Public Interest and Orphan Data Sets:** hierarchical ownership model that automatically transfers ownership, to maintain important data sets when individuals leave the university.

Find Your Fit - A Guide to UC Berkeley Data Storage

Is your data highly sensitive or subject to regulation?

For data, such as protected health information, that requires higher levels of data security and compliance.



Use SRDC

Built for working with and storing highly sensitive data, classified as P4 by UC Berkeley.

Do you need High-Performance Computing (HPC) integration?

For large scale simulation, modeling, or compute intensive data analysis.



Use Savio HPC

Optimized for high throughput computation and large, concurrent workloads.

Do you need to publish citable datasets with a DOI?

For sharing research data to meet FAIR principles and funder mandates.



Use Dataverse

Designed for archiving, publishing, and enhancing the discovery of research assets.

Is real-time collaboration with external users a top priority?

For projects that require easy file sharing and co-authoring with partners outside of UC Berkeley.



Use Berkeley Box

Ideal for file sharing, remote access, and integration with productivity tools.

None of the above needs apply?

For general storage and backup needs.



Use Cloudian

A scalable, object storage solution for longer-term storage and backup needs.

Library Service: Dataverse

The **role of Dataverse**: positioned as the solution for Data Sharing & Publishing to meet FAIR principles and funder mandates.

Use cases:

- Transitioning faculty away from self-hosting data to more sustainable, archived solutions
- Meeting funder and publisher requirements
- Incentivizing reproducibility and ensuring researchers receive academic credit through dataset citations
- Complements other repository offerings (ex. Dryad, ICPSR, etc.)

Future directions



Long-Term Planning:

Data Curator has been hired.

Transitioning the pilot into a sustained, ongoing campus service.

Developing formal guidance for data retention and archiving.

Summer survey

Thanks!

Many thanks to our colleagues in Berkeley IT and Research IT!

Contact info: research-storage@berkeley.edu

Library Team (in addition to maría):

Erin Foster: Research Data Management Program Service Lead

Carolyn Caizzi: AUL, Digital Information & Information Technology

Lynne Grigsby: Head of Library IT

Sam Teplitzky: Open Science Librarian

Anna Sackmann: Data Services Librarian

The problem...

What campus sees...

“UC Berkeley campus and higher education is at a moment of crisis with storage for research data. Cloud storage providers such as Google [and] Box...no longer offer unlimited storage. Their revised contracts will result in skyrocketing costs for UC Berkeley unless current levels of storage with those providers are reduced. Researchers, under the guidance of research support and IT staff, are scrambling to find new homes for their data, scholarly notes, collaborative files...”

excerpt from 2024 Research IT white paper

What a faculty member sees...

Re: ???!!!!??Re: YOUR GOOGLE ACCOUNT WILL BE FROZEN in 5 days unless you act now Inbox x

Library Service: Dataverse



Publish and share research data generated at UC Berkeley

[UC Berkeley Library Dataverse](#) > [UC Berkeley Faculty Research Data](#) >

Supplemental Material [Videos] -- 2D Active Bubbles

Version 1.0



Omar, Ahmad K.; Langford, Luke, 2026, "Supplemental Material [Videos] -- 2D Active Bubbles", <https://doi.org/10.60503/D3/L58J4P>, UC Berkeley Library Dataverse, V1

[Cite Dataset](#) ▼

Learn about [Data Citation Standards](#).

Library Service: Dataverse

Use Cases:

- Transitioning faculty away from self-hosting data to more sustainable, archived solutions
- Meeting funder requirements
- Incentivizing reproducibility and ensuring researchers receive academic credit through dataset citations
- Complements other repository offerings (ex. Dryad)

Technical details

Open Source Infrastructure: Dataverse is a robust open-source tool with a sizeable global community.

Upstream Contributions: Potential opportunities for UC staff developers to contribute back to the Dataverse upstream community; small contributions to date

Shared Practice: Aligning with UCLA library (and other peer institutions) already utilizing Dataverse