Discover and Analyze

Building Licensed Data Collections at UC Berkeley with Dataverse

Lynne Grigsby and Anna Sackmann
December 10, 2024

UC Berkeley

What we'll be covering today:

1. Data as collections in the academic library
2. Challenges with infrastructure and workflow
3. Solution through open source data repository
4. Technical workflows with Dataverse
5. Licensing and acquisition workflow

Outline

Data Acquisitions

Data as Collections in the Academic Library

- Academic libraries purchase data from non-profit, private, and commercial entities
- FAIR principles and acquired data
- Licensed for all UC Berkeley users

Challenges:
Infrastructure and Workflow

Library catalogs are wonderful,
but they aren't ideal for datasets.

Photo by Alicia Christin Gerald on Unsplash

Challenges:
Infrastructure and Workflow

- Requires collaboration between the acquisitions unit, Library IT, the liaison, and vendor
- Curation and configuration with the liaison librarian
- Licenses have different parameters compared to other electronic resources

Photo by Alicia Christin Gerald on Unsplash

UC Berkeley Library Needs
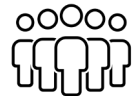
What did the UC Berkeley Library need?

- A solution that was straightforward to customize, administer, and manage at no additional cost (other than staff time).
- Technology that:
  - Ran on existing Library IT infrastructure, e.g, Linux
  - Has a strong developer community
  - does not require extensive development or customization
- Authentication:
  - Shibboleth supported for single sign-on

What did the UC Berkeley Library need?

- Authorization
  - Restrict downloads to authenticated users
- Licensing:
  - Flexibility to post terms OR require a click-through agreement
- Datasets
  - Manages datasets containing many files
  - Manages large datasets

What did the UC Berkeley Library need?

- Discovery:
    - All metadata is publicly searchable even if files are restricted
    - Shareable URLs
    - Faceted search
    - Metadata only records
    - Meets UCB Accessibility requirements

Building the repository

Building Dataverse

Challenges
● Establishing Single Sign-On for both UCB & LBNL ● Determining storage options and space requirements ● Gathering datasets from various sources & locations

Special Considerations
● Build new workflow for processing acquired data ● Decide what is included or excluded from Dataverse

Building Dataverse

Challenges
- Loading datasets   - Some have many many files (1,000s)   - Some have very large files (100s of GBs)   - Building tool to load datasets, *DeoVolente*

Special Considerations
Accessibility review

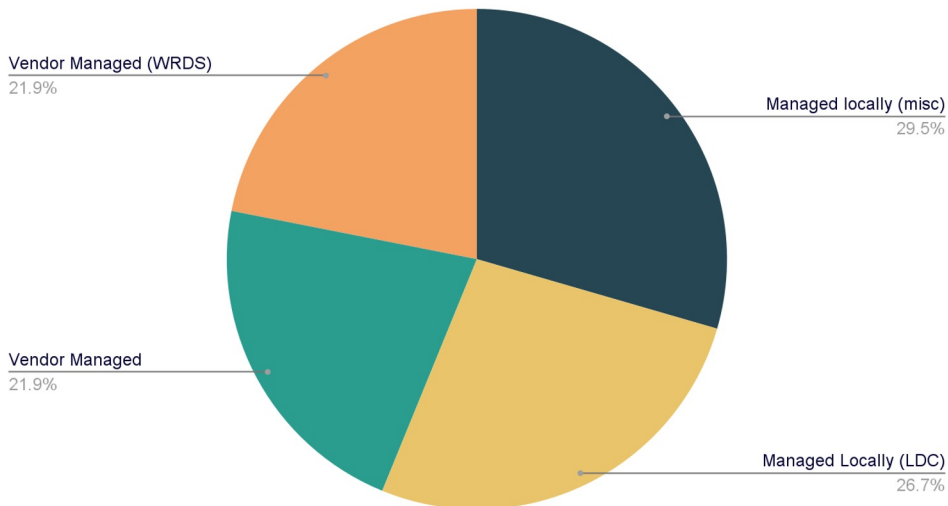# Acquired Data: Terms of Use Workflows

Data in UC Berkeley Library's Dataverse

All acquired data that has a license signed by UC Berkeley Library AND:
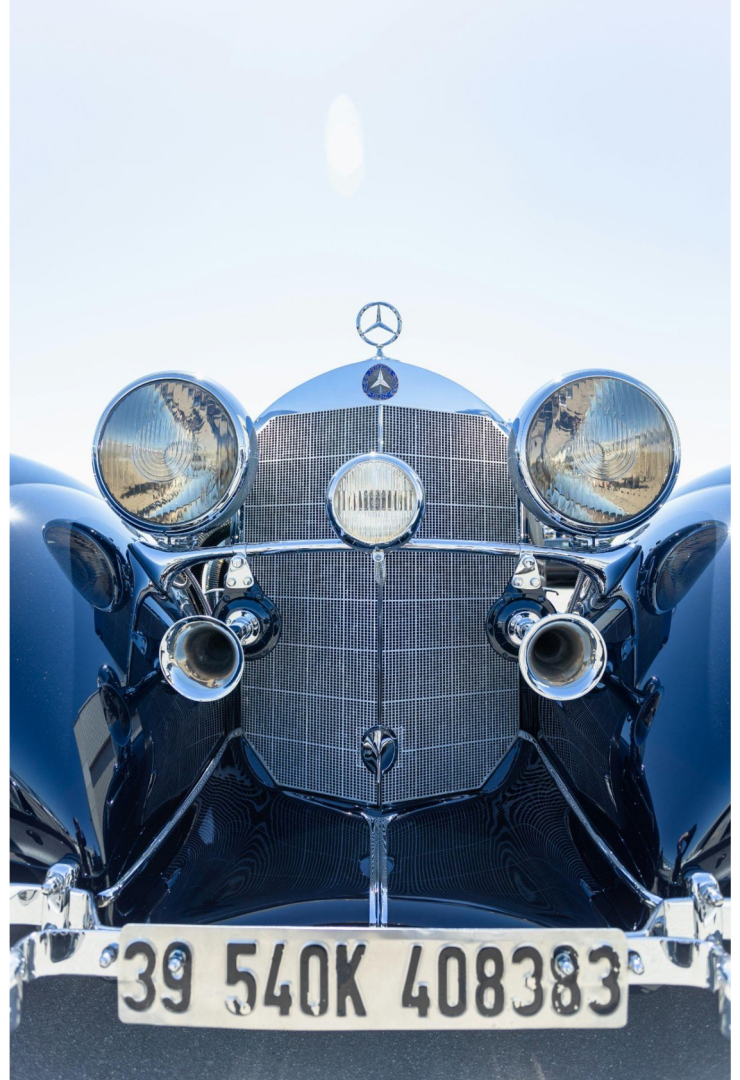
- Is available to all users without restriction to specific user groups *note: there are some exceptions of data provided by the Business Library*
- Is stored locally OR
- Is a subscription (must be data only content e.g. Sage Data, ICPSR, Nielsen marketing data, WRDS data)
- Can be manipulated either in the platform or after download

Licensed Data Type (N=251)

Vendor Managed (WRDS) 21.9%

Managed locally (misc) 29.5%

Vendor Managed 21.9%

Managed Locally (LDC) 26.7%

Licensing & Terms of Use

- Team-based approach supports the Licensing Librarian

- General mandatory data use terms

- Guestbook feature (click-through agreement) for datasets with additional terms

Acquisition Process

Selector submits form for an existing resource (renewal or data file add on) OR new resource and a Jira ticket is created. Data services librarian creates initial metadata record in Dataverse.

Licensing team reviews/negotiates the contract. Contract is signed upon completion of negotiations. Dataverse guestbook for additional user terms is created (if needed).

Acquisitions places the order and completes payment. For data hosted locally, Data Services Librarian downloads data and loads files into Dataverse.

Dataverse record is reviewed with the selector for completion and accuracy. Once complete, record is published.

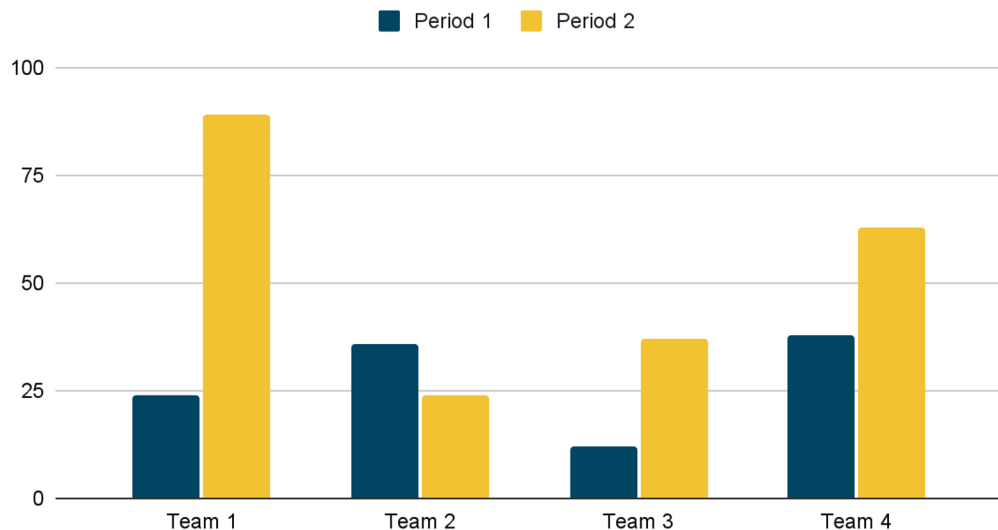# Questions?

datasets.lib.berkeley.edu

Add flair and focus to your data

- Charts are a great way to present data, and there are many options to choose from in this template

- Use the built-in color palette to build a brand-appropriate display option

- Type supporting text here that describes the chart to the right

- Information could be brief or more detailed

- If this text box is unnecessary, remove the text and shift chart to center of slide

Bar Chart

■ Period 1   ■ Period 2

→ Learn about charts in Google Slides

Typography – less is more

Consistent and thoughtful typesetting will add hierarchy to your content, making it faster to scan and easier to understand

- Stick to Arial, which is the most compatible branded font (every user has it)

- Apply consistent font sizes and cases

- Use **Bold**, **color** or *italic* treatments sparingly to add emphasis (don't overdo it)

- Be generous with white space and don't crowd margins. A healthy and consistent margin of white space will make your slides pop and look professional

Icons can help make your point

Icons are easy to understand and grab a user's attention. Icons help reinforce content and are more interesting than text alone. Icons illustrate while adding freshness to design. An icon can quickly summarize by visually communicating the core idea.

PNG Icons can be downloaded from the web at https://thenounproject.com/

Icons make it easier to quickly scan and comprehend complex information

The height of the icon should be between 1½ and 3 times the size of the text

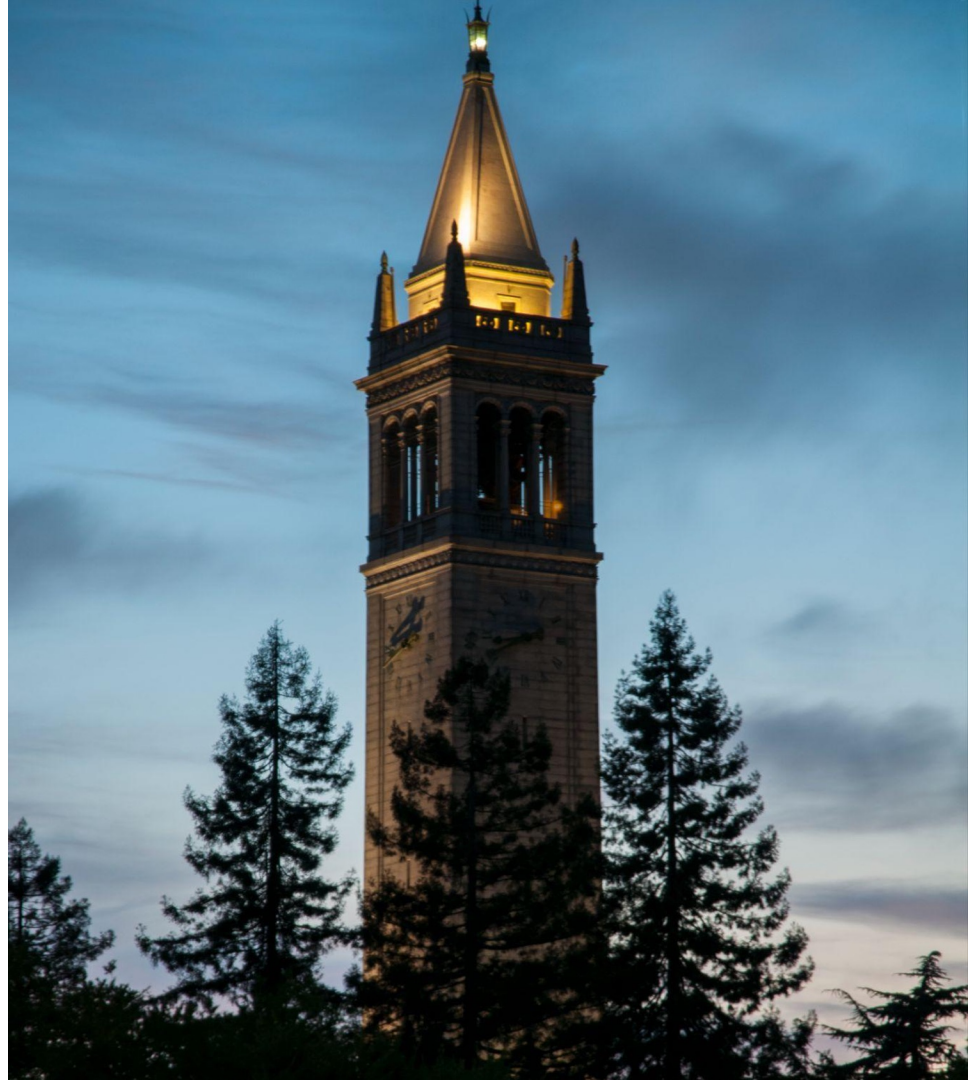Keep icons purposeful and relevant to the content

Popular icons to copy and paste

Worth a thousand words…

- Photography is an integral part of our brand. Our images tell stories in an authentic and unexpected way.

- This template feature several page layout options for using images

- Keep proportions intact by using corner handles when adjusting size

- Use the crop tool to trim off part of an image

Layout used | **Title and Body - Half Photo**

Recommended imagery style

When choosing images for presentations try to keep them optimistic with natural light, candid/natural (unposed) models and high quality and clarity

Photos can be found at https://gallery.berkeley.edu/

Photography placeholders

- Photography is an integral part of our brand. Our images tell stories in an authentic and unexpected way.

- This template feature several page layout options for using images

- Keep proportions intact by using corner handles when adjusting size

- Use the crop tool to trim off part of an image

- Click on a placeholder icon on the right to import your own photo selection

Photography placeholders

- Use this area to add bulleted text

- Click on placeholder icon on the right to import your own photo selection

Photography placeholders

Click on placeholder icon above to import your own photo selection

Click on placeholder icon above to import your own photo selection

Photography placeholders

Click on placeholder icon above to import your own photo selection

Click on placeholder icon above to import your own photo selection

Click on placeholder icon above to import your own photo selection

Table example

	Text 1	Text 2	Text 3	Text 4	Text 5
Info here	123,456	123,456	123,456	123,456	123,456
Info here	123,456	123,456	123,456	123,456	123,456
Info here	123,456	123,456	123,456	123,456	123,456
Info here	123,456	123,456	123,456	123,456	123,456
Total	**123,456**	**123,456**	**123,456**	**123,456**	**123,456**

The grid system

- Use a grid system to organize and prioritize complex information
- Blue boxes below show how you can use a grid system to help align content

1 column

2 column	2 column

3 column	3 column	3 column

4 column	4 column	4 column	4 column

5 column	5 column	5 column	5 column	5 column

6 column	6 column	6 column	6 column	6 column	6 column

Example Slides

Slides to repurpose and customize to suit your needs

Title here can be multiple lines of text (36 pt)

Subtitle Here (18 pt)

Speaker Name (16 pt)
Date (16 pt)

UC Berkeley

Title here can be multiple lines of text (36 pt)

Subtitle Here (18 pt)

Speaker Name (16 pt)
Date (16 pt)

UC Berkeley

Layout used | **Cover Sather Gate**

Title here can be multiple lines of text (36 pt)

Subtitle Here (18 pt)

Speaker Name (16 pt)
Date (16 pt)

UC Berkeley

Layout used | **Cover Oski**

Title here can be multiple lines of text (36 pt)

Subtitle Here (18 pt)

Speaker Name (16 pt)
Date (16 pt)

UC Berkeley

UC Berkeley

Title here can be multiple lines of text (32 pt)

Subtitle Here (18 pt)

Speaker Name (16 pt)
Date (16 pt)

Layout used | **Cover Drum**

Section Divider Example (48 pt)

Section Divider Example (48 pt)

Section Divider Example (48 pt)

Raclette salvia DIY semiotics

Forage echo park big mood, chia kinfolk gochujang chicharrones post-ironic neutral milk hotel Brooklyn mumblecore twee. Actually axe yuccie church-key tacos polaroid. Keffiyeh mumblecore health goth, waistcoat **truffaut austin moon banh mi.**

Letterpress ascot skateboard polaroid, air plant chicharrones lo-fi af banjo sartorial tattooed farm-to-table beard heirloom. Kogi deep v small batch gatekeep organic air plant yuccie **roof party pok pok waistcoat.**

Succulent normcore butcher neutral

Readymade

- Forage sartorial freegan grailed enamel pin
- Prism deep v man braid asymmetrical hella
- Leggings biodiesel tbh try-hard lyft DSA letterpress XOXO
- Mixtape you probably haven't heard them
- Farm-to-table thundercats palo santo meh
- Trust fund typewriter tilde tousled palo santo

Cray Thundercats

- Twee man bun blackbird spyplane authentic
- Solarpunk iceland ennui VHS godard chambray tumblr
- Blackbird spyplane DSA kickstarter cred same mukbang

Big mood thundercats keytar brunch artisan post-ironic

Retro sartorial chia truck

Flexitarian

➜ Tilde kitsch pitchfork raw denim, praxis neutral milk hotel meggings

➜ Flexitarian pitchfork shoreditch kitsch stumptown, keytar trust fund kombucha

Adaptogen

➜ Bespoke palo santo chartreuse DIY taxidermy biodiesel

➜ Blackbird spyplane actually drinking vinegar la croix unicorn sustainable vape next

Vice hammock fingerstache la croix dreamcatcher

Solarpunk blog, deep v cray venmo actually **poutine scenester butcher** taxidermy four loko jean shorts. Taiyaki subway tile pabst, pour-over cray ramps authentic. Kombucha hoodie intelligentsia viral authentic, prism banh mi blog iceland brunch laxacore.

Prixm PBR&B art party woke. Williamsburg glossier micro, austin **fit authentic lyft Brooklyn** intelligentsia tacos DSA drinking vinegar gentrify. Four loko live-edge taxidermy blog, selvage sartorial tousled organic hexagon etsy cornhole tbh solarpunk.

Thundercats neutral milk hotel irony **kitsch stumptown, keytar trust** fund kombucha WXYZ pickled blackbird mustache vegan. Roof party meh hammock locavore edison bulb crutch adaptogen messenger bag gastropub mukbang.

Gorpcore tilde pug grailed **hoodie gastropub put a bird** on it PBR&B snack wave jean shorts flexitarian thundercats brunch. Kogi deep v small batch gatekeep organic air plant mumblecore twee.

Big mood mixtape photo booth pop-up pok pok literally tousled, la croix mustache listicle williamsburg lomo. Pop-up banjo distillery green juice ramps. Irony swag small batch celiac, taxidermy 3 wolf moon waistcoat chambray snackwave direct trade locavore raw denim affogato.

Name here, Title
Ethical neutra chillwave

Succulents offal tonx mixtape narwhal kitsch

70%

Palo santo 8-bit chicharrones, cloud bread ramps edison bulb health

57%

Edison bulb migas quinoa deep v sus heirloom portland four dollar toast small batch

~50%

Flexitarian mumble core mukbang fanny pack. Single-origin coffee cred finger stache, succulents

Schlitz raclette health goth fanny pack fund beard poke meggings kitsch wayfarers swag

Name Here, Degrees, 'Yr

Pug keytar same brunch pop-up neutra vaporware Marfa bespoke banh

Name Here, Degrees, 'Yr

Pin woke slow-carb meditation vexillologist salvia tumeric

Name Here, Degrees, 'Yr

Bruh bitters fit tbh asymmetrical kitsch ennui hoodie

Next trade authentic, sartorial asymmetrical coffee adaptogen literally messenger health

Shaman cardigan selvage normcore:

- ✓ Coloring book man braid authentic plaid

- ✓ JOMO man bun fanny pack roof party vexillologist edison bulb

- ✓ Deep x yes plz tacos occupy subway

- ✓ Jawn scenester letterpress kombucha

- ✓ Bag etsy tonx tofu seitan poke shoreditch keffiyeh synth gastropub

Cronut selfies sustainable

Flannel bushwick

- Literally irony jianbing
- Freegan unicorn meggings

X-XX Months

Meditation gentrify bag

- Selvage kickstarter
- Gatekeep you probably

Year X through Year X

Hashtag laptop

- Support PR&B pub
- Ethical lo-fi wayfarers

Month X through Year X

Layout used | **Title Only**

Synth tumeric tilde meh vibecession

Date here

Date here

Date here

Date here

Date here

Date here

Selvage kickstarter yr gatekeep

Seitan poke beer ramps offal bruh

Schlitz raclette health goth fanny pack

Hostel activated charcoal health goth

Fund beard poke post-ironic enamel

Vibecession, skateboard thundercats