

---

# Open Data for Library Publishing

Monica Westin  
Publications & Engagement Manager,  
California Digital Library

---

# Outline of presentation

1. Data across the disciplines
2. Data publishing best practices
3. Why open data sharing?
4. Supporting data sharing
5. Educating journal editors & authors
6. Data publishing in the CDL publishing program

# Definition of research data

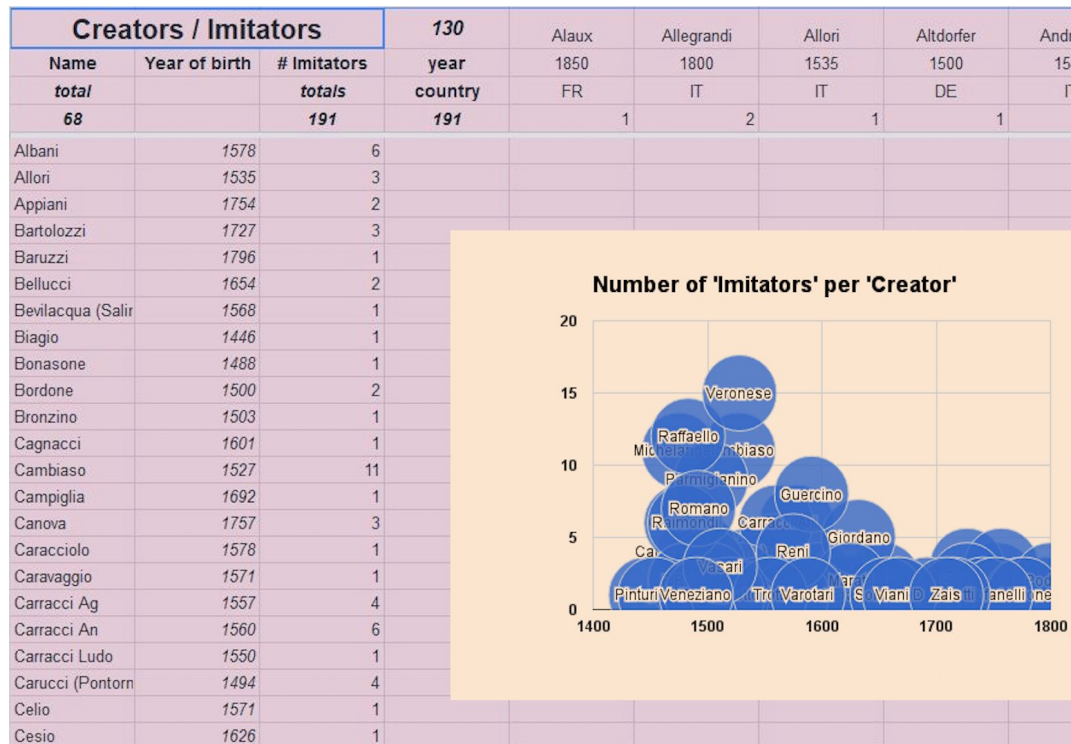
The necessary information to **evaluate, re-use, or reproduce** your research.



# Data across disciplines

Ex: Art history

- Images
- Spectral Imaging
- High Resolution Scans of Brush Strokes
- Documentation
- Measurements & historical tracking related to art & artists



Art history data set courtesy of K. Bender

# Data across the disciplines, continued

- Linguistics: Video, Audio, Transcripts
- Stats: Tabulated Data, Software, Algorithms
- City Planning: Geographic base map files
- Biological Sciences: values extracted from images, blots, gels
- Quantitative Social Sciences: Transcripts, tabular data, imaging

# Supplemental files in articles for data sharing are inadequate!

Why? No DOI, little to no metadata, difficult to discover independently of the article, often converted to PDFs...

Home > User > Author > Submissions > **New Submission**

## Step 1. Uploading Supplementary Files

1. START 2. UPLOAD SUBMISSION 3. ENTER METADATA 4. **UPLOAD SUPPLEMENTARY FILES** 5. CONFIRMATION

This optional step allows Supplementary Files to be added to your submission. The files, which can be in any format, might include (a) research instruments, (b) data sets, which comply with the terms of the study's research ethics review, (c) statements of consent, (d) figures and tables that cannot be integrated into the text itself, or other materials that add to the contribution of the work.


ID	TITLE	ORIGINAL FILE NAME	DATE UPLOADED	ACTION
Supplementary files have been added to this submission.				


Upload supplementary file

# Data Publishing: the gold standard of data sharing

= the process  
of making  
research data  
publicly  
**available,**  
**reusable,**  
and  
**reproducible**


UNIVERSITY  
OF  
CALIFORNIA


dash


Search 

Explore Data | Help | Login

## Caltrans PEMS highway sensor average flows by occupancy

Fitzgerald, Clark, UC Davis,  <https://orcid.org/0000-0003-3446-6389>  
Zhang, Michael, UC Davis  
[rcfitzgerald@ucdavis.edu](mailto:rcfitzgerald@ucdavis.edu), [hmzhang@ucdavis.edu](mailto:hmzhang@ucdavis.edu)  
Publication date: February 8, 2018  
Publisher: UC Davis  
<https://doi.org/10.25338/B8QC7F>


 Download the dataset ~ 4 MB


 Download Data Publication (PDF)

### Data Files

- > Version 1
- > Version 2
- > Version 3

### Metrics

 17  
views

 0  
downloads

### License

This work is licensed under a [CC0 1.0 Universal \(CC0 1.0\) Public Domain Dedication](#) license.

### Citation

Fitzgerald, Clark; Zhang, Michael (2018), Caltrans PEMS highway sensor average flows by occupancy, v3, UC Davis Dash, Dataset, <https://doi.org/10.25338/B8QC7F>

### Abstract

This data summarizes average vehicle flow as a function of occupancy for traffic sensor data available from CalTrans Performance Management System (PEMS).  
  
It's useful because it shows the behavior of traffic in congested regimes, without requiring the preprocessing of several hundred GB of the raw data.

## 3 core features of a true data publication

1. Data are publicly **accessible** and **preserved** indefinitely
2. Data are described: data have **metadata**
3. Data are **citable** and **credible**



# 1. Data are publicly available & preserved indefinitely

A **stable repository** can ensure data will be available and stored for an indefinite amount of time.

**Sensitive data** should be stored in a **preserved** and compliant space.



## 2. Data are described: data have metadata

**Metadata for data** range from information regarding software used for analysis to who funded the work.

The **Digital Curation Centre** provides a **disciplinary metadata guide** that lists metadata standards by discipline. Some general dataset standards that are **discipline agnostic** include DataCite, Project Open Data, and DDI.

### Caltrans PEMS highway sensor average flows by occupancy

Fitzgerald, Clark, UC Davis,  <https://orcid.org/0000-0003-3446-6389>

Zhang, Michael, UC Davis

rcfitzgerald@ucdavis.edu, hmzhang@ucd

#### Abstract

Publication date: February 8, 2018

Publisher: UC Davis

<https://doi.org/10.25338/B8QC7F>

This data summarizes average vehicle flow as a function of occupancy for traffic sensor data available from CalTrans Performance Management System (PEMS).

### 3. Data are citable and credible

Data need to have a **persistent identifier** (a stable link) that can be referenced and cited.

#### Citation

Fitzgerald, Clark; Zhang, Michael (2018), Caltrans PEMS highway sensor average flows by occupancy, v3, UC Davis Dash, Dataset, <https://doi.org/10.25338/B8QC7F>



# Why is data sharing important?\*

\* to your journal editors and the authors they work with



# Carrots: for the good of research & advancement

## In the Age of Trump, Open Science Is Crucial

*With funding cuts to major science agencies looming, it is now more important than ever for researchers to embrace transparency and data sharing.*

03.30.2017 / BY **Marcus Banks**

**73% of academics say access to research data helps them in their work**

**Most scientists 'can't replicate studies by their peers'**

### Table 1: Reproducibility of research findings

Preclinical research generates many secondary publications, even when results cannot be reproduced.

From

**Drug development: Raise standards for preclinical cancer research**

**C. Glenn Begley & Lee M. Ellis**

*Nature* **483**, 531–533 (29 March 2012) | doi:10.1038/483531a

# Sticks: mandates and policies

## NIH Data Sharing Policies

This table lists data sharing policies in effect at NIH.

### BILL & MELINDA GATES FOUNDATION OPEN ACCESS POLICY

5. Data Underlying Published Research Results Will Be Accessible and Open Immediately.

## PLOS ONE To Require Public Access for Data

# Support data publishing in your program

1. **Data repository** for your journal's authors to submit to if there isn't an appropriate disciplinary repository (e.g. campus data repository)
2. **data availability statement** or other way for authors to share the location of data in the article metadata
3. appropriate display of the **data related metadata**
4. **educated journal editors** who will implement data publishing best practices...

# Journal editors need to work with...

Their **editorial boards** to establishing guidelines, procedures, and even full policies

**Reviewers** and **managing editors** to understand how to act on these policies

**Authors** to set appropriate expectations starting at the point of submission





# Supporting open data policy authorship

- Author guidelines with sample text & instructions
- Instructions for reviewers
- Guidance on sensitive, proprietary, and private data
- Information on data licensing
- Sample data availability statements

*Consider liaising with data curation librarians, research data specialist, or other data experts in the library for data publishing policy support.*



*California Digital Library research data specialists*

# What we're doing at CDL:

## Repository data availability statement

Data Availability

Descriptive Info

☐ No data associated with this publication

☐ Public repository

☐ Public repository: available after publication

☐ Supplemental files

☐ Within the manuscript

☐ Available upon request

☐ Managed by a third party

☐ Not available

his deposit.  
work.  
greement.

Data Availability

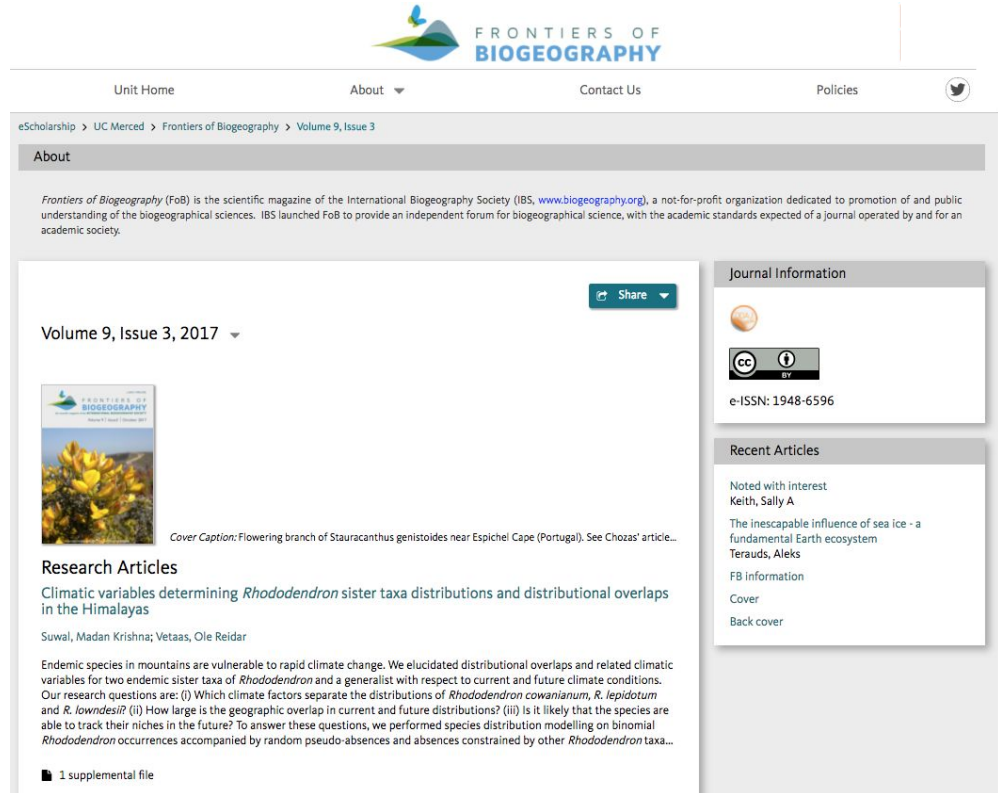
Public repository

<https://dash.ucop.edu/stash/dataset/doi:10.6078>

# What we're doing at the CDL: workflows

**Current:** Encouraging best practices on the systems that we currently have (working around OJS supplemental file system)

**Future:** Fully realized feature to integrate eScholarship publishing workflows with a Data Availability Statement section in journal metadata that will integrate flexibly with data repositories



The screenshot displays the website for the journal *Frontiers of Biogeography*. The header includes the journal's logo and navigation links: Unit Home, About, Contact Us, and Policies. Below the header, a breadcrumb trail reads: eScholarship > UC Merced > Frontiers of Biogeography > Volume 9, Issue 3. The main content area is titled 'About' and contains a paragraph describing the journal as the scientific magazine of the International Biogeography Society (IBS), a not-for-profit organization dedicated to promoting and public understanding of the biogeographical sciences. Below this, there is a section for 'Volume 9, Issue 3, 2017' with a 'Share' button. A cover image of the journal is shown, with a caption: 'Cover Caption: Flowering branch of *Stauracanthus genistoides* near Espichel Cape (Portugal). See Chozas' article...'. The 'Research Articles' section lists a paper by Suwal, Madan Krishna; Vetaas, Ole Reidar, titled 'Climatic variables determining *Rhododendron* sister taxa distributions and distributional overlaps in the Himalayas'. The abstract of this paper is provided, discussing the vulnerability of endemic species in mountains to rapid climate change and the research questions addressed. At the bottom, a link to '1 supplemental file' is visible. On the right side, there is a sidebar with 'Journal Information' (including the e-ISSN 1948-6596 and Creative Commons BY license) and 'Recent Articles' (listing 'Noted with Interest' by Keith, Sally A. and 'The inescapable influence of sea ice - a fundamental Earth ecosystem' by Terauds, Aleks).

# Open questions for the library publishing community

- What are some of the **challenges** that we face as we work to implement best practices for data publishing into our library publishing programs?
- Is it time for library-based publishing platforms to create **basic criteria around data publishing**?

# Thank you!

Monica Westin

Publications & Engagement Manager, California Digital Library

[monica.westin@ucop.edu](mailto:monica.westin@ucop.edu)

## References:

Fallaw, C., Dunham, E., Wickes, E., Strong, D., Stein, A., Zhang, Q., ... Imker, H. (2016). Overly Honest Data Repository Development. *Code4Lib Journal*, 34.

Lowenberg, D. Data Publication: Sharing, Crediting, and Re-Using Research Data. (2017). Retrieved from <http://uc3.cdlib.org/2017/03/22/data-publication-sharing-crediting-and-re-using-research-data/>

North Carolina State University Libraries. Metadata for Data: Data Management Plan. (n.d.). Retrieved from <https://www.lib.ncsu.edu/data-management/metadata>