Learning About Data Sharing

Katie Coburn, Ph. D.
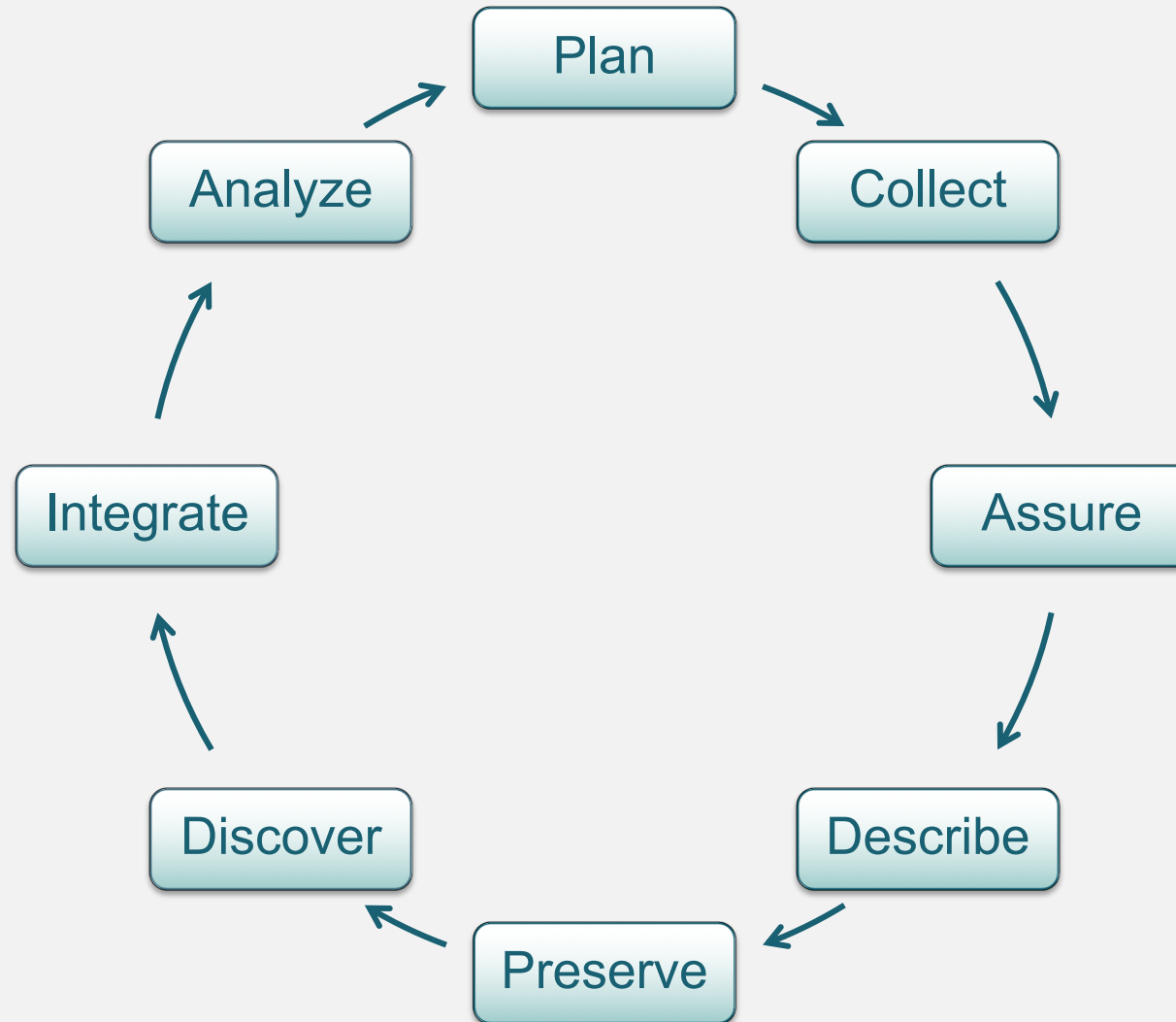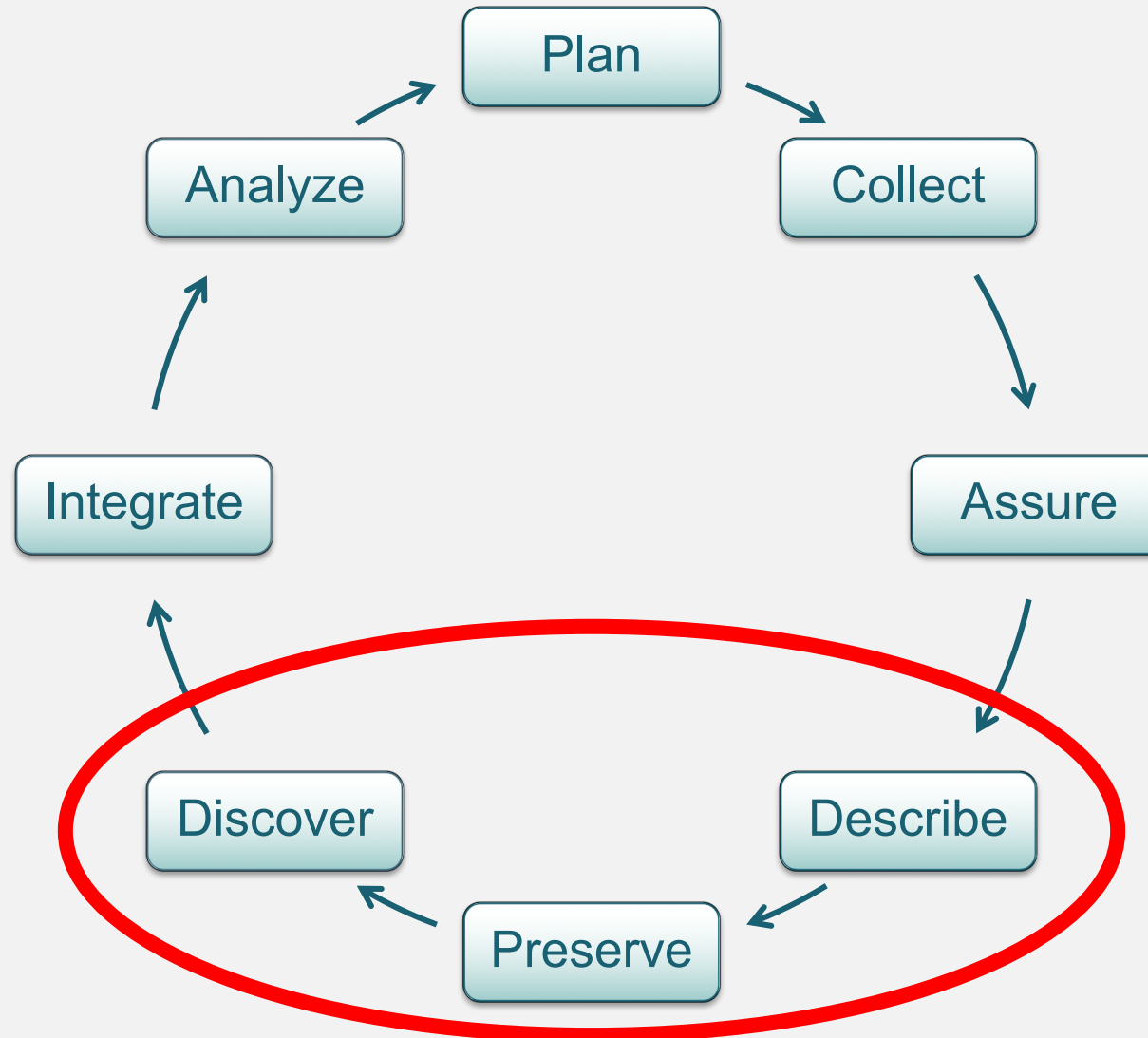**UC Merced Library**

ABOUT DATA

DATA ARE…

DATA ARE...

- Behind averages and standard deviations

- Images, blots, gels

- Qualitative information, answers to interviews and surveys

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

DATA LIFE CYCLE

Plan

Collect

Analyze

Integrate

Assure

Discover

Describe

Preserve

The DataONE Data Life Cycle

DATA LIFE CYCLE

TERMINOLOGY

Open Data

Part of the Open Access (OA) movement. Data that can be freely used, re-used, re-distributed (under CC-BY license)

Data Publication

Sharing your data publicly (open data) in a repository or platform that grants DOIs or stable identifiers for the purpose of re-use
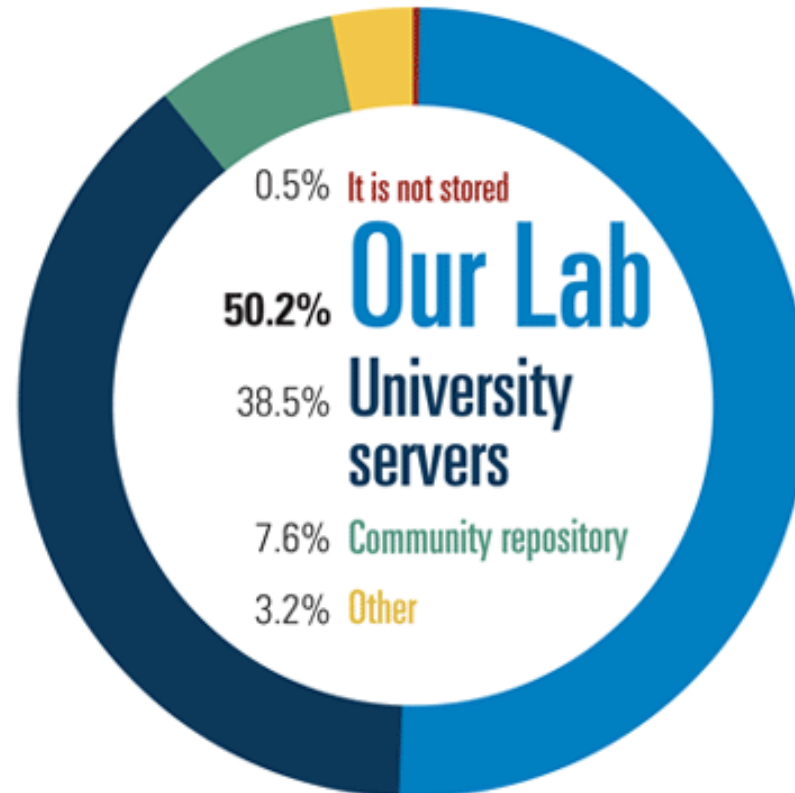
Data Repository

Place where data can be preserved, accessible, and citable
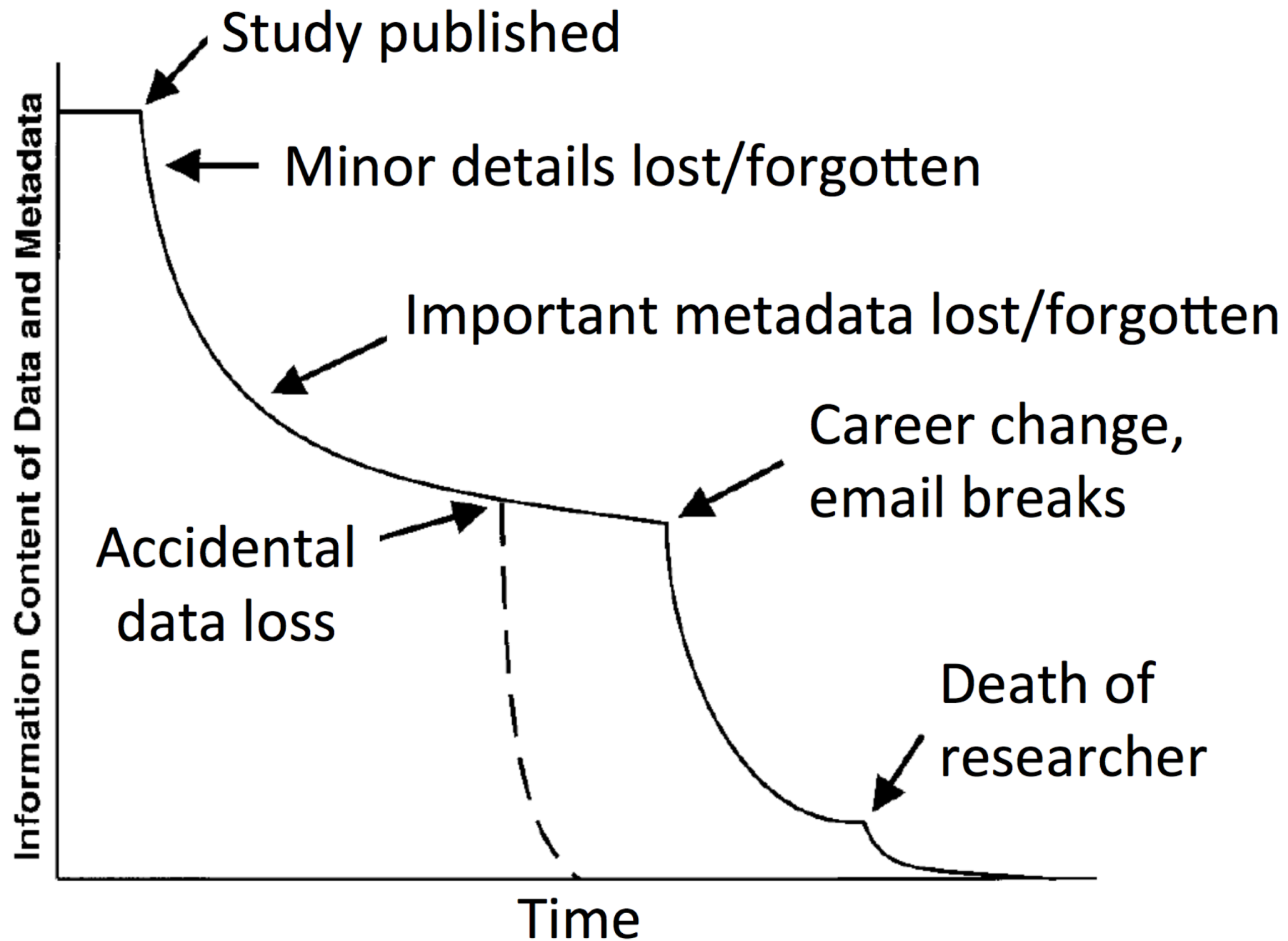
SHARING DATA

WORKING STORAGE =/= PRESERVATION

Where do you archive most of the data generated in your lab or for your research?
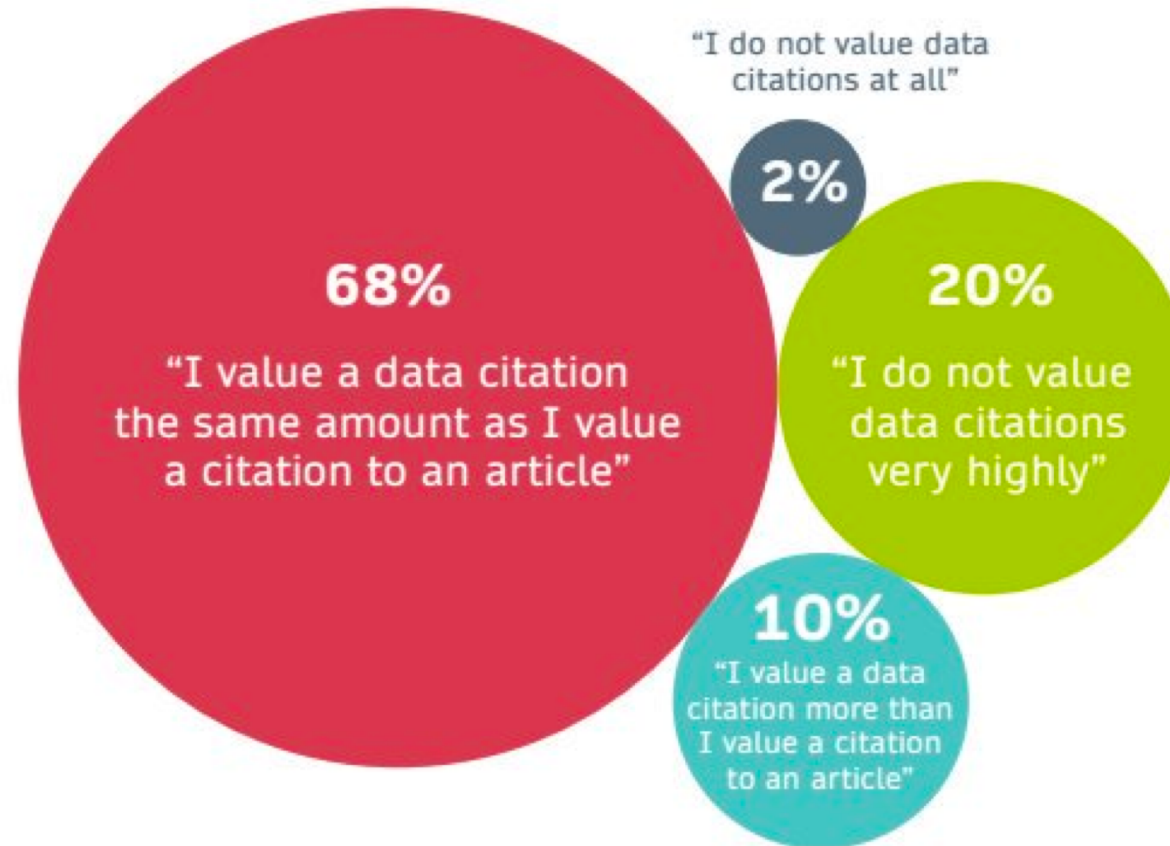
" Even within a single institution **there are no standards for storing data**, so each lab, or often each fellow, uses ad hoc approaches. "
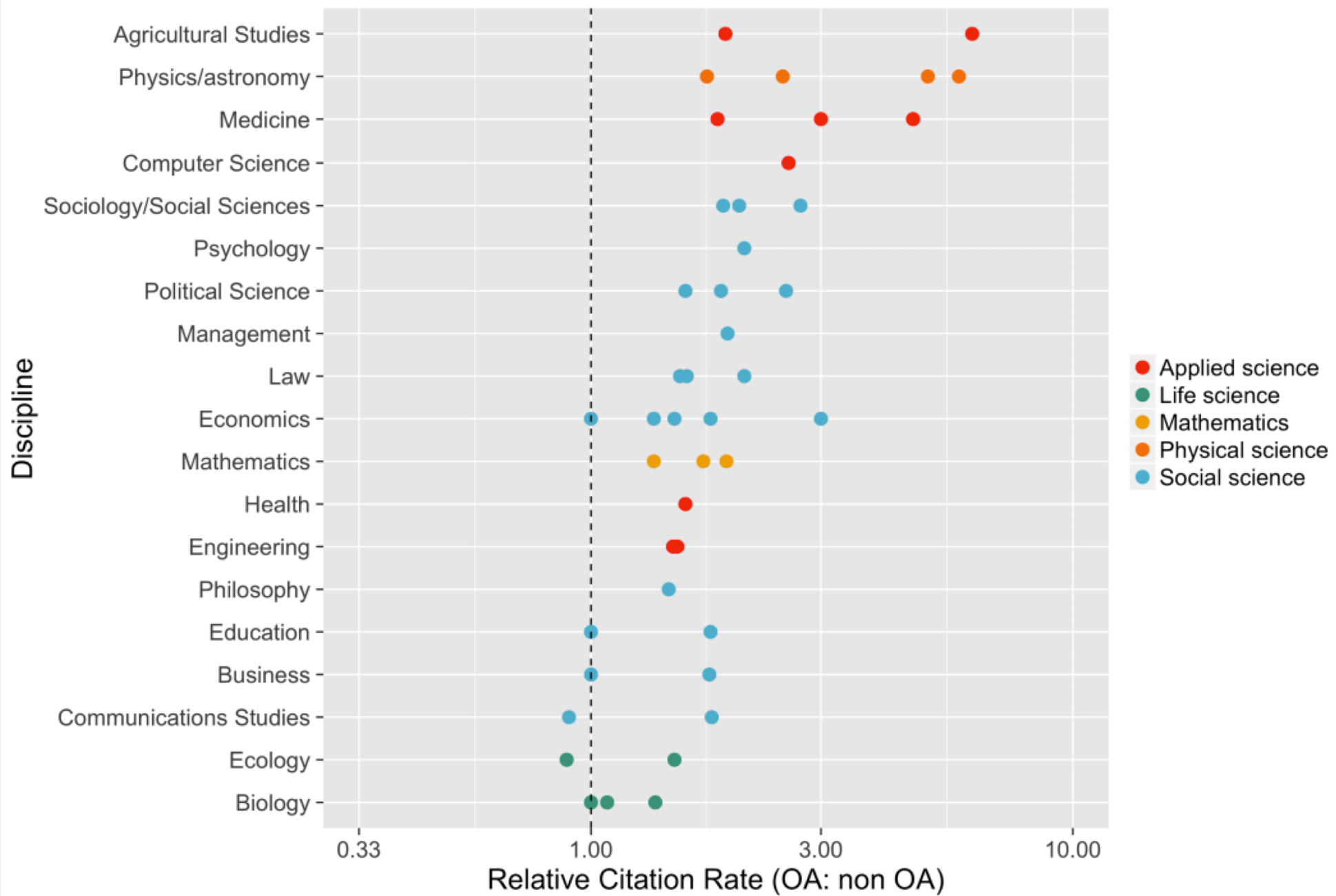
- 0.5% It is not stored
- 50.2% **Our Lab**
- 38.5% University servers
- 7.6% Community repository
- 3.2% Other

Challenges and Opportunities (2011). *Science.*

Michener *et al* (1997). Nongeospatial metadata for the ecological sciences.

GET CREDIT FOR YOUR DATA

80% of researchers value data citation as much as, or more than, article citation

68% "I value a data citation the same amount as I value a citation to an article"

2% "I do not value data citations at all"

20% "I do not value data citations very highly"

10% "I value a data citation more than I value a citation to an article"

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

BILL & MELINDA GATES FOUNDATION OPEN ACCESS POLICY

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

NIH Data Sharing Policies

This table lists data sharing policies in effect at NIH.

PLOS ONE To Require Public Access for Data

What is open data?

Elsevier supports the principle ↗ that "Raw research data should be made freely available to all researchers" and authors should be free to publically post their raw research data (see Author Rights for more details).

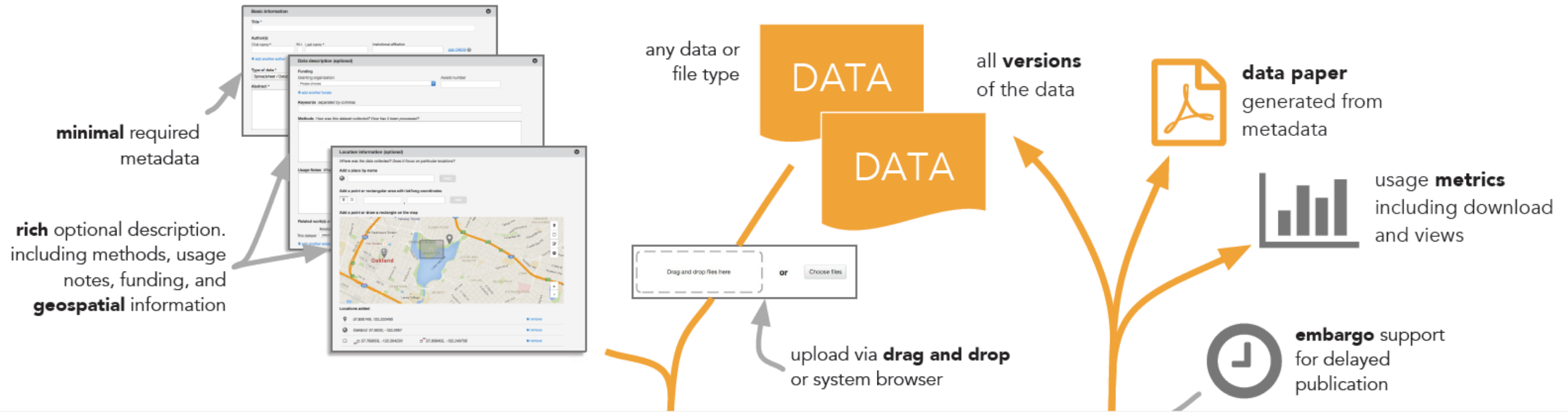## 73% of academics say access to research data helps them in their work

ABOUT DASH

- UC Curation Center (UC3) at California Digital Library (CDL) developed Dash to meet scholarly needs for data sharing

- Overlay for standards-compliant repositories (Merritt)

- **Open-source**

- Intuitive self-service interface for research data management

"Dash: Data Sharing Made Easy at the University of California"

FEATURES

minimal required metadata

rich optional description. including methods, usage notes, funding, and **geospatial** information

any data or file type

DATA

DATA

upload via **drag and drop** or system browser

all **versions** of the data

data paper generated from metadata

usage **metrics** including download and views

embargo support for delayed publication
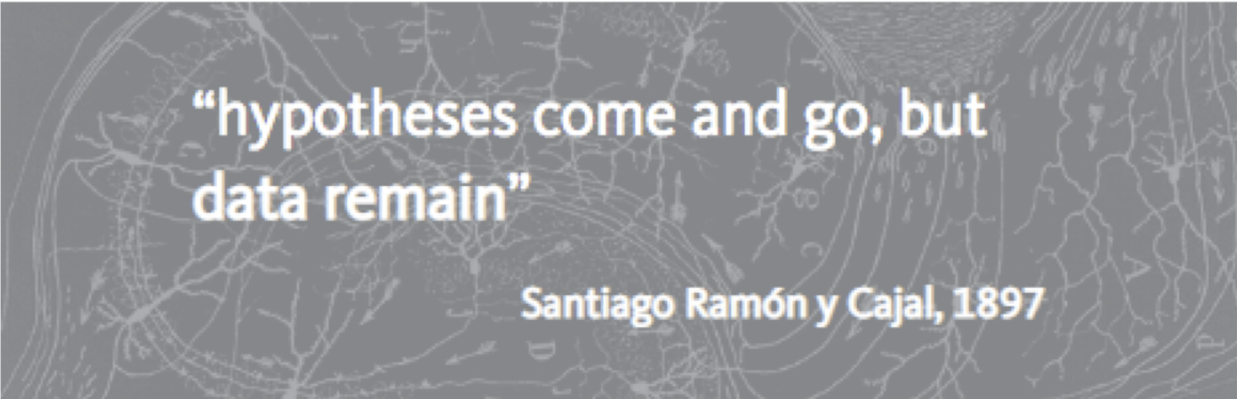
BENEFITS OF DASH

- Data are archived indefinitely

- Data can be published at any time

- Data can be versioned

- Data can be uploaded directly or through a "manifest" (list of URLs)

- Up to 100 GB of data can be uploaded at once

- No limit on file type

- Related works can be linked

- All data receive a DOI

- Link to ORCID ID

- Dash is a **UC project**

A QUICK TUTORIAL

An easy-to-use data publication service

"hypotheses come and go, but data remain"

Santiago Ramón y Cajal, 1897

Search 493 datasets

Search

Login with
your ORCID iD

- or -

Login using Single Sign-on

DataONE ⬍ Submit

My Datasets

+ Start New Dataset

In Progress

You do not have any datasets in progress.

Submitted

You do not have any submitted datasets.

Describe Your Datasets

Basic Information

Title *

[   ]

Author(s)

First Name *   Last Name   Institutional Affiliation *   Author Email *

[Kathleen]   [Coburn]   [   ]   [kcoburn@ucmerced.edu]

[+ Add Author]   [iD Connect your ORCID iD]   ?

Abstract *

Source | ✂ 📋 📋 📋 📋 | ↶ ↷ | 🔗 🔗 ⚑ | ▦ ▤ Ω | ⅓≣ ≔ | ⇥≣ ⇤≣ | 🙶 | DIV | ≣ ≣ ≣ ≣

Format   ▾ | A▾ A▾ | B I U S x₂ x² | Tₓ

Data Description (optional) ▲

Funding

Granting Organization

Award Number

+ add another funder

Keywords

Methods: How was this dataset collected? How has it been processed?

Source | ✂ ▢ ▢ ▢ ▢ | ↶ ↷ | ⌕ ⌕ ▶ | ▦ ▤ Ω | ⅟≡ ⋮≡ | ⋮≡ ⋮≡ | ❞ ▧ | ≡ ≡ ≡ ≡

Format ▾ | A▾ A▾ | B I U S x₂ x² | I_x

Usage Notes: What else would someone need to know to use this dataset? Are there any missing values? Please upload any files necessary for re-use (i.e. ReadMe files) and note them.

Source | ✂ ▢ ▢ ▢ ▢ | ↶ ↷ | ⌕ ⌕ ▶ | ▦ ▤ Ω | ⅟≡ ⋮≡ | ⋮≡ ⋮≡ | ❞ ▧ | ≡ ≡ ≡ ≡

Format ▾ | A▾ A▾ | B I U S x₂ x² | I_x

Related Work(s): Are there any articles (posted or published), datasets, or code that are related to this Data Publication? Please choose the relation type and identifier type from the drop downs and enter your unique permanent identifier below.

Relationship

cites

ID Type

DOI

Identifier

+ add another related work

Location Information (optional) ▲

Where was the data collected? Does it focus on particular locations?

Add a point or rectangular area with lat/long coordinates
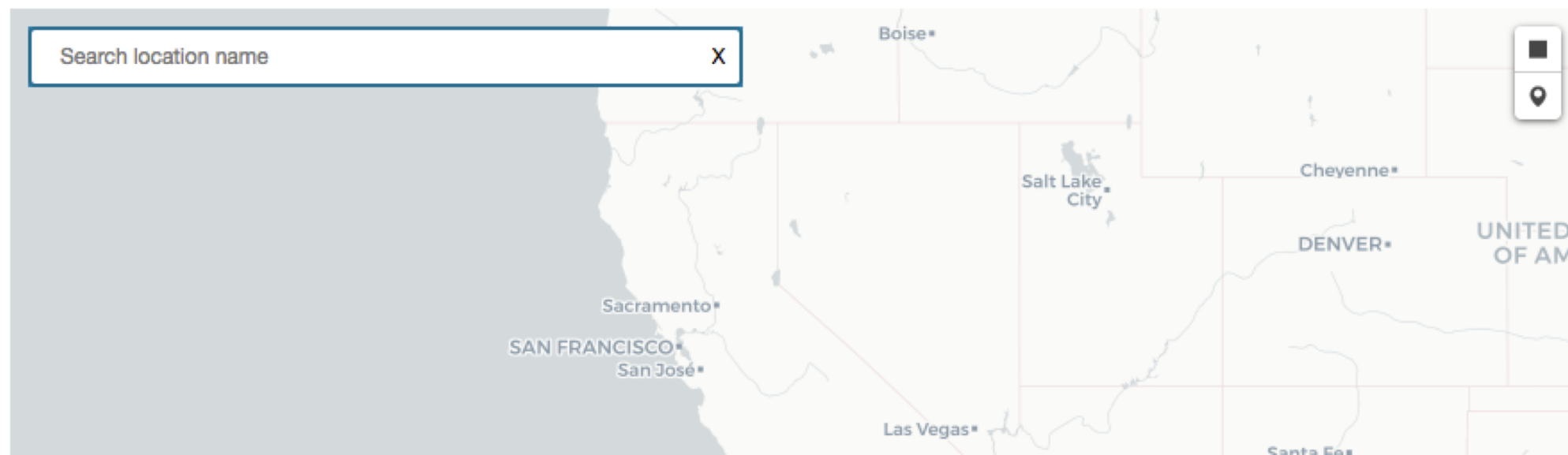
📍 Point ☐ Bounding Box 📍 | e.g., 37.801₂ | , | e.g., -122.25 | Add

Add a point or draw a rectangle on the map

Search location name   X

■
📍

Boise

Salt Lake City

Cheyenne

Sacramento

Denver

UNITED OF AM

SAN FRANCISCO
San José

Las Vegas

Santa Fe

Upload Your Files (optional)

Step 1: Choose File Upload Technique

You may upload data via two mechanisms: directly from your computer, or from a URL on an external server. Please note that for this version of your submission you may only use one of these two upload mechanisms.

Size limits: 10 GB of new files per version may be uploaded from your computer or 100 GB in total by URL or submitting multiple versions.

🔘 Drag and drop or choose files from your computer   ⚪ Enter URLs of file locations

Step 2: Choose Files

Drag and drop files here

or

Choose Files

Size limits: 10 GB of new files per version may be uploaded from your computer or 100 GB in total by URL or submitting multiple versions.

○ Drag and drop or choose files from your computer   ● Enter URLs of file locations

Step 2: Enter Files

```
https://ucmerced.app.box.com/folder/27007520129
```

Place each URL on a new line.

☐ * I confirm that no Personal Health Information or Sensitive Data are being uploaded with this submission.

Validate Files

Choose Publication Date

If you would like to keep your dataset private for peer review, for up to 180 days, please select a date below.

A private reviewer URL will be available for sharing until your dataset is public. Your permanent DOI for citation use will be available upon submission on the data landing page. If you would like to preserve data privately for longer than six months for non-peer review purposes, please contact us at curation@ucmerced.edu.

Make my dataset publicly available:

- ● Today
- ○ On 11 / 5 / 2018

Agree to Terms

Annual Survey of Orange County 1986

Baldassare, Mark

Publication date: October 31, 2014

Publisher: UC Irvine

https://doi.org/10.7280/D1059F

Citation

Baldassare, Mark (2014), Annual Survey of Orange County 1986, UC Irvine Dash, Dataset,
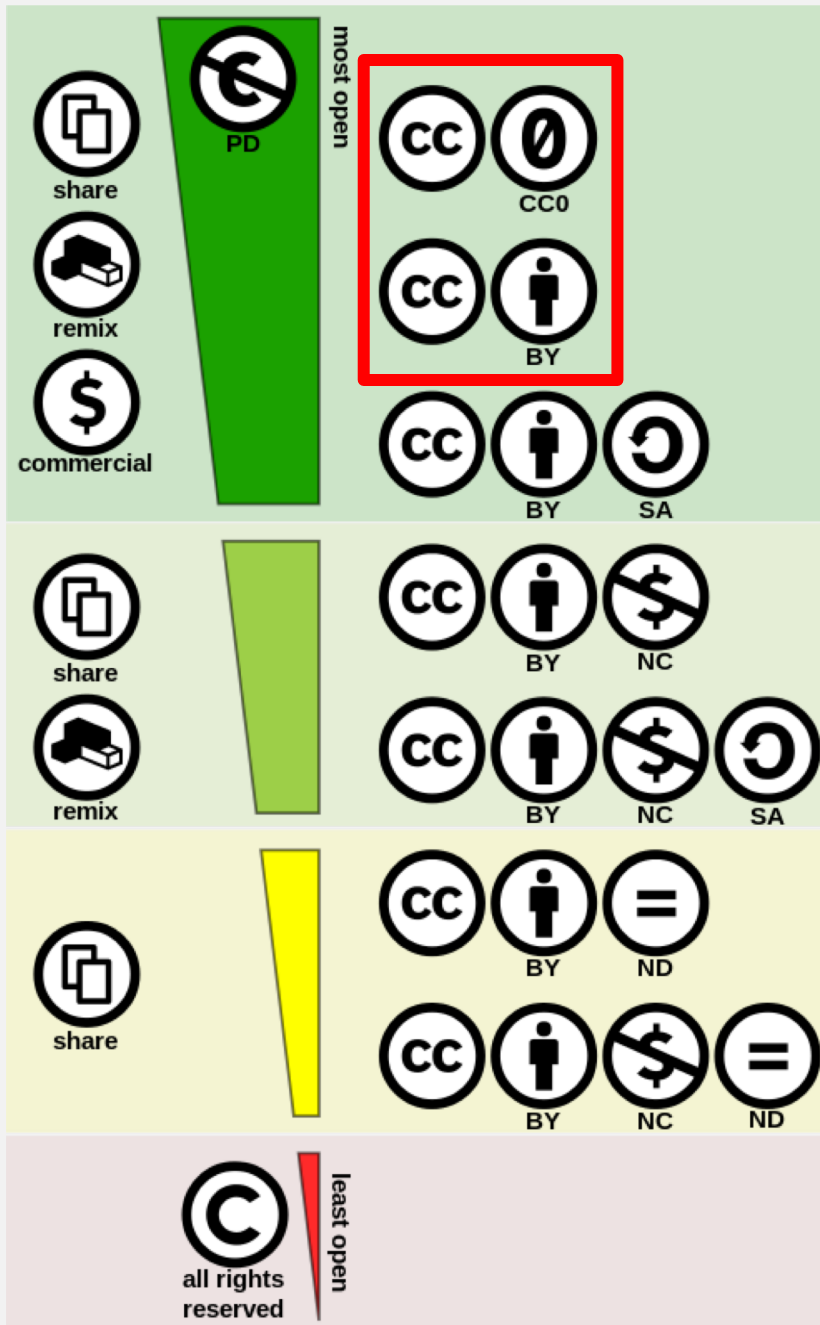
https://doi.org/10.7280/D1059F

Abstract

This is the fifth year of the Orange County Annual Survey. The 1986 report asks questions found in national surveys and compares Orange County to the nation. It examines key trends over time with questions repeated annually in our survey. The design, analysis, and interpretation of the 1986 survey are guided by goals (1) to provide decisionmakers in the private and public sectors with valid and current information on Orange County residents; (2) to track shifts in important attitudes and population characteristics over time; (3) to study the social, economic and political issues of Orange County from a non-advocacy positions and (4) to help establish public discussion and enlightened debate in future public policy. The sample size is 1,008 Orange County adult residents.
Online data analysis & additional documentation in Link below.

Methods

The Orange County Annual Survey was conducted by Mark Baldassare, an associate professor of social ecology at UC Irvine. The telephone interviews with 1,008 adult residents took place between September 2 and 19. Telephone interviews provide representative data in Orange County, because more than 97 percent of the households have telephones. The calls were made on weekend days and weekday nights using a random sample of listed and unlisted telephone numbers. Telephone numbers were randomly generated by computer from a list of working blocks of telephone exchanges. Within a household, respondents were chosen for interview using the Troldahl-Carter method, which randomly selects a household member from a grid. The field work was conducted by the Center for Survey Research, Public Policy Research Organization, at UC Irvine. The sample was generated by Pijacki Associates of Shoreham, New York. Each interview contained 90 questions and took an average of 20 minutes to complete. The surveys were designed in

---

Versions

∨ August 3, 2017

AnnualSurve...OC1986.pdf	3.13 MB
ocas86.dat	113.90 kB
ocas86.ddl	30.75 kB
ocas86.sas	17.36 kB
ocas86.sps	15.28 kB
ocas86.tab	158.09 kB
ocas86.xml	120.84 kB
ocas86cdbk.txt	69.72 kB
ocas86st.txt	19.05 kB

Metrics

SENSITIVE DATA

- [Creative Commons](#) licenses allow you to [choose](#) restrictions for your data
- Dash only allows the **first two licenses**
- The user (you) are responsible for anonymizing your data if necessary

MISCONCEPTIONS & SOLUTIONS

- De-identification should be within the realm of your IRB approvals
- If data cannot be shared publicly, data should be available upon request
- Subject area does not imply sensitive:
  - Ex. Abortion (laws) v Abortion (personal responses)
- Institutions & repositories working on solutions for sharing and storing sensitive data

QUESTIONS

Contact Us

Melinda Laroco Boehm, Ph.D.
Senior Research Development Officer
Office of Research Development
mboehm2@ucmerced.edu
(209) 382-4301

Emily S. Lin
Head, Digital Curation and Scholarship
Library
elin@ucmerced.edu
(209) 658-7146

Katie Coburn
CLIR Postdoctoral Fellow: Data Curation Specialist
Library
kcoburn@ucmerced.edu
(209) 205-6178

Jerrold Shiroma
Digital Scholarship Librarian
Library
jshiroma@ucmerced.edu
(209) 756-0237