

# Building Communities & Services for Data at CDL

Carly Strasser | @carlystrasser

DataCite  
25 August 2014



BY



University of California

**CDL**

California Digital Library



University of California

**CDL**

California Digital Library





University of California

**CDL**

California Digital Library

**stu**



**s**

UNIVERSITY  
OF  
CALIFORNIA



University of California

**CDL**

California Digital Library

UNIVERSITY  
OF  
CALIFORNIA



University of California

**CDL**

California Digital Library

**10** campuses

**233,000** students

receive **8%** of all research  
funding in the US

**2-3%** of all peer-reviewed articles  
published in the world every year

UNIVERSITY  
OF  
CALIFORNIA



University of California  
**CDL**  
California Digital Library



**UC3**  
UC Curation Center

UNIVERSITY  
OF  
CALIFORNIA



University of California  
**CDL**  
California Digital Library



**UC3**  
UC Curation Center

**Data curation**

**Archiving  
scholarship**

**Data  
management**

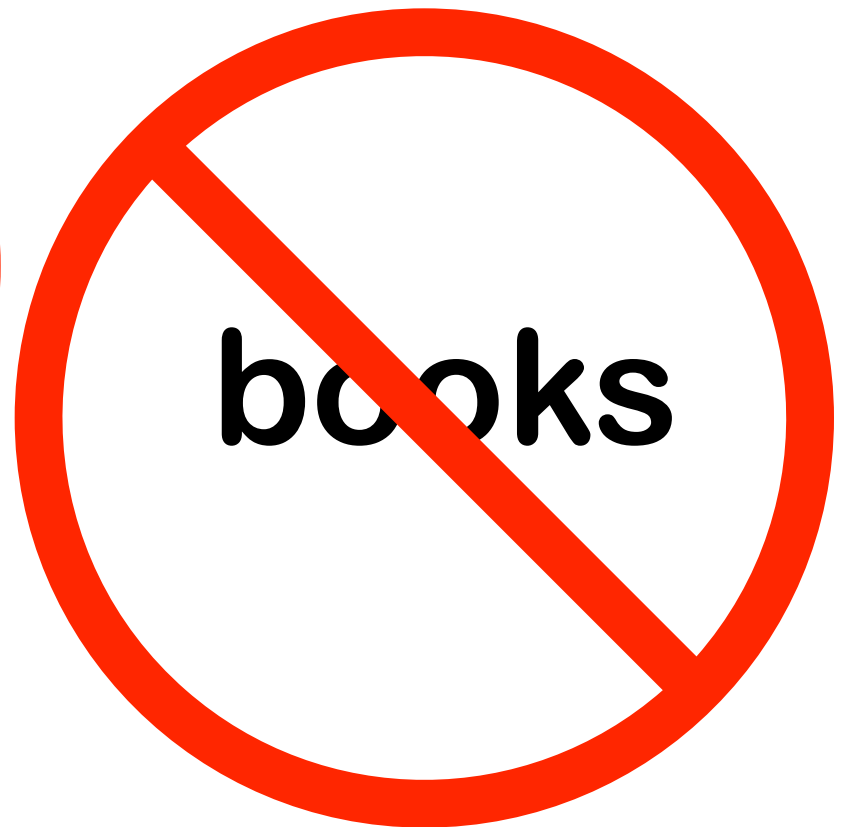
**Data sharing**

**Access to research  
products**

**Open  
research**



# How?



*From Flickr by mhlradio*

# Outreach & Education

# Tools & Services

# Participation & partnerships



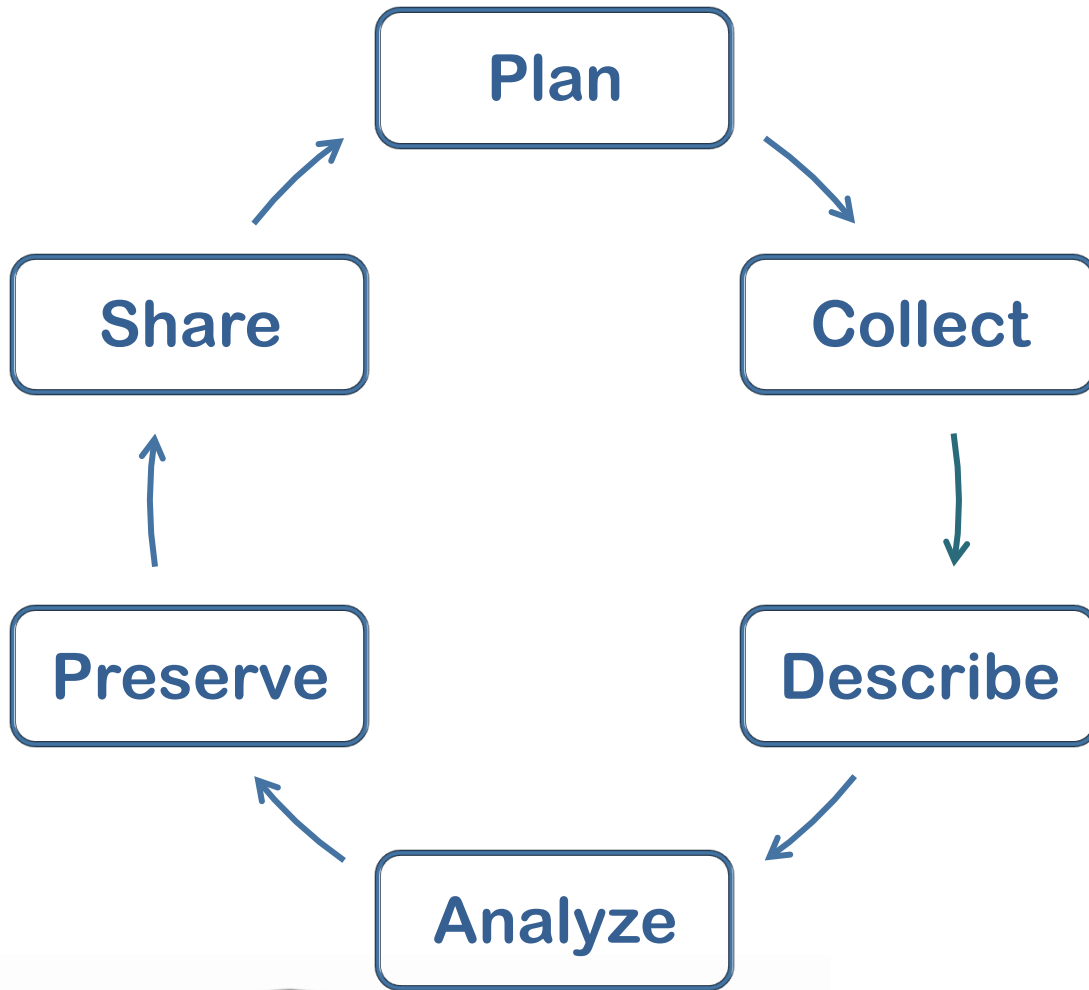
*From Flickr by gsagostinho*

*From Flickr by Boston Public Library*



# Tools & Services

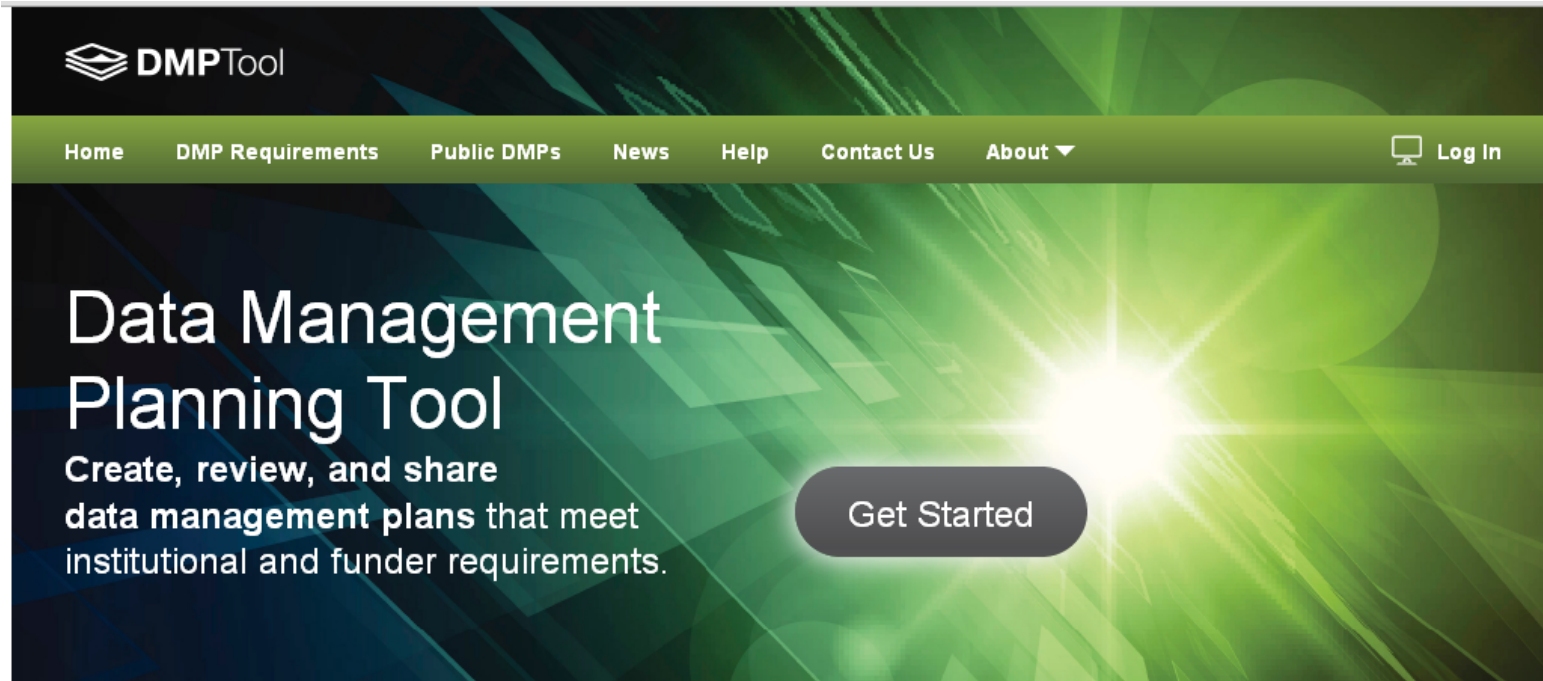
1



# Data Life Cycle

Plan > Collect > Describe > Analyze > Preserve > Share

# DMPTool: dmptool.org



**DMPTool**

Home DMP Requirements Public DMPs News Help Contact Us About ▼ Log In

## Data Management Planning Tool

Create, review, and share data management plans that meet institutional and funder requirements.

[Get Started](#)



### PUBLIC DMPs

List of sample data management plans provided by DMPTool users.



### DMPTOOL NEWS

Latest information about data management and the DMPTool.

- » [DMPTool2 Development: User Experience T...](#)
- » [Hang A DMPTool Poster!](#)
- » [DMP tools workshop at #IDCC14 by Sarah ...](#)
- » [DMPTool2 Release Schedule Update](#)
- » [DMPTool 2 Feature Highlight: Guidance a...](#)



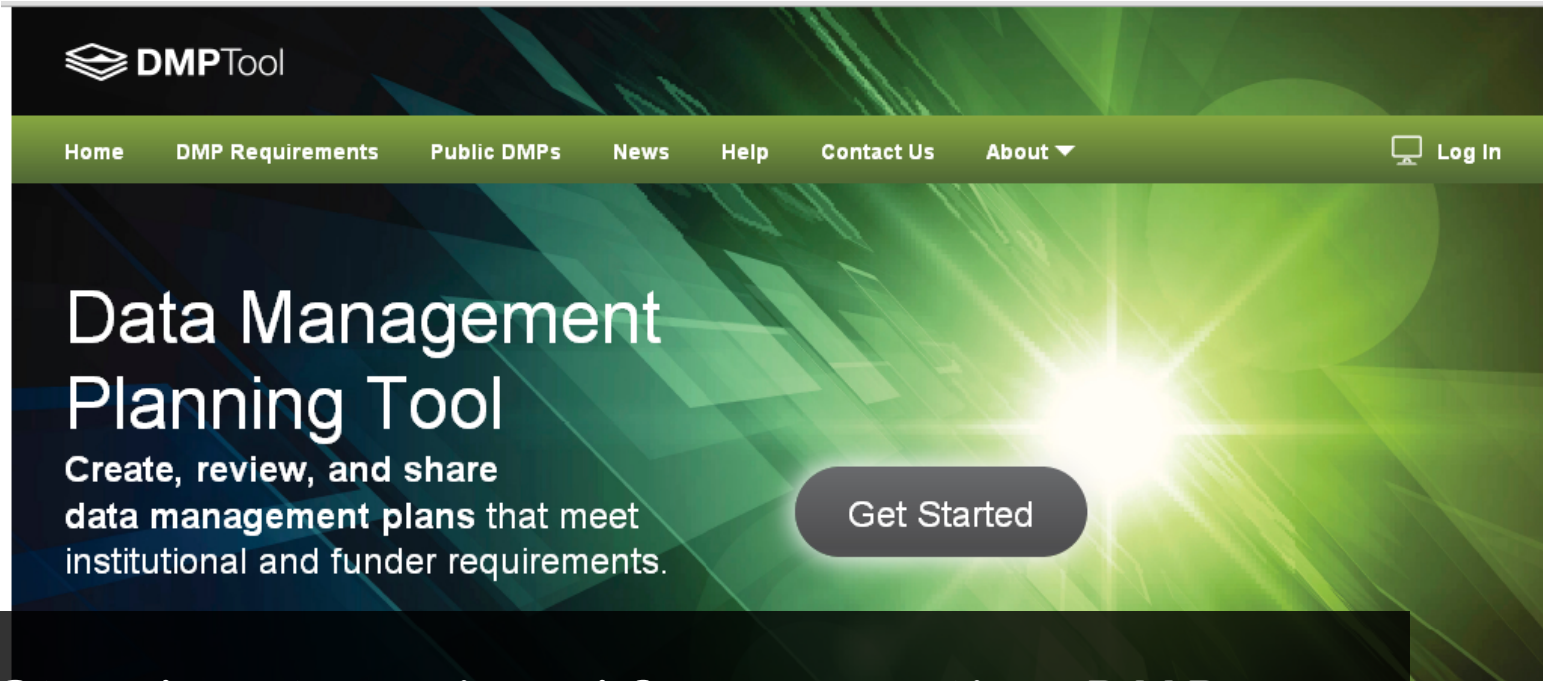
### DMPTOOL HELP

Overview of how to use the tool, plus resources and guidance on data management.

- » [Frequently Asked Questions](#)
- » [Create a DMP](#)
- » [Administer the DMPTool](#)
- » [Data management guidance](#)
- » [Community resources](#)

Plan > Collect > Describe > Analyze > Preserve > Share

# DMPTool: dmptool.org



Step-by-step wizard for generating DMP

Create | edit | re-use | share | save | export

Open to community

## PUBLIC DMPs

List of sample data management plans provided by DMPTool users.

## DMPTOOL NEWS

Latest information about data management and the DMPTool.

- > DMPTool2 Development: User Experience T...
- > Hang A DMPTool Poster!
- > DMP tools workshop at #IDCC14 by Sarah ...
- > DMPTool2 Release Schedule Update
- > DMPTool 2 Feature Highlight: Guidance a...

## DMPTOOL HELP

Overview of how to use the tool, plus resources and guidance on data management.

- > Frequently Asked Questions
- > Create a DMP
- > Administer the DMPTool
- > Data management guidance
- > Community resources



## DMP DETAILS

### Gordan and Betty Moore Foundation

#### Outline

- Summary of research and data**
- Approval requirements
- Data restrictions
  - Restriction 1
  - Restriction 2
- Additional documentation
- Intellectual property rights
- Tools used
  - Collecting data
  - Generating data
- Metadata
- Storage and access
  - Subsection 1
    - Subsection 1.1
- Long-term planning and archiving
- Depositing data in a repository
- Data management plan review

#### Instructions

#### Resources

#### Suggested Response

##### Instructions

Summarize the intended purpose of the research, the type of data to be collected or generated, the approximate dates when the data will be generated or collected, and the anticipated volume of data.


##### Guidance

Use this section to summarize the purpose of the research and the data that will be collected or generated in the course of the project. Make sure to include the anticipated or estimated volume of data. Consider these questions:

- What is the purpose of the research?
- What data will be collected and/or used in this project? (include only a brief statement here)

Hide Guidance

##### Your Response (Mandatory)

**B I U**    

Cancel Changes

Save Responses

Plan > Collect > **Describe** > Analyze > Preserve > Share



**Create persistent identifiers**

**Manage identifiers & associated metadata**

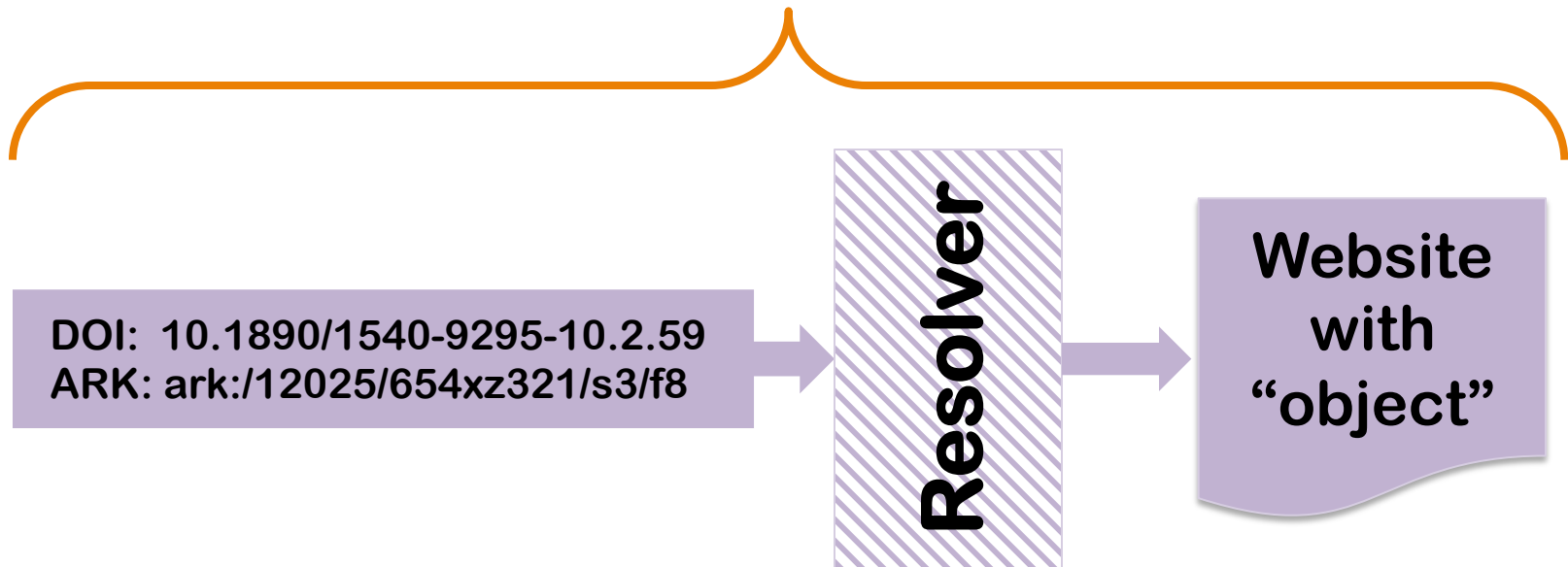
**Resolve identifiers**



Plan > Collect > **Describe** > Analyze > Preserve > Share



long-term identifiers made easy

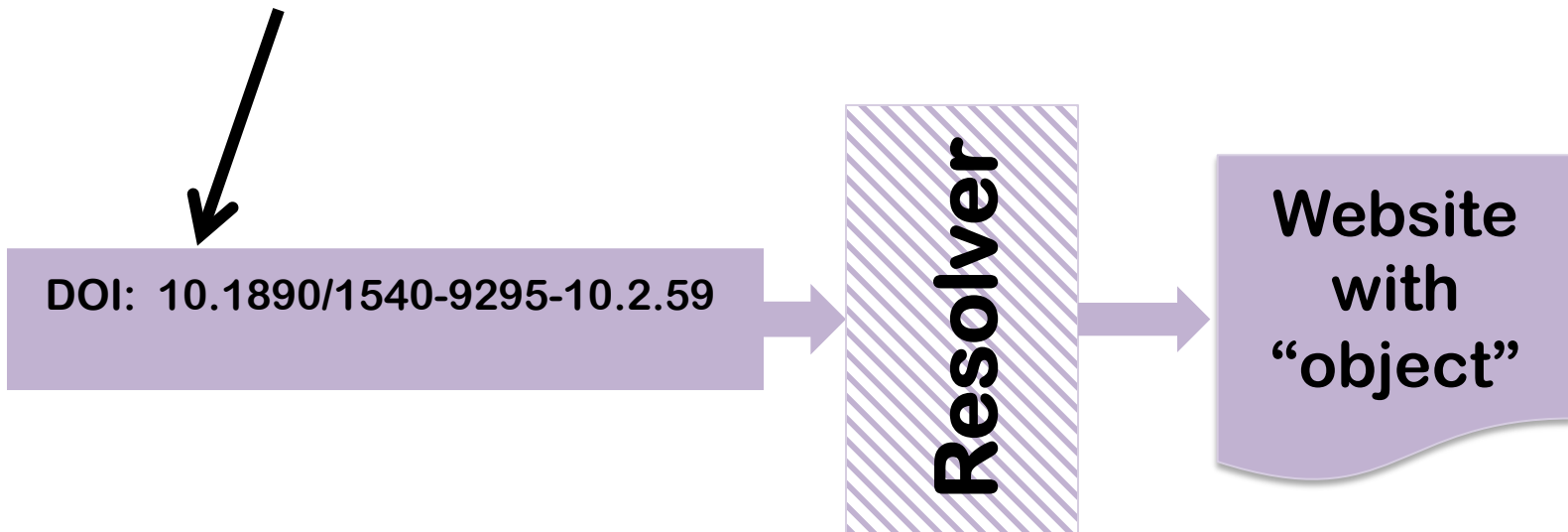


Plan > Collect > **Describe** > Analyze > Preserve > Share



DataCite

Helping you to find,  
access, and reuse data



Sidebar



University of California

**CDL**

California Digital Library

+

**cross** **ref**

# Sidebar



University of California

# CDL

California Digital Library



OA publications that get **DataCite DOIs**  
via **EZID** benefit from **CrossRef**  
metadata services & subscribers

Plan > **Collect** > Describe > Analyze > Preserve > Share



**Web Archiving Service**

Collect, manage & preserve websites

## **Preserve & store websites**

**Analysis tools**

**Full-text search**

**10,772 web sites**

Plan > Collect > Describe > Analyze > **Preserve** > Share



Share, manage & preserve your content

**Institutional repository for preservation  
& access to digital assets**

Plan > Collect > Describe > Analyze > **Preserve** > Share



Share, manage & preserve your content

**Institutional repository for preservation  
& access to digital assets**

## **About**

- **Launched 2010**
- **Content-agnostic**
- **Discipline-agnostic**
- **Deposit via API or UI**

Plan > Collect > Describe > Analyze > **Preserve** > Share



Share, manage & preserve your content

## Institutional repository for preservation & access to digital assets

### About

- Launched 2010
- Content-agnostic
- Discipline-agnostic
- Deposit via API or UI

### But...

- Account setup not self-service
- UI not user-friendly for novices



Plan > Collect > Describe > Analyze > **Preserve** > **Share**



Plan > Collect > Describe > Analyze > **Preserve** > Share



**Simplify data deposit and search for  
UC researchers**



**Free public access to research products**



## Search for Data

Go

OR

Browse all data



## Share Data

- ✓ Make your data count
- ✓ Get credit for your data
- ✓ Ensure reproducibility
- ✓ Promote reuse
- ✓ Meet funder requirements

Learn more

## Select a Dataset...

Faceted  
search &  
browse

### Campus

- UC Berkeley (8)
- UC Office of the President (7)
- UC Los Angeles (2)
- UC San Francisco (2)

### Author

- Joel (2)
- Strong, Marisa (2)
- Willett, Perry (2)
- Aguilar, Fernando (1)
- Banerjee, Utpal (1)
- Beylkin, Gregory (1)

*more*

### Keyword

- Amino acid sensing (1)
- Apartment houses (1)
- Chess Opening (1)
- Concrete columns (1)
- Data management (1)

*more*

 [Clear search](#)

Sorted by:

## Phylogenomic inference of protein molecular function: supplementary data

by Sölander, Kimmen  
at UC Berkeley  
> Show abstract

## DMPTool usage

by Willett, Perry | Willett, Perry  
at UC Berkeley  
> Show abstract

## Approximating a wavefunction as an unconstrained sum of Slater determinants: dataset

by Beylkin, Gregory | Mohlenkamp, Martin J. | Pérez, Fernando  
at UC Berkeley

# Phylogenomic inference of protein molecular function: supplementary data

## Citation

**Citation** Sölander, Kimmen (2003): Phylogenomic inference of protein molecular function: supplementary data. [ark:/99999/fk4893jsd](https://doi.org/10.1002/prot.10483)

**Title** Phylogenomic inference of protein molecular function: supplementary data

**By** [Sölander, Kimmen](#)

**Description** Protein families evolve a multiplicity of functions through gene duplication, speciation and other processes. As a number of studies have shown, standard methods of protein function prediction produce systematic errors on these data. Phylogenomic analysis—combining phylogenetic tree construction, integration of experimental data and differentiation of orthologs and paralogs—has been proposed to address these errors and improve the accuracy of functional classification. The explicit integration of structure prediction and analysis in this framework, which we call structural phylogenomics, provides additional insights into protein superfamily evolution.

**Methods** [Show Methods](#)

**Keywords** [Protein](#) | [Phylogenomic analysis](#)

**Identifier** [ark:/99999/fk4893jsd](https://doi.org/10.1002/prot.10483)

Download 159.6 Kb  
dataset





Content in Dash is offered "as is." By downloading files, you agree to the [Dash Terms of use](#). To the extent possible under law, the authors have waived all copyright and related or neighboring rights to these data by agreeing to a [CC0](#) license.


## License info

# Describe Your Dataset

\* Required Field

\* Dataset Title 

\* Institution 

\* Data Type 

\* Data Creator(s)   [+Additional Creator](#)

Keywords 

[+Additional Keyword](#)

Abstract 

Methods 

[Save](#)

[Save And Continue](#)

**DataCite metadata for now; plans for more**

# Upload Your Dataset(s)

Drag & drop files here

...or you can [+ Add files...](#)

[Go Back Edit Description](#)

[Review and Submit](#)

## Upload FAQ

- See [Policies](#) for more information on uploading datasets to Dash.
- All file formats are accepted by Dash, although it is good practice to share data using open formats. See the UK Data Archive for a [list of optimal file formats](#).
- Include any files that may help others to use your data. This includes readme files, formal metadata files, or other critical information.
- Any data submitted via Dash will be under a [Creative Commons Attribution 4.0 \(CC BY-4.0\)](#) license. We do not currently support any other license types, nor do we allow for restrictions on data access or use.
- It is your responsibility to ensure your data are being shared responsibly and ethically. Please be careful of sharing sensitive data and ensure you are complying with institutional and governmental regulations.

# Down the Road



Open source community  
Integrate with standard protocols  
Establish instances outside of UC



# Home

strasser edited this page on May 23 · 10 revisions

[github.com/cdluc3/dash/wiki](https://github.com/cdluc3/dash/wiki)



## Campus Implementation Documents

- [Implementation tasks](#)
- [Generic FAQ text](#)
- [Generic About text](#)
- [Agreements](#)

## Links & Documents

- [UC3 Dash: Plan of Work, Requirements, and Implementation](#) (google doc)
- [One-Page Summary](#) (google doc)
- [Feb 6 Webinar Slides](#) (slideshare)
- [UCSF DataShare website](#)
- Code repositories for UCSF DataShare project: [datashare](#) and [datashare-ingest](#)

Pages 7

- [agreements](#)
- [Campus Contacts](#)
- [Generic About for Dash](#)
- [Generic FAQ for Dash](#)
- [Home](#)
- [Policies](#)
- [Tasks for Campuses to Implement Dash](#)

+ Add a custom sidebar

Clone this wiki locally

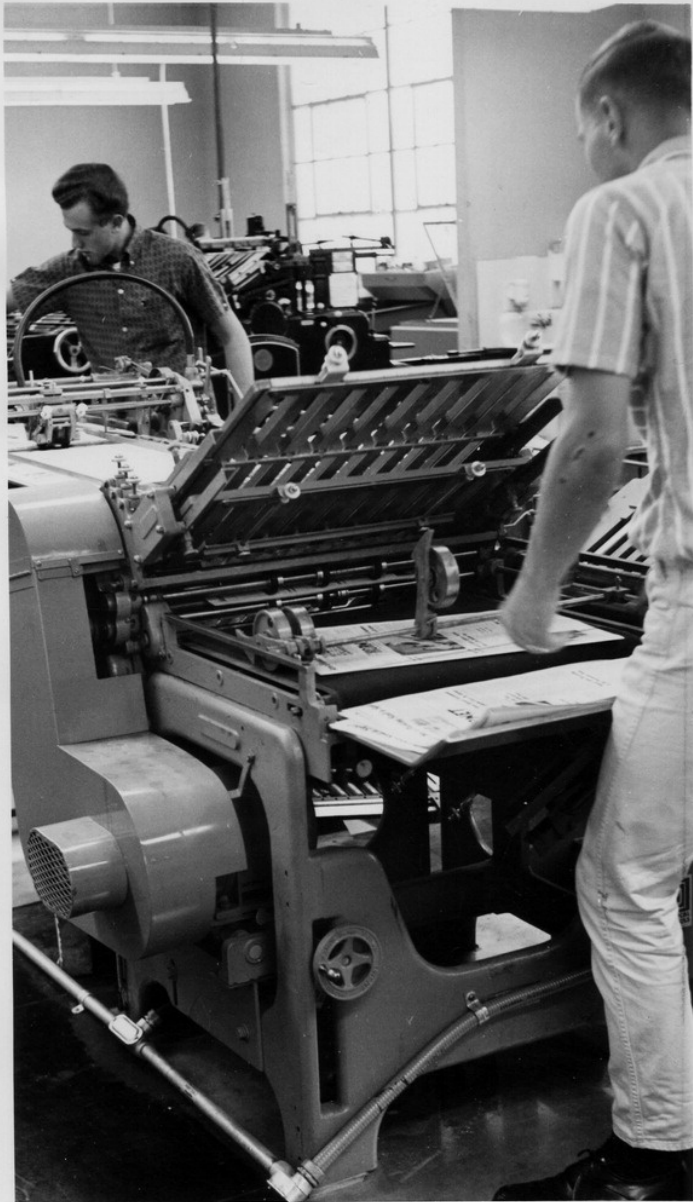
<https://github.com/CDLUC3/dash/wiki>

Clone in Desktop

Navigation sidebar with icons for: code, search, home, table of contents, back, forward, refresh, and other utility icons.

Plan > Collect > Describe > Analyze > Preserve > Share

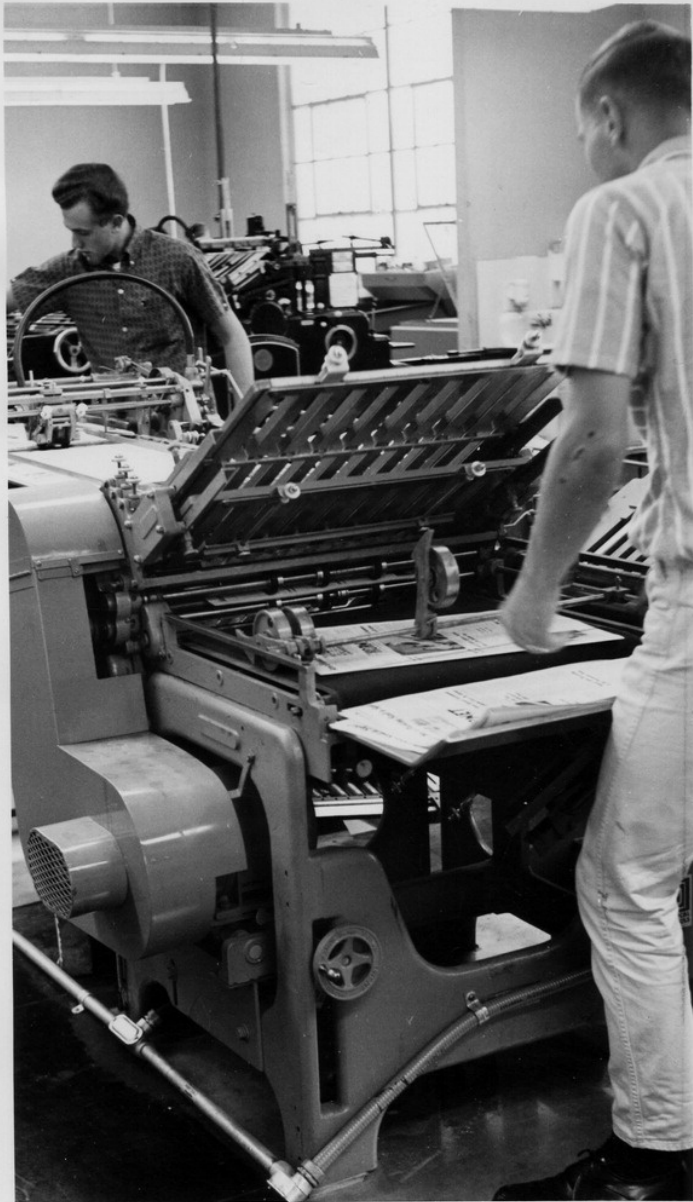
# Data Publication



Plan > Collect > Describe > Analyze > Preserve > Share

# Data Publication

What should data  
publication look  
like at UC?





eScholarship  
University of California



**John Kratz,  
Postdoc**

**What should data  
publication look  
like at UC?**



REVIEW

**REVISED** Data publication consensus and controversies [v2; ref status: indexed, <http://f1000r.es/3hi>]

John Kratz, Carly Strasser

California Digital Library, University of California Office of the President, Oakland, CA, 94612, USA

**v2** First published: 23 Apr 2014, 3:94 (doi: [10.12688/f1000research.4264](https://doi.org/10.12688/f1000research.4264))  
 Latest published: 16 May 2014, 3:94 (doi: [10.12688/f1000research.4518](https://doi.org/10.12688/f1000research.4518))

**Abstract**

The movement to bring datasets into the scholarly record as first class research products (validated, preserved, cited, and credited) has been inching forward for some time, but now the pace is quickening. As data publication venues proliferate, significant debate continues over formats, processes, and terminology. Here, we present an overview of data publication initiatives underway and the current conversation, highlighting points of consensus and issues still in contention. Data publication implementations differ in a variety of factors, including the kind of documentation, the location of the documentation relative to the data, and how the data is validated. Publishers may present the data as supplemental material to a journal article, with a descriptive “data paper,” or independently. Complicating the situation, different initiatives and communities use the same terms to refer distinct but overlapping concepts. For instance, the term “published” means that the data is publicly available and

**doi: 10.12688/f1000research.4518**

standards and processes encompass the full range employed in reviewing the

**Article Status Summary**

**Referee Responses**

Referees	1	2	3
<b>v1</b> published 23 Apr 2014	 report 1		
<b>v2</b> published 16 May 2014 <b>REVISED</b>		 report	 report

1 **Mark Parsons**, Research Data Alliance USA, **Peter Fox**, Rensselaer Polytechnic Institute USA

2 **Mark Costello**, University of Auckland

**“Available”**

**“Publish”**

**“Paper”**

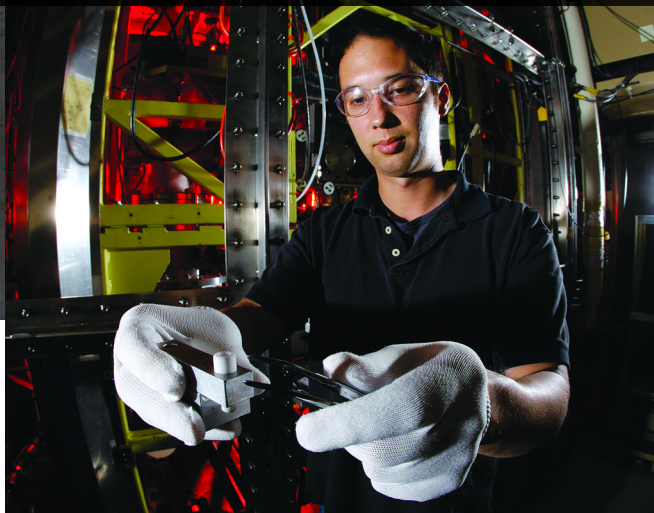
**“Sharing”**

**“Peer review”**

**“Article”**

**“Publication”**

# What do researchers think of data publication?



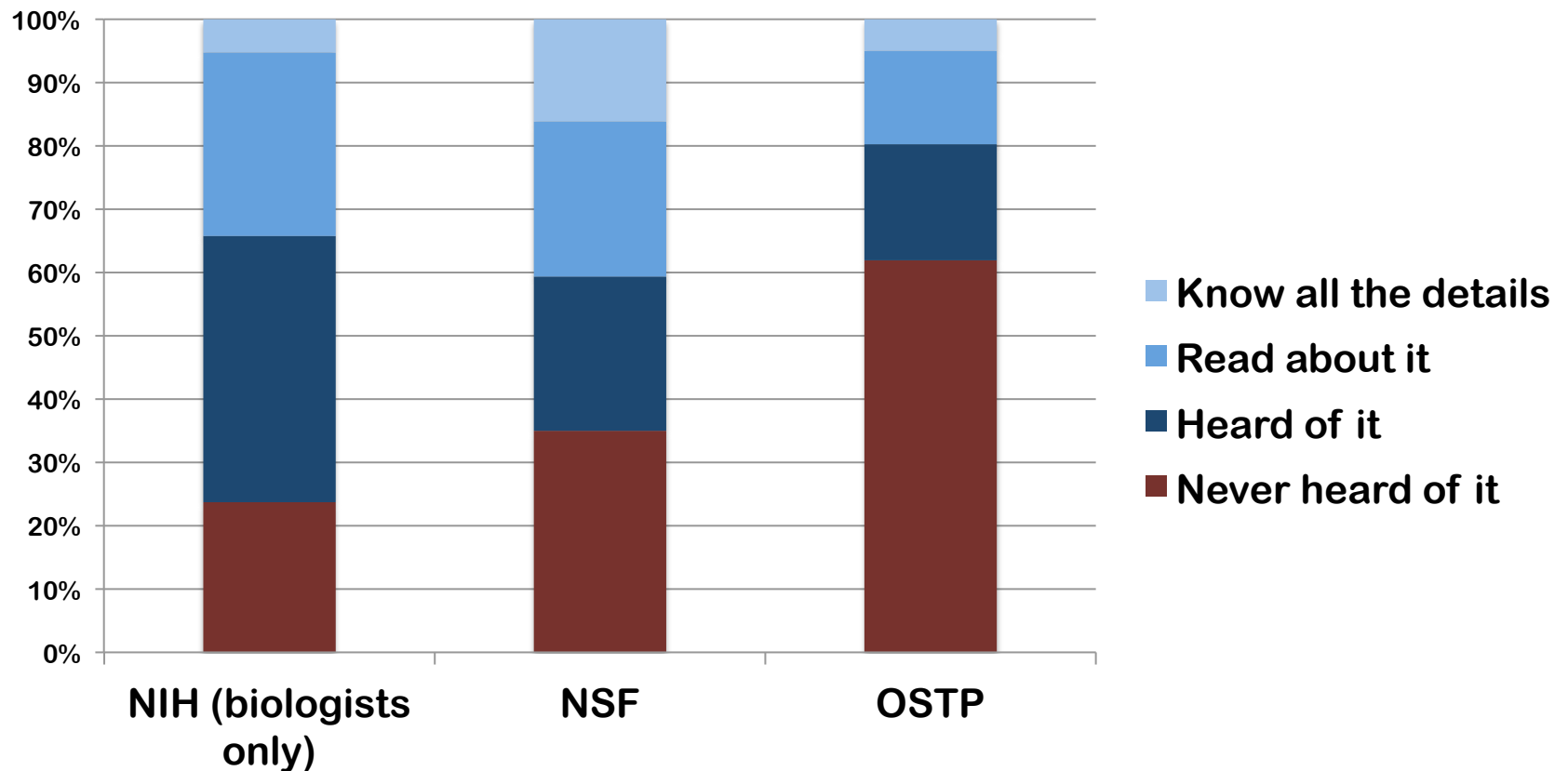
C. Strasser

From Flickr by Sandia Labs

# Survey of researchers

N=274

“How familiar are you with these data policies?”

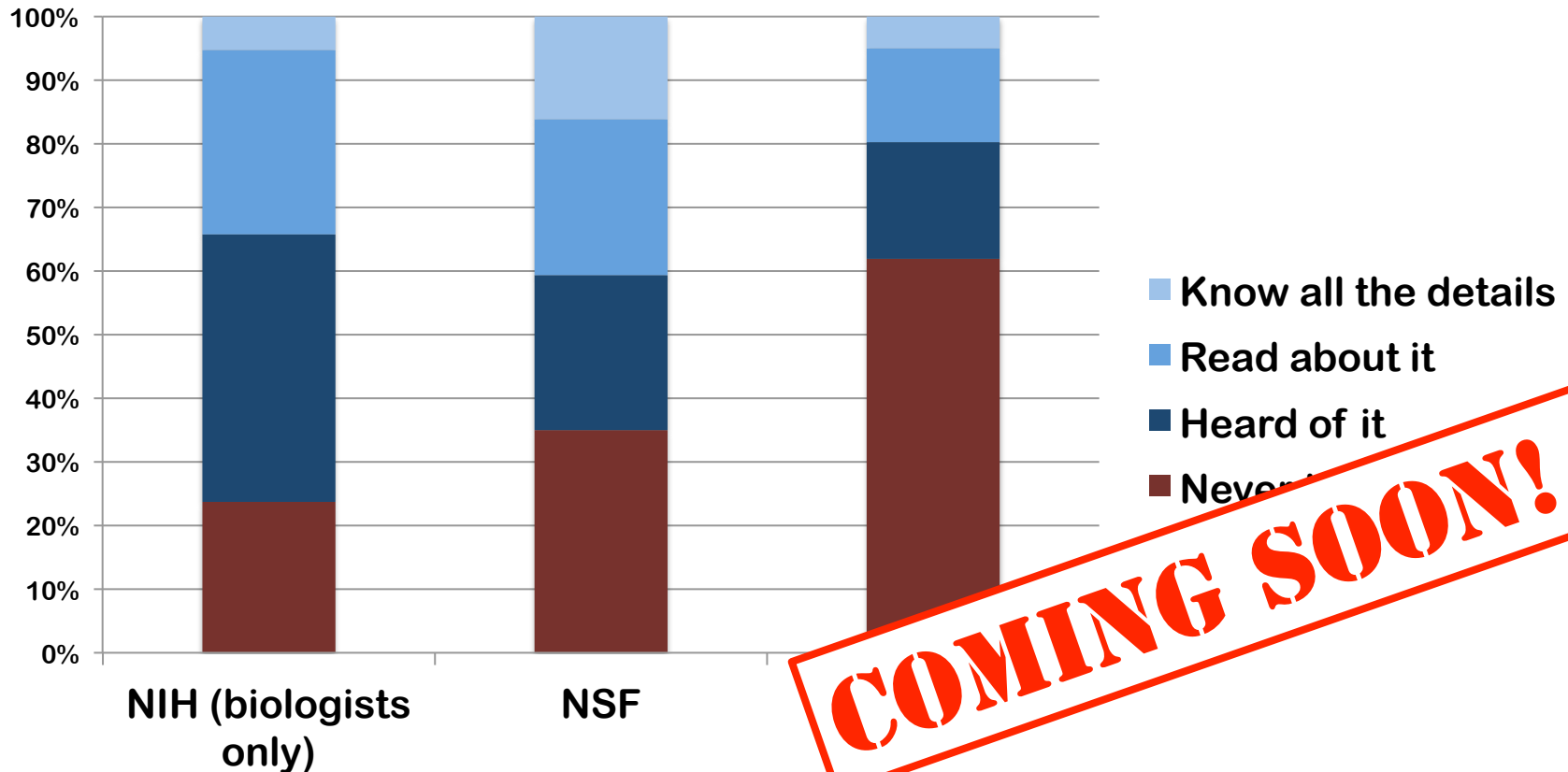




# Survey of researchers

N=274

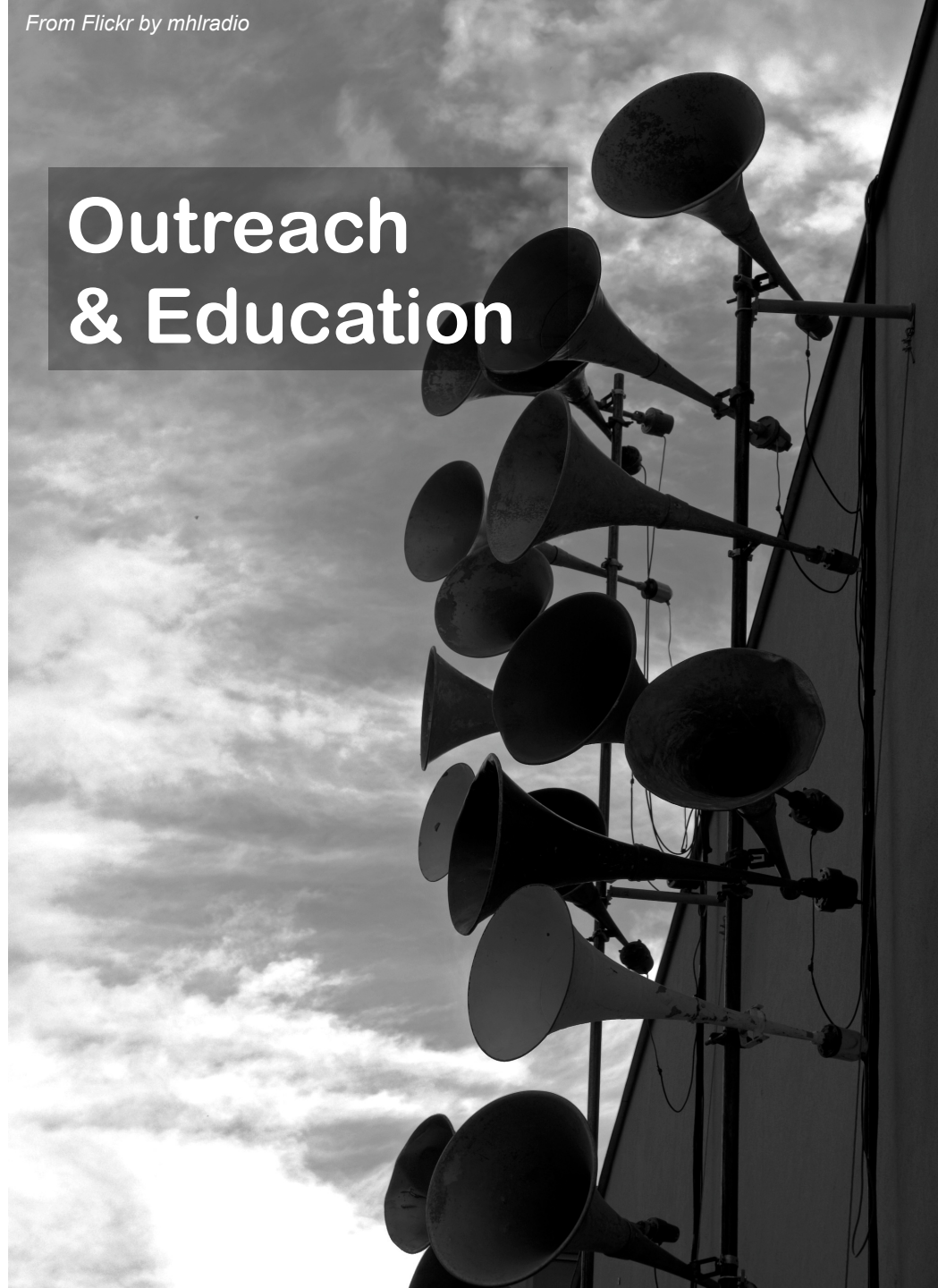
“How familiar are you with these data policies?”



**COMING SOON!**

# Outreach & Education

# 2



# Help Researchers...

**care & prioritize**

**document data**

**write good DMPs**



**Conferences**

**Campus visits**

**Workshops**

**Guest lectures**

**SOAIP**

**Conferences**

**Campus visits**

**Workshops**

**Guest lectures**

**SOAIP**

**Data management best practices**

**Open science**

**Tools & tips**

# Support Librarians in Helping Researchers



**Researchers**

**Librarians**

**CDL**

# DATA PUB

Blog about all things data from the California Digital Library

DMP WORKSHOP DATA CITATION HOME

**JUN 16 2014**  
**LEAVE A COMMENT**

BY CARLY STRASSER

LIBRARIES, MEETINGS &  
CONFERENCES

## IT TAKES A DATA MANAGEMENT VILLAGE

A couple of weeks ago, information scientists, librarians, social scientists, and their compatriots gathered in Toronto for the [2014 IASSIST meeting](#). IASSIST is, of course, an acronym which I always have to look up to remember – [International Association for Social Science Information Service & Technology](#). Despite its forgettable name, this conference is one of the better meetings I've attended. The conference leadership manages to put together a great couple of days, chock full of wonderful plenaries and interesting presentations, and even arranged a hockey game for the opening reception.

Although there were many interesting talks, and I'm still processing the great discussions I had in Toronto, a couple really rang true for me. I'm going to now shamelessly paraphrase one of these talks (with permission, of course) about



# Presentations & workshops

- **Data management education**
- **Tools and services**
- **Open data and open science**
- **The role of libraries**



COLORADO  
**ALLIANCE**  
OF RESEARCH LIBRARIES

**IDCC**



**Council on  
Library and  
Information  
Resources**



**Research Data  
Access & Preservation Summit**





INSTITUTE *of*  
**Museum** and **Library**  
SERVICES

## Improving Data Stewardship with the **DMPTool**: Helping Libraries Promote Data Management



INSTITUTE *of*  
**Museum** and **Library**  
SERVICES

## Improving Data Stewardship with the **DMPTool**: Helping Libraries Promote Data Management

- Data management guidance
- Promotional materials for DMPTool:  
slides, posters, postcards
- Improvements to DMPTool website

# 3

## Participation & partnerships





University of California

**CDL**

California Digital Library



DataCite



**Founding member of DataCite**

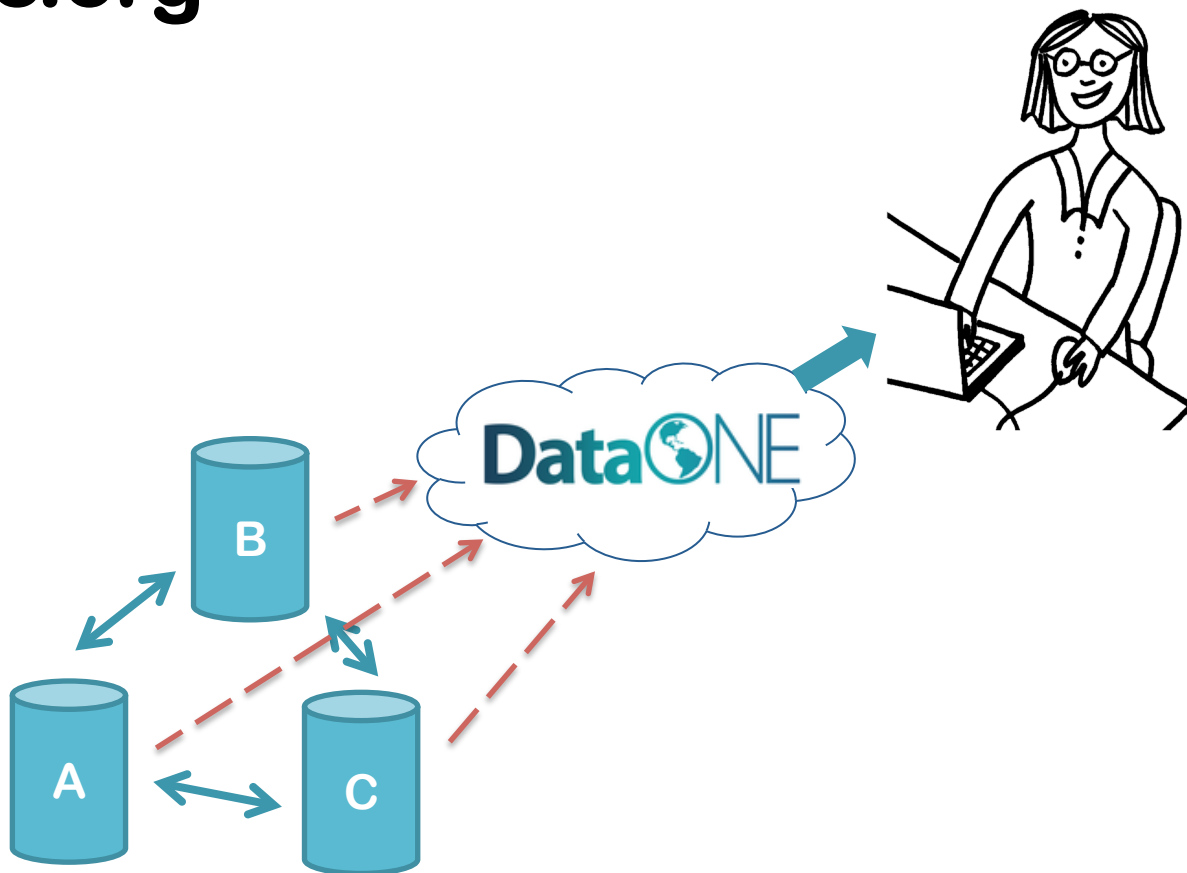
**Received DataCite Gold Award last 2 years**

**Starr & Cruse are active participants**



NSF-funded DataNet Project  
Office of Cyberinfrastructure

[www.dataone.org](http://www.dataone.org)

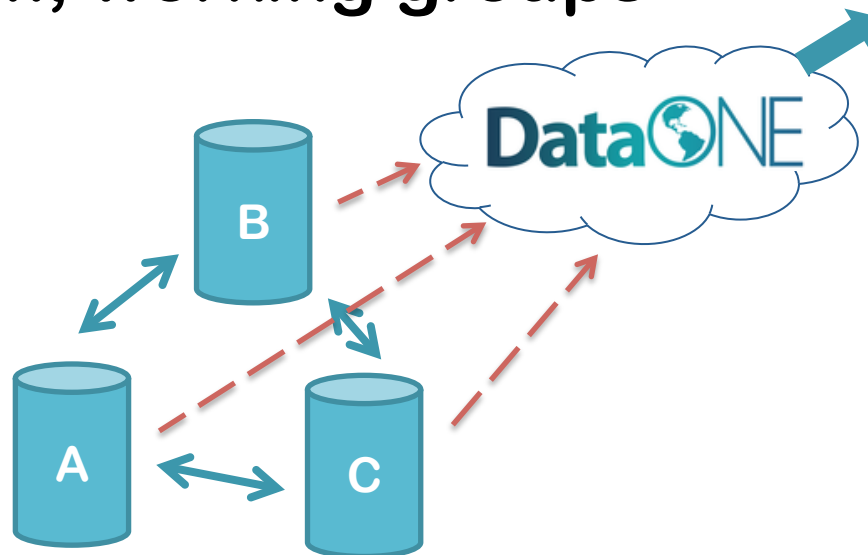




NSF-funded DataNet Project  
Office of Cyberinfrastructure

[www.dataone.org](http://www.dataone.org)

**UC3 member involvement:**  
leadership team, working groups





Research Data Sharing  
without barriers



- Home
- About
- Organisation
- Working and Interest Groups
- News & Events
- Plenary Meetings
- RDA in the Press



**Elections for RDA Technical Advisory Board members are now open! Candidate profiles are available [here](#)**



### Certification of Digital Repositories IG

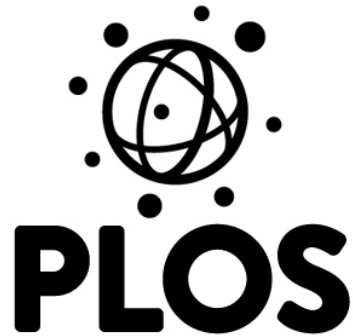
Status: Recognised & Endorsed Joint RDA/WDS Interest Group  
This Working Group will build on previous work in the area of certification and deliver a global overview and a set of recommendations and requirements that allow the effective implementation of certification of digital repositories on a national, European and even global level.

- Agricultural Data Interoperability IG
- Big Data Analytics IG
- Brokering IG
- Certification of Digital Repositories IG**
- Community Capability Model WG
- Data Citation WG
- Data Foundation and Terminology WG





# Partners



[nature.com](http://nature.com) : Publications A-Z index



**Website**

**carlystrasser.net**

**Email**

**carlystrasser@gmail.com**

**Twitter**

**@carlystrasser**

**Slides**

**slideshare.net/carlystrasser**



From Flickr by compujeramey

