

UC Office of the President
CDL Staff Publications

Title

Supporting research data management at the University of California

Permalink

<https://escholarship.org/uc/item/51r4q7kw>

Author

Abrams, Stephen

Publication Date

2017-12-16

AXIES Conference, Hiroshima, 14-15 December 2017

Supporting Research Data Management at the University of California



Stephen Abrams

California Digital Library
University of California

Stephen.Abrams@ucop.edu

ORCID [0000-0003-2326-6672](https://orcid.org/0000-0003-2326-6672)



Research data management (RDM)

Effective management of scholarly research data is necessary

- To ensure integrity and accountability
- To avoid needless duplication of effort
- To enable scholarly inquiry, innovation, and advancement
- To promote public awareness and informed discourse



The library's role in research data management

The continuation of the long-standing mission to provide effective stewardship of the University's intellectual capital by its libraries, archives, and museums



www.flickr.com/photos/janoma/4754744617

University of California

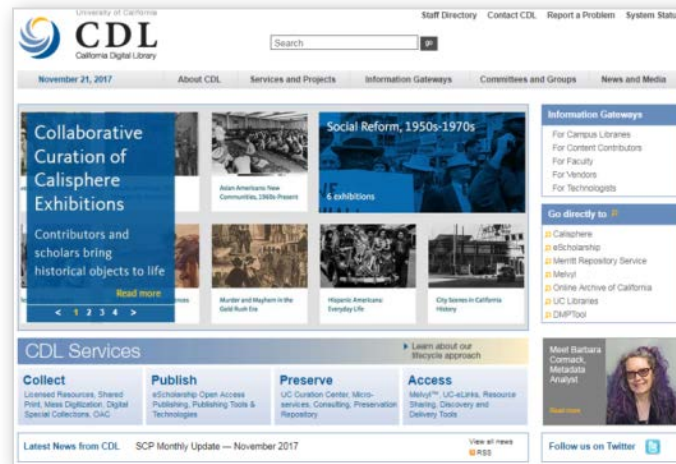
A large and diverse research University system

- 10 campuses
 - ▶ 238,000 students;
 - ▶ 190,000 faculty and staff
 - ▶ 150 academic disciplines;
 - ▶ 900 graduate programs
- 5 medical centers
- 3 national laboratories
- \$4.97 billion (¥559 billion) in annual funded research



California Digital Library (CDL)

Providing transformative digital library services, grounded in campus partnerships and extended through external collaborations, that amplify the impact of the libraries, scholarship, and resources of the University of California



www.cdlib.org

UC Curation Center (UC3)

Providing innovative solutions for active curation and long-term preservation of the University's digital resources



uc3.cdlib.org
www.cdlib.org/uc3

What do we mean by curation and preservation?

Curation

“Maintaining, preserving and adding value to digital research data throughout its lifecycle”

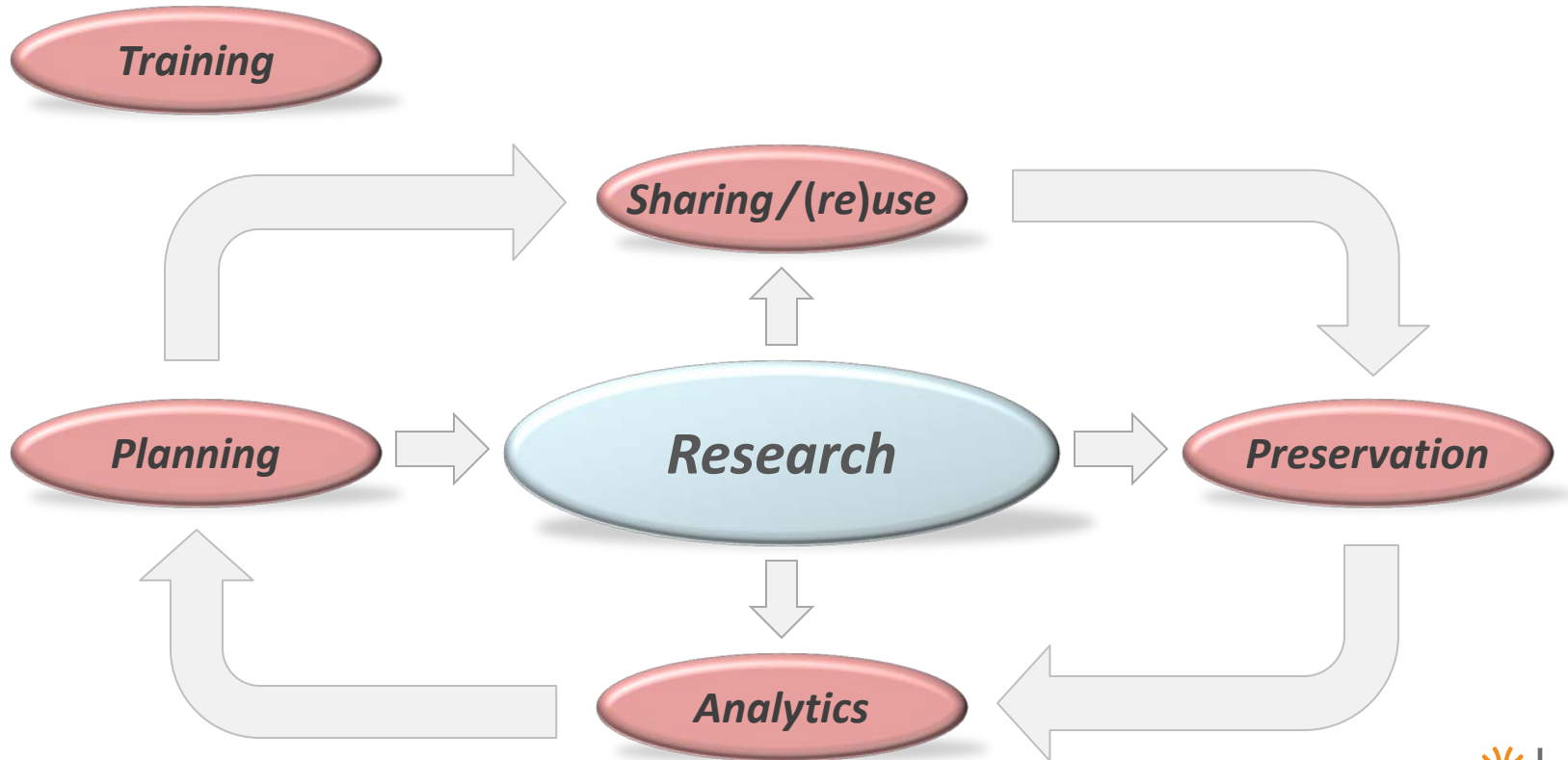
www.dcc.ac.uk/digital-curation/what-digital-curation

Preservation

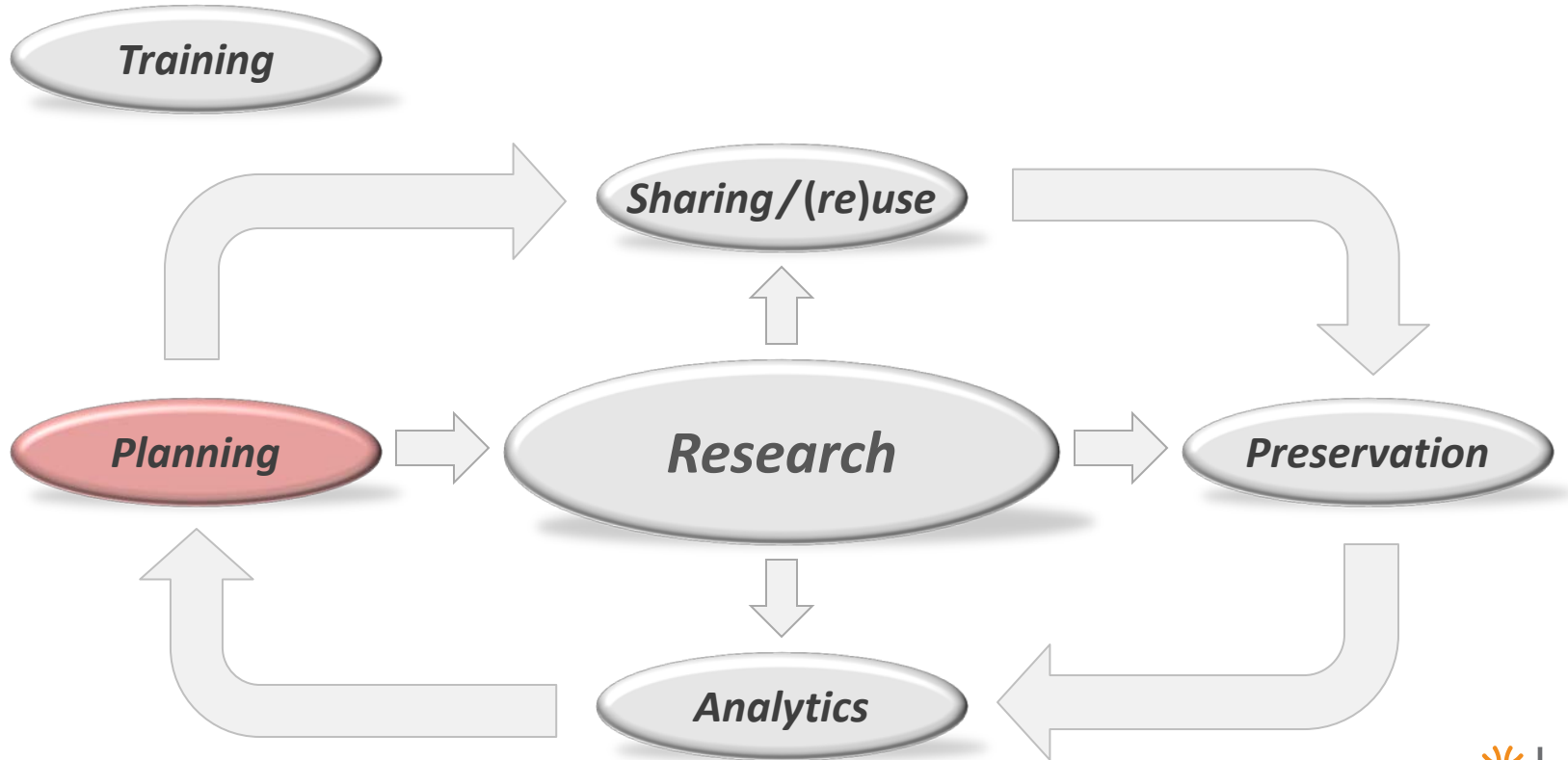
“Policies, strategies and actions that ensure access to digital content over time”

www.ala.org/alcts/resources/preserv/defdigpres0408

UC3's initiatives in research data management



Planning



Planning

Ideally, data management decisions should be planned before a research investigation starts

Formal data management plans (DMPs) are now required for funding proposals by all US federal agencies and many private foundations

- Even if not required, data management planning should be encouraged as a scholarly best practice
- It is better to be proactive, rather than reactive
- It is better to be deliberate, rather than ad hoc

DMPTool

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.

- » UNDERSTANDING THE ROLE OF PHYSICIAN INTEGRATION WITHIN NURSING HOMES IN POST-ACUTE CARE OUTCOMES
- » "A Microgravity-Themed Collaborative Intervention Promoting Student Selection of a STEM Career Pathway"
- » A Political Ecology of Value: A Cohort-Based Ethnography of the Environmental Turn in Nicaraguan Urban Social Policy

[View All](#)

DMPTOOL NEWS

Latest information about data management and the DMPTool.

- » Roll up, roll up. Get yer DMP update here!
- » NSF EAGER Grant for Actionable DMPs
- » DMPRoadmap summer camp news
- » On the right track(s) – DCC release dra...
- » RDA-DMP movings and shakings

[More News](#)

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

[View All](#)

dmptool.org

- *Create and share plans conforming to funder requirements*
- *Customized for public and private funders*
- *Customized with institutional resources and guidance*
- *Optional institutional review*
- *Public sample plans*

Next steps

roadmap

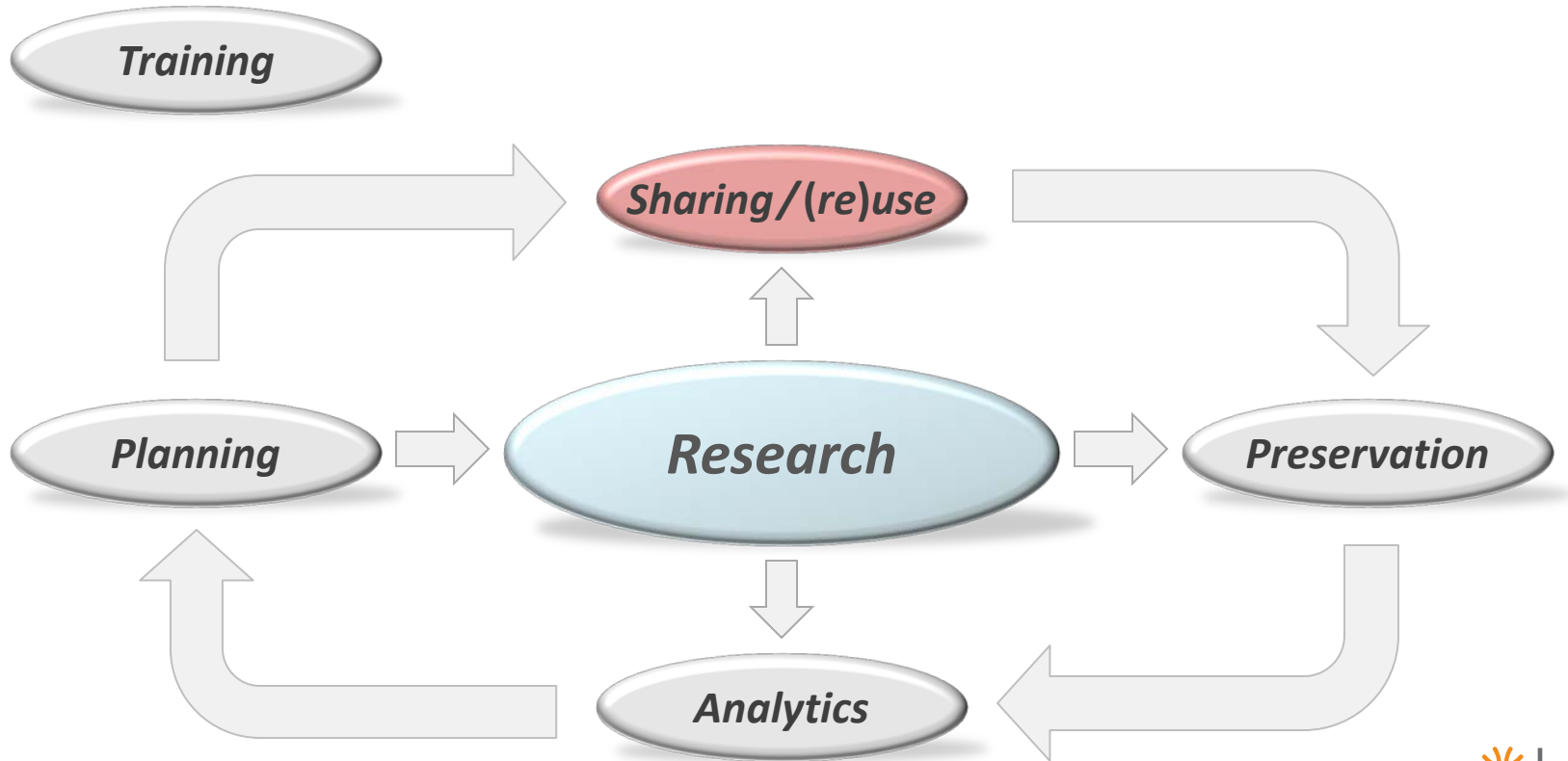
Collaborating with UK Digital Curation Centre (DCC) on DMP Roadmap, a consolidated internationalized platform

- blog.dmptool.org/2017/08/17/dmproadmap-summer-camp-news/
- github.com/DMPRoadmap/roadmap

Machine-actionable DMPs for improved capacity planning and compliance checking

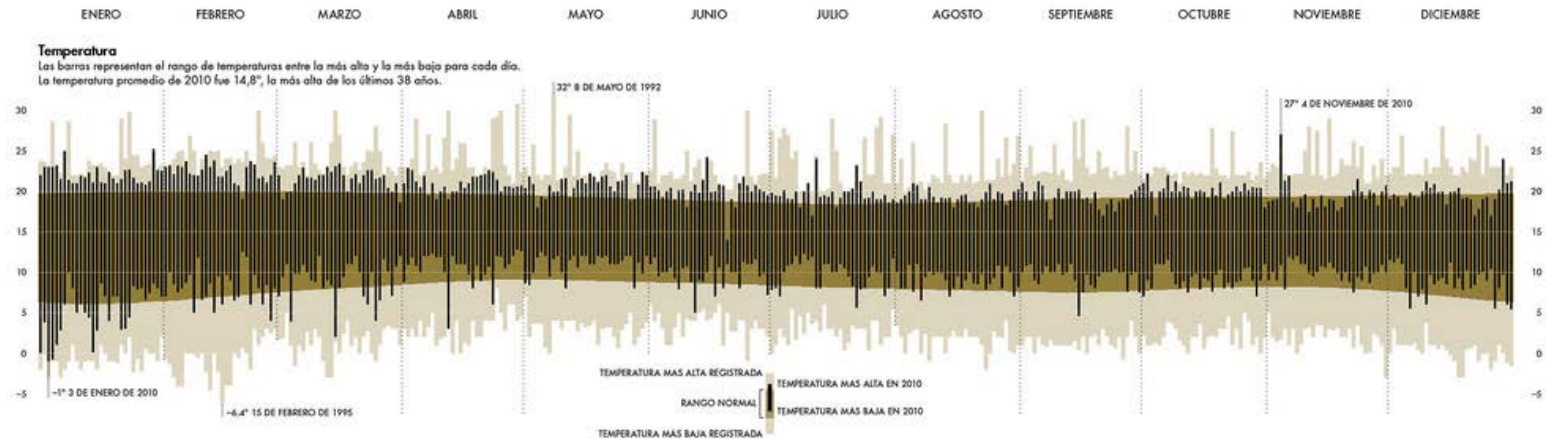
- blog.dmptool.org/2017/09/18/nsf-eager-grant-for-making-dmps-actionable/
- www.force11.org/group/fairdmp
- www.rd-alliance.org/groups/active-data-management-plans.html
- www.rd-alliance.org/groups/dmp-common-standards-wg

Sharing/(re)use



Sharing/(re)use

Data should be published so that they are available for review, re-analysis, and the starting point for new inquiry



The most significant step to making data reusable

Being managed by an appropriate curatorial program and system

- Minimally-curated data in a managed system can be enhanced over time
- Data that are not actively managed will become unusable sooner or later



www.flickr.com/photos/practicalowl/504522097

Dash data publication service

UNIVERSITY OF CALIFORNIA | dash

Search

Explore Data | Help | Login

An easy-to-use data publication service

"hypotheses come and go, but data remain"
Santiago Ramón y Cajal, 1897

Search 358 datasets

Search

Why share your data?

- Your journal requires it.
Many publishers — including PLOS, eLife, and Springer Nature — require authors to publish the data underlying papers in a suitable preservation repository.
- Your funder requires it.
Agencies that fund research — including all US federal agencies — are beginning to require grantees to publish data from funded projects.
- It's the right thing to do.
By offering clear description of your data including methods and instructions for reuse, you are promoting efficiency, transparency, and reproducibility in the research community.

dash.ucop.edu

- Self-service operation by researchers
- FORCE11 FAIR principles www.force11.org/group/fairgroup/fairprinciples
- DataCite DOIs and metadata schema.datacite.org/meta/kernel-4.0/
- Overlay layer sitting on top of any standards-compliant repository
- Multi-tenant UI
- Curatorial interface

Adoption of Dash has been slow

Problem

Everyone agrees on the benefit of data publication, but no one wants to do it if it means additional work

Solution

Integrate data publication as a side-effect of other activities that researchers are already doing, e.g., article publication

We're working to integrate Dash into journal publication workflows

Sharing/(re)use

Sharing and reuse occurs not only through open publication, but also between collaborators, and within laboratories



www.flickr.com/photos/uniformed_services_university/6306334990

Dat-in-the-Lab

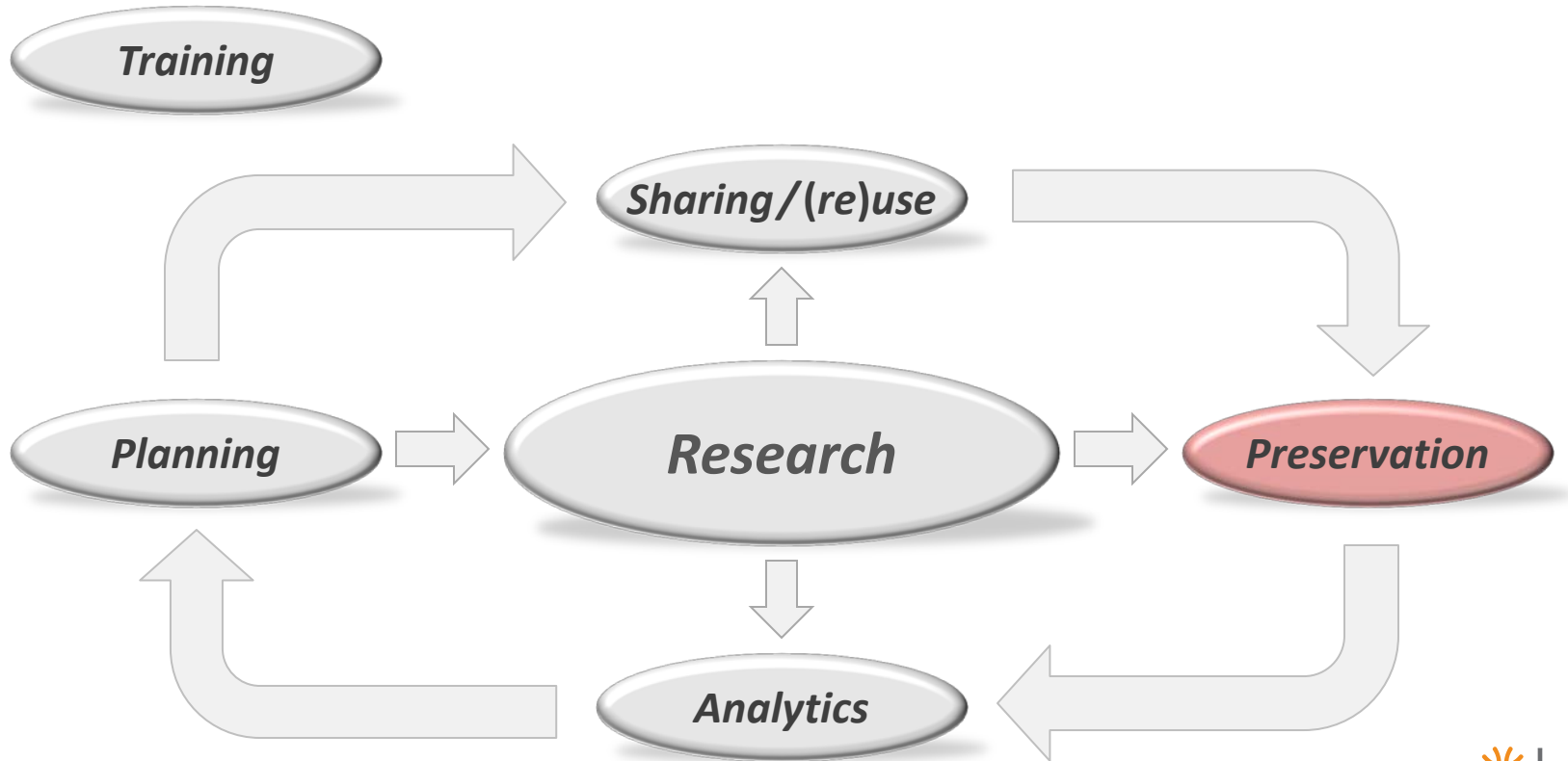
Prototype use of the Dat peer-to-peer data sharing technology (datproject.org) in two University of California research laboratories

Streamlining data workflows for research and publication, and afterwards

- *Collaboration between UC3 and Code for Science & Society (CSS) codeforscience.org*

uc3.cdlib.org/2017/09/27/moore-foundation-supports-uc3-research-data-management-project/

Preservation



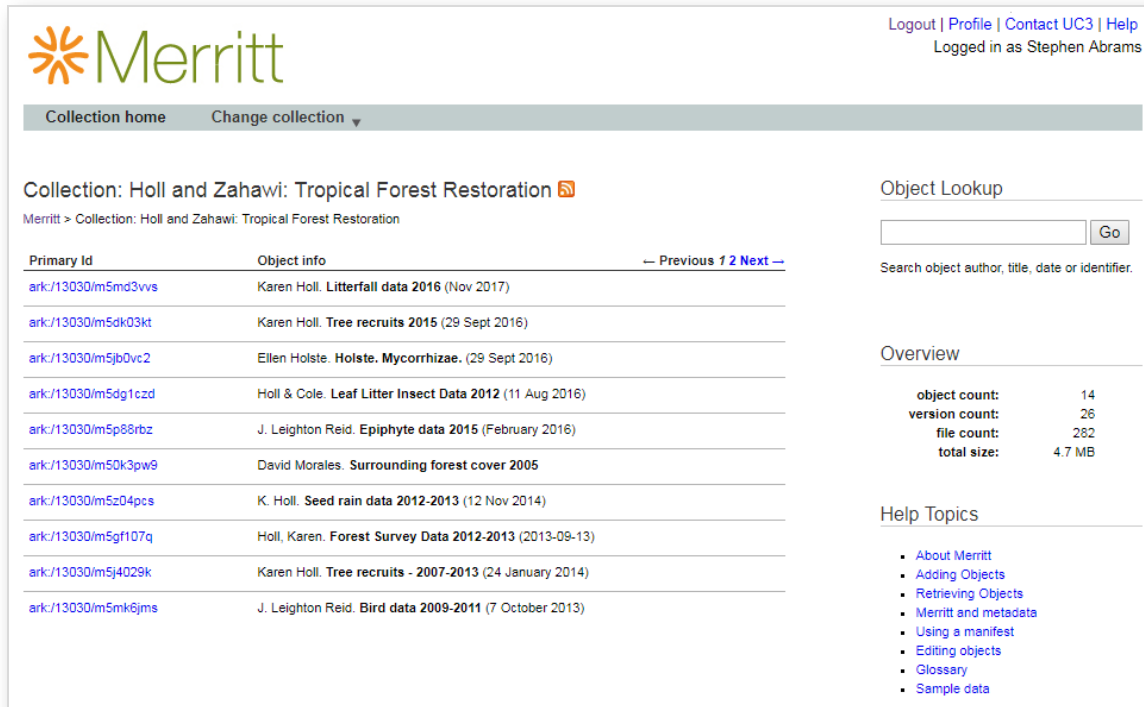
Preservation

Ensuring that data remain accessible and usable by scholars and researchers now and in the future




[commons.wikimedia.org/wiki/File:Documents_on_shelves_in_the_archive_store_\(History_Centre\)_at_the_Herbert.jpg](https://commons.wikimedia.org/wiki/File:Documents_on_shelves_in_the_archive_store_(History_Centre)_at_the_Herbert.jpg)

Merritt repository



The screenshot shows the Merritt repository interface. At the top left is the Merritt logo. To the right are links for 'Logout | Profile | Contact UC3 | Help' and a status 'Logged in as Stephen Abrams'. Below the logo is a navigation bar with 'Collection home' and 'Change collection'. The main content area displays the collection title 'Collection: Holl and Zahawi: Tropical Forest Restoration' and a table of objects. On the right side, there is an 'Object Lookup' search box, an 'Overview' section with statistics, and a 'Help Topics' section with a list of links.

Collection: Holl and Zahawi: Tropical Forest Restoration 

Merritt > Collection: Holl and Zahawi: Tropical Forest Restoration

Primary Id	Object info	— Previous 1 2 Next —
ark:/13030/m5md3vvs	Karen Holl. Litterfall data 2016 (Nov 2017)	
ark:/13030/m5dk03kt	Karen Holl. Tree recruits 2015 (29 Sept 2016)	
ark:/13030/m5jb0vc2	Ellen Holste. Holste. Mycorrhizae , (29 Sept 2016)	
ark:/13030/m5dg1czd	Holl & Cole. Leaf Litter Insect Data 2012 (11 Aug 2016)	
ark:/13030/m5p88rbz	J. Leighton Reid. Epiphyte data 2015 (February 2016)	
ark:/13030/m50k3pw9	David Morales. Surrounding forest cover 2005	
ark:/13030/m5z04pcs	K. Holl. Seed rain data 2012-2013 (12 Nov 2014)	
ark:/13030/m5gf107q	Holl, Karen. Forest Survey Data 2012-2013 (2013-09-13)	
ark:/13030/m5j4029k	Karen Holl. Tree recruits - 2007-2013 (24 January 2014)	
ark:/13030/m5mk6jms	J. Leighton Reid. Bird data 2009-2011 (7 October 2013)	

Object Lookup

Search object author, title, date or identifier.

Overview

object count: 14
version count: 26
file count: 282
total size: 4.7 MB

Help Topics

- [About Merritt](#)
- [Adding Objects](#)
- [Retrieving Objects](#)
- [Merritt and metadata](#)
- [Using a manifest](#)
- [Editing objects](#)
- [Glossary](#)
- [Sample data](#)

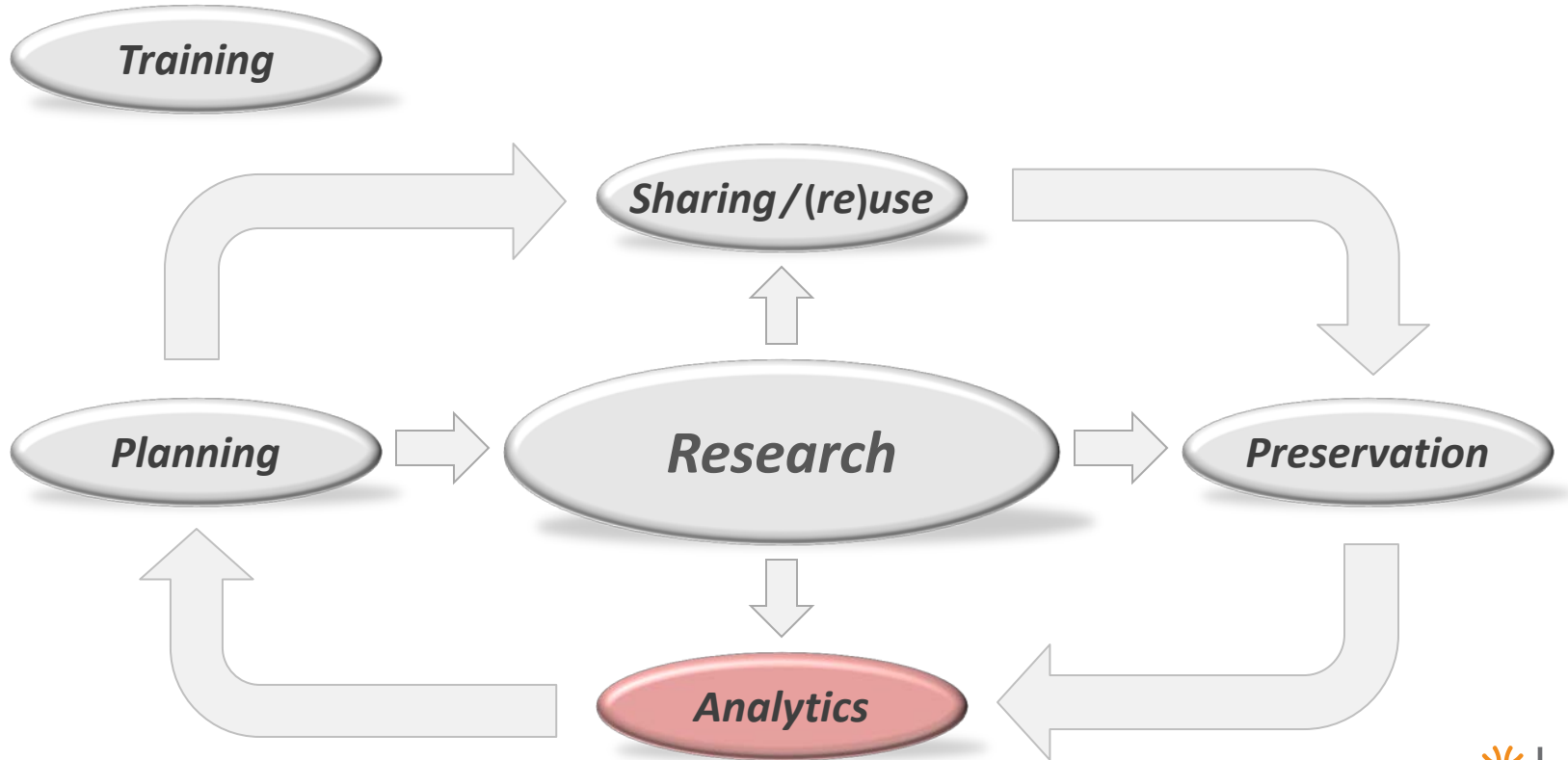
■ Currently used primarily for cultural heritage material

■ But with growing data collections, including all Dash datasets

■ Integrated with the DataONE network www.dataone.org

■ CoreTrustSeal self-audit underway www.coretrustseal.org

Analytics

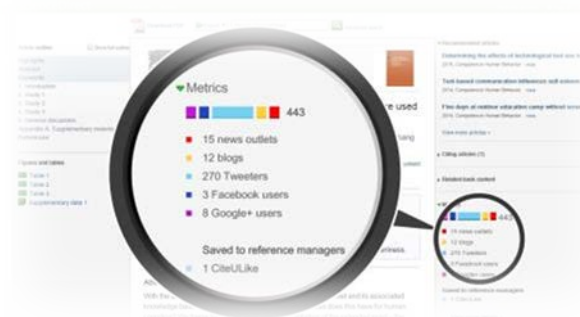


Analytics

If data are to be considered first-class research outputs alongside traditional publications, it is important to quantify their impact

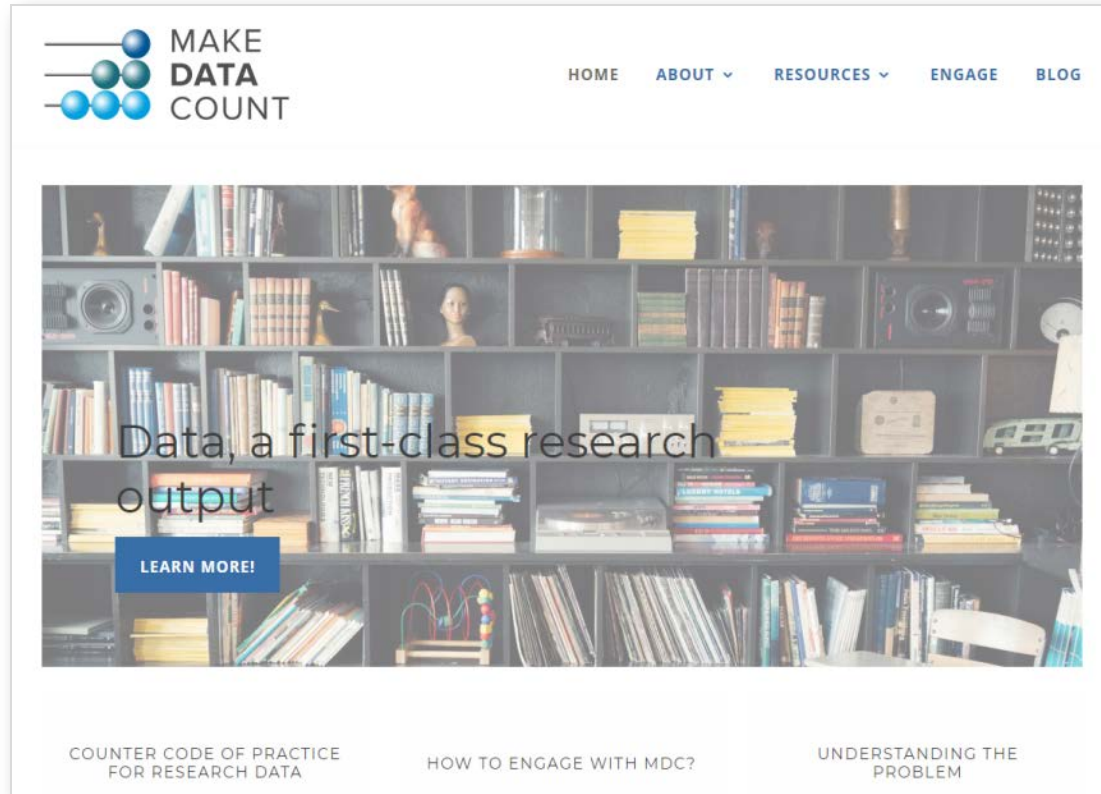
We need an infrastructure for tracking and reporting usage similar to that in place for the published literature

- Kratz and Strasser (2015), “Making data count,” *Scientific Data* 2 [dx.doi.org/10.1038/sdata.2015.39](https://doi.org/10.1038/sdata.2015.39)



blog.sciencedirect.com/topics/altmetrics

Make Data Count (MDC)

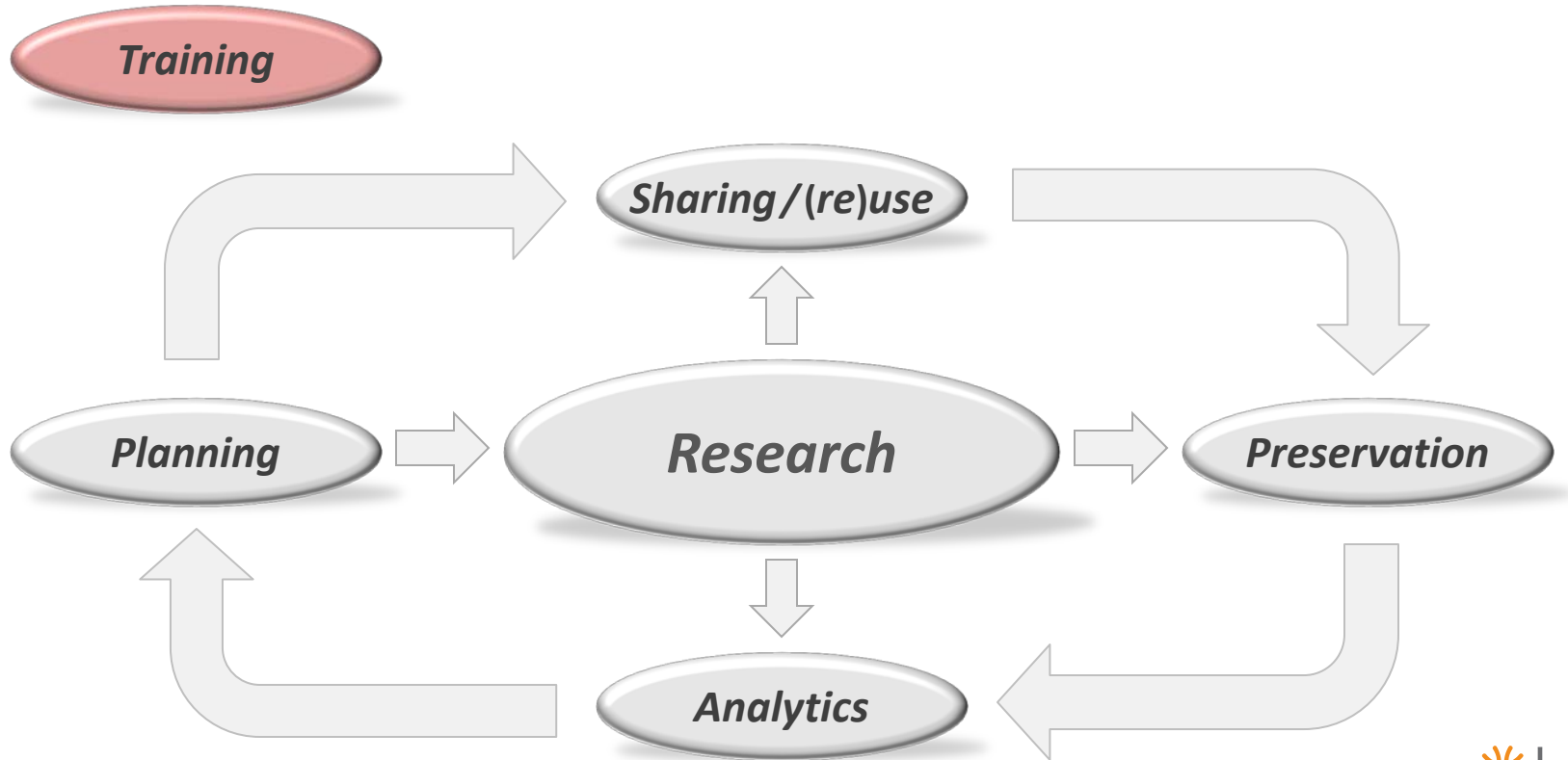


The screenshot shows the homepage of the Make Data Count website. At the top left is the logo, which consists of four blue circles of varying sizes arranged in a cluster, followed by the text "MAKE DATA COUNT". To the right of the logo is a navigation menu with the following items: "HOME", "ABOUT" (with a dropdown arrow), "RESOURCES" (with a dropdown arrow), "ENGAGE", and "BLOG". Below the navigation is a large banner image of a bookshelf filled with books and decorative objects. Overlaid on the bookshelf is the text "Data, a first-class research output" in a large, white, sans-serif font. Below this text is a blue button with the white text "LEARN MORE!". At the bottom of the page, there are three columns of text: "COUNTER CODE OF PRACTICE FOR RESEARCH DATA", "HOW TO ENGAGE WITH MDC?", and "UNDERSTANDING THE PROBLEM".

- *Collaboration between UC3, DataONE, and DataCite*
- *COUNTER code of practice for data-level metrics (DLM)*
www.projectcounter.org
- *Extending DataCite/ Crossref EventData to support DLM*
www.datacite.org/eventdata.html

makedatacount.org

Data literacy training



Data literacy training



Most scholars and researchers have never received any data literacy training

They do not view the library as the natural place to turn for advice and guidance



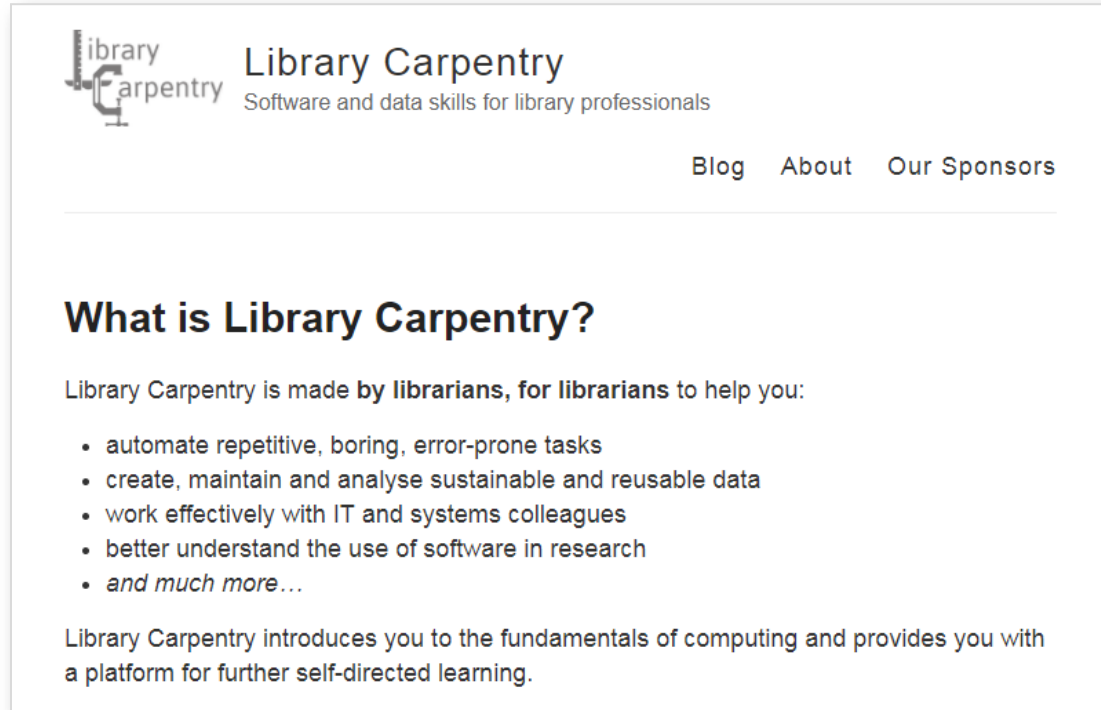
www.flickr.com/photos/rosengrant/3847862890

Self-assessment maturity guide

Practice Maturity 				
Planning for data	When it comes to my data, I have an informal "way of doing things" but not a formal plan.	I have a formal DMP document that outlines how I plan collect, manage, and save my data.	I have an active data management plan that is revisited throughout my project's lifecycle.	My data management plan is the hub of all of my research activity.
Saving data	I save my data only on my local machine(s).	I save my data on an external hard drive, server, or in the cloud.	I save my data in multiple locations.	
Organizing data	I have identified the data I need to keep organized, including data types, sources, and file formats.	I apply consistent naming and structuring schemes to all of the files associated with my data.	I apply community or discipline-specific schemes when naming and structuring my files.	I work to ensure my data exists in a form that is interoperable and suitable for long term preservation.
Preparing data	I format my data consistently. I use the same units and formats across variables.	I document the format and structure of my data in a data dictionary, codebook, or readme.	I describe my data using a discipline-appropriate metadata schema.	Because of my well-defined organizational practices, my data is <i>already</i> prepared for analysis.
Analyzing data	I keep notes on the parameters, procedures, and protocols applied throughout my data analysis workflow.	I maintain a lab notebook that documents the specifics of my analysis workflow as well as my decision making process.	My protocol, lab notebook, or analysis workflow is collocated with the results of my analyses.	
Sharing data	I communicate my data via tables and figures in a poster, presentation, or paper.	Any description of my data includes either a data availability statement or my data as a supplementary material.	I deposit my data in a database, repository, or system that provides a persistent identifier.	
Ad hoc, Non-reproducible 				
Optimized, Standardized				

- *Intended to assess individual researchers, not institutions*
- *Informative, not prescriptive*

Library carpentry



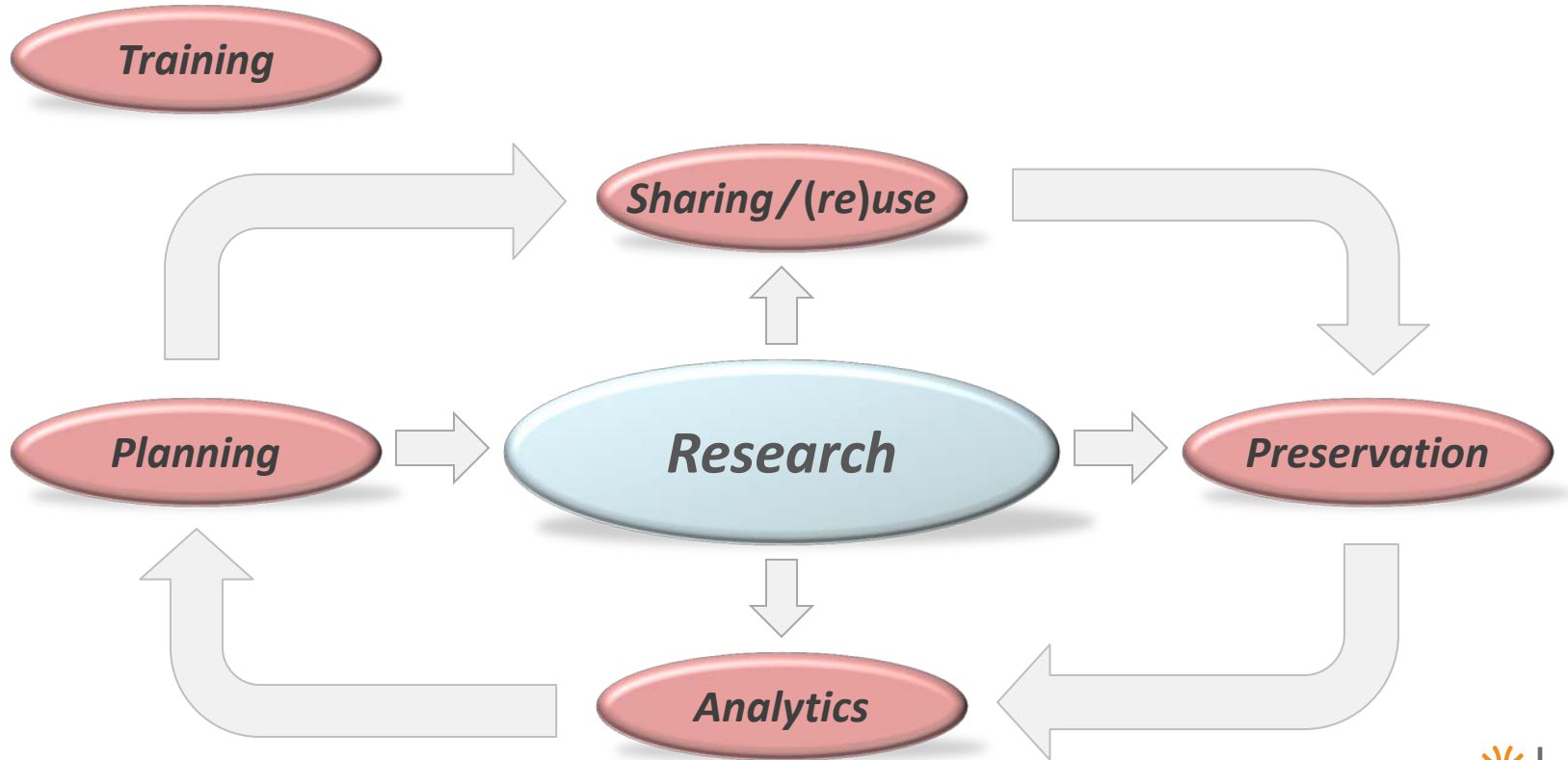
The screenshot shows the Library Carpentry website. At the top left is the logo, which consists of the words 'library' and 'carpentry' stacked vertically, with a stylized 'C' that incorporates a hammer and chisel. To the right of the logo, the text reads 'Library Carpentry' in a large font, followed by 'Software and data skills for library professionals' in a smaller font. On the right side of the page, there are three navigation links: 'Blog', 'About', and 'Our Sponsors'. Below the navigation links is a horizontal line. The main heading is 'What is Library Carpentry?'. Below this heading is a paragraph: 'Library Carpentry is made by librarians, for librarians to help you:'. This is followed by a bulleted list of five items: '• automate repetitive, boring, error-prone tasks', '• create, maintain and analyse sustainable and reusable data', '• work effectively with IT and systems colleagues', '• better understand the use of software in research', and '• and much more...'. At the bottom of the page, there is another paragraph: 'Library Carpentry introduces you to the fundamentals of computing and provides you with a platform for further self-directed learning.'

- *Increase awareness, develop new training modules, and coordinate national activities*

librarycarpentry.github.io

uc3.cdlib.org/2017/11/06/skills-training-for-librarians-expanding-library-carpentry/

Supporting research data management



Supporting research data management

at the University of California

UC Curation Center
California Digital Library
University of California

www.cdlib.org/uc3

uc3.cdlib.org

uc3@ucop.edu

[@uc3cdl](https://twitter.com/uc3cdl)

Stephen.Abrams@ucop.edu

ORCID 0000-0003-2326-6672

