Title
Supporting Research Data Management at the University of California

Permalink
https://escholarship.org/uc/item/86p3x8hw

Author
Abrams, Stephen

Publication Date
2017-12-12

Kyoto University, 12 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

UC3
UC Curation Center

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

ucop.webdamdb.com/bp/#/folder/201490/

UC3
UC Curation Center

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

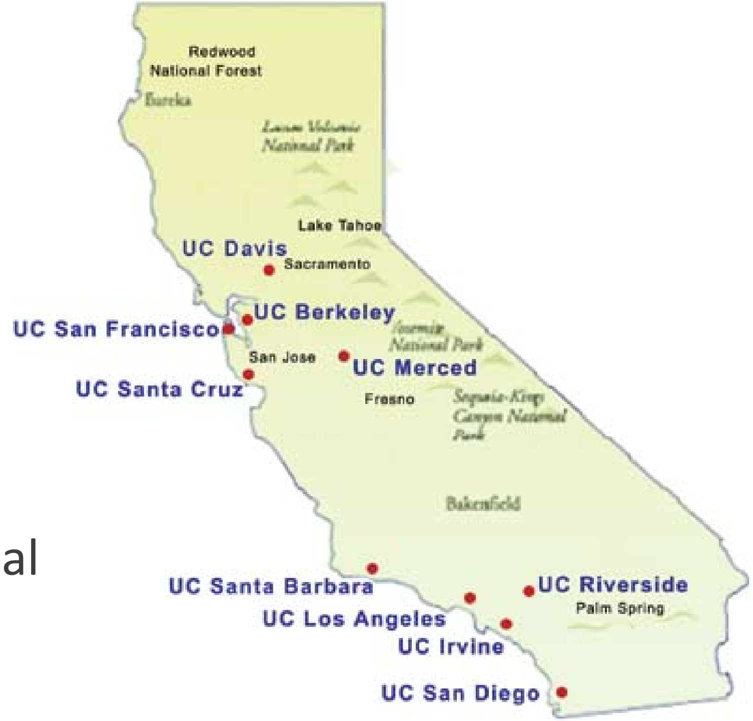
www.flickr.com/photos/janoma/4754744617

UC3
UC Curation Center

University of California

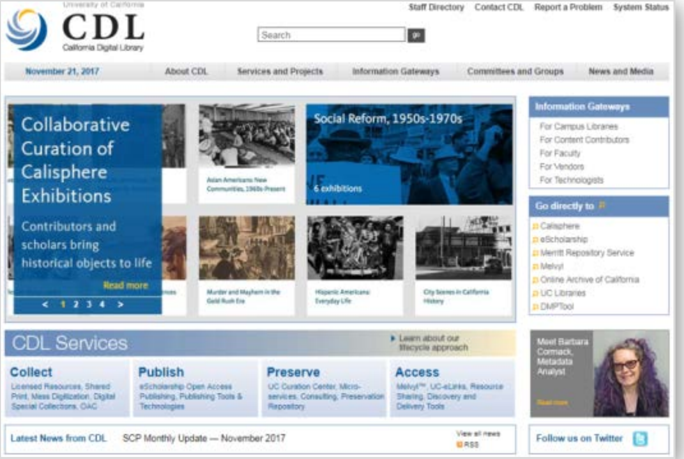
A large and diverse research University system

- 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



University of California
CDL
California Digital Library

Staff Directory Contact CDL Report a Problem System Status

Search go

November 21, 2017 | About CDL | Services and Projects | Information Gateways | Committees and Groups | News and Media

Collaborative Curation of Calisphere Exhibitions

Contributors and scholars bring historical objects to life

Read more

< 1 2 3 4 >

Asian Americans New Communities, 1960s-Present

Social Reform, 1950s-1970s

6 exhibitions

Murder and Mayhem in the Gold Rush Era

Hispanic Americans Everyday Life

City Scenes in California History

Information Gateways

- For Campus Libraries
- For Content Contributors
- For Faculty
- For Vendors
- For Technologists

Go directly to

- Calisphere
- eScholarship
- Merritt Repository Service
- Melvyl
- Online Archive of California
- UC Libraries
- DMPTool

CDL Services

Learn about our lifecycle approach

Collect
Licensed Resources, Shared Print, Mass Digitization, Digital Special Collections, OAC

Publish
eScholarship Open Access Publishing, Publishing Tools & Technologies

Preserve
UC Curation Center, Micro-services, Consulting, Preservation Repository

Access
Melvyl™, UC-eLinks, Resource Sharing, Discovery and Delivery Tools

Meet Barbara Cormack, Metadata Analyst

Read more

Latest News from CDL SCP Monthly Update — November 2017

View all news
RSS

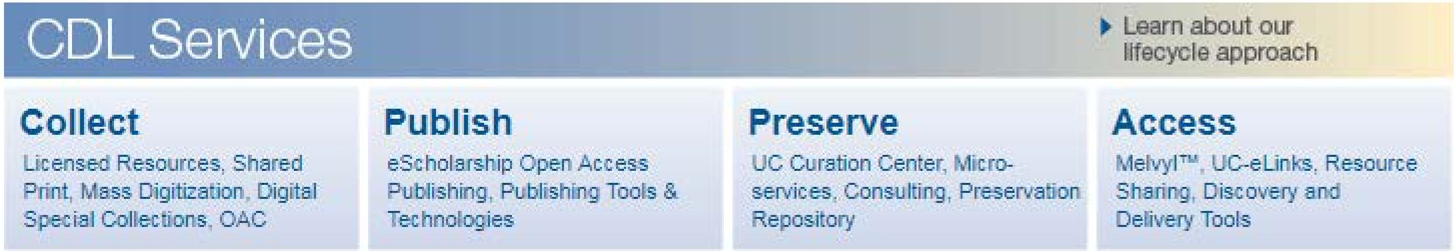
Follow us on Twitter

www.cdlib.org

UC3
UC Curation Center

Four CDL programs

- Bibliographic services
- Open access publishing and special collections
- Licensing and mass digitization
- Curation and preservation (UC Curation Center)



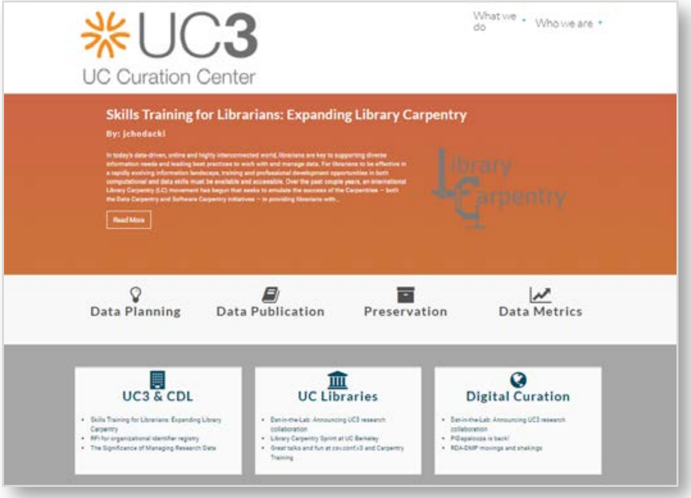
CDL Services

▶ Learn about our lifecycle approach

Collect	Publish	Preserve	Access
Licensed Resources, Shared Print, Mass Digitization, Digital Special Collections, OAC	eScholarship Open Access Publishing, Publishing Tools & Technologies	UC Curation Center, Micro-services, Consulting, Preservation Repository	Melvyl™, UC-eLinks, Resource Sharing, Discovery and Delivery Tools

UC Curation Center (UC3)

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



Skills Training for Librarians: Expanding Library Carpentry

By: jchodacki

In today's data-driven, online and highly interconnected world, librarians are key to supporting diverse information needs and leading best practices to work with and manage data. For librarians to be effective in a rapidly evolving information landscape, training and professional development opportunities in both computational and data skills must be available and accessible. Over the past couple years, an international Library Carpentry (LC) movement has begun that seeks to emulate the success of the Carpentries – both the Data Carpentry and Software Carpentry initiatives – in providing librarians with...

UC3 & CDL

- Skills Training for Librarians: Expanding Library Carpentry
- RFI for organizational identifier registry
- The Significance of Managing Research Data

UC Libraries

- Dat-in-the-Lab: Announcing UC3 research collaboration
- Library Carpentry Sprint at UC Berkeley
- Great talks and fun at csv,conf,v3 and Carpentry Training

Digital Curation

- Dat-in-the-Lab: Announcing UC3 research collaboration
- PIDapalooza is back!
- RDA-DMP movings and shakings

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

UC3
UC Curation Center

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

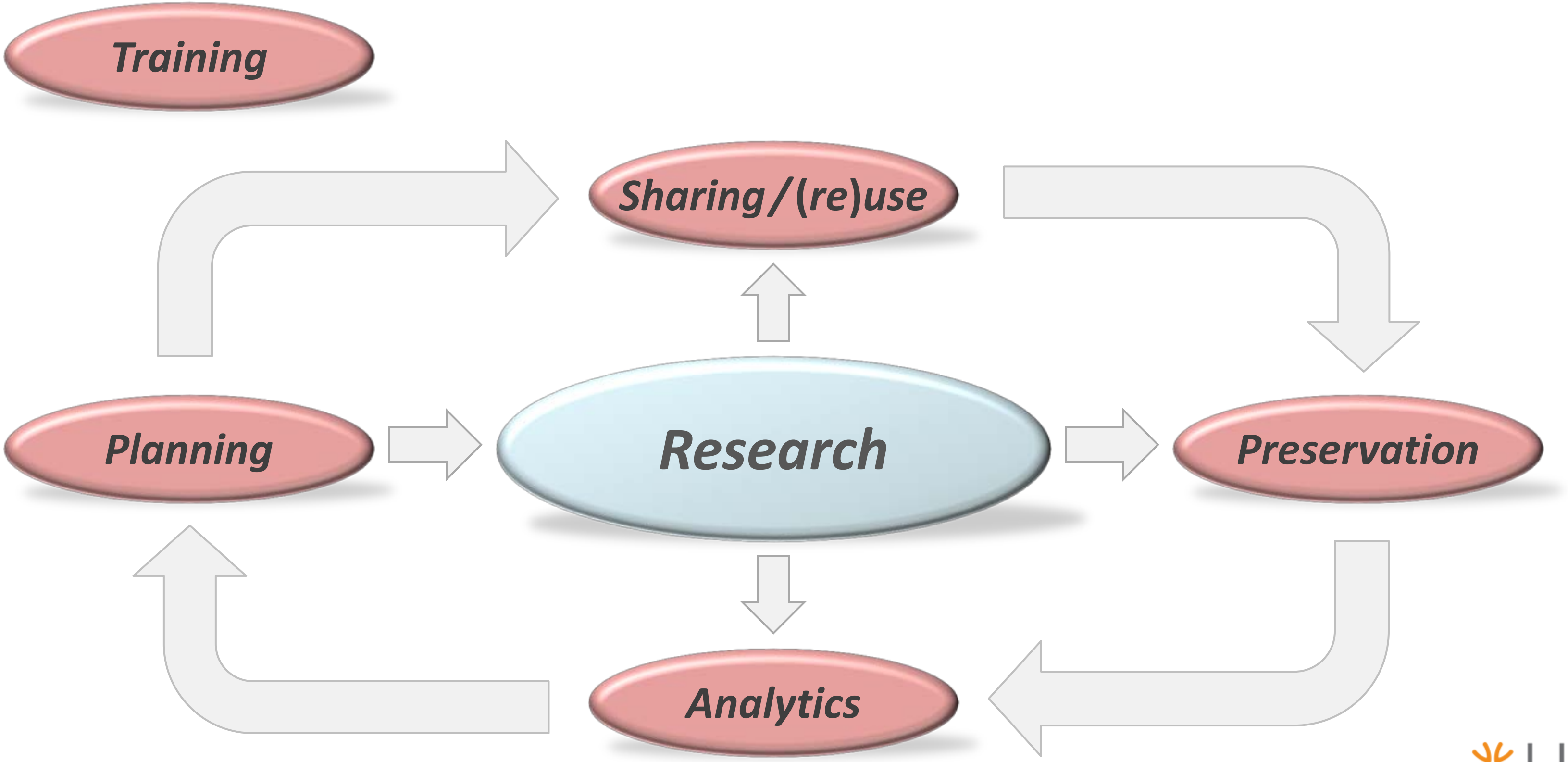
Preservation

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

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

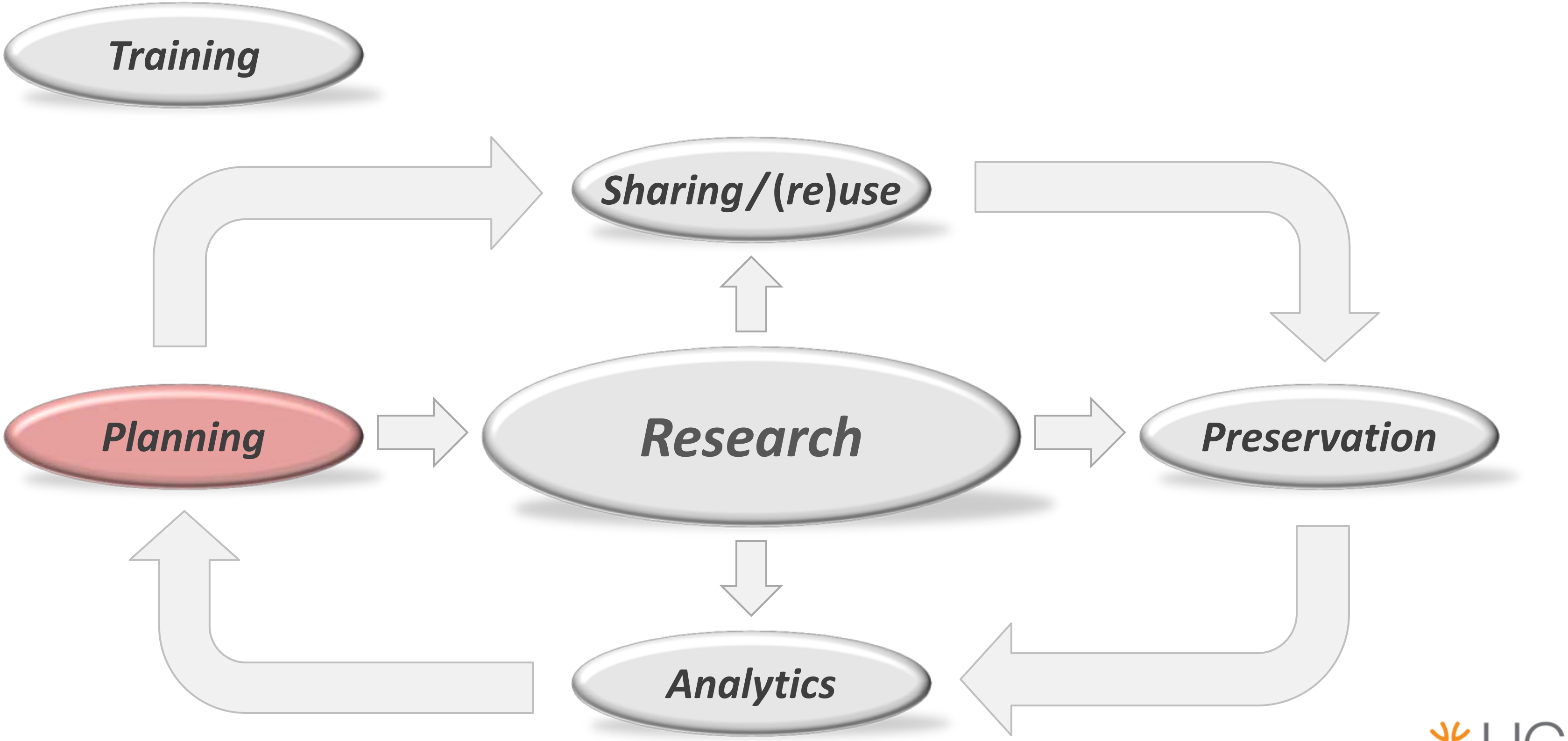
UC3
UC Curation Center

UC3's initiatives in research data management



UC3
UC Curation Center

Planning



UC3
UC Curation Center

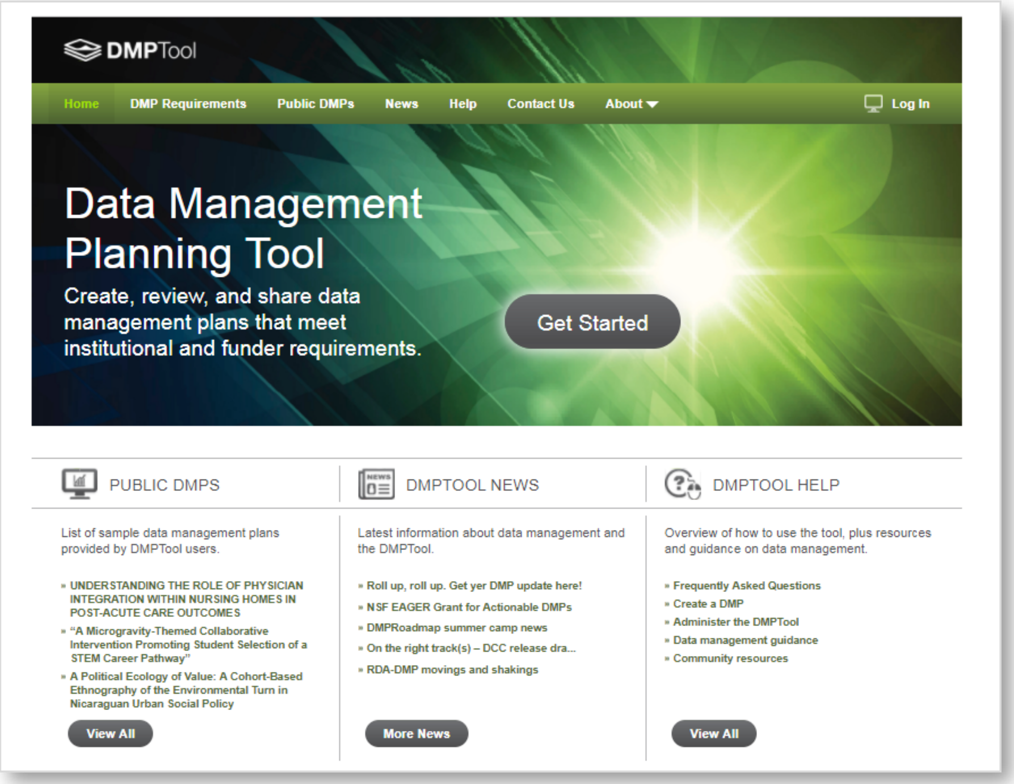
Planning

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



Data Management Planning Tool

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

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

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

DMPTOOL HELP

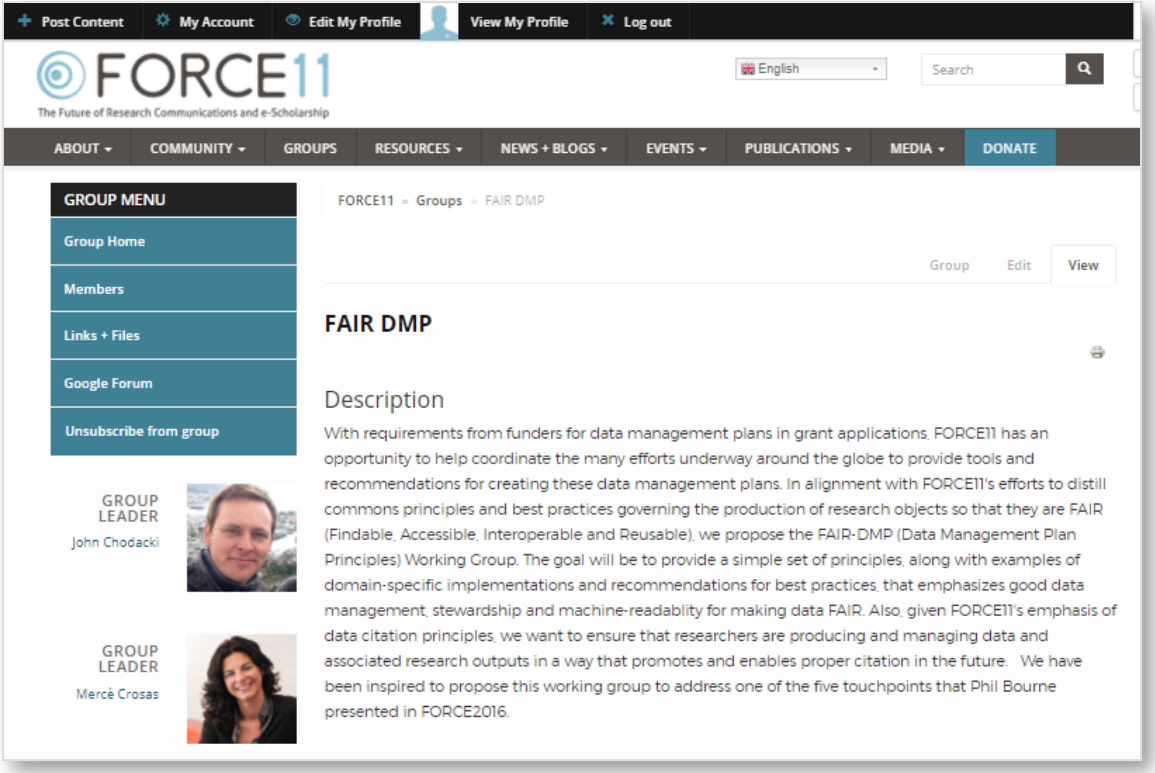
Overview of how to use the tool, plus resources and guidance on data management.

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

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*

FAIR DMP



FAIR DMP

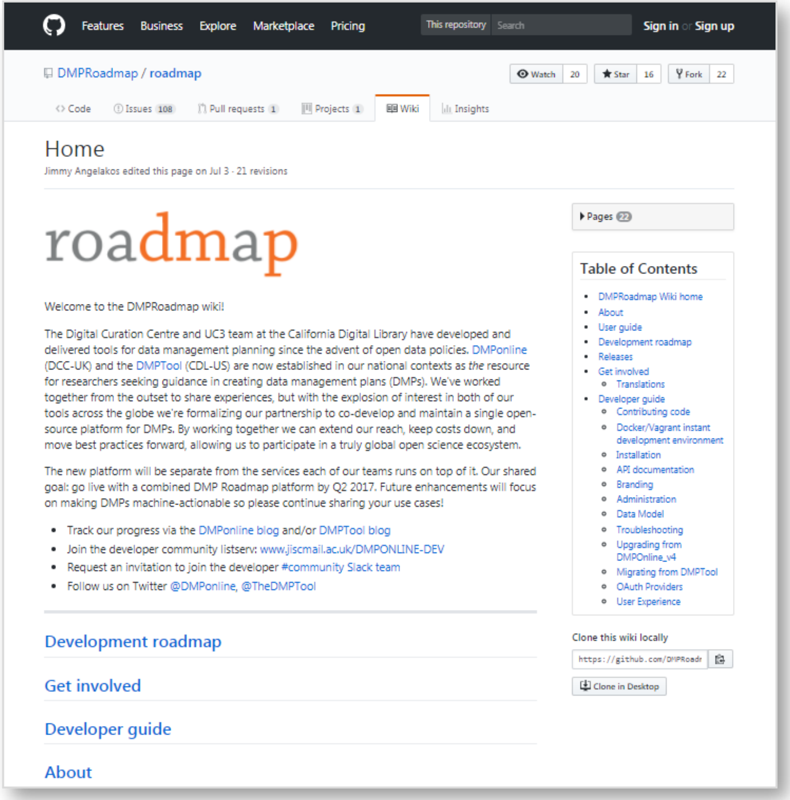
Description

With requirements from funders for data management plans in grant applications, FORCE11 has an opportunity to help coordinate the many efforts underway around the globe to provide tools and recommendations for creating these data management plans. In alignment with FORCE11's efforts to distill commons principles and best practices governing the production of research objects so that they are FAIR (Findable, Accessible, Interoperable and Reusable), we propose the FAIR-DMP (Data Management Plan Principles) Working Group. The goal will be to provide a simple set of principles, along with examples of domain-specific implementations and recommendations for best practices, that emphasizes good data management, stewardship and machine-readablity for making data FAIR. Also, given FORCE11's emphasis of data citation principles, we want to ensure that researchers are producing and managing data and associated research outputs in a way that promotes and enables proper citation in the future. We have been inspired to propose this working group to address one of the five touchpoints that Phil Bourne presented in FORCE2016.

www.force11.org/group/fairdmp

- *Best practices for data management planning that will produce FAIR data*
- *Findable, accessible, interoperable, reusable*

DMP/Roadmap



Home

Jimmy Angelakos edited this page on Jul 3 · 21 revisions

roadmap

Welcome to the DMPRoadmap wiki!

The Digital Curation Centre and UC3 team at the California Digital Library have developed and delivered tools for data management planning since the advent of open data policies. DMPonline (DCC-UK) and the DMPTool (CDL-US) are now established in our national contexts as *the* resource for researchers seeking guidance in creating data management plans (DMPs). We've worked together from the outset to share experiences, but with the explosion of interest in both of our tools across the globe we're formalizing our partnership to co-develop and maintain a single open-source platform for DMPs. By working together we can extend our reach, keep costs down, and move best practices forward, allowing us to participate in a truly global open science ecosystem.

The new platform will be separate from the services each of our teams runs on top of it. Our shared goal: go live with a combined DMP Roadmap platform by Q2 2017. Future enhancements will focus on making DMPs machine-actionable so please continue sharing your use cases!

- Track our progress via the DMPonline blog and/or DMPTool blog
- Join the developer community listserv: www.jiscmail.ac.uk/DMPONLINE-DEV
- Request an invitation to join the developer #community Slack team
- Follow us on Twitter @DMPonline, @TheDMPTool

Development roadmap

Get involved

Developer guide

About

Table of Contents

Clone this wiki locally

https://github.com/DMPRoadr

- *Collaboration with UK Digital Curation Centre*
- *Common platform consolidating DMPTool and DMPonline*
- *Internationalization*
- *github.com/DMPRoadmap/roadmap*

blog.dmptool.org/2017/08/17/dmproadmap-summer-camp-news/

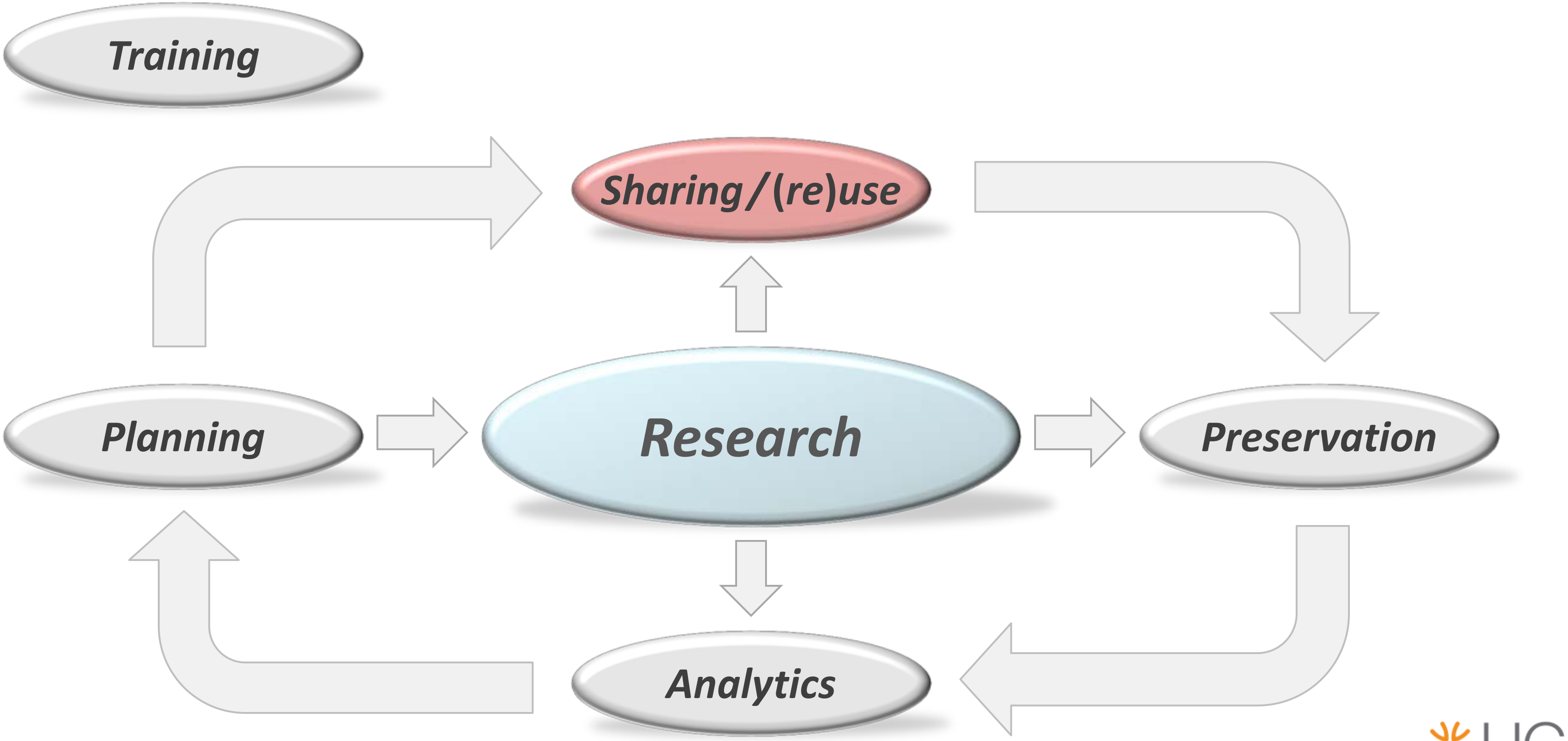
maDMPs – Machine-actionable DMPs

NSF-funded project to explore the transformation of DMPs from static documents into ““information structured in a consistent way so that machines, or computers, can be programmed against the structure”

- *Capacity planning*
- *Compliance checking*
- *FORCE11 FAIR DMP*
- *RDA Active DMPs interest group and Common Standards working group*

blog.dmptool.org/2017/09/18/nsf-eager-grant-for-making-dmps-actionable/

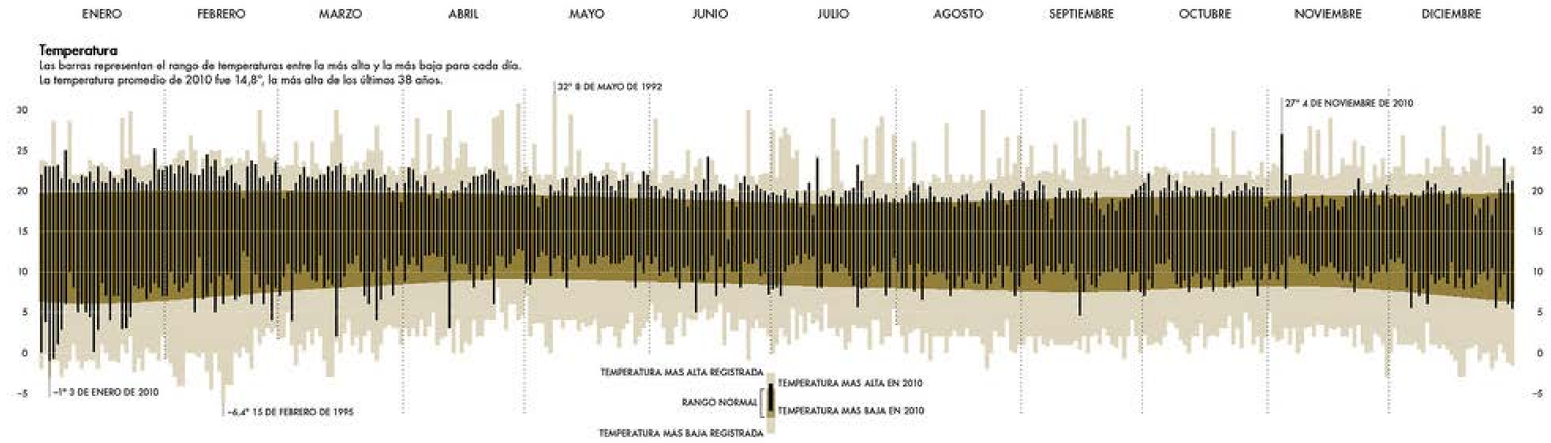
Sharing/(re)use



UC3
UC Curation Center

Sharing/(re)use

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



www.flickr.com/photos/juglardelzipa/5582942966

What makes data reusable?

- Findable
- Accessible
- Interoperable
- Reusable

www.force11.org/group/fairgroup/fairprinciples

The most significant step to making data reusable

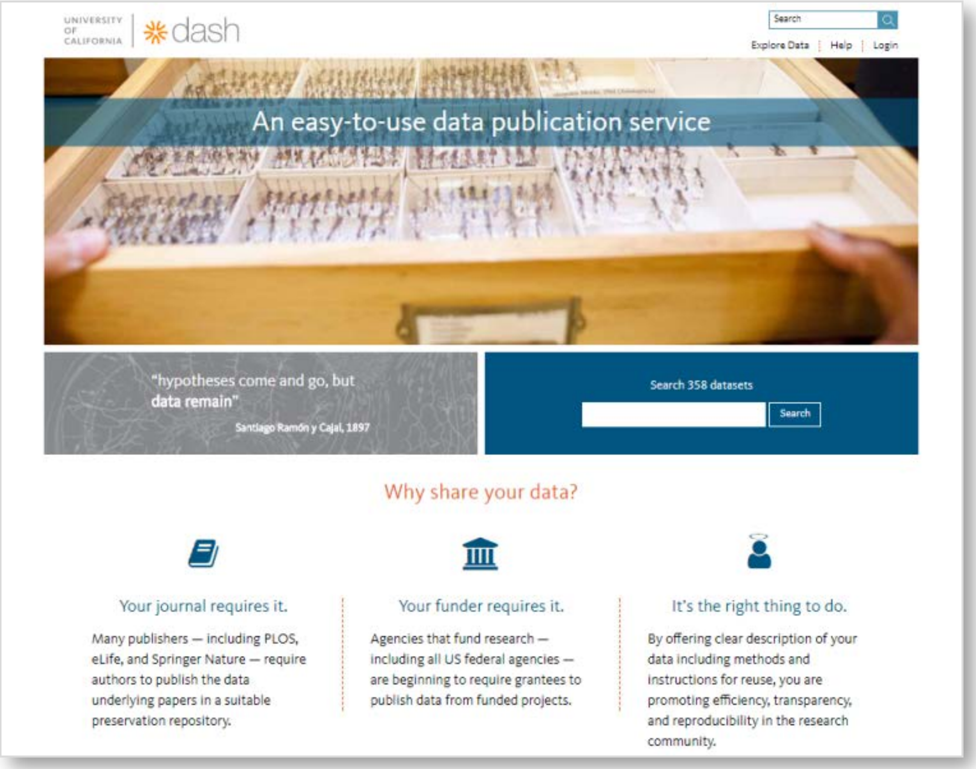
Being managed by an appropriate curatorial program and system

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

www.flickr.com/photos/practicalowl/504522097

UC3
UC Curation Center

Dash data publication service



An easy-to-use data publication service

"hypotheses come and go, but data remain"

Santiago Ramón y Cajal, 1897

Search 358 datasets

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*
- *DataCite metadata* *schema.datacite.org/meta/kernel-4.0/*
- *Overlay layer sitting on top of any standards-compliant repository*
- *Multi-tenant UI*
- *Curatorial interface*

Dash data publication service

- Findable
  - DOIs assigned and indexed with DataCite *www.datacite.org*
  - Authors/contributors, title, abstract, methods, usage notes, funders, geospatial locations
  - Formatted citations conforming to Joint Declaration of Data Citation Principles *www.force11.org/datacitationprinciples*
- Accessible
- Interoperable
- Reusable

Dash data publication service

- Findable
- Accessible
  - DOIs point to permanent dataset landing pages
  - Download data or data paper
  - Data download can be embargoed, while metadata remains accessible
- Interoperable
- Reusable

Dash data publication service

- Findable
- Accessible
- Interoperable
  - DataCite metadata *schema.datacite.org*
  - Optional references to citing articles and related datasets and data packages
- Reusable

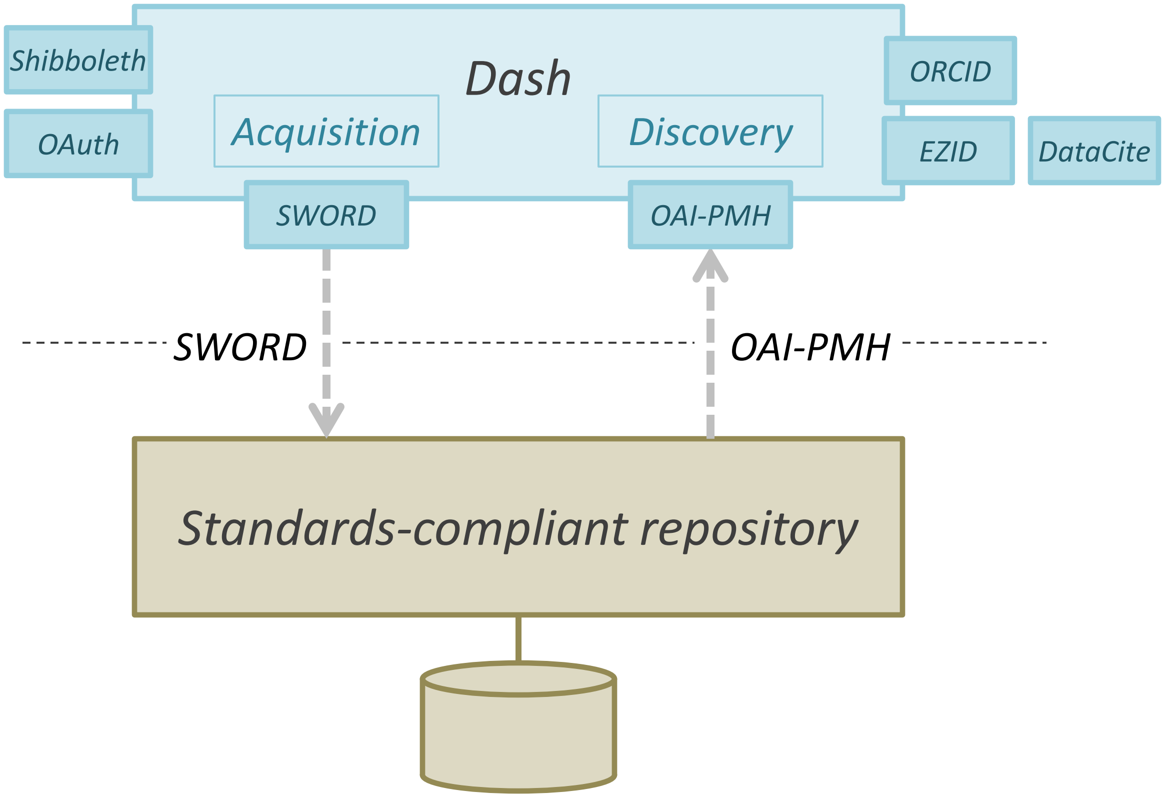
Dash data publication service

- Findable
- Accessible
- Interoperable
- Reusable
  - CC0 or CC-BY licensing
  - Complete version history

Additional Dash features

- Multi-tenant UI with institutional branding
- Login with ORCiD or institutional credentials
- Upload via drag-and-drop or from external hosting sites, e.g., Box, Dropbox, Google Drive, laboratory server, etc.
- Curation interface to add value to managed data
- Usage metrics

Dash overlay architecture



- *Ruby/Rails application*
- *Loosely coupled to an underlying repository*
- *Communication via standard protocols*
- *Can be integrated with any standards-compliant repository*

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/

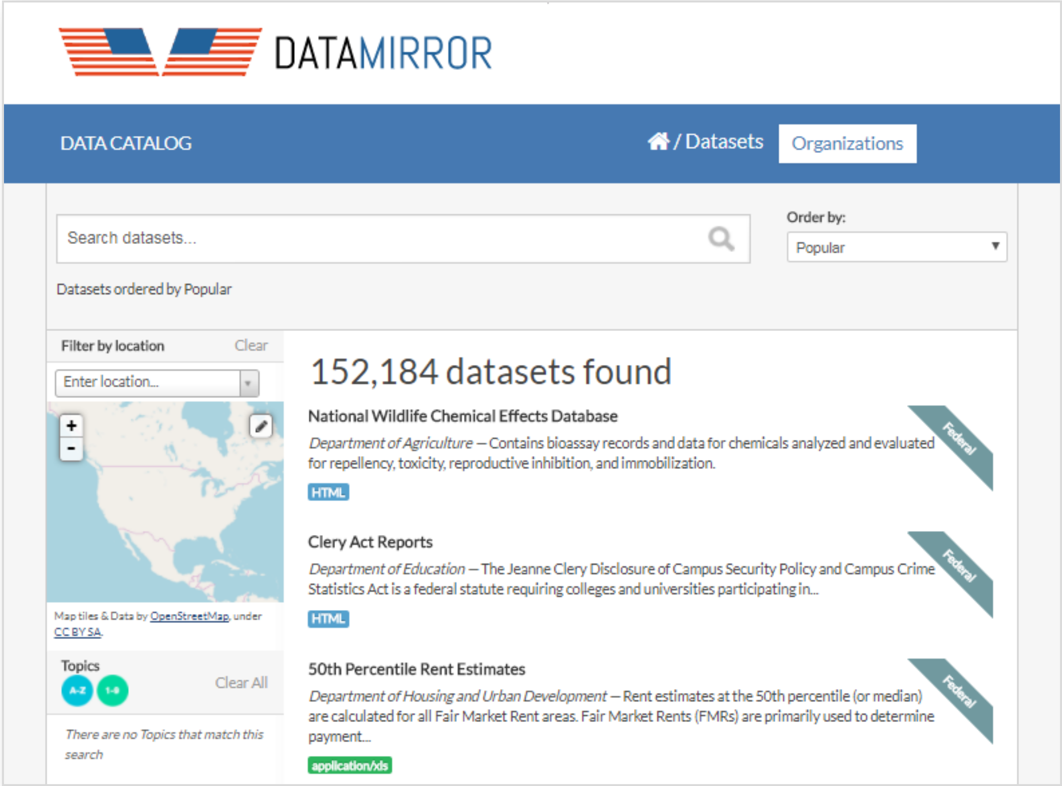
Sharing/(re)use

Federal government agencies and laboratories perform about 11% of all research in the US

The *data.gov* open data portal provides access to over 197,000 (42 TB) government datasets

The ongoing commitment to sustain this access is unclear

Datamirror



Datasets ordered by Popular

Map tiles & Data by OpenStreetMap, under CC BY SA.

There are no Topics that match this search

152,184 datasets found

National Wildlife Chemical Effects Database
Department of Agriculture — Contains bioassay records and data for chemicals analyzed and evaluated for repellency, toxicity, reproductive inhibition, and immobilization.
HTML
Federal

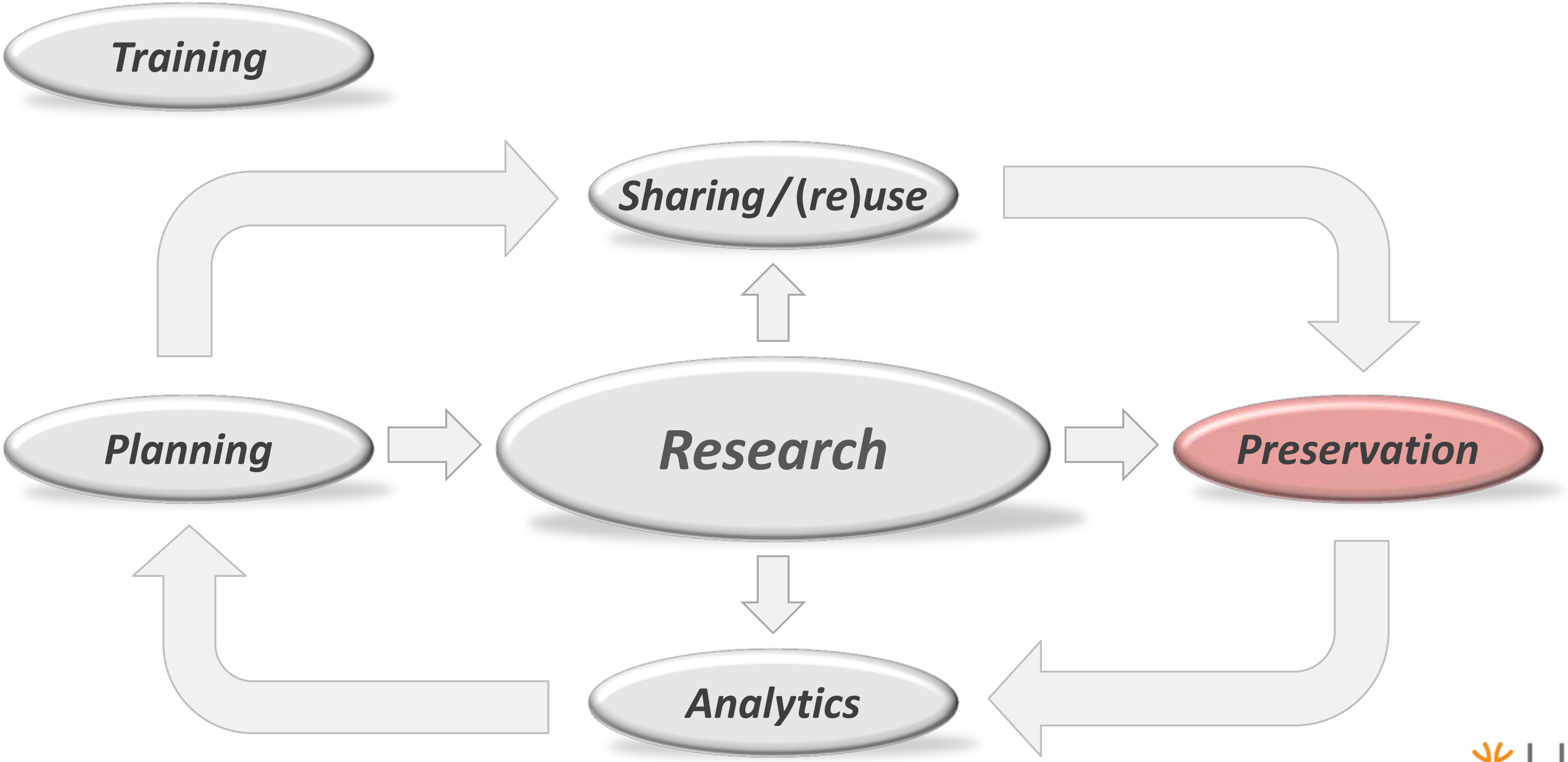
Clery Act Reports
Department of Education — The Jeanne Clery Disclosure of Campus Security Policy and Campus Crime Statistics Act is a federal statute requiring colleges and universities participating in...
HTML
Federal

50th Percentile Rent Estimates
Department of Housing and Urban Development — Rent estimates at the 50th percentile (or median) are calculated for all Fair Market Rent areas. Fair Market Rents (FMRs) are primarily used to determine payment...
application/xls
Federal

datamirror.org

- *Mirror of the US federal government open data portal, data.gov*
- *If these data are critical to the University's mission, then it is our responsibility to help steward it*

Preservation



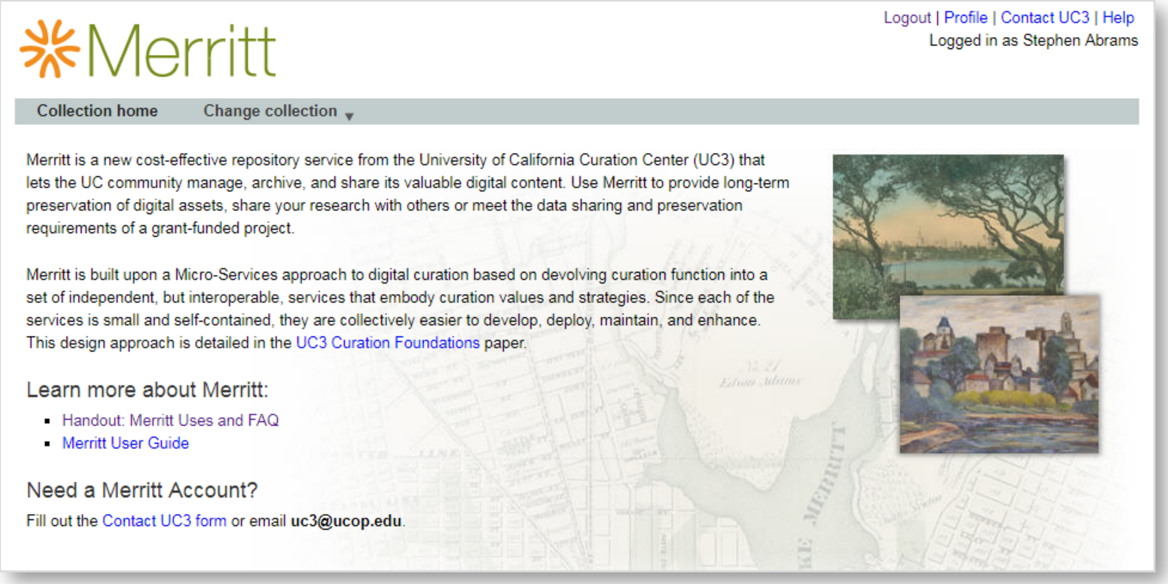
UC3
UC Curation Center

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

Merritt



Merritt
Logout | Profile | Contact UC3 | Help
Logged in as Stephen Abrams
Collection home
Change collection

Merritt is a new cost-effective repository service from the University of California Curation Center (UC3) that lets the UC community manage, archive, and share its valuable digital content. Use Merritt to provide long-term preservation of digital assets, share your research with others or meet the data sharing and preservation requirements of a grant-funded project.

Merritt is built upon a Micro-Services approach to digital curation based on devolving curation function into a set of independent, but interoperable, services that embody curation values and strategies. Since each of the services is small and self-contained, they are collectively easier to develop, deploy, maintain, and enhance. This design approach is detailed in the UC3 Curation Foundations paper.

Learn more about Merritt:

- Handout: Merritt Uses and FAQ
- Merritt User Guide

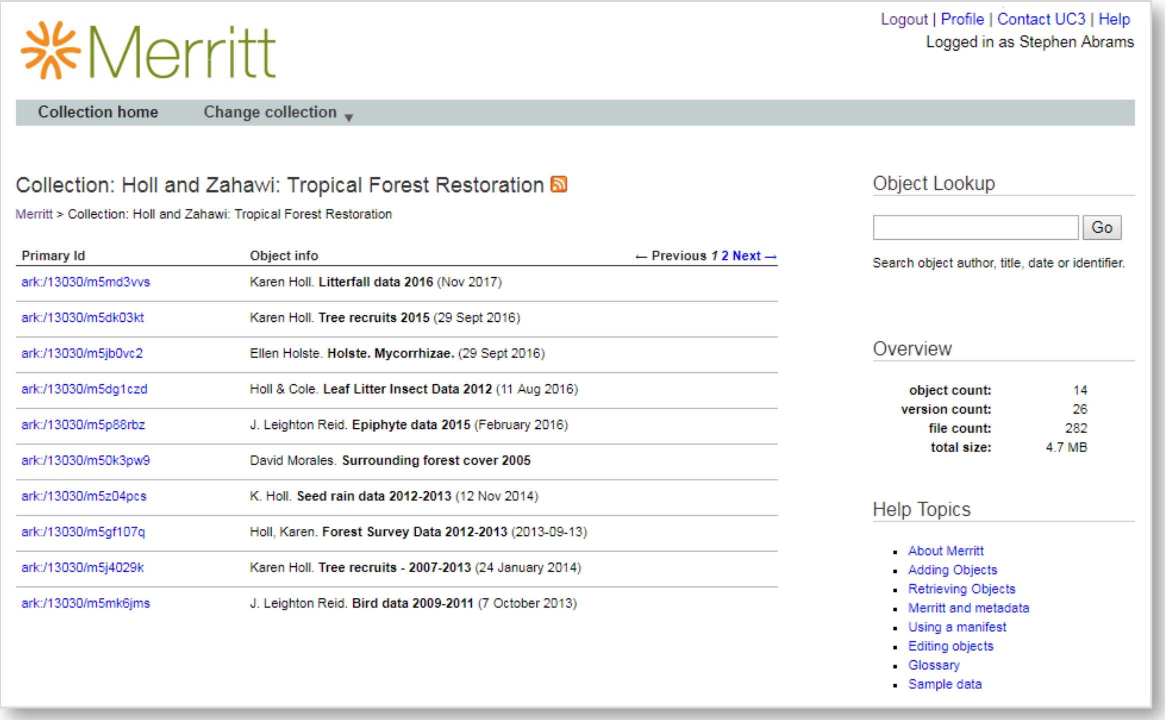
Need a Merritt Account?

Fill out the Contact UC3 form or email **uc3@ucop.edu**.

- *Preservation and access*
- *All 10 campuses,*
35 curatorial units,
403 collections
- *2.7 million objects,*
3.4 million versions,
40.9 million files
89.3 TB
- *Partial cost recovery for preservation storage*

merritt.cdlib.org

Merritt


Merritt
Logout | Profile | Contact UC3 | Help
Logged in as Stephen Abrams
Collection home
Change collection
Go

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*

Preservation at the network level

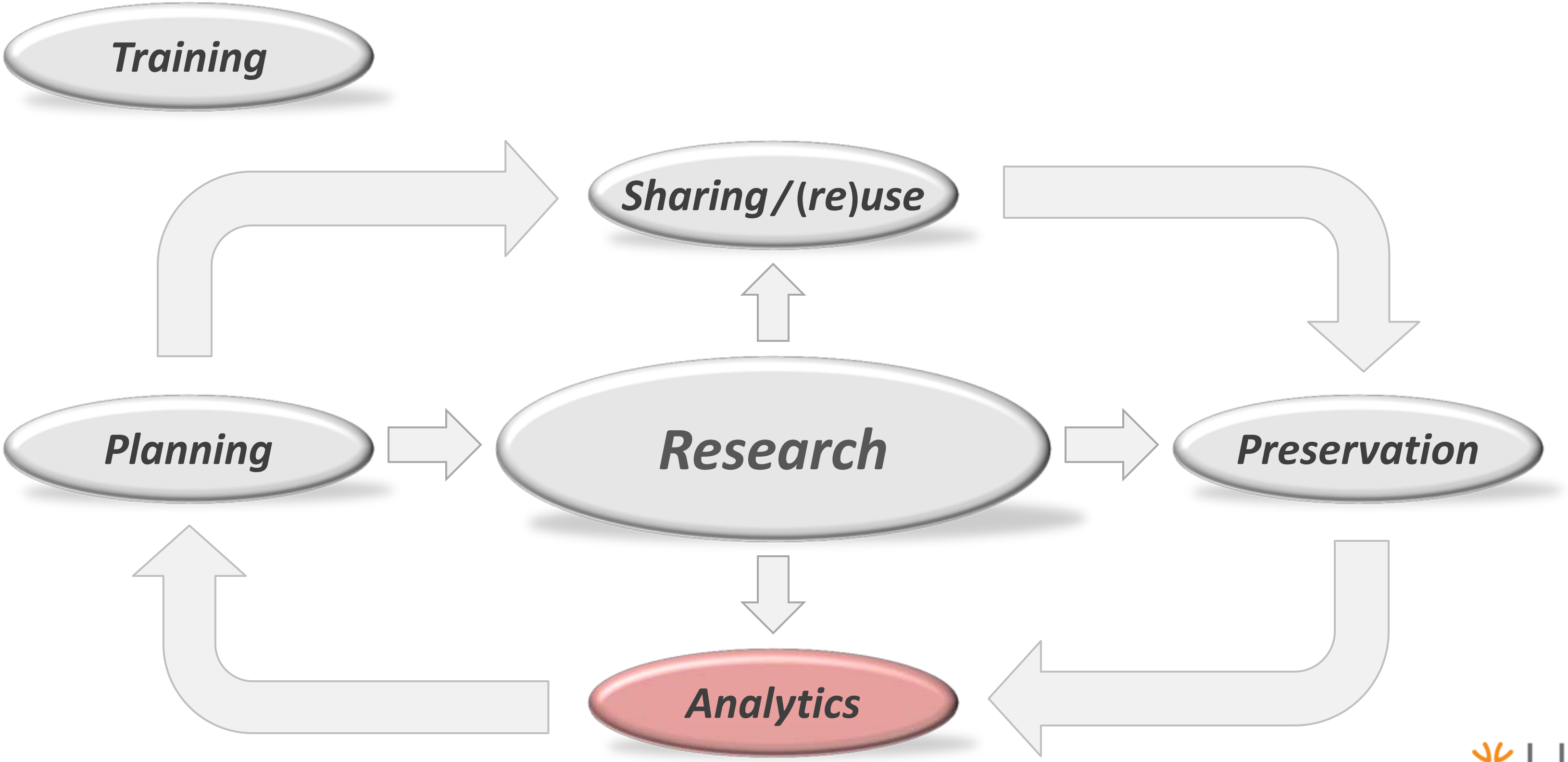
Digital Preservation Network (DPN) – additional storage replication
dpn.org

HathiTrust – mass digitization of serials and monographs
www.hathitrust.org

International Internet Preservation Consortium (IIPC) – web archives
netpreserve.org

National Digital Stewardship Alliance (NDSA) – advocacy
ndsa.org

Analytics



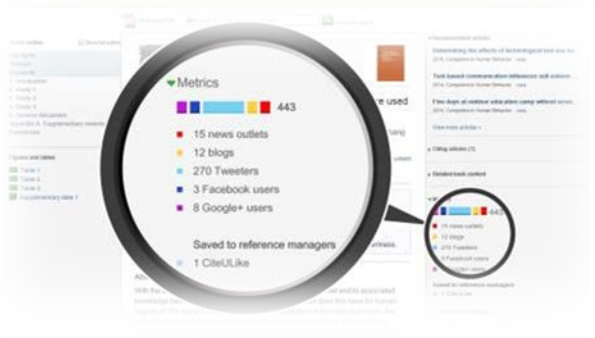
UC3
UC Curation Center

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

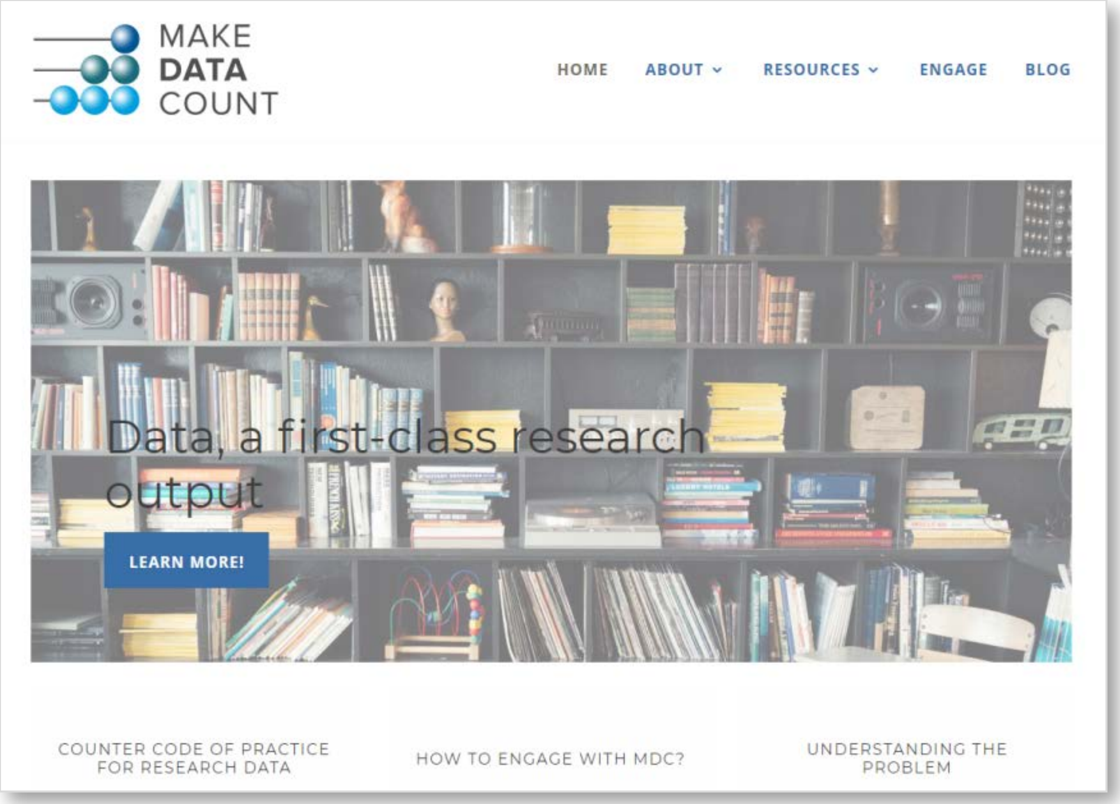


Metrics
443
- 15 news outlets
- 12 blogs
- 270 Tweeters
- 3 Facebook users
- 8 Google+ users

Saved to reference managers
- 1 CiteULike

blog.sciencedirect.com/topics/altmetrics

Make Data Count (MDC)



MAKE DATA COUNT

HOME ABOUT ⌄ RESOURCES ⌄ ENGAGE BLOG

Data, a first-class research output

LEARN MORE!

COUNTER CODE OF PRACTICE FOR RESEARCH DATA

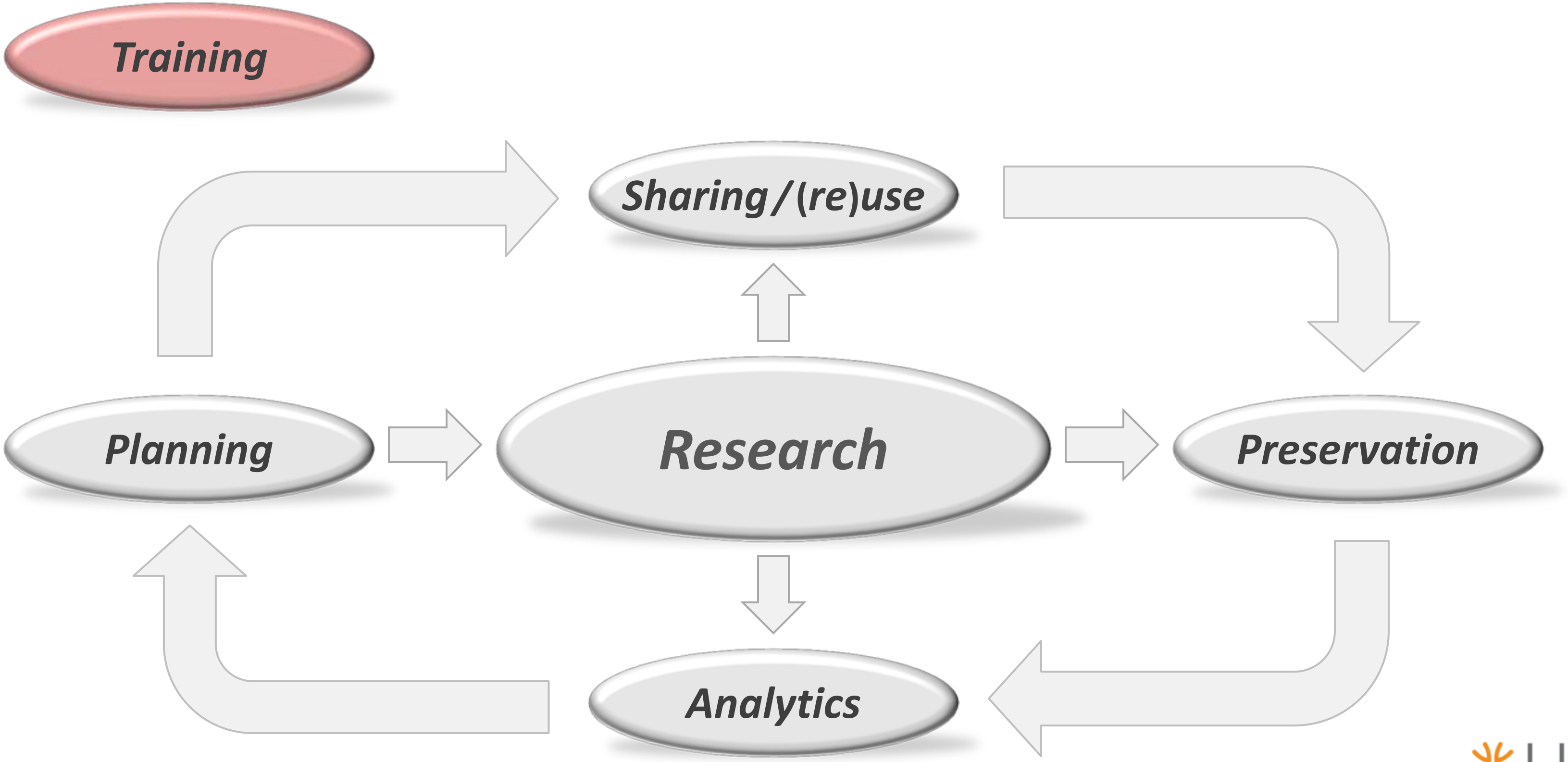
HOW TO ENGAGE WITH MDC?

UNDERSTANDING THE PROBLEM

makedatacount.org

- *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*

Data literacy training



UC3
UC Curation Center

Data literacy training

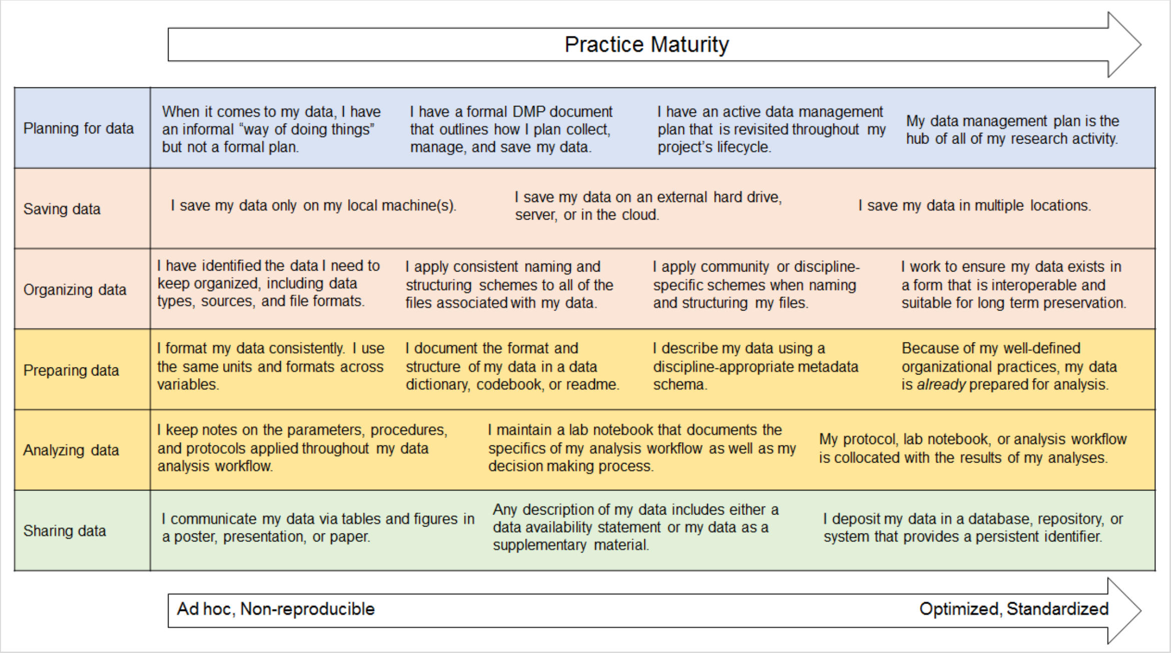
Most scholars and researchers have never received any data literacy training

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

www.flickr.com/photos/rosengrant/3847862890

UC3
UC Curation Center

Self-assessment maturity guide



Practice Maturity

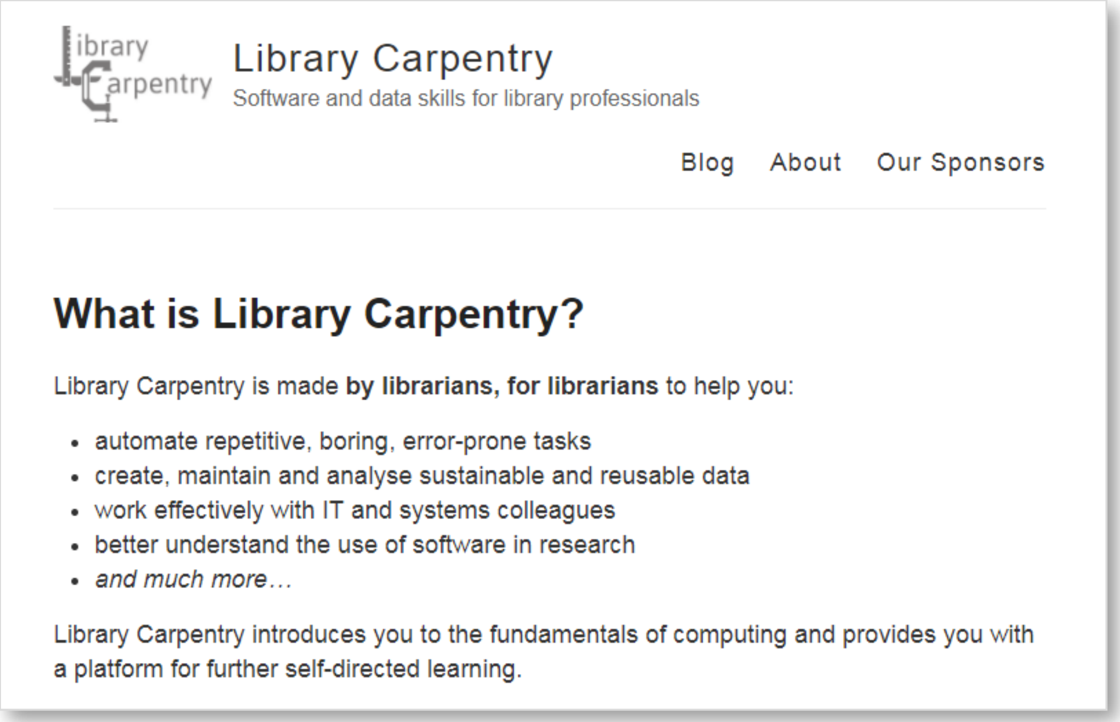
Planning for data	When it comes to my data, I have an informal "way of doing things" but not a formal plan.	I have a formal DMP document that outlines how I plan collect, manage, and save my data.	I have an active data management plan that is revisited throughout my project's lifecycle.	My data management 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 *already* 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*

uc3.cdlib.org/2016/09/12/building-a-user-friendly-rdm-maturity-model/

Library carpentry



Library Carpentry

Software and data skills for library professionals

What is Library Carpentry?

Library Carpentry is made **by librarians, for librarians** to help you:

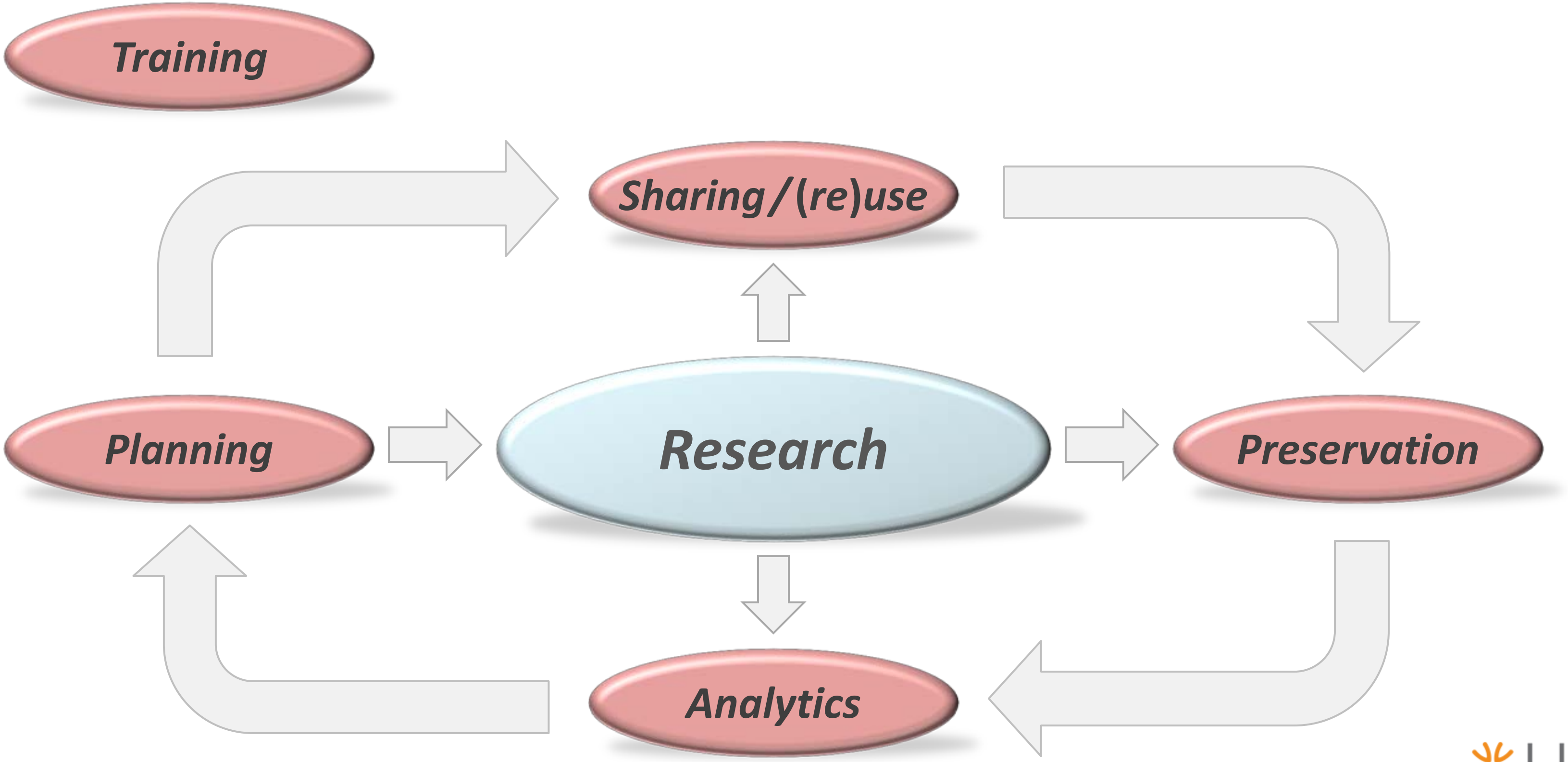
- automate repetitive, boring, error-prone tasks
- create, maintain and analyse sustainable and reusable data
- work effectively with IT and systems colleagues
- better understand the use of software in research
- *and much more…*

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



UC3
UC Curation Center

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

Stephen.Abrams@ucop.edu
ORCID 0000-0003-2326-6672

UC3
UC Curation Center