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
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
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
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
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

- Training
- Planning
- Research
- Analytics
- Sharing/(re)use
- Preservation
Planning

- Training
- Planning
- Research
- Analytics
- Sharing/(re)use
- Preservation
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

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

dmptool.org
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

Training

Planning

Research

Analytics

Preservation
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
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
Dash data publication service

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

[dash.ucop.edu](dash.ucop.edu)
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
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
Preservation

Training

Planning

Research

Sharing/(re)use

Preservation

Analytics
Preservation

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


dx.doi.org/10.1038/sdata.2015.39
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

Make Data Count (MDC)

makedatacount.org
Data literacy training

- Training
- Planning
- Research
- Analytics
- Sharing/(re)use
- Preservation
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.
# Self-assessment maturity guide

## Practice Maturity

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

**Ad hoc, Non-reproducible**

**Optimized, Standardized**

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

Library carpentry

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

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.

librarycarpentry.github.io
uc3.cdlib.org/2017/11/06/skills-training-for-librarians-expanding-library-carpentry/
Supporting research data management

- Training
- Planning
- Analytics
- Sharing/(re)use
- Preservation

Research
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