

Nov 14th, 10:15 AM - 11:00 AM

The Library's Role in Data Management Support: A Perspective from UC Libraries

Amy Koshoffer
University of Cincinnati - Main Campus

Follow this and additional works at: <https://corescholar.libraries.wright.edu/irday>

Repository Citation

Koshoffer, Amy, "The Library's Role in Data Management Support: A Perspective from UC Libraries" (2017). *Ohio IR Day*. 2.
https://corescholar.libraries.wright.edu/irday/fall_2017/november14/2

This Event is brought to you for free and open access by the University Libraries' Events at CORE Scholar. It has been accepted for inclusion in Ohio IR Day by an authorized administrator of CORE Scholar. For more information, please contact library-corescholar@wright.edu.

The Library's Role in Research Data Management Support

A perspective from UC Libraries

Amy Koshoffer
Science Informationist

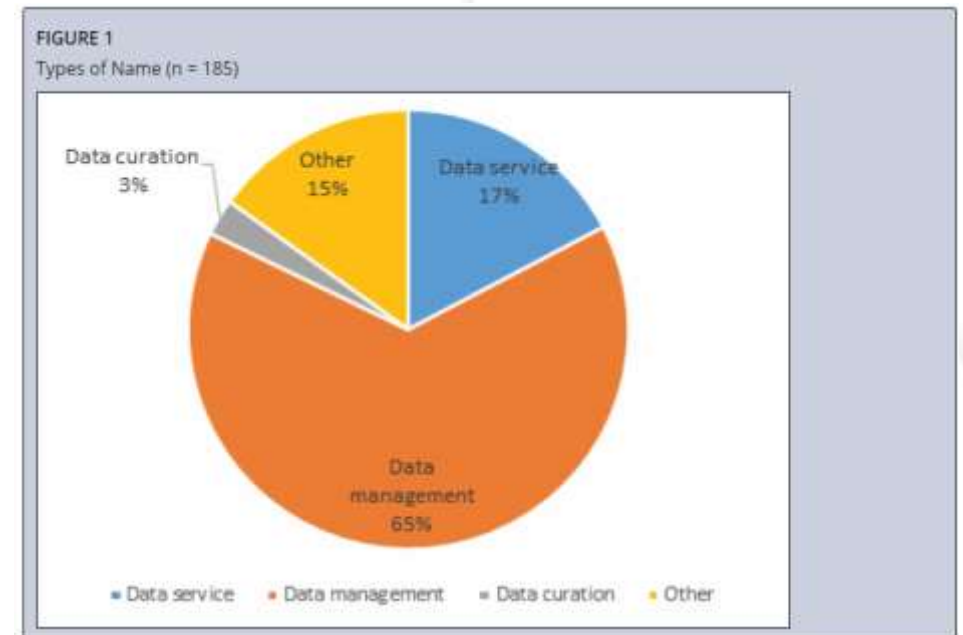
www.libraries.uc.edu



A study of 185 library websites indicated that **65% use the term “data management”** on their webpage. (Yoon & Schultz 2017)

A recent survey of 80 American Research Libraries (ARL) institutions found that **80% had data curation and repository services** in place or planned to provide them (Hudson-Vitale et al., 2017).

Survey of 73 ARL libraries found that **74% offer some data management support** (Fearon et al., 2013).



Yoon, A., & Schultz, T. (2017). Research Data Management Services in Academic Libraries in the US: A Content Analysis of Libraries' Websites. *College & Research Libraries*, 78(7), 920. doi:<https://doi.org/10.5860/crl.78.7.920>

What is Research Data Management?

All activities focused on helping

describe, explain, locate, make it easier to retrieve, use, or keep track of *

data

- Researcher Generated Data
- Resources Researchers need to conduct scholarship
- Digital
- Key to reproducibility of research
- Key to data sharing

Leads to ...

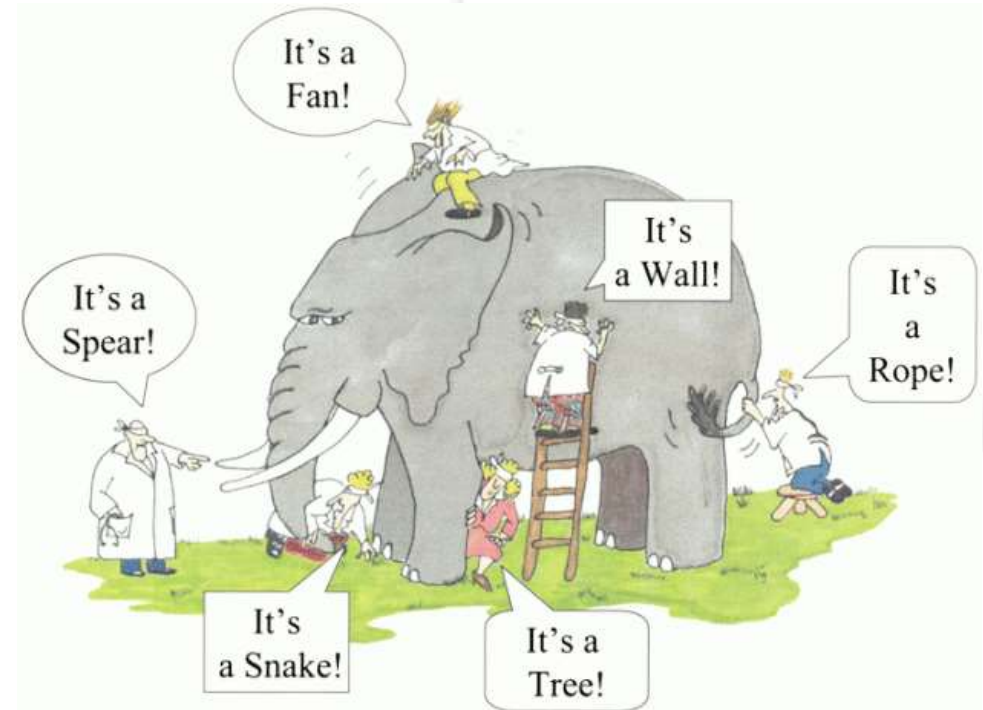
- Easier research workflow
- More grants
- More money and time to do research
- **Happier life**

www.libraries.uc.edu

The logo for the University of Cincinnati Libraries. It features a red stylized 'UC' monogram above the text 'University of CINCINNATI LIBRARIES' in a serif font. The word 'LIBRARIES' is in a larger, bold, red font.

What is Research Data?

- Research data, unlike other types of **information**, is collected, observed, or created, for purposes of analysis to produce **original research results**". (University of Edinburgh)
 - Information collected to answer new questions
- "The recorded factual material commonly accepted in the research community as necessary to **validate research findings**". (Excerpted from OMB Circular A-110 36.d.2.i)
 - Information collected to answer old questions



Research Data Associated with Most Disciplines:



UC: Research of Kristina Neumann:
Google map of ancient Antioch with coin data overlay

- Numerical measurements
- Text from Survey/Focus groups
- Images
- Video
- Mapping/GIS data

Why Manage Research Data?

For the Scientific Community = **Sharing**

- Meet grant & publishing requirements
 - Support Open Access Policies
- Facilitate new discoveries

For the Research Group = **Sharing**

- Save time, increase research efficiency
- Maintain data integrity

For the Individual = **Management**

- Preserve your data
- Increase your research impact

From MIT Libraries' [Reasons to Manage & Publish Your Data](http://www.libraries.uc.edu) www.libraries.uc.edu

Data management Plans vs. Data Management Planning

- Plan

- List of activities
- Contract with funding agency
- Specific to each funder

- Planning

- On-going activities

Planning makes creating the plan easy

Questions to be Addressed in a DMP

- What are the **funding agency** requirements for data management?
- What **data** will be created during the research project?
- How will the **data files be named** and organized?
- How will **changes** in the data files be **tracked**?
- Which **file format(s)** will be used for the data?
- How will the data be **documented**?
- Where will the data be **stored**?
- **How long** will the data be stored? (e.g., 2 years? 10 years?)
- Should the data be placed in a repository or an archive for **sharing** and long-term storage?
- Who should have **access** to the data?

Meet Grant & Publishing Requirements

Funder	Policy Coverage		Policy Stipulations					Support Provided			
	Published Outputs	Data	Time Limits	DMP	Access / Sharing	Long-term curation	Monitoring	Guidance	Article Repository	Data Repository	Costs
AHRQ	●	●	●	●	●	◐	◐	◐	●	◐	◐
ASPR	●	●	●	●	●	●	●	◐	●	◐	◐
CDC	●	●	●	●	●	◐	●	◐	●	◐	◐
DOD	●	●	●	●	●	◐	●	◐	●	**	◐
DOE	●	●	●	●	●	◐	◐	◐	●	◐	◐
DOT	●	●	●	●	●	◐	●	◐	●	◐	●
FDA	●	●	●	●	●	◐	●	◐	●	◐	◐
NASA	●	●	●	●	●	◐	●	◐	●	◐	◐
NIH	●	●	●	●	●	◐	◐	◐	●	◐	◐
NIST	◐	●	◐	●	●	●			●	**	
NOAA	●	●	●	●	●	●	●	◐	●	◐	◐
NSF	●	●		●	●	◐	●		●	●	●
SI	●	●	●	◐	●	◐	◐	◐	◐	◐	◐
USDA	●	◐				◐	◐	◐	●	**	●
USAID	◐	●	◐		◐		◐	◐	◐	**	
USGS	●	●	●	●	●	●	●	●	●	**	●
VA	◐	◐	◐	●	**	◐	●	◐	●	**	●

Last updated: October 18, 2016

● Full Coverage ◐ Partial Coverage
 ○ No Coverage □ Not Enough Information
 ** For more information, see <http://bit.ly/FedOASummary>

www.libraries.uc.edu



When policies go into effect -	Jan. 2016 (for both Article and Data)
Data Management Plan required with funding proposal -	Yes
Timeline for making articles public -	No later than 12 months from publication
Recommended article repository -	NSF Public Access Repository (NSF-PAR) https://par.nsf.gov/
Timeline for making data public -	https://www.nsf.gov/bfa/dias/policy/dmp.jsp

The Results of Poor Research Data Management



Data Sharing and Management Snafu in 3 Short Acts

http://www.youtube.com/watch?v=N2zK3sAtr-4&feature=player_embedded

www.libraries.uc.edu

Issues that arise in Video:

1. Perception that all is in the article – interpretation but not data
2. Ignoring Funding or Publisher requirements
3. Not sure of location of ONLY copy of data
4. Storage of data on a USB, not backed-up in a more reliable location
5. Off campus storage – no instructions for access
6. Proprietary program that is no longer accessible
7. Non-open format of file type
8. Abbreviations used – meaningless variable and file names
9. No explanation included with data set
10. Lost communication with original data generator

Main issues - Data has no Context

Research Data Management Best Practices

- Give files meaningful names
- Organized within a folder structure
- Use open formats that will be accessible in the future
- Back up data – “The 3-2-1Rule”
- Prepare for long term preservation
- Provide context –
 - Metadata
 - Readme describing file relationship
 - Protocols
- Provide contact info for data generator
- Name a steward
- Chose a repository

Journal Requirements



Data Availability

The data underlying the findings of research published in PLOS journals must be made publicly available. Rare exceptions may apply and must be agreed to with the Editor. Data should be de-identified where appropriate (see Human Subjects and Animal Research).

= the **'minimal dataset'** consists "of the dataset used to reach the conclusions drawn in the manuscript with related metadata and methods, and any additional data **required to replicate the reported study findings in their entirety.**

<https://www.plos.org/editorial-publishing-policies>
http://wiki.datadryad.org/Journal_instructions

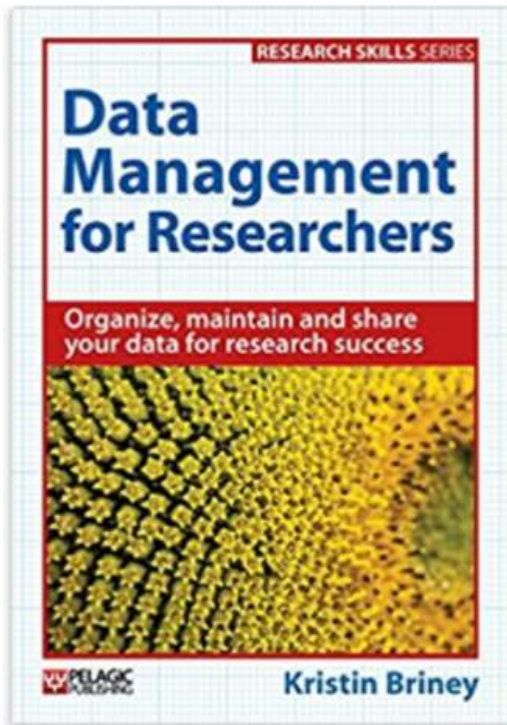
www.libraries.uc.edu


University of
CINCINNATI
LIBRARIES

CC BY-NC

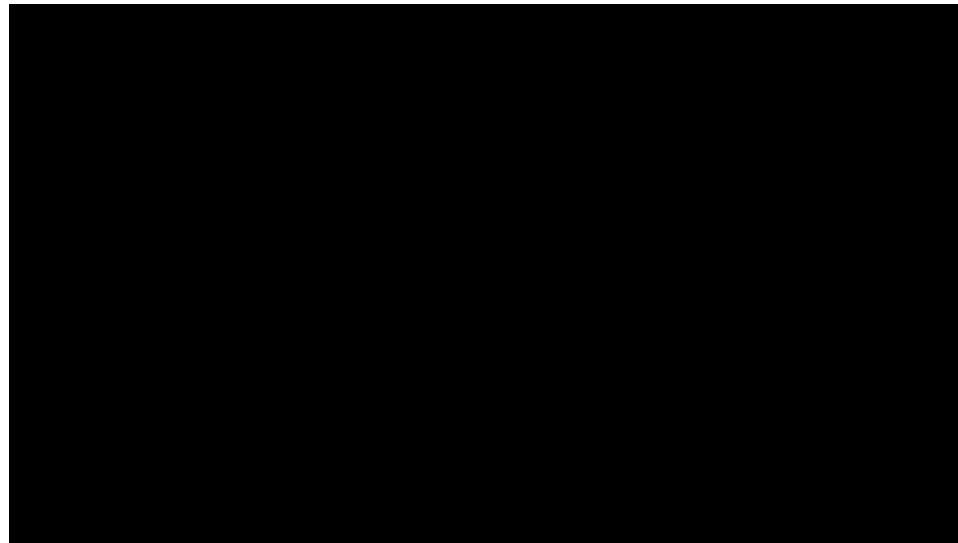
Data or it did not happen

-Kristen Briney



<https://pelagicpublishing.com/products/data-management-for-researchers-briney>

TEDx TEDx Talks
Published on Nov 10, 2015



<https://www.youtube.com/watch?v=dXKbkpilQME>

<http://uwm.edu/libraries/dataservices/>



[DATA AB INITIO](#)

University of
CINCINNATI
LIBRARIES

www.libraries.uc.edu

CC BY-NC

Sharing Data – Dr. Christy Holland

PLOS article link - <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0177786>

scholar@uc

The UC digital repository, part of the UC Research Hub.

Type keywords in here



University of
CINCINNATI

Browse

About ▾

Contact

Help ▾



Clot on a Suture Dataset

Collection submitted by: Shenwen Huang

[Browse this collection](#)

- Clot On a Suture Dataset: Human clots in porcine plasma, Plasma Only (Huang, Shenwen; Shekhar, Himanshu; Holland, Christy)
- Clot On a Suture Dataset: Human clots in porcine plasma, rt-PA (3.15 ug/ml) (Huang, Shenwen; Shekhar, Himanshu; Holland, Christy)
- Clot On a Suture Dataset: Human clots in porcine plasma, rt-PA (3.15 ug/ml), Definity, & US (Huang, Shenwen; Shekhar, Himanshu; Holland, Christy)
- Clot On a Suture Dataset: Porcine clots in human plasma, Plasma Only (Huang, Shenwen; Shekhar, Himanshu; Holland, Christy)
- Clot On a Suture Dataset: Porcine clots in human plasma, rt-PA (3.15 ug/ml) (Huang, Shenwen; Shekhar, Himanshu; Holland, Christy)
- Clot On a Suture Dataset: Porcine clots in human plasma, rt-PA (3.15 ug/ml), Definity, & US (Huang, Shenwen; Shekhar, Himanshu; Holland, Christy)
- Clot On a Suture Dataset: Porcine, Plasma Only (Huang, Shenwen; Shekhar, Himanshu; Holland, Christy)
- Clot On a Suture Dataset: Porcine, rt-PA (3.15 ug/ml) (Huang, Shenwen; Shekhar, Himanshu; Holland, Christy)
- Clot On a Suture Dataset: Porcine, rt-PA (3.15 ug/ml), Definity, & US (Huang, Shenwen; Shekhar, Himanshu; Holland, Christy)
- Porcine and Human Clot SEM Images (Huang, Shenwen; Shekhar, Himanshu; Holland, Christy)

www.libraries.uc.edu

<https://scholar.uc.edu/collections/sufia:bc386p70c>



Library input –
Readme files

University of
CINCINNATI
LIBRARIES

CC BY-NC

Data Management Makes Sharing Data Easier

Research communities want data sharing

Applications in Plant Sciences 5(10):1700111. 2017

<https://doi.org/10.3732/apps.1700111>

The Frontier of Data Discoverability: Why We Need to Share Our Data

[Theresa M. Culley](#)



Theresa Culley – Professor of Biology



www.libraries.uc.edu

<http://www.bioone.org/doi/full/10.3732/apps.1700111>

CC BY-NC

Scholar@UC



Signed in successfully.

what is scholar@uc?

Scholar@UC enables the UC community to share research and scholarly works with a worldwide audience. Scholar@UC is a collaborative between UC Libraries and IT@UC, with support from the Office of Research.
[Learn more about Scholar@UC](#)

- Samvera Community
- In house Developers
- Partnership with IT@UC

- Self Submission
- 6 required metadata fields
 - 21 possible fields



FEATURED RESEARCHER



FEATURED COLLECTION



FEATURED WORK

<https://scholar.uc.edu/>
<https://samvera.org/>

www.libraries.uc.edu

University of
CINCINNATI
LIBRARIES

Sharing Data Example

scholar@uc

The UC digital repository, part of the UC Research Hub.

Type keywords in here



University of
CINCINNATI

Browse

About ▾

Contact

Help ▾


Cleistogamy Review Dataset - Species Level (Dataset)

Digital Object Identifier

[doi:10.7945/C2V305](https://doi.org/10.7945/C2V305)

This DOI is the best way to cite this dataset.

Files

File	Filename	Date Uploaded	Visibility	Actions ?
 CL_SpeciesList_100205.xls	CL_SpeciesList_100205.xls	2016-08-26	Open Access	Download

Library input –
Encouraged Submission to IR
DOI

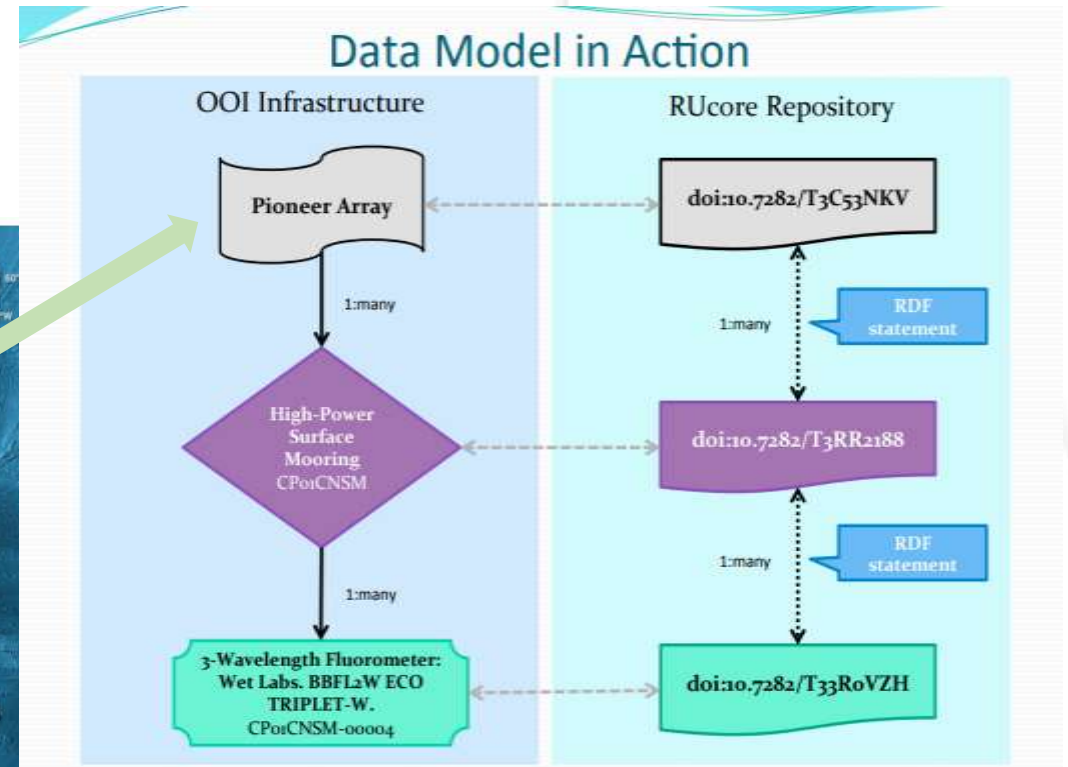
University of
CINCINNATI
LIBRARIES

www.libraries.uc.edu

Sharing Data Example

Collecting data to monitor health of the Atlantic and Pacific Oceans

- DM by Cyberinfrastructure group @ Rutgers University
- 800+ instruments
- 75 sites
- Over 10,000 data products in 1st year
- Data deposited in RUcore
- Dynamic Data
- Metadata and DOI strategy developed by RU Librarians
- Based on instrument recalibration



<http://oceanobservatories.org/>
<https://www.cni.org/topics/e-science/an-ocean-of-data-a-metadata-and-doi-strategy-for-large-dynamic-data-about-the-worlds-oceans>
<https://rucore.libraries.rutgers.edu/rutgers-lib/46629/#related>

www.libraries.uc.edu



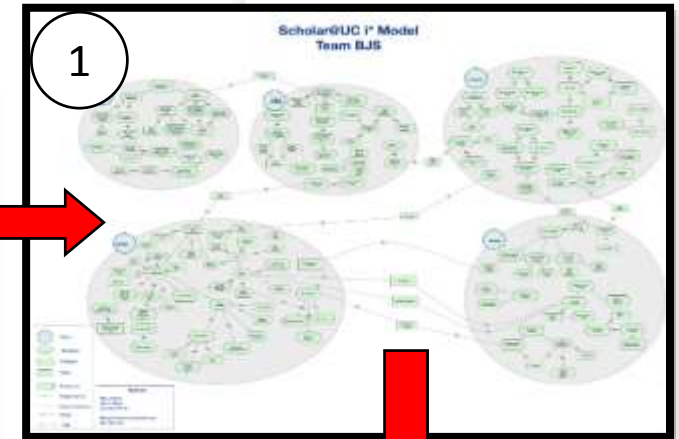
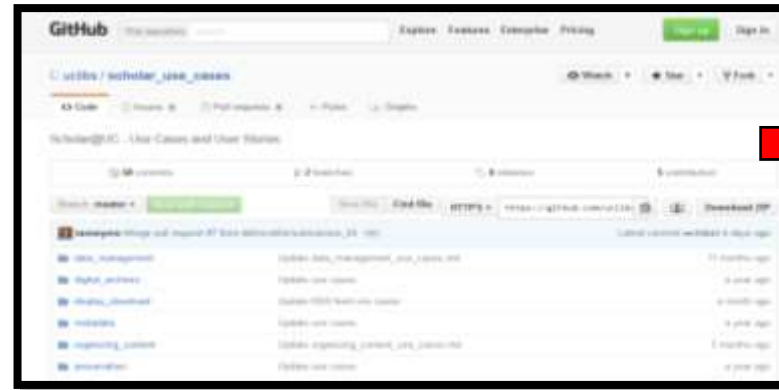
CC BY-NC

Cool Data Examples

Shared by Data Librarians around the nation

- **University of British Columbia** - Social housing for pigs: <http://hdl.handle.net/11272/10219>
- **Stanford University** - The Revs Vehicle Dynamics Database: <https://purl.stanford.edu/cv083fq4881>
- **University of Minnesota** - Frogs Exploit Statistical Regularities in Noisy Acoustic Scenes to Solve Cocktail-Party-Like Problems: <https://doi.org/10.13020/D6ZP4H>
- **Brown University Library** - Dramatic fighting by male cuttlefish for a female mate: <https://doi.org/10.7301/Z0PR7SX4>
- **Brown University Library** - Laura the camel from the Cleveland Zoo's brain: <https://doi.org/10.7301/Z00C4SPV>
- **Brown University Library** - 3D print your own skeleton of a 2,000 year old Ibis mummy in RISD's museum: <https://doi.org/10.7301/Z0PZ56R5>
- **American University** - List of U.S. Military Bases Abroad" (<http://dx.doi.org/10.17606/M6H599> (downloaded over 3168 times)
- **Virginia Tech** - How do sea birds dive without injury? <https://data.lib.vt.edu/collections/rn3011410>
- **Research Data Archive** - International Comprehensive Ocean-Atmosphere Data Set (ICOADS) - <https://rda.ucar.edu/datasets/ds548.0>
- **Research Data Archive** - NCEP GDAS/FNL 0.25 Degree Global Tropospheric Analyses and Forecast Grids - <https://rda.ucar.edu/datasets/ds083.3>

Partnerships – Dr. Nan Niu, Requirement Engineering



1) Teaching Tool

Repository Use cases were the sources of data for a class project

2) Research Tool

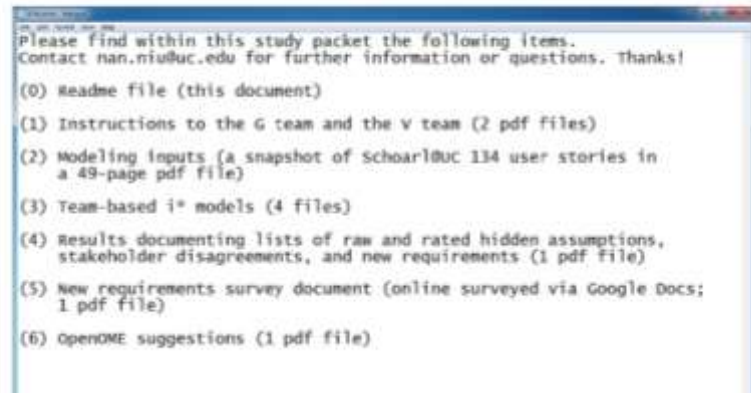
Replication study of viewpoint modeling to better understand traceability impact on team-level requirements modeling

3) Data Archive

Put all data related to paper into Scholar@UC

RE'16 Viewpoint Merging Study Materials (Generic Work)

3



Link to supporting documentation for Research Paper in Scholar

https://scholar.uc.edu/works/generic_works/05741s72s

Library Blog Posts

<https://libapps.libraries.uc.edu/libblog/2016/01/scholaruc-goes-to-class/>

<http://libapps.libraries.uc.edu/libblog/2016/09/paper-featuring-scholaruc-gets-best-research-paper-award/>

www.libraries.uc.edu

CC BY-NC

February 22, 2013

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: John P. Holdren *JPH*
Director

SUBJECT: Increasing Access to Federally Funded Scientific Research

I. Policy Principle

The Administration is committed to ensuring that, to the greatest extent and with the fewest constraints possible and consistent with law and the objectives set forth below, the direct results of federally funded scientific research are made available to the public, industry, and the scientific community. Such results include peer-reviewed publications and digital data.

Scientific research supported by the Federal Government has produced innovative breakthroughs that drive our economy. The results of that research become the source of new insights and are assets for progress in areas such as health, energy, the environment, agriculture, and national security.

Access to digital data sets resulting from federally funded research allows companies to focus resources and efforts on understanding and developing new products and services. For example, open weather data underpins the forecasting industry, and publicly available genome sequences has spawned many biotechnology innovations. In addition, wider availability of peer-reviewed publications and scientific data in digital formats will create innovative economic markets for services related to curation, preservation, analysis, and visualization. Policies that mobilize these publications and data for re-use through open access and broader public access also maximize the impact and accountability of the Federal Government's research investment. These policies will accelerate scientific breakthroughs and innovations, promote entrepreneurship, and enhance economic growth and job creation.

The Administration also recognizes that librarians provide valuable services, including the coordination of peer review, that are essential to ensuring the high quality and integrity of many scholarly publications. It is critical that these services continue to be made available. It is also important that Federal policy not adversely affect opportunities for researchers who are not funded by the Federal Government to publish any analysis or results of their research.

To achieve the Administration's commitment to increase access to federally funded published research and digital scientific data, Federal agencies investing in research and development must have clear and coordinated policies for increasing such access.

The MEMO -

Federal
Requirements:

Subject: Increasing Access to the
Results of Federally Funded
Scientific Research

Efforts to capture at risk data



<http://www.ppehlab.org/datarefuge>

Sharing rescued data



<http://climatemirror.org/>

As mirrored data

landsat-pds.s3.amazonaws.com | [HTTP](http://) | [US](https://) | amzn-noc-contact@amazon.com

This mirror is operated by Amazon AWS. It is currently hosting the following datasets:

- earthexplorer.usgs.gov | [TBA GB](#) [TBA GB](#)

<https://climate.daknob.net/>

www.libraries.uc.edu

Eric Holthaus  @EricHolthaus

Scientists: Do you have a US .gov climate database that you don't want to see disappear?

Add it here:

11:22 PM - 10 Dec 2016

<https://twitter.com/EricHolthaus/status/807802801506041856>



Eric Holthaus 

@EricHolthaus

LIBRARIES+ NETWORK



<https://libraries.network/>

www.libraries.uc.edu



Why should Libraries have a role in Research Data Management?

We are masters at giving context to, preserving, and providing access to information.

We are integral to the research process.

Data management activities expand on our existing partnership with researchers.

Relevant to our mission

How did we get started at UC?

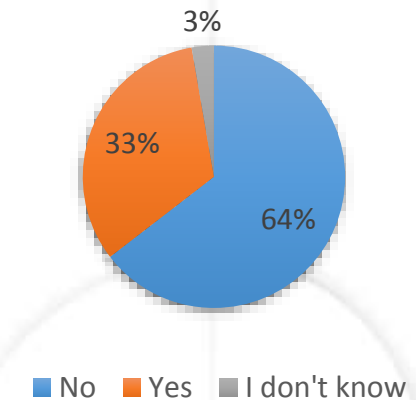
Pre-planning

- The Dean of Libraries, Director of the Science and Engineering Libraries, and Chief Information Officer attended the ARL-DuraSpace E-Science Institute in Dec 2012.
- Strategic Planning
- \$\$\$\$
- Partnerships
- Built/Building Team
- Developed/Developing Services
- New spaces
- Resources
- Work in Progress....

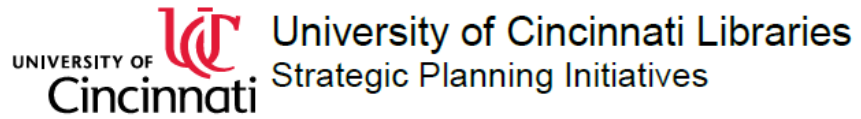
Should you conduct a needs assessment?

- There are lots of examples of surveys
- Could follow these examples
- At UC we did conduct a survey
 - 55.76% response rate
 - Researchers express need for
 - Data Management Consultation
 - Assistance Archiving Data

Have you written a data management plan?



Strategic Initiatives



INITIATIVE CHARTER

1. General Information

Initiative Name:	Developing a Suite of Programs for UC Researchers (Research, Implement)
------------------	---

Partnerships

- Central IT
- UC Libraries – Liaison Librarians, Digital Scholarship Center
- Researchers
- Office of Research
 - AVP Integrated Research – Head of DAAP Library
- Colleges
- Other centers, organizations within UC

Partnerships



Changjoo Kim -
GIS Partner



Tom Beck -
HPC/Data Day Partner



Geoffrey Pinski
Office of Research - IP office



Littisha Bates
Data Day Partner



Dylan Ward and Craig Dietsch -
GIS Partners



John Wallrodt -
Repository/Data Day/
GIS Partner



Michael Richardson
Data Day Partner



Yan Yu
Data Day Partner



Members of the Consortium on Cultural Diversity in Chemistry
COS Partners



Brett Kottmann
IT@UC - HPC/UCScienceNet



Jane Combs
IT@UC - Research and Computing

And so many more...



Building the Team

Research Data Services Team

- Senior Associate Dean & Directory of Health Sciences Library
- Assistant Director for Research and Informatics
- Director of Science and Engineering Libraries
- **Informationists**
- Metadata Librarians
- Liaison Librarians
- IT@UC Representatives
- Scholarly Communications & Digital Publishing Strategist

Collaborators

- **Repository Team**
- Digital Scholarship Center



Building the Team

Professional development

- Science Boot Camps
- Research Data and Access Preservation Conference (RDAP) <https://www.asist.org/events/rdap-summit/>
- Midwest Data Librarian Symposium - <https://mwdatalibrariansymposium.wordpress.com/>
- Subject Specific
 - iDigBio Digital Data Conference - <https://www.idigbio.org/tags/berkeley>
- ACRL – Data Management Roadshows <http://www.ala.org/acrl/conferences/roadshows/rdmroadshow>

Professional Organizations

- Research Data Alliance <https://www.rd-alliance.org/node>
- Association for Information Science and Technology (ASIST) <https://www.asist.org/>
- SLA Data Caucus <http://data.sla.org/>
- International Association for Social Science Information Services and Technology (IASSIST) <http://iassistdata.org/>

Building Services

- Set up a website

<http://libraries.uc.edu/digital-scholarship/data-services.html>

- Specialized Emails

ASKDATA@UC.EDU , ASKGIS@UC.EDU

Building Services

- Health Informatics
- Data Management Planning
- Geographic Information Systems support
- Data Visualization
- Specialized workshops
 - **Data management planning**
 - Bioinformatics databases
 - Data best practices
 - OSF, REDCap, LabArchives,
 - Intro to GIS, R



Some of our Accomplishments

- Established Health Informatics Lab
 - Hired Graduate Assistant – teaches R workshops, support
- GIS Support - <http://guides.libraries.uc.edu/gis>
 - Established a Data & GIS Collab
 - Student workers – support
 - GIS Learning Community
- Center for Open Science
 - **Role of Ambassador**
 - **Reproducibility Workshops**
 - OSF.UC.EDU
- NCBI workshops



Some of our Accomplishments

- Held two [UC Data Day Events](#) – Next will be **March 6, 2018**
 - Keynote by Patty Brennan
 - Director of NLM and NIH Interim Associate Director for Data Science
- DMPTool Integration
- ORCID Institutional Membership
- Repository Support - Scholar@UC Camps
- **Grant support**
 - 2 NIH Informationist Supplements
 - Replicating the NIH Model at UC
 - College of Medicine – awarded grants to COM faculty
 - College Engineering and Allied Science considering model



Student Poster Session, UC Data Day 2017

Challenges

- Data Management is not sexy no matter how you package it
- Developing good use cases
- Avoiding the perception of mandates
- Building partnerships takes time
- Getting time with researchers
- Social Science research support
- Budgets are never big enough

Our Next Steps:

Connect with more researchers

Data Science Support - Symposium

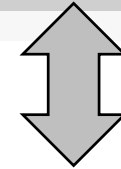
Data Visualization Support – Curriculum/DSC

Integration into researcher workflows – Purdue example

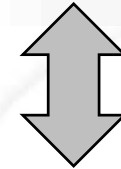
Interoperability of Systems

Digital Preservation Policy

scholar@uc



[ORCID](#) / Datacite DOI's



 University of
CINCINNATI | Research Directory

Internal Researcher Profile system
<https://researchdirectory.uc.edu/>

How do you get started?

Take a moment of Zen



How do you get started?

Are you a team of one or many?

- Focus on the small steps - best practices
- Focus on existing partnerships
- Make use of the existing community

Resources

Great Institutional Resources

U of Minnesota - <https://www.lib.umn.edu/datamanagement>

Massachusetts Institute of Technology - <https://libraries.mit.edu/data-management/>

Get on Newsletters

University of Illinois <https://www.library.illinois.edu/rds/> get a Data Nudge

Stanford - <https://library.stanford.edu/research/data-management-services>

Professional or Institution Courses and Resources

DataOne - <https://www.dataone.org/data-management-planning>

U of Massachusetts – NECDMC <https://library.umassmed.edu/necdmc/dmp>

Mantra - <http://mantra.edina.ac.uk/>

SLA Data Caucus – Toolkit - <http://data.sla.org/data-tool-kit/>

Books *Data Management: A Practical Guide for Librarians* – Margaret E. Henderson

978-1-4422-6439-7 • eBook



Amy Koshoffer, MS, MLIS
Science Informationist

Science and Engineering Libraries
University of Cincinnati Libraries
Office: 240C Braunstein Hall
PO Box 210153
Cincinnati, OH 45221-0153

Tel: (513) 556-1310
Email: amy.koshoffer@uc.edu
Web: libraries.uc.edu

Thank You

Questions?

Answers?

Slides available at <https://osf.io/7v6e3/>

www.libraries.uc.edu



CC BY-NC