Putting a Good Data Management Plan into Action and making Impact

Amy Koshoffer
Asst. Dir. Research & Data Services | UC Libraries

University of CINCINNATI LIBRARIES

Research & Data Services | UC Libraries

Data Management

These data services align with researchers at the time of proposal planning and continue through the course of the project. These services include consultations with researchers regarding collecting/generating, storing, sharing and preserving research data.

Data Analytics

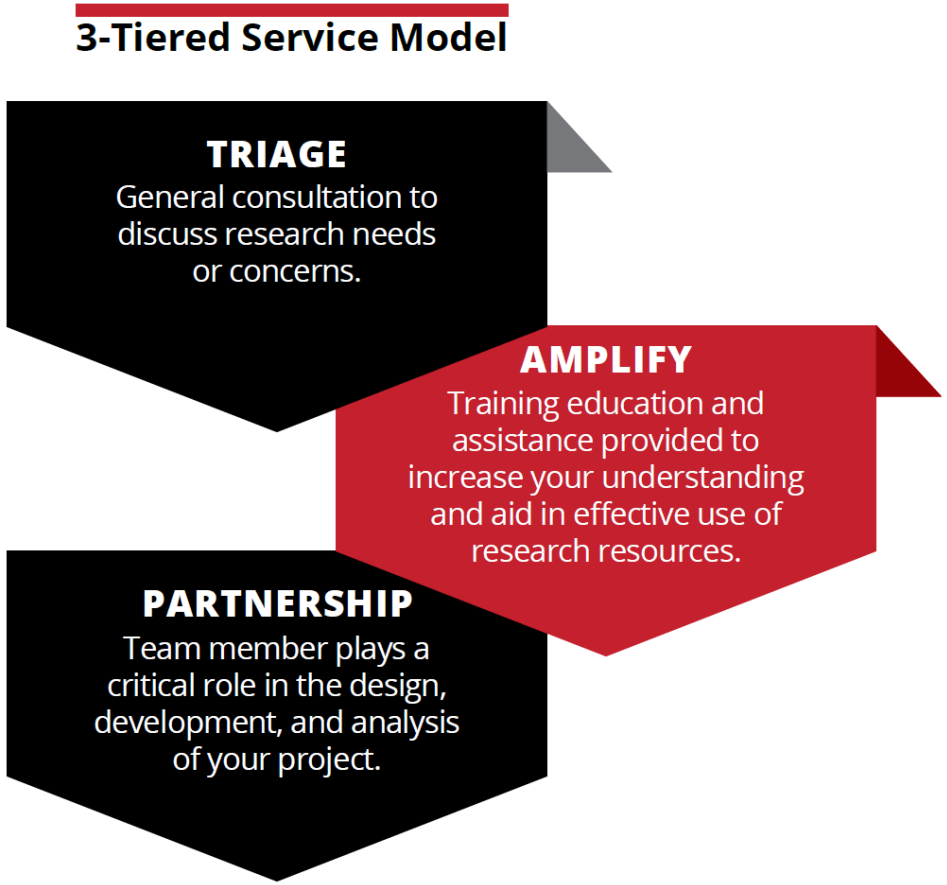
Performed by a combination of librarians and graduate students, these services will assist researchers in generating insight from data to inform decisions and research outcomes. Our analytics services can assist researchers in identifying appropriate analyses, explaining statistical methods, interpretation of results, data wrangling and statistical computation, and implementing models and algorithms in
supported software environments.

Biomedical Informatics

These services are designed to assist researchers navigate biomedical databases and acquire genomic and proteomic data for research purposes.

Geographic Information System

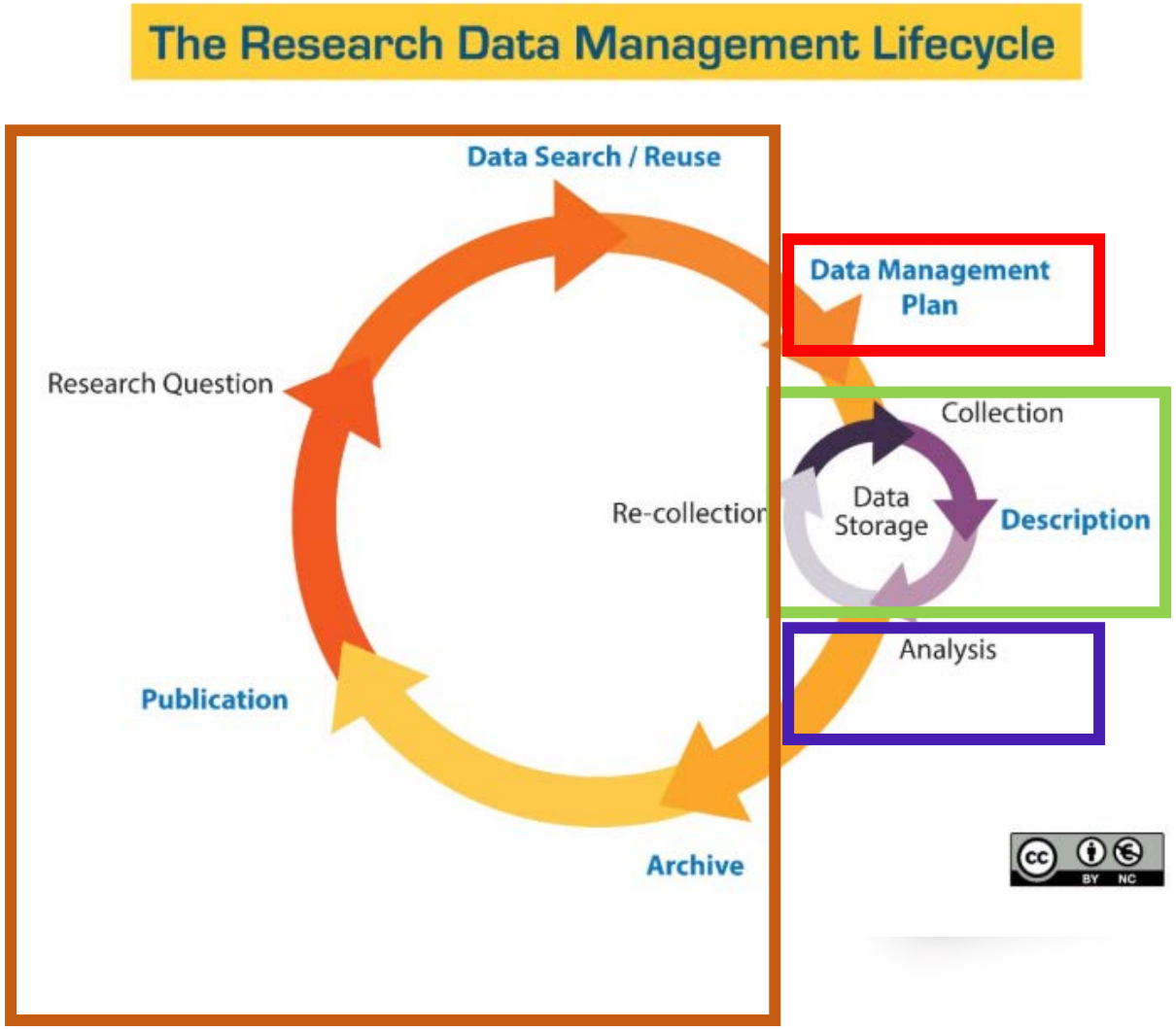
Performed by a combination of librarians and graduate students, these services are centered on the utilization of spatial and/or geographic data. Researchers can gain assistance on gathering data, spatial analysis, map design, and web mapping applications.



The Research & Data Services group provides innovative and effective research partnerships by engaging researchers in data services through collaboration, consultation, and training that aims to amplify the global UC research profile.

www.libraries.uc.edu

University of CINCINNATI LIBRARIES



Data management planning

Storing data
- Onedrive

Processing /Analyzing data
- ARC

Archiving, sharing data
- scholar@uc

University of CINCINNATI LIBRARIES

A data management plan should answer the following:

- What data will be produced?
- What standards will be used to document the data?
- How will the data be protected especially sensitive and restricted data
- How will the data be archived and preserved?
- How will reuse of the data and access to the data , and IP rights, if applicable, be facilitated?

https://figshare.com/articles/Data_Management_Plan_Checklist/1130852
Source- Kristin Briney

University of CINCINNATI LIBRARIES

Who requires a data management plan?

Federal Funding agencies, and other grant issuing agencies -

- Agency for Healthcare Research & Quality (AHRQ)
- Centers for Disease Control and Prevention (CDC)
- Department of Defense (DOD)
- Department of Energy (DOE)
- Department of Transportation (DOT)
- Food and Drug Administration (FDA)
- National Aeronautics and Space Administration (NASA)
- National Institutes of Health (NIH)
- National Institute of Standards and Technology (NIST)
- National Oceanic and Atmospheric Administration (NOAA)
- National Science Foundation (NSF)
- US Agency for International Development (USAID)
- US Department of Agriculture (USDA)
- US Geological Survey (USGS)
- More info on Federal Funding Agencies' policies

https://www.lib.ncsu.edu/data-management/funding_agencies

University of CINCINNATI LIBRARIES

When writing the plan :

- Be specific
- Think and plan ahead

When the project is running:

- Put the DMP into action
- Reference the plan and adjust as the project evolves

University of CINCINNATI LIBRARIES

What data will be produced?

	Complete / Detailed	Addressed issue, but incomplete	Does not address the issue	NSF Directorates Req
Describes what types of data will be captured, created, or collected	This project will produce: observational data in binary format, which is then converted by proprietary instrumentation software into ASCII text; Matlab mat-files for working data during the analysis phase; CSV and/or HDF5 files for data sharing and preservation.”	"We will collect zooplankton data from net tows. We will also collect elemental ratio data from preserved samples."		All
	Clearly defines data type(s). E.g. text, spreadsheets, images, 3D models, software, audio files, video files, reports, surveys, patient records, samples, final or intermediate numerical results from theoretical calculations, etc. Also defines data as: observational, experimental, simulation, model output or assimilation	*Some details about data types are included, but DMP is missing details or wouldn’t be well understood by someone outside of the project*	*No information about data types is included; fails to adequately describe data types*	

How will reuse of the data and access to the data...?

		Complete / Detailed	Adressed issue, but incomplete	Does not address	Directorate w/ this req.
3.1	Provides details on whenthe data will be made publicly available	clearly specifies when the data will be made available to people outside of the project. *"Data generated from our research will be shared incrementally throughout the time periodof the project, and the full dataset will be available no later than one year after the project end"*	Verifies that the data will be made available outside of the project but does not identify when, such as a time frame (e.g., duration of the project, or for a period after the conclusion of the project). *"It will be the policy of the project to publish relevant findings expeditiously in the peer-reviewed literature".*	Does not specify when the data will be made available outside of the project.	All
3.2	Describes howthe data will be made publicly available	Includes specific details on the means by which the data will be made available. E.g., this may include a publically accessible data repository or a description of how the researcher or a 3rd party will provide access *"The data we collect will be made publicly accessible via submission to our university's repository service, which has capability for data set ingest and review"*	Provides vague or limited information on how the data will be made available, or details about sharing can be inferred from the mention of a repository or archive that will be used for depositing the data. *"The data will be made accessible to others upon request".*	Includes no details on the means by which the data will be made available.	All
3.3	If the data are deemed to be of a "sensitive" nature, describes what protections will be put into place to protect privacy or confidentiality of research subjects	Actions that will be taken to address the sharing of sensitive data are clearly described *"Because of confidentiality issues, each subject will be assigned an arbitrary code. Personal information will be permanently removed prior to data analysis. All data will be stored and backed-up by the PI. The de-identified electronic data will be preserved on DVDs and external hard drives. Copies of these data will also be preserved offsite at a university storage facility. Completed surveys, consent forms, and written analyses of students' lessons will be stored in a locked file cabinet accessible only to the PI"*	Actions that will be taken to address the sharing of sensitive data are vaguely described. *"Prior to publication, the data will be maintained confidential and stored in the lab, and in the PI's machine."*	Actions that will be taken to address the sharing of sensitive data are not described	All

Using data management plans as a research tool - https://osf.io/kh2y6/

https://doi.org/10.1371/journal.pcbi.1004525

PLOS COMPUTATIONAL BIOLOGY

Ten Simple Rules for Creating a Good Data Management Plan

William K. Michener

Published: October 22, 2015 • https://doi.org/10.1371/journal.pcbi.1004525

DMPTool

Build your Data Management Plan

https://dmptool.org/

University of CINCINNATI LIBRARIES

DMPTool

Build your Data Management Plan

https://dmptool.org/

Table of Contents

Choose option 1 – institution – University of Cincinnati Main Campus
Login with 6+2

Create a new plan

Before you get started, we need some information about your research project to set you up with the best DMP template for your needs.

* What research project are you planning?

mock project for testing, practice, or educational purposes

* Select the primary research organization

University of Cincinnati Main Campus (UC)

- or -

No research organization associated with this plan or my research organization is not listed

* Select the primary funding organization

Begin typing to see a filtered list

- or -

No funder associated with this plan or my funder is not listed

Create plan | Cancel

Project planning – Data Management Plan

Project begins

All the hard work

Project end – Publishing/Archiving/Sharing

Journal Article/
Data - Repository

www.libraries.uc.edu

University of CINCINNATI LIBRARIES

https://commons.wikimedia.org/wiki/File:Wikimedia_Audience_Research_Project_Internal_Synthesis_Workshop_14.jpg

Data Sharing

University of CINCINNATI LIBRARIES

Best place to share your data:

- Subject repository
- Repository that…
  - Is easy to use
  - Accepts many types of files
  - Provides description - Metadata
  - Allows content to be indexed, searched and retrieved
  - Provides permanent access – PURL, DOI
  - Has analytics – page views, downloads
- Sensitive Data needs a gatekeeper
  - ICPSR - https://www.icpsr.umich.edu/icpsrweb/
  - *Center for Health Informatics - https://www.med.uc.edu/chi

https://www.re3data.org/
https://www.nature.com/sdata/policies/repositories

* You can get data, but not necessarily store here

University of CINCINNATI LIBRARIES

UC's data sharing solution > scholar@uc

http://scholar.uc.edu

www.libraries.uc.edu

University of CINCINNATI LIBRARIES

scholar@uc

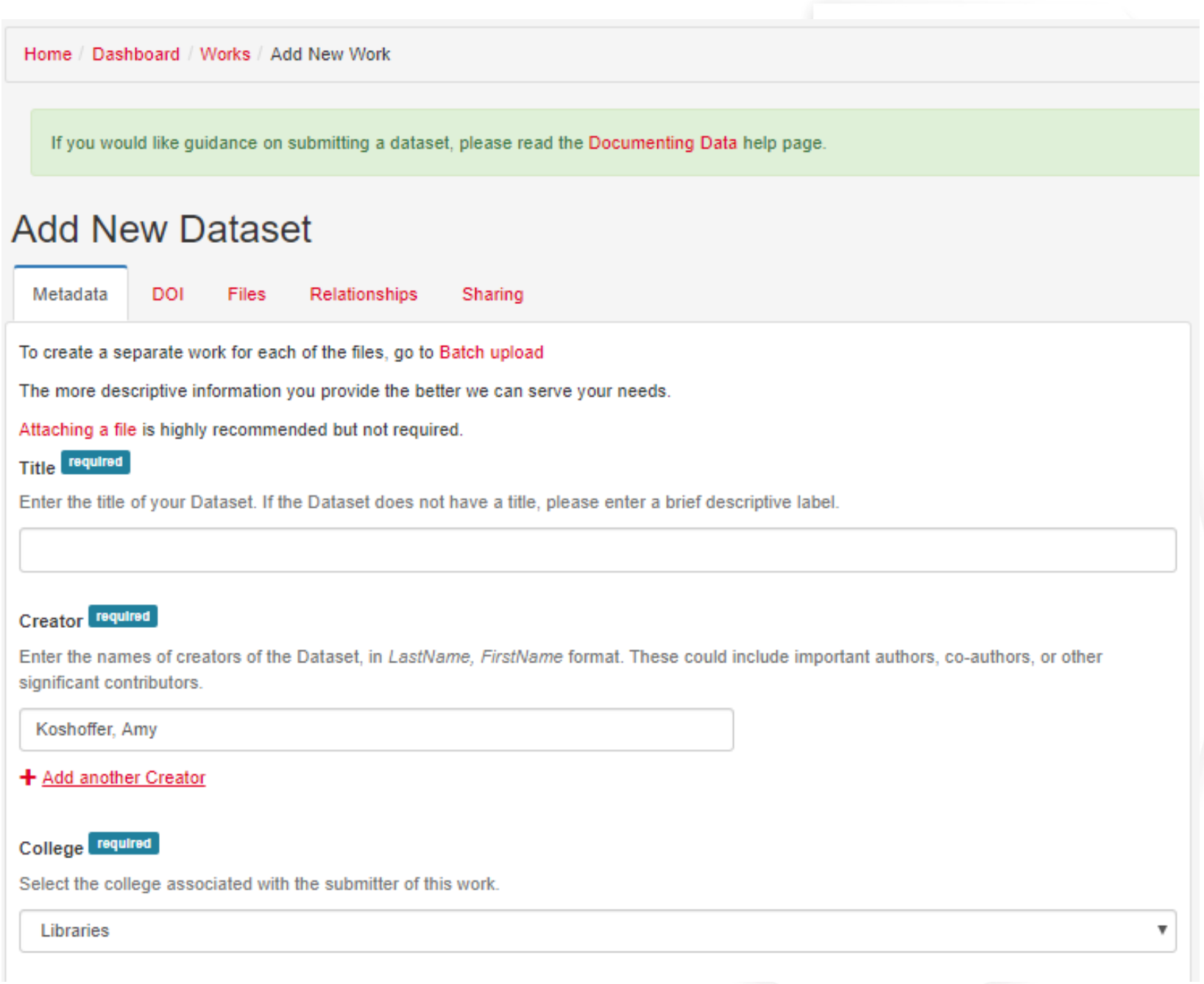
Log in with 6+2

Under the works menu, choose add new work

Choose work type - dataset

Follow the wizard
- Add metadata
- Choose to create /link DOI
- Add files
- Links to other works in Repo
- Allow access for editing
- License
- Visibility
  - private, UC only, embargoed, public

To see the UC only content, you need to be logged in. If off campus, use the VPN.



Suggested Documentation for data

- Read Me file = instruction sheet – how to use the data
  - File relationship
  - File naming convention
  - Data dictionary
    - Variable explanation
  - Terms that are unique to the dataset
  - Keywords
  - Omissions and errors explained
  - Methodology - protocol
  - Metadata following a suggested schema if appropriate
  - Contact information

Readme File Template from U of Minnesota:

https://drive.google.com/file/d/0B5Dm3XFQIoc4RDY4VEM4OFJobUk/view

Sharing Data – Dr. Christy Holland

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

https://scholar.uc.edu/collections/bc386p70c

My Collections / Clot on a Suture Dataset

Clot on a Suture Dataset Open Access

No Description

Total Items: 10
Size: 84.7 GB
Creator: Huang, Shenwen

Submitter: huangsw@mail.uc.edu
Rights: All rights reserved

Items in this Collection

Search Collection Go

Sort By: relevance Show 10 per page Refresh

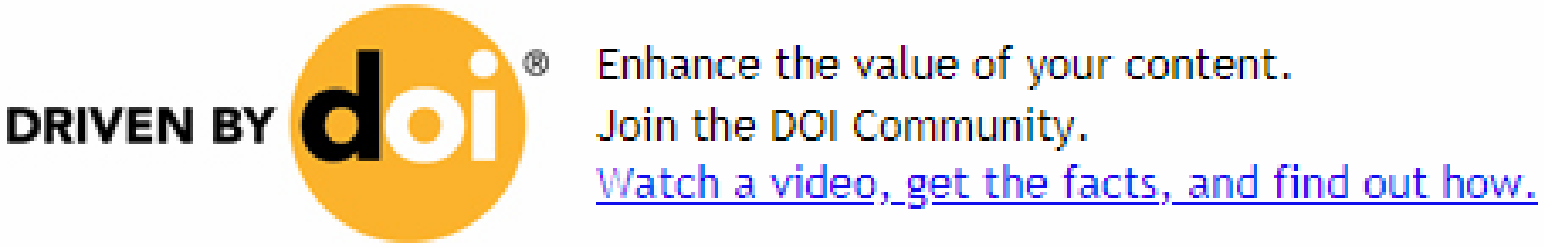
Title	Date Uploaded	Visibility
Porcine and Human Clot SEM Images	04/11/2017	Open Access

. . .

www.libraries.uc.edu

University of CINCINNATI LIBRARIES

Discoverability is...



DOI allows you to connect other resources to your work

- ORCID
- Citations

DOI Providers

- Crossref
- Datacite

University of CINCINNATI LIBRARIES

Giving Your Content a DOI - From Scholar@UC to your ORCID Profile

Attributes

Attribute Name	Values
Creator	Koshoffer, Amy
Submitter	Amy Koshoffer
College	Libraries
Department	Geolog-Math-P
Publisher	University of C
Rights	Attribution-Nor

DOI

doi:10.7945/C2NP4P

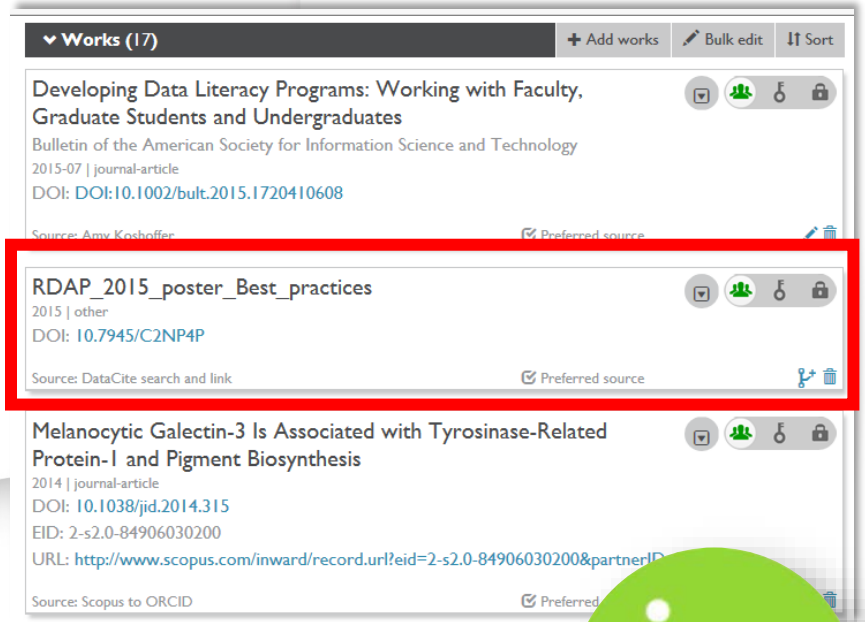
This DOI link is the best way for others to cite your work.

ORCID
Connecting Research and Researchers

Amy Koshoffer

ORCID iD

https://orcid.org/0000-0001-8130-103X



University of CINCINNATI LIBRARIES

www.libraries.uc.edu

http://orcid.org/ http://research.uc.edu/directory.aspx

Data Citation:

System can generate a statement such as “When citing these data, please use the following citation”:

Author, Year of publication, Dataset Title, Data Repository or Archive, Version of data set, Global Persistent Identifier.”

-Derived from Metadata entered

DataCite 2014:

Sarah Callaghan’s presentation (PPTX - 3.1 Mb))
(https://www.force11.org/node/4771)

University of CINCINNATI LIBRARIES

scholar@uc

Features

Multiple file type accepted
3 Gb per file upload using browser max – larger files consult Repo support
Metadata for discovery
Multiple editors
Ability to form collections of related works
Create DOIs (Datacite)
PURLs
Check sum for data integrity
Levels of access

University of CINCINNATI LIBRARIES

Resources

University of CINCINNATI LIBRARIES

UC Offices that affect Data Policies

1) Board of Trustees - http://www.uc.edu/trustees.html
2) General Counsel - http://www.uc.edu/gencounsel.html
3) Office of Research - http://research.uc.edu/home.aspx
4) Intellectual Property Office - http://www.ipo.uc.edu/
5) Information Security - http://www.uc.edu/infosec/about.html
6) Export control http://researchcompliance.uc.edu/ExportControls/ExportControls.aspx

University of CINCINNATI LIBRARIES

Data Policies

3361:10-43-18

Records: responsibilities and rights concerning ownership, access to and maintenance of original scientific records.
https://www.uc.edu/trustees/rules.html > http://codes.ohio.gov/oac/

(6) Provision of a record sufficient to demonstrate compliance by the university with applicable laws and regulations.

UC must retain the scientific record in sufficient detail to accomplish these goals for a period of not less than five years beyond the completion of the research unless statute or regulation requires a longer retention period.

Policy 9.1.1

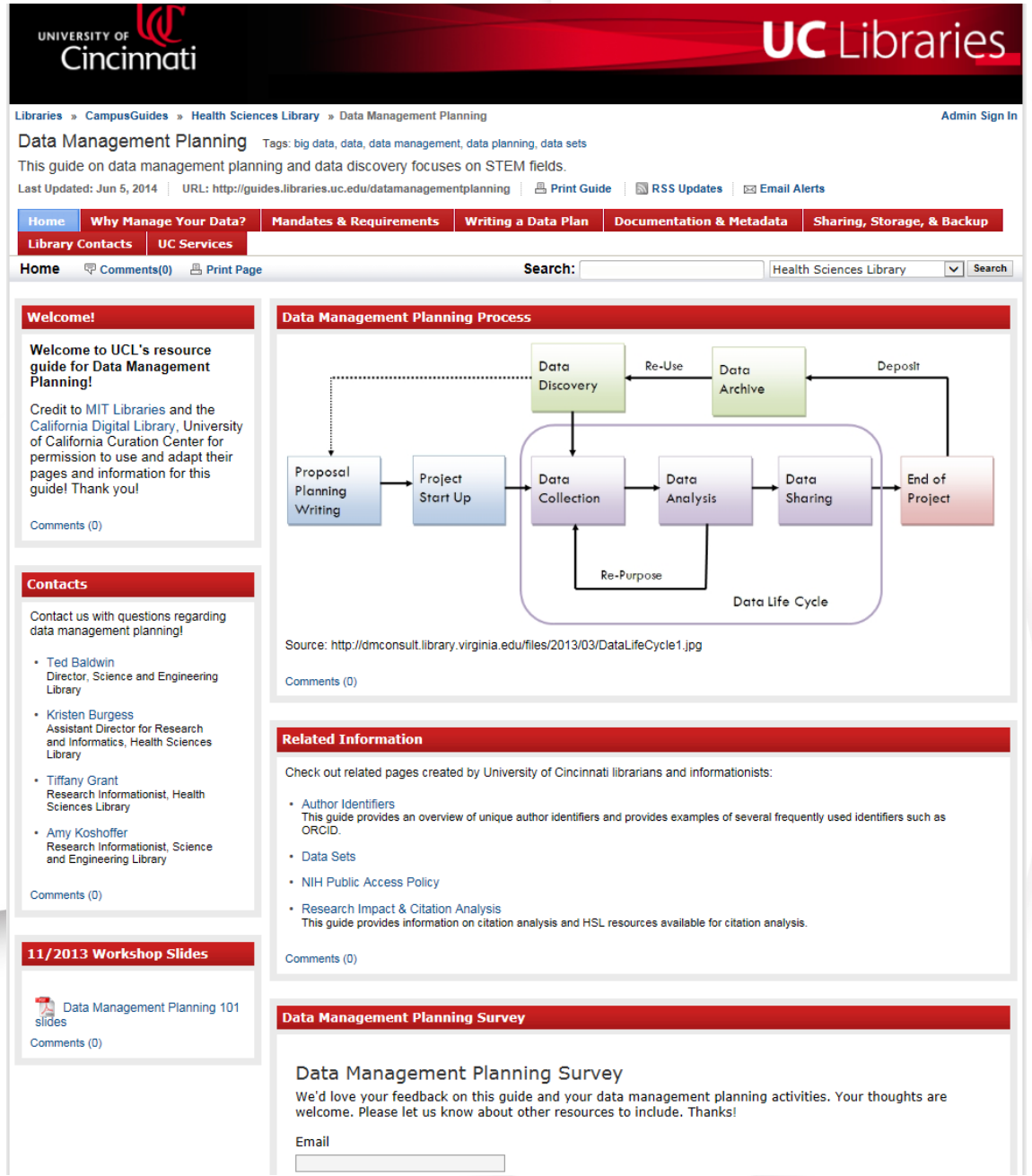
Data Protection – http://www.uc.edu/infosec/compliance.html

University of CINCINNATI LIBRARIES

The Library can help with….

- Best practices
  - Checklists
- Tools
- Workshops
- Consultation

http://guides.libraries.uc.edu/datamanagementplanning



ASKDATA@UC.EDU

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

The University has two supported storage strategies for unstructured data.

OneDrive

- Enterprise Cloud Storage - preferred solution, leverages our investment in O365 technologies
- No additional cost passed on to the individual or department
- Unlimited storage
- It is not designed to be a Departmental share - you can share files and folders but it's tied to the individuals account
- Data is secure and encrypted during sync
- Export Control Compliant - Please refer to the Universities Export Controls Office for policy.

Isilon Storage

- On Premise Enterprise SAN for unstructured data
- $.05 per GB - monthly
- Ability to grow data footprint as needed
- Data is synchronized to SOCC - off site copy
- Share permissions are managed by your local IT support or IT@UC
- Requires AD authentication
    - Externally accessible via VPN
- No backend encryption
    - Can be encrypted by mounting Isilon CIFS\NFS share to server that runs encryption software
        - Additional cost for Server
- Export Control Compliant - Please refer to the Universities Export Controls Office for policy.