




Representative Documents

Data Repositories

BRITISH COLUMBIA RESEARCH LIBRARIES' DATA SERVICES

Abacus Dataverse Network

<http://dvn.library.ubc.ca/dvn/>



British Columbia Research Libraries'
Data Services

Hosted at the University of
British Columbia

Abacus home | Login | Participating Institutions

POWERED BY THE **Dataverse Network** PROJECT v. 3.6

Create Account | Log In

Abacus Dataverse Network

Advanced Search | Tips

The **Abacus Dataverse Network** is the research data repository of the **British Columbia Research Libraries' Data Services**, a collaboration involving the Data Libraries at **Simon Fraser University (SFU)**, the **University of British Columbia (UBC)**, the **University of Northern British Columbia (UNBC)** and the **University of Victoria (UVic)**. To retrieve datasets which are licensed to all participating institutions, as well as datasets unique to your own institution, please go to the login page above and follow the **Affiliated User** instructions. For assistance, please contact the Data Librarian at your participating ...[more >>](#)

Dataverses

13 Dataverses

A Dataverse is a container for research data studies, customized and managed by its owner.


BROWSE DATERVERSES BY INSTITUTION

SFU (1)	UNBC (2)
ABACUS (2)	UVic (3)
UBC (7)	

RECENTLY RELEASED DATERVERSES

UBC Michael Smith Laboratories	Mar 29, 2017
UNBC Research Data Collection	Feb 28, 2017
UBC Faculty of Forestry	Oct 6, 2016
Canada Health Infoway Data Collection	Aug 16, 2016
UBC Library Open Data	Sep 15, 2015

[View More >](#)

a place of mind


UBC Dataverses
Dataverses: 5
University of British Columbia collection

Studies

1,965 Studies, **31,526** Files, **35,690** Downloads

A study is a container for a research data set. It includes cataloging information, data files and complementary files.

RECENTLY RELEASED STUDIES

Protease-inhibitor interaction predictions: Lessons on the complexity of protein-protein interactions by Fortelny, Nikolaus; Butler, Georgina; Overall, Christopher; Pavlidis, Paul	Apr 3, 2017
Noisy TIMIT Speech by Abdulaziz, Azhar; Kepuska, Veton	Mar 31, 2017
[Census of Canada.] Digital Boundary Files, Census Year 2016	Mar 31, 2017
GALE English-Chinese Parallel Aligned Treebank -- Training by Li, Xuansong; Grimes, Stephen; Strassel, Stephanie; Ma, Xiaoyi; Xue, Nianwen; Marcus, Mitch; Taylor, Ann	Mar 31, 2017
BOLT Chinese Discussion Forum Parallel Training Data by Song, Zhiyi; Garland, Jennifer; Walker, Christopher; Strassel, Stephanie	Mar 31, 2017


[View More >](#)

MOST DOWNLOADED STUDIES

Census of Canada. Cartographic Boundary Files, 2011	2106
CanMap Content Suite, v2016.3	1202
Canadian Community Health Survey, 2011-2012 [2013]	1037
Vancouver's Old Streams, 1880-1920 by Lesack, Paul; Proctor, Sharon J.	980
General Social Survey, Cycle 25: Family, 2011 [2013]	928

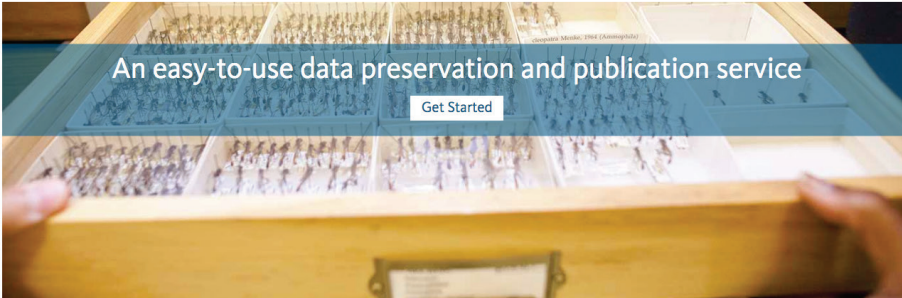
[View More >](#)

Abacus Dataverse Network - British Columbia Research Library Data Services - Hosted at the University of British Columbia © 2015



Explore Data | My Datasets | Login | More Dash Sites

Search



Get Started


"hypotheses come and go, but data remain"

Santiago Ramón y Cajal, 1897

Search 286 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.


Why use our service?




Straightforward compliance. Submit your data to satisfy publisher and funder requirements for preservation and availability with a minimum of effort. Optionally link data to the article and funding organization.



Any field. Any format. Submit data in any file format from any field of research. Share all of the data from a project in one place.



Get discovered. Each landing page and dataset is optimized for search engines. Include any relevant geo-location information to take advantage of our built-in geospatial search.



Credit for your work. Get an informative landing page to facilitate re-use of your data and a citable DOI so you can get credit. Add your ORCID identifier and we will automatically send the information to your ORCID profile.

Steps to share

- Prepare.** Gather your data and the information that others would need to use it. (Submission Basics)
- Describe.** Document your dataset. You can choose to describe your dataset a little or a lot — including geospatial locations and links to publications and datasets. (Metadata Basics)
- Upload.** Add your dataset through upload form or drag-and-drop. Each file can be up to 2GB; datasets can be up to 10GB total. (Upload Basics)

66 Representative Documents: Data Repositories

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN LIBRARY

Illinois Data Bank

<https://databank.illinois.edu/>

UNIVERSITY LIBRARY UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Illinois Data Bank [+ Deposit Dataset](#) [Find Data](#) [Policies](#) [Help](#) [Log in with NetID](#)

The Illinois Data Bank is a public access repository for publishing research data from the University of Illinois at Urbana-Champaign

You are ready to deposit data if:

- your data is in a final state and not expected to undergo revisions.
- you have removed any private, confidential, or other legally protected information from your data.
- you are a faculty member, staff member, or graduate student at the University of Illinois at Urbana-Champaign.
- you have permission to publicly distribute data from all creator(s) and/or copyright owner(s).


[Learn how to publish your data](#)


Published data:

- is open to anyone in the world.
- receives a stable identifier ([DOI](#)) for easy reference and citation.
- is readily available for anyone to access for a minimum of 5 years.
- is located in a stable environment that complies with many funder and publisher requirements.

[Review our policies](#)

The Illinois Data Bank is a product of the [Research Data Service](#) at the University Library. [See our Access and Use Policy](#). [Contact us](#) for questions and to provide feedback.


ILLINOIS
© 2016 University of Illinois Board of Trustees



Deep Blue Data

Login

HomeAboutHelpContact

Enter search terms

GoAll


Deep Blue Data is a repository offered by the University of Michigan Library that provides access and preservation services for digital research data that were developed or used in the support of research activities at U-M.

Browse

Deposit Your Work

Featured Works

Neighborhood Effects...




Deposited by:

tveinot@umich.edu

Keywords:

Recreation Centers, Census tract level, Recreation Areas, Park Land, Spatial Measures, Michigan, Metropolitan Detroit


The Observation and...



Deposited by:

dcsoztrk@umich.edu

Citations to Open and...




Deposited by:


helpop@umich.edu

HelpHomePolicy & Terms of Use

© 2015 Regents of the University of Michigan

**LIBRARIES**
digital conservancy

Search Browse Help Sign in

Data Repository for U of M

The Data Repository for University of Minnesota (DRUM)

DRUM is a publicly available collection of digital research data generated by U of M researchers, students, and staff. Anyone can search and download the data housed in the repository, instantly or by request.

The Data Repository accepts submissions from University affiliates for digital archiving and access. [Learn more](#) about depositing to the Data Repository and other services to manage your data.

*U of M affiliates only | [How to submit](#)

How to Upload


- 1. Prepare Data**


Data should be free of identifying or sensitive information and include adequate documentation. Not sure? [Contact us for help!](#)
- 2. Upload**


Have your files ready (up to 2GB each) and use the [upload form](#) to fill out metadata about your data.
- 3. Curatorial Review**

Our data experts will consult with you to ensure that your data is in a format and structure that best facilitates long-term access, discovery, and reuse.

Features

-  **Flexible Access Options**

Choose to make your data immediately accessible to everyone, or moderate access to your data upon request.
-  **Meet Grant Requirements**

Comply with federal mandates for data management planning (DMP) and sharing. [Read more.](#)
-  **Persistent Access**

A DOI for your data guarantees no more broken links when others include this persistent link to your data in scholarly articles as a bibliographic citation.

Our Services


- Data Management Plan Assistance**

We offer personalized assistance for drafting your next grant's Data Management Plan. [Contact us](#) for assistance during your planning process.
- Metadata Consultation**

We can help structure your data using disciplinary best practices to ensure the best organization of your data.
- Training and Workshops**

The library offers free online and drop-in workshops on data management best practices periodically throughout the year.

To include DRUM in your next Data Management Plan, contact us to get boilerplate text and more information on how the Data Repository can be incorporated into your next grant.



[Help](#)
[Contact Us](#)
[My Account](#)

RUCore: Rutgers University Community Repository

[Search](#)
[Services](#)
[Collections](#)

[Home](#)
[RUresearch Data Portal](#)

RUresearch Data Portal

The **RUresearch** Data Portal provides a place for Rutgers researchers to share their research data with the global scholarly community. **RUresearch** leverages all the capabilities of RUCore and adds additional tools and services specific to research data.

Search [More...](#)

[Browse All Projects](#) | [Advanced Search](#)

Learn More

RUresearch Services

- **Free assistance and advice** to design your data strategy and prepare the data Management component of a grant proposal.
- A **customized search and retrieval portal** for your data.
- **Ongoing management and support** for your data
- **Is there a fee for placing my data in RUresearch?**
- **Guidance for the NSF Data Management Grant Requirement**

What does RUresearch Provide

- **Robust preservation** and long term access
- **Metadata** customized to your project's needs or community practice (e.g., DDI, Darwin Core, etc.)
- **Access control**, including flexible embargo periods and statements of use
- **All associated information**, presented in Context, such as codebooks, lab notes, Images, software, related publications, etc.
- **Contact the RUresearch Data Team**

About RUresearch

[RUresearch Data Team](#)

[NSF Data Management](#)

[Data Management and RUresearch Presentation](#)

Highlighted RUresearch Collections

[Video Mosaic Collaborative](#)

[Ralston Horse Behavior Research](#)

About Us

[How does RUCore work?](#)

[Policies](#)

[Services](#)

[Participate in RUCore](#)

[Collections](#)

Infrastructure

[Trusted Repository](#)

[Preservation](#)

[Understanding Metadata](#)

[Technical Glossary](#)

Open Source

[OpenWMS](#)

[OpenETD](#)

[OpenMIC](#)

[OpenWAAND](#)


Developers

[Reference Materials](#)

[Web Services/APIs](#)

[Schemas](#)

[Harvesting](#)


[Statistical Profile](#)

• [Version 8.2.3](#)

[Site Search](#)

• [Privacy Policy](#)


[libraries.rutgers.edu](#)
Copyright ©2017 Rutgers, The State University of New Jersey(Further Copyright Information)

SCHOLARS PORTAL

Dataverse

<https://dataverse.scholarsportal.info/>

The screenshot shows the homepage of the Scholars Portal Dataverse. At the top, there is a navigation bar with the text "Scholars Portal Dataverse" on the left, a search bar with the placeholder "Search all dataverses..." and a red "Search" button, and a menu of links: "About", "Guides", "Support", "Sign Up", "Log In", and a "Fr (Beta)" link on the right. Below the navigation bar is a large teal banner. On the left side of the banner is a white geometric line pattern. On the right side, the text "SCHOLARS PORTAL RESEARCH DATA PLATFORM" is displayed in white. Below this, the text "Publish and track your data, discover and reuse others' data!" is written in a large, white, sans-serif font. At the bottom of the banner, there are two buttons: a red "Create My Dataverse" button and a white "or Explore" button with a magnifying glass icon. Below the banner is a horizontal carousel of four university logos and their corresponding dataverse names: "University of Windsor", "Western University", "LAURIER Wilfrid Laurier University", and "YORK U York University". The carousel has left and right arrow navigation buttons. Below the carousel, a paragraph of text states: "Canada's federal research agencies are strong advocates for making publically-funded research data as accessible as possible. In 2016, the Tri-Agency released a [Statement of Principles on Digital Data Management](#) that outlines expectations and responsibilities for research data management and open data sharing."


**Texas Digital Library**

HOMEABOUTNEWSMEMBERSSERVICESTRAININGSUPPORT

[Home](#) › [Texas Data Repository](#)

Texas Data Repository

The [Texas Data Repository](#) is a platform for publishing and archiving datasets (and other data products) created by faculty, staff, and students at Texas higher education institutions. The repository (<https://dataverse.tdl.org/>) is built in an open-source application called Dataverse, developed and used by Harvard University.



The repository is hosted by the Texas Digital Library, a consortium of academic libraries in Texas with a proven history of providing shared technology services that support secure, reliable access to digital collections of research and scholarship.

Benefits of a Texas Data Repository

- **Compliance with funding requirements.** The Texas Data Repository helps researchers comply with funder mandates for data archiving and sharing, and supports research grant-seekers by having infrastructure available at the time of proposal.
- **Reliable, managed access for data.** The Texas Data Repository provides a convenient and reliable place to collect and share data. And by depositing data there, researchers benefit from the Texas Digital Library's focus on long-term access and preservation of scholarly content.
- **Increase scholarly impact.** By publishing their data in the Texas Data Repository, researchers give their data credibility through a unique, citable, and persistent online identifier (i.e., a Digital Object Identifier), which makes it easy for others to cite reliably.
- **Collaboration with research teams.** Some situations may necessitate restricting access to data, at least for a period of time. The Texas Data Repository allows researchers to share their data with a select group of colleagues, version the data, and publish it when they're ready.
- **Access to local support through their institution's library.** Along with robust technical support from the TDL, users of the Texas Data Repository can rely on trained librarians at their home institution to assist with multiple phases of the research cycle, including data management planning preparation for data publishing, and long-term curation.
- **Efficient use of resources.** By pooling resources across multiple institutions, the Texas Data Repository realizes cost savings through a shared infrastructure while showcasing local contributions through university-branded data collections and local library services. Each institution can focus its resources on unique services that meet local research community needs.

How the Texas Data Repository Works

The Texas Data Repository is designed for regular to mid-sized data sets (individual file sizes up to 2 GB), which comprises the majority of research data. These data can include:

- Data from any scholarly discipline and in any file type
- Materials such as codebooks and other supplementary documentation
- Data that does NOT contain confidential or sensitive information (like social security numbers or other identifiers)

Researchers affiliated with participating TDL member institutions will be able to:

- Store and organize data sets and upload files
- Maintain multiple versions data sets
- Share data sets online with trusted colleagues OR release data for public access online
- Get recognition and proper academic credit for scholarly work through a data citation with a persistent identifier (i.e., a DOI, or digital object identifier)

Library faculty or staff at each of TDL's participating member institutions will provide local assistance to researchers at their institution as they prepare and deposit their data.

- Each participating university will have its own branded "dataverse" within the overall repository, which it can use to showcase its researcher contributions.

Participate in the Texas Data Repository


Institutions interested in participating in the Texas Data Repository must be an institution of higher learning in Texas and a member of the Texas Digital Library. To find out more about membership opportunities, please see the [Membership](#) section of our website.

If your TDL member institution decides to participate, all faculty, staff, and students at your institution will be able to deposit their datasets. Anyone may view or download datasets in the Texas Data Repository, but only individuals from a participating TDL member institution may deposit datasets.


TDL members should contact the TDL (info@tdl.org) to begin utilizing this new service. The process includes:


- Sign a Memorandum of Understanding
- Establish authentication systems on your campus (e.g., Shibboleth or Two Factor)
- Identify a Texas Data Repository liaison on your campus



 VirginiaTech
Invent the Future®


VTechData

Search VTechData...  All ▾

Login 

HOME ABOUT HELP CONTACT

Welcome to VTechData!





Featured Datasets Recently Uploaded


Featured Researcher Explore Tags

Featured Dataset

Featured Researcher
[View other featured researchers](#)


PURPOSE


ACCESS POLICY


DEPOSIT POLICY

The purpose of VTechData is to highlight, preserve, and provide access to the work of faculty, staff and students, as well as the intellectual output of the world's community through the discovery and dissemination of new knowledge.

VTechData was designed to make data and other research products openly available to the general public; however, we recognize that temporary access restrictions may be required for certain research situations.

By depositing data or other research materials into Virginia Tech's Data Repository, you affirm that the deposit represents your own work or the work of your collaborators; any work that is not your own must be properly cited.

WASHINGTON UNIVERSITY
OPEN SCHOLARSHIP

[Home](#) [About](#) [FAQ](#) [My Account](#)

Enter search terms:

in this series ⌵

[Advanced Search](#)

[Notify me via email or RSS](#)

Links

[Data Management Research Guide](#)

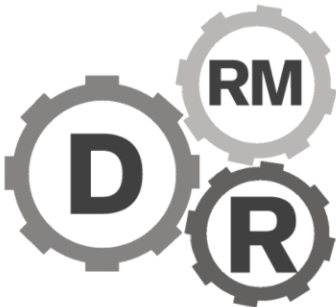
Browse

[Collections](#)
[Disciplines](#)
[Authors](#)

Author Corner

[Author FAQ](#)
[Policies](#)
[Submission Guidelines](#)
[Submit Research](#)

Home > [DATA](#)



DIGITAL RESEARCH MATERIALS
(DATA & SUPPLEMENTAL FILES)

Digital Research Materials Repository

This Digital Research Materials series is a place for WUSTL affiliates, including faculty, students, and affiliated researchers, to share and publish digital data and supplemental files for long-term access and future use. While some academic disciplines have established research data repositories, many fields of research do not have easily available options for archiving and online access. Benefits include:

- **Flexible Access Options:** Make your data immediately accessible to all, or moderate access to your data upon request.
- **Long-term Access:** Persistent links and [identifiers](#) (DOIs) make it easy for others to cite your data.
- **Analytics:** Track how often your data are viewed and downloaded.
- **Meet Grant Requirements:** Comply with federal mandates for data management planning (DMP) and sharing (see a list on our [Data Management Research Guide](#)).
- **Maximize Reusability:** Our data experts will consult with you to ensure that your data are in a format and structure that best facilitates long-term access, discovery, and reuse.

To get started, determine if your data are ready to upload by reviewing the [Policies](#) and [Submission Guidelines](#). If you need help, contact your [subject librarian](#) or use the tools available on our [Data Management Research Guide](#). Once you are ready, sign-in and begin uploading your data to the Data Collection. A data curator will email you with any questions about your upload and next steps within two working days.

SPEC Kit 354: Data Curation

75

Data Curation Services

LOCATIONS

CONNECT FROM OFF-CAMPUS

HOURS

ACCESSIBILITY

ACCOUNTS

GIVING

UCI Libraries

FIND

Books & Collections

SERVICES

Service Points & Tools

ABOUT

Visiting & Contacts

HELP

Ask a Librarian & Guides

Search for books, journals, guides, web pages

DIGITAL SCHOLARSHIP

What We Do

DSS collaborates with UCI faculty members, students, and administrators to transform research and scholarly communication using new media and digital technologies. DSS manages content, infrastructure, and user interfaces for all stages of the research lifecycle, including research planning, execution, dissemination, discovery, preservation, and impact.

Scholarly Communication

Make your scholarly work accessible and interactive. We help you manage your information through the research process and communicate research results to the broader scholarly community and to the public:

- Use tools to manage citations and research output
- Comply with [UC Open Access Policies](#)
- Deposit publications and other research results into open-access [repositories](#)
- Publish in open-access journals
- Understand your rights as author
- Use online tools to manage your research impact and scholarly reputation

Go to page

DATA Curation

Data curation is the active management of data to maintain and extend its value over time. It includes effectively organizing data for access, documenting context for reproducibility, and securely preserving the physical integrity of the work. We help you with all stages of data management required by funding agencies:

- Write grant winning [Data Management Plans](#)
- Deposit data into [repositories](#) for access and preservation
- Capture [metadata](#) to allow re-use
- Create permanently resolvable hyperlinks
- Connect your data with your publications

Go to page

Digital Production

Create or support creation of themed digital collections that are openly accessible for scholarly use, including text, image, audio, video, or other media types. We can help you:

- Appraise and digitize content in traditional media for discovery and online access
- Design [metadata](#) solutions suiting the nature of content and the purpose of project
- Manage digital assets and make them accessible online in user-friendly ways

Go to page

Digital Preservation

Digital preservation is the active management of digital content over time. It combines policies, strategies, and actions to ensure ongoing access to and accurate rendering of authentic reformatted and born-digital content. We can help you:

- Migrate content in obsolete or fragile formats to sustainable formats
- Store content in trusted [repositories](#) for long-term access
- Use reliable tools to harvest, build, and preserve collections of web content

Go to page

DSS Home

What We Do

Who We Are

Projects & Activities

Training & Events

FAQ

Schedule DSS Consultation

Email DSS

Site Feedback

SPEC Kit 354: Data Curation

77

LOCATIONS
CONNECT FROM OFF-CAMPUS

HOURS
ACCESSIBILITY
ACCOUNTS
GIVING

UCI Libraries

FIND
Books & Collections

SERVICES
Service Points & Tools

ABOUT
Visiting & Contacts

HELP
Ask a Librarian & Guides

Search for books, journals, guides, web pages
DIGITAL SCHOLARSHIP

Data Curation

Data curation is the active management of data to maintain and extend its value over time. It includes effectively organizing data for access, documenting context for reproducibility, and securely preserving the physical integrity of the work. We can help you with all stages of data management required by funding agencies:



DMPTool

Explore

Create, review, and share [data management plans](#) that meet institutional and funder requirements. Examples:

- NEH Office of Digital Humanities
- NIH Generic
- NSF Generic
- NSF Biological Sciences
- NSF Engineering
- NSF Social, Behavioral, Economic Sciences



dash

DATA SHARING MADE EASY

Explore

A simple self-service tool to archive and share your research data for accelerated advancement of knowledge. Examples:

- The Nature Conservancy San Diego Field Office (2015): Oak Woodland Monitoring for the Irvine Open Space Easements (4). UC Irvine. Dataset. doi:10.7280/d1001p
- Baldassare, Mark (2014): Annual Survey of Orange County 1986. UC Irvine. Dataset. doi:10.7280/d1059f
- Bannan, Caitlin C; Calabró, Gaetano; Kyu, Daisy Y; Mobley, David L (2016): Simulation input and out and data analysis for calculating partition coefficients of small molecules in octanol/water and cyclohexane/water. UC Irvine. Collection. doi:10.7280/d15k5m



EZID

Explore

A service making it easy to create and manage long-term, globally unique identifiers for your data and sources, ensuring their future discoverability and avoiding link rot. Examples:

- [ark:/87278/s6bv81tf](#)
- [doi:10.5061/DRYAD.HM482/1](#)

[Learn about ARK](#) | [Learn about DOI](#)

DSS Home

What We Do

Who We Are

Projects & Activities

Training & Events

FAQ





UC SANTA BARBARA LIBRARY

[HOURS](#) | [LOGIN](#) | [GIVING](#)

[Research](#) [Services](#) [Visit](#) [About](#) [Q](#)

Data Curation

Home

- Data Curation
- Resources
- Related Services
- Contact Us

Contact Us
datacuration@library.ucsb.edu

The Data Curation Program supports active and ongoing management of data produced by UCSB researchers throughout the research process to cultivate broader awareness of the continued utility of data within, and possibly beyond disciplinary origins. We aim to ensure that data are well-described and findable; accessible and usable; citable; and sustainably preservable.

We offer the following services:

Planning

- Data management planning and proposal assistance
- Funder data management requirements
- Using the [DMPTool](#)

Day-to-day management

- Best practices in data management
- Data management tools and services
- Metadata standards
- Metadata population and extraction strategies

Publication

- Data publication journals
- Obtaining persistent identifiers for data
- Data citation formats
- Data copyright and licenses

Preservation/archival

- Selection and usage of data repositories

Typically, we offer consultation and teaching though we are open to other forms of engagement.

INFORMATION FOR:

Undergraduates	Faculty	Users with Disabilities	Library Employees
Graduate Students	Staff	Visitors	Terms of Use

UC SANTA BARBARA LIBRARY
SANTA BARBARA, CA 93106-9010

UCSB UCSB LIBRARY - MAIN (805) 893-2478 | MUSIC LIBRARY (805) 893-2641
COPYRIGHT © 2010-2017, THE REGENTS OF THE UNIVERSITY OF CALIFORNIA, ALL RIGHTS RESERVED.

[f](#) [t](#) [t](#) [v](#)

JOHNS HOPKINS LIBRARIES | **Data Management Services**

Home Consulting Training Archiving Data Management Resources ▾ About Us ▾ 🔍

Data Management Services

~~ Serving Researchers on Homewood and Medical Campuses ~~

Effective data management can increase the pace of the research process, contribute to the soundness of research results, and meet funding agency requirements by making research data easy to share.

Supporting data management at every point in your research

↓

Where are you in your project?

Just Starting	In Progress	Almost Done
Plan Your Research Project	Manage and Analyze Data	Publish and Share Data / Software

Our Services

Consulting	Training	Archiving
Assistance with writing data sharing or management plans	Best practices for data management and sharing	Archive your research data for sharing & preservation

LIBRARY

UNIVERSITY OF MICHIGAN

[Accessibility](#)
[My Account](#)
[★ Favorites](#)
[Log in](#)

[Services](#)
[Libraries & Departments](#)
[About](#)

[MGet It](#)
[Search Tools](#)
[Catalog \(Mirlyn\)](#)

Search

[MLibrary](#)
[ArticlesPlus](#)
[Catalog](#)

Search

Browse

[- Recommended Resources by Subject -](#)

[Ask](#)

[Get Help](#)

Research Data Services

[Data Management Planning](#)
[Federal Funding Agency Requirements](#)
[Data Management Plan Tools](#)
[DMP Review Service \(CoE\)](#)

[Discovery & Access](#)
[Data Organization & Management](#)
[Metadata & Documentation](#)

[Data Sharing & Publication](#)
[Deep Blue Data Repository](#)

[Preservation](#)
[Data Visualization](#)

[Home](#) > [Services](#) > Research Data Services

Research Data Services

Your research data are important. Research Data Services is a network of services throughout the Library to assist you during all phases of the research data lifecycle. For questions about research data or to schedule a consultation, please get in touch with your [subject librarian](#) or [email us](#).

We provide or are planning to provide services in the following areas:

Data Management Planning: helping plan for managing, sharing and curating data and develop Data Management Plans (DMPs) that meet funder requirements.

Discovery & Access: assisting in discovering, accessing, and acquiring different types of research materials, including data.

Data Organization & Management: helping researchers to understand, develop and apply strategies for organizing and managing their data.

Metadata & Documentation: locating standards for documentation that capture the details of generating, processing and analyzing data so it can be discovered, understood and reused.

Data Sharing & Publication: helping disseminate research data for discovery, access and reuse in ways that enable researchers to receive credit for their work.


Preservation: taking action to sustain the accessibility and scholarly value of data over time.

Data Visualization: a rich and diverse set of practices, methodologies and tools from hand drawn charts to interactive web maps to immersive 3-D environments.

Images courtesy of Dmitry Sunseifer, anbilero adaleru, Zlatko Najdenovski, Norbert Kucsera, Giorgio Ubaldi, and Pascal Conil-lacoste via the [Noun Project](#).

Now Available!

[Learn more about Deep Blue Data](#)




LIBRARIES
digital conservancy

Q Search


Browse ▾

Help ▾

Sign In



Data Repository for U of M



Our Suite of Services

DRUM is part of a broad suite of services, offered at no cost, that support research data management on campus. Larger data projects may need additional support; please [request a consultation](#).

We also refer you to other campus-based services. See <http://lib.umn.edu/datamanagement> for additional information.

</tbody> </table>

Plan

Support for Writing Data Management Plans (DMP)

We provide one-on-one assistance, templates, and consultation when writing DMPs, now required by some grant funding agencies (e.g. NSF and NIH).

- Grant deadline? [contact us](#) for timely consultation and feedback on your next DMP.
- Not sure where to start? Use our [online templates](#) to help draft a DMP.
- Sensitive data? Learn how to create [IRB content forms](#) that facilitate sharing later.

Manage

Metadata Consultation and Data Management Training

Data management consultants will help with topics such as:

- [Metadata advice and consultation](#). Larger projects may want to include library staff as cost-share on grant-funded projects (limited availability).
- Providing on-demand [training and education modules](#) to get you and your staff up to speed on data management best practices.
- [Referral to campus services, such as storage, backup, analysis](#) and tools for [preparing sensitive data for sharing](#) (e.g., de-identification).

Share

Data Repository and Curation Services

UMN researchers may deposit data to the Libraries' Data Repository for U of M (DRUM), subject to our [collection policies](#). Data sets submitted to the Data Repository are reviewed by data curation staff to ensure that data is in a format and structure that best facilitates long-term access, discovery, and reuse.

- Read more [about the Data Repository for U of M \(DRUM\)](#)
- Browse the [collection of available data](#) accepted to DRUM.

Grant Partner

Full Life-cycle Support for Data

The Libraries offer a full suite of support for PIs that seek support for data management, data sharing, and long-term preservation for data generated by funded grants. Costs determined per grant-funded project; [contact us](#) to discuss your project.

- Want to include the Data Repository for U of M in your next grant? Use our [boilerplate language for your Data Management Plan](#).

On request

DRUM

About the Data Repository

Services for Your Data

DRUM Policies

Contact Us

82 Representative Documents: Data Curation Services

UNIVERSITY of NOTRE DAME

HESBURGH LIBRARIES

CurateND

Search CurateND

ABOUTLOG INFAQ

ABOUT

Curating Notre Dame's Research and Scholarship for Study throughout Time

MANAGEPRESERVEDISCOVERSHARE

Manage

- Are you required to have a data management plan?
- Do you work with a team or need to regulate levels of access?
- Do you want to link your work to supplemental materials of multiple formats?
- Do you want to curate and elevate your entire portfolio of work?

CurateND is a research portal that allows for the management of your research portfolio on several levels. In response to funding mandates, portal features and our related [Center for Digital Scholarship](#) services offer streamlined consultation for the creation of data management plans. Upon deposit, you can select different access privileges and embargo periods that allow you to share research results with groups and regulate access to your works while in progress.

In the digital age, research "data" comes in multiple formats, often within a single published work. CurateND is designed to accept, manage and securely preserve files of any format including: datasets, articles, images, video, whitepapers, presentations, and more. And, you can deposit and illuminate as much of your portfolio of work and supplemental materials as you choose.

Get Started >>

MANAGEPRESERVEDISCOVERSHARE

Preserve

- Does your funding mandate a plan for long-term preservation?
- Do you want assurances that you can locate your research and related works in the future?
- Do you want to preserve conference proceedings?
- Is it important to steward your research for future scientists and scholars?

CurateND employs preservation standards that meet the requirements of funding agencies for long-term preservation and curation over the life-cycle of research. Whether or not a project is grant-funded, our preservation standards will give you security and peace of mind that you will find your work in the future and it will be guarded against corruption.

And, should the need arise, the platform will migrate files to new formats for continued access and usability. All this is to ensure that your work and the work of your students is preserved for future study throughout time.

Get Started >>

MANAGE PRESERVE DISCOVER SHARE

Discover

- Do you want help to illuminate your published research globally?
- Are you interested in increasing your discoverability and citations?
- Do you want to share working papers or negative results?
- Do you use non-traditional publishing platforms?

CurateND was built to provide a first-class search experience through the portal. And, with well-structured metadata and optimization, we help ensure that publicly available content is listed prominently by outside search engines. Global access through one main platform increases discoverability and allows central tracking for downloads and citations at the individual, department, college or university level.

If you have other working papers and research results important to share with your research community around the world, CurateND is a perfect solution. And, if your research involves multi-media and non-traditional publishing platforms, CurateND will feature these works and seamlessly link to all related materials.

Get Started >>

MANAGEPRESERVEDISCOVERSHARE

Share

- Are you required to share your research, data and related works?
- Do you need to create a DOI for citing and sharing your data?
- Do you have images, posters, presentations, collections, white papers or datasets that you want to share?
- Do you want to highlight the work of graduate and undergraduate research?

For those with grant-funded research and data sharing mandates, CurateND puts your front-end data management plan into action. You have the ability (rights not withstanding) to share content at any level—from restricted access, to lab or campus access, to open access for the world.

CurateND can create a DOI on demand, linking to works on your behalf. A DOI is a convenient (and often required) way to cite your data in publications and it makes it easy for others to cite your work. You can share all of the associated work and multiple data formats that are not supported by the publishing platform.


It is equally valuable for featuring the important contributions of undergraduate and graduate research across all disciplines. All members of the campus community can create an account and contribute to intellectual fabric that is Notre Dame.


Get Started >>

CurateND is a service of the Hesburgh Libraries of Notre Dame.
Questions? Call (574) 631-6258 or email curate@nd.edu.

[Help](#)

Copyright © 2017 University of Notre Dame

UNIVERSITY OF
NOTRE DAME
Hesburgh Libraries



RUTGERS
University Libraries

RUcore: Rutgers University Community Repository

Help Contact Us My Account

Search Services Collections

Home > **RUresearch Data Portal** > About


About RUresearch

Free Assistance and Advice

We will assist you with preparing grant proposals and designing your data strategy. The RUresearch team consists of experienced digital information professionals who work with data, write and manage grants and serve as peer reviewers and consultants for granting agencies, including the National Science Foundation.

We can offer data management advice in:

1. Identifying your data model. What data are you capturing and how does it interact with other data in your research environment?
2. Designing a metadata strategy. How can I describe my data to ensure that colleagues in my field and those in other disciplines can find and reuse the data?
3. Capturing your data. What is the best methodology for capturing my data, for further analysis and for sharing with others? Is this a spreadsheet, a database, an XML document, or something else? If my community already has a data format, how can I readily transfer the data I collect to that format?
4. Making your data discoverable in your search portal. RUcore offers a search portal to your data that can be easily incorporated into your project website. As an example, see the [Video Mosaic Collaborative](#) and the [Equine Science Center](#) collection. We can help you select the right data elements and record displays to ensure that end users can find and use your research.



Data Management and RUresearch Presentation by Ryan Womack

Customized Search and Retrieval Portal

Finding resources that meet your information needs depends on the metadata, or descriptive information, used. This metadata should reflect the terminology of your field as well as information that is meaningful to enable you to find and select the best information for your needs. [Learn more about metadata.](#)

At the same time, metadata should be standardized, consistent and enable your data to be shared with a global, multidisciplinary audience.

RUcore employs a sophisticated, flexible metadata strategy that can customize metadata to support your primary audience yet still be compatible with prevailing metadata standards. For an example, see the [Video Mosaic Collaborative](#), an NSF-funded mathematics education video collection. Metadata is customized to reflect mathematics education practices and to support core audiences of mathematics education faculty, researchers and practicing teachers.

RUresearch includes a portal application that enables you to select metadata elements to filter a search and to display in search results. The search and retrieval portal is easily incorporated in your project's website using a technology known as "iFrames."

Ongoing Management and Support for Your Data

RUcore, which includes RUresearch, is an important, core service for the Rutgers University Libraries. Many library faculty and staff are engaged in its support. The Rutgers University Libraries are recognized leaders among their peer institutions in digital repository development and have contributed significant open source software to the field. We are committed to the long time persistence and availability of your data and are continuously developing new tools and services as well as upgrades to the RUcore platform to manage the digital resources we support. You can be confident in the long-term sustainability of data deposited in RUcore.



[Help](#)
[Contact Us](#)
[My Account](#)

RUCore: Rutgers University Community Repository

[Search](#)
[Services](#)
[Collections](#)

[Home](#)
[RUresearch Data Portal](#)
[About](#)

Is there a fee for placing my data in RUresearch?

RUCore accepts all types of resources that represent the significant intellectual output of the university. This includes faculty journal articles and other scholarly publications, theses and dissertations for degrees awarded by Rutgers University, and resources such as data sets that result from the research process. Individual resources, such as individual data sets that involve simple cataloging and storage, such as the example data sets currently available in the RUresearch portal, can be accepted at no cost. The same is true for electronic journal article preprints and post prints.

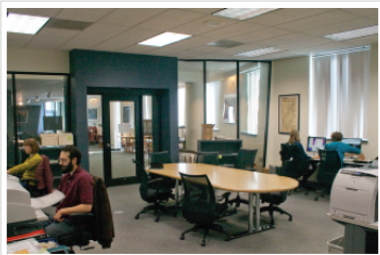
The Library will consult on your data management plan or grant at no cost, but managing data for a large research project, such as projects generally funded by grants, involves significant work and planning that will generally require a fee for service. The services we offer include customizing metadata and providing both ongoing cataloging and storage and management of data and associated documents and software. This fee can be accommodated through cost recovery charges in the grant budget, either as a data management fee or through the involvement of library faculty and staff as co-P.I.s or researchers on the grant, with associated line item cost recovery. This will be a one-time, cost recovery only fee that can be incorporated into the grant proposal budget. Data will be preserved and made accessible for the long term at no additional cost to the project beyond the one-time initial cost. However, that initial cost, although negotiable, will be based on the amount of work and effort anticipated for the life of the project.

Robust Preservation


The Rutgers University Libraries' RUCore initiative includes a Data Curation Research Center and a Data Curator who participates actively in digital preservation research and development. The Rutgers University Libraries are internationally recognized as being on the forefront for digital preservation standards and practices, particularly for digital video. We currently employ "industry best practices" for digital file preservation, including:

- Multiple backups and restoration practices, including online, nearline, off line and offsite storage of files.
- Continuous file integrity checks, such as checksum assignment and checking
- Persistent identifiers that use metadata to continuously locate a file, even if it is moved during routine storage reallocation. When you reference a citation URL, you can be confident that the file will be retrieved.
- Storage of files in multiple formats. One or more canonical formats that are vendor independent and conform to non-proprietary standards are employed whenever feasible. The original file format is also always maintained. We are currently transcoding most numeric data sets to comma separated values (CSV) format. We are also currently investigating XML (eXtensible Markup Language) and RDF (Resource Description Framework) for web based canonical formats, as well as community specific data standards such as the DDI (Data Documentation Initiative) for social science and survey data, and SensorML for sensor data. If your community uses a specific data storage format, we will explore its use with you.

[Learn more about preservation](#)



Data Curation Research Center



RUTGERS
University Libraries

RUcore: Rutgers University Community Repository

[Help](#)
[Contact Us](#)
[My Account](#)

[Search](#)
[Services](#)
[Collections](#)

[Home](#)
[RUresearch Data Portal](#)
[About](#)

Metadata

Metadata is simply "data about data." Metadata helps the data owner organize and manage the data he or she creates. It's primary role, however, is to make sure that data can be discovered and reused by others. Well-designed metadata should support four core user needs, known by the acronym "FISO", for "find," "identify," "select" and "obtain."

- Can the user **FIND** the information he is seeking?
- Can the user **IDENTIFY** what she has found? E.g., if the user is looking for a video, does the metadata record clearly indicate that the described resource is a video?
- Can the user **SELECT** the most appropriate resource, when several are retrieved, based on the metadata records. E.g., if the user is looking for air quality sampling that measures nitrous oxide levels, can he determine which resource among many air pollution data sets includes nitrous oxide sampling?
- Can the user **OBTAIN** the resource quickly and easily from the metadata record?

Good metadata is responsive to the information needs of its user community. It captures the information most important for that community, using terminology that is accurate, current and meaningful to that community. It also needs to be consistently applied and shareable with a broader community. Metadata standards evolved to enable consistency and broader sharing of information. One of the oldest and most famous is *Dublin Core*, a 15-element metadata standard that is widely used. Many research communities have evolved their own standards, such as *Darwin Core* for biological specimens and DDI (data documentation initiative) for survey-based data. RUcore employs a very flexible, sophisticated event-based metadata implementation that supports many different metadata standards but is largely independent of any one standard. We can display and export records in many different standards, including the standard *your* community uses. We can also design customized metadata that can support many standards or serve as a community standard, specific to your project's needs.

[Learn more about metadata](#)

Access Control

RUcore can assist you with controlling access to your data. Creator(s) of data own the copyright to that data. The rights holder has the right to determine access and reuse of the data. Because of this, you will need to provide RUcore with a non-exclusive license to manage your data and make it available for others to use. We currently offer two methods of access control. We will work with you on a rights statement explaining what use others may make of your data. This provides important information for end users about reuse of your data. We can also embargo your data for a time period of your choosing. Metadata about your data will appear in the portal, but an embargo note will indicate that the data is not currently available for re-use. This will raise scholarly awareness of your data, so that others will not duplicate your work unnecessarily or will contact you directly for further information about your data and its availability.

Associated Information

The research process is a complex ecosystem of information. Research frequently begins with a grant proposal or other methodology for establishing a hypothesis and a research proposal. Data is collected, and throughout the collection process, documents are created, such as lab or experiment notes, survey questionnaires, images, video, etc. Instruments such as sensors, particulate collectors, telescopes, microscopes are used to collect and analyze data. Maintenance or calibration records for those instruments can be important when collection practices must be justified or explained. Specific software code may be written to process and analyze the data. At the end of the research process, peer reviewed publications share the conclusions and extensively reference the research data. From conception to collection to publication, the entire research process produces valuable data that should be collected and made available for others.

[Help](#)
[Contact Us](#)
[My Account](#)

[RUcore: Rutgers University Community Repository](#)
[Search](#)
[Services](#)
[Collections](#)

[Home](#)
[RUresearch Data Portal](#)
[About](#)

Associated Information

The research process is a complex ecosystem of information. Research frequently begins with a grant proposal or other methodology for establishing a hypothesis and a research proposal. Data is collected, and throughout the collection process, documents are created, such as lab or experiment notes, survey questionnaires, images, video, etc. Instruments such as sensors, particulate collectors, telescopes, microscopes are used to collect and analyze data. Maintenance or calibration records for those instruments can be important when collection practices must be justified or explained. Specific software code may be written to process and analyze the data. At the end of the research process, peer reviewed publications share the conclusions and extensively reference the research data. From conception to collection to publication, the entire research process produces valuable data that should be collected and made available for others.

The Compound Data Object

```

graph TD
    subgraph "Compound Data Object"
        direction TB
        subgraph "Aggregate Level"
            Data((Data))
            PD((Project Documents))
            SW((Software))
            INST((Instrumentation))
        end
        subgraph "Simple Objects with content models"
            ND1((Numeric Data))
            ND2((Numeric Data))
            SURV((Survey))
            INTV((Interview))
        end
        Data -- "hasPart/isPartOf" --> ND1
        Data -- "hasPart/isPartOf" --> ND2
        PD -- "hasPart/isPartOf" --> SURV
        PD -- "hasPart/isPartOf" --> INTV
    end

    ViewObject((View Object))
    ProjectCollection((Project Collection))


    ProjectCollection -- "hasMember" --> ViewObject
    ViewObject -- "isMemberOf" --> ProjectCollection
    ViewObject -- "hasMember/isMemberOf" --> Data
    ViewObject -- "hasMember/isMemberOf" --> PD
    ViewObject -- "hasMember/isMemberOf" --> SW
    ViewObject -- "hasMember/isMemberOf" --> INST

    External[Relationships External to Object  
Dissertation in RUetd;  
Article in RUcore (or elsewhere)]
    ViewObject --> External
  
```

Capturing this associated data is more complex than storing and providing access to the data itself. Capturing the entire life cycle or information ecosystem of a data set is a three step process.

Step One - the Infrastructure

The first step is to build relationships between data objects within the repository architecture. RUresearch leverages the Fedora open source repository architecture by creating a compound object that pulls together the **data**, **associated documents**, such as code books, lab notes, images, **associated software** (such as MATLAB, R, or SAS scripts) intended to support the data, and **instrumentation records**, such as maintenance or calibration records. Resources that are dependent on the data for meaning and do not stand alone are included in the compound object for the project. Associated objects that are separately cataloged and may have different or additional creators, such as analyses, articles, books and presentations, are also related to the compound object. The repository infrastructure provides the groundwork for providing meaningful context



RUTGERS
University Libraries

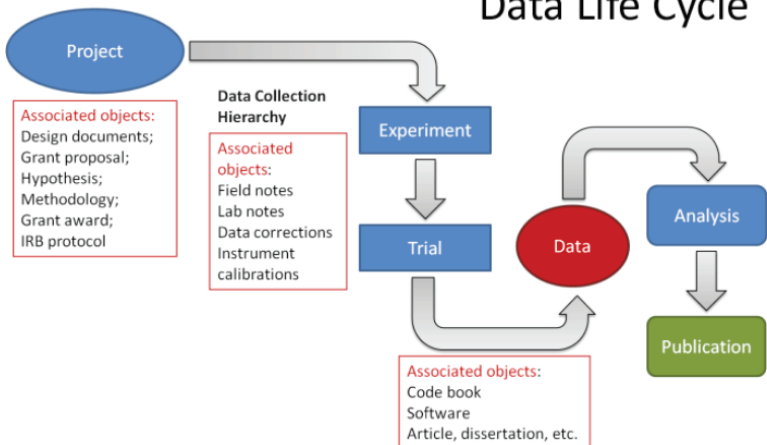
RUcore: Rutgers University Community Repository

Help Contact Us My Account

Search Services Collections

Home
RUresearch Data Portal
About

Data Life Cycle



```

graph TD
    Project([Project]) --> Experiment[Experiment]
    Experiment --> Trial[Trial]
    Trial --> Data((Data))
    Data --> Analysis[Analysis]
    Analysis --> Publication[Publication]
    Data --> Experiment
    Data --> Trial
    
```

Project
Associated objects:
Design documents;
Grant proposal;
Hypothesis;
Methodology;
Grant award;
IRB protocol

Data Collection Hierarchy
Associated objects:
Field notes
Lab notes
Data corrections
Instrument calibrations

Experiment

Trial

Data
Associated objects:
Code book
Software
Article, dissertation, etc.

Analysis

Publication

Step Two - Capturing Associated Objects

Associated objects are captured in digital form and stored in the repository as part of the compound digital object for the project. They are preserved along with the data so all the effort is involved in the front end—capturing and uploading—but ongoing management is very light.

Step Three – Providing Context

There are many different objects that can be captured, at any stage of the research process. Organizing these and making them available in a meaningful way is deceptively complex. Think of your own most recent research project. Can you immediately locate your IRB protocol if a question comes up about making personally identifiable data accessible? Can you find the lab notes created by the student who graduated two years ago took—the ones you kept because his insights were so valuable? Assuming you can find those notes, do you remember which of the three graduate students that year actually took the insightful notes? Do you have documentation of his permission to share those notes, and can you credit him for their creation? Valuable information often gets lost or discarded because it is just too hard to manage it all and remember the context of its creation. Even when you remember the context, you may not want to see it all the time, or to share all context with everyone. It's important to you that you have publicity releases for all the graduate students who appear in your video but sharing those releases with the world at large violates the privacy you are at pains to protect. RUcore's answer lies in its innovative and unique data model and metadata implementation. RUcore uses a metadata implementation that captures information useful for finding information but also information useful for managing information. Rights metadata, one of the types of information collected, is largely kept hidden from the general user. Documents associated with rights, such as publicity releases or IRB protocols, are not available for public display but are available to RUcore administrators and will soon be available to collection owners. The context surrounding any information, such as a research project, is situated in place and time. Separate objects can also have separate access controls for availability to different audiences, allowing for both public-use and restricted use versions of data which may contain sensitive information. RUcore uses metadata "events" to document the "who, what, when, and where" of context about research and its supplementary materials.

[Help](#)
[Contact Us](#)
[My Account](#)

[Home](#)
[RUresearch Data Portal](#)
[About](#)

Rights Event Example

Equine Science Center Videos

Rights Event

Type Permission of License

Date Time 2009-12-09

Associated Entity

Role Contributor

Name Elyse Conway

Associated Object

Type Publicity release

Name Model release

Detail rutgers-lib:27494

Data Life Cycle Event Example

Data Life Cycle Event

Type Related publication

Label Article references the data set, Major League Baseball and Performance Data, 1986

Name Pazzani, Michael J. and Bay, Stephen D. (1999) The Independent Sign Bias: Gaining Insight from Multiple Linear Regression in Proceedings of the Twenty First Annual Conference of the Cognitive Science Society.

Identifier <http://hdl.rutgers.edu/1782.1/rucore30016700001.Manuscript.000056844>

Data is central

The data itself remains central for discovery and use. RUresearch offers flexible portals that can configure metadata displays to show different levels of context. Creators of data may need different information than users. Users in the primary research domain may need different information than users in a broader multidisciplinary context. The libraries will work with researchers to present information about research data in ways that are meaningful and clear.

Data-centric View

```

graph TD
    Project[Project] -- "Is created in" --> Data((Data))
    Data -- "Is analyzed in" --> AnalysisEvent[Analysis event]
    AnalysisEvent -- "Is published in" --> PublicationEvent[Publication event]
    Project --> CreationEvent1[Creation event  
e.g., Experiment]
    CreationEvent1 --> CreationEvent2[Creation event  
e.g., Trial]
  
```

Events

- Situate each lifecycle event in place and time, with associated objects and agents.
- Events can be displayed **or not** in different portals
- Data is disambiguated from its context for more efficient reuse. But the context is always there to be retrieved,
- **No limit to the number and type of events that can be added.**

About Us

How does RUcore work?

Policies

Feedback

Infrastructure

Trusted Repository

Preservation

Understanding Metadata

Open Source

OpenWMS

OpenETD


OpenMIG

Developers

Reference Materials

Web Services/APIs

Edman



VTechData

Q

All ▾

Login

HOME

ABOUT

HELP

CONTACT

Virginia Tech's Data Repository (VTechData) is a platform for openly publishing datasets or other research products created by Virginia Tech faculty, staff, and students.

Content Policy

The purpose of VTechData is to highlight, preserve, and provide access to work produced by the Virginia Tech community and the intellectual output of the university in its land-grant mission. VTechData and Virginia Tech serve the Commonwealth of Virginia, the nation, and the world's community through the discovery and dissemination of new knowledge.

Data and associated materials will be accepted in any language and in a variety of forms and formats. Through the deposit process, depositors are encouraged to provide adequate documentation to ensure usability and accessibility, and should include discipline-specific documentation in a separate file, where appropriate. VTechData staff may be able to provide some of these services.

VTechData Publication and Curation services include:

- minting DOIs for datasets that can be included in articles or other publications
- assistance with file conversion for preservation and long-term reuse
- assistance with organizing and documenting complex file structures and project data
- assistance with generating contextual or disciplinary metadata

If you wish to take advantage of these services, please contact data services at vtchdata@vt.edu.

Access Policy

The general intent of VTechData is to make data and other research products openly available to the general public; however, we recognize that certain research situations necessitate restricting access to content for a certain period of time. Restricted objects can be deposited in VTechData if they can, within a well-defined and reasonably short period of time, be made openly accessible to the general public via de-identification or anonymization processes. For assistance, please contact data services at vtchdata@vt.edu.

Restricted objects will be discoverable but not accessible under the terms of deposit. Do not deposit sensitive or confidential data in VTechData; while we make every effort to maintain the privacy of restricted data, any server can be vulnerable. Sensitive data should not be stored online. If you have questions, please contact us.

Deposit Policy

By depositing data or other research materials into VT's Data Repository, you affirm that:

- the deposit represents your own work or the work of your collaborators; any work that is not your own must be properly cited.

Please read our [Deposit Agreement](#) before depositing your work.

Members of the Virginia Tech community own copyright in their scholarly or educational works as described in [VT Policy 13000](#) and [VT Policy 13015](#). Copyright owners depositing their works in VTechData retain their copyright while granting a non-exclusive license to Virginia Tech's Libraries for access and preservation purposes. All depositors must agree to the non-exclusive distribution license or place their works in the public domain.

Copyright owners may also elect to license their work via the Creative Commons or Open Data License. This option is available as part of the submission process.

Copyright © 2015 Virginia Polytechnic Institute and State University

Special Collections | Art + Architecture | Veterinary Medicine | Northern Virginia Center Resource Center | Mobile Library Website

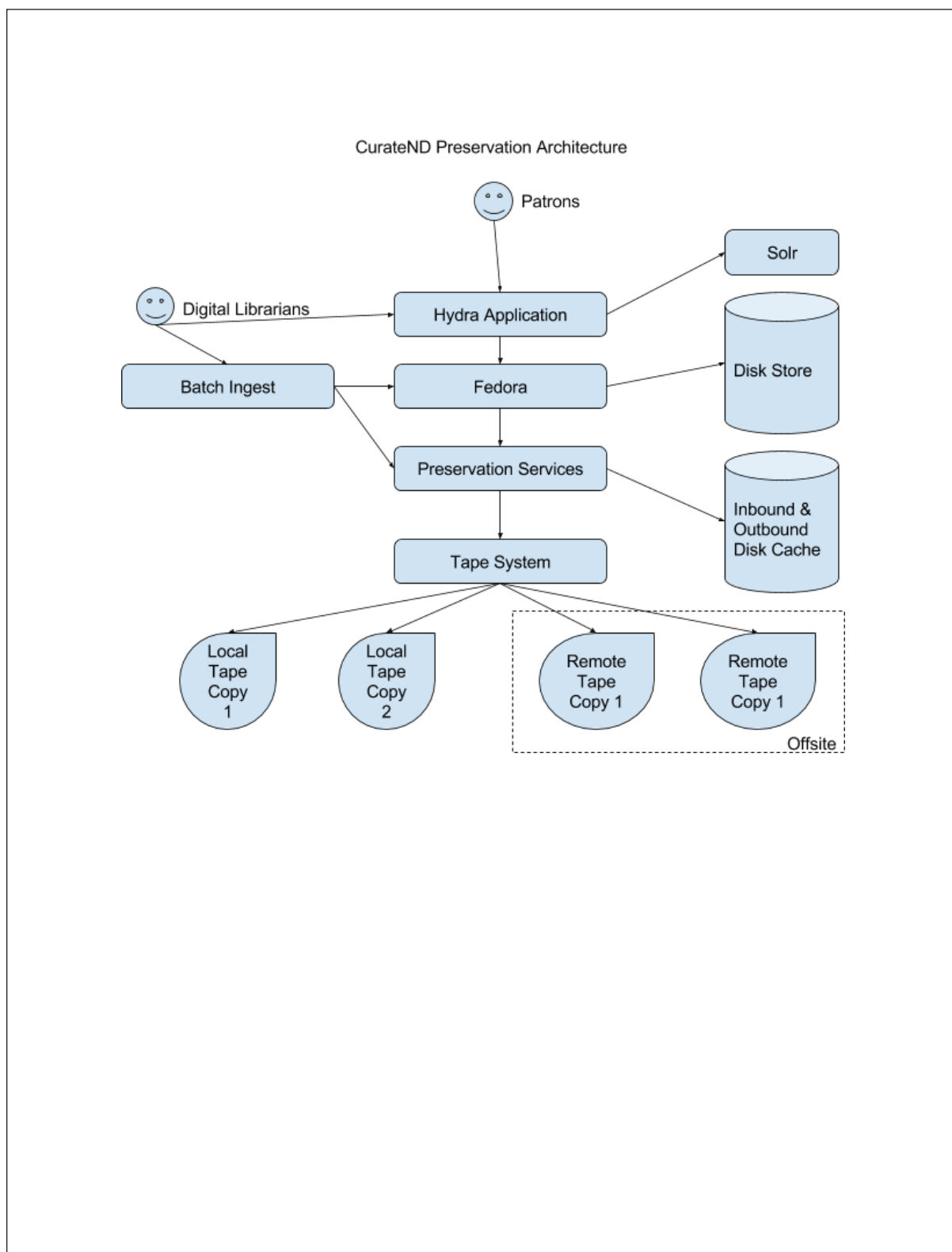
VTechData follows Virginia Tech's Privacy Policy


Data Curation Infrastructure

CurateND Data Curation Infrastructure

CurateND uses a Hydra-based discovery application. It uses Fedora Commons 3.x as the object registry and metadata store and Apache Solr as an index. Using both Fedora and Solr is common for Hydra applications. Self-deposit items go through the Hydra application. There is also a batch ingest ability, which deposits items directly into the preservation store as well as Fedora. Objects in Fedora contain pointers to our preservation store. The preservation store is a custom application that puts content into BagIt bags for storage on tape; maintains a disk cache of content; provides a URL for each preserved file; and runs fixity checks on the content. The data is ultimately all stored on tape, with two copies kept locally and two remotely. The tape appliance handles the replication.

Digital Librarians can deal with the batch ingest directly via a networked filesystem. Content is staged on the filesystem, where it can also be reviewed, assessed, and described. When it is ready, the librarian can start an ingest, which copies the data into the preservation system, the metadata into the preservation system, and a copy of the metadata into Fedora. It then asks the Hydra application to index the new content.





A Service of the Ontario Council of University Libraries

[Scholars Portal Home](#) / [Guides](#) / [Scholars Portal Dataverse Guide](#) / [About Dataverse](#)

Scholars Portal Dataverse Guide

About Dataverse

[Scholars Portal Dataverse](#)
[FAQ](#)
[Related Resources](#)
[Scholars Portal Dataverse Account Terms of Use](#)

Your Dataverses

Your Research Data

Finding and Using Data





Help & Support

Go to Dataverse

Advanced Dataverse Guide

[Dataverse Project Advanced User Guide](#)

Library Contacts

Scholars Portal Dataverse

The Scholars Portal Dataverse is a repository primarily for research data collected by researchers and organizations affiliated with Ontario universities, although anyone in the world is welcome to use Scholars Portal Dataverse to deposit, share, and archive data.

Dataverse is an open-source tool developed by the Institute for Quantitative Social Sciences (IQSS) at Harvard University. Scholars Portal Dataverse is provided as a shared [Ontario Council of University Libraries](#) service.

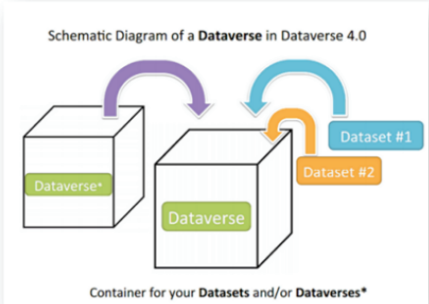
Researchers can use Dataverse to directly deposit data, create metadata, release and share data openly or privately, visualize and explore data, and search for data.

For more information about Scholars Portal Dataverse, please contact dataverse@scholarsportal.info

What is a Dataverse?














A Dataverse is a container for one or more Datasets or Dataverses. Each Ontario University has a Dataverse that contains many Dataverses and Datasets. Researchers can create Dataverses for their own research data and projects, and/or directly deposit Datasets within their Institutional Dataverse.

A Dataverse accepts all kinds of data files: tabular, text, image, etc. All file formats are accepted.



Schematic Diagram of a Dataverse in Dataverse 4.0

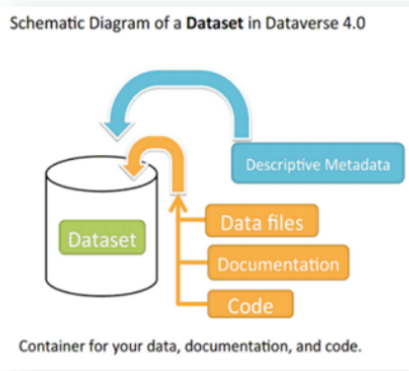
The diagram shows a central box labeled 'Dataverse' with a green label. To its left is a box labeled 'Dataverse*' with a green label. A purple curved arrow points from 'Dataverse*' to 'Dataverse'. To the right of 'Dataverse' are two boxes labeled 'Dataset #1' (blue) and 'Dataset #2' (orange). A blue curved arrow points from 'Dataverse' to 'Dataset #1', and an orange curved arrow points from 'Dataverse' to 'Dataset #2'. Below the diagram is the text: 'Container for your Datasets and/or Dataverses*'

What is a Dataset?

A Dataset is a container for a particular research data set (this can include research data, code, and documentation).

Datasets have an associated metadata record (also referred to as cataloging information or data documentation). This metadata provides contextual information on the dataset. Please see [here](#) for more information on creating metadata for datasets.



Why use Dataverse?

Some key benefits to using Dataverse to manage your research data include:

- **Secure data management.** Dataverse supports the creation of terms of use and restrictions if you want to limit the use of or access to data. It also provides a backup copy for safekeeping.
- **Effective sharing.** Dataverse is a convenient way to disseminate your data, and can facilitate your research team's collaboration within a secure space.
- **Track changes.** Dataverse provides increased control over managing changes to a project without overwriting any part of that project, an especially useful feature when working on a team.
- **Long-term access and preservation.** Persistent identification to your data ensures reliable protection and prevention from data obsolescence.
- **Organization and compatibility.** Create your own personal web data archive that conforms to metadata standards to maximize system compatibility and searchability.
- **Save time.** Dataverse has an easy to use interface for uploading and searching through your data.
- **Increase research visibility.** Increase scholarly recognition for your work beyond your research publications.
- **Meet grant requirements.** Many funding agencies now *require* that researchers deposit data which collected as part of their research project into an archive.

References


Crosas M. *The Dataverse Network: An Open-source Application for Sharing, Discovering and Preserving Data*. D-Lib Magazine. 2011;Volume 17(1/2).

King, Gary. 2007. An Introduction to the Dataverse Network as an Infrastructure for Data Sharing. *Sociological Methods and Research* 36: 173–199. Available at <http://j.mp/iHJcAa>

TEXAS DIGITAL LIBRARY

Texas Data Repository | About | How Dataverse Works

<http://data.tdl.org/about/>

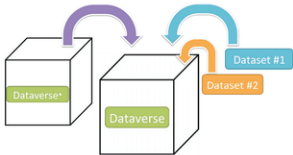


[About](#) [Documentation](#) [FAQs](#) [Log In](#) [Help](#)

How Dataverse Works

A Dataverse is a container for datasets (research data, code, documentation, and metadata) and other Dataverses, which can be setup for individual researchers, departments, journals, and organizations.

Schematic Diagram of a **Dataverse** in Dataverse 4.0



Container for your **Datasets** and/or **Dataverses***

* Dataverses can now contain other Dataverses (this replaces Collections & Subnetworks)

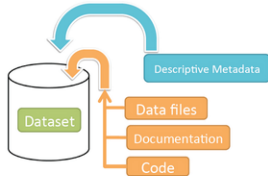
Each Dataverse contains datasets. Each dataset includes:

- Data files
- Metadata that describes the data files

Each dataset may also include:

- Code associated with the data files
- Additional documentation describing the data files and project from which they derive

Schematic Diagram of a **Dataset** in Dataverse 4.0




Container for your data, documentation, and code.

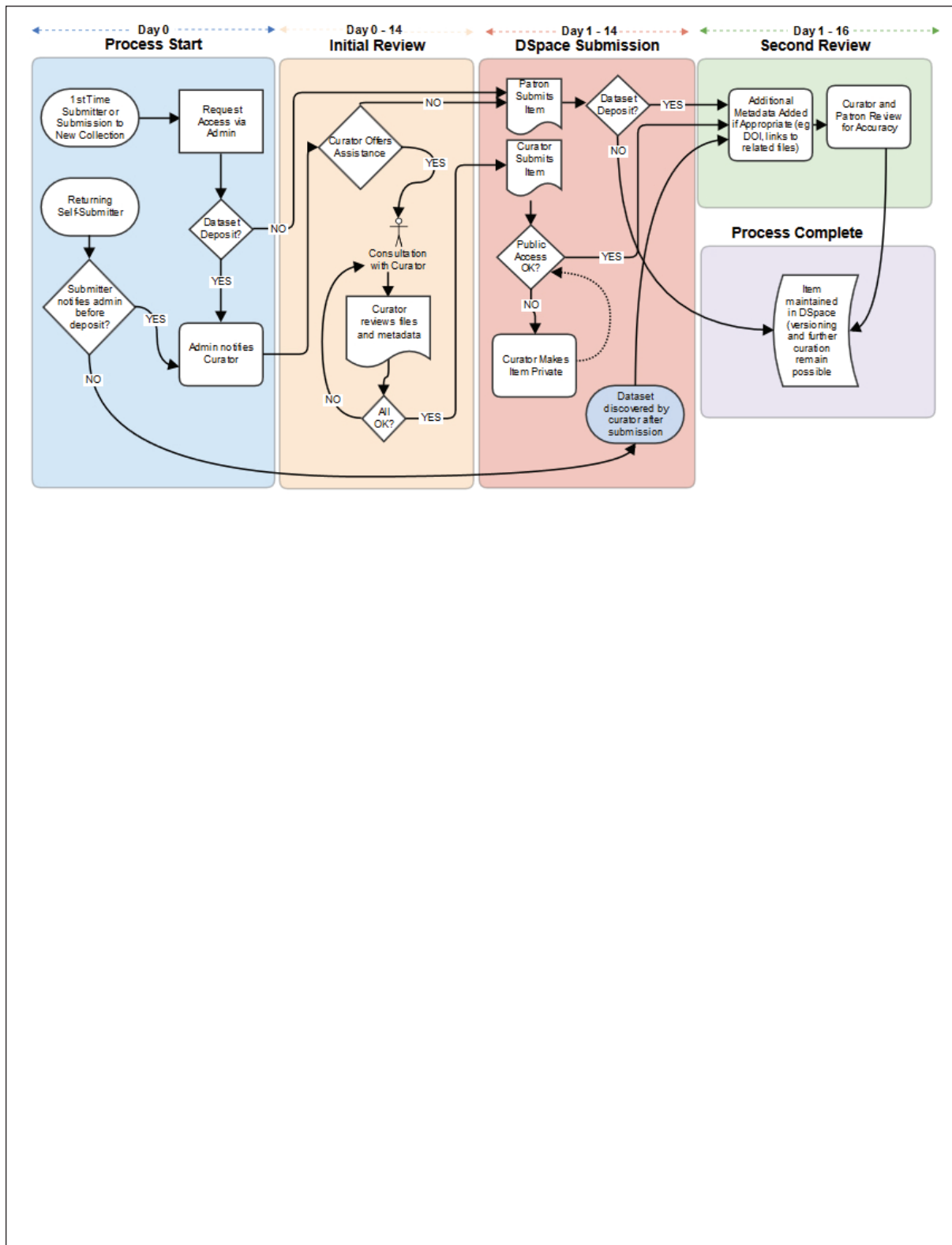
Help

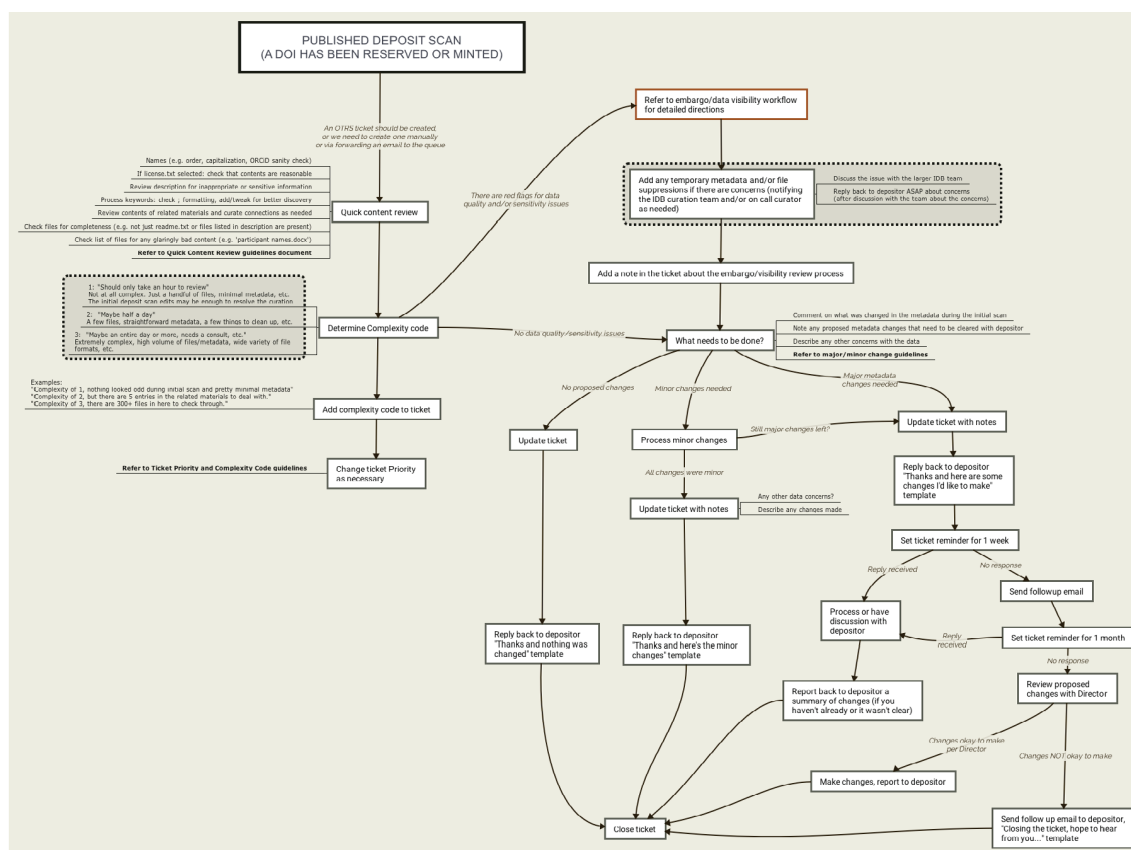
- [User Guide](#)
- [Helpdesk and Library Support](#)

Hosted by



Data Curation Workflows

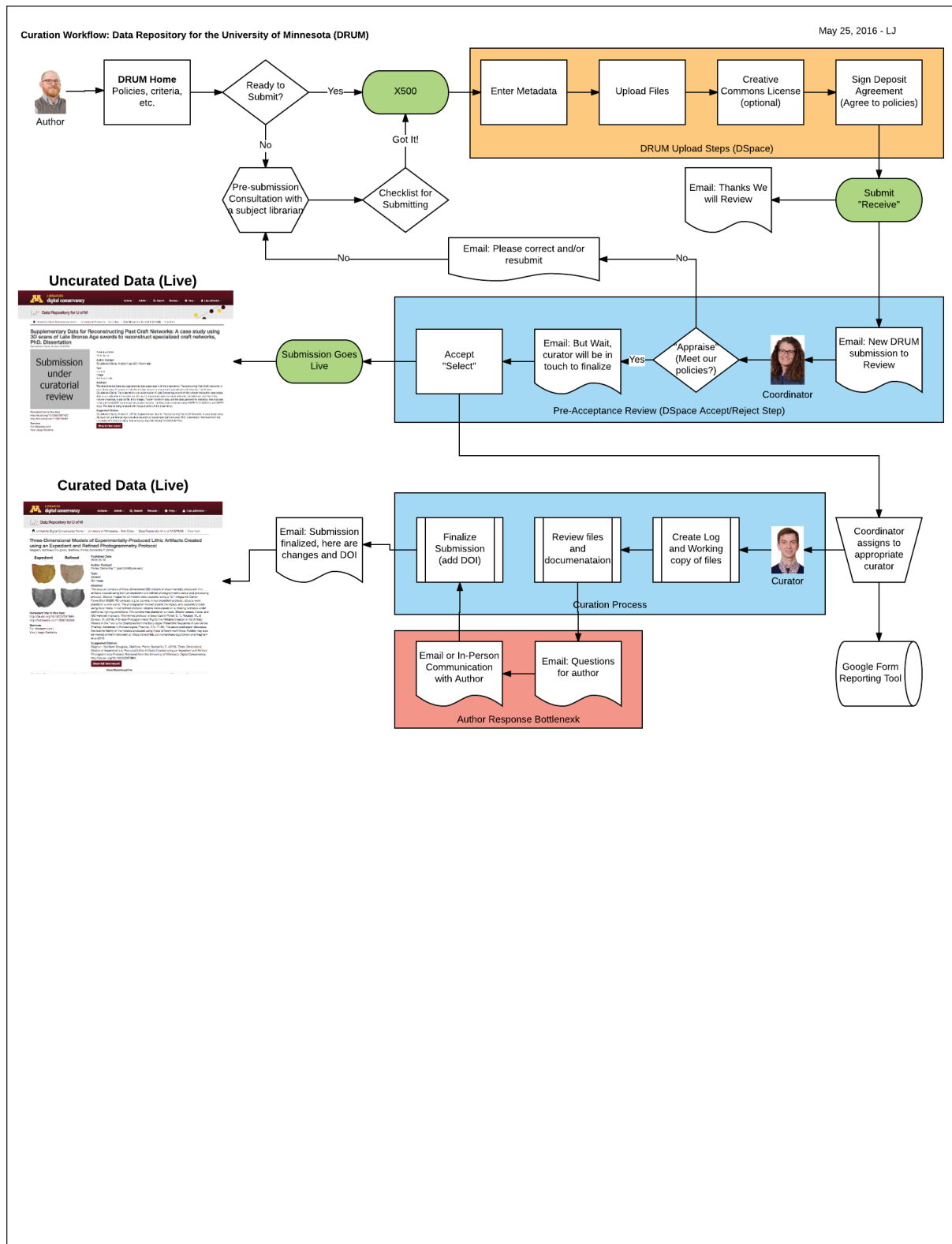


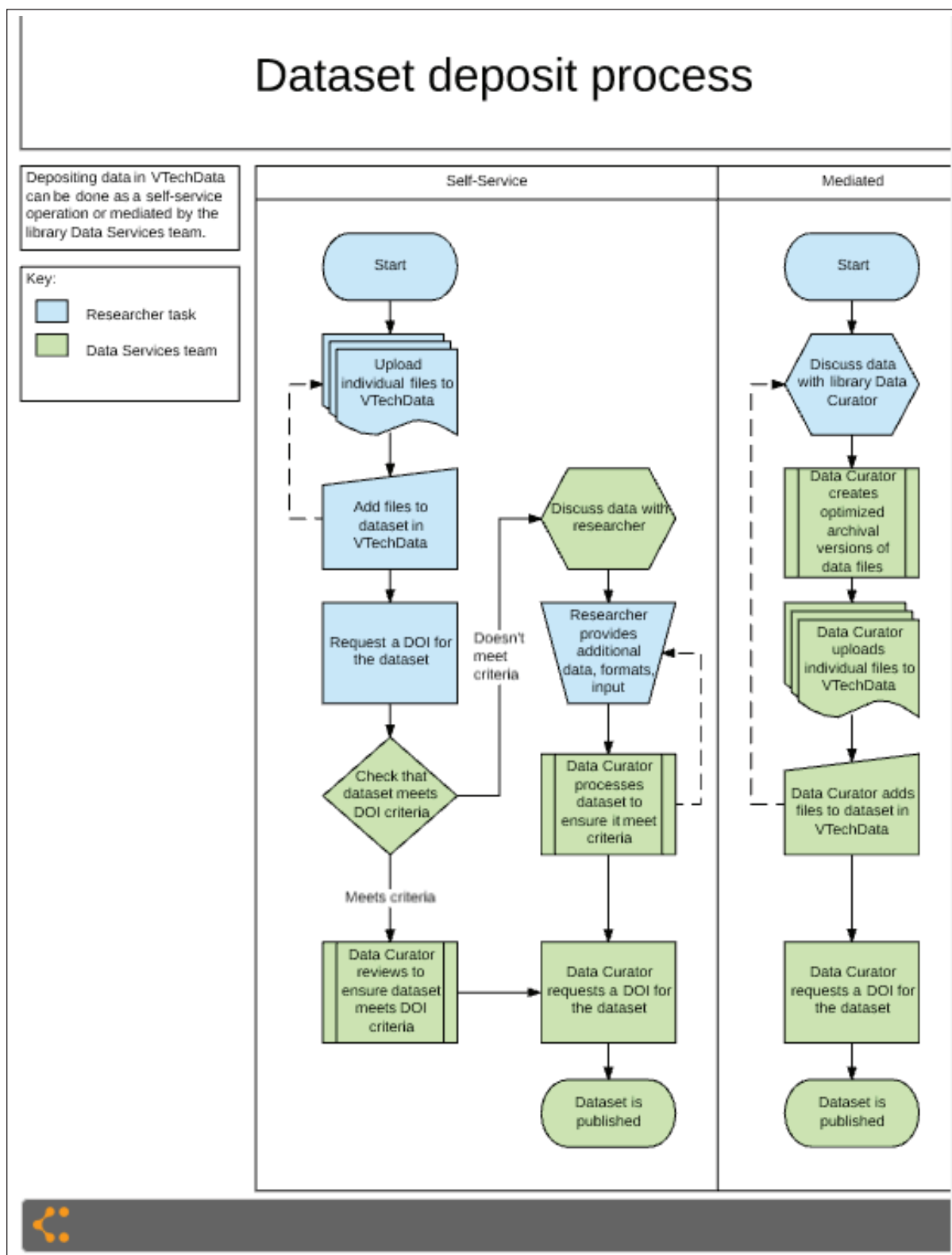


UNIVERSITY OF MINNESOTA LIBRARIES

Curation Workflow (DRUM)

<https://drive.google.com/file/d/0B5Dm3XFQloc4d0Q0Nk5ISW92TGs/view?usp=sharing>

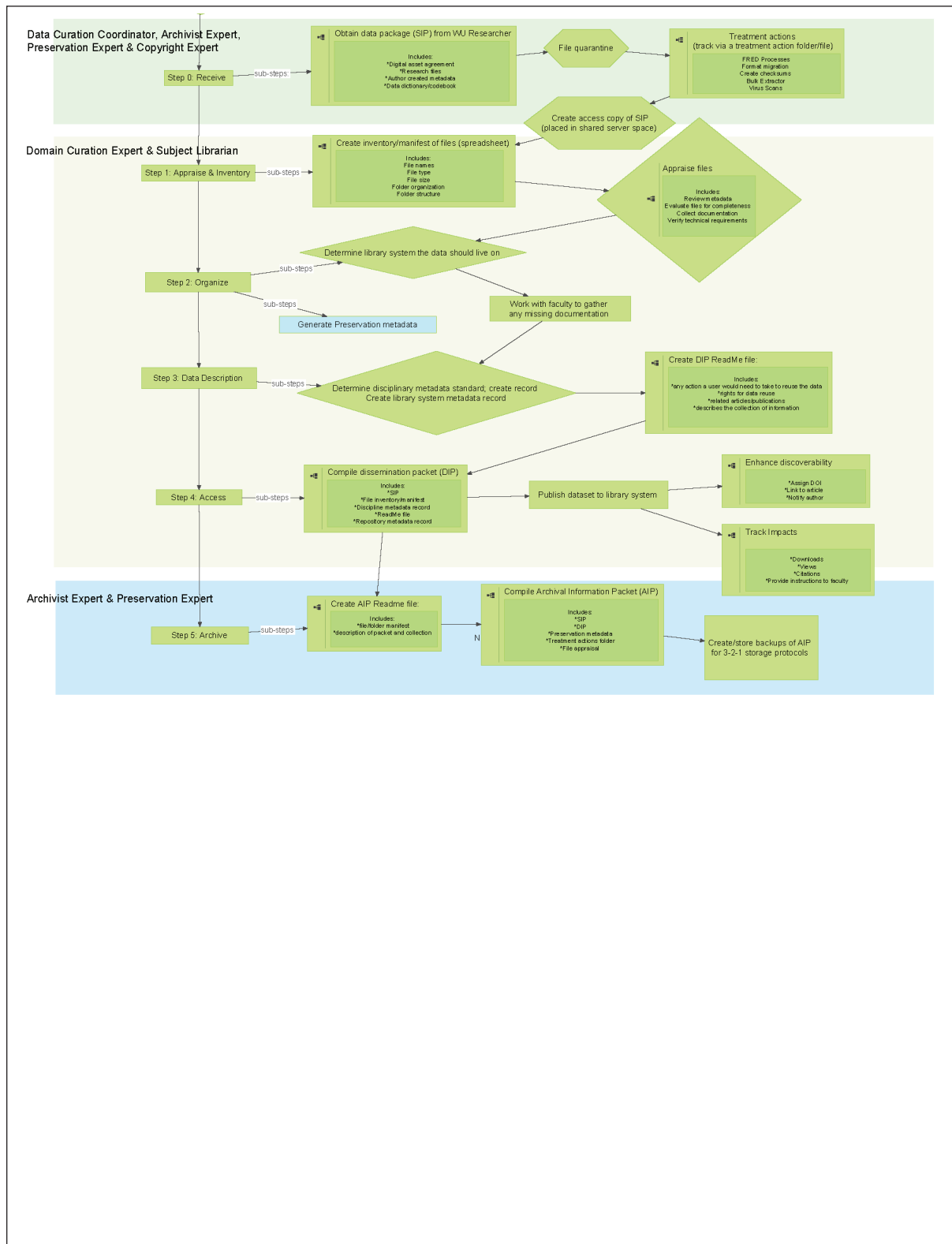




WASHINGTON UNIVERSITY IN ST. LOUIS LIBRARIES

Digital Research Materials Repository Curation Workflow

<https://drive.google.com/file/d/0B5Dm3XFQloc4UTRtZHnQ09QNnc/view?usp=sharing>



Data Models and Metadata Schemas

CurateND Data Model

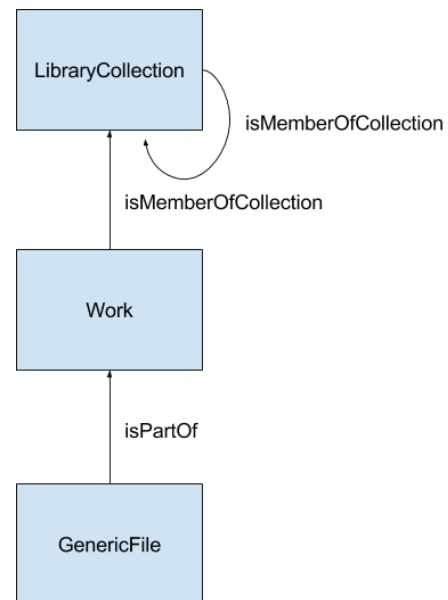
Structural Relationships

CurateND uses an early version of the PCDM for structural relationships and a Dublin Core with extensions for the descriptive metadata. Objects have one of three types: LibraryCollection, Work, or Generic File. In practice, while there is a single type of *LibraryCollection* and *GenericFile*, there are many types of *Works*.

All the predicates are in the Fedora Commons 3 external relation namespace, i.e. `info:fedora/fedora-system:def/relations-external#`.

Descriptive Metadata

The descriptive metadata is based on Dublin Core, but has freely added extensions when needed.



UNIVERSITY OF NOTRE DAME, HESBURGH LIBRARIES
CurateND Metadata Model

Predicate	Display label	Content Type	Input description	Cardinality (Y=many, N=one)
http://purl.org/dc/terms/alternative	Alternative Title	String, title alternative form.	Already in input page but not displaying?	many
http://purl.org/dc/terms/contributor(unqualified)	Contributor	String, generally personal name. e.g. "Butler, Octavia"	This is also in the input page, but not displaying for books.	many
http://purl.org/dc/terms/contributor#artist	Contributing Artist	String, generally personal name. e.g. "Butler, Octavia"	An entity responsible for creating artistic works within the resource, other than illustrations.	many
http://purl.org/dc/terms/contributor#author	Coauthor	String, generally personal name. e.g. "Butler, Octavia"	An authorial entity who contributed to the resource.	many
http://purl.org/dc/terms/contributor#editor	Contributing Editor	String, generally personal name. e.g. "Butler, Octavia"	An entity responsible for editing the resource.	many
http://purl.org/dc/terms/contributor#illustrator	Contributing Illustrator	String, generally personal name. e.g. "Butler, Octavia"	An entity responsible for illustrating the resource.	many
http://purl.org/dc/terms/contributor#photographer	Contributing Photographer	String, generally personal name. e.g. "Butler, Octavia"	An entity responsible for creating photographic works within the resource.	many
http://purl.org/dc/terms/creator	Inventor	String, generally personal name. e.g. "Butler, Octavia"	An entity listed on the patent as a creator.	Y
http://purl.org/dc/terms/creator (unqualified)	Creator	String, generally personal name. e.g. "Butler, Octavia"	An entity responsible for the resource's creation.	many

UNIVERSITY OF NOTRE DAME, HESBURGH LIBRARIES
CurateND Metadata Model

http://purl.org/dc/terms/creator#administrative_unit	Department	String	Relevant academic departments	Y
http://purl.org/dc/terms/creator#artist	Artist	String, generally personal name. e.g. "Butler, Octavia"	An entity responsible for art works in a resource which consists primarily of art works (e.g. an art book).	many
http://purl.org/dc/terms/creator#author	Author	String, generally personal name. e.g. "Butler, Octavia"	An entity responsible for significant authorial work within the resource.	many
http://purl.org/dc/terms/creator#editor	Editor	String, generally personal name. e.g. "Butler, Octavia"	An entity responsible for significant editorial work in creating the resource.	many
http://purl.org/dc/terms/creator#illustrator	Illustrator	String, generally personal name. e.g. "Butler, Octavia"	An entity responsible for illustrations of a resource which consists primarily of illustrations (e.g. a children's picture book).	many
http://purl.org/dc/terms/creator#local	n/a	String, generally personal name. e.g. "Butler, Octavia"	Creators who are (or were) associated with the local institution. People are to be listed here in addition to being listed in dc:creator.	Y
http://purl.org/dc/terms/creator#photographer	Photographer	String, generally personal name. e.g. "Butler, Octavia"	An entity responsible for photography in a resource which consists primarily of photographs, (e.g. a collection of a photographer's work).	many
http://purl.org/dc/terms/date#application	Application Date	String. (date?) Has form "YYYY-MM-DD"	The date of the initial submission of the application for this patent.	N
http://purl.org/dc/terms/date#prior_publication	Prior Publication Date	String. (date?) Has form "YYYY-MM-DD"	Date of prior publication (?)	N
http://purl.org/dc/terms/date#copyrighted	Copyright Date	Should be in form YYYY. More granular dates unlikely	The resource's copyright date	one
http://purl.org/dc/terms/date#submitted	Date Added	Date	Date object was created in CurateND	N

UNIVERSITY OF NOTRE DAME, HESBURGH LIBRARIES
CurateND Metadata Model

http://purl.org/dc/terms/description	Description	String	Description of patent, may contain abstract.	Y
http://purl.org/dc/terms/description#table_of_contents	Table of Contents	String. May be chapter titles separated by a space, two hyphens, and a space, e.g. " -- ". Does not need to be parsed specially, can simply be displayed as a string.	A listing of the chapters or sections of a resource as taken from the resource's contents listing.	one
http://purl.org/dc/terms/extent	Extent	String. Will probably be page length, e.g. "368 pages"	The number of pages in the resource, the resource's size, or the resource's duration	many
http://purl.org/dc/terms/extent#claims	Claims	String.	The number of claims in this patent. Usually an integer, but has type string to handle any possible special cases.	N
http://purl.org/dc/terms/identifier#isbn	ISBN	Structured alphanumeric string (regex test: {0-9x}), either 10 or 13 characters. They often contain hyphens, but can be normalized to not.	The resource's ISBN.	many
http://purl.org/dc/terms/identifier#local	Local Identifier	String. For most digitized books, will be the call number, but may be another kind of local identifier used to shelve or handle books.	The resource's local identifier, e.g. call number.	many
http://purl.org/dc/terms/identifier#other_application	Other Application	String.	A prior submittal of this patent for review. (?)	Y
http://purl.org/dc/terms/identifier#patent	Patent Number	String. May contain spaces. E.g. "US 1234567890123 B2"	The patent number for this resource. Probably refers to the USPTO but not restricted to US patents.	N

UNIVERSITY OF NOTRE DAME, HESBURGH LIBRARIES
CurateND Metadata Model

http://purl.org/dc/terms/identifier#prior_publication	Prior Publication Number	String.	Identifier for the prior publication of this patent.	Y
http://purl.org/dc/terms/isPartOf	Published in	String (unfortunately)	The title of the journal, book, or other work in which the Article was published	N
http://purl.org/dc/terms/issued	Publication Date	Should be in form YYYY. More granular dates unlikely	The resource's publication date	one
http://purl.org/dc/terms/issued	Publication Date	Can we test as YYYY or YYYY-MM or YYYY-MM-DD	The article's publication date as year, year-month, or year-month-day, eg. 2015 or 2015-05 or 2015-05-31	N
http://purl.org/dc/terms/issued	Date Issued	String. (See date discussion above)	Date the patent was issued.	N
http://purl.org/dc/terms/isVersionOf#edition	Edition	String. Will probably be a number and additional text: 2, 2nd edition, etc.	The resource's edition.	one
http://purl.org/dc/terms/language	Language	String		Y
http://purl.org/dc/terms/modified	Date Modified	Date	Date object was last modified in CurateND	N
http://purl.org/dc/terms/publisher	Publisher	String		Y
http://purl.org/dc/terms/rights	Rights	String	Intellectual Rights	Y
http://purl.org/dc/terms/rightsHolder	Assignee	String	Assignee of the patent.	Y
http://purl.org/dc/terms/source	USPTO Link	URL.	Link to the patent at the USPTO website (or other patent office websites).	N
http://purl.org/dc/terms/subject#cpc	Classification (CPC)	String.	Cooperative Patent Classification codes.	Y

UNIVERSITY OF NOTRE DAME, HESBURGH LIBRARIES
CurateND Metadata Model

http://purl.org/dc/terms/subject#ipc	Classification (IPC)	String.	International Patent Classification codes.	Y
http://purl.org/dc/terms/subject#lcsb	Subject (Library of Congress)	String. Taken from the book's bib record. E.g. "Amnesia -- Fiction."	A topic of the content of the resource as taken from the Library of Congress Subject Headings.	many
http://purl.org/dc/terms/subject#uspc	Classification (US Patent)	String.	US Patent Classification codes.	Y
http://purl.org/dc/terms/title	Title	String	Title of the Patent	N
http://purl.org/dc/terms/type	n/a	String	The subtype of Work this item is.	N
http://purl.org/ontology/bibo/eissn	e-ISSN	e-ISSN validation?	The e-ISSN of the publication in which the article appears	N
http://purl.org/ontology/bibo/isbn	ISBN	ISBN validation from Book Type?	If the article is published in a volume with an ISBN, the volume's ISBN	N
http://purl.org/ontology/bibo/issn	ISSN	ISSN validation?	The ISSN of the publication in which the article appears	N
http://purl.org/ontology/bibo/issue	Issue	String	The number(s) or name of the issue in which the article appears.	N
http://purl.org/ontology/bibo/numPages	Number of pages	Integer	The total number of pages as an integer	N
http://purl.org/ontology/bibo/pageEnd	Last page	String	The number or other identifier of the article's final page.	N
http://purl.org/ontology/bibo/pageStart	First page	String	The number or other identifier of the first page on which the article appears, e.g. "42" or "E594"	N
http://purl.org/ontology/bibo/volume	Volume	String	The number or name of the volume in which the article appears.	N



QUICK START GUIDE: SIMPLE CREATE

Using EZID's UI, you can quickly and easily create ARKs and DOIs. If you do not know any of the values for the properties outlined below, see the Quick Start Guide "What to do if required information is unavailable."

FOR ARKS

Property	Description	Examples
Object location URL	The current location (URL) of the identified object.	http://merritt.cdlib.org/m/ark%3A%2F13030%2Fqt5np807ch http://opencontext.org/subjects/199ED3F0-8CA2-4BBD-FA14-468133255587 http://www.coredu.fr/repository/OAIHandler?verb=GetRecord&metadataPrefix=lom&identifier=oai:editors.coredu.fr:31779
Who	<p>The name of an entity (person, organization, or service) responsible for creating the content or making it available, e.g. author, creator.</p> <p>Put name parts in "sort-friendly" order. Separate multiple names with ",". Append one or more final commas (",") to indicate that one or more internal commas can be used as inversion points to recover natural word order (if different from sort-friendly word order).</p>	<p>Kim, JH,; Cho, J,; Keane, TD,</p> <p>Virginia Department of Historic Resources (VA-DHR); Open Context Editors</p> <p>Canal Educatif à la Demande</p>
What	A name or other human-oriented identifier given to the resource, e.g. a title.	<p>Political fragmentation and land use changes in the Interior Plains</p> <p>Virginia Site Files: 44WR0079 (Site)</p> <p>Vidéos Sciences & Innovation de Canal Éducatif à la Demande</p>
When	A point or period of time (date range) important in the lifecycle of the resource, often when it was created, modified, or made available. Use ";" to separate entries and "~" to indicate approximation.	<p>10/4/2015</p> <p>2014-07-31T00:00:00-07:00</p> <p>1/1/2007</p>



FOR DOIs

Property	Description	Examples
Object location URL	The current location (URL) of the identified object.	https://lilliput.figshare.com/articles/Impact_of_Task_Performance_Fraud_Risk_Assessment_on_Forensic_Skills_and_Mindsets_Experience_from_Nigeria/2002749 http://doi.virtualbrain.org/lp/10.5072/FK2028TW8Z http://mdsoarstage.lib.umd.edu/handle/11603-STAGE/4859
Creator	The main researchers involved in producing the data, or the authors of the publication in priority order. May be a corporate, institutional, or personal name. In personal names, list family name before given name.	George, Christopher Worth, A [MGH] Owens, Allessia P.
Title	A name or title by which the data or publication is known.	Impact of Task Performance Fraud Risk Assessment on Forensic Skills and Mindsets: Experience from Nigeria Internet Brain Segmentation Repository Mentoring African American males
Publisher	A holder of the data (e.g., an archive) or the institution which submitted the work. In the case of datasets, the publisher is the entity primarily responsible for making the data available to the research community.	Figshare MGH CMA Maryland Shared Open Access Repository
Publication year	The year when the data was or will be made publicly available. If an embargo period is in effect, use the year when the embargo period ends.	2015 2015 2008
Resource type	The general type of the data.	Dataset Dataset Text



QUICK START GUIDE: ADVANCED CREATE FOR DOIS

Using EZID's UI to create a DOI, you must provide DataCite metadata.

Mandatory DataCite properties are indicated with an asterisk (*).

Property	Description	Notes
Creator* (repeats)	The main researchers involved in producing the data, or the authors of the publication, in priority order. Mandatory	Personal, corporate, or institutional name(s)
Title* (repeats)	A name or title by which a resource is known. Mandatory	Free text
Publisher*	The name of the entity that holds, archives, publishes, prints, distributes, releases, issues, or produces the resource. Mandatory	Free text
PublicationYear*	The year when the data was or will be made publicly available. Mandatory	YYYY
ResourceType	A description of the resource. Uses a controlled vocabulary. Recommended, but <i>will become mandatory in next version</i> .	See Quick Start Guide for controlled list
Subject (repeats)	Subject, keyword, classification code, or key phrase describing the resource. Recommended	Free text
Contributor (repeats)	The institution or person responsible for collecting, managing, distributing, or otherwise contributing to the development of the resource. Recommended	See Quick Start Guide for controlled list. Works with ORCIDs.
Date (repeats)	Different dates relevant to the work. Recommended	Uses W3CDTF formats
Language	The primary language of the resource. Optional	Allowed values are taken from IETF BCP 47, ISO 639-1 language codes
AlternateIdentifier (repeats)	An identifier or identifiers other than the primary Identifier applied to the resource being registered. Optional	Free text
RelatedIdentifier (repeats)	Identifiers of related resources. (Must be globally unique.) Recommended	See Quick Start Guide for controlled list
Size (repeats)	Unstructured size information about the resource. Optional	Free text
Format (repeats)	Technical format of the resource. Optional	Free text
Version	The version number of the resource. Suggested practice: track major_version.minor_version. Optional	Free text
Rights (repeats)	Any rights information for this resource. Optional	Free text
Description (repeats)	All additional information that does not fit in any of the other categories. May be used for technical information. Recommended	Abstract strongly suggested
GeoLocation (with point and box sub-properties)	Spatial region or named place where the data was gathered or about which the data is focused. Recommended	Can use WGS 84 (World Geodetic System) coordinates or free text


For details about field constraints and all sub-properties, see <http://schema.datacite.org>

Data Deaccessioning Policies

DATVERSE PROJECT

Deaccessioning Your Dataset [not recommended]

<https://dataverse.scholarsportal.info/guides/en/4.5/user/dataset-management.html#deaccession-your-dataset-not-recommended>

Dataverse Project

AboutCommunityBest PracticesSoftwareContact

Search

User Guide

- Account Creation & Management
- Finding and Using Data
- Dataverse Management
- Dataset + File Management
- Tabular Data File Ingest
- Data Exploration Guide
- Appendix
- Installation Guide
- API Guide
- Developer Guide
- Admin Guide


Deaccession Your Dataset [not recommended]

Deaccessioning a dataset or a version of a dataset is a very serious action that should only occur if there is a legal or valid reason for the dataset to no longer be accessible to the public. If you absolutely must deaccession, you can deaccession a version of a dataset or an entire dataset. To deaccession, go to a dataset you've already published (or add a new one and publish it), click on Edit Dataset, then Deaccession Dataset. If you have multiple versions of a dataset, you can select here which versions you want to deaccession or choose to deaccession the entire dataset. You must also include a reason as to why this dataset was deaccessioned from a dropdown list of options. There is also a free-text box to add more details as to why this was deaccessioned. If the dataset has moved to a different repository or site you are encouraged to include a URL (preferably persistent) for users to continue to be able to access this dataset in the future.

If you deaccession the most recently published version of the dataset but not all versions of the dataset, you are able to go in and create a new draft for the dataset. For example, you have a version 1 and version 2 of a dataset, both published, and deaccession version 2. You are then able to edit version 1 of the dataset and a new draft will be created.

Important Note: A tombstone landing page with the basic citation metadata will always be accessible to the public if they use the persistent URL (Handle or DOI) provided in the citation for that dataset. Users will not be able to see any of the files or additional metadata that were previously available prior to deaccession.

Data Science at The Institute for Quantitative Social Science | Code available at

Created using Sphinx 1.4.6
GitHub

Back to top

Version 4.5 | Last updated on Sep 28, 2016

© Copyright 2016, The President & Fellows of Harvard College

Illinois Data Bank

+ Deposit Dataset

Find Data

Policies

Help

Log in with NetID

Policy Framework and Definitions

Access and Use Policy

Accession Policy

Deposit Agreement

Preservation Policy

Preservation Review Procedures

Withdrawal Guidelines

Preservation Review Guidelines

Preservation Review, Retention, Deaccession, Revision, and Withdrawal Procedure

Purpose of this Procedure

This document outlines the procedures for reviewing, revising, retaining, Deaccessioning, and Withdrawing Data Files, Metadata Files, and Descriptive Metadata published in the Illinois Data Bank.

Preservation Review

The long-term viability of Datasets published in the Illinois Data Bank will be assessed using a robust set of review criteria. The Illinois Data Bank is committed to transparency, accountability, and collaborative decision-making regarding assessments of the long-term preservation status of research data. While a variety of unique factors influence decisions made about the Disposition of Datasets, the criteria outlined in the [Preservation Review Guidelines](#) provide a basis for assessing Datasets.

Preservation Review Roles and Responsibilities

Assessment decisions are a shared responsibility and are often influenced by discipline-specific factors. The Research Data Service staff are responsible for developing and leading the assessment process of Datasets and will consult with ad hoc "Assessment Teams" comprising functional and subject specialists as well as domain experts outside of the Library as appropriate. The "Assessment Team" may also incorporate input from other stakeholders as necessary.

Retention

The Illinois Data Bank anticipates that the majority of Preservation Reviews will result in Dataset retention. The decision to retain a Dataset will typically indicate that the preservation viability of the Dataset is acceptable given the determined long-term value of the Dataset, and that Illinois Data Bank resources being deployed to steward the Dataset are at a level that is proportional to its long-term value.

The Illinois Data Bank will commit resources to escalating preservation efforts for Datasets determined to have remarkable value that are suffering preservation risk or are not available in the most usable states. Examples of escalated preservation procedures include file format migration, enhancing Descriptive Metadata/Metadata Files, or improving access and/or use services by developing data-type-specific viewers/emulators.

Deaccession

A decision to Deaccession the Data Files and/or Metadata Files associated with a Dataset will only occur if it is determined that the Dataset is not of long-term value to its research community and/or its inclusion in the Illinois Data Bank detrimentally affects the Illinois Data Bank's ability to steward effectively other resources whose research value and preservation viability are evident.

Upon deciding to Deaccession the Data Files and/or Metadata Files associated with a Dataset, the Illinois Data Bank will consider one of these options:

- Transfer to a repository more appropriately situated to steward the Data Files and/or Metadata Files.
- Transfer Data Files and/or Metadata Files back to the Long-Term Contact Person indicated in the Descriptive Metadata.

For any Data Files and/or Metadata Files that are to be Deaccessioned, a good faith effort to contact the Long-Term Contact Person will be made by notifying them at the email address the Illinois Data Bank has on record. The notification will outline the Illinois Data Bank's Deaccessioning decision.

If the Illinois Data Bank does not receive a response from the Long-Term Contact Person after 90 days, the Illinois Data Bank will transfer or discard the Data Files and/or Metadata Files according to the practices and security standards in place at the time of Deaccessioning.

The Illinois Data Bank will not Deaccession any Data Files and/or Metadata Files before the initial commitment period ends, currently five years.

The Illinois Data Bank currently plans to retain Descriptive Metadata persistently for all Datasets deposited in the Illinois Data Bank regardless of the Disposition of Data File(s) and/or Metadata File(s) except in rare circumstances as determined by the Director of the Research Data Service.

Illinois Data Bank
+ Deposit Dataset
Find Data
Policies
Help
Log in with NetID

Policy Framework and Definitions
Access and Use Policy
Accession Policy
Deposit Agreement
Preservation Policy
Preservation Review Procedures
Withdrawal Guidelines
Preservation Review Guidelines

Revisions

Depositors are expected to confirm the validity of all content prior to publishing a Dataset. However, should an error in the Descriptive Metadata be discovered, the original Depositor or Research Data Service staff may make a revision to the Descriptive Metadata, which is tracked via a public change log.

If a Creator of a Dataset finds that a file in their published Dataset contains an error, they must contact the Research Data Service staff to submit the corrected file. A new version of the entire Dataset will be created and a new DOI will be assigned. Research Data Service staff will see that the Descriptive Metadata associated with the Dataset makes apparent which version is most recent and what changes occurred. Research Data Service staff will refer to the [Illinois Data Bank Withdrawal Guidelines](#) when determining whether to remove the erroneous Dataset from public view. Depositors are expected to limit the need for versioning by not publishing erroneous Datasets; as such, Datasets are limited to no more than six versions.

Withdrawal of Deposited Datasets

The Illinois Data Bank may Withdraw a published Dataset from the repository before the current five year commitment period ends for a compelling reason. Compelling reasons include, but are not limited to, failure to meet the Criteria for Depositing outlined in the [Illinois Data Bank Accession Policy](#), detection of malware in deposited files, violations of copyright or publisher policy, violations of contracts (e.g., Nondisclosure Agreement, Material Transfer Agreement, etc.), research misconduct (e.g., plagiarism, fabrication or falsification of data, etc.), legal requirements, national security, or situations that violate the University Code of Conduct.

Datasets may not be Withdrawn because the Depositor or Creator is moving to another institution. Creators have the right to provide additional copies to other institutions under the non-exclusive Deposit Agreement.

All Withdrawal requests must be submitted to databank@library.illinois.edu. These will be reviewed by Research Data Service staff who may contact the requestor for more information. If the request is submitted by a third party or the decision to Withdraw is made by the Research Data Service staff, the Long-Term Contact Person and the Depositor will be notified of the request via the email addresses the Illinois Data Bank has on record. Research Data Service staff are not responsible for resolving legal disputes, but will refer University of Illinois community members to the University of Illinois Office of University Counsel at <http://www.legal.uillinois.edu/>.

At minimum, Data Files and/or Metadata Files associated with a Withdrawn Dataset are removed from the public view and are no longer available for download. Research Data Service staff will add a statement of Withdrawal to the associated Dataset's Descriptive Metadata. In many cases, Withdrawal results in suppression of public access to Data Files and/or Metadata Files, even when the entire Dataset will be retained within our systems for the sake of provenance. In rarer cases, the Research Data Service staff may be compelled to delete all or part of a Dataset altogether. The Research Data Service staff will refer to the [Illinois Data Bank Withdrawal Guidelines](#) to respond to the varying circumstances under which Withdrawal may occur.

Withdrawal of Dataset Drafts

In order to ensure sustainability of technological and storage resources, the Research Data Service staff retain the right to delete initiated Datasets that have remained in draft state in the Illinois Data Bank in excess of 12 months. A good faith effort to contact the Depositor via the email address on record in the Illinois Data Bank will be made prior to deletion.

Acknowledgements

- Carnegie Mellon University Research Showcase. Research Showcase @ CMU Policies, Revision and Withdrawal Policy. <http://repository.cmu.edu/policies.html>
- Illinois Digital Environment for Access to Learning and Scholarship (IDEALS). IDEALS Withdrawal Policy. <http://hdl.handle.net/2142/3744>

Contact Research Data Service staff with questions or to request an addition or revision to this policy.

SPEC Kit 354: Data Curation

119

3. Collections & Content

Defining Research Data

For the purposes of Deep Blue Data, research data are defined as representations of observations, objects, or other entities used as evidence of phenomena for the purposes of research or scholarship. In practical terms, Deep Blue Data will accept data that were developed or used in the support of research activities of U-M faculty, students and staff.

Data Formats

As the intent of the Deep Blue Data data repository is to make data as openly available as possible for discovery, understanding, and reuse, we strongly encourage the submission of data in formats that are open and nonproprietary.

If data cannot be converted to nonproprietary formats, we then encourage data submission in formats that are widely used.

Deep Blue Data will accept data in proprietary formats provided that these formats are appropriate for the research communities who are likely to have an interest in the data. However, it may not be possible to provide as high a level of preservation service for proprietary formats (see Preservation Policy).

Retention Review

Data submitted to Deep Blue Data will be reviewed after 10 years to determine if a data set should be retained and be subject to further, periodic, reviews thereafter. The goal of these reviews is to identify and possibly remove data that have reached the end of their use and reuse life cycle, or have become inaccessible (e.g. because of format obsolescence). The retention review will be conducted by the Data Curation Librarian, appropriate subject librarian(s), and, whenever possible, the depositor. The retention decision will be driven by a determination of the ongoing value to the research community. Long-term retention will also be determined by file format based preservation levels assigned upon deposit. Any data removed from the repository will be returned to the depositor whenever possible and documented with a tombstone record, which is the remaining metadata from a deleted record kept for the purposes of permanence.

Removing work from Deep Blue Data

Depositors can remove their work from Deep Blue Data with the assistance of and after consultation with staff if there is a mutual determination that the work is not appropriate for the service. Whenever work is removed, a tombstone record will remain.

If the depositor requests that the data be withdrawn from Deep Blue Data, the Library will take the following factors into consideration:

- If the data has been shown to contain inaccurate or faulty information
- If there is evidence of the data being used, cited, or downloaded

The Library also reserves the right to remove any deposit for reasons including:

- It was not appropriate for deposit (e.g. it contains sensitive information, viruses or other malware, or if we receive a verified complaint that it contains materials determined to be an infringement of copyright)
- It is no longer of active interest as described below (see the Retention Review section)

In such cases we will make reasonable attempts to contact the depositor so they can arrange for a new home for the data. A tombstone record will always remain for any deposit that is removed.

Copyright and Take-Down Notification

Please refer to the [library](#) and [University policy and procedures on copyright and take-down](#).

<https://www.dropbox.com/sh/r1o7aok5q06bv8v/cbFsK7KFA/sampling.doc>

UNIVERSITY of NOTRE DAME

HESBURGH LIBRARIES

CurateND

Search CurateND

ABOUT

LOG IN

FAQ

[Return to Policies Listing](#)

Retention and Review of CurateND Policy

Effective Date: June 10th, 2016

CurateND is intended to provide permanent access to and preservation of content deposited in the repository. As a member of a global family of institutional repositories, CurateND also aspires to provide uncensored open access to its content. There are instances, however, when content may need to be removed from the repository.

Content Preserved in Perpetuity

Any materials deposited in CurateND meeting the following conditions will be retained in perpetuity without need of review.

- Persistent identifiers and related metadata:** All persistent identifiers such as DOIs and PURLs, related records, and other metadata (i.e. descriptive information) will be preserved regardless of continued preservation decisions of any associated files.
- Work types the University is committed to preserve:** Certain works such as theses and dissertations, articles, library curated collections, or related files will be retained in perpetuity regardless of other conditions laid out here.
- Open access files (or sets of files) smaller than CurateND Size threshold:** Any files or sets of files attached to a publicly accessible work (and thus publicly accessible metadata) smaller than the size threshold defined below will be retained in perpetuity.

Content Reviewed Periodically

Any content that does not meet the above conditions will be reviewed for retention by the CurateND team, as summarized in this table. More information about the review periods and the types of review is included below.

Metadata access	Content access	Review period	Types of review
Private	Private	Every 5 years	Scholarly value; sponsor requirements; increase access
Univ. of Notre Dame	Private	Every 5 years	Scholarly value; sponsor requirements; increase access
	Univ. of Notre Dame	Every 10 years	Scholarly value; sponsor requirements; increase access
Open access	Private	Every 5 years	Scholarly value; sponsor requirements; increase access
	Univ. of Notre Dame	Every 10 years	Scholarly value; sponsor requirements; increase access
	Open access	Every 10 years	Size threshold; scholarly value; sponsor requirements

Review Periods

Content (and associated metadata) not retained in perpetuity will be reviewed every 5 or 10 years, as indicated above. The clock starts from the year that the content was deposited. For example, if private content with private metadata is deposited in CurateND in 2017, it will be reviewed in 2022, 5 years after depositing.

Types of Review

Scholarly Value Determination

In consultation with the content owner, if possible, a library subject specialist, related campus department, or other domain expert will make a determination for continued retention. If none of these individuals or groups can be contacted or make a determination, the University Committee on Libraries (UCL) will be consulted. The determination to continue preservation will be made based on at least the following criteria:

- Have past usage rates via CurateND been high or low?
- Is content likely to be used, or continue to be used in the future (i.e., has the content been superseded by other scholarship)?
- Is the content deemed especially rare, ephemeral, unique, or significant?

Please note: Any work with metadata marked private (and thus with private content) may be removed if further preservation requirements are not documented or cannot be proven by content owners.

Also note: CurateND will make reasonable efforts to contact content owners based on available information. If content owners cannot be reached, the Hesburgh Libraries will make the final determination as to whether content should continue to be preserved.

Size Threshold

CurateND will maintain a size threshold for a single file. Currently, this size threshold is 50 GB. This size threshold itself will be periodically reviewed and extended based on technological advancements.

Research Sponsor Retention Requirements

If research funding or other project requirements deem that content needs to be preserved for a specified amount of time, the CurateND team will do so depending on the resource support needed for that content.

Please note: If the content has preservation requirements, but does not meet the scholarly value determination, exceeds the current size threshold, and was deposited more than 10 years ago, the CurateND team may seek compensation from the content owner in order to continue preserving the content.

Increase Access

For all content reviewed that is not open access (it is assumed open access content will also have open access metadata), the content owner or proxy will be asked to increase access to at least the next level, if copyright or other circumstances allow. For example, can private metadata be made accessible to the University of Notre Dame? Or if content is accessible to the University, can it be made open access?

CurateND is a service of the Hesburgh Libraries of Notre Dame.

Questions? Call (574) 631-6258 or email curate@nd.edu.

Help

Copyright © 2017 University of Notre Dame



VII. Deaccessioning Data

Items may be deaccessioned from the repository for the following reasons:

- copyright violation
- legal requirements and proven violations
- national security
- falsified research
- confidentiality concerns etc.

Items may also be deaccessioned from the repository by the depositor. Deaccessioning a dataset or a version of a dataset is a very serious action that should only occur if there is a legal or valid reason for the dataset to no longer be accessible to the public. If you absolutely must deaccession, you can deaccession a version of a dataset or an entire dataset. To deaccession, go to a dataset you've already published (or add a new one and publish it), click on Edit Dataset, then Deaccession Dataset. If you have multiple versions of a dataset, you can select here which versions you want to deaccession or choose to deaccession the entire dataset. You must also include a reason as to why this dataset was deaccessioned from a dropdown list of options. There is also a free-text box to add more details as to why this was deaccessioned. If the dataset has moved to a different repository or site you are encouraged to include a URL (preferably persistent) for users to continue to be able to access this dataset in the future.

Important Note: A tombstone landing page with the basic citation metadata will always be accessible to the public if they use the persistent URL (Handle or DOI) provided in the citation for that dataset. Users will not be able to see any of the files or additional metadata that were previously available prior to deaccession.

Should a dataset be removed by either the repository or the depositor, TDL reserves the right to retain its citation metadata record in the repository as trace of the dataset. Additionally, the citation metadata of withdrawn items will be searchable.

References


DISC-UK DataShare Project, "Policy-making for Research Data in Repositories: A Guide," <https://www.coar-repositories.org/files/guide.pdf>

Dataverse Project, "User Guide: Dataset + File Management," <http://guides.dataverse.org/en/latest/user/dataset-management.html>

Footnotes

1. These General Terms of Use are adapted from Harvard Dataverse generic best practices templates created for these purposes. For original, see: <http://best-practices.dataverse.org/harvard-policies/harvard-terms-of-use.html>
2. The Privacy Policy is adapted from Harvard Dataverse best practices generic templates created for these purposes. For the original, please see: <http://best-practices.dataverse.org/harvard-policies/harvard-privacy-policy.html>
3. Adapted from <https://creativecommons.org/publicdomain/zero/1.0/>
4. Adapted from the Data Citation Synthesis Group, "Joint Declaration of Data Citation Principles": <https://www.force11.org/group/joint-declaration-data-citation-principles-final>
5. The Texas Data Repository Community Norms are adapted from Harvard Dataverse best practices templates created for these purposes. For original templates, please see <http://best-practices.dataverse.org/harvard-policies/community-norms.html>. Important modifications to this section include more extensive use of the Joint Declaration of Data Citation Principles.
6. Adapted from Data Citation Synthesis Group: Joint Declaration of Data Citation Principles. Martone M. (ed.) San Diego CA: FORCE11; 2014 [datacitation].
7. The Data Usage Agreement is adapted from the Harvard best practices templates created for these purposes. For original template, please see <http://best-practices.dataverse.org/harvard-policies/sample-dua.html>

Data Curation Job Descriptions



[Advanced Search](#) [Job Map](#) [Employer Panel](#)


[POST A JOB](#) [FIND A JOB](#)

Job


Data Services Manager

at Johns Hopkins University ([view profile](#))


Location

 Baltimore, MD


Date Posted

 January 23, 2017


Category

 Academic

Job Type


 Full-time

Apply Here

 <https://jobs.jhu.edu/hujobs/jobview.cfm?reqId=312736&postId=13364>

Description

The Sheridan Libraries have recently created a Data Management Directorate encompassing data infrastructure, applications and services. Reporting to the Associate Dean of Research Data Management, Sayeed Choudhury, this new directorate recognizes explicitly the role of data as a new form of collections. We are recruiting a Data Services Manager as part of this new organization.



The Data Services Manager manages the operations of the JHU Data Management Services (DMS) and Geographical Information Systems (GIS) unit including: a) team of four data management specialists and three GIS specialists that provide data management planning, training, archiving and geospatial data services and support to researchers across Johns Hopkins University, and b) new and ongoing collaboration and partnerships with other departments and units across JHU. In particular, the Manager must work closely with the Sheridan Libraries' Academic Liaisons and the Digital Research and Curation Center and the Welch Medical Library.

The Sheridan Libraries and University Museums are strongly committed to diversity. A strategic goal of the Libraries and Museums is to work toward achieving diversity when recruiting new staff and promoting existing staff. The Libraries and Museums prize initiative, creativity, professionalism, and teamwork.

About DLF Jobs

Postings are free for [DLF member institutions](#), and \$200 for non-members, which includes friendly, customized promotion of your position online. Let us help you attract great candidates! [More](#).

Sign up for job alerts

Keyword

E-mail

[CREATE EMAIL ALERT](#)

SPEC Kit 354: Data Curation

125



University of Michigan Library

POSITION DESCRIPTION

Job Description

The Research Data Curation Librarian will advance the library's mission to create and sustain data services for the campus that support the mission of the University of Michigan researchers through the library's Research Data Services (RDS) unit. A key focus of this position will be to contribute to the development of the data repository in collaboration with colleagues and stakeholders, in the library and across campus.

Date: 8/2015

Department: Science, Engineering, Clark Library and Research Data Services

Working Title: Research Data Curation Librarian

University Classification: <Librarian>

Position Summary:

The University of Michigan Library has embarked on an aggressive and exciting initiative to address research data management and curation needs at the University.

RDS is responsible for strategic planning, coordination, and deployment of research data services directed at facilitating the research lifecycle. This includes creating and implementing data management assistance for the campus, outreach to faculty in collaboration with librarian subject specialists, informationists, training, and assessment of RDS programs and services. RDS operates in 4 key areas: 1) Education, Awareness and Community Building, 2) Technical Infrastructure, 3) Policy and Strategy, and 4) Consultation and Services.

The responsibilities of the Research Data Curation Librarian will fall in all four of the above areas, with a particular focus on developing and maintaining the services offered through the research data repository in collaboration with colleagues and stakeholders, in the library and across campus.

Reporting Structure:

Reports to the Research Data Services Manager

Supervisory Experience:

This is a largely collaborative position that requires negotiation of relationships across the library

and the University. As such, it will require student supervision and deployment experience, but has no FTEs reporting to it.

Responsibilities (essential functions):

While partnering with colleagues at the U-M Office of Research, Information Technology Services, Advanced Research Computing, as well as academic programs, institutes, departments, and colleges across campus, the Research Data Curation Librarian will:

Work with researchers to curate and archive data (30%)

The Librarian will work with researchers to identify, recruit, ingest and deposit data in the library's digital repository, adhering to local policies, national and international standards and best practices. The incumbent will play a significant role in outreach to the research community to deposit data in both the digital repository or an appropriate subject repository, as well as creating training programs, help guides and web resources for Data Education and RDS for internal and external audiences. When necessary the Librarian will consult with researchers on their specific needs such as adopting metadata standards or data sensitivity characterization.

Create, support and sustain technical infrastructure (20%)

In collaboration with key partners, the incumbent will act as the point person for the data repository, investigate integrative infrastructures to connect campus needs to the repository, design and implement workflows, and execute technical processes involved in managing the lifecycle of digital datasets including data transformation projects.

Work with campus stakeholders on larger data collections issues (15%)

In addition to serving as a consultant to researchers and librarians on data issues and services, performs data management planning with principle investigators and researchers; assists in the development and delivery of training and instructional materials on data curation; provides guidance and instruction on discovery, acquisition and use of research data in the public domain.

Engage and participate in all aspects of the RDS and library services as appropriate (25%)

The Research Data Curation Librarian will participate in developing RDS within the Library and actively working to promote and advance the components of RDS amongst librarians. This includes the development of resources, documentation and instructional content about data curation, participating in selected cross-library working groups to create and improve services. Other duties as assigned.

Professional Development (10%)

Pursue research and professional development activities individually and as appropriate to the position. Engage with the library community and communities of practice beyond the library.

Required Qualifications:

- ALA-accredited Master's degree or an equivalent combination of a relevant advanced degree and experience
- Demonstrated knowledge of or direct experience managing and curating research data
- Knowledge of information technologies, standards and best practices prevalent in digital or data curation
- Ability to articulate roles in the research data ecosystem
- Knowledge of technologies for data management and curation, and familiarity with preservation principles and practices
- Ability to work independently and effectively with others as a team within a complex and fluid organization. Ability to work well in a multicultural and collaborative environment
- Possess excellent written and oral communication skills; ability to present and share ideas clearly and effectively to a diverse audience

Desired Qualifications

- Experience working with digital repository or content management systems
- Experience documenting workflows and procedures
- Knowledge of metadata formats, including Dublin Core, MODS, METS, and data exchange protocols such as SWORD and OAI-PMH.
- Experience in identifying researcher information needs and in creating effective services to meet those needs
- Demonstrated experience in the acquisition and management of born-digital or digitized library, archival, or research materials
- Demonstrated time management and project completion skills
- Demonstrated commitment to customer service

Digital Library Data Curation Developer

The Hesburgh Libraries is seeking a passionate software developer to join our Digital Library Technology Unit in support of digital library and research data curation services. With an emphasis on data curation, the individual will design and develop digital library frameworks and applications in areas such as controlled vocabularies, digital collections, digital content harvesting. Within science, engineering, and the social sciences, the individual will work with librarians, campus partners, and researchers to embed research data curation tools and workflows into active research projects for archiving and sharing data in our institutional repository CurateND (<http://curate.nd.edu>), or other relevant community repositories. This will involve combining data tool and architecture design with development of automated data extraction utilities and linked data technologies to apply domain specific metadata. The individual will also develop web based user clients for researchers to manage and browse research data. Additionally, the individual will contribute to our digital library frameworks and applications in areas such as controlled vocabularies, digital collections, digital content harvesting, and general support of digital library applications.

This position includes the opportunity to join us in a vibrant open source project called Hydra (<http://projecthydra.org>) in which we have partnered with several other universities and organizations to create advanced digital library applications and services.

Job duties include:

- Design and develop digital library applications supporting digital library and data curation services
- Provide technical leadership in data architecture and design for digital library data projects in collaboration with the Digital Library Technology Unit
- With campus partners, develop services and web clients to manage, archive, and share research data
- Create APIs and processes to integrate other campus systems with CurateND from groups like Engineering Science and Computing, Center for Research Computing, and Digital Production.
- Work with librarians and campus partners through our Center for Digital Scholarship to develop data models and tools to tag and describe data and collections with domain specific ontologies
- Provide software development support for research projects involving computational analysis or scientific data. This may involve manipulating or analyzing data with a statistical/computational package (e.g. R, SciPy, Matlab, Mathematica, STATA)
- Support digital humanities projects as needed with automated text analysis, topic modeling, and other methods

Minimum Qualifications

- Bachelors degree in Computer Science or related discipline, or equivalent software development work experience.
- At least 2 years experience working with at least one programming language (such as Python, Ruby on Rails, C, C++, Java, Python).
- At least 2 years experience creating relational databases using Oracle, MySQL, Postgres, or other modern RDBMS.
- Experience developing web based user interfaces and/or applications

- Experience designing and implementing APIs or middleware related services
- Excellent personal skills in order to work closely with customers throughout the research lifecycle

Preferred Qualifications

- Experience developing against digital repository systems such as Hydra, Islandora, Fedora Commons, or DSpace
- Experience with search indexes such as Solr, Lucene, and ElasticSearch
- Experience with research ontologies, RDF, or other linked data technologies
- Experience developing search, browse, or other visualization interfaces for research data
- Experience with computational and statistical packages such as R, Matlab, SPSS, SAS, and STATA.
- Applied research experience as either a member or in support of a science or engineering research project involving data computation or analysis
- Experience with digital humanities computational techniques such as text mining, or topic modeling