Representative Documents
Data Repositories
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

1. Prepare. Gather your data and the information that others would need to use it. (Submission Basics)
2. 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)
3. 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)
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.
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.

Featured Works

© 2015 Regents of the University of Michigan
UNIVERSITY OF MINNESOTA LIBRARIES
Data Repository for U of M
http://conservancy.umn.edu/handle/11299/166578

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 ensures 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.
Canada's federal research agencies are strong advocates for making publicly-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 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 on 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:
Texas Data Repository
https://tdl.org/texas-data-repository/

- 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 TDLs 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

Texas Data Repository
Welcome to VTechData!

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

Purpose

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.

Access Policy

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.

Deposit Policy

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.
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 (DOI's) 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.
Data Curation Services
What We Do

DSS collaborates with UC 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.

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 output and scholarly reputation

Data curation is the active management of digital 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
- Thematize data into repositories for access and preservation
- Capture metadata to allow retrieval
- Create permanent, replaceable hyperlinks
- Connect your data with your publications

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

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, reformatable, and born-digital content. We can help you:

- Selects content to stabilize or fragile formats to surmount formats
- Store content in trusted repositories for long-term access
- Use reliable tools to harvest, build, and preserve collections of web content
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.

DMP Tool

Create, review, and share data management plans that meet institutional and funder requirements. Examples:
- NIH Office of Digital Humanities
- NIH Generic
- NSF Generic
- NSF Biological Sciences
- NSF Engineering
- NSF Social, Behavioral, Economics Sciences

dash

A simple self-service tool to archive and share your research data for accelerated advancement of knowledge. Examples:
- Bannister, Caitlin C; Colabotti, Gaetano A; Deyo, Daniel; McIlroy, David L (2015) Simulation input and output data analysis for calculating partition coefficients of small molecules in octanol/water and cyclohexane/water. UC Irvine. Collection. doi:10.7280/b195e6n

EZID

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:8727/Milov81f
  doi:10.5641/00000001
  Learn about ARK | Learn about DOI

UNIVERSITY OF CALIFORNIA, IRVINE LIBRARIES
Digital Scholarship | Data Curation
http://www.lib.uci.edu/dss/data-curation
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.
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
  - Plan Your Research Project

- In Progress
  - Manage and Analyze Data

- Almost Done
  - Publish and Share Data / Software

Our Services

- Consulting
  - Assistance with writing data sharing or management plans

- Training
  - Best practices for data management and sharing

- Archiving
  - Archive your research data for sharing & preservation
Research Data Services

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 Dimitry Sunseifer, antilieu adaliku, Zlatko Nadjenovski, Norbert Kucsera, Giorgio Libiod, and Pascal Conil- tacoste via the Noun Project.
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://bit.ly/2ysmtul for additional information.

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 consent forms that facilitate sharing later. (Use the table to start.

Manage
- **Metadata Consultation and Data Management Training**
  - Metadata advice and consultation. Larger projects may want to include library staff as cost-sharers 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.
  - Refer 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**
  - UW-M researchers may deposit data to the Library’s 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.
  - Socially engaged about the Data Repository for U of M (DRUM)
  - Browse the collection of available data accepted to DRUM.

Grant Partner
- **Full Lifecycle Support for Data**
  - On request

The Librarians 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 bozemanite language for your Data Management Plan.
ABOUT

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

MANAGE PRESERVE DISCOVER SHARE

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
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
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
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
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.

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 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.
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 postprints.

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, offline and offline 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 DOI (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
RUTGERS UNIVERSITY LIBRARIES
About RUresearch
https://rucore.libraries.rutgers.edu/research/about/

RUCore: Rutgers University Community Repository

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.
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, particular 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.

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.
**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 it all 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.
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

- Is created in
- Is analyzed in
- Is published in

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.
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 vtechdata@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 vtechdata@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 13009 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.
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.
Scholars Portal Dataverse Guide

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 Datasets and Dataverses. 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.
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 of 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


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

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 form 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
Data Curation Workflows
Curation Workflow: Data Repository for the University of Minnesota (DRUM)
https://drive.google.com/file/d/0B5Dm3XFQloc4d0Q0Nk5ISW92TGs/view?usp=sharing
Dataset deposit process

Depositing data in VTechData can be done as a self-service operation or mediated by the library Data Services team.

Key:
- Blue: Researcher task
- Green: Data Services team

Self-Service

Start

Upload individual files to VTechData

Add files to dataset in VTechData

Request a DOI for the dataset

Check that dataset meets DOI criteria

Doesn't meet criteria

Meet criteria

Data Curator reviews to ensure dataset meets DOI criteria

Data Curator requests a DOI for the dataset

Dataset is published

Mediated

Start

Discuss data with library Data Curator

Data Curator creates optimized archival versions of data files

Data Curator uploads individual files to VTechData

Data Curator adds files to dataset in VTechData

Data Curator processes dataset to ensure it meet criteria

Researcher provides additional data, formats, input

Discuss data with researcher

Data Curator requests a DOI for the dataset

Dataset is published

VIRGINIA TECH UNIVERSITY LIBRARIES
Dataset Deposit Process
http://guides.lib.vt.edu/c.php?g=465788&p=3688869

Representative Documents: Data Curation Workflows
Digital Research Materials Repository Curation Workflow

https://drive.google.com/file/d/0B5Dm3XFQloc4URTtZHnQ09QNnc/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.
### Predicate Display label Content Type Input description Cardinality

<table>
<thead>
<tr>
<th>Predicate</th>
<th>Display label</th>
<th>Content Type</th>
<th>Input description</th>
<th>Cardinality</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://purl.org/dc/terms/alternative">http://purl.org/dc/terms/alternative</a></td>
<td>Alternative</td>
<td>String, title alternative form.</td>
<td>Already in input page but not displaying?</td>
<td>many</td>
</tr>
<tr>
<td><a href="http://purl.org/dc/terms/contributor">http://purl.org/dc/terms/contributor</a></td>
<td>Contributor</td>
<td>String, generally personal name. e.g. &quot;Butler, Octavia&quot;</td>
<td>This is also in the input page, but not displaying for books.</td>
<td>many</td>
</tr>
<tr>
<td><a href="http://purl.org/dc/terms/contributor#artist">http://purl.org/dc/terms/contributor#artist</a></td>
<td>Contributing Artist</td>
<td>String, generally personal name. e.g. &quot;Butler, Octavia&quot;</td>
<td>An entity responsible for creating artistic works within the resource, other than illustrations.</td>
<td>many</td>
</tr>
<tr>
<td><a href="http://purl.org/dc/terms/contributor#author">http://purl.org/dc/terms/contributor#author</a></td>
<td>Coauthor</td>
<td>String, generally personal name. e.g. &quot;Butler, Octavia&quot;</td>
<td>An authorial entity who contributed to the resource.</td>
<td>many</td>
</tr>
<tr>
<td><a href="http://purl.org/dc/terms/contributor#editor">http://purl.org/dc/terms/contributor#editor</a></td>
<td>Contributing Editor</td>
<td>String, generally personal name. e.g. &quot;Butler, Octavia&quot;</td>
<td>An entity responsible for editing the resource.</td>
<td>many</td>
</tr>
<tr>
<td><a href="http://purl.org/dc/terms/contributor#illustrator">http://purl.org/dc/terms/contributor#illustrator</a></td>
<td>Contributing Illustrator</td>
<td>String, generally personal name. e.g. &quot;Butler, Octavia&quot;</td>
<td>An entity responsible for illustrating the resource.</td>
<td>many</td>
</tr>
<tr>
<td><a href="http://purl.org/dc/terms/contributor#photographer">http://purl.org/dc/terms/contributor#photographer</a></td>
<td>Contributing Photographer</td>
<td>String, generally personal name. e.g. &quot;Butler, Octavia&quot;</td>
<td>An entity responsible for creating photographic works within the resource.</td>
<td>many</td>
</tr>
<tr>
<td><a href="http://purl.org/dc/terms/creator">http://purl.org/dc/terms/creator</a></td>
<td>Inventor</td>
<td>String, generally personal name. e.g. &quot;Butler, Octavia&quot;</td>
<td>An entity listed on the patent as a creator.</td>
<td>Y</td>
</tr>
<tr>
<td><a href="http://purl.org/dc/terms/creator(unqualified)">http://purl.org/dc/terms/creator(unqualified)</a></td>
<td>Creator</td>
<td>String, generally personal name. e.g. &quot;Butler, Octavia&quot;</td>
<td>An entity responsible for the resource’s creation.</td>
<td>many</td>
</tr>
<tr>
<td><a href="http://purl.org/dc/terms/creator#administrative_unit">http://purl.org/dc/terms/creator#administrative_unit</a></td>
<td>Department</td>
<td>String</td>
<td>Relevant academic departments</td>
<td>Y</td>
</tr>
<tr>
<td>--------------------------------------------------</td>
<td>------------</td>
<td>--------</td>
<td>------------------------------</td>
<td>---</td>
</tr>
<tr>
<td><a href="http://purl.org/dc/terms/creator#artist">http://purl.org/dc/terms/creator#artist</a></td>
<td>Artist</td>
<td>String, generally personal name, e.g. &quot;Butler, Octavia&quot;</td>
<td>An entity responsible for art works in a resource which consists primarily of art works (e.g. an art book).</td>
<td>many</td>
</tr>
<tr>
<td><a href="http://purl.org/dc/terms/creator#author">http://purl.org/dc/terms/creator#author</a></td>
<td>Author</td>
<td>String, generally personal name, e.g. &quot;Butler, Octavia&quot;</td>
<td>An entity responsible for significant authorial work within the resource.</td>
<td>many</td>
</tr>
<tr>
<td><a href="http://purl.org/dc/terms/creator#editor">http://purl.org/dc/terms/creator#editor</a></td>
<td>Editor</td>
<td>String, generally personal name, e.g. &quot;Butler, Octavia&quot;</td>
<td>An entity responsible for significant editorial work in creating the resource.</td>
<td>many</td>
</tr>
<tr>
<td><a href="http://purl.org/dc/terms/creator#illustrator">http://purl.org/dc/terms/creator#illustrator</a></td>
<td>Illustrator</td>
<td>String, generally personal name, e.g. &quot;Butler, Octavia&quot;</td>
<td>An entity responsible for illustrations of a resource which consists primarily of illustrations (e.g. a children's picture book)</td>
<td>many</td>
</tr>
<tr>
<td><a href="http://purl.org/dc/terms/creator#local">http://purl.org/dc/terms/creator#local</a></td>
<td>n/a</td>
<td>String, generally personal name, e.g. &quot;Butler, Octavia&quot;</td>
<td>Creators who are (or were) associated with the local institution. People are to be listed here in addition to being listed in dc:creator.</td>
<td>Y</td>
</tr>
<tr>
<td><a href="http://purl.org/dc/terms/creator#photographer">http://purl.org/dc/terms/creator#photographer</a></td>
<td>Photographer</td>
<td>String, generally personal name, e.g. &quot;Butler, Octavia&quot;</td>
<td>An entity responsible for photography in a resource which consists primarily of photographs, (e.g. a collection of a photographer's work).</td>
<td>many</td>
</tr>
<tr>
<td><a href="http://purl.org/dc/terms/date#application">http://purl.org/dc/terms/date#application</a></td>
<td>Application</td>
<td>String (date?) Has form &quot;YYYY-MM-DD&quot;</td>
<td>The date of the initial submission of the application for this patent.</td>
<td>N</td>
</tr>
<tr>
<td><a href="http://purl.org/dc/terms/date#prior_publication">http://purl.org/dc/terms/date#prior_publication</a></td>
<td>Prior</td>
<td>String (date?) Has form &quot;YYYY-MM-DD&quot;</td>
<td>Date of prior publication (?)</td>
<td>N</td>
</tr>
<tr>
<td><a href="http://purl.org/dc/terms/date#copyrighted">http://purl.org/dc/terms/date#copyrighted</a></td>
<td>Copyright</td>
<td>Should be in form YYYY, More granular dates unlikely</td>
<td>The resource's copyright date</td>
<td>one</td>
</tr>
<tr>
<td><a href="http://purl.org/dc/terms/date#created">http://purl.org/dc/terms/date#created</a></td>
<td>Date Added</td>
<td>Date</td>
<td>Date object was created in CurateND</td>
<td>N</td>
</tr>
<tr>
<td>Property</td>
<td>Description</td>
<td>Type</td>
<td>Notes</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-------------</td>
<td>------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>description</td>
<td>String. May contain abstract</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>table_of_contents</td>
<td>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.</td>
<td>one</td>
<td></td>
<td></td>
</tr>
<tr>
<td>extent</td>
<td>String. Will probably be page length, e.g. “368 pages”</td>
<td>many</td>
<td></td>
<td></td>
</tr>
<tr>
<td>claims</td>
<td>String. The number of claims in this patent. Usually an integer, but has type string to handle any possible special cases.</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>isbn</td>
<td>Structured alphanumeric string (regex text: \d+-\d+), either 10 or 13 characters. They often contain hyphens, but can be normalized to not.</td>
<td>many</td>
<td></td>
<td></td>
</tr>
<tr>
<td>local_identifier</td>
<td>String. For most digitized books, will be the call number, but may be another kind of local identifier used to shelve or handle books.</td>
<td>many</td>
<td></td>
<td></td>
</tr>
<tr>
<td>other_application</td>
<td>String. A prior submission of this patent for review. (?)</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>patent_number</td>
<td>String. May contain spaces. E.g. “US 1234567890123 B2”</td>
<td>N</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### CurateND Metadata Model

<table>
<thead>
<tr>
<th>URI</th>
<th>Description</th>
<th>Type</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://purl.org/dc/terms/identifier#prior_publication">http://purl.org/dc/terms/identifier#prior_publication</a></td>
<td>Prior Publication Number</td>
<td>String</td>
<td>Identifier for the prior publication of this patent. Y</td>
</tr>
<tr>
<td><a href="http://purl.org/dc/terms/published_in">http://purl.org/dc/terms/published_in</a></td>
<td>Published in</td>
<td>String</td>
<td>The title of the journal, book, or other work in which the Article was published N</td>
</tr>
<tr>
<td><a href="http://purl.org/dc/terms/issued">http://purl.org/dc/terms/issued</a></td>
<td>Publication Date</td>
<td>Should be in form YYYY. More granular dates unlikely</td>
<td>The resource's publication date one</td>
</tr>
<tr>
<td><a href="http://purl.org/dc/terms/issued">http://purl.org/dc/terms/issued</a></td>
<td>Publication Date</td>
<td>Can we test as YYYY or YYYY-MM or YYYY-MM-DD</td>
<td>The article's publication date as year, year-month, or year-month-day, eg. 2015 or 2015-05 or 2015-05-31 N</td>
</tr>
<tr>
<td><a href="http://purl.org/dc/terms/issued">http://purl.org/dc/terms/issued</a></td>
<td>Date Issued</td>
<td>String (See date discussion above)</td>
<td>Date the patent was issued. N</td>
</tr>
<tr>
<td><a href="http://purl.org/dc/terms/language">http://purl.org/dc/terms/language</a></td>
<td>Language</td>
<td>String</td>
<td>Y</td>
</tr>
<tr>
<td><a href="http://purl.org/dc/terms/modification_date">http://purl.org/dc/terms/modification_date</a></td>
<td>Date Modified</td>
<td>Date</td>
<td>Date object was last modified in CurateND N</td>
</tr>
<tr>
<td><a href="http://purl.org/dc/terms/publisher">http://purl.org/dc/terms/publisher</a></td>
<td>Publisher</td>
<td>String</td>
<td>Y</td>
</tr>
<tr>
<td><a href="http://purl.org/dc/terms/rights">http://purl.org/dc/terms/rights</a></td>
<td>Rights</td>
<td>String</td>
<td>Intellectual Rights Y</td>
</tr>
<tr>
<td><a href="http://purl.org/dc/terms/assignee">http://purl.org/dc/terms/assignee</a></td>
<td>Assignee</td>
<td>String</td>
<td>Assignee of the patent. Y</td>
</tr>
<tr>
<td><a href="http://purl.org/dc/terms/uspto">http://purl.org/dc/terms/uspto</a></td>
<td>USPTO Link</td>
<td>URL</td>
<td>Link to the patent at the USPTO website (or other patent office websites). N</td>
</tr>
<tr>
<td><a href="http://purl.org/dc/terms/classification">http://purl.org/dc/terms/classification</a></td>
<td>Classification (CP)</td>
<td>String</td>
<td>Cooperative Patent Classification codes. Y</td>
</tr>
<tr>
<td>URI</td>
<td>Description</td>
<td>Type</td>
<td>Value</td>
</tr>
<tr>
<td>--------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------</td>
<td>---------</td>
<td>----------------</td>
</tr>
<tr>
<td><a href="http://purl.org/dc/terms/subject#lcsh">http://purl.org/dc/terms/subject#lcsh</a></td>
<td>Subject (Library of Congress) String. Taken from the book's bib record. E.g. Amnesia – Fiction.</td>
<td>many</td>
<td></td>
</tr>
<tr>
<td><a href="http://purl.org/dc/terms/title">http://purl.org/dc/terms/title</a></td>
<td>Title String Title of the Patent</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td><a href="http://purl.org/dc/terms/type">http://purl.org/dc/terms/type</a></td>
<td>n/a String The subtype of Work this item is.</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td><a href="http://purl.org/ontology/bibo/eIssn">http://purl.org/ontology/bibo/eIssn</a></td>
<td>e-ISSN e-ISSN validation? The e-ISSN of the publication in which the article appears</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td><a href="http://purl.org/ontology/bibo/issue">http://purl.org/ontology/bibo/issue</a></td>
<td>Issue String The number(s) or name of the issue in which the article appears.</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td><a href="http://purl.org/ontology/bibo/numPages">http://purl.org/ontology/bibo/numPages</a></td>
<td>Number of pages Integer The total number of pages as an integer</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td><a href="http://purl.org/ontology/bibo/lastPage">http://purl.org/ontology/bibo/lastPage</a></td>
<td>Last page String The number or other identifier of the article's final page.</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td><a href="http://purl.org/ontology/bibo/firstPage">http://purl.org/ontology/bibo/firstPage</a></td>
<td>First page String The number or other identifier of the first page on which the article appears, e.g. &quot;42&quot; or &quot;E594&quot;</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td><a href="http://purl.org/ontology/bibo/volume">http://purl.org/ontology/bibo/volume</a></td>
<td>Volume String The number or name of the volume in which the article appears.</td>
<td>N</td>
<td></td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Object location URL</td>
<td>The current location (URL) of the identified object.</td>
<td><a href="http://merritt.cdlib.org/o/ark%3A%2F13030%2Fqt5np807ch">http://merritt.cdlib.org/o/ark%3A%2F13030%2Fqt5np807ch</a></td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="http://opencontext.org/subjects/199ED3F0-8CA2-4BBD-FA14-468132555877">http://opencontext.org/subjects/199ED3F0-8CA2-4BBD-FA14-468132555877</a></td>
</tr>
<tr>
<td>Who</td>
<td>The name of an entity (person, organization, or service) responsible for creating the content or making it available, e.g. author, creator. Put name parts in &quot;sort-friendly&quot; order. Separate multiple names with &quot;;&quot;. Append one or more final commas (&quot;;&quot;) 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).</td>
<td>Kim, JH; Cho, J.; Keane, TD;Virginia Department of Historic Resources (VA-DHR); Open Context EditorsCanal Educatif à la Demande</td>
</tr>
<tr>
<td>What</td>
<td>A name or other human-oriented identifier given to the resource, e.g. a title.</td>
<td>Political fragmentation and land use changes in the Interior PlainsVirginia Site Files: 44WR0079 (Site)Vidéos Sciences &amp; Innovation de Canal Éducatif à la Demande</td>
</tr>
<tr>
<td>When</td>
<td>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 &quot;;&quot; to separate entries and &quot;~&quot; to indicate approximation.</td>
<td>10/4/20152014-07-31T00:00:00-07:001/1/2007</td>
</tr>
</tbody>
</table>
# EZID Identifiers Made Easy

**FOR DOIs**

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Creator</strong></td>
<td>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.</td>
<td>George, Christopher Worth, A [MGH] Owens, Allessia P.</td>
</tr>
<tr>
<td><strong>Title</strong></td>
<td>A name or title by which the data or publication is known.</td>
<td>Impact of Task Performance Fraud Risk Assessment on Forensic Skills and Mindsets: Experience from Nigeria Internet Brain Segmentation Repository Mentoring African American males</td>
</tr>
<tr>
<td><strong>Publisher</strong></td>
<td>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.</td>
<td>Figshare MGH CMA Maryland Shared Open Access Repository</td>
</tr>
<tr>
<td><strong>Publication year</strong></td>
<td>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.</td>
<td>2015 2015 2008</td>
</tr>
<tr>
<td><strong>Resource type</strong></td>
<td>The general type of the data.</td>
<td>Dataset Dataset Text</td>
</tr>
</tbody>
</table>
Using EZID’s UI to create a DOI, you must provide DataCite metadata. Mandatory DataCite properties are indicated with an asterisk (*).

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

For details about field constraints and all sub-properties, see [http://schema.datacite.org](http://schema.datacite.org)
Data Deaccessioning Policies
Deaccessioning Your Dataset [not recommended]

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 ofDatasets 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 ofDatasets 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 use services by developing data type-specific viewers/simulators.

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.
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., Non-disclosure 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.illinois.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 suspension 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 conditions 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


Contact Research Data Service staff with questions or to request an addition or revision to this policy.
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.library.umich.edu/docs/oa/oa2020info/oa-copyright-policy.pdf
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 the table. More information about the review periods and the types of review is included below.

<table>
<thead>
<tr>
<th>Metadata access</th>
<th>Content access</th>
<th>Review period</th>
<th>Types of review</th>
</tr>
</thead>
<tbody>
<tr>
<td>Private</td>
<td>Private</td>
<td>Every 5 years</td>
<td>Scholarly value; sponsor requirements; increase access</td>
</tr>
<tr>
<td>Univ. of Notre Dame</td>
<td>Private</td>
<td>Every 5 years</td>
<td>Scholarly value; sponsor requirements; increase access</td>
</tr>
<tr>
<td>Univ. of Notre Dame</td>
<td>Univ. of Notre Dame</td>
<td>Every 10 years</td>
<td>Scholarly value; sponsor requirements; increase access</td>
</tr>
<tr>
<td>Open access</td>
<td>Private</td>
<td>Every 5 years</td>
<td>Scholarly value; sponsor requirements; increase access</td>
</tr>
<tr>
<td>Univ. of Notre Dame</td>
<td>Open access</td>
<td>Every 10 years</td>
<td>Scholarly value; sponsor requirements; increase access</td>
</tr>
</tbody>
</table>

https://curate.nd.edu/policies/retention-review
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?
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**


**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](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](http://best-practices.dataverse.org/harvard-policies/harvard-privacy-policy.html)

3. Adapted from [https://creativecommons.org/publicdomain/zero/1.0/](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](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](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](http://best-practices.dataverse.org/harvard-policies/sample-dua.html)
Data Curation Job Descriptions
JOHNS HOPKINS UNIVERSITY LIBRARIES
Data Services Manager
https://jobs.diglib.org/job/data-services-manager/

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/jhujobs/jobs/view.cfm?repid=312736&postid=13304

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 will manage 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.
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
UNIVERSITY OF NOTRE DAME, HESBURGH LIBRARIES
Digital Library Data Curation Developer

• 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