

University of Massachusetts Medical School

eScholarship@UMMS

University of Massachusetts and New England
Area Librarian e-Science Symposium

2016 e-Science Symposium

Apr 6th, 12:00 AM

Data Repositories: The Answer that Actually Came with a Question

Lisa Johnston

University of Minnesota - Twin Cities

Follow this and additional works at: https://escholarship.umassmed.edu/escience_symposium



Part of the [Scholarly Communication Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-Share Alike 4.0 License](#).

Repository Citation

Johnston, L. (2016). Data Repositories: The Answer that Actually Came with a Question. *University of Massachusetts and New England Area Librarian e-Science Symposium*. <https://doi.org/10.13028/m22a-h336>. Retrieved from https://escholarship.umassmed.edu/escience_symposium/2016/program/5

Creative Commons License



This work is licensed under a [Creative Commons Attribution-Noncommercial-Share Alike 4.0 License](#). This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in University of Massachusetts and New England Area Librarian e-Science Symposium by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.

Data Repositories

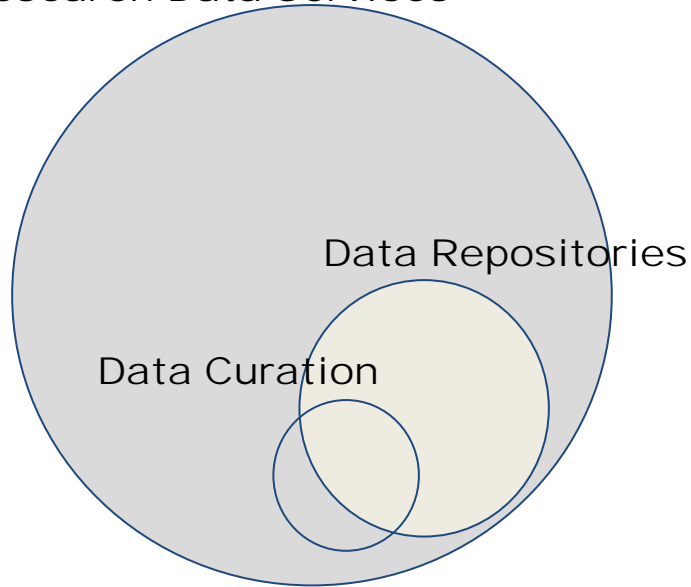
The answer that actually came with a
question

Lisa Johnston

Research Data Management/Curation Lead,
University of Minnesota

What are data repositories?

Research Data Services



“A repository (also referred to as a data repository or digital data repository) is a **searchable** and **queryable** interfacing entity that is able to store, manage, maintain and curate Data/Digital Objects. A repository is a **managed** location (destination, directory or ‘bucket’) where digital data objects are registered, permanently stored, made **accessible** and **retrievable**, and curated. Repositories preserve, manage, and provide access to many types of digital material in a variety of formats. Materials in online repositories are **curated** to enable search, discovery, and reuse. There must be sufficient control for the digital material to be authentic, reliable, accessible and usable on a continuing basis. (p3, footnote 2).”

Why are data repositories useful?



- Governments have open data initiatives
 - Democratize research results
 - Release the potential of valuable/\$\$ data
- Federal/private funders increasingly require data sharing
 - Public access
 - Return on \$\$ investment ⇒ do new research
- Journals have data sharing policies
 - Increase transparency
 - Facilitate reproducibility
- Researcher/disciplinary culture shift in digital age
 - Ease of sharing ⇒ greater potential for reuse
 - Citation impact, reputation building

What kinds of data repositories exist ?

Disciplinary Data Repositories



CERN *OpenData*: CMS data



And 1516 more listings at
<http://www.re3data.org/>

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

What kinds of data repositories exist ?

General Data Repositories



<http://openICPSR.org>



What kinds of data repositories exist?

Institutional Data Repositories

Running...

- **Dataverse**
- Digital Commons
- DSpace
- EPrints
- Fedora
- HubZero
- Hydra
- Islandora

The screenshot shows the Harvard Dataverse website interface. At the top, there is a navigation bar with a search icon, 'About', 'Guides', 'Support', 'Sign Up', and 'Log In'. Below this is the Harvard-Smithsonian Center for Astrophysics (CfA) logo and the text 'CfA HARVARD-SMITHSONIAN CENTER FOR ASTROPHYSICS'. A breadcrumb trail shows 'CfA Dataverse (Harvard-Smithsonian Center for Astrophysics) Home Page'. The main content area is titled 'Harvard Dataverse > CfA Dataverse' and contains introductory text about the repository's mission and the Astronomy Dataverse Network. A search bar is present with a 'Find' button and a link to 'Advanced Search'. Below the search bar, there are filters for 'Dataverses (34)', 'Datasets (111)', and 'Files (1,841)'. The 'Dataverse Category' filter shows 'Researcher (5)' and 'Research Project (2)'. The 'Publication Date' filter shows years from 2011 to 2012. The 'Author Name' filter lists several authors. The search results section shows '1 to 10 of 145 Results' and lists three results, each with a document icon, a title, a date, and a brief description.

Dataverse

Q About Guides - Support Sign Up Log In

CfA HARVARD-SMITHSONIAN CENTER FOR ASTROPHYSICS

CfA Dataverse (Harvard-Smithsonian Center for Astrophysics) Home Page

Harvard Dataverse > CfA Dataverse

This is the Astronomy data repository at Harvard. It is currently open to all scientific data from astronomical institutions worldwide. Administration and support is provided by the Harvard-Smithsonian Center for Astrophysics (CfA) in collaboration with Harvard Library (HL) and the Institute for Quantitative Social Science (IQSS). Infrastructure is provided by Harvard University Information Technology Services.

The Astronomy Dataverse Network plays an important role in fulfilling your Data Management Plan requirements (e.g. as mandated by NSF), and for providing data re-use and citation opportunities. Find out more about our team by exploring the [Seamless Astronomy](#) and [Wolbach Library](#) teams at the CfA. We are also connecting the Astronomy Dataverse to the indexing services provided by the [SAO/NASA Astrophysical Data Service \(ADS\)](#).

Search this dataverse... Q Find Advanced Search

Dataverses (34)
 Datasets (111)
 Files (1,841)

Dataverse Category
Researcher (5)
Research Project (2)

Publication Date
2011 (69)
2013 (36)
2014 (22)
2015 (9)
2012 (7)

Author Name
Thaddeus, P. (38)
Dame, T. M. (25)
COMPLETE team (20)
Hartmann, Dap (19)
Bronfman, L. (7)

1 to 10 of 145 Results II Sort -

APEX CMZ SHFI-1 survey
Aug 20, 2015 - APEX CMZ H2CO Dataverse
Ginsburg, Adam, 2015, "APEX CMZ SHFI-1 survey", <http://dx.doi.org/10.7910/DVN/27601>, Harvard Dataverse, V5
APEX SHFI-1 survey of the CMZ

Replication data for: Deep 3.8 Micron Observations of the Trapezium Cluster
Jun 23, 2015 - August Muench Dataverse
Muench, August; Alves, Joao; Lada, Charles; Lada, Elizabeth, 2015, "Replication data for: Deep 3.8 Micron Observations of the Trapezium Cluster", <http://dx.doi.org/10.7910/DVN/28977>, Harvard Dataverse, V2
This is the data behind the paper, "Deep 3.8 Micron Observations of the Trapezium Cluster," by Lada et al. (2004). It includes FITS image files and reference comparison files that would prove useful for interpreting the FITS image files. A note on the images: two images are given...

Replication Data for: YSOVAR: Mid-infrared Variability in the Star-forming Region Lynds 1688
May 15, 2015 - Hans Moritz Günther Dataverse
Guenther, Hans Moritz, 2015, "Replication Data for: YSOVAR: Mid-infrared Variability in the Star-forming Region Lynds 1688", <http://dx.doi.org/10.7910/DVN/LNRC6A>, Harvard Dataverse, V1
This dataset contains all YSOVAR lightcurves of L1688 as published in YSOVAR: Mid-infrared Variability in the Star-forming Region Lynds 1688. The data has been processed and cleaned as described in that article. Please note that additional electronic data (e.g. a table of source...

What kinds of data repositories exist?

Institutional Data Repositories

Running...

- Dataverse
- **Digital Commons**
- DSpace
- EPrints
- Fedora
- HubZero
- Hydra
- Islandora

The screenshot shows the Washington University Open Scholarship website. At the top, the text reads "WASHINGTON UNIVERSITY OPEN SCHOLARSHIP" next to the university's logo. Below this is a navigation bar with links for "Home", "About", "FAQ", and "My Account".

The main content area is divided into a left sidebar and a main right section. The sidebar contains a search box with the text "Enter search terms:" and a "Search" button. Below the search box is a dropdown menu labeled "in this series" and a link for "Advanced Search". There is also a link to "Notify me via email or RSS". The sidebar has three sections: "Links" with a link to "Data Management Research Guide", "Browse" with links for "Collections", "Disciplines", and "Authors", and "Author Corner" with links for "Author FAQ", "Policies", "Submission Guidelines", and "Submit Research".

The main right section has a breadcrumb trail "Home > DATA" and a heading "DATA AND DATA SETS". Below the heading is a paragraph of text: "This data series is a place for WUSTL affiliates, including faculty, students, and affiliated researchers, to share and publish digital data 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:" followed by a bulleted list of three items: "Flexible Access Options", "Long-term Access", and "Meet Grant Requirements". Below this list is another paragraph: "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." At the bottom right of the main content area is a blue "Follow" button.

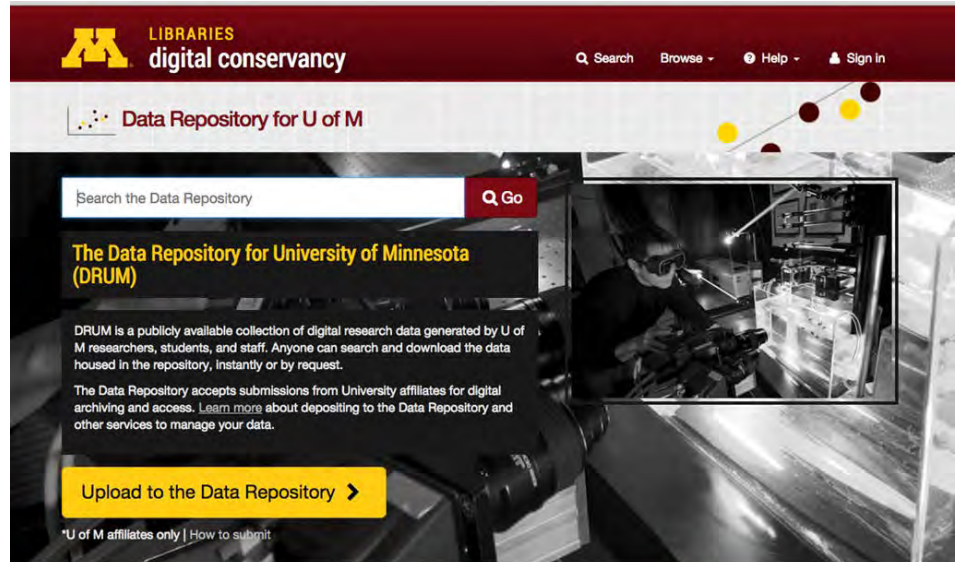
At the bottom of the page, there is a footer with the Digital Commons logo (powered by bepress), navigation links for "Home | About | FAQ | My Account | Accessibility Statement", and the Washington University in St. Louis logo next to a "LIBRARIES" logo.

What kinds of data repositories exist ?

Institutional Data Repositories

Running...

- Dataverse
- Digital Commons
- **DSpace**
- EPrints
- Fedora
- HubZero
- Hydra
- Islandora



LIBRARIES
digital conservancy

Search Browse Help Sign In

Data Repository for U of M

Search the Data Repository **Go**

The Data Repository for University of Minnesota (DRUM)

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

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

Upload to the Data Repository >

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

How to Upload

1. Prepare Data

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

2. Upload

Have your files ready (up to 2GB each) and use the upload form to fill out metadata about your data.

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.](#)

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

What kinds of data repositories exist ?

Institutional Data Repositories

Running...

- Dataverse
- Digital Commons
- DSpace
- **EPrints**
- Fedora
- HubZero
- Hydra
- Islandora

The screenshot shows the homepage of the University of Bath Research Data Archive. At the top, there is a blue navigation bar with the University of Bath logo and the text 'UNIVERSITY OF BATH'. Navigation links for 'Home', 'About', and 'Browse' are present. Below the navigation bar, there are links for 'Login' and 'Create Account', and a search input field with a 'Search' button. The main content area features a large heading 'Welcome to University of Bath Research Data Archive' and a sub-heading 'Find data from our world class research'. Below this is a search bar with a 'Search' button. A 'Latest Additions' box is also visible. At the bottom, there are three columns of links: 'About this Archive', 'Browse Archive', and 'Archive Policies'. The footer contains information about OAI 2.0 support and social media icons for Atom and RSS.

What kinds of data repositories exist?

Institutional Data Repositories

Running...

- Dataverse
- Digital Commons
- DSpace
- EPrints
- **Fedora**
- HubZero
- Hydra
- Islandora

The screenshot shows the RUresearch Data Portal website. At the top, there is a navigation bar with links for 'Help', 'Contact Us', and 'My Account'. Below this, the 'RUcore' logo is displayed, followed by 'Rutgers University Community Repository'. A search bar and buttons for 'Search', 'Services', and 'Collections' are also present. The main content area features a 'Home' link and the title 'RUresearch Data Portal'. A descriptive paragraph states that the portal provides a place for Rutgers researchers to share their research data. Below this is a search section with a search box and a 'Search' button, along with links for 'Browse All Projects' and 'Advanced Search'. A 'Learn More' section follows, detailing 'RUresearch Services' such as free assistance, customized search, ongoing support, and guidance for NSF requirements. A 'What does RUresearch Provide' section lists services like robust preservation, customized metadata, access control, and information presentation. On the right side, there is a blue graphic of a brain composed of various data-related icons, and a sidebar with links to 'About RUresearch', 'RUresearch Data Team', 'NSF Data Management', 'Data Management and RUresearch Presentation', 'Highlighted RUresearch Collections', 'Video Mosaic Collaborative', and 'Ralston Horse Behavior Research'.

What kinds of data repositories exist ?

Institutional Data Repositories

Running...

- Dataverse
- Digital Commons
- DSpace
- EPrints
- Fedora
- **HubZero**
- Hydra
- Islandora

PURDUE UNIVERSITY Purdue University Research Repository

Login Register Have a Question?

Home Publications Projects Get Started Policies Contact Us

Search

Public Access to Federally Funded Research

The [Office of Science and Technology Policy](#) has directed federal agencies with more than \$100M in R&D expenditures to develop plans to make the published results of federally-funded research, including datasets, freely available to the public within one year of publication.

Learn More



- Create a Data Management Plan**
Learn about the detailed requirements for your data management plan (DMP). Funding agency requirements are very specific and our DMP resources can help you to clear up any confusion. [Get Started](#)
- Upload Research Data to Your Project**
Create a project to upload and share your data with collaborators using our step-by-step form to guide you through the process. Invite collaborators from other institutions to join your project. [Create a Project](#)
- Publish your Dataset**
Package, describe, and publish your dataset with a Datacite DOI. Publishing will ensure your dataset is citable, reusable, and archived for the long-term. [See Published Datasets](#)

LARS - Vegetation and Soils Field Research Data

Laboratory for Applications of Remote Sensing (LARS)
Purdue University



The collection of 127 downloadable datasets, collected between 1975 and 1991, consists of calibrated spectral data and the associated biophysical and meteorological data.

Username: OPTIONAL

Name:

E-mail:

Question:

Submit

What kinds of data repositories exist?

Institutional Data Repositories

Running...

- Dataverse
- Digital Commons
- DSpace
- EPrints
- Fedora
- HubZero
- **Hydra**
- Islandora

The screenshot shows the Penn State ScholarSphere website. At the top, there is a dark blue header with the Penn State logo and the text "ScholarSphere". A "Login" button is in the top right corner. Below the header is a navigation bar with links for "Home", "About", "Help", and "Contact". A search bar is located on the right side of the navigation bar, with a "Go" button and a dropdown menu set to "All".

The main content area features a large banner image with a scientific theme. Overlaid on this image are two white text boxes. The first box asks, "Need help with research data management?" and the second box says, "Consult the Libraries' guide on RDM services. Questions? Contact the Research Data Management Team." To the right of these boxes is a green button that says "Share Your Work" with a circular arrow icon, and a link for "Terms of Use".

Below the banner, there are three main sections:

- Featured Works:** This section lists three items with their titles, depositors, and keywords. The first item is "Analytical Comparisons of..." by Mohannad Ahmed Abdurahman, with keywords "radio, astronomy, telescope". The second is "4 1/2 stars: Creating a..." by Jennie Anne Levine Knies, with keywords "Libraries, Undergraduate research, Penn State Wilkes-Barre". The third is "The Uptake of Vitamin A..." by Joanna Kalina Urbanek, with keywords "vitamin A, brain, mathematical modeling".
- Recent Additions:** This section is currently empty.
- Featured Researcher:** This section features a profile for Dr. Laura Guertin, Professor of Earth Science at Penn State Brandywine. It includes a photo of her and a short bio. The bio mentions her pedagogical research agenda and her use of ScholarSphere. Below the bio, there is a paragraph explaining why she uses ScholarSphere: "At scientific conferences, more and more presentations are being given in a poster format, on 4 foot by 8 foot boards. A poster is an excellent way for me to capture highlights of my work with text and images, but it is difficult for colleagues and other researchers to take away all of the information after viewing the poster."

What kinds of data repositories exist ?

Institutional Data Repositories

Running...

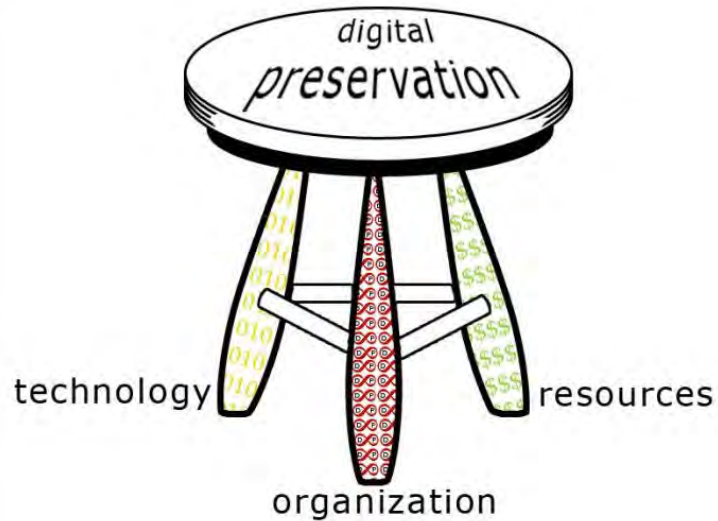
- Dataverse
- Digital Commons
- DSpace
- EPrints
- Fedora
- HubZero
- Hydra
- **Islandora**



The screenshot shows the SFU Research Data Repository (Radars) website. The header includes the SFU logo, the text "SIMON FRASER UNIVERSITY ENGAGING THE WORLD", and the title "Research Data Repository". A search bar and navigation links for "A-Z DIRECTORY", "This site", and "SFU.ca" are also present. The main content area is titled "Radars: SFU's Research Data Repository" and includes a welcome message: "Welcome to Radars, SFU's Research Data Repository. Here, you can:". Below this, there are three main sections: "Deposit data" with a red icon of a folder and a downward arrow, "Browse data" with a blue icon of a folder and a magnifying glass, and "Access your private data workspace" with a dark grey icon of a folder and a padlock. A sidebar on the left contains a menu with options: "BROWSE RESEARCH DATA", "PRIVATE DATA WORKSPACE", "DESKTOP SYNC", "DEPOSIT DATA", "REPOSITORY SEARCH", "DEPOSIT TERMS", "SUPPORT AND SERVICES", and "ADMIN LOGIN". A "Log in" button is located at the bottom of the sidebar. At the bottom of the main content area, there is a paragraph of text: "In addition to providing a place for you to deposit your research data, the SFU Library can assist in managing your data. Do you want to publish a completed data set alongside a research paper, need some help in formatting, or just looking for extra storage so that you can cope with the volume of your own backups? If you have any questions, we'd love to talk to you about your data needs. Please see our guides on [research data management](#) or contact us at lib-data@sfu.ca for more information."

How do I build an institutional data repository?

Blueprint for Institutional Data Repositories

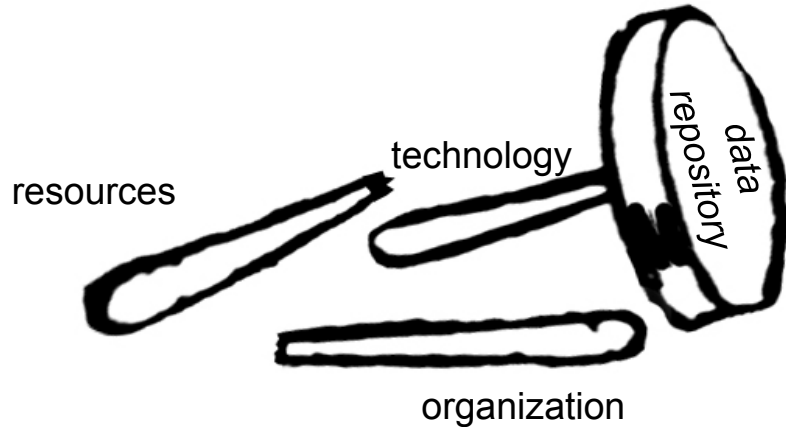


Nancy McGovern's Three-Legged Stool of building an organization's digital preservation program...

- technological infrastructure
- organizational infrastructure
- resources framework

How do I build an institutional data repository?

Blueprint for Institutional Data Repositories



Lisa's three-legged stool for building a data repository (basically the same...)

- technological infrastructure
- organizational infrastructure
- resources framework

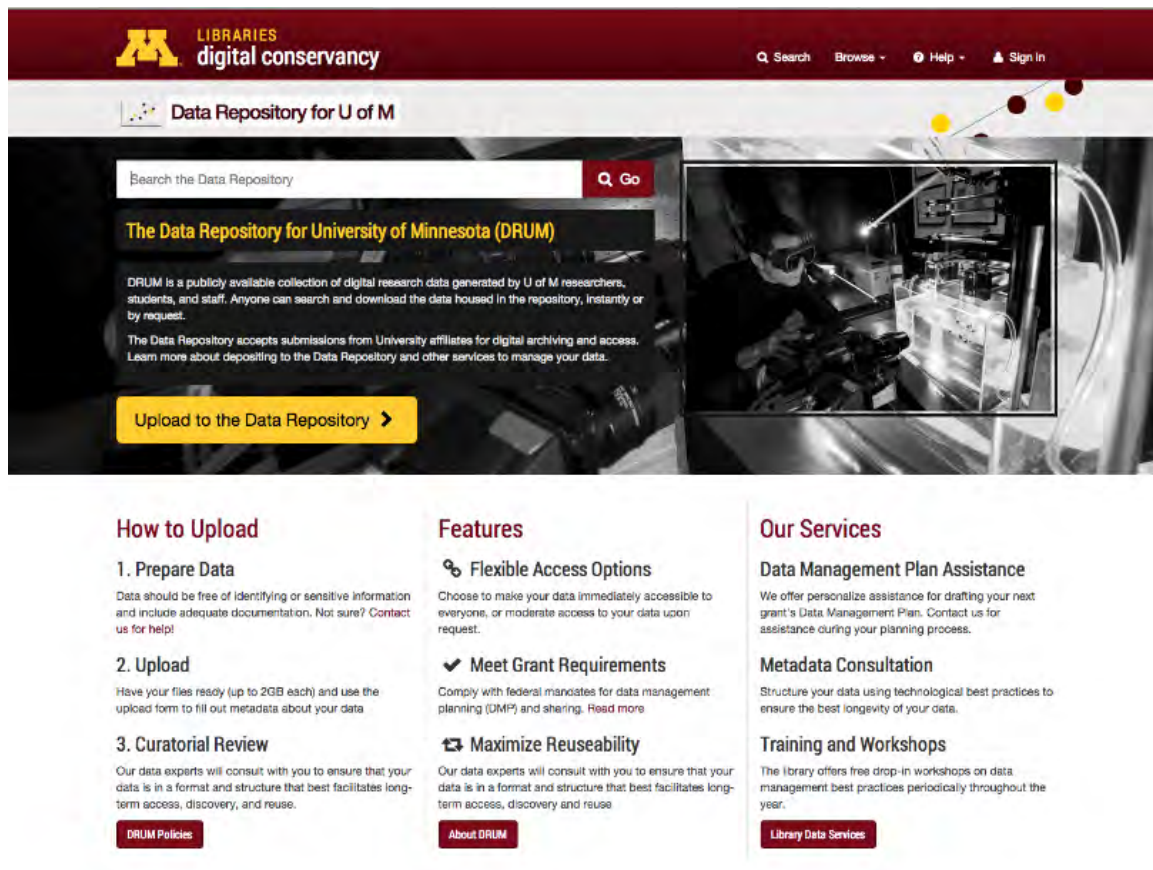
DRUM

<http://z.umn.edu/drum>

Launched Nov 2014

Available to U of M researchers and provides:

- Open access
- Curation services
- Permanent identifiers (DOI)
- Flexible Licenses
- File download analytics
- Preservation



The screenshot shows the DRUM website homepage. At the top, there is a dark red header with the logo for 'LIBRARIES digital conservancy' and navigation links for 'Search', 'Browse', 'Help', and 'Sign In'. Below the header, the main content area features a search bar with the text 'Search the Data Repository' and a 'Go' button. A prominent yellow button labeled 'Upload to the Data Repository' is positioned below the search bar. To the right, there is a large image of a person in a laboratory setting, possibly using a microscope. Below the main content area, there are three columns of text: 'How to Upload' with three numbered steps (1. Prepare Data, 2. Upload, 3. Curatorial Review), 'Features' with three bullet points (Flexible Access Options, Meet Grant Requirements, Maximize Reuseability), and 'Our Services' with three sections (Data Management Plan Assistance, Metadata Consultation, Training and Workshops). Each section includes a brief description and a link to more information.

LIBRARIES digital conservancy

Search Browse Help Sign In

Data Repository for U of M

Search the Data Repository **Go**

The Data Repository for University of Minnesota (DRUM)

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

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

Upload to the Data Repository >

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.
[DRUM Policies](#)

Features

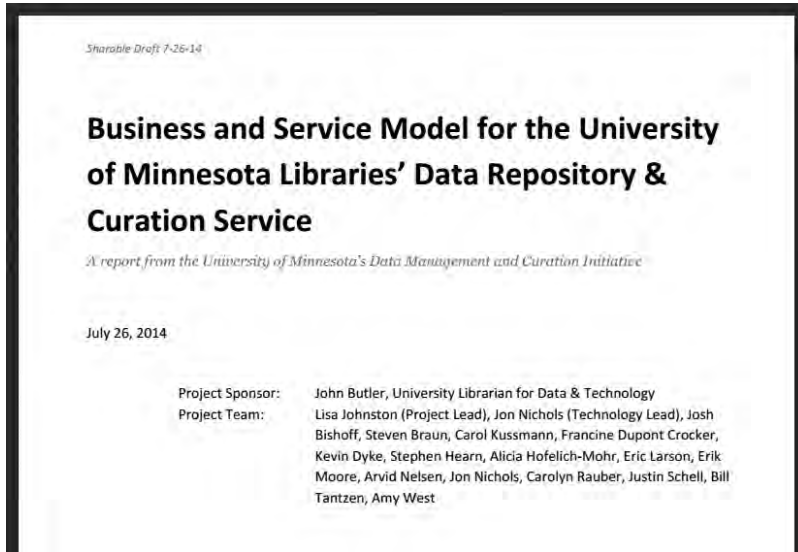
- Flexible Access Options**
Choose to make your data immediately accessible to everyone, or moderate access to your data upon request.
- Meet Grant Requirements**
Comply with federal mandates for data management planning (DMP) and sharing. [Read more](#)
- Maximize Reuseability**
Our data experts will consult with you to ensure that your data is in a format and structure that best facilitates long-term access, discovery and reuse.
[About DRUM](#)

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**
Structure your data using technological best practices to ensure the best longevity of your data.
- Training and Workshops**
The library offers free drop-in workshops on data management best practices periodically throughout the year.
[Library Data Services](#)

How do I build my institutional data repositories?

Organizational Infrastructure



- Libraries DM+C Initiative 2014-2015 with dedicated resources and a 19-person team
 - **Business model**
 - Policies
 - Services
- Existing libraries digital preservation framework
- Campus: New research data management policy
- Marketing: Official launch in March 2015

How do I build my institutional data repositories?

Organizational Infrastructure

DRUM Policies and Terms of Use

Housed within the University of Minnesota Digital Conservancy, the Data Repository for U of M (DRUM) is subject to the [policies of the Digital Conservancy](#) and, in addition, include the following. Note: Policies and Terms are subject to change.

Policies related to DEPOSIT of Materials in DRUM

- [Data Collection Policy](#)
- [End-user Access Policy](#)
- [Preservation Policy](#)
- [Deposit License](#)

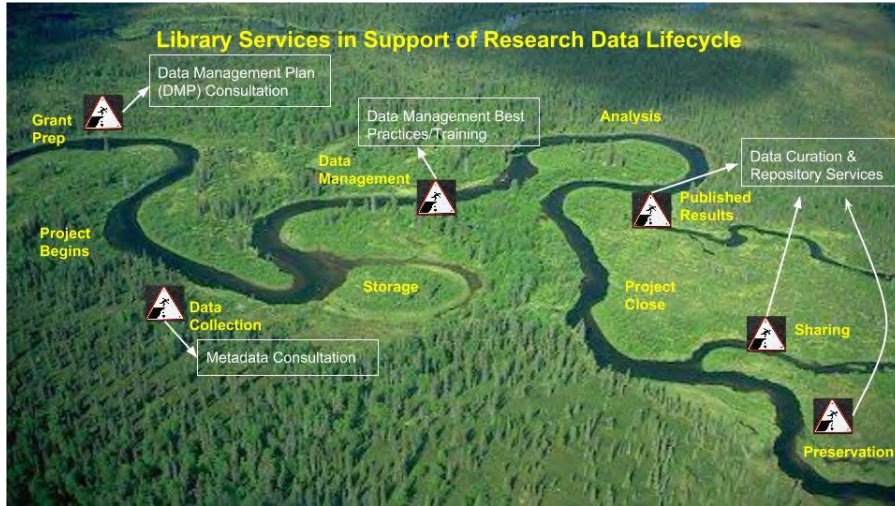
Policies related to USE of Materials Available in DRUM

- [Terms of Use](#)
- [End-user Access Policy](#)

- Libraries DM+C Initiative 2014-2015 with dedicated resources and a 19-person team
 - Business model
 - **Policies**
 - Services
- Existing libraries digital preservation framework
- Campus: New research data management policy
- Marketing: Official launch in March 2015

How do I build my institutional data repositories?

Organizational Infrastructure



- Libraries DM+C Initiative 2014-2015 with dedicated resources and a 19-person team
 - Business model
 - Policies
 - **Services**
- Existing libraries digital preservation framework
- Campus: New research data management policy
- Marketing: Official launch in March 2015

How do I build my institutional data repositories?

Organizational Infrastructure



- Libraries DM+C Initiative 2014-2015 with dedicated resources and a 19-person team
 - Business model
 - Policies
 - Services
- **Existing libraries digital preservation framework**
- Campus: New research data management policy
- Marketing: Official launch in March 2015

How do I build my institutional data repositories?

Organizational Infrastructure



The screenshot shows a webpage titled "ADMINISTRATIVE POLICY" with a sub-header "Research Data Management: Archiving, Ownership, Retention, Security, Storage, and Transfer". Below the title is a yellow box containing a table of contents and metadata. The table of contents lists items such as "Policy Statement", "Reason for Policy", "Procedures", "Forms/Instructions", "Appendices", "FAQ", "Additional Contacts", "Definitions", "Responsibilities", "Related Information", and "History". The metadata includes "Effective Date: Draft, Under review", "Last Updated: Month, YYYY", "Responsible University Officer: Vice President for Research", "Policy Owner: Associate Vice President for Research", and "Policy Contact: Frances Lawrenz". Below the yellow box is a section titled "POLICY STATEMENT" which begins with "The Office of the Vice President for Research (OVPR); the Office of the Vice President and Chief Information Officer (OVPCIO); the Academic Health Center (AHC); the University Libraries and system campus libraries; the colleges, schools and system campuses; and the Principal".

- Libraries DM+C Initiative 2014-2015 with dedicated resources and a 19-person team
 - Business model
 - Policies
 - Services
- Existing libraries digital preservation framework
- **Campus: New research data management policy**
- Marketing: Official launch in March 2015

How do I build my institutional data repositories?

Organizational Infrastructure

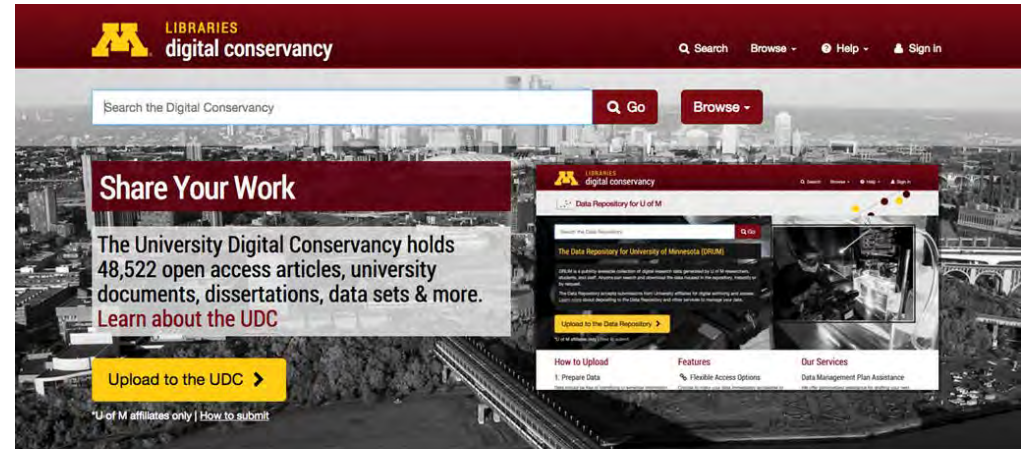


- Libraries DM+C Initiative 2014-2015 with dedicated resources and a 19-person team
 - Business model
 - Policies
 - Services
- Existing libraries digital preservation framework
- Campus: New research data management policy
- **Marketing: Official launch in March 2015**

Case Study: Building the Data Repo for the U of M

Technical Infrastructure

- **DRUM part of existing IR (DSpace):**
 - Metadata schema
 - Collection home page
 - Record view
- Meet federal funding requirements
- Curation procedures
 - Sensitive Data
 - Readme Template
 - Transform File Formats



Features

Free & Open Access

Deposit your material for free. Satisfy requirements for open access to research articles and support the University's Open Access Policy.

Increase Visibility

Receive higher ranking in Google and Google Scholar. Increase your exposure and potential for higher impact.

Data Repository for U of M

Share, publish, and preserve your digital data for long-term access and future use in the Data Repository.

Full-text Search

Search across the full text of your works using Google, Google Scholar, or a Digital Conservancy search box embedded on your own website.

Persistent Access

A persistent url for each work guarantees no more broken links. The UDC ensures your works will remain accessible and usable over time.

Long-term Preservation

Items in the UDC are managed by the U of M Libraries, which is committed to long-term access and preservation of the collection (see our policy).

Case Study: Building the Data Repo for the U of M

Technical Infrastructure

- DRUM part of existing IR (DSpace):
 - **Metadata schema**
 - Collection home page
 - Record view
- Meet federal funding requirements
- Curation procedures
 - Sensitive Data
 - Readme Template
 - Transform File Formats

University Digital Conservancy Item Submission

[Describe](#) [Describe](#) [Upload](#) [CC License](#) [License](#) [Confirm](#) [Complete](#)

Describe Item

Author(s)
Last name, e.g. *Smith* or *University of Minnesota* First, MI e.g. *Barb H.* or *Robert Jr.* [Add More](#)

List the author(s) of the data set. Use complete names eg. "James Ryan Swanson"

Group Author
 [Add More](#)

If applicable for datasets produced by a group. Ex. "University of Minnesota Department of Chemistry, Crystallography Research Lab"

Title (Required)

Titles should be descriptive and include the what, where, when and scale of your data.

Date Completed (Required)
Year (required) Month (optional) Day (optional)

Date the data was finalized for distribution.

Contact (Required)
Last name, e.g. *Smith* or *University of Minnesota* First, MI e.g. *Barb H.* or *Robert Jr.*

Name of the principal contact for the data.

Contact Email (Required)

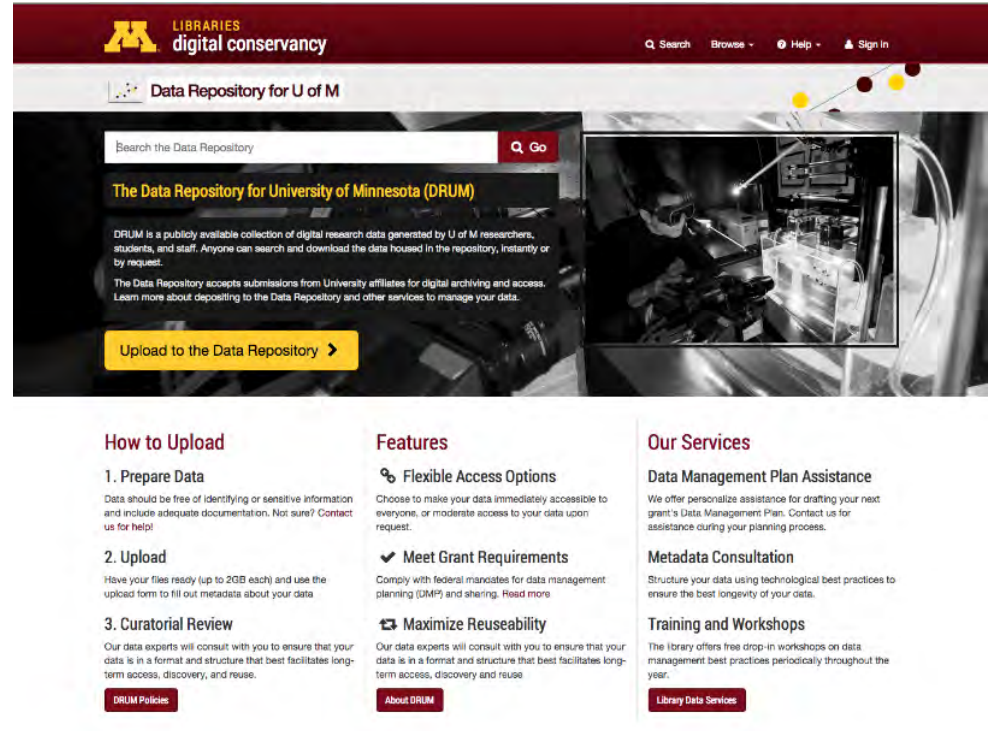
Email address of the principal contact for the data.

[Save & Exit](#) [Next >](#)

Case Study: Building the Data Repo for the U of M

Technical Infrastructure

- DRUM part of existing IR (DSpace):
 - Metadata schema
 - **Collection home page**
 - Record view
- Meet federal funding requirements
- Curation procedures
 - Sensitive Data
 - Readme Template
 - Transform File Formats



The screenshot shows the homepage of the Data Repository for U of M (DRUM). The header features the University of Minnesota Libraries Digital Conservancy logo and navigation links for Search, Browse, Help, and Sign In. The main content area includes a search bar, a description of DRUM as a publicly available collection of digital research data, and a prominent yellow button labeled "Upload to the Data Repository". Below the main content, there are three columns of information: "How to Upload" with three numbered steps (Prepare Data, Upload, Curatorial Review), "Features" with three bullet points (Flexible Access Options, Meet Grant Requirements, Maximize Reuseability), and "Our Services" with three sections (Data Management Plan Assistance, Metadata Consultation, Training and Workshops). Each section has a corresponding "About DRUM" or "Library Data Services" button.

Schema published: "The Supporting Documentation for Implementing the Data Repository for the University of Minnesota (DRUM): A Business Model, Functional Requirements, and Metadata Schema" at <http://hdl.handle.net/11299/171761>.

Case Study: Building the Data Repo for the U of M

Technical Infrastructure

- DRUM part of existing IR (DSpace):
 - Metadata schema
 - Collection home page
 - **Record view**
- Meet federal funding requirements
- Curation procedures
 - Sensitive Data
 - Readme Template
 - Transform File Formats

The screenshot shows the DRUM record view page for a dataset. The header includes the University of Minnesota Libraries Digital Conservancy logo and navigation links. The main content area displays the title 'White-tailed deer density estimates across the eastern United States, 2008' by Walters, Brian, F. Woodall, Christopher, W. Russell, Matthew, B. (2016). A map of the eastern United States is shown, color-coded by deer density. The page includes metadata such as 'Published Date', 'Author Contact', 'Data Type', and an 'Abstract'. A 'Suggested Citation' is provided, along with a 'Persistent link to this item' and 'Services' like 'Full Metadata (XML)' and 'View Usage Statistics'. A table at the bottom lists files for download, including 'deer_density_QDMA.zip' and 'README.txt'.

File View/Open	Description	Size	Format
deer_density_QDMA.zip	GIS data-single shapefile	3.792Mb	application/zip
README.txt	README text file with basic description of data	1.287Kb	text/plain

Case Study: Building the Data Repo for the U of M

Technical Infrastructure

- DRUM part of existing IR (DSpace):
 - Metadata schema
 - Collection home page
 - Record view
- **Meet federal funding requirements**
- Curation procedures
 - Sensitive Data
 - Readme Template
 - Transform File Formats

Resources and Examples

The Libraries have organized the following resources that may be useful for you as you write your DMP:

- [UMN DMP template \(google doc\): a template you can use to begin writing your DMP](#)
- [Data Management Checklist](#) (two page pdf)
- Boilerplate text that you can use in your DMP and project proposal:

A long-term data sharing and preservation plan will be used to store and make publicly accessible the data beyond the life of the project. The data will be deposited into the Data Repository for the University of Minnesota (DRUM), <http://hdl.handle.net/11299/166578>. The University Libraries' hosted institutional data repository is an open access platform for dissemination and archiving of university research data. Data files in DRUM are written to an isilon storage system with two copies, one local to each of the two geographically separated University of Minnesota Data Centers. The local Isilon cluster stores the data in such a way that the data can survive the loss of any two disks or any one node of the cluster. Within two hours of the initial write, data replication to the 2nd Isilon cluster commences. The 2nd cluster employs the same protections as the local cluster, and both verify with a checksum procedure that data has not altered on write. In addition, DRUM provides long-term preservation of digital data files for at least 10 years using services such as migration (limited format types), secure backup, bit-level checksums, and maintains a persistent DOIs for data sets, facilitating data citations. In accordance to DRUM policies, the (deidentified, if applicable) data will be accompanied by the appropriate documentation, metadata, and code to facilitate reuse and provide the potential for interoperability with similar data sets.

Case Study: Building the Data Repo for the U of M

Technical Infrastructure

- DRUM part of existing IR (DSpace):
 - Metadata schema
 - Collection home page
 - Record view
- Meet federal funding requirements
- **Curation procedures**
 - Sensitive Data
 - Readme Template
 - Transform File Formats

Submission
under
curatorial
review

Case Study: Building the Data Repo for the U of M

Technical Infrastructure

- DRUM part of existing IR (DSpace):
 - Metadata schema
 - Collection home page
 - Record view
- Meet federal funding requirements
- Curation procedures
 - **Sensitive Data**
 - Readme Template
 - Transform File Formats

The image shows a 'Wizard' window for identifying sensitive data in DRUM. The window is titled 'Wizard' and has a 'Confirmation' section. Below the confirmation text, there are two sections: 'Identities' and 'Locations'. Each section contains a list of data types with checkmarks indicating their status.

Confirmation
Please review your selections below and click Finish to begin your search. To make changes, click Back.

Identities

<input checked="" type="checkbox"/> Social Security Numbers	<input checked="" type="checkbox"/> Dates of Birth	<input checked="" type="checkbox"/> Australia: TFN
<input checked="" type="checkbox"/> Credit Card Numbers	<input checked="" type="checkbox"/> Phone Numbers	<input checked="" type="checkbox"/> Canada: SIN
<input checked="" type="checkbox"/> Password Entries	<input checked="" type="checkbox"/> E-Mail Addresses	<input checked="" type="checkbox"/> United Kingdom: NINO
<input checked="" type="checkbox"/> Bank Account Numbers	<input checked="" type="checkbox"/> Personal Addresses	<input checked="" type="checkbox"/> United Kingdom: NHS
<input checked="" type="checkbox"/> Driver Licenses	<input checked="" type="checkbox"/> Passport Numbers	

Locations

<input checked="" type="checkbox"/> E-Mails	<input checked="" type="checkbox"/> Browsers	<input checked="" type="checkbox"/> Windows Registry
<input checked="" type="checkbox"/> Files and Compressed Files	<input checked="" type="checkbox"/> My Documents and Settings	

Buttons: < Back, Finish, Cancel, Help

The background document is titled 'Procedure for Identifying Sensitive, Private, or Legally Protected Data in DRUM' and is from the University of Minnesota. It includes a 'Purpose' section stating that DRUM has clear policies on what types of data may be accepted, and an 'Evaluation and Review of DRUM Submissions' section with two steps: Step 1: U of M Researcher submits their data to DRUM at z.umn.edu/drum; Step 2: Data curators in the library receive notification that a new submission has occurred, but has not been reviewed.

Case Study: Building the Data Repo for the U of M

Technical Infrastructure

- DRUM part of existing IR (DSpace):
 - Metadata schema
 - Collection home page
 - Record view
- Meet federal funding requirements
- Curation procedures
 - Sensitive Data
 - **Readme Template**
 - Transform File Formats

```
This readme.txt file was generated on <YYYYMMDD> by <Name>
```

GENERAL INFORMATION

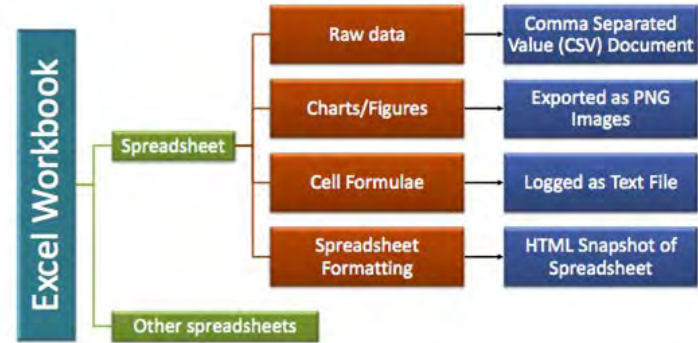
1. Title of Dataset:
2. File Information:
 - A. Filename:
 - B. Short description:
 - C. Filename:
 - D. Short description
 - E. Filename:
 - F. Short description:
 - G. If data set includes multiple files related to one another, include relationship here:
3. Principal Investigator Contact Information
 - A. Name:
 - B. Institution:
 - C. Address:
 - D. Email:
4. Associate or Co-investigator Contact Information
 - A. Name:
 - B. Institution:
 - C. Address:
 - D. Email:

Case Study: Building the Data Repo for the U of M

Technical Infrastructure

- DRUM part of existing IR (DSpace):
 - Metadata schema
 - Collection home page
 - Record view
- Meet federal funding requirements
- Curation procedures
 - Sensitive Data
 - Readme Template
 - **Transform File Formats**

Excel Archival Tool



File View/Open

Data For Gamma-toxin.xlsx

Archived_Data_for_Gamma_Toxin.zip

Combined Figures.pzf

Combined_Figures.xml

Description

Experiment Data Readings

Archival Version of Data for Gamma-toxin.xlsx

Data Analysis and Figures in Prism

Archival Version of Combined Figures.pzf

Case Study: Building the Data Repo for the U of M

Resources Infrastructure



- Funding for DataCite DOIs
- Staffing Model
- Training for new curation staff

Web of Science™ InCites™ Journal Citation Reports® Essential Science Indicators™ EndNote™

WEB OF SCIENCE™

Search Return to Search Results

Save to EndNote online Add to Marked List

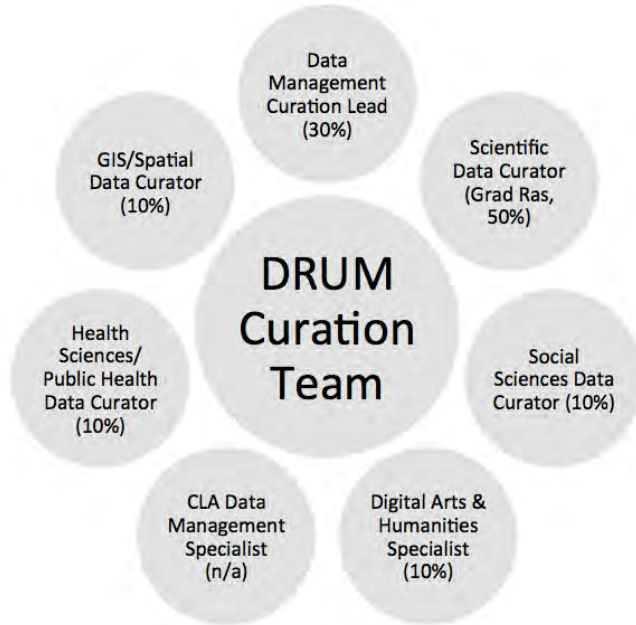
Data and visualizations of air quality impacts of conventional and alternative light-duty transportation in the United States

From Repository: Data Repository for the University of Minnesota
By: Tessum, Christopher W; Hill, Jason D; Marshall, Julian D

Data Repository for the University of Minnesota
DOI: <http://dx.doi.org/10.13020/D6159V>
Viewed Date: 05 Jan 2015
Published: 2014

Case Study: Building the Data Repo for the U of M

Resources Infrastructure



- Funding for DataCite DOIs
- **Staffing Model**
- Training for new curation staff

Case Study: Building the Data Repo for the U of M

Resources Infrastructure

DMCI Scientific Data Curator Training Guide 2015

University Libraries Data Management and Curation Initiative (DMCI)

Please connect to and familiarize yourself with the [DMCI Staff Handbook](#) (2014-) where we will keep our shared documents. These include:

- DMCI Data Curation Procedures (Curator Handbook)
- Data Curators' Position Description
- This document

Table of Contents

[1.0 Responsibilities](#)

[1.1 Duties of the Scientific Data Curator](#)

[1.2 Daily Tasks of the Scientific Data Curator](#)

[1.3 Regular Meeting to Attend as the Scientific Data Curator](#)

[2.0 Research](#)

18 Training Activities

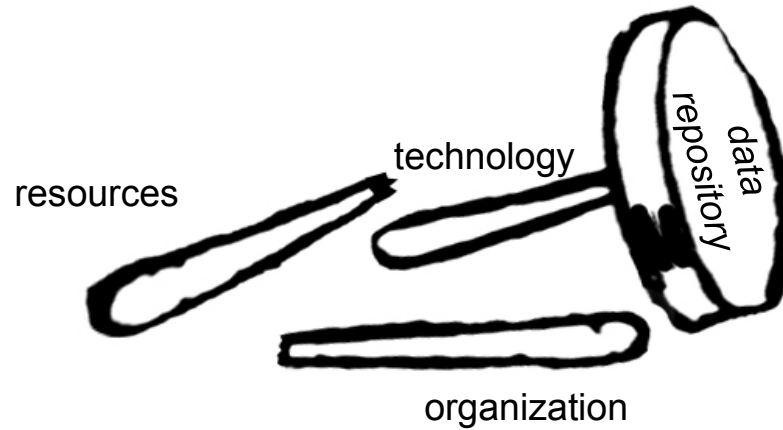
- Funding for DataCite DOIs
- Staffing Model
- Training for new curation staff

Conclusions



- Greater exposure of library services on campus
- Data repositories open up new conversations and opportunities (e.g., big data, campus networks, policies)
- Rewarding work to partner with researchers to publish their data sets.

Thanks and Questions



Keep building those stools...