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

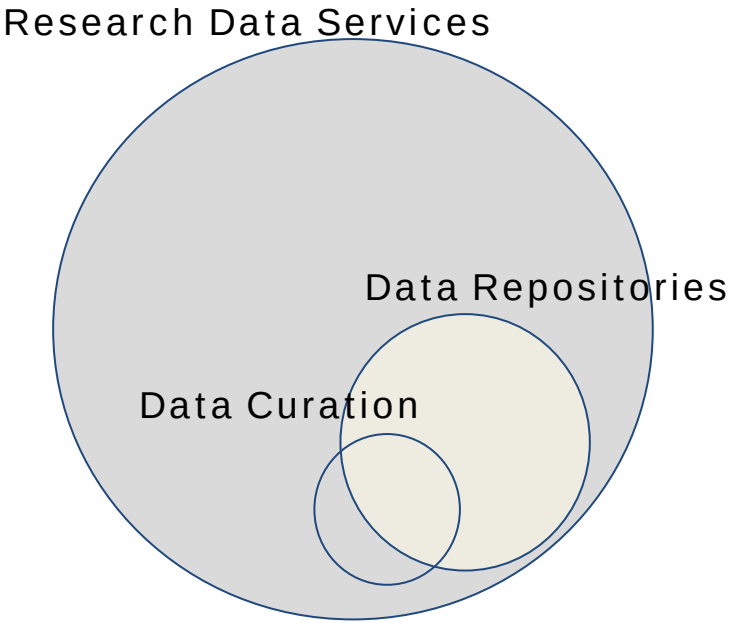
The answer that actually came with a question

Lisa Johnston

Research Data Management/Curation Lead,
University of Minnesota

LIBRARIES
UNIVERSITY OF MINNESOTA

What are data repositories?



"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)."

Citation: The RDA Data Foundations and Terminology working group has a growing dictionary of data related terms that is searchable at http://smw-rda.esc.rzg.mpg.de/index.php/Main_Page

Why are data repositories useful?

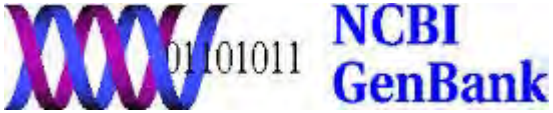
- Governments have open data initiatives
  - Democratize research results
  - Release the potential of valuable/$$ data
- Federal/private funders increasingly require data sharing
  - Public access
  - Return on $$ investment ⇒ do new research
- Journals have data sharing policies
  - Increase transparency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

Citation: These ideas expanded in Johnston, Lisa R. (in press). “Introduction.” Curating Research Data Volume 1: Practical Strategies for Your Digital Repository. ACRL Press.

LIBRARIES
UNIVERSITY OF MINNESOTA

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

figshare
credit for all your research

zenodo
Research. Shared.

MENDELEY DATA Beta

The Dataverse Project

OPEN SCIENCE FRAMEWORK

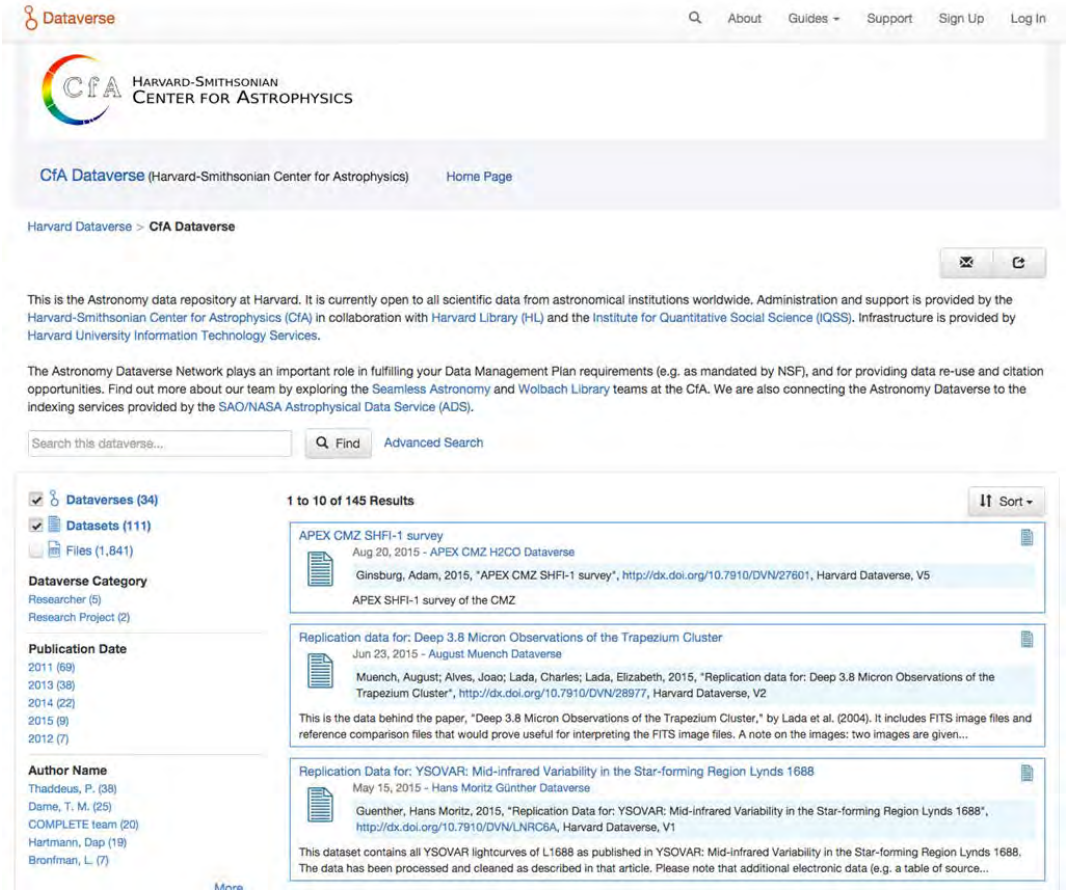
http://openICPSR.org
open ICPSR

What kinds of data repositories exist?

Institutional Data Repositories

Running...

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

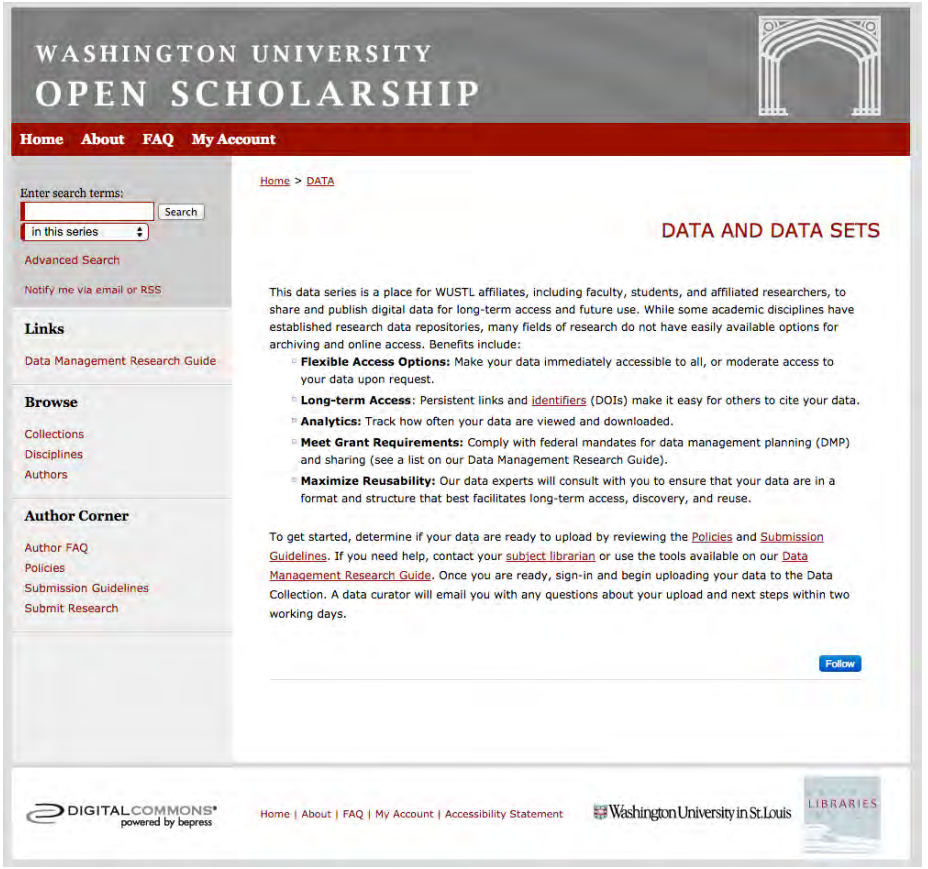


What kinds of data repositories exist?

Institutional Data Repositories

Running...

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

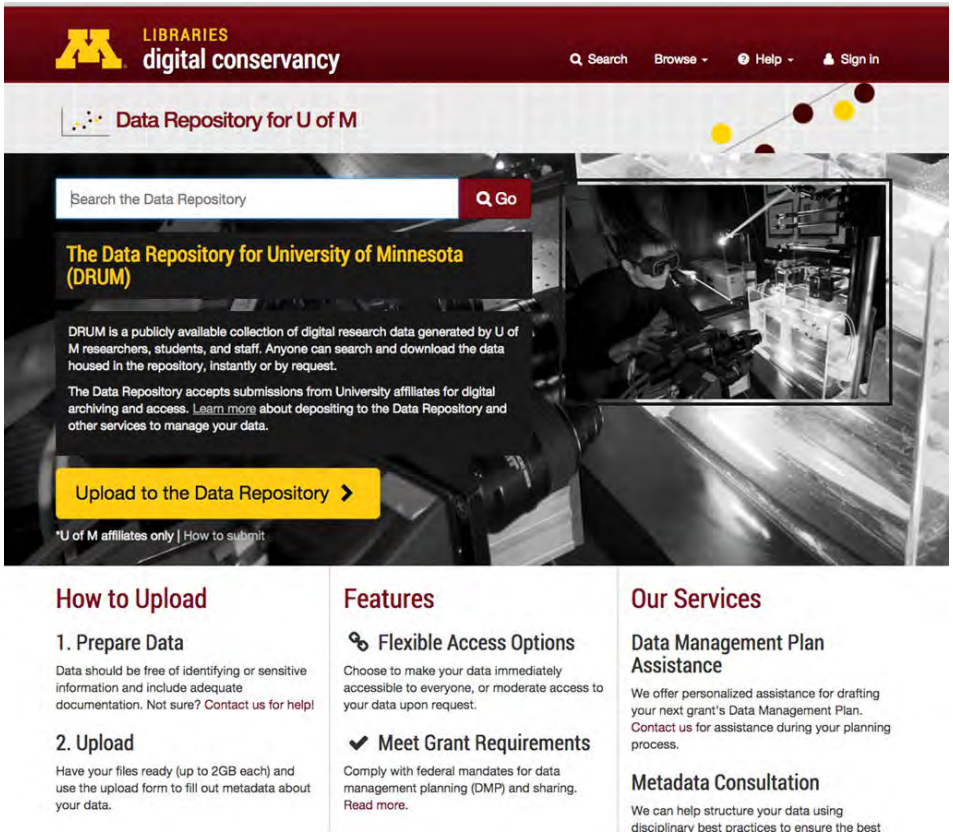


What kinds of data repositories exist?

Institutional Data Repositories

Running...

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



What kinds of data repositories exist?

Institutional Data Repositories

Running...

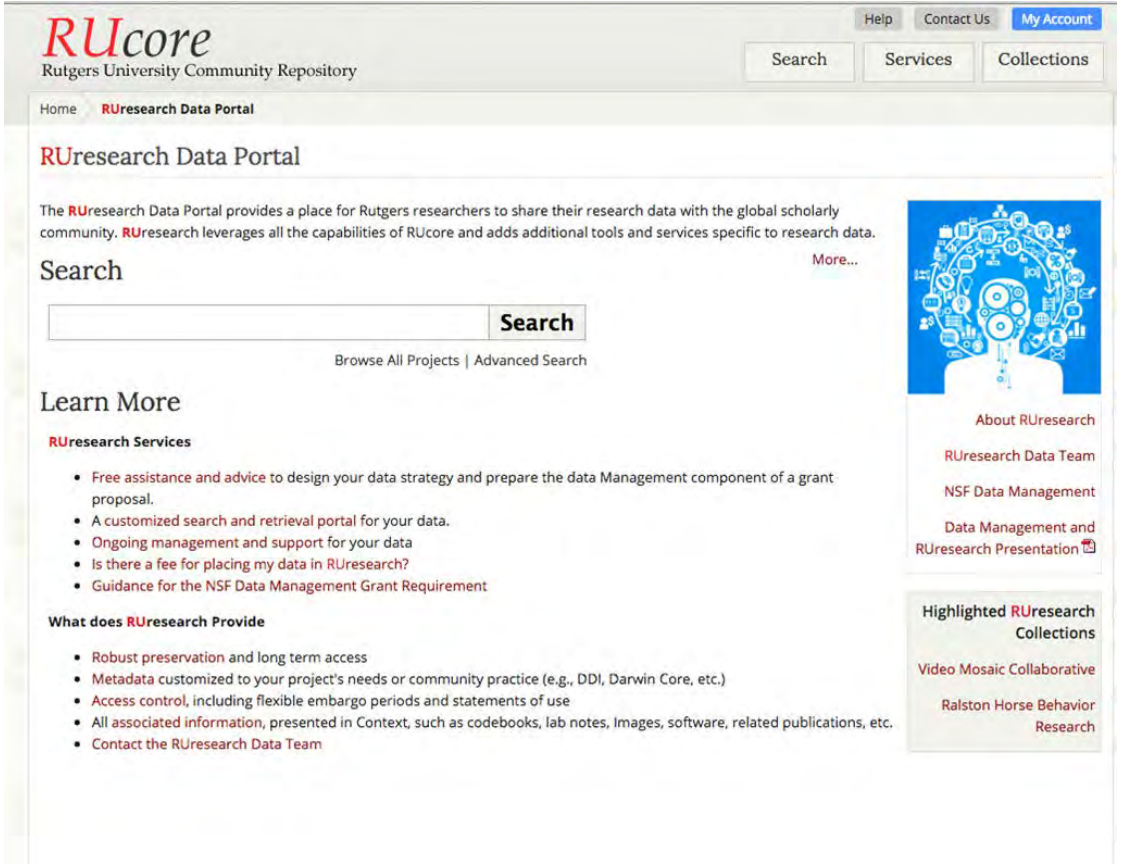
- Dataverse
- Digital Commons
- DSpace
- **EPrints**
- Fedora
- HubZero
- Hydra
- Islandora

What kinds of data repositories exist?

Institutional Data Repositories

Running...

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



What kinds of data repositories exist?

Institutional Data Repositories

Running...

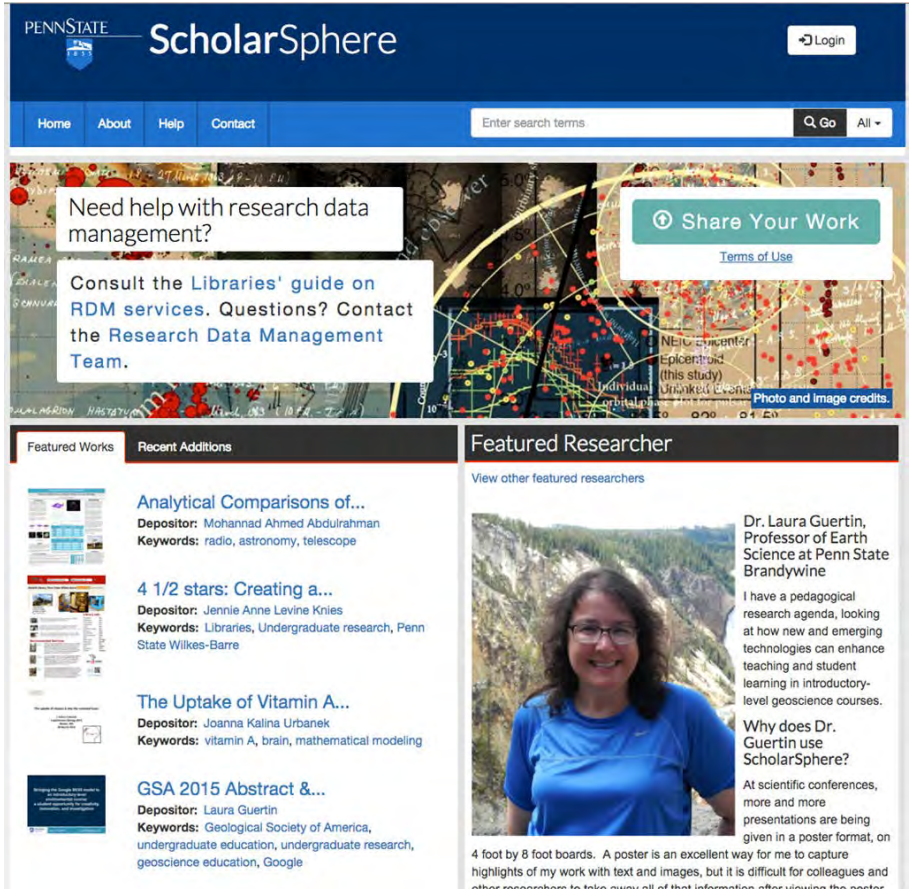
- Dataverse
- Digital Commons
- DSpace
- EPrints
- Fedora
- **HubZero**
- Hydra
- Islandora

What kinds of data repositories exist?

Institutional Data Repositories

Running...

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



What kinds of data repositories exist?

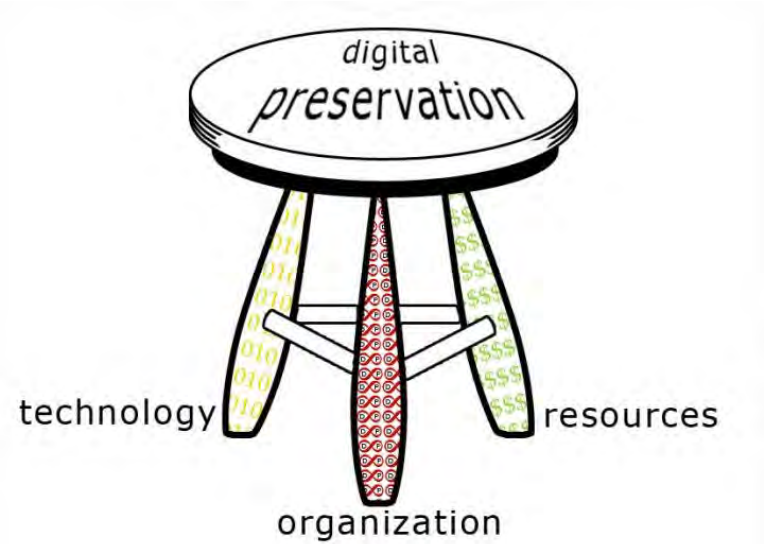
Institutional Data Repositories

Running...

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

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…

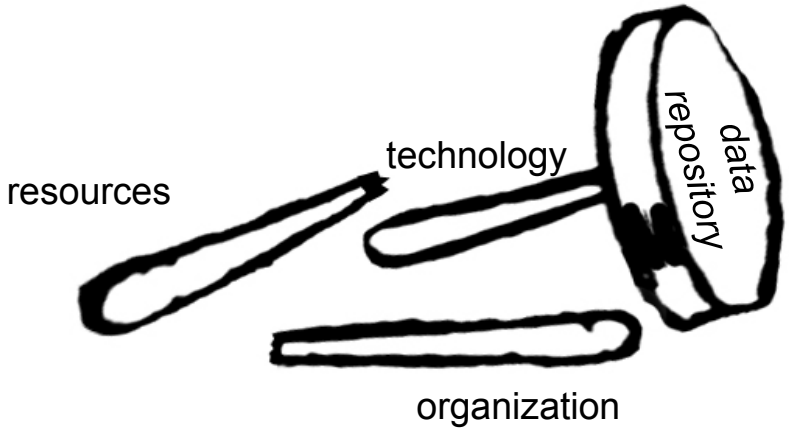
- technological infrastructure
- organizational infrastructure
- resources framework

Citation: http://www.iassistdata.org/downloads/2006/f2_mcgovern.pdf

LIBRARIES
UNIVERSITY OF MINNESOTA

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

Image: https://placeshakers.files.wordpress.com/2010/08/stool-parody.jpg

LIBRARIES
UNIVERSITY OF MINNESOTA

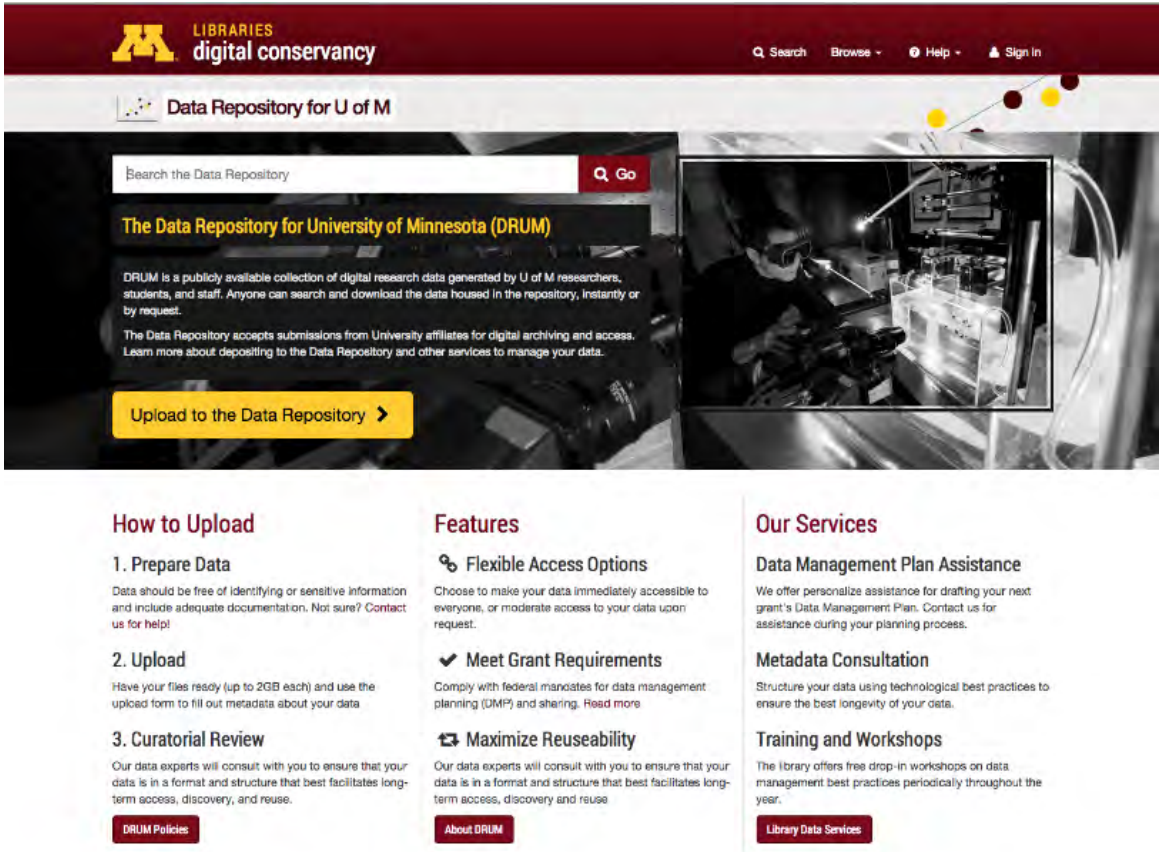
DRUM

http://z.umn.edu/drum

Launched Nov 2014

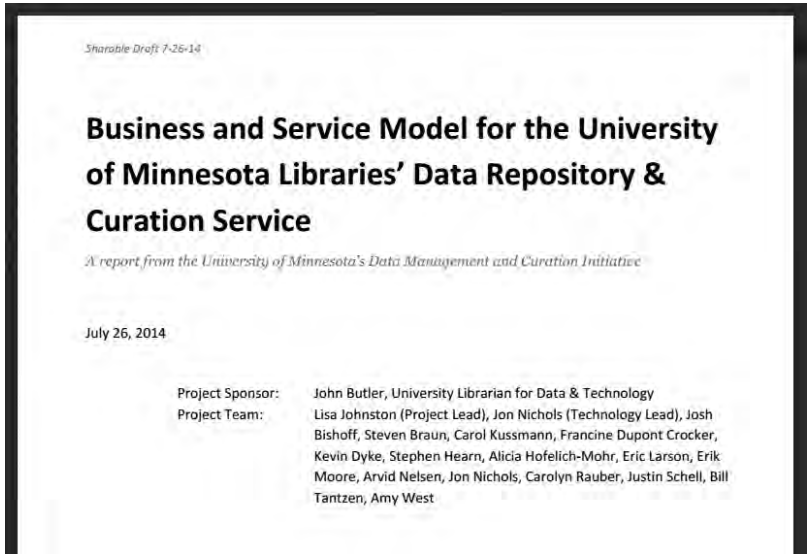
Available to U of M researchers and provides:

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



How do I build my institutional data repositories?

Organizational Infrastructure



Sharable Draft 7-26-14

Business and Service Model for the University of Minnesota Libraries' Data Repository & Curation Service

A report from the University of Minnesota's Data Management and Curation Initiative

July 26, 2014

Project Sponsor: John Butler, University Librarian for Data & Technology
Project Team: Lisa Johnston (Project Lead), Jon Nichols (Technology Lead), Josh Bishoff, Steven Braun, Carol Kussmann, Francine Dupont Crocker, Kevin Dyke, Stephen Hearn, Alicia Hofelich-Mohr, Eric Larson, Erik Moore, Arvid Nelsen, Jon Nichols, Carolyn Rauber, Justin Schell, Bill Tantzen, Amy West

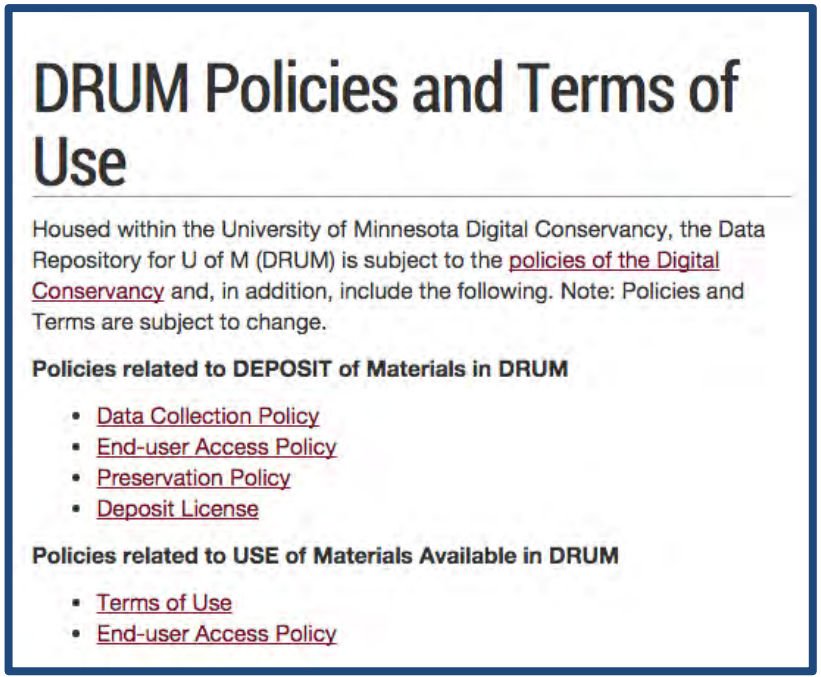
- Libraries DM+C Initiative 2014-2015 with dedicated resources and a 19-person team
  - **Business model**
  - Policies
  - Services
- Existing libraries digital preservation framework
- Campus: New research data management policy
- Marketing: Official launch in March 2015

Citation: "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.

LIBRARIES
UNIVERSITY OF MINNESOTA

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

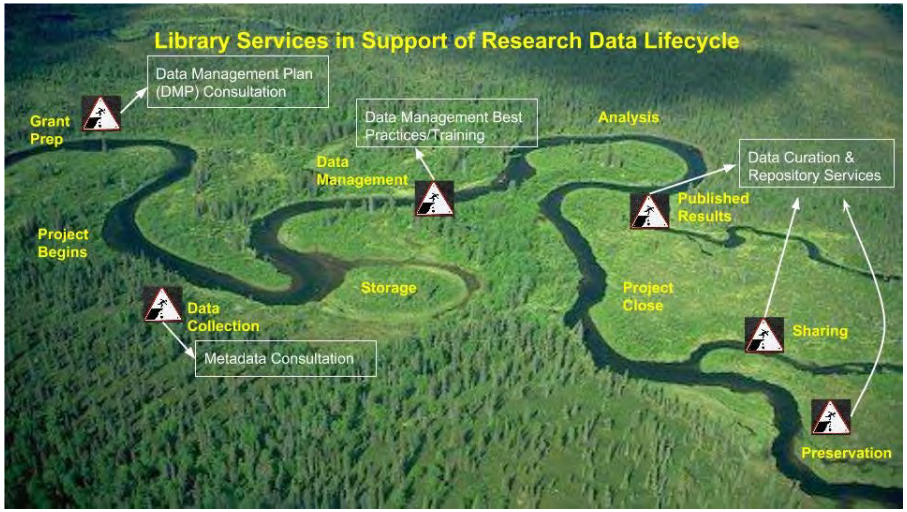
- Libraries DM+C Initiative 2014-2015 with dedicated resources and a 19-person team
  - Business model
  - **Policies**
  - Services
- Existing libraries digital preservation framework
- Campus: New research data management policy
- Marketing: Official launch in March 2015

Citation: "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.

LIBRARIES
UNIVERSITY OF MINNESOTA

How do I build my institutional data repositories?

Organizational Infrastructure



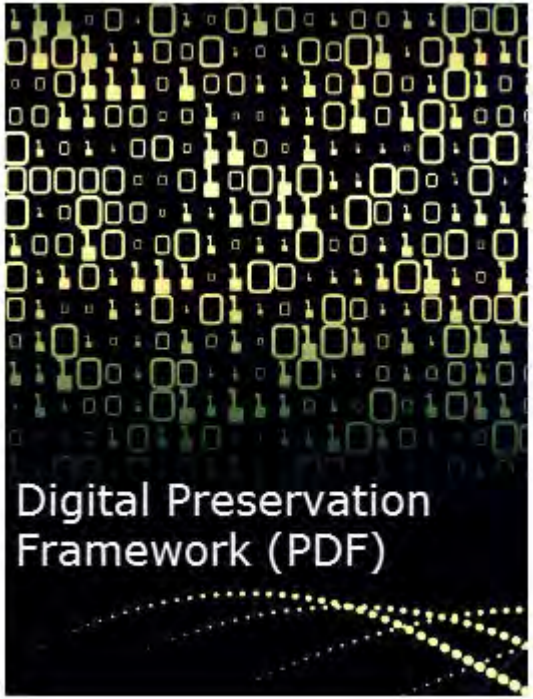
- Libraries DM+C Initiative 2014-2015 with dedicated resources and a 19-person team
  - Business model
  - Policies
  - **Services**
- Existing libraries digital preservation framework
- Campus: New research data management policy
- Marketing: Official launch in March 2015

Citation: “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.

LIBRARIES
UNIVERSITY OF MINNESOTA

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

Citation: https://www.lib.umn.edu/dp/digital-preservation-framework

LIBRARIES
UNIVERSITY OF MINNESOTA

How do I build my institutional data repositories?

Organizational Infrastructure

ADMINISTRATIVE POLICY

Research Data Management: Archiving, Ownership, Retention, Security, Storage, and Transfer

Policy Contents

- Policy Statement
- Reason for Policy
- Procedures
- Forms/Instructions
- Appendices
- FAQ
- Additional Contacts
- Definitions
- Responsibilities
- Related Information
- History

Effective Date: Draft Under review

Last Updated: Month, YYYY

Responsible University Officer:

- Vice President for Research

Policy Owner:

- Associate Vice President for Research

Policy Contact:

- Frances Lawrenz

POLICY STATEMENT

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

Citation: https://policy.umn.edu/research/researchdata

LIBRARIES
UNIVERSITY OF MINNESOTA

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**

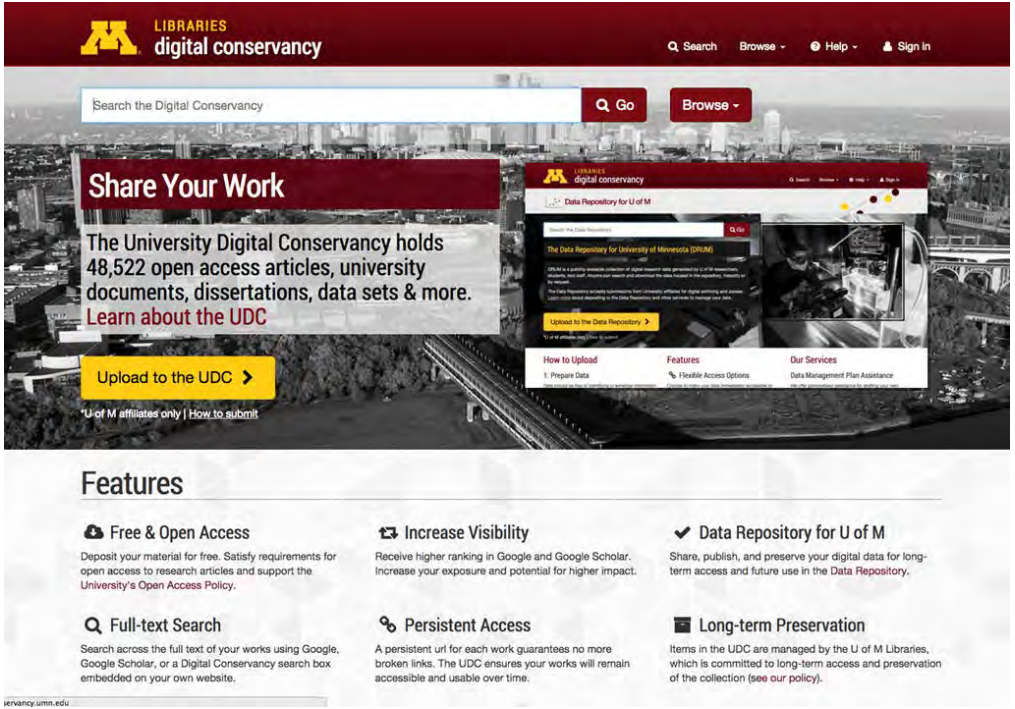
Image: http://continuum.umn.edu/email/2015/drum/

LIBRARIES
UNIVERSITY OF MINNESOTA

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



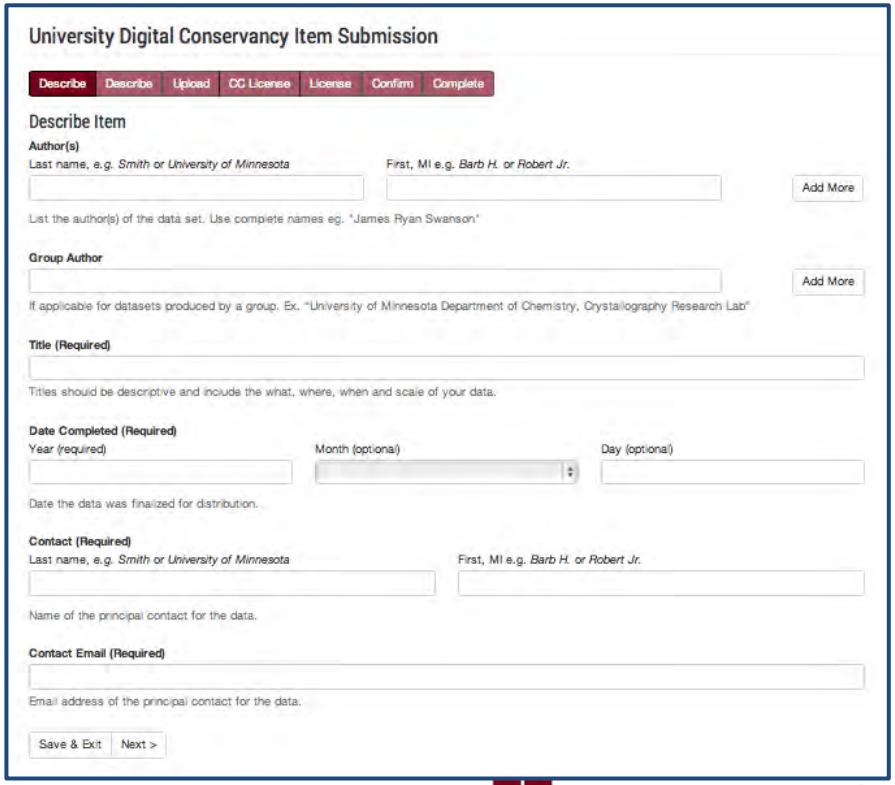
IR: https://conservancy.umn.edu/

LIBRARIES
UNIVERSITY OF MINNESOTA

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



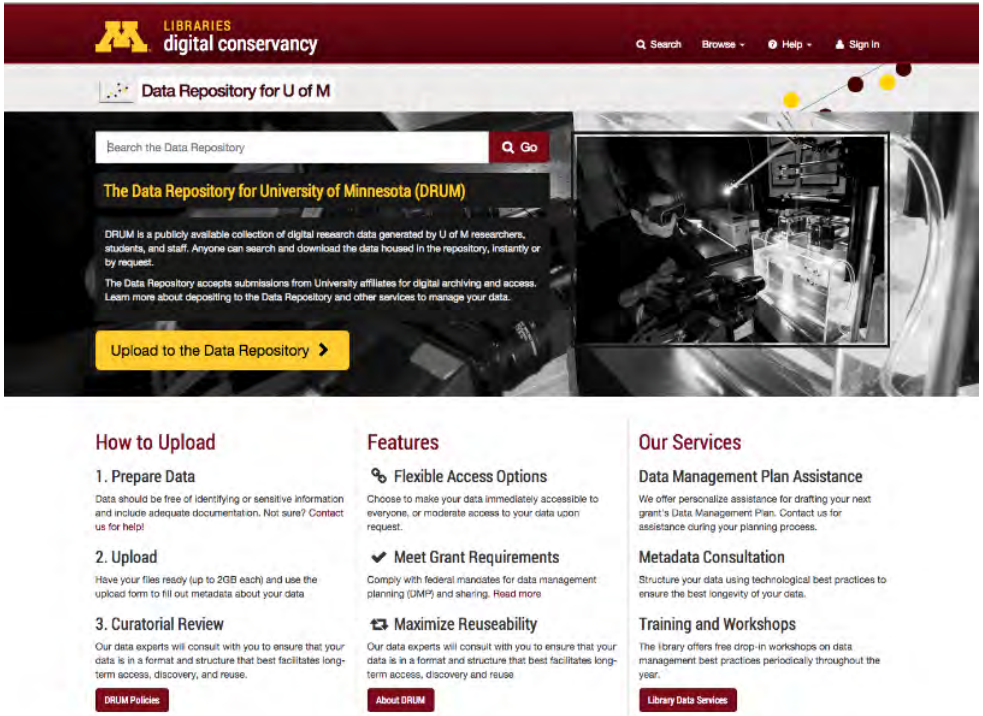
Schema published: "The Supporting Documentation for Implementing the Data Repository for the University of Minnesota (DRUM): A Business Model, Functional Requirements, and Metadata Schema" at http://hdl.handle.net/11299/171761.

LIBRARIES
UNIVERSITY OF MINNESOTA

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



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.

LIBRARIES
UNIVERSITY OF MINNESOTA

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

Technical Infrastructure

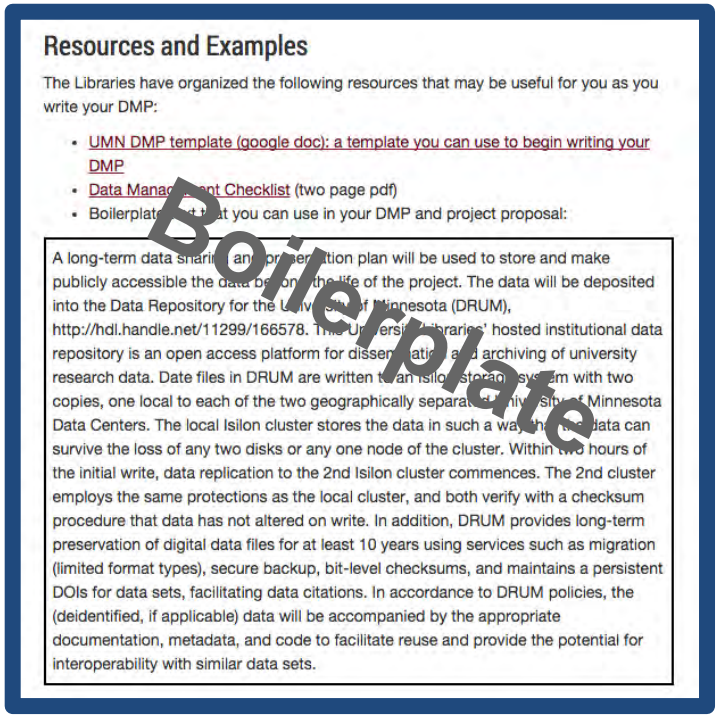
- DRUM part of existing IR (DSpace):
  - Metadata schema
  - Collection home page
  - **Record view**
- Meet federal funding requirements
- Curation procedures
  - Sensitive Data
  - Readme Template
  - Transform File Formats

LIBRARIES
UNIVERSITY OF MINNESOTA

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



IR: https://conservancy.umn.edu/

LIBRARIES
UNIVERSITY OF MINNESOTA

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

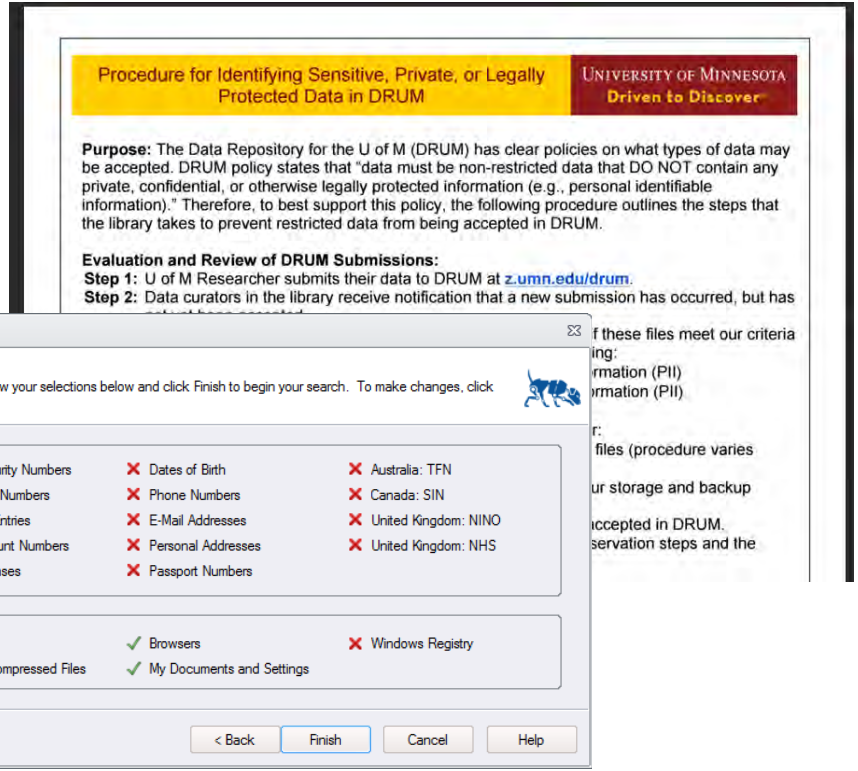
IR: https://conservancy.umn.edu/

LIBRARIES
UNIVERSITY OF MINNESOTA

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



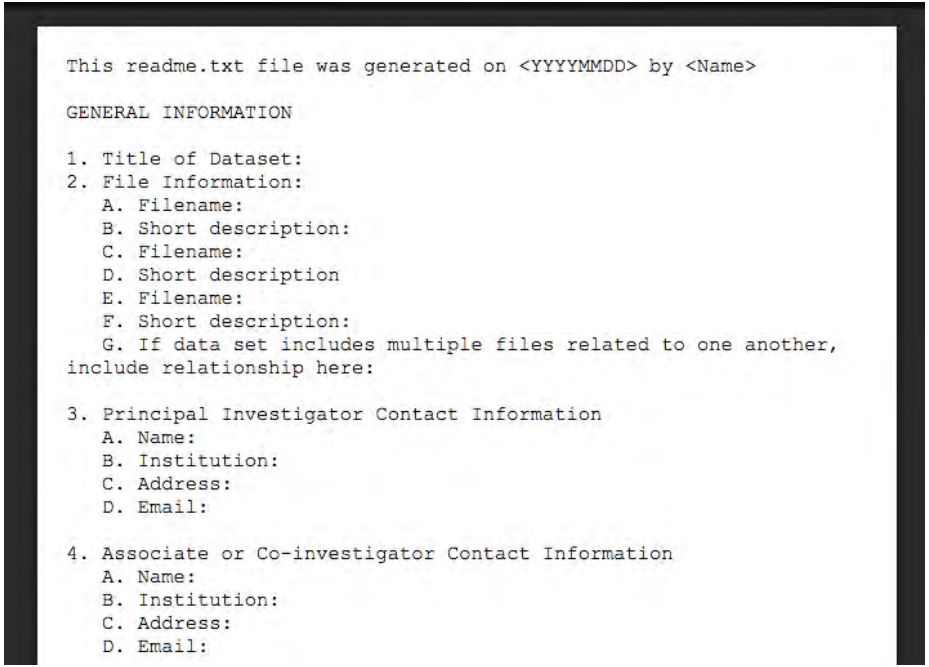
Handout: https://www.lib.umn.edu/datamanagement/sensitive

LIBRARIES
University of Minnesota

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:
```

Download: z.umn.edu/readme

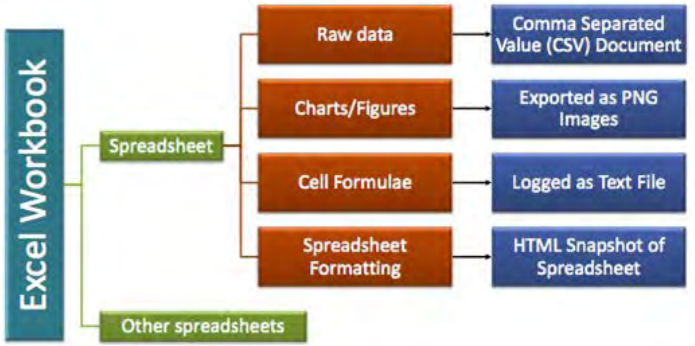
LIBRARIES
UNIVERSITY OF MINNESOTA

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	Description
Data For Gamma-toxin.xlsx	Experiment Data Readings
Archived_Data_for_Gamma_Toxin.zip	Archival Version of Data for Gamma-toxin.xlsx
Combined Figures.pzf	Data Analysis and Figures in Prism
Combined_Figures.xml	Archival Version of Combined Figures.pzf

Download: Excel Archival Tool Software (Github download) http://z.umn.edu/exceltool

LIBRARIES
UNIVERSITY OF MINNESOTA

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

Resources Infrastructure

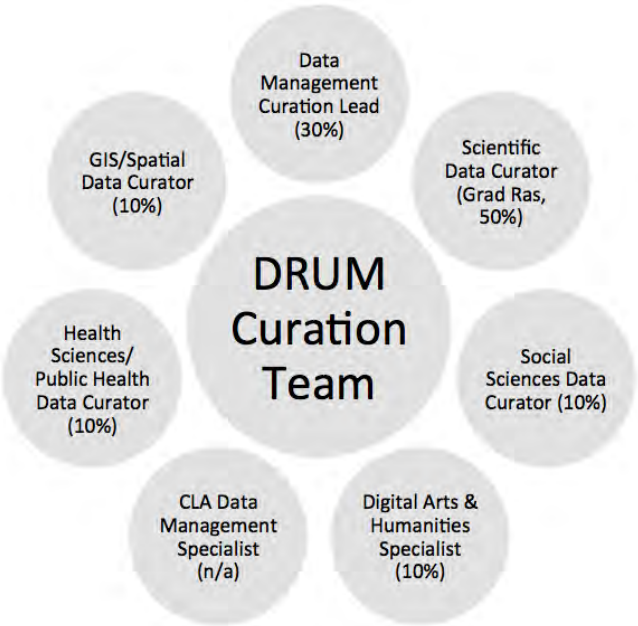
DataCite
FIND, ACCESS, AND REUSE DATA

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

LIBRARIES
UNIVERSITY OF MINNESOTA

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

Resources Infrastructure



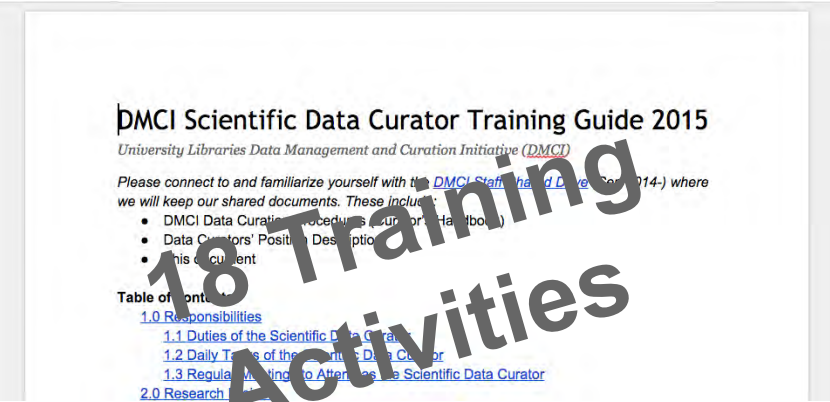
- Funding for DataCite DOIs
- **Staffing Model**
- Training for new curation staff

Staffing model 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.

LIBRARIES
UNIVERSITY OF MINNESOTA

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

Resources Infrastructure



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

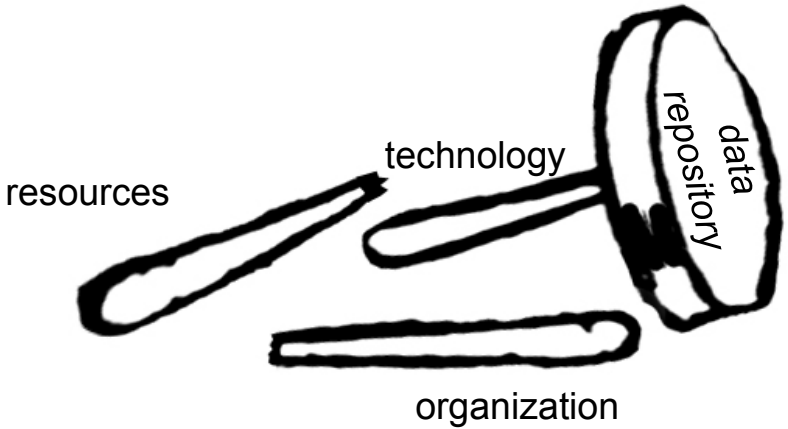
LIBRARIES
UNIVERSITY OF MINNESOTA

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.

LIBRARIES
UNIVERSITY OF MINNESOTA

Thanks and Questions



Keep building those stools...

LIBRARIES
UNIVERSITY OF MINNESOTA