

NISO Webinar: Research Data Curation Part 2: E-Science Librarianship

September 18, 2013

Speakers:

Lisa Johnston


Research Services Librarian, Co-Director of the University Digital Conservancy,
University of Minnesota Libraries

Sayeed Choudhury

Associate Dean for Research Data Management,
Sheridan Libraries of Johns Hopkins University

Carly Strasser

Data Curation Project Manager,
UC Curation Center (UC3), California Digital Library



Research is digital, and this presents some challenges...





Research data can also be big and hard to manage...



UNIVERSITY OF MINNESOTA
Driven to DiscoverSM

Federal agencies must improve public access to digital research data (OSTP Memo, Feb 22, 2013).





Preservation, organization, and public access?
Sounds like a job for the libraries!

Google | google.com/datacenters



UNIVERSITY OF MINNESOTA
Driven to DiscoverSM



Challenge: How to create sustainable and
scaleable *data curation programs* on campus.



UNIVERSITY OF MINNESOTA
Driven to DiscoverSM



1. Helping researchers develop better data management skills.



UNIVERSITY OF MINNESOTA
Driven to DiscoverSM

Data Management Training Initiatives by the University of Minnesota Libraries (2010 - present)

Visits to
<http://lib.umn.edu/datamanagement>

47,862

Faculty and Staff attendees to drop-in
Workshop “Creating a Data
Management Plan” (RCR CE credit)

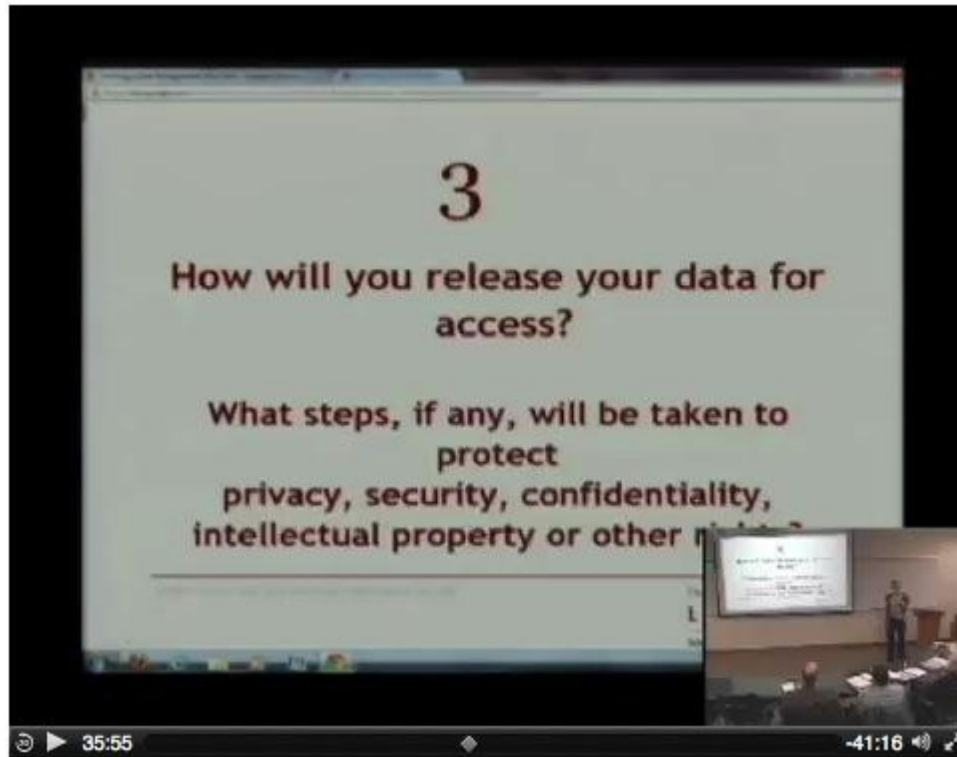
360

Graduate students enrolled in open
online “Data Management Course”
(non-credit Fall 2012 and Spring 2013)

58



Interactive Workshops Facilitate Discussion



["Training Researchers on Data Management: A Scalable, Cross-Disciplinary Approach,"](#) Available in the Journal of eScience Librarianship (Vol. 1: Iss. 2)



UNIVERSITY OF MINNESOTA
Driven to DiscoverSM

Data Information Literacy: IMLS-funded project 2011-14



In-Depth Interviews

- 90-120 minutes
- 4 graduate students
- 1 faculty member
- Interview tools:
z.umn.edu/dil



Johnston, L. and Jeffryes, J. (2013). "Data Management Skills Needed by Structural Engineering Students: A Case Study at the University of Minnesota." J. Prof. Issues Eng. Educ. Pract., DOI:10.1061/(ASCE)EI.1943-5541.0000154 (Feb. 13, 2013).



UNIVERSITY OF MINNESOTA
Driven to DiscoverSM

Online Course for Graduate Students reaches across campus

 UNIVERSITY OF MINNESOTA
Driven to Discover™

Search this site

Data Management Course

Civil Engineering Section

Modules

1. Introduction to Data Management
2. Data to be Managed
3. Organization and Documentation
4. Data Access and Ownership
5. Data Sharing and Re-use
6. Preservation Techniques
7. Complete Your DMP

Resources

Data Analysis

UMN Services

About

Instructors

Syllabus

News & Updates

Civil Engineering Section



This short course on data management is designed for graduate students in the civil and structural engineering disciplines who seek to prepare themselves as "data literate e-scientists" in the digital research environment. Detailed videos and writing activities will help you prepare for the specific and long-term needs of managing your research data. Experts in digital curation will describe current sharing expectations of federal funding agencies (like NSF, NIH) and give advice on how to ethically share and preserve research data for long-term access and reuse.

Students will get out of this course:

- Seven web-based lessons that you can watch anytime online or download to your device.
- A Data Management Plan (DMP) template with tips on how to complete each section. Your completed DMP can be used in grant applications or put into practice as a protocol for handling data individually or within your research group or lab.
- Feedback and consultation on your completed DMP by research data curators in your field.

Data Management Course Enrollment

Name

Email

Powered by [Google Docs](#)

[Report Abuse](#) - [Terms of Service](#) - [Additional Terms](#)

 [Open Data Management Course Enrollment](#)

Jeffryes, J. and Johnston, L. (2013). "An E-Learning Approach to Data Information Literacy Education." 2013 ASEE Annual Conference (Atlanta). <http://www.asee.org/public/conferences/20/papers/6956/view>



Five “Flipped Classroom” Workshops Coming Fall 2013



Data Management Series 5: How to Digitally Preserve Your Data for the Future

What happens to your data after the project is complete? What about when you graduate? Will you be able to use your data 10 years from now? Who is going to maintain the data for future use? This session of the Fall 2013 Data Management Series will demonstrate preservation and curation techniques used by information professionals who manage digital information for long-term access.

Flipped classroom: before attending, participants will watch a [six minute video on preservation techniques](#) so that in class, we can focus on hands-on participation and activities. Participants that attend all management sessions (Oct 11-Nov 8) will receive a Certificate for their UMN training records. See information on our course website at <http://z.umn.edu/datamgmt13>. Audience: Graduate students enrolled at the UMN.

Fri, 11/08/2013 - 2:00pm - 3:00pm

Location: **310 Walter Library**

Instructor(s):

[Jon Jeffries](#)

[Lisa Johnston](#)

Status: Open

[Register for this Course](#)

Fri, 11/08/2013 - 3:30pm - 4:30pm

Location: **310 Walter Library**

Instructor(s):

[Jon Jeffries](#)

[Lisa Johnston](#)

Status: Open

[Register for this Course](#)



Personal Archiving Skills Transferable to Data Info Lit

UNIVERSITY LIBRARIES

Welcome: ljohnsto | [MyLibrary](#) | [Logout](#)

[Home](#) [About Us](#) [Services](#) [How to Find](#) [Ask Us!](#) [Chat](#) [Email](#) [Phone](#) [FAQ](#)

Current Awareness & Personal Information Management

- Managing Your Information - Home
- Citation Management
- Productivity Tools
- Alerts and Feeds
- Personal Archiving
 - Documents & PDFs
 - Electronic Notes
 - Photos**
 - Research Data
- Social Networking

Archiving Your Photos

Our phones, digital cameras, tablets, and computers all contain the digital record of our lives. How do we keep track of our vast photo libraries? Here are two solutions to get you started.

A Savvy Solution Using Dropbox

Engineering Librarian, Jan Fransen, shares her method of archiving photos using cloud based storage and Wi-fi. [View this Diagram](#) of her process and watch the video tutorial.

Which Photo Software Should I Choose?

Use this [Matrix of Photo Software](#) tools, such as Flickr, iPhoto, Picasa, and Facebook, to see which one meets your needs. Contact [Scott Spicer](#) with photo management questions.

Photo Workflow: Eye-Fi, Dropbox, and Picasa

Shoot

When I* take a picture with my camera, the **Eye-Fi** card sends it to my phone when they're both connected to the same **WiFi**.

Store

My phone's **Eye-Fi** app sends images to my computer when they're connected to the same **WiFi** (usually at home).

Find

When I open **Picasa** on my computer, it looks for new photos in the folders Eye-Fi and Dropbox use, organizes by date, and tags people for me when it recognizes them.

Backup

Don't forget to backup! For photos, I synchronize to an online **PicasaWeb** (Google) account. I use **CrashPlan** to back up everything on my computer.

Picasa

Picasa includes photo editing tools, collage and video creators, and tools for ordering prints online from your favorite photo service. (I use **SnapFish**.)

*For Jan Fransen, an engineering librarian in Walter Library





2. Listening to the Evolving Campus Need



UNIVERSITY OF MINNESOTA
Driven to DiscoverSM

Data service needs will vary (and evolve) across disciplines



Scientific Data:

- Aerospace Engineering
- Astronomical images
- Institute for Health Informatics
- Chemical Engineering Research Lab
- Botany Images of the Bell Museum Herbarium collection

GIS Data:

- Minnesota Geological Survey
- USpatial and TerraPop

Social Sciences/Survey data:

- Office of Institutional Research
- Climate Change Working Group

Arts & Humanities:

- DAH Symposium
- Ojibwe Conversation Data



Build partnerships with others who are tackling these issues

Office of Information Technology, <http://www.oit.umn.edu/services-category/index.htm>

- Computing and network services include: storage, backup, server support, and database administration.
- Tutorials and workshops available throughout the semester.
- Security tools and guidance (<http://www.oit.umn.edu/security>)

Office of the Vice President for Research, <http://www.research.umn.edu/forresearchers/resources.html>

- “Research Data Management” and “Intellectual Property” online workshops.
- Minnesota Supercomputing Institute tools and training, <http://www.msi.umn.edu/>

Academic Health Center, Information Systems <http://www.health.umn.edu/facultystaff/ahcis>

- Secure, on-site data storage, and anonymization techniques.
- Tools and data analysis with the CTSI Portal, <http://www.ctsi.umn.edu/research/tools-software/index.htm>

Office for Technology Commercialization, <http://www.license.umn.edu/>

- Advising on keeping records and lab notebooks.
- Support for patentable research.

College of Liberal Arts, Office of Information Technology, <http://claoit.umn.edu/research>

- CLA services include: file server and database hosting; survey design and analysis.
- Support and advise on Digital Art & Humanities projects.



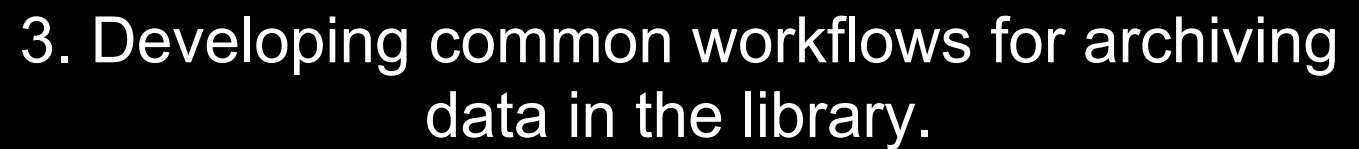
Bring together data service providers in an informal way

Sun 9/15	Mon 9/16	Tue 9/17
		
	9 – 11 Research Data Management Community of Practice (All welcome!)	9 – 11 PSE Staff Meeting
	12p – 1p Lunch meeting ii/li (Janice)	

Discussion Topics:

- Data Storage Options on Campus
- Metadata Standards
- Spatial Data
- Best Practices for De-identifying Research Data
- Data Repositories (Local, National)
- Data Services at the Supercomputing institute
- Practical Examples for Managing data (Sciences)





Libraries offer data archiving tools and repositories



- Institutional Repository for self-deposit of datasets



- Digital library collections open to user-submissions for image, audio, and video.



- GIS data initiative on campus, library partnership.



Example Data Archived in the UDC

digitalconservancy

UNIVERSITY OF MINNESOTA

[University of Minnesota Digital Conservancy >](#)
[University of Minnesota - Twin Cities >](#)
[Unmanned Aerial Vehicle \(UAV\) Research Group >](#)
[Control Law Flight Data >](#)

Please use this permanent URL to cite or link to this item: <http://purl.umn.edu/107827>

Title: Thor Flight 15

Authors: Murch, Austin

Issue Date: 22-Jun-2011

Description: Flight test data, Thor Flight 15 on 06/06/11, in two formats: MATLAB and plain text

Permanent URL: <http://purl.umn.edu/107827>

Appears in Collections: [Control Law Flight Data](#)

Files in This Item:

File	Description	Size	Format	
thor_flight15_loworderAW_2011_06_06.txt	Flight Data, plain text format	11441Kb	Text	View/Open
thor_flight15_loworderAW_2011_06_06.mat	Flight Data, MATLAB format	1113Kb	MATLAB	View/Open
FlightReports_2011_06_06.txt	Flight Report for Test Date	2Kb	Text	View/Open
flight_data_dictionary_thor.pdf	Metadata	196Kb	PDF	View/Open

[Show full item record](#)

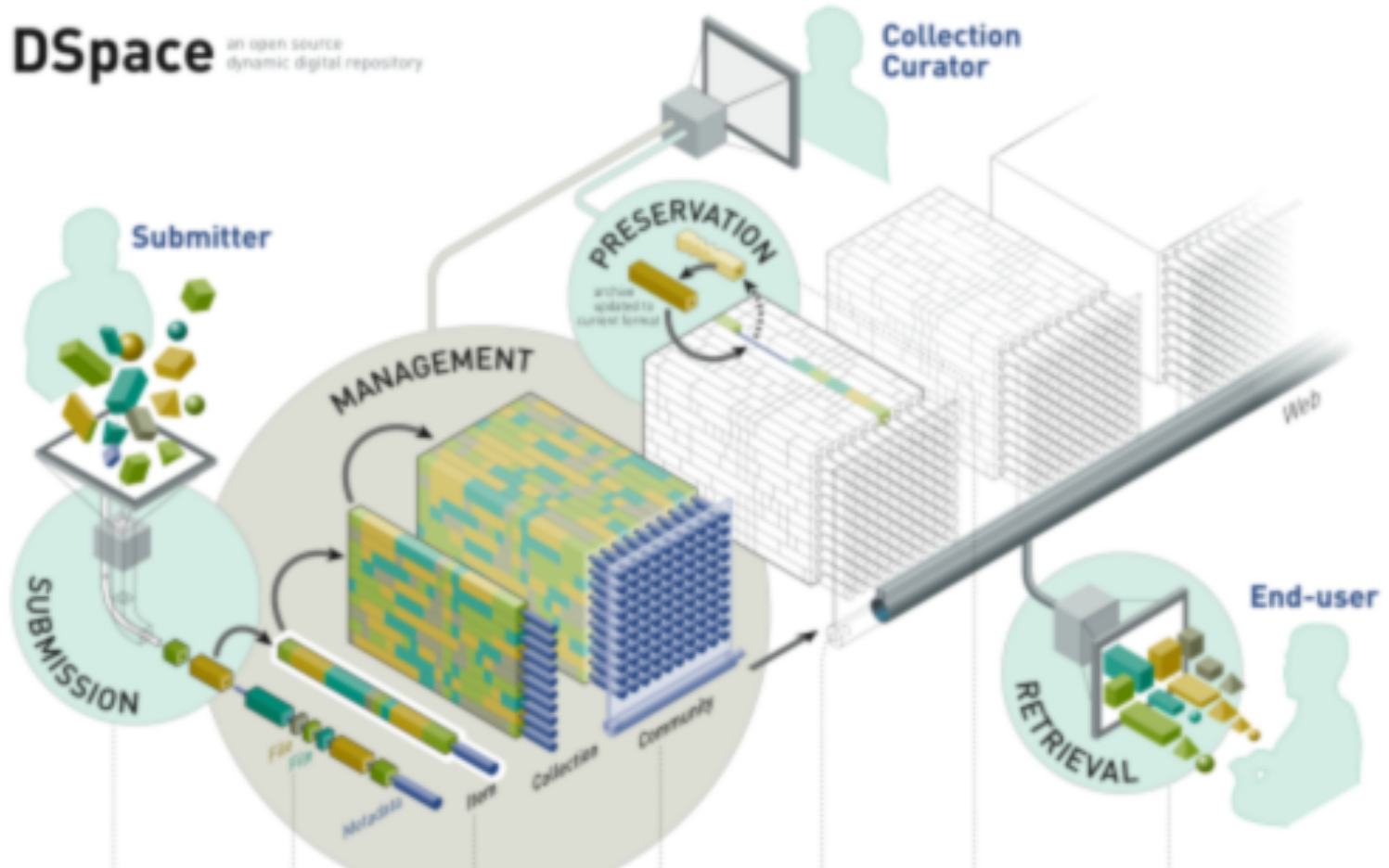
[Recommend this item](#)

[Stats Display](#)

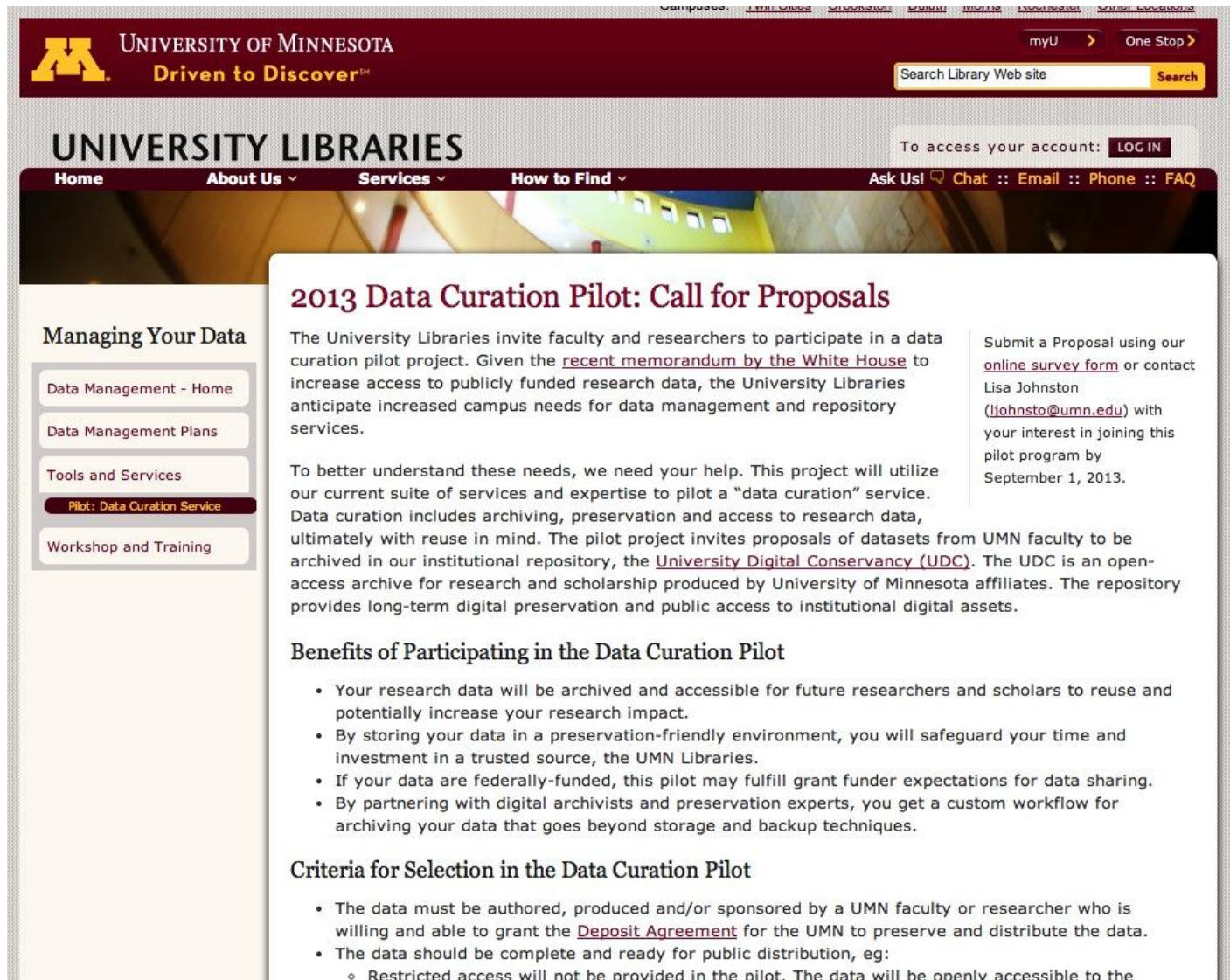


UNIVERSITY OF MINNESOTA
Driven to DiscoverSM

Data Curation Pilot 2013 will develop a workflow for data curation as a service.



Pilot will examine six example data sets from a variety of disciplines.



The screenshot shows the University of Minnesota Libraries website. The header includes the University of Minnesota logo and the tagline "Driven to Discover". A search bar is located in the top right. The main navigation bar includes links for Home, About Us, Services, and How to Find. A secondary navigation bar includes links for Ask Us!, Chat, Email, Phone, and FAQ. The left sidebar is titled "Managing Your Data" and contains links for Data Management - Home, Data Management Plans, Tools and Services, Pilot: Data Curation Service (highlighted), and Workshop and Training. The main content area features a large heading "2013 Data Curation Pilot: Call for Proposals". Below this heading, there is a paragraph introducing the pilot project, a paragraph explaining the need for the pilot, a section titled "Benefits of Participating in the Data Curation Pilot" with a bulleted list, and a section titled "Criteria for Selection in the Data Curation Pilot" with a bulleted list. A sidebar on the right contains contact information for Lisa Johnston and a link to the online survey form.

2013 Data Curation Pilot: Call for Proposals

The University Libraries invite faculty and researchers to participate in a data curation pilot project. Given the [recent memorandum by the White House](#) to increase access to publicly funded research data, the University Libraries anticipate increased campus needs for data management and repository services.

To better understand these needs, we need your help. This project will utilize our current suite of services and expertise to pilot a "data curation" service. Data curation includes archiving, preservation and access to research data, ultimately with reuse in mind. The pilot project invites proposals of datasets from UMN faculty to be archived in our institutional repository, the [University Digital Conservancy \(UDC\)](#). The UDC is an open-access archive for research and scholarship produced by University of Minnesota affiliates. The repository provides long-term digital preservation and public access to institutional digital assets.

Submit a Proposal using our [online survey form](#) or contact Lisa Johnston (ljohnsto@umn.edu) with your interest in joining this pilot program by September 1, 2013.

Benefits of Participating in the Data Curation Pilot

- Your research data will be archived and accessible for future researchers and scholars to reuse and potentially increase your research impact.
- By storing your data in a preservation-friendly environment, you will safeguard your time and investment in a trusted source, the UMN Libraries.
- If your data are federally-funded, this pilot may fulfill grant funder expectations for data sharing.
- By partnering with digital archivists and preservation experts, you get a custom workflow for archiving your data that goes beyond storage and backup techniques.

Criteria for Selection in the Data Curation Pilot

- The data must be authored, produced and/or sponsored by a UMN faculty or researcher who is willing and able to grant the [Deposit Agreement](#) for the UMN to preserve and distribute the data.
- The data should be complete and ready for public distribution, eg:
 - Restricted access will not be provided in the pilot. The data will be openly accessible to the




Data Curation Pilot Faculty Responses to “Why Participate”

“I recognize that I'm not the only person in this predicament of storing larger sets of data and that figuring out how to do this well and sustainably will help many, many folks around the University.”

“Data curation goes beyond backup and storage. Meanwhile, how to archive, preserve, and provide access to (sometimes large) datasets is still new to many researchers”





We can help by training researchers to create better data, bringing together campus partners, and implementing curation techniques that scale.

Google | google.com/datacenters



UNIVERSITY OF MINNESOTA
Driven to DiscoverSM

Thanks

<http://z.umn.edu/datapilot13>



UNIVERSITY OF MINNESOTA
Driven to DiscoverSM

The Library's Role in Enabling Data Interaction for Researchers

NISO Webinar: Research Data Curation
Part 2: Libraries and Big Data
Sayeed Choudhury



DataConservancy



Data Management Services

- Johns Hopkins University Data Management Services (JHUDMS) – <http://dmp.data.jhu.edu>
- Culmination of over a decade of R&D starting with Sloan Digital Sky Survey (SDSS)
- Implementation of Data Conservancy technology development, educational, workforce development and sustainability programs



Two Stages of JHUDMS

- Pre-proposal consultation and assistance with data management plan preparation for NSF proposals – though rapidly expanding beyond NSF and into other use cases
- Post-proposal data management through JHU Data Archive
- First stage paid for directly by JHU; second stage paid for through line items within NSF proposal budgets



Data Management Layers

Layers	Characteristics	Implication for PI	Implication relative to NSF
Curation	Adding value throughout life-cycle	<ul style="list-style-type: none">• Feature Extraction• New query capabilities• Cross-disciplinary	<ul style="list-style-type: none">• Competitive advantage• New opportunities
Preservation	Ensuring that data can be fully used and interpreted	<ul style="list-style-type: none">• Ability to use own data in the future (e.g. 5 yrs)• Data sharing	<ul style="list-style-type: none">• Satisfies NSF needs across directorates
Archiving	Data protection including fixity, identifiers	<ul style="list-style-type: none">• Provides identifiers for sharing, references, etc.	<ul style="list-style-type: none">• Could satisfy most NSF requirements
Storage	Bits on disk, tape, cloud, etc. Backup and restore	<ul style="list-style-type: none">• Responsible for:<ul style="list-style-type: none">• Restore• Sharing• Staffing	<ul style="list-style-type: none">• Could be enough for now but not near-term future



“Big Data”

- What is Big Data?
- There are definitions based on the “V’s” of Big Data (e.g., volume, velocity, variety)
- What is clear is that it’s fundamentally different from “spreadsheet science”
- For me, if a (designated) community’s ability to deal with data is overwhelmed, it’s “Big Data”
- SDSS lessons learned:
<https://wiki.library.jhu.edu/x/eY1XAQ>



Libraries and Big Data

- My assertion is that our community has been overwhelmed by data
- While it's essential to leverage existing capability, we need to be aware that this is the beginning of the journey
- The goal is not about supporting libraries – it is about supporting scholarship
- Data repositories need to coalesce into infrastructure
- Interaction with data should become seamless



Acknowledgements

- NSF Award OCI-0830976
- Sheridan Libraries and JHU financial support
- Data Conservancy colleagues for slides
- <http://dataconservancy.org>
- <http://dmp.data.jhu.edu> -- JHU DMS
- <http://www.dlib.org/dlib/september12/mayernik/09mayernik.html> -- blueprint document
- <https://www.youtube.com/watch?v=F6iYXNvCRO4> -- data management layer stack model

NISO Webinar: Research Data Curation Part 2: Libraries and Big Data

Questions?

All questions will be posted with presenter answers on the NISO website following the webinar:

http://www.niso.org/news/events/2013/webinars/data_curation



THANK YOU

Thank you for joining us today.
Please take a moment to fill out the brief online survey.

We look forward to hearing from you!