WHAT'S THE DEAL WITH DATA?

UNDERSTANDING RESEARCH DATA MANAGEMENT & SUPPORT AT UVM

Elizabeth A. Berman
OPEN SCHOLARSHIP

Open Access

Open Data
National Institute of Health (NIH)
National Science Foundation (NSF)
*Office of Science & Technology (OSTP)
Department of Energy (DOE)
Agency for Healthcare Research and Quality (AHRQ)
Centers for Disease Control and Prevention (CDC)
Department of Defense (DOD)
Food and Drug Administration (FDA)
NASA
National Institute of Standards and Technology (NIST)
United States Geological Survey (USGS)
National Oceanic and Atmospheric Administration (NOAA)
United States Department of Agriculture (USDA)
“World-class excellence in disciplinary and multi-disciplinary research is a priority for UVM as one of the nation’s premier small research universities and a leader in creating and translating knowledge in service to the people of Vermont and the nation.”
– Office of the Vice President of Research

1,098 FT Faculty
266 PT Faculty
615 Grants & Contracts
$128M External Funding
70% Federal Grants

http://www.uvm.edu/medicine/discovery/?Page=factsfigures.php
RESEARCH DESIGN

PHASE 1: Environmental Scan

• Data management tools & software
• Different models of service and support

PHASE 2: Mixed Methods Research

• How do faculty at UVM manage their research data? How do they document and describe data? Where do they store data? How do they provide access to or share their data?
• Are there any challenges or barriers UVM faculty face in effectively managing their research data?
• Is there a need or support for data management planning services on campus?
PHASE ONE

ENVIRONMENTAL SCAN
Data management

Services

We help MIT faculty and researchers manage, store, and share data they produce.

Assistance with creating data management plans

Many research funders have requirements for data sharing and data management plans. We can help you to create these plans, assess the data management needs of your project, and help you to identify data management solutions.

Individual consultation

We are available to help you, your lab, or center understand your data management needs and recommend best practices for keeping your data usable, now and into the future.

Workshops

Our workshops teach you how to manage your data more efficiently and help you to share your data with others. See a schedule of upcoming workshops.

Questions?

MIT users:
Email data-management@mit.edu
Learn how we can help you

Visitors:
See information for visitors
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

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

Our Services

Data Management Plan Assistance
We offer personalized assistance for crafting 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

http://conservancy.umn.edu/handle/11299/166578
Metadata Production and Consulting Services

Metadata is structured data that facilitates access, administration and preservation of resources.

Want help? Contact us!

We provide expert metadata advice to academic libraries, journal publishers, non-profit organizations, agencies and Cornell University units across multiple fields.

Services include:

Consultation
- Review project requirements and make recommendations
- Analyze existing metadata and recommend transformation strategies, when appropriate
- Educate and train metadata creators on best practices

Development
- Identify community metadata standards to best describe and serve resources
- Create data models that foster seamless access, interoperability and reuse
- Determine types of access and appropriate repositories that fit project needs
- Establish workflows for metadata creation or capture

Production
- Reformat metadata to meet project specifications
- Modify metadata to conform with community standards or project guidelines
- Provide quality control throughout project lifecycle

Guiding Principles:
Data Visualization

Using Data Analysis & Visualization Software

- Browse the data visualization example gallery and locate related tools
- View software available in the Data and Visualization Services computer lab
- Find help on specific software or topics:
  - Learn how to use Tableau for easy, interactive charts and maps
  - Learn how to use OpenRefine for data processing
  - Get tips for text analysis projects
  - View additional tutorials on the Data and Visualization Services blog
  - Compare various mapping tools

Getting Help with Visualization

- Meet with a Data and Visualization Services Consultant
- Attend a Data and Visualization Services Workshop
- Get tips on designing effective charts, figures and posters
- Visit the Multimedia Project Studio for graphic design help

Engaging with the Visualization Community at Duke

- Attend the Visualization Friday Forum
- View our Flickr Visualization Gallery
- Submit your project to our Data Visualization Contest
- Visit or propose a data visualization project for the LINK MediaWalls

WHAT IS DATA VISUALIZATION?

In a nutshell, data visualization is the term we use to describe all of the ways people transform data into visual representations. This could be a map, a bar chart, a timeline or an artistic rendering of data. For more definitions, examples and helpful data visualization tools, see our Duke Data Visualization LibGuide.

Mapping and GIS

Mapping is a special subset of data visualization. Researchers use Geographic Information System (GIS) software to conduct complicated geographical data analyses and to develop high-quality maps. The Data and Visualization Services department includes staff with special training in GIS and mapping who have developed a series of workshops and training guides to help with various types of mapping projects.

Some of the tools we can introduce you to are:

- ArcGIS for professional GIS and mapping
- Tableau for easy, interactive maps
- Online mapping tools:
  - GeoCommons for open analysis, interactive online mapping
  - Google Earth for 3D maps and animated tours
  - Google Fusion Tables for online maps based on spreadsheets
Think ahead. Plan ahead.

Discover the new, easy-to-use tool designed to build quality data management plans — for free

Go to the DMPTool

RESOURCES & TRAINING
- Data management guides
- Webinars
- Slides and Marketing materials

UC BY THE NUMBERS
- 862 DMPs created by UC researchers
- 980 users across UC

UC CONTACTS
- Have a question?
- Get local campus help

RECENT DMP NEWS
- Get the latest information about data management and the DMPTool

https://dmp.cdlib.org/
PHASE TWO

MIXED METHODS RESEARCH
RESEARCH DESIGN: SEQUENTIAL MIXED METHODS

Document Analysis

In-Depth Interviews

Survey

Results

Target Population

- University of Vermont faculty
- Received a NSF grant between January 1, 2011 – December 31, 2014
- Required to submit a data management plan (DMP)

Population = 49
Sample = 35
• Biology (CALS)
• Plant Biology (CALS)
• Plant & Soil Science (CALS)
• Anthropology (CAS)
• Chemistry (CAS)
• Geology (CAS)
• Geography (CAS)
• Physics (CAS)
• Psychology (CAS)
• Mathematics & Statistics (CEMS)
• School of Engineering (CEMS)
• Rubenstein
• EPSCoR: Experimental Program to Stimulate Competitive Research (OTHER)
Data Management Plans Coding

- What types of data will be produced? **DATA TYPE**
- What metadata standards do you need to follow for documentation? **METADATA**
- What physical and/or cyber resources and facilities (including third party resources) will be used to store the data? **STORAGE**
- Do any considerations need to be made to protect sensitive information, including study participant confidentiality and intellectual property protection? **RESTRICTIONS**
- What policies do you need to follow with respect to data sharing and reuse? **SHARING**
- How will you ensure archiving and preservation of the data you will produce? **PRESERVATION**
DOCUMENT ANALYSIS

- Experimental data
- Samples & specimens
- Computer codes
- Software
- Images
- Computational data
- Lab notebooks
- Geospatial data
- Field notes
- Modeling data
- Crowdsourced data
- Interview transcripts
- Surveys
- Music recordings
- Curriculum materials
- Generic data
51% (n=18) addressed data description
28% (n=9) addressed specific metadata schema
- Darwin Core, Water Markup Language (WaterML), Federal Geographic Data Committee Geospatial Metadata Standards, Systems Biology Markup Language (SBML), Ecological Metadata Language (EML)
• **83% (n=29)** addressed storage
  - computer hard drive, external hard drive, UVM server, external server, third-party cloud storage (e.g. DropBox), USB drives
• **40% (n=14)** addressed storage & back-up protocols

---

**51%**

**83%**

**80%**

**100%**

**Metadata & Documentation**

**Data Storage & Back-up**

**Data Sharing & Re-Use**

**Data Preservation & Archiving**
DOCUMENT ANALYSIS

- 80% (n=28) addressed sharing data
- 20% (n=7) addressed restrictions
- 11% (n=4) addressed sharing via open-source/creative commons licenses
- 23% (n=8) addressed sharing via direct request

51% Metadata & Documentation
83% Data Storage & Back-up
80% Data Sharing & Re-Use
100% Data Preservation & Archiving
100% (n=35) addressed data preservation and access
49% (n=17) addressed preservation via repositories
  - publications (n=20); conferences & meetings (n=13); research group/project website (n=15); personal website (n=4); collaborative wiki space (n=1)
Number of Words in Data Management Plan

- 0-133: 2
- 134-267: 4
- 268-401: 7
- 402-535: 7
- 536-669: 2
- 670-803: 1
- 804-937: 7
- 938-1071: 3
- 1072-1205: 1
- 1206-1339: 1
Target Population

- University of Vermont faculty
- Received a NSF grant between January 1, 2011 – December 31, 2014
- Required to submit a data management plan (DMP)
- Purposeful sampling:
  - academic department
  - academic rank
  - gender

Population = 49
Sample = 6
IN-DEPTH INTERVIEWS

Pre-Interview Questionnaire

- Data types & formats
- Metadata & description
- Size of dataset
- Data storage & back-ups
- Lifespan of data
- Data sharing methods & restrictions

Semi-Structured Interview

- Data management guidance
- Data management challenges
- UVM support of data management
Data Access, Data Sharing & Researcher Attitudes

- Data access vs. data sharing
- Infrastructure: subject repositories, general data repositories, institutional data repositories
- Education and outreach

Institutional Support & Collaboration

- Institutional policies
- Funding concerns
- Decentralized institutional model: misconceptions and misunderstandings about available infrastructure and services
FURTHER RESEARCH

• Broad survey of faculty data management practices.

• Better understand researchers’ attitudes towards data management planning:
  o Do researchers use DMP for guidance during research projects?

• Data management planning as tool to engage with students
  o Intersection between student pedagogy, student learning, information literacy, and campus infrastructure?

• What is the capacity of UVM Libraries to provide data management services or support?