

# The Open Data Imperative

## The Role of Academic Libraries and Archives in Research Data Management Support

Presented by Julia Frankosky, Lisa Schmidt,  
& Hailey Mooney

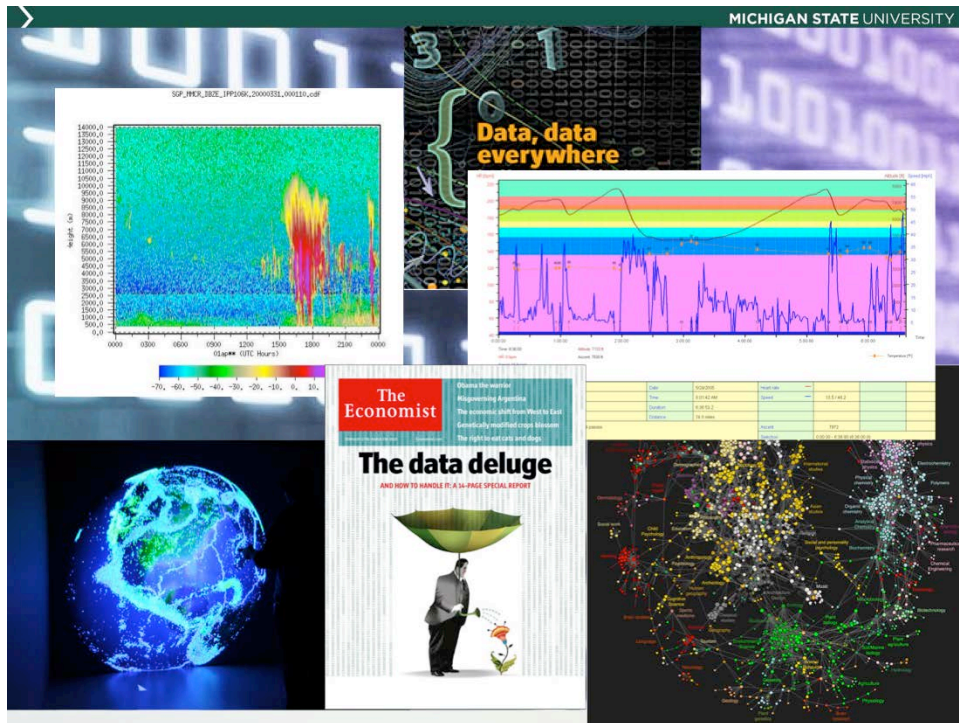
with contribution by Aaron Collie

Michigan Library Association Annual Conference - October 16, 2013

## Overview

- Research data, scholarly communications, the open access movement, and federal requirements
- Research data management best practices and guidelines
- The MSU example

Good morning everyone. My colleagues and I are going to talk to you about the role that academic libraries and archives can play in regards to helping researchers manage their research data. I'm going to provide some background and context about research data, like what's going on in the scholarly communication arena and with regards to the open access movement, as well as federal mandates concerning this data. My colleague Lisa will provide some best practices and guidelines for research data management, and Hailey will talk about the services and activities that the MSU Libraries and archives have implemented.



Getting started, I want to make sure that we all have the same understanding of data and what it means when we talk about research data. Research data is not new by any means, but it has definitely moved into the spotlight in the past few years, often when people are discussing the abundance of data: a data deluge that has befallen us and the issues of managing and storing big data. But what exactly are people talking about? Data means different things to different people depending on their disciplines. Data is simply information and research data is any sort of information that is generated during the course of research. Some examples include survey results, gene sequences, photos, videos, audio files, transcripts, specimen, and sensor readings.



Open access is the practice of providing free access online to peer-reviewed scholarly research, predominantly in the form of journal articles, though sharing other items like book chapters and theses is becoming more popular. So what does the open access movement have to do with research data?

## Scholarly Communication



Scholarly communication lifecycle model from Western Libraries: <http://www.lib.uwo.ca/scholarship/scholarlycommunication.html>

This is where the scholarly communication cycle comes into play. Scholarly communication is the act of conducting research and then sharing your findings, which allows others to discover what you've found and use it in their own studies. In our current scholarly communication model, the research data that is shared has been synthesized by the researcher in a journal article that contains snippets of the data in charts, graphs, or tables. These articles are increasingly being disseminated in open access journals and there is a growing movement to include the actual research data behind the articles within these publications. Data publication would allow additional synthesis of this data by other researchers who may be able to interpret the data in different ways. It also could help decrease the cost of some research projects since the original research would have already been generated so the costly data collection stage can be skipped.

But deciding whether or not you as the researcher would like to share your data openly is not always a decision that you get to make. Major grant funding agencies now require that grant applications, which is how many of these research projects are paid for, now come with a data management plan that must include not only how you plan on saving and storing your data, but also how and where you plan on sharing it.

## National Institute of Health (NIH)

"Starting with the October 1, 2003 receipt date, investigators submitting an NIH application seeking \$500,000 or more in direct costs in any single year are expected to include a plan for data sharing or state why data sharing is not possible."

"Final NIH Statement on Sharing Research Data," February 26, 2003, accessed September 12, 2013, <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html>.

(read slide text)

NIH asks that the data sharing plan consist of a brief paragraph and can include a schedule for data sharing, format for the final dataset, what documentation you'll provide, what analytic tools, if any, will be included, how you'll share the data. These are just some suggestions and NIH states that what you include in your plan may vary.

## National Science Foundation (NSF)

“Beginning January 18, 2011, proposals submitted to NSF must include a supplementary document of no more than two pages labeled "Data Management Plan" (DMP). This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results.”

"NSF Data Management Plan Requirements," accessed September 12, 2013, <http://www.nsf.gov/eng/general/dmp.jsp>

(read slide text)

Like NIH, NSF also offers general guidelines as to what to include in your data management plan, and different directorates have different guidelines. This is helpful to an extent, but the lack of specifics can lead to some confusion as to how detailed should a NSF data management plan be.

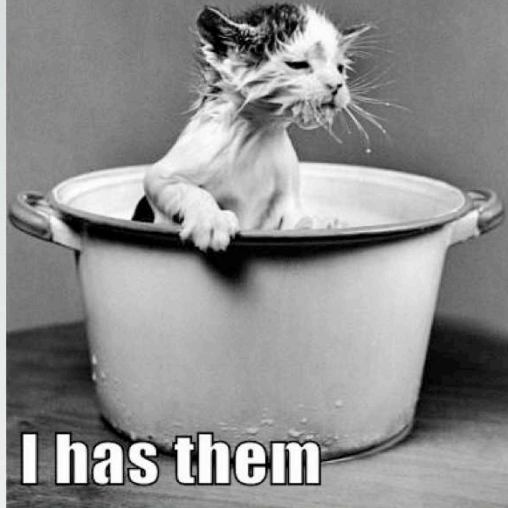
## White House Memorandum

"The Office of Science and Technology Policy (OSTP) hereby directs each Federal agency with over \$100 million in annual conduct of research and development expenditures to develop a plan to support increased public access to the results of research funded by the Federal Government. This includes any results published in peer-reviewed scholarly publications that are based on research that directly arises from Federal funds..."

John Holdren, Memorandum for the Heads of Executive Departments and Agencies: Increasing Access to the Results of Federally Funded Scientific Research. February 22, 2013, accessed September 12, 2013, [http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp\\_public\\_access\\_memo\\_2013.pdf](http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf)

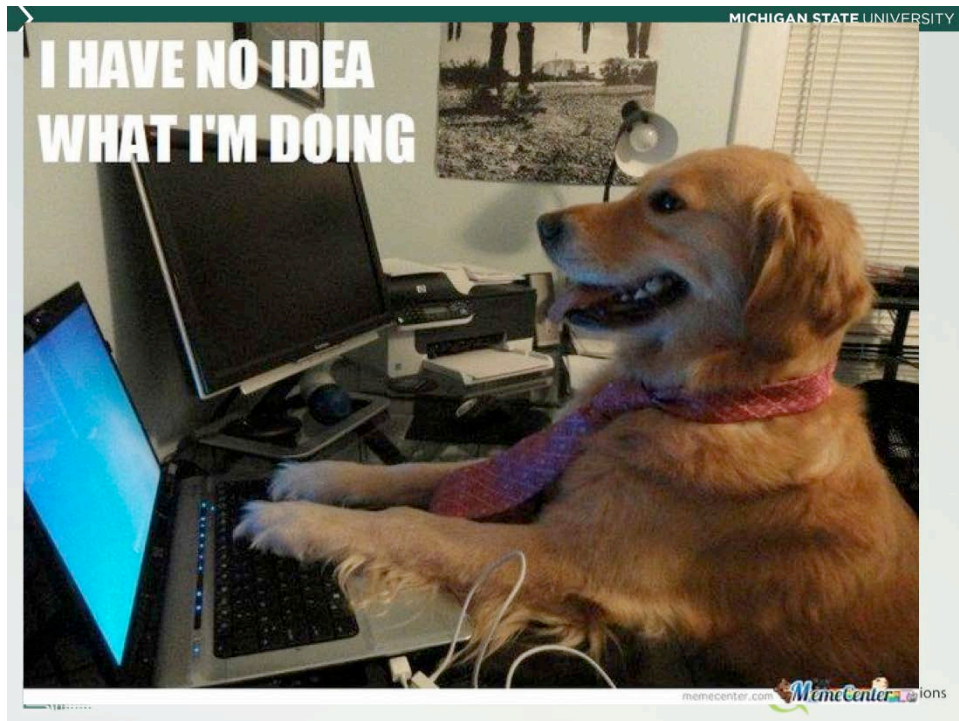
[Read slide text]

**Problems:**



**I has them**

Agencies and institutions, and especially their researchers find themselves facing some issues when it comes to actually figuring out how to manage and share their data.



How exactly do you write a good data management plan? Other than the suggested elements provided by the grant funder, what else do you need to include? What infrastructure is in place at my institution to help with things like data storage and sharing tools? Where can I go to get help with all these questions?

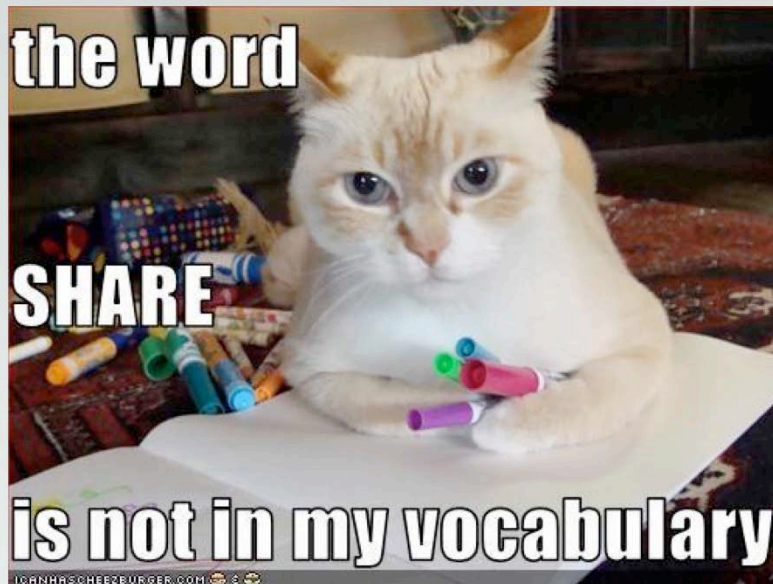
Just writing a sound data management plan can be intimidating. Researchers may feel that they have no idea what they're doing. And that while the guidelines offered by the funding agency are helpful, they aren't necessarily enough of a roadmap to help a research write a viable data management plan.



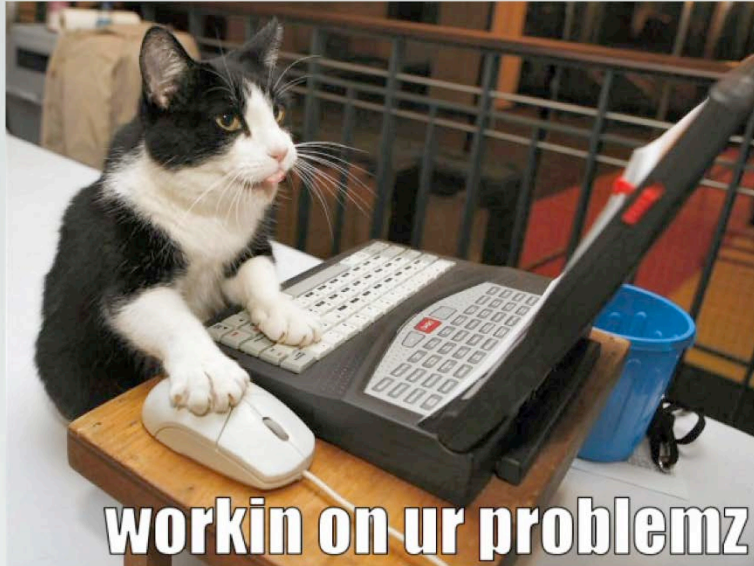
How do you even organize your data so that it makes sense to you in 5 or 10 years as well as someone who was not involved with this project? How often do you come across a file on your computer that has an incredibly vague name? Now multiply that by one hundred (or more) and you have a horrible situation faced by many researchers. How can anyone, or even yourself, reuse your data if they have no idea what it is or how to find it in your files?



What's even more terrifying than having a hard time locate a file on your computer? How about having your computer crash, wiping out all of your files forever. Having your data backed up is a very real threat to researchers. You may be saving your data to your computer and an external harddrive, but are you really saving ALL of your data? Are you keeping the backup drive in the same location as the original? Do you even have enough storage on your computer to complete this data collection or are you going to have half saved on your computer, another part on an external hard drive, and the rest saved in cloud storage?



Along with more technical issues, there is also a discipline-specific culture issue that your researchers may encounter: whether or not data sharing is encouraged. Some disciplines, like astronomy have a culture of data sharing since the data they use is massive and expensive to collect. But this is not the case for all disciplines, especially in the humanities and social sciences. Sharing your data, especially before your scholarly article on the topic is able to make it through the peer-review and publication process, can be a scary prospect. What if someone uses the data you painstakingly collected and publishes the same findings as you, but before you could get the chance? How will this affect your tenure and promotion prospects? How often your data is cited in other articles isn't taken into account the way that one counts article citation, so there doesn't seem to be as much incentive from that standpoint to share.



There's a lot of questions that arise in the minds of your researchers, and this is where we as librarians and archivists can step in and assist with providing answers and best practices to assist your researchers with sorting these problems out.

# RESEARCH DATA MANAGEMENT GUIDELINES AND BEST PRACTICES