

# Managing Research Data in The 'Active' Phase

MacDATA Workshop: March 16, 2016

---

J. Brodeur, D. Askey  
McMaster University Library



**MAPS  
DATA  
GIS**

# Overview

1. Research Data Management: The larger context
2. Storing and backing up data

# Workshop Objectives

## **The aim of this workshop is to:**

1. Encourage a conversation about data management challenges, questions and solutions,
2. Present and discuss approaches to managing data throughout the 'active' research phase; and,

# The Larger Context: Research Data Management

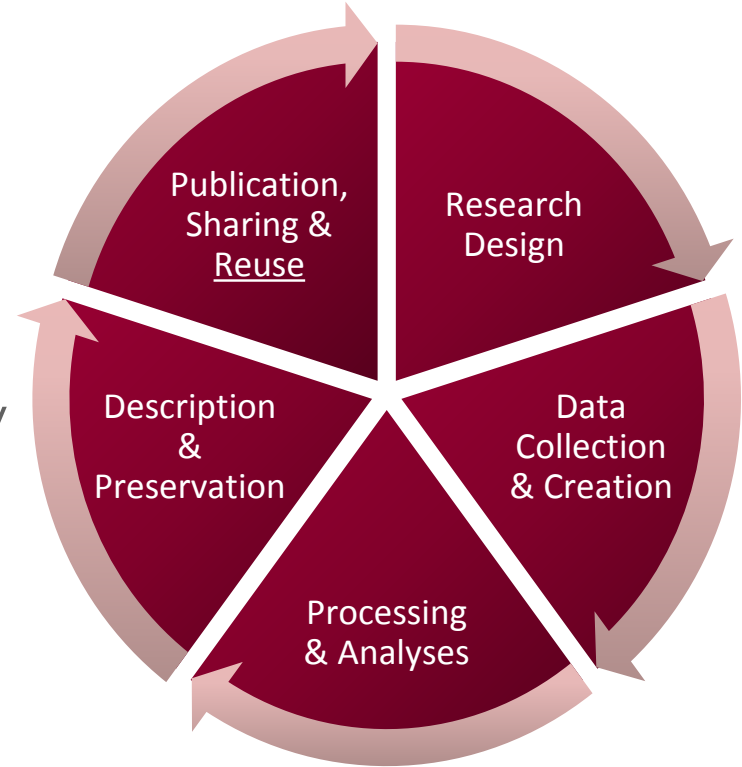
---

# Research Data Management is...

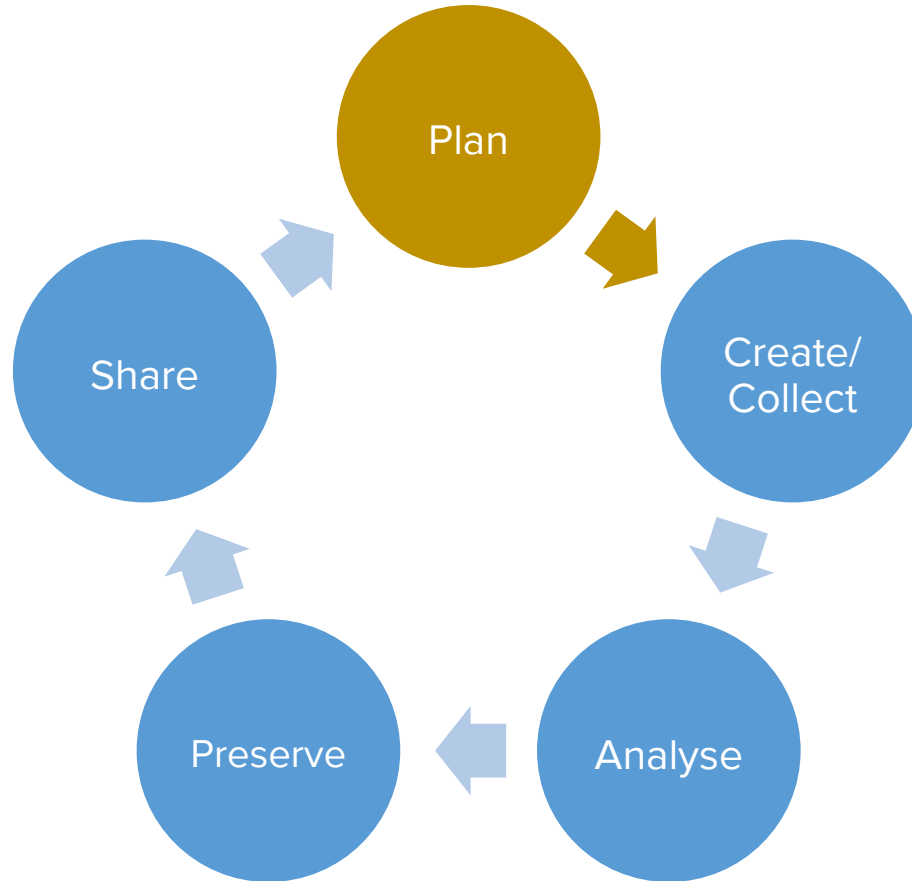
... the **active organization** & maintenance of data

... the **application of best practices** to ensure data security, accessibility, usability, and integrity

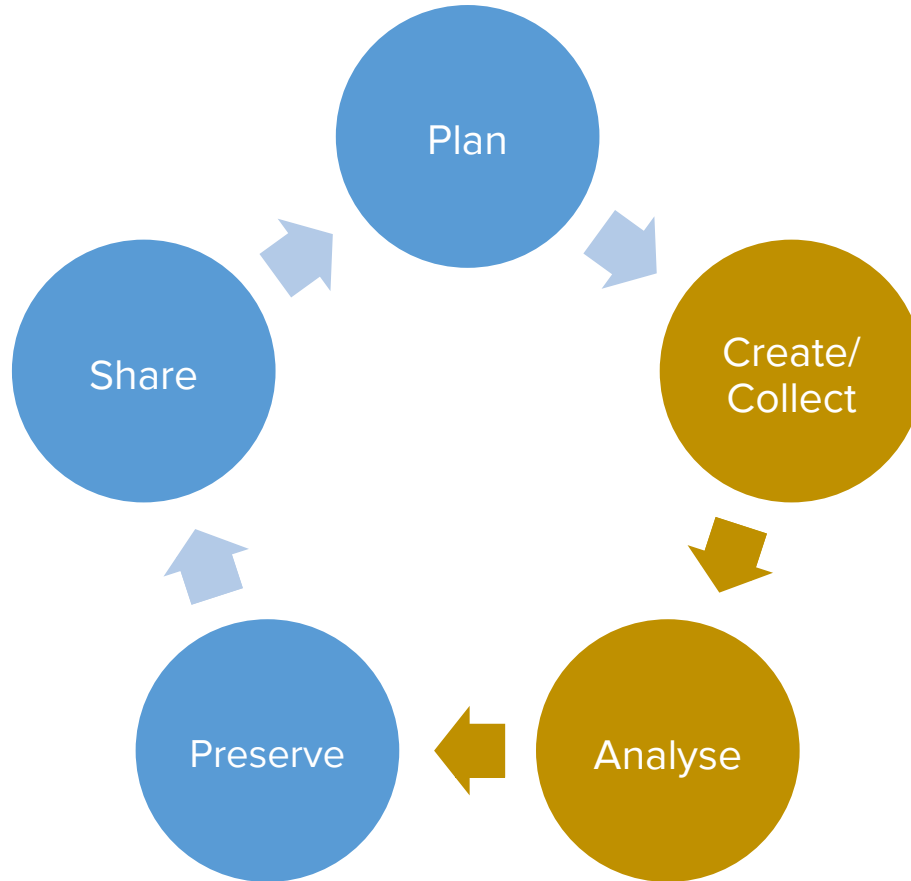
... a **set of activities** resulting in self-describing data sets that can be discovered and reused.



# Last Workshop



# Today's Workshop



If you shared your data another researcher or collaborator, would they be able to:

- a. Interpret and understand it?
- b. Use it in new analyses?

Would someone (including you) be able to find, interpret and use your data 20 years from now?



# Tri-Agency Statement of Principles on Digital Data Management

## Expectations:

- Data management planning
- Constraints and obligations
- Collection and storage
- Metadata
- Preservation, Retention and Sharing
- Timeliness
- Efficient and cost-effective



# Tri-Agency Statement of Principles on Digital Data Management

## Expectations:

- Data management planning
- Constraints and obligations
- Collection and storage
- Metadata
- Preservation, Retention and Sharing
- Timeliness
- Efficient and cost-effective

# Tri-Agency Statement of Principles on Digital Data Management

## Constraints and obligations

- ✧ “Research data must be managed in conformity with all commercial, legal and ethical obligations”.

## Collection and storage

- ✧ “... using software and formats that ensure **secure** storage and enable **preservation** of and **access** to the data...”



Draft Tri-Agency Statement of Principles on Digital Data Management

<http://www.science.gc.ca/default.asp?lang=En&n=83F7624E-1>

# Tri-Agency Statement of Principles on Digital Data Management

## Preservation, Retention and Sharing

- ✧ “All research data resulting from agency funding should normally be preserved in a **publicly accessible, secure** and **curated** repository...”

## Timeliness

- ✧ Data should be shared **as early as possible** in the research process...”
- ✧ “...data should be shared **no later than upon the publication** of results”

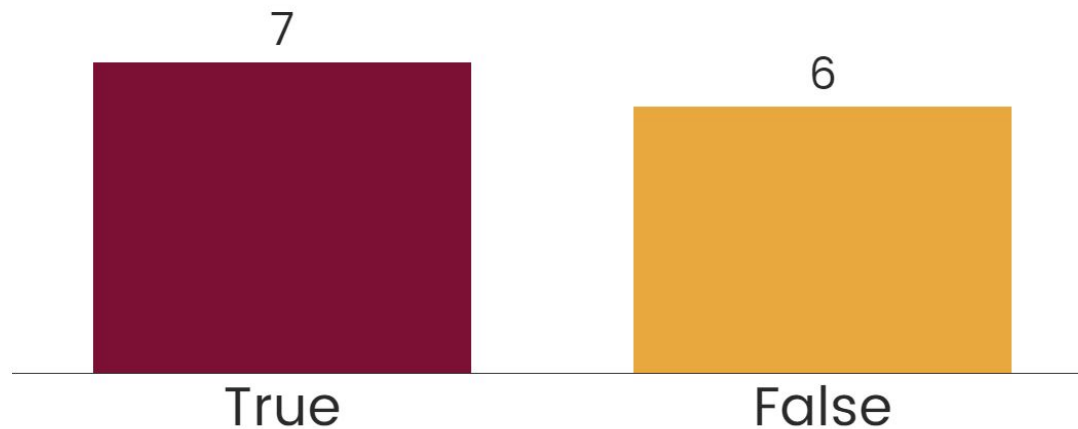


Draft Tri-Agency Statement of Principles on Digital Data Management

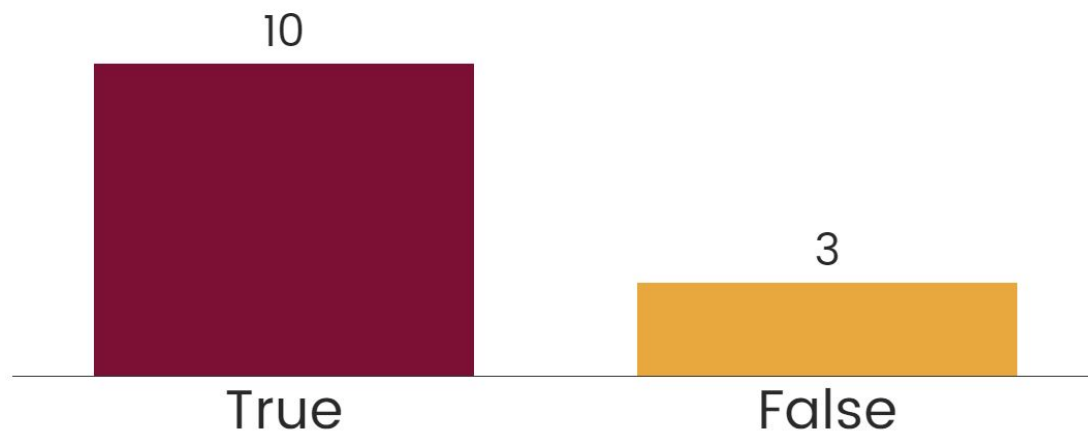
<http://www.science.gc.ca/default.asp?lang=En&n=83F7624E-1>

# Audience Poll #1

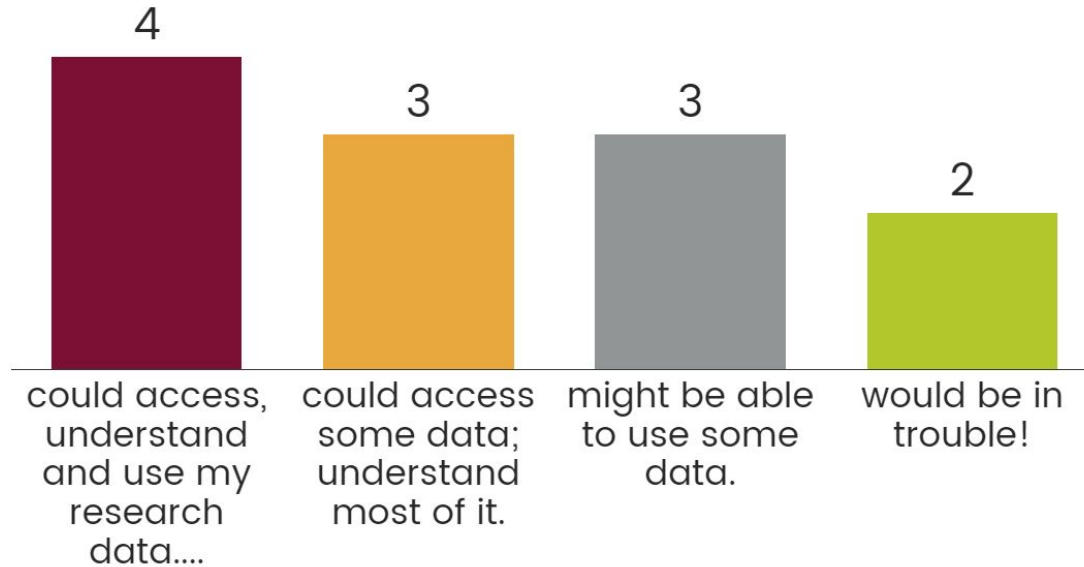
# I have lost important data/files in the past...



I could access and use my research data from 5/10/15 years ago...



# If I were abducted today, my collaborators...





# Considering my data, I struggle with...

Strongly disagree

Keeping up with the pace of collection.

2.5

Finding sufficient storage space.

2.3

Keeping it all backed up.

2.4

Managing appropriate access to the data.

2.7

Training others how to manage and use it.

3.9

Describing the data and methods appropriately

3.5

Strongly agree

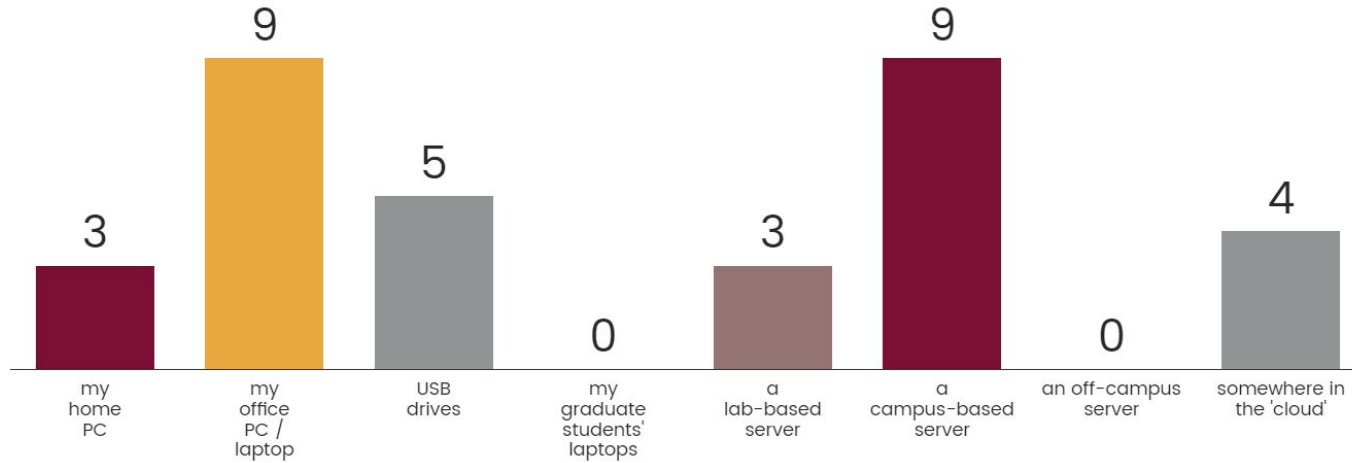


# Data Storage and Backup

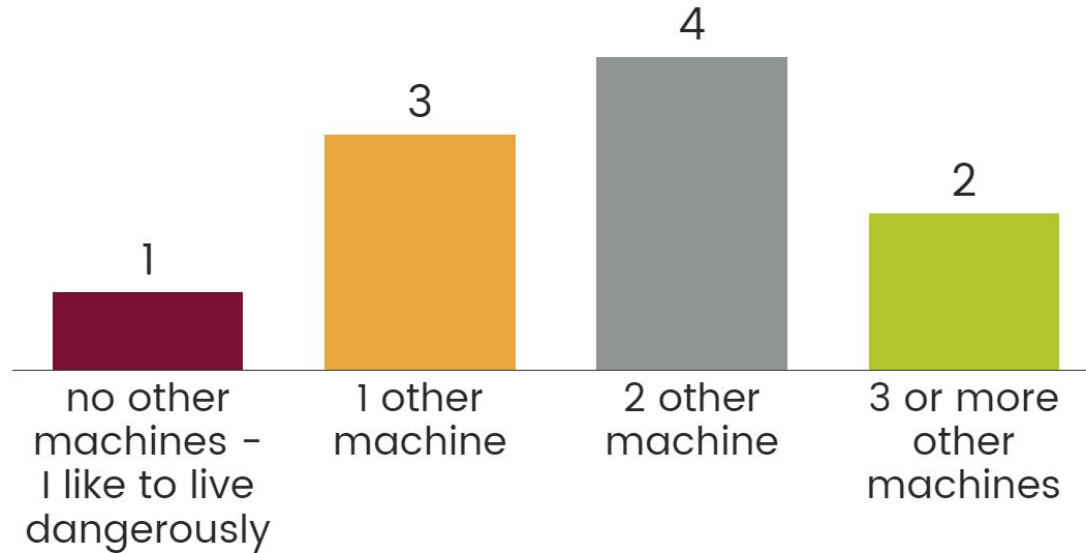
---

# Audience Poll #1

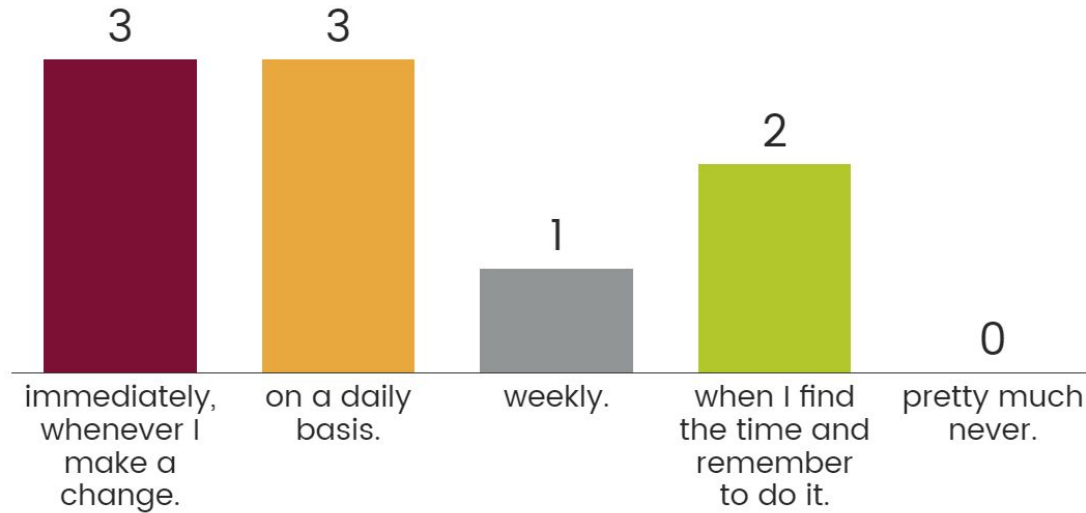
# My research data resides on...



# My research data is replicated on...



# My data is backed up...



# Challenges: Storage and Backup

- ✧ Volume
- ✧ Documentation
- ✧ Managing Access
- ✧ Ethical / Privacy Issues
- ✧ Security and Integrity
- ✧ Cost

# Guiding Questions for Sharing & Backup

What resources are available?

- ✧ Financial, infrastructure & services

How important is the data?

How much data is there?

Who needs access to data?

What level of data security is required?

- ✧ Ethical / legal requirements



# Discussion

With your neighbours, take a few minutes to discuss the benefits and drawbacks of storing your data on:

- ✧ PCs / laptops / mobile devices
- ✧ External storage devices (hard drives, optical, USB)
- ✧ Networked drives
- ✧ Cloud services (Dropbox, Google Drive, etc.)

# Digital Data is Fragile

...but it can be handled appropriately

3

2

1

3

copies of your data

2

1

3

copies of your data

2

copies are on-hand (easily accessible)

- ✦ 1 “**production**” (working) copy
- ✦ 1 “**production backup**” copy

1

3

copies of your data

2

copies are on-hand (easily accessible)

- ✦ 1 “**production**” (working) copy
- ✦ 1 “**production backup**” copy

1

copy is in another location (“off-site”),  
with a ***trusted*** service provide

“**Production**” copy → Where you work with the data

- ✧ PC, laptop, mobile device, etc.

“**Production backup**” copy → Easily accessible (+ versioning?) backup

- ✧ External hard drive with backup software
- ✧ MacDrive (seafiler): <https://macdrive.mcmaster.ca>
- ✧ Dropbox, Google Drive, etc.

# Off-site “Archived” Backup

321

## Providers / Services:

- ✧ Campus / Consortium-hosted (RHPCS)
- ✧ Remote, Commercial
  - Backblaze, Iron Mountain, JustCloud, etc.



# Considerations for “Archived” Backup

321

- ✧ Security (physical and electronic)
- ✧ Automation
- ✧ Availability (and time to recover)
- ✧ Versioning
- ✧ Integrity-checking and error correction
- ✧ Data storage (locational) requirements
- ✧ Cost

# Why Distance is Important

Gustavus Adolphus College

March 29, 1998

F3 tornado



# Why Distance is Important

Lynn Jamieson,  
University of Edinburgh



# Resources

University of Edinburgh MANTRA RDM Training Kit

- ✧ <http://datalib.edina.ac.uk/mantra/libtraining.html>

McMaster Library's Data Management Webpage

- ✧ <http://library.mcmaster.ca/rdm/collecting>

CARL Portage Network's National RDM Information Page

- ✧ <https://portagenetwork.ca/>
- ✧ \*\*\* Full information page to be released in mid-April \*\*\*

# Thank You.

**RDM**  
@McMaster

Check out <http://library.mcmaster.ca/rdm> for more information

Contact us at [rdmgmt@mcmaster.ca](mailto:rdmgmt@mcmaster.ca)