

# Managing your research data

---

Resources, tools, strategies

Vivek Jadon ([vivek@mcmaster.ca](mailto:vivek@mcmaster.ca)) & Jay Brodeur ([brodeujj@mcmaster.ca](mailto:brodeujj@mcmaster.ca))

McMaster University Library

FHS Graduate Student Workshop | 06-November, 2018

# Overview

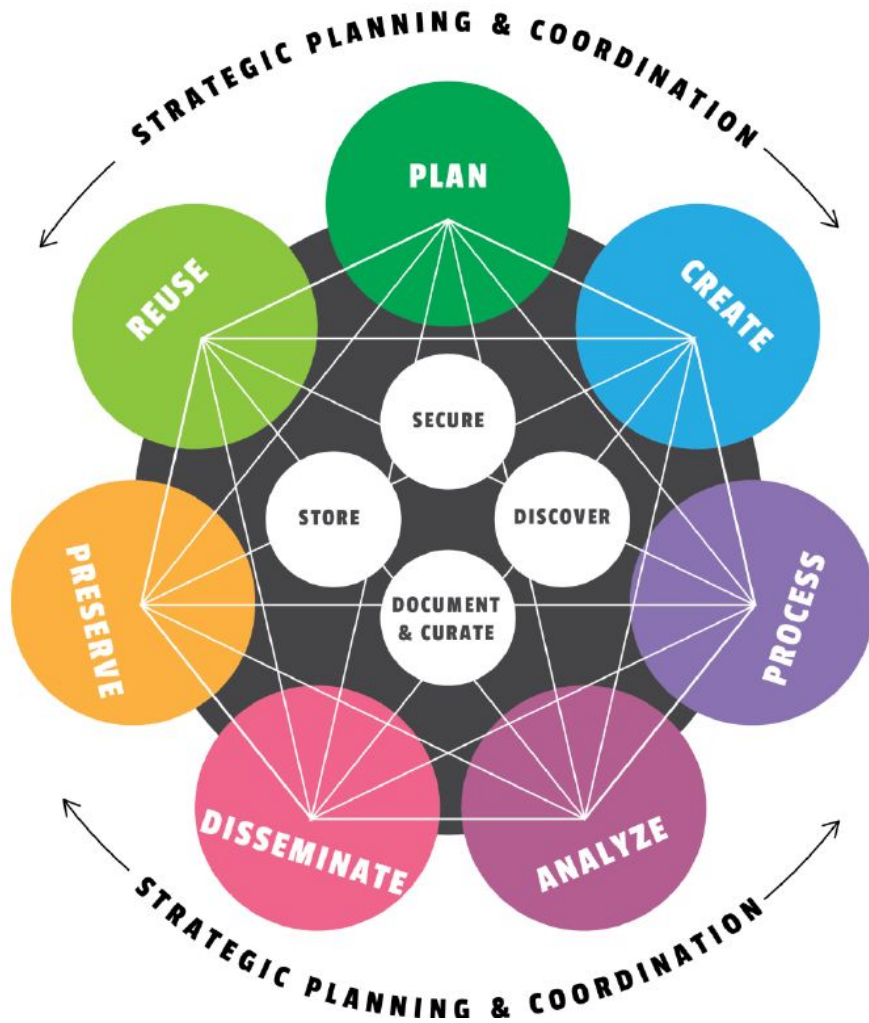
- ✧ A brief Research Data Management primer (10 mins)
- ✧ Data management planning (20 mins)
- ✧ Active data management
  - Data storage & backup (10 mins)
  - In-project data management & collaboration (15 mins)
- ✧ Data archiving & sharing (20 mins)
- ✧ Helpful data management resources (5 mins)

**Research data** are contents that are used as primary sources to support research, scholarship, artistic activity or research-creation, and that are used as evidence in the research process and commonly accepted in the research community as necessary to validate research findings and results.

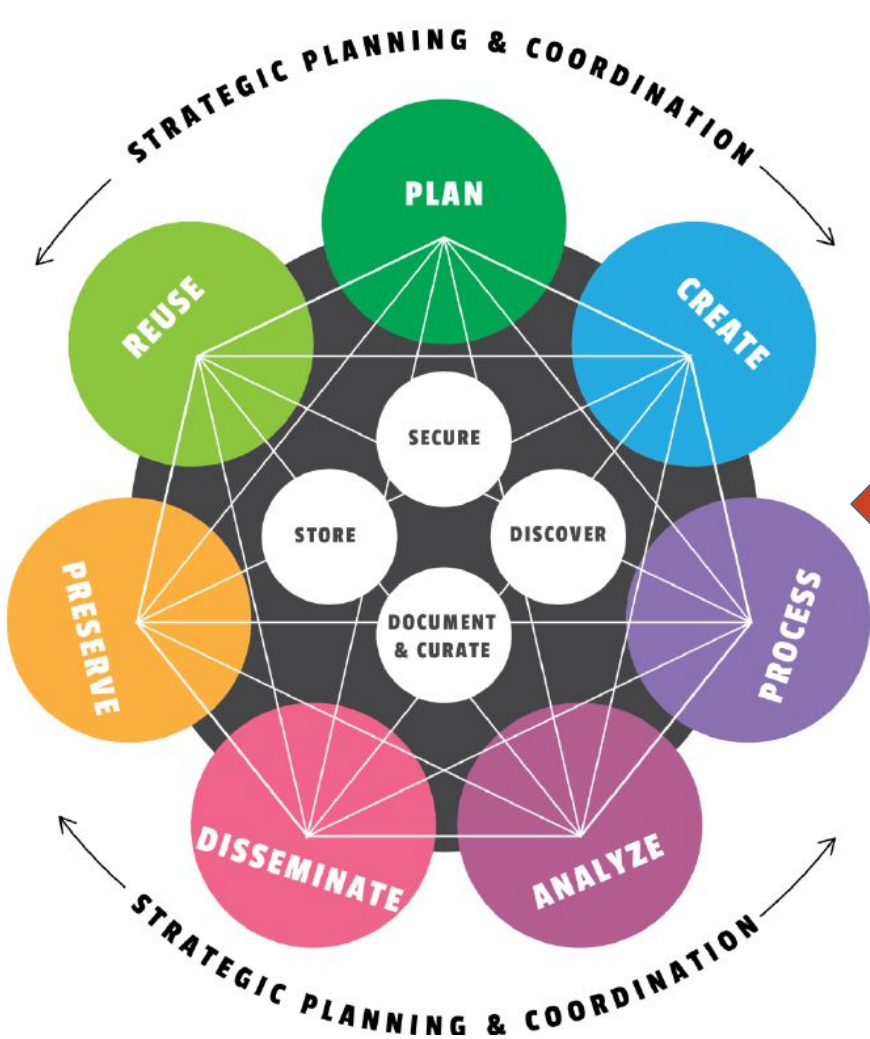
Canadian Tri-Agency

**Research Data Management** is the active organization & maintenance of data throughout the research cycle to ensure its security, accessibility, usability, and integrity

# The research data lifecycle



Courtesy: Leadership Council for Digital Research Infrastructure (LCDRI) Data Management Position Paper, 2017-08-31



# Research Data Management

is the active organization & maintenance of data throughout the **research data lifecycle** to ensure its security, accessibility, usability, and integrity

# Applying RDM best practices will benefit...

## **Researchers and their collaborators**

- ✧ Improves research efficiency and productivity
- ✧ Provides extra credit for research work
- ✧ Increases research impact
- ✧ May help to meet funding requirements

## **Research Communities**

- ✧ Accelerates discovery
- ✧ Enables validation and verification

## **Funders, governments and the public**

- ✧ Improves return on investment
- ✧ Increases research transparency
- ✧ Data as a public good

# Requirements and Expectations

## Funding Agencies and Governments



## Publishers and Research Communities





### Partnerships and Projects

- Meet Canada's scientists and professionals
- Canadian High Arctic Research Station
- Northern Contaminants Program
- Canada's Extended Continental Shelf Program
- Health Canada: Centre for Vaccine Evaluation
- Science and Technology Data
- Research Funding Collaboration
  - Co-operative Funding Programs
  - Policies and Guidelines
    - Open Access
    - Selecting the Appropriate Federal Granting Agency
    - Research Ethics
    - Responsible Conduct of Research
    - Public Communications

## Tri-Agency Statement of Principles on Digital Data Management

### 1. Preamble

The [Canadian Institutes of Health Research \(CIHR\)](#), the [Natural Sciences and Engineering Research Council of Canada \(NSERC\)](#), and the [Social Sciences and Humanities Research Council of Canada \(SSHRC\)](#) (the agencies) are federal granting agencies that promote and support research, research training, knowledge transfer and innovation within Canada.

As publicly funded organizations, the agencies are strong advocates for making the results of the research they fund as accessible as possible. In promoting access to research results, they aspire to advance knowledge, avoid research duplication and encourage reuse, maximize research benefits to Canadians and showcase the accomplishments of Canadian researchers. These aspirations align with the Government of Canada's commitment to open science, as described in [Seizing Canada's Moment: Moving Forward in Science, Technology and Innovation](#)(2014).

Research data include observations about the world that are used as primary sources to support scientific and technical inquiry, scholarship and research-creation, and as evidence in the research process.<sup>1</sup> Research data are gathered through a variety of methods, including experimentation, analysis, sampling and repurposing of existing data. They are increasingly produced or translated into digital formats. When properly managed and responsibly shared, these digital resources enable researchers to ask new questions, pursue novel research programs, test alternative hypotheses, deploy innovative methodologies and collaborate across geographic and disciplinary boundaries. The ability to store, access, reuse and build upon digital research data has become critical to the advancement of science and scholarship, supports innovative solutions to economic and social challenges, and holds tremendous potential for Canada's productivity, competitiveness and quality of life.

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: **Expectations**

- Data management planning
- Constraints and obligations
- Adherence to standards
- Collection and storage
- Metadata
- Preservation, retention and sharing
- Timeliness
- Acknowledgement and citation
- Efficient and cost-effective

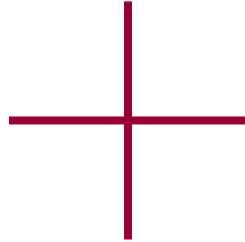
# Tri-Agency Statement of Principles on Digital Data Management: **Responsibilities**

## Researchers

- incorporating best practices
- developing DMPs
- adhering to policies and standards

## Research Communities

- developing & promoting standards
- fostering excellence
- selecting repositories



## Research Institutions

- supporting best practices
- providing access to resources
- creating guidance and policies

## Research Funders

- developing policies & guidance
- promoting data management
- providing peer reviewers

# Draft Tri-Agency RDM Policy

Tri-Agency [draft data management policy](#)<sup>1</sup> — June, 2018

- ✧ Applies to grant recipients and institutions administering tri-agency funds.
- ✧ June-Sep, 2018: consultation period; feedback to inform policy

3 pillars:

1. **Institutions:** Institutional Strategy
2. **Researchers:** Data Management Plans
3. **Researchers:** Data Deposit

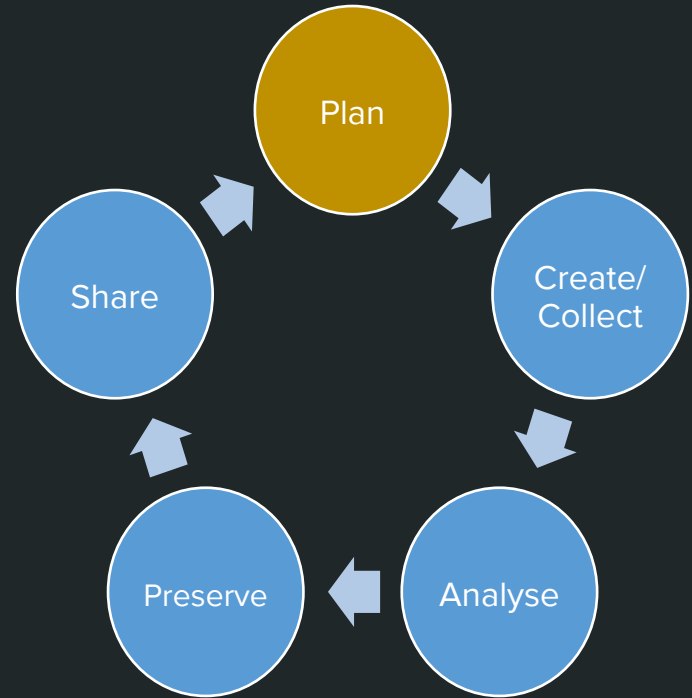
Phased and incremental implementation

<sup>1</sup>[http://www.science.gc.ca/eic/site/063.nsf/eng/h\\_97610.html](http://www.science.gc.ca/eic/site/063.nsf/eng/h_97610.html) [shortened: <https://bit.ly/2OsJB8w>]

# Planning for RDM

---

- Considerations
- DMP Assistant





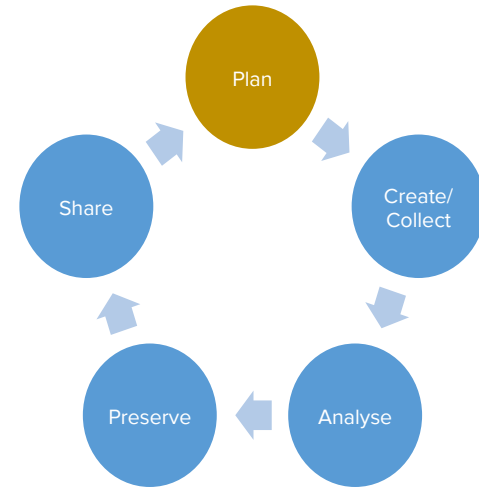
## ...it all starts with a (good) plan

A research **Data Management Plan (DMP)** should:

- ✧ Describe how you will manage data through all stages of your research
- ✧ Communicate a strategy for creating share-worthy data products

A **good** DMP will:

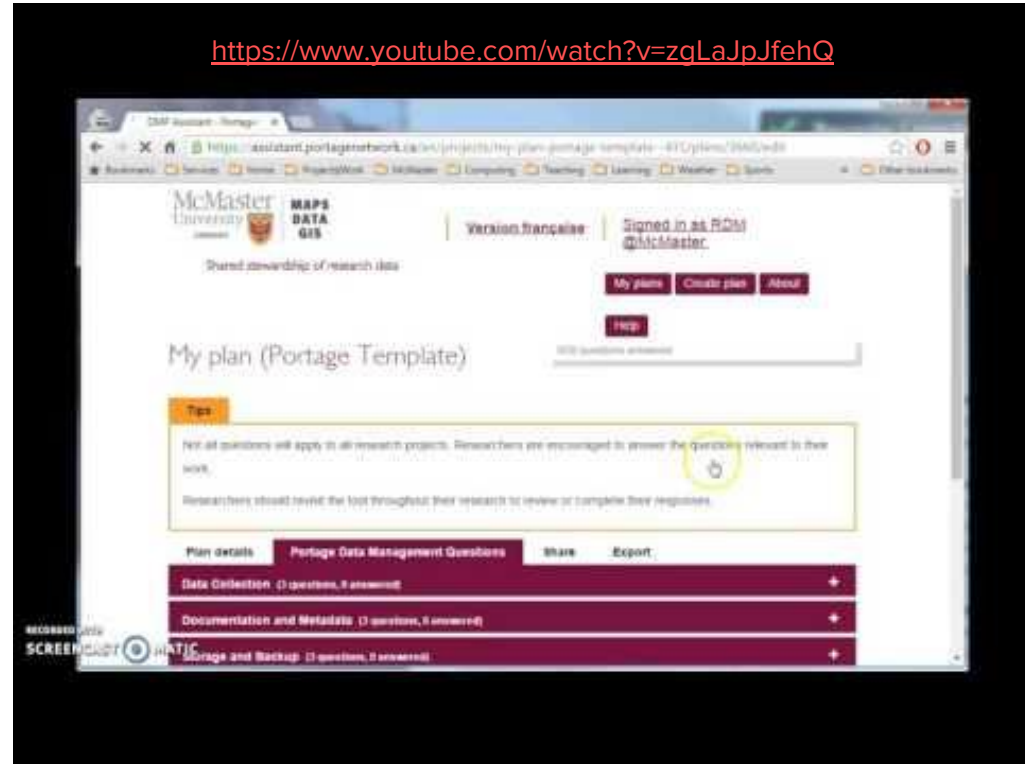
- ✧ Be completed at the time of study design
- ✧ Ensure compliance with policies / obligations
- ✧ Document and organize research activities
- ✧ Help identify support requirements
- ✧ *(Likely)* evolve with your study...



# Portage DMP Assistant

- ✧ A web-based, bilingual data management planning tool.
- ✧ Available to all researchers in Canada.
- ✧ A guide for best practices in data stewardship.
- ✧ Exportable data management plans.

<https://www.youtube.com/watch?v=zgLaJpJfehQ>



<https://assistant.portagenetwork.ca/>

# Activity - Develop a DMP (~15 minutes)

Using the **CARL Portage DMP Assistant**:



- ✦ Navigate to [assistant.portagenetwork.ca](https://assistant.portagenetwork.ca)
- ✦ Create an account & login (check email for verification)
  - Or login with user: **mcm.rdm.demo@gmail.com** | pass: **demopass**
- ✦ Develop a DMP for a current, upcoming or prospective research project
- ✦ Browse through the DMP questions
- ✦ Complete the answerable questions in rough form (or internally)
- ✦ Note any confusing, problematic or difficult questions

# Navigate to the DMP Assistant:

[assistant.portagenetwork.ca](https://assistant.portagenetwork.ca)

Sign in with your email / password



Your account was successfully confirmed. Please sign in.

DMP Assistant is a bilingual tool for preparing data management plans (DMPs). The tool follows best practices in data stewardship and walks researchers step-by-step through key questions about data management.

**Step 1** Sign up with DMP Assistant

**Step 2** Sign in and select a template under Organizations. The Portage template is the default.

**Step 3** Answer the questions that are relevant to your work. Guidance and examples are provided.

**Step 4** Revisit the tool throughout your research to review or revise your

### Sign in

If you have an existing account with DMP Assistant or previous version of DMP Builder.

  
  
[Forgot your password?](#)

Remember me

Campus ID single sign-in authentication coming soon!

Click “Create plan”



Set organization to “McMaster University”



### Create a new plan

Please select from the following drop-downs so we can determine what questions and guidance should be displayed in your plan.

If you aren't responding to specific requirements from a funder or an institution, you can choose the **Portage Data Stewardship Template**. The Portage Data Stewardship Template is based on internationally accepted standards and best practices. It has been prepared and is maintained by a group of research data management experts from research libraries across Canada.

To see institutional questions and/or guidance, select your organization.

You may leave blank or select a different organization to your own. If you leave blank, default Portage DMP template will be [Not applicable/not listed](#), used.

Choose a template

There are a number of possible templates you could use. Please choose one.

Tick to select any other sources of guidance you wish to see.

McMaster

Select the “Portage” template



[assistant.portagenetwork.ca](http://assistant.portagenetwork.ca)

Enter relevant personal, organizational and study information



Expand boxes to complete questions for each section



Provide answers in the left-hand box  
Use right-hand box for guidance/comments



The screenshot displays the 'My plan (Portage Template)' interface. At the top, there are tabs for 'Plan details', 'Portage Data Management Questions', 'Share', and 'Export'. Below the tabs, a form is shown with fields for 'Plan name', 'ID', 'Grant number', 'Principal Investigator/Researcher', 'Principal Investigator/Researcher ID', 'Plan data contact', and 'Description'. A 'Save' button is visible at the bottom right of the form.

Below the form, there is a section titled 'This plan is based on:' followed by a list of sections: 'Data Collection (0 questions, 0 answered)', 'Documentation and Metadata (1 question, 0 answered)', 'Storage and Backup (1 question, 0 answered)', 'Preservation (2 questions, 0 answered)', 'Sharing and Reuse (1 question, 0 answered)', and 'Responsibilities and Resources (3 questions, 0 answered)'. Each section has a plus sign to its right, indicating it can be expanded.

Below the list, there is a 'Tips' section with a yellow background. The text reads: 'Not all questions will apply to all research projects. Researchers are encouraged to answer the questions relevant to their work. Researchers should revisit the tool throughout their research to review or complete their responses.'

Below the tips, there is a 'Portage Data Management Questions' section. It has tabs for 'Plan details', 'Portage Data Management Questions', 'Share', and 'Export'. The 'Portage Data Management Questions' tab is active. Below the tabs, there is a section titled 'Data Collection (0 questions, 0 answered)'. The text reads: 'This section addresses data collection issues such as data types, file formats, naming conventions, and data organisation – factors that will improve the usability of your data and contribute to the success of your project.'

Below the text, there is a text input area with a rich text editor toolbar. The toolbar includes buttons for bold, italic, underline, strikethrough, bulleted list, numbered list, link, unlink, and table. Below the input area, there is a 'Save' button.

On the right side of the 'Data Collection' section, there is a 'Guidance' box with a yellow background. It has a title 'Portage Guidance' and a plus sign to its right. The text inside reads: 'Examples: numeric, images, audio, video, text, tabular data, modeling data, spatial data, instrumentation data.'

# Reflections and Discussion

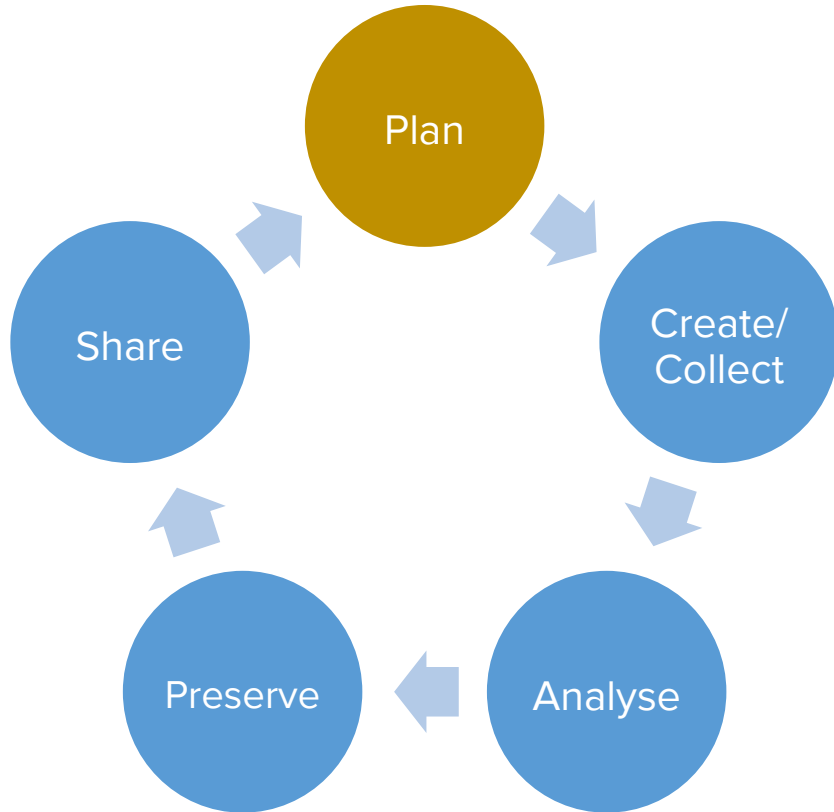
## **What are your reflections on:**

- ✧ The overall experience?
- ✧ The interface and its content?

What sections / questions were particularly challenging? Howso?

What kind of support or resources would help you complete this plan?

# Planning



What are the stipulations in institutional, funder or publisher data policies to be followed?

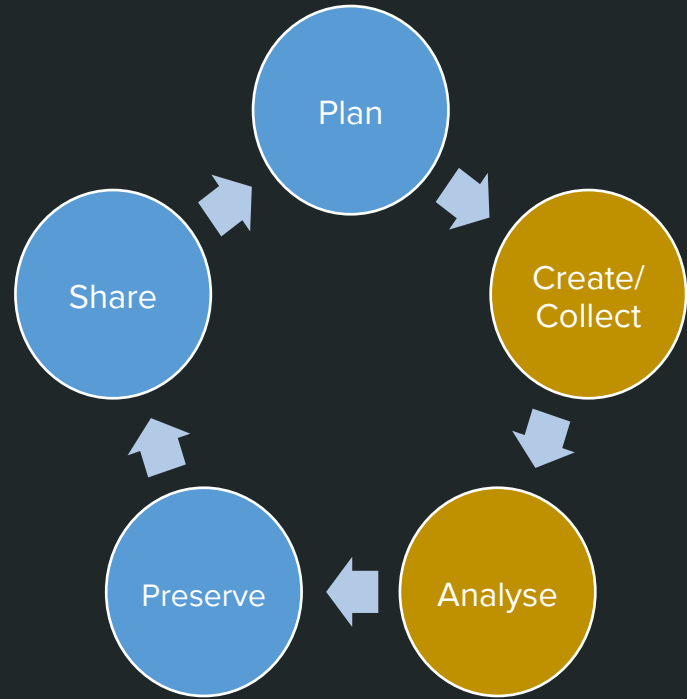
What resources do you require to manage your data?

Who is responsible for data management and long-term stewardship?

# In-project management

---

- Storage & backup solutions
- Managing files
- Managing collaboration



# Things to consider...

What types of data (and how much) will you collect?

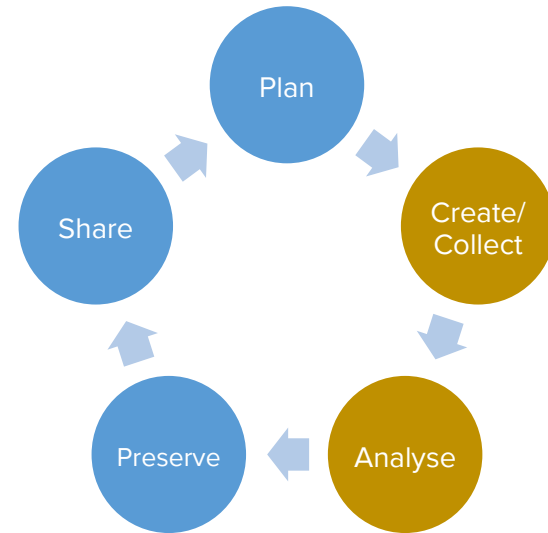
How will you **organize, secure, and backup** your data?

Are there **ethical** or commercial restrictions?

How will you **describe** your data so that others understand it?

How will you control **access** to your data?

How will you manage data versions?



# Quiz time!

Go to **menti.com** and use the code: **45 20 33**

# Portage Network's "Good Enough RDM"

<https://bit.ly/2zv54Eg>

## "Good Enough" Research Data Management

*(a brief guide for busy people)*

This brief guide presents a set of good data management practices that researchers can adopt, regardless of their data management skills and levels of expertise.

1

### Save your raw data in original format

- 1.1 Don't overwrite your original data with a cleaned version.
- 1.2 Protect your original data by locking them or making them read-only.
- 1.3 Refer to this original data if things go wrong (as they often do).

2

### Backup your data

- 2.1 Use the 3-2-1 rule: Save three copies of your data, on two different storage mediums, and one copy off site.
- 2.2 Do not backup or store sensitive data on a commercial cloud (Dropbox, Google Drive, etc.).

3

### Describe your data

- 3.1 **Machine Friendly:** Describe your dataset with a metadata standard for discovery.
- 3.2 **Human Friendly:** Describe your variables, so your colleagues will understand what you meant. Data without good metadata is useless. Give your variables clear names.
- 3.3 Do not leave cells blank - use numeric values clearly out of range to define missing (e.g. '99999') or not applicable (e.g. '88888') data, and describe these in your data dictionary.
- 3.4 Convert your data to open, non-proprietary formats.
- 3.5 Name your files well with basic metadata in file names.

4

### Process your data

- 4.1 Make each column a variable.
- 4.2 Make each row an observation.
- 4.3 Store units (e.g. kg or cm) as metadata (in their own column).
- 4.4 Document each step processing your data in a README file.

5

### Archive and preserve your data

- 5.1 Submit final data files to a repository assigning a persistent identifier (e.g. handles or DOIs).
- 5.2 Provide good metadata for your study so others could find it (use your discipline's metadata standard, e.g. Darwin Core, DDI, etc.).

# 1

## ***Save your raw data in original format***

- 1.1 Don't overwrite your original data with a cleaned version.
- 1.2 Protect your original data by locking them or making them read-only.
- 1.3 Refer to this original data if things go wrong (as they often do).



## 2

### *Backup your data*

- 2.1 **Use the 3-2-1 rule:** Save three copies of your data, on two different storage mediums, and one copy off site.
- 2.2 Do not backup or store sensitive data on a commercial cloud (Dropbox, Google Drive, etc.).

# 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 (seafile): <https://macdrive.mcmaster.ca>
- MacDrop
- Dropbox, Google Drive, etc.

# Considerations for “Archived” Backup

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

321

	RHPCS - Backup Services	RHPCS - Hosted Server Packages	MacDrive	Microsoft OneDrive / Teams	MacDrop
<b>Storage Quota</b>	1 TB; more available for a fee	1 TB; more available for fee	300 GB per account	1 TB per account; up to 5 TB by request	2 GB for free; more available for a fee.
<b>Rates / cost</b>	\$500 / yr + one time set up fee (\$125 / machine) Additional space: \$300 / TB Restore services: \$125 / hour	\$500 - \$4000 / yr Setup fee: \$500 - \$1000 Additional space: \$450 / TB	No cost to users	No cost to users	\$10/month for 100GB
<b>Backups / versioning</b>	Nightly, 14-day rotating cycle; Restore services through RHPCS	Nightly, 14-day rotating cycle; Restore services through RHPCS Nextcloud sync service available.	Ongoing real-time sync 4-month version history Full Library restore through UTS	Ongoing real-time sync Unlimited version history (?)	Ongoing real-time sync Versions are kept indefinitely so long as versions don't exceed 50% of user's quota
<b>Who can use this service?</b>	Any subscribing users or research group	Any subscribing users or research group	McMaster Faculty and Staff Graduate students can obtain zero-quota accounts	All McMaster faculty, staff and students	All FHS users
<b>Server location</b>	A.B. Bourns building	A.B. Bourns building	Replicated clusters in Gilmour Hall and JHE	<b>OneDrive:</b> Canadian servers <b>Teams:</b> Soon in Canadian servers only	Health Science Centre
<b>Other notes</b>			Supports encrypted libraries, file and directory sharing, Desktop client, web interface	Supports file and directory sharing, Desktop client, web interface	
<b>More info</b>	<a href="https://rhpcs.mcmaster.ca/current-rates">rhpcs.mcmaster.ca/current-rates</a>	<a href="https://rhpcs.mcmaster.ca/current-rates">rhpcs.mcmaster.ca/current-rates</a>	<a href="https://macdrive.mcmaster.ca/">macdrive.mcmaster.ca/</a> Documentation: <a href="https://goo.gl/AvRGwx">https://goo.gl/AvRGwx</a>	<a href="https://portal.office.com/">portal.office.com/</a> Documentation: <a href="https://mcmaster.ca/uts/licensing">mcmaster.ca/uts/licensing</a>	<a href="https://www.csu.mcmaster.ca/service-catalogue/network-and-communications/macdrop">https://www.csu.mcmaster.ca/service-catalogue/network-and-communications/macdrop</a>

Access the matrix: [goo.gl/45iy38](https://goo.gl/45iy38)

# 3

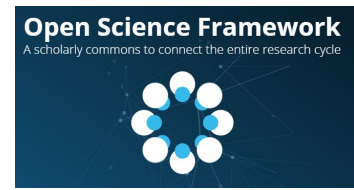
## *Describe your data*

- 3.1 **Machine Friendly:** Describe your dataset with a metadata standard for discovery.
- 3.2 **Human Friendly:** Describe your variables, so your colleagues will understand what you meant. Data without good metadata is useless. Give your variables clear names.

# 4

## *Process your data*

- 4.1 Make each column a variable.
- 4.2 Make each row an observation.
- 4.3 Store units (e.g. kg or cm) as metadata (in their own column).
- 4.4 Document each step processing your data in a README file.



# Open Science Framework

<https://osf.io/>

- ✧ Free, open source web application
- ✧ Connects and supports the research workflow
- ✧ Integrates with existing services (e.g. Google Drive, Dropbox, etc.)
- ✧ Supports collaborative workflows
- ✧ Facilitates data publishing to selected data repositories

<https://youtu.be/2TV21qOzfhw>

GETTING STARTED WITH THE  
OPEN SCIENCE FRAMEWORK  
(OSF)

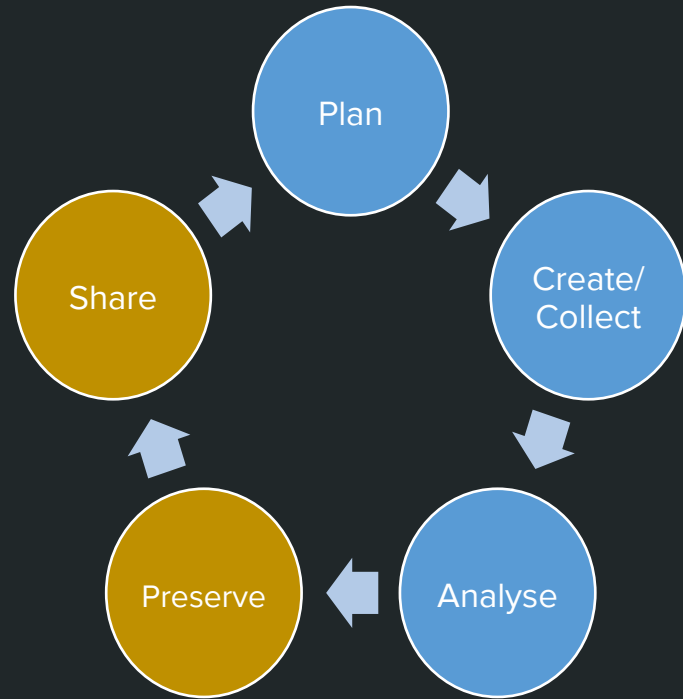
Workshop demo: <https://osf.io/muq5f>



# Archiving and Sharing

---

- FAIR principles
- Data repositories
- Scholars Portal Dataverse



# 5

## *Archive and preserve your data*

- 5.1 Submit final data files to a repository assigning a persistent identifier (e.g. handles or DOIs).
- 5.2 Provide good metadata for your study so others could find it (use your discipline's metadata standard, e.g. Darwin Core, DDI, etc.).

# Things to consider...

How will your data products be stored in the long-term?

- ✧ How to ensure that it remains *integral* and *secure*?
- ✧ Who will assume long-term *responsibility* for your data?

How will others access your data products?

- ✧ What data (if any) can/should be shared? Who should have access?
- ✧ How will you manage legal, commercial & ethical constraints?

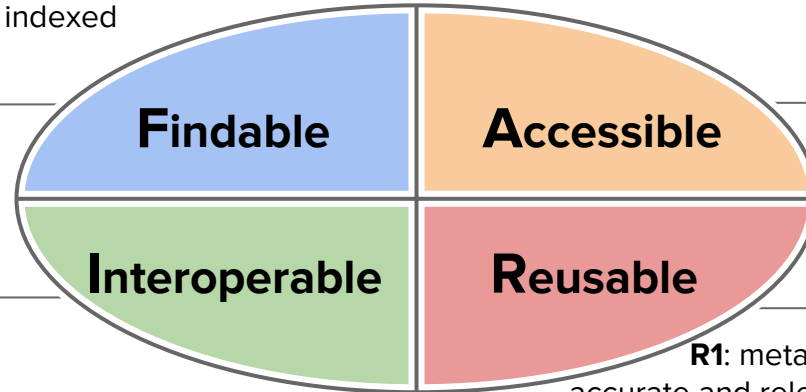
How to maximize credit for sharing your data?

- ✧ In which repository should you deposit your data?
- ✧ How to ensure that your data is **FAIR**  
(*findable, accessible, interoperable and reusable*)?

# The FAIR Guiding Principles

**F1:** (meta)data have a globally unique and eternally persistent identifier  
**F2:** data are described with rich metadata  
**F3:** metadata clearly and explicitly includes the ID of the data it defines  
**F4:** (meta)data are registered and indexed in a searchable resource

**A1:** (meta)data retrievable by their ID using a standardized protocol  
**A1.1:** protocol is open, free and universally implementable  
**A1.2:** protocol allows for AuthT/ AuthZ where needed  
**A2:** metadata is always accessible



**I1:** (meta)data use a formal, accessible, shared, broadly applicable language for knowledge rep.  
**I2:** (meta)data use vocabularies that follow FAIR principles  
**I3:** (meta)data include qualified references to other (meta)data

**R1:** meta(data) richly described with accurate and relevant attributes  
**R2:** (meta)data released with a clear and accessible data usage license  
**R3:** (meta)data associated with detailed provenance  
**R4:** (meta)data meet domain-relevant community standards

# Domain vs. General Repositories

**Domain-specific Repositories:** Archives of digital and/or digitized information related to an area of research. They can be broad in scope, such as ICPSR at University of Michigan, or more tightly focused, such as Milkulski Archive at Space Telescope (MAST) at the Space Telescope Science Institute.

**General Repositories:** The archives of digital and/or digitized information is not related to an area of research and hence it is general in nature. Scholars Portal Dataverse is an example of general repository.

re3data.org

Registry of research data repositories

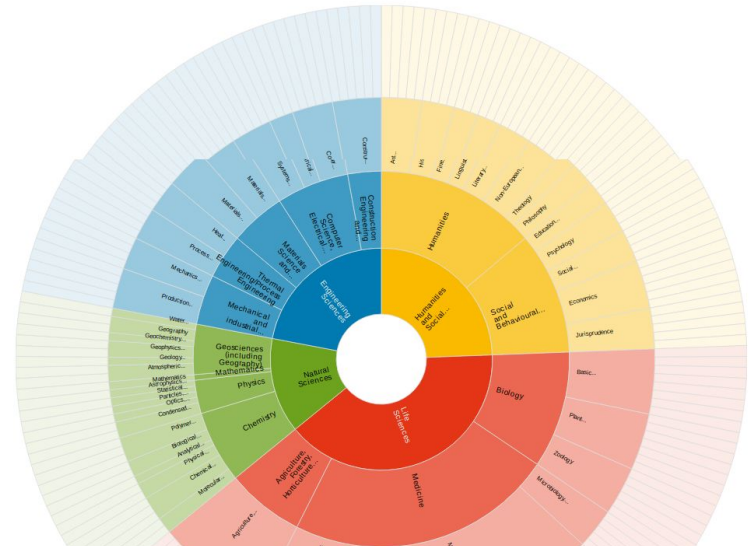
re3data.org  
REGISTRY OF RESEARCH DATA REPOSITORIES

Home Search Browse Suggest FAQ About Schema API Contact Legal notice / Impressum

### Browse by subject

Graphical  Text

click to zoom into subjects or to select a bottommost subject in the hierarchy as filter for the re3data search page  
ctrl + click on a top subject to select it as filter

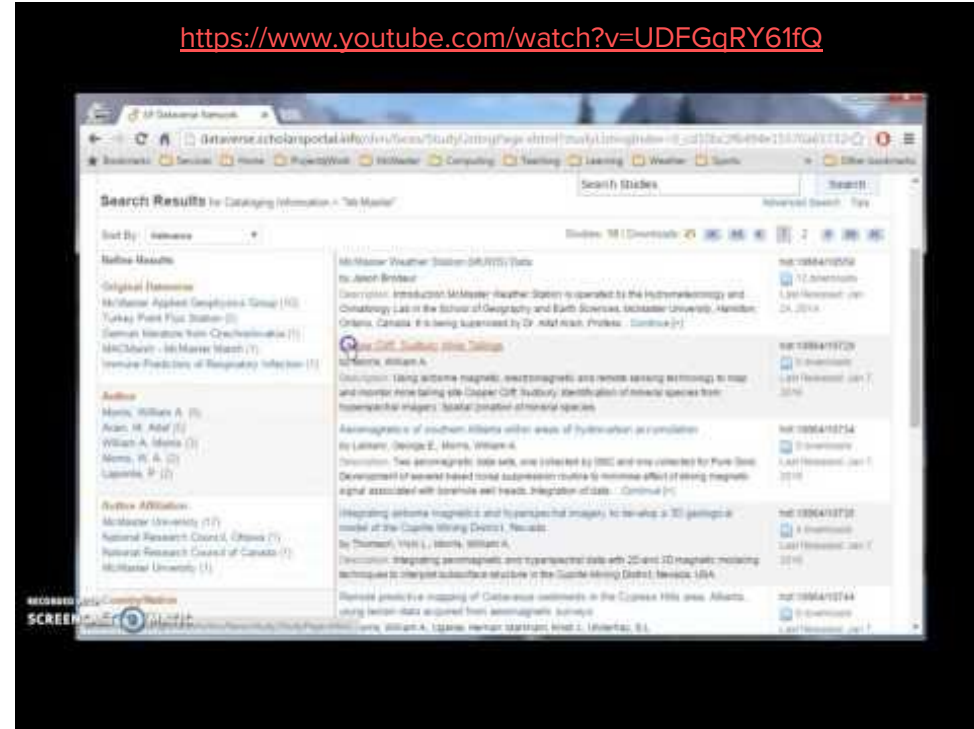


# Scholars Portal Dataverse



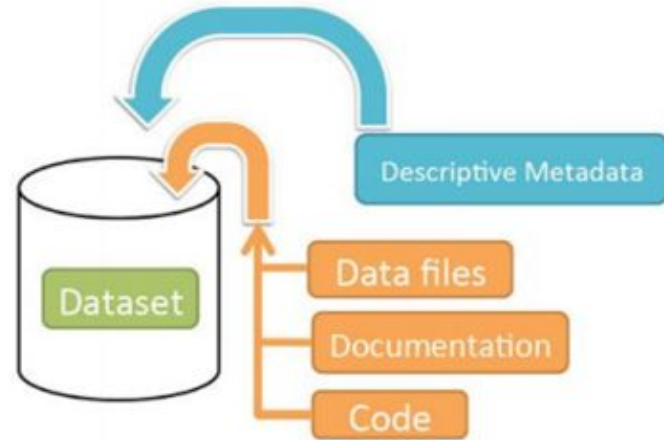
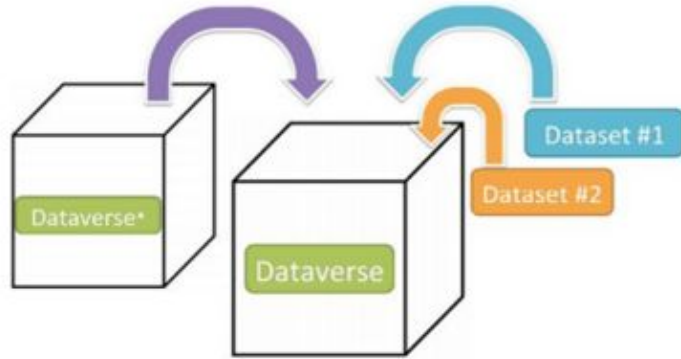
- ✧ A data repository for researchers at Ontario's universities -- **free** and **open** for all researchers in Canada
- ✧ An online platform to share, preserve, cite, explore and analyze research data.
- ✧ Allows researchers to control how they share their data.
- ✧ Supports data DOI registration through Datacite Canada.

<https://www.youtube.com/watch?v=UDFGqRY61fQ>



<http://dataverse.scholarsportal.info>

# Dataverses, Datasets, Files





# Create your own (demo) Dataverse

- Scholars Portal Demo Dataverse site:  
<https://demodv.scholarsportal.info>

# Sign-up & Institutional Affiliation

Scholars Portal Dataserve   [About](#) [User Guide](#) [Support](#) [Sign Up](#) [Log In](#) Fr (Beta)

Scholars Portal Dataserve > [Create Account](#)

**Why have a Dataserve account?** To create your own dataserve and customize it, add datasets, or request access to restricted files.

Account Information

**Username \*** Create a valid username of 2 to 60 characters in length containing letters (a-z), numbers (0-9), dashes (-), underscores (\_), and periods (.).

**Password \*** Create a password that is maximum six characters long and uses at least one letter and number.

**Retype Password \***

**Given Name \***

**Family Name \***

**Email \***

**Affiliation \***

**Position**

**General Terms of Use \***

**Scholars Portal Dataserve Network Account Terms of Use**

**Acceptance of Terms**

The following terms and conditions govern all use of the Scholars Portal Dataserve website (the Site) and the services available on or at the Site (taken together, the Service). The Service is operated by the Ontario Council of University Libraries (OCUL). The Service is offered subject to acceptance without modification of all of the terms and conditions contained herein (the Terms of Use) and all other operating rules, policies and procedures that may be published from

**Email \***

**Affiliation \***

**Position**

**General Terms of Use \***

The following terms and conditions govern all use of the Scholars Portal Dataserve website (the Site) and the services available on or at the Site (taken together, the Service). The Service is operated by the Ontario Council of University Libraries (OCUL). The Service is offered subject to acceptance without modification of all of the terms and conditions contained herein (the Terms of Use) and all other operating rules, policies and procedures that may be published from

Scholars Portal Dataserve   [About](#) [User Guide](#) [Support](#) [Sign Up](#) [Log In](#) Fr (Beta)

**Ryerson University**

Ryerson University Dataserve

Scholars Portal Dataserve - Ryerson University Dataserve

[Advanced Search](#)

**1 to 4 of 4 Results**

International Meeting Dictionary  
Jan 11, 2017 - Royal Holloway, Uxbridge, UK

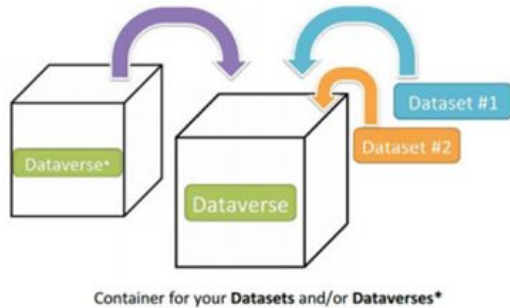
Book - K. Reed, David, Steady, 2017 "International Meeting Dictionary" Jan 11, 2017 (PDF) 2017, 2017  
100% OCUL member-owned/controlled

The Dataserve was developed/created for the following reasons: an update to an English monograph series (MOS) on available with an associated user interface and a copyright transfer of user work and alterations (e.g., PDF, HTML, etc.) stored for the original MOS.

Book - International Meeting Dictionary  
Jan 11, 2017

**Author Name**  
View Full Text

# Institutional Dataverse



Container for your Datasets and/or Dataverses\*

Screenshot of the Scholars Portal Dataverse interface. The page displays search results for datasets, including a list of datasets and their associated metadata. The search results show two entries for 'LaurelCreek\_Water\_Biological\_2017' and 'LaurelCreek\_Water\_Biological\_2011', both dated Jan 29, 2019, and associated with the University of Waterloo Ecology Lab. The description for both entries is: 'Students of a Field Ecology course conducted analysis of water from Laurel Creek at 11 different locations on the University of Waterloo campus, ON, Can. Water was analyzed for benthic invertebrates.'

You can have as many Dataverses within your Institution Dataverse as you like!

# Customize your Dataverse

- Customizable features:
  - Branding/logo
  - Featured Dataverses
  - Facets
  - Contact

The screenshot displays the 'Scholars Portal Dataverse' interface. At the top, there is a search bar and navigation links. The main header area features the Queen's University logo and the text 'Queen's University Dataverse', which is circled in red. To the right of this header, the text 'Custom space -slogan, links, logo, etc.' is overlaid. Below the header, two featured dataverse tiles are shown, also circled in red. The first tile is 'Queen's University Library Research Data Archive Dataverse' and the second is 'Queen's University Biological Station Data Archive Dataverse'. To the right of these tiles, the text 'Featured Dataverses' is overlaid. The bottom section of the interface shows a search bar for the current dataverse, a list of facets (Dataverses: 8, Datasets: 71, Files: 753), and a list of search results, including a result for 'Research Data Management (RDM) Survey of Queen's University's Faculty of Health Sciences'.

# Linking Datasets and Dataverses



Carleton University Dataverse

Root Dataverse > Carleton University Dataverse

Search this dataverse...

**Dataverses (3)**

**Datasets (4)**

**Files (3)**

**Publication Status**

Unpublished (6)

1 to 7 of 7

Sub-test



### Link Dataverse

Choose which of your dataverses you would like to link this dataverse to.

Your Dataverses

- Botanical Test Dataverse
- Aileen Liu Dataverse
- Algoma University Dataverse
- Andrea Kosavic Dataverse
- Audrey Tam Dataverse
- Botanical Test Dataverse**
- Brandon University
- Brewster's Dataverse
- Brock University Dataverse
- Camilla Gilis-Adelman Dataverse
- Carleton Test Dataverse
- Child and Youth Care Dataverse
- Collaboration Dataverse
- Dan Jakubek Dataverse
- Dataception
- Define the Square Dataverse
- Depths of the Dataverse
- Dirty Dan Dataverse
- Elias Vitari Dataverse
- French testing Dataverse
- Gaby Wynn Dataverse

Save Linked Dataverse

Contact Share **Link** Edit

Link Dataverse  
Link Search  
Add Data

Sort

# Self-service Deposit Model

- Anyone with an account can publish to Dataverse
- SP Terms of Use covers removal of data if necessary

## Scholars Portal Dataverse Account Terms of Use

### Acceptance of Terms

The following terms and conditions govern all use of the Scholars Portal Dataverse website (the Site) and the services available on or at the Site (taken together, the Service). The Service is operated by the Ontario Council of University Libraries (OCUL). The Service is offered subject to acceptance without modification of all of the terms and conditions contained herein (the Terms of Use) and all other operating rules, policies and procedures that may be published from time to time on this Site by OCUL.

The Service is available only to individuals who are at least 13 years old or possess legal parental or guardian consent. If you do not so qualify, do not attempt to use the Service. OCUL may refuse to offer the Service to any person or entity at any time and may change its eligibility criteria, at any time, in its sole discretion. *Modification of Terms of Use*

OCUL reserves the right, at its sole discretion, to modify or replace any of the Terms of Use at any time. It is your responsibility to check the Terms of Use periodically for changes. Your continued use of the Service following the posting of any changes to the Terms of Use constitutes acceptance of those changes.

More...

### Rules and Conduct

As a condition of use, you promise not to use the Service for any purpose that is prohibited by the Terms of Use. For purposes of the Terms of Use, "Content" includes, without limitation, any information, data, text, software, scripts, graphics, and interactive features generated, provided, or otherwise made accessible by OCUL, or its partners on or through the Service.

By way of example, and not as a limitation, you shall not (or permit others to) either (a) take any action or (b) upload, download, post, submit or otherwise distribute or facilitate distribution of any content using any communications service or other service available on or through the Service, that:

- infringes any patent, trademark, trade secret, copyright, right of publicity or other right of any other person or entity;
- is unlawful, threatening, abusive, harassing, defamatory, libelous, deceptive, fraudulent, invasive of another's privacy, tortious, obscene, offensive, or profane;
- constitutes unauthorized or unsolicited advertising, junk or bulk e-mail ("spamming");
- contains software viruses or any other computer codes, files, or programs that are designed or intended to disrupt, damage, limit or interfere with the proper function of any software, hardware, or telecommunications equipment or to damage or obtain unauthorized access to any system, data or other information of Scholars Portal or any third party

<http://guides.scholarsportal.info/dataverse>

# Storage Size Limits

- Currently SP Dataverse supports
  - 2 GB file size upload
  - Will be increased to 10 GB in the near future
  - Unzipping, tabular ingest, processing intense

# Dataset Management / File Versioning



Canadian Media Concentration Research Project Dataverse (Carleton University)

CMCRP Project Site

Scholars Portal Dataverse > Carleton University Dataverse > Canadian Media Concentration Research Project Dataverse >

**The Growth of the Network Media Economy in Canada, 1984-2016**

Metrics

3 Downloads

Contact Share



**The Growth of the Network Media Economy in Canada, 1984-2016** Version 2.0

Winseck, Dwayne, 2017, "The Growth of the Network Media Economy in Canada, 1984-2016",  
doi:10.5683/SP/15ZHPR, Scholars Portal Dataverse, V2

Cite Dataset

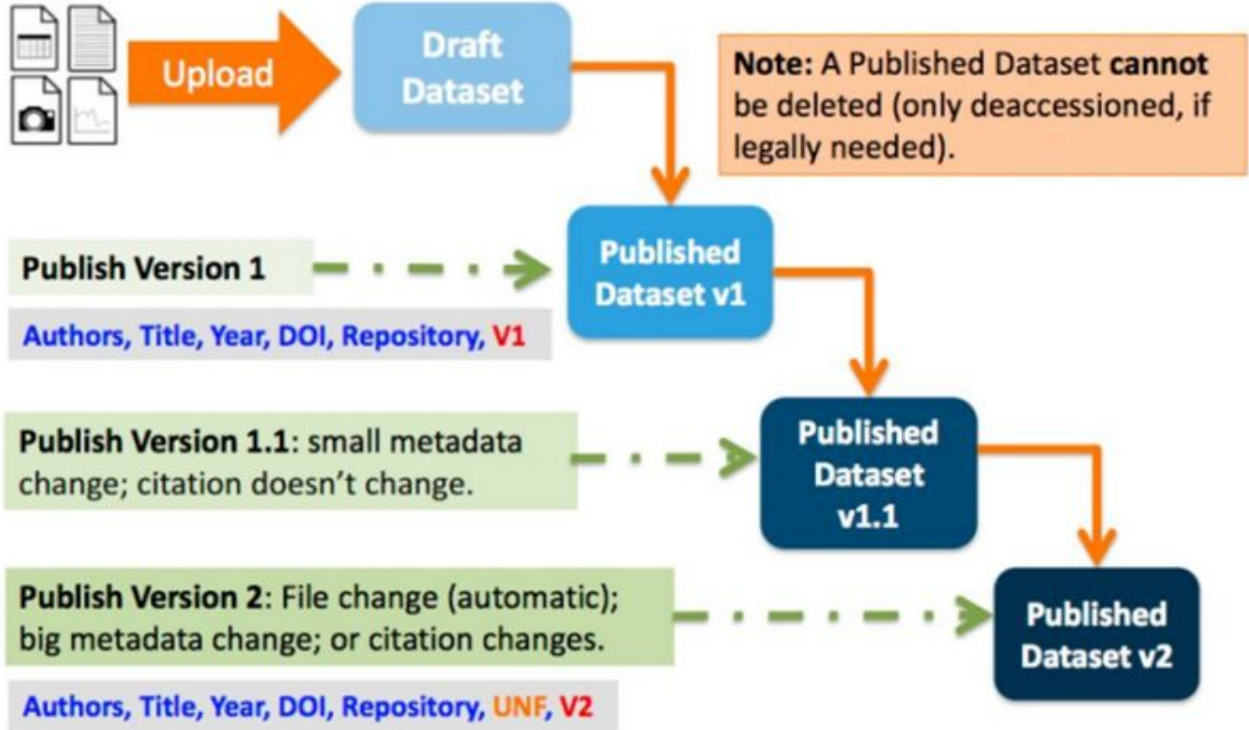
Learn about Data Citation Standards.

Files Metadata Terms **Versions**

Dataset	Summary	Contributors	Published
2.0	<b>Files (Added: 2);</b> View Details	MacOdrum Library Cataloguing and Collections	December 13, 2017
1.0	This is the first published version.	MacOdrum Library Cataloguing and Collections, MacOdrum Library	December 7, 2017



# Dataset Management / File Versioning



# Dataset and File Permissions

- User/groups and roles
  - Assign permissions for collaborators, curators, file downloaders (access)
  - Granular file-level permissions
  - IP Group based permissions

The screenshot shows the 'Dataset Permissions' interface for the 'Canadian Media Concentration Research Project Database'. A menu is open over the 'Permissions' option, listing actions like 'Files (Upload)', 'Metadata', 'Terms', 'Permissions', 'Private URL', and 'Thumbnails + Widgets'. Below the menu, a table lists the users and groups that have access to the dataset.

User/Group Name (Affiliation)	ID	Role	Action
CMCR Project (Carleton University)	@cmcrp	Admin	Role assigned at Canadian Media Concentration Research Project Database
George Dumovich (Carleton University, MacOdrum Library)	@pdmsr	Admin	Role assigned at Canadian Media Concentration Research Project Database
MacOdrum Library (Carleton University)	@cudata	Contributor	<input checked="" type="checkbox"/> Remove Assigned Role
MacOdrum Library (Carleton University)	@cudata	Curator	Role assigned at Canadian Media Concentration Research Project Database
MacOdrum Library Cataloguing and Collections (Carleton University)	@cucat	Curator	Role assigned at Canadian Media Concentration Research Project Database

# Licensing


- Default to CC0 (Open Data)
- OR
- Custom “Data Usage Agreement”

**CC0 1.0 Universal (CC0 1.0)  
Public Domain Dedication**


This is a human-readable summary of the [Legal Code \(read the full text\)](#)

[Disclaimer](#)

**No Copyright**



The person who associated a work with this deed has **dedicated** the work to the public domain by waiving all of his or her rights to the work worldwide under copyright law, including all related and neighboring rights, to the extent allowed by law.



You can copy, modify, distribute and perform the work, even for commercial purposes, all without asking permission. See [Other Information](#) below.

## Sample Data Usage Agreement

*(Note: This is a sample DUA for datasets that have de-identified human subject data)*

This is an agreement (“Agreement”) between you the downloader (“Downloader”) and the owner of the materials (“User”) governing the use of the materials (“Materials”) to be downloaded.

### I. Acceptance of this Agreement

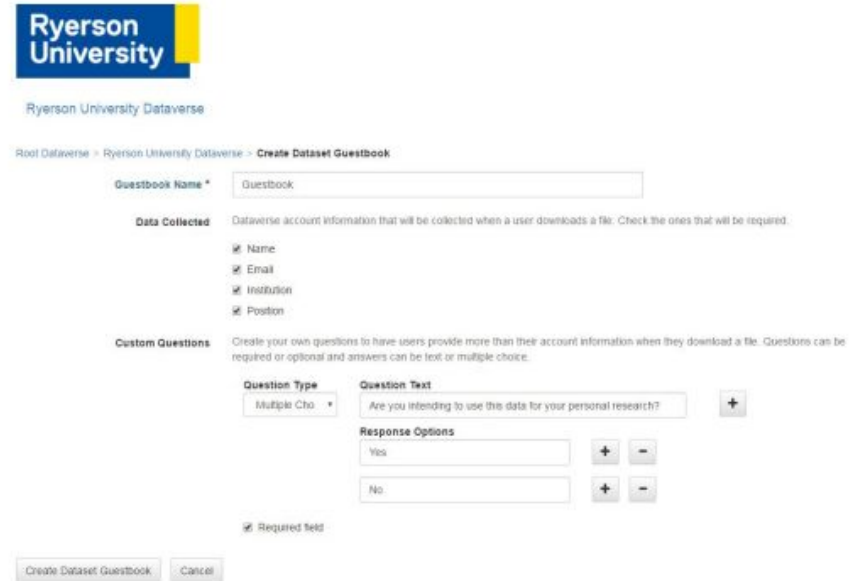
By downloading or otherwise accessing the Materials, Downloader represents his/her acceptance of the terms of this Agreement.

### II. Modification of this Agreement

Users may modify the terms of this Agreement at any time. However, any modifications to this Agreement will only be effective for downloads subsequent to such modification. No modifications will supersede any previous terms that were in effect at the time of the Downloader’s download.

# Guestbooks

- Who is downloading my data?
  - User fills out guestbook form
- Owner downloads Guestbook report in Excel format
- Can be used to mediate access / approval of access to data



**Ryerson University**

Ryerson University Dataverse

Root Dataverse > Ryerson University Dataverse > **Create Dataset Guestbook**

**Guestbook Name \***

**Data Collected** Dataverse account information that will be collected when a user downloads a file. Check the ones that will be required.

- Name
- Email
- Institution
- Position

**Custom Questions** Create your own questions to have users provide more than their account information when they download a file. Questions can be required or optional and answers can be text or multiple choice.

Question Type	Question Text	
Multiple Cho.	Are you intending to use this data for your personal research?	+
	Response Options	
	Yes	+ -
	No	+ -

Required field

# Metadata Standards

- DataCite, Dublin Core, DDI (Citation)
- DDI (Social Science)
- Virtual Observatory VOResource (Astronomy)
- ISA-Tab (Life Science)

Citation Metadata ▾

Geospatial Metadata ▾

Social Science and Humanities Metadata ▾

Astronomy and Astrophysics Metadata ▲

Type	<input checked="" type="checkbox"/> Image <input type="checkbox"/> Mosaic <input type="checkbox"/> EventList <input type="checkbox"/> Spectrum <input type="checkbox"/> Cube <input type="checkbox"/> Table
Facility	<input type="text"/> +
Instrument	<input type="text"/> +
Object	<input type="text"/> +
Spatial Resolution	<input type="text"/>

# Metadata Templates

- Setup templates available at the Dataverse level to reduce need to copy information
- Can be set to default template, or can be copied

Ryerson University Dataverse

Root Dataverse > Ryerson University Dataverse > Create Dataset Template

Template Name \*

\*Asterisks indicate metadata fields that users will be required to fill out while adding a dataset to this dataverse.

Citation Metadata ▲

Title *	<input type="text" value="Enter title..."/>	
Subtitle	<input type="text"/>	
Alternative Title	<input type="text"/>	
Alternative URL	<input type="text" value="Enter full URL, starting with http://"/>	
Other ID	Agency <input type="text"/>	Identifier <input type="text"/>
		<input type="button" value="+"/>
Author *	Name * <input type="text" value="FamilyName, GivenName or Organization"/>	Affiliation <input type="text"/>
		<input type="button" value="+"/>

# Data Citation and Journal Cross-referencing

- Get a DOI / data citation before publishing data
  - DataCite Canada DOI minting and indexing
  - Cite data in a paper, access for review process
- ORCID ID field (coming soon... ORCID sign-in)
- Reference a publication that uses data



# Widgets support for Research Promotion

Scholars Portal **Dataverse** Fr (Beta)

Search all dataverses...   [About](#) [User Guide](#) [Support](#) [Sign Up](#) [Log In](#)

Queen's University Library Research Data Archive Dataverse (Queen's University)

Scholars Portal Dataverse > Queen's University Dataverse > Queen's University Library Research Data Archive Dataverse > **Analysis of Strainbursts in the Sudbury Region and Numerical Modelling of Destress Blasting**

[Metrics](#) [98 Downloads](#) [Contact](#) [Share](#)

**Analysis of Strainbursts in the Sudbury Region and Numerical Modelling of Destress Blasting** [View HTML](#)

Gingras Little, Kristopher, McKinnon, Steve, Moreau-Ventura, Lindsay, McDonald, Andrew, 2017, "Analysis of Strainbursts in the Sudbury Region and Numerical Modelling of Destress Blasting", doi:10.5683/SF14RF1BLJ, Scholars Portal Dataverse, V1, UNF:6:6RLU0960vVasehebbCcIQ== [Cite Dataset](#) [Learn about Data Citation Standards](#)

**Description**  
The occurrences of strainbursting in underground mining have long presented a risk to workers underground. The investigation of these events occurred in the Sudbury basin between 2013 and 2015, yielding results in the nature of observations, numerical models, and a database of events. Included in the data are videos of borehole observations, instrumentation data, code to replicate the numerical models used in the analysis of destress blasting, and a database containing the author's interpretation of many strainburst event reports. All of the data contained was used in the author's Master's of Applied Science thesis, "The Analysis of Strainbursts in the Sudbury Region And Numerical of Blasting". The findings from this research and associated data set contained herein, will add to the continuous process of improving the understand and prevention of strainbursts in mines.

**Subject**  
Engineering

**Keyword**  
Borehole observations, Destress blasting, Mining, Strainbursts

[Files](#) [Metadata](#) [Terms](#) [Versions](#)

Search this dataset

**GARY KING** [BIO & C.V.](#) [WRITINGS](#) [RESEARCH AREAS](#) [SOFTWARE](#) [DATAVERSE](#) [RESEARCH GROUP](#) [TEACHING](#) [CONTACT](#)

**HOME /**

## Dataverse

Below is my dataverse, which is comprised of data sets and replication data sets associated with my published articles and books. For information about the dataverse project (which I created and run), see [this article](#) and the [dataverse.org](#) project website.

Harvard Dataverse > Gary King Dataverse

Search this dataverse...   [Advanced Search](#)

**Dataverses (0)**  
**Datasets (59)**  
**Files (1,366)**

**Publication Date**  
2017 (24)  
2009 (8)  
2010 (6)  
2011 (4)  
2014 (4)  
[More...](#)

**Subject**  
Social Sciences (46)

**Author Name**  
King, Gary (47)  
Gery King (10)  
Zeng, Langche (8)  
Gelman, Andrew (4)  
Sanofi, Samir (4)  
[More...](#)

**Keyword Term**  
Causal Inference: FFRD: M4R4New

**1 to 10 of 59 Results** [Sort -](#)

**Replication Data for: Why Propensity Scores Should Not Be Used for Matching**  
Jun 12, 2018  
King, Gary; Nielsen, Richard, 2018, "Replication Data for: Why Propensity Scores Should Not Be Used for Matching", doi:10.7910/DVN/CDD8AE, Harvard Dataverse, V1  
We show that propensity score matching (PSM), an enormously popular method of preprocessing data for causal inference, often accomplishes the opposite of its intended goal – increasing imbalance, inefficiency, model dependence, and bias. PSM supposedly makes it easier to find me...

**Replication Data for: How the News Media Activates Public Expression and Influences National Agendas**  
Nov 13, 2017  
King, Gary; Schrewer, Benjamin; White, Ariel, 2017, "Replication Data for: How the News Media Activates Public Expression and Influences National Agendas", doi:10.7910/DVN/1EM87K, Harvard Dataverse, V1  
We demonstrate that the news media causes Americans to take public stands on issues, join national policy conversations, and express themselves publicly more often than they would otherwise — all key components of democratic politics. We recruited 48 mostly small media outlets...

**Replication data for: How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, not Engaged Argument**  
Apr 8, 2017  
King, Gary; Pan, Jennifer; Roberts, Margaret E., 2017, "Replication data for: How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, not Engaged Argument", doi:10.7910/DVN/Q52MFO, Harvard Dataverse, V1, UNF:6:3WEJ2mcy2m+vfsm00Jig==  
The Chinese government has long been suspected of hiring as many as 2,000-300 people to surreptitiously insert huge



# Data Analysis and Visualization

- Data Explorer (cross tabulations)

## Drone Awareness and Perceptions: A Three Country Study 2014

Angus Reid Global, 2014, "Drone Awareness and Perceptions: A Three Country Study 2014", hdl:10864/10798, Scholars Portal Dataverse, V5

Q Results Download

Chart View Table View

ID	Name	Label
104579	PanelIdQuestion	PanelIdQuestion
104625	gender	Gender
104461	Q5_Federal_Vote_2011	Did you vote in the last FEDERAL election in May 2011?
104501	Q6_Federal_PartyChoice_2011	Who did you vote for in the last FEDERAL Election?
104666	Age_Rollup_broad	Age_Rollup_broad
104709	Income	Income
104442	HH_Income	Household Income
104512	Education_Level	Education
104538	Region_Rollup	Region_Rollup
104553	MACRO_REGION	MACRO_REGION

First 1 2 3 4 5 Last Records Per Page 10

Who did you vote for in the last FEDERAL Election?	Region_Rollup	%
Conservative Party of Canada	BC	4.8
	Alberta	3.5
	SK	1.7
	MB	1.3
	ON	14.0
	QC	8.8
	Atlantic	3.6
	Territories	0.0
Liberal Party of Canada	BC	2.3
	Alberta	1.9
	SK	0.7
	MB	0.9
	ON	7.0
	QC	4.8
	Atlantic	1.6
	Territories	0.0
BC	4.4	
Alberta	3.2	
SK	1.1	

Cross  
tabulation

# Data Analysis and Visualization

## Drone Awareness and Perceptions: A Three Country Study 2014

Angus Reid Global, 2014, "Drone Awareness and Perceptions: A Three Country Study 2014", hdl:10864/10796, Scholars Portal Dataverse, V5

Q [x] Results Download

Chart View Table View

ID	Name	Label
104579	PanelistidQuestion	PanelistidQuestion
104625	gender	Gender
104461	Q5_Federal_Vote_2011	Did you vote in the last FEDERAL election in May 2011?
104501	Q6_Federal_PartyChoice_2011	Who did you vote for in the last FEDERAL Election?
104666	Age_Rollup_broad	Age_Rollup_broad
104709	Income	Income
104442	HH_Income	Household Income
104512	Education_Level	Education
104538	Region_Rollup	Region_Rollup
104553	MACRO_REGION	MACRO_REGION

First 1 2 3 4 5 Last Records Per Page 10

**Variable Q5\_Federal\_Vote\_2011: Did you vote in the last FEDERAL election in May 2011?**

Values	Categories	N
1	Yes	2566
2	No	479

Summary Statistics

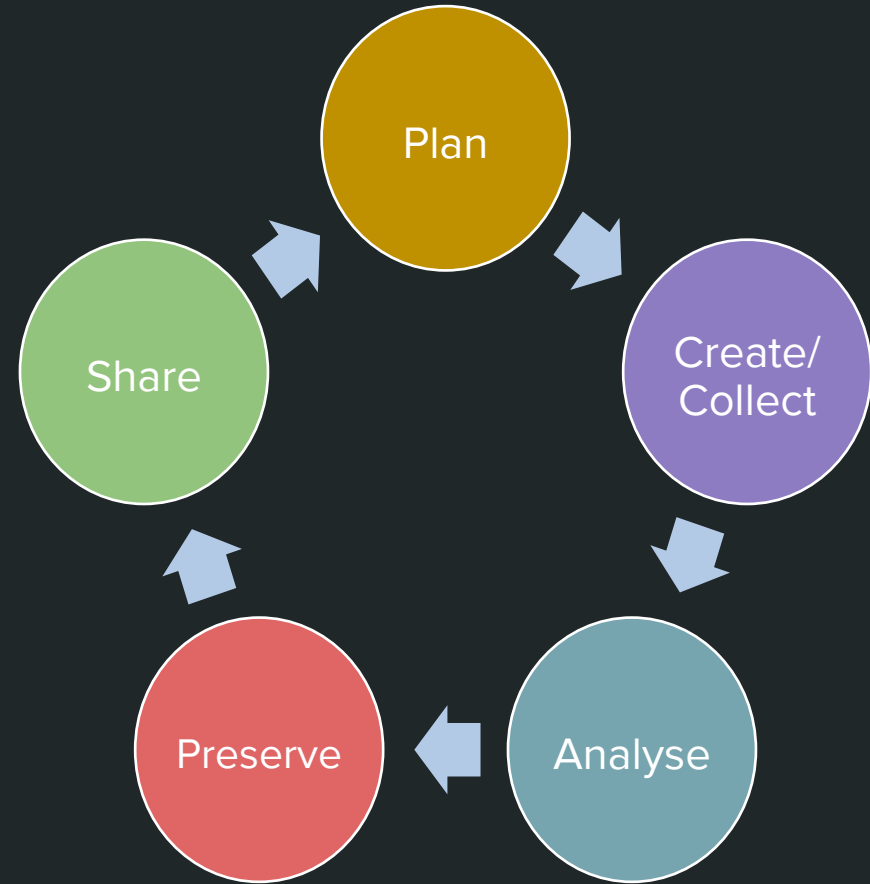
Chart view

# Preservation – Archivemata Integration

- Free and open-source digital preservation system designed to maintain standards-based, long-term access to collection of digital objects
- Currently work is underway to support Archivemata integration with Dataverse

# Helpful data management resources

---



# Helpful resources

- ✧ [CIHR's RDM Learning Module](#) (beta)
- ✧ Portage's ["Good Enough" Research Data Management Guide](#)
- ✧ [UBC Library DataGuide](#)
- ✧ [MANTRA Research Data Management Training](#)
- ✧ Scholars Portal Dataverse [User Guide](#)



# Thank You.

For more information:

**RDM**  
@McMaster

[library.mcmaster.ca/services/rdm](https://library.mcmaster.ca/services/rdm)  
[rdmgmt@mcmaster.ca](mailto:rdmgmt@mcmaster.ca)

McMaster Library's RDM webpage  
and primary contact

**portage**  
SERVICES PARTAGÉS POUR LES DONNÉES DE RECHERCHE  
SHARED STEWARDSHIP OF RESEARCH DATA

[portagenetwork.ca/](https://portagenetwork.ca/)

CARL Portage Network page: Access to  
RDM information and DMP Assistant