

Managing your research data

Resources, tools, strategies

Vivek Jadon (vivek@mcmaster.ca) & Jay Brodeur (brodeujj@mcmaster.ca)

McMaster University Library

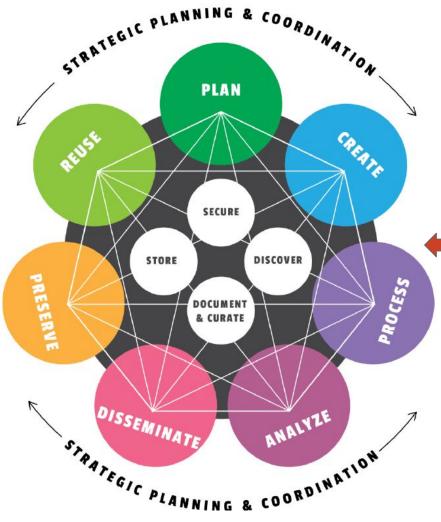
CRESS Research in Focus | 06-March, 2018



Overview

- ♦ A (very) brief Research Data Management primer (5 mins).
- ♦ Data management planning (20 mins)
- ♦ Data storage & backup (10 mins)
- In-project data management (10 mins)
- ♦ Data archiving & sharing (30 mins)
- Helpful data management resources (5 mins)

In our previous episode...



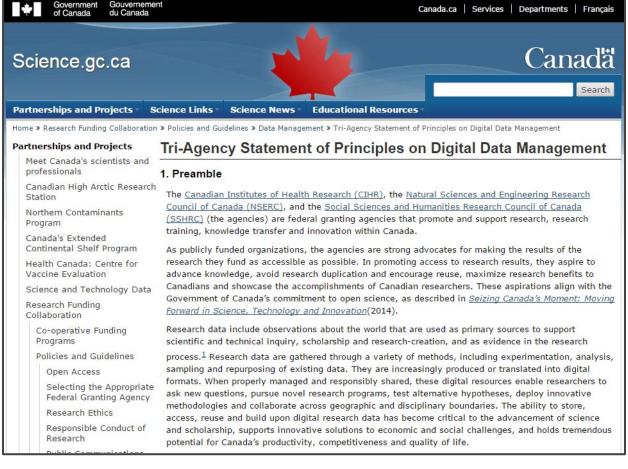


Research Data Management

is the active organization & maintenance of data throughout the

research data lifecycle to ensure its security, accessibility, usability, and integrity

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







Draft Tri-Agency RDM Policy



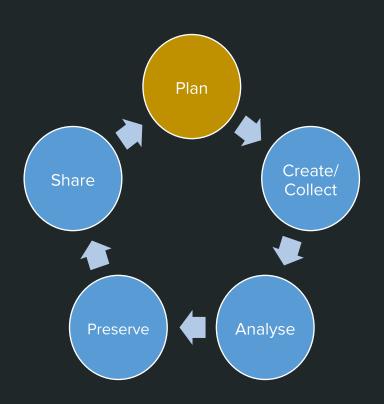
- ♦ Tri-Agency draft data management policy expected April, 2018
- ♦ 6-month consultation period; feedback will inform policy

Proposed policy — 3 possible requirements:

- 1. **Institutions**: Institutional Strategy
- 2. **Researchers**: Data Management Plans
- 3. **Researchers**: Data Deposit^{a,b,c,d,e,f,g}
- Phased and incremental implementation
 <u>Link to M. Lucas' slides Portage Network Event</u> (30-Jan, 2018)

Planning for RDM

- Considerations
- DMP Assistant





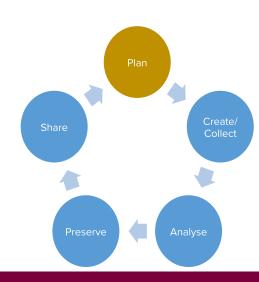
...it all starts with a (good) plan

A research Data Management Plan (DMP) should:

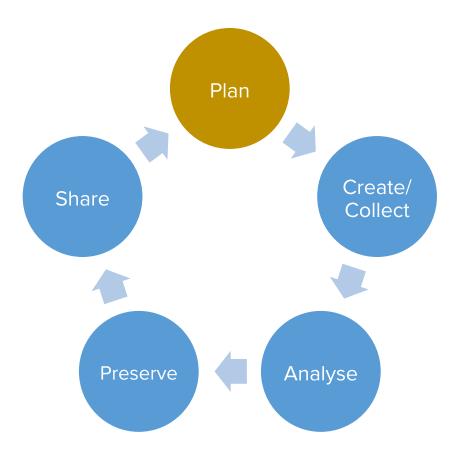
- ♦ Describe how you will manage data through all stages of your research
- ♦ Communicate a strategy for creating share-worthy data products

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...



Planning



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

What resources do you require to manage your data?

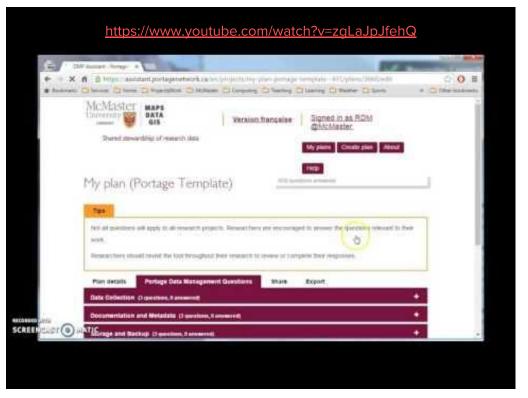
Who is responsible for data management and long-term stewardship?



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://assistant.portagenetwork.ca/



Activity - Develop a DMP (~15 minutes)

Using the **CARL Portage DMP Assistant:**



- Navigate to <u>assistant.portagenetwork.ca</u>
- 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

Sign in with your email / password



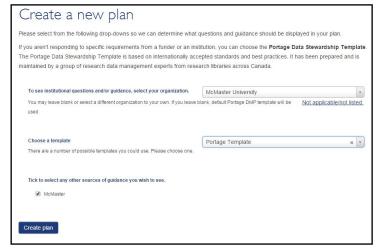
Click "Create plan"

Set organization to "McMaster University"
Select the "Portage" template









assistant.portagenetwork.ca

Enter relevant personal, organizational and study information

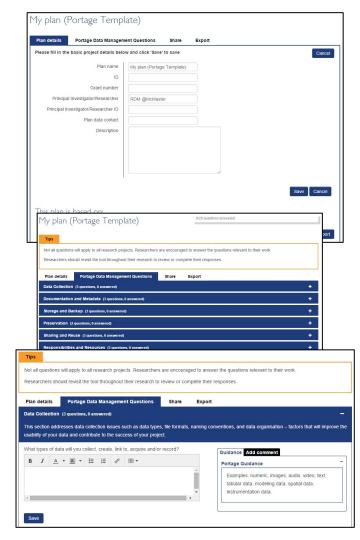


Expand boxes to complete questions for each section



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







Reflections and Discussion

What are you reflections on:

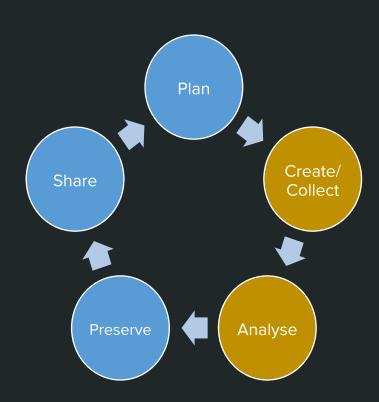
- The overall experience?
- The interface and its content?

What sections / questions were particularly challenging? Howso?

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

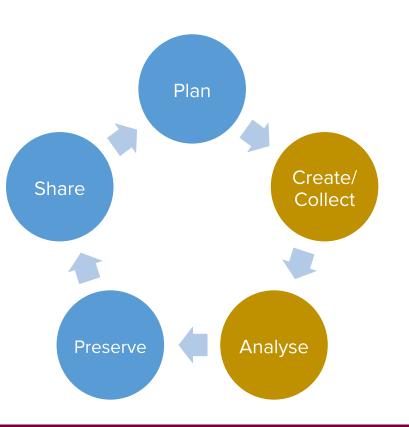
In-project management

- Storage & backup solutions
- 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: 10 42 71



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
- Dropbox, Google Drive, etc.

Considerations for "Archived" Backup

McMaster University LIBRAD

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

	RHPCS - Backup Services	RHPCS - Hosted Server Packages	MacDrive	Microsoft OneDrive / Teams	
Storage Quota	1 TB; more available for fee	1 TB; more available for fee	300 GB per account	1 TB per account; up to 5 TB by request	
Rates / cost	\$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	
Backups / versioning	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 (?)	
Who can use this service?	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	
Server location	A.B. Bourns building	A.B. Bourns building	Replicated clusters in Gilmour Hall and JHE	OneDrive: Canadian servers Teams: Soon in Canadian servers only	
Other notes			Supports encrypted libraries, file and directory sharing, Desktop client, web interface	Supports file and directory sharing, Desktop client, web interface	
More info	rhpcs.mcmaster.ca/current-rates	rhpcs.mcmaster.ca/current-rates	macdrive.mcmaster.ca/ Documentation: https://goo.gl/AvRGwx	portal.office.com/ Documentation: mcmaster.ca/uts/licensing	

Access the matrix: goo.gl/45iy38





<u>Open Science Framework</u>

https://osf.io/

<>	Free,	open	source	web	application

Connects and supports the

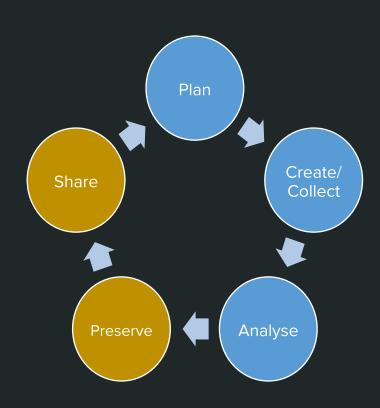
- research workflow
- Integrates with existing services (e.g. Google Drive, Dropbox, etc.)
- Supports collaborative workflows
- Facilitates data publishing to selected data repositories

https://youtu.be/2TV21gOzfhw

GETTING STARTED WITH THE OPEN SCIENCE FRAMEWORK (OSF)

Archiving and Sharing

- FAIR principles
- Data repositories
- Scholars Portal Dataverse





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

Findable

Interoperable

Accessible

Reusable

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

I1: (meta)data use a formal, accessible, shared, broadly applicable language for knowledge rep.

12: (meta)data use vocabularies that follow FAIR principles

13: (meta)data include qualified references to other (meta)data

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Bouwman, J. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific data, 3. *doi:10.1038/sdata.2016.18*



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





Federated Research Data Repository (FRDR)

- Compute Canada + Canadian Association of Research Libraries (CARL)
- A scalable, federated platform for digital research data management and the discovery of Canadian research data
- National data repository
- Framework for federating existing and future data repositories

https://www.frdr.ca

Discovery Deposit Preservation

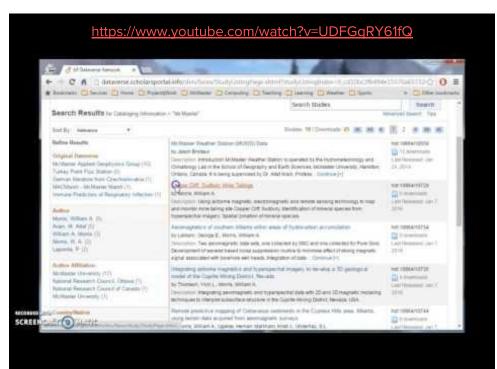




Scholars Portal Dataverse

Scholars Portal
A Service of the Ontario Council of University Libraries

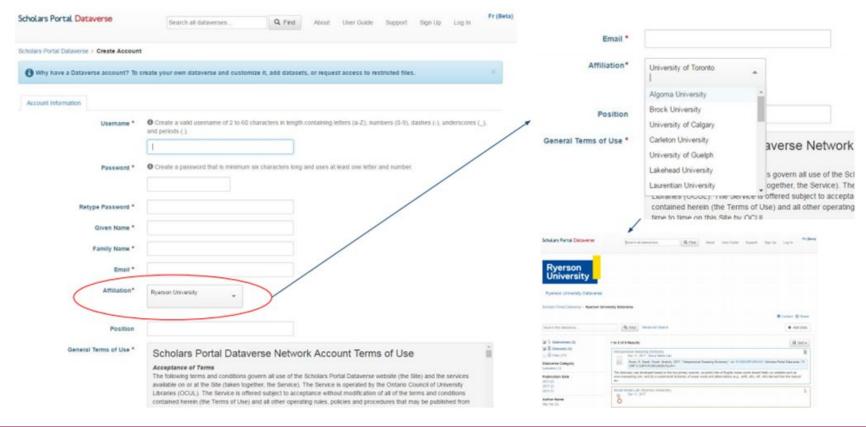
- A data repository for researchers at Ontario's universities -- free and open for all researchers in Canada
- An online platform to share, preserve, cite, explore and analyze research data.
- Allows researchers to control how they share their data.
- Supports data DOI registration through Datacite Canada.



http://dataverse.scholarsportal.info

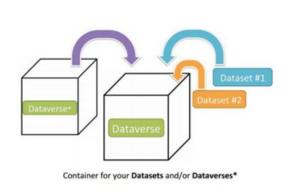


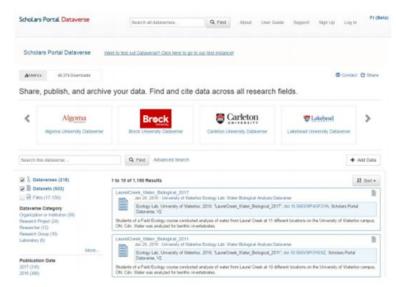
Sign-up & Institutional Affiliation





Institutional Dataverse



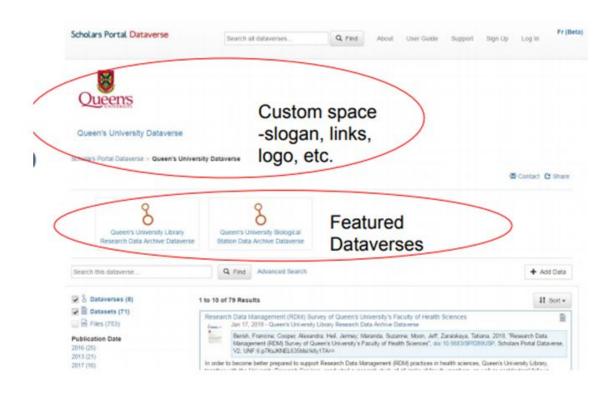


You can have as many Dataverses within your Institution Dataverse as you like!



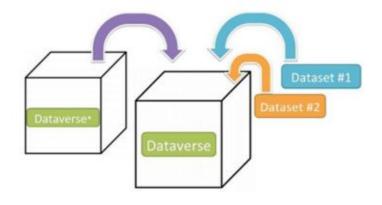
Customize your Dataverse

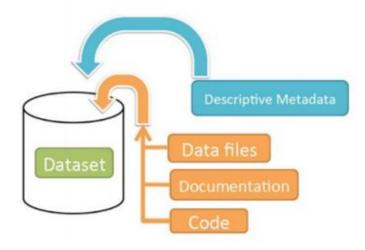
- Customizable features:
 - Branding/logo
 - Featured Dataverses
 - Facets
 - Contact





Dataverses, Datasets, Files



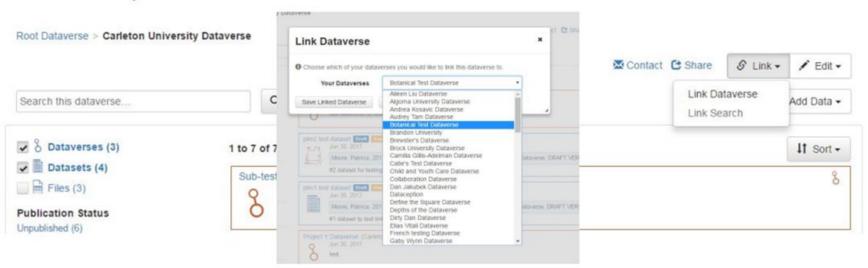




Linking Datasets and Dataverses



Carleton University Dataverse





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 Cataverse website (the Site) and the services available on or at the Site (taken together, the Service). The Service is operated by the Critario 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 distribute on 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, ties, 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

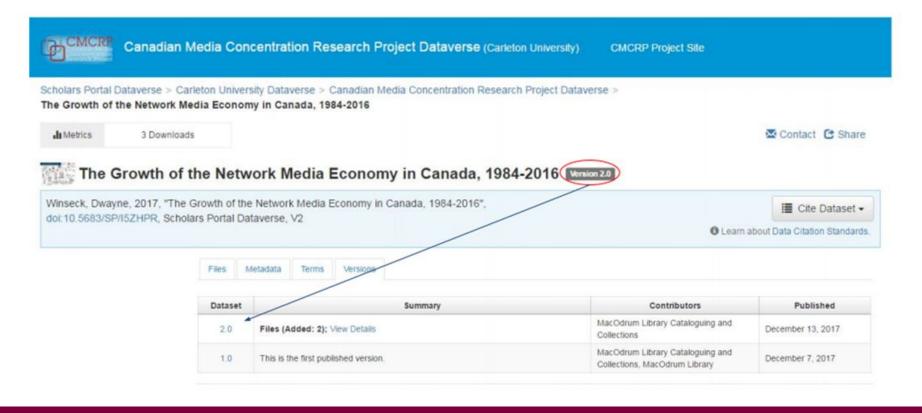


Storage Size Limits

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

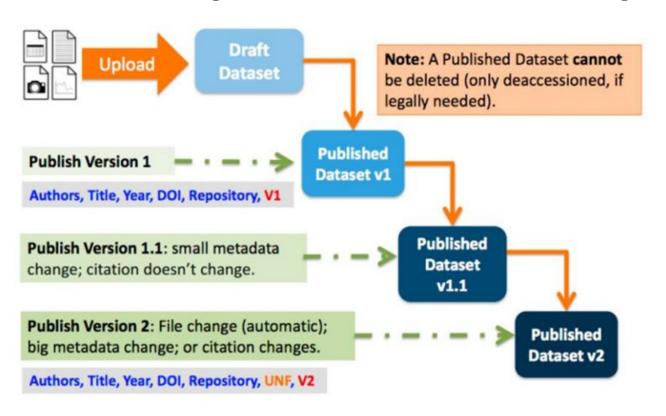


Dataset Management / File Versioning





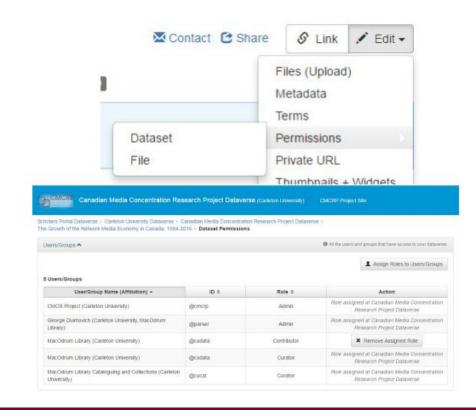
Dataset Management / File Versioning





Dataset and File Permissions

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



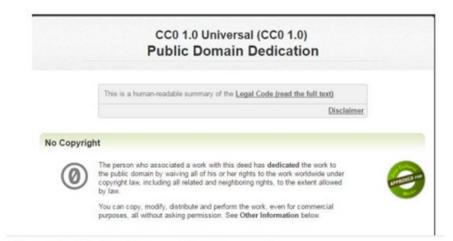


Licensing

Default to CC0 (Open Data)

OR

 Custom "Data Usage Agreement"



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.

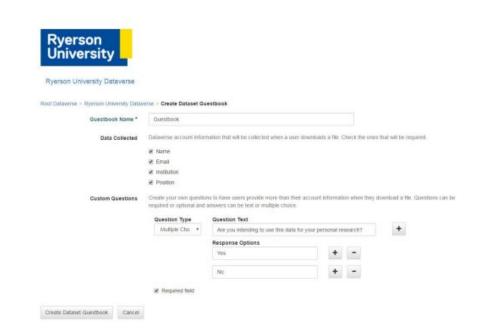
II. Modification of this Agreement

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





Metadata Standards

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

Ottetion Metadata Y		
Geospatial Metadata 💙		
ocial Science and Humanities M	nadota 🗸	
stronomy and Astrophysics Meta	Sida ^	
Туре	Image Mosaic EventList Spectrum Cube Tarrie	
Facility		•
Instrument		•
Object		•
Spatial Resolution		



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

Template Name * Port	tage template		
	Portage template		
nies indicate metaciata fields th	all users will be required to fill out while	adding a detaset to this detoverse.	
tion Metadata A			
itie "	Shifes (Ma		
LOTTE			
Uternative Title			
Vismative URL	Short full LEE, starting with Higgs (
Other ID	Agency	Martiflar	



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

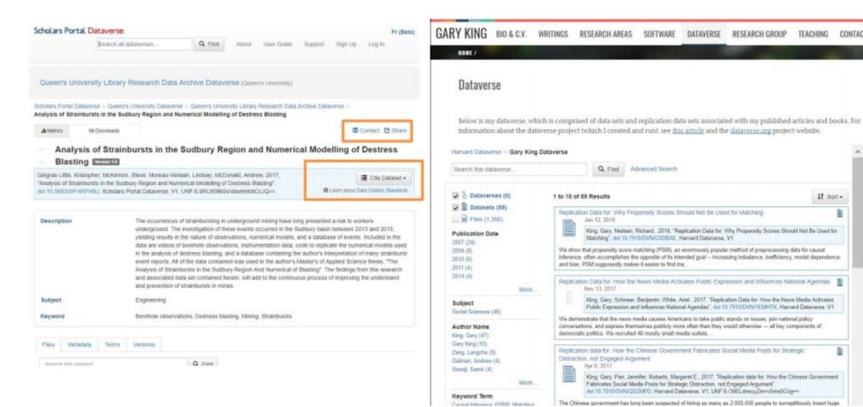




CONTACT

If Son+

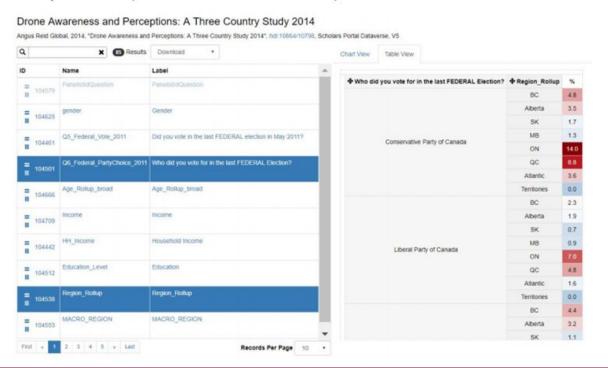
Widgets support for Research Promotion





Data Analysis and Visualization

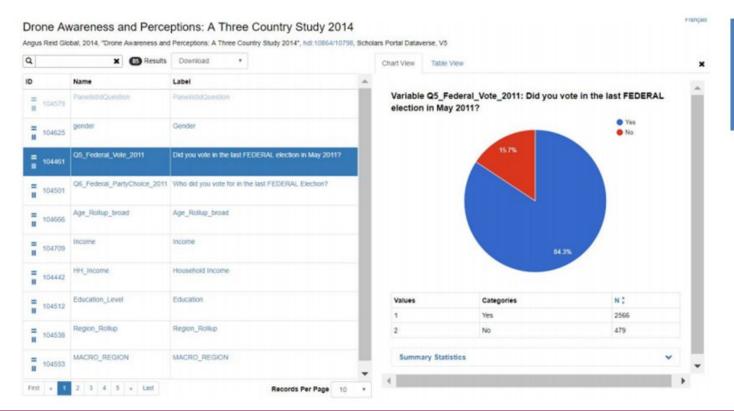
Data Explorer (cross tabulations)







Data Analysis and Visualization







Preservation – Archivematica 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 Archivematica integration with Dataverse

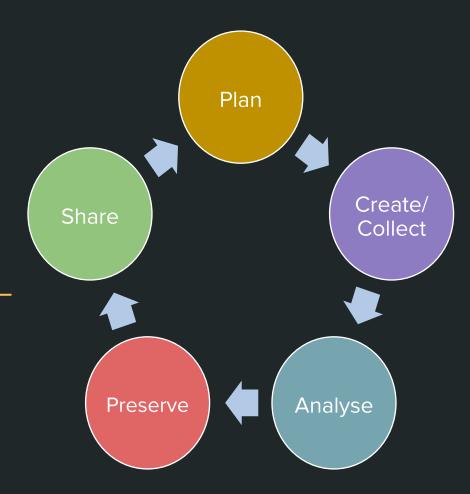


Create your own (demo) Dataverse

Scholars Portal Demo Dataverse site:

http://demodv.scholarsportal.info

Helpful data management resources





Helpful resources

- Portage's "Good Enough" Research Data Management Guide
- ♦ UBC Library DataGuide
- MANTRA Research Data Management Training
- Scholars Portal Dataverse <u>User Guide</u>





Thank You.

For more information:



<u>library.mcmaster.ca/rdm</u> <u>rdmgmt@mcmaster.ca</u>

McMaster Library's RDM webpage and primary contact



portagenetwork.ca/

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