Managing Your Research Data from Cradle to Grave

Resources for Research at Queen's (R4R@Q)

March 13, 2019

Chris MacPhee, Assistant Director, Centre for Advanced Computing

Courtney Matthews, Head, Open Scholarship Services, Library

Agenda

Introduction to research data management

Managing your data throughout the research lifecycle

Best practices and tools

Resources

What is research data management (RDM)?

Processes applied throughout the **lifecycle of a research project** to guide the collection, documentation, storage, sharing, and preservation of research data.

Practices that are **integral to conducting responsible research** and can help researchers save resources by ensuring their data is complete, understandable, and secure.

What is research data management (RDM)?

Practices that follow institutional and funding agency guidelines like the **DRAFT Tri-Agency Research Data Management Policy**

Enables the broader research community to derive maximum value from research data that can be accessed, shared, reused and repurposed.

SSHRC CRSH

CIHR IRSC

NSERC CRSNG

What are research data?

Primary sources supporting research, scholarship or artistic endeavours

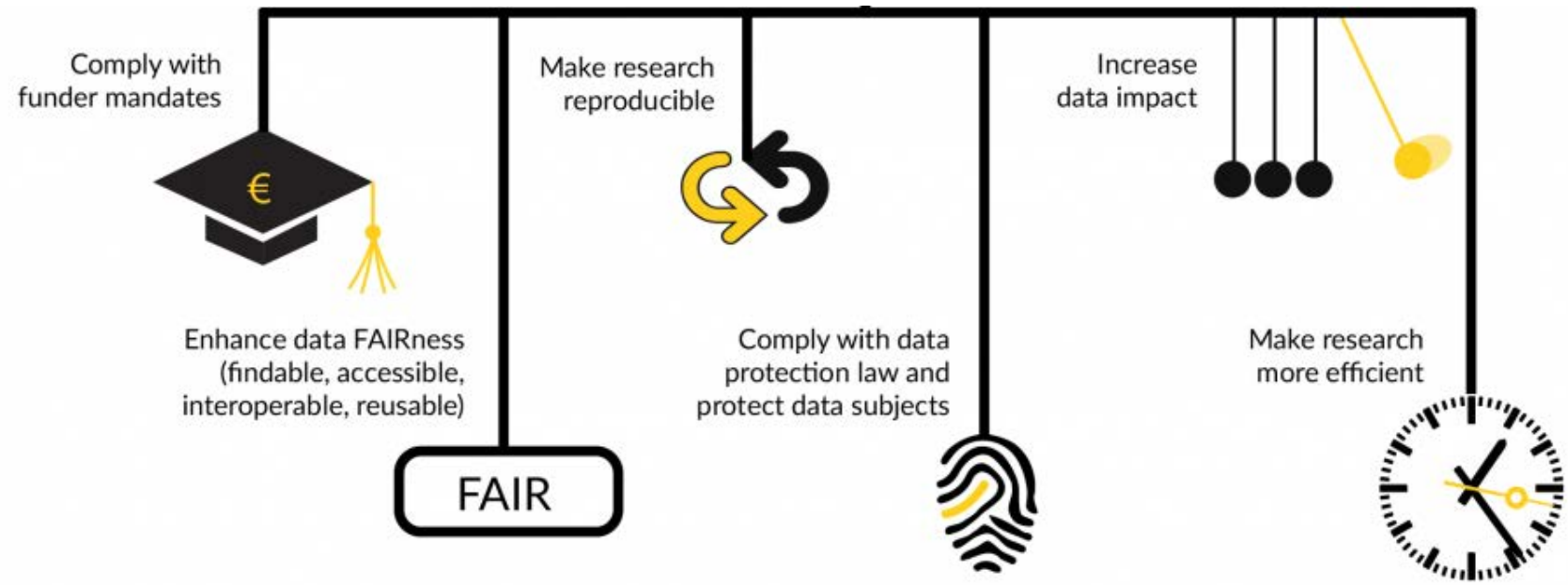
Can be used as evidence to validate findings and results

Is experimental data, observational data, operational data, third party data, public sector data, monitoring data, processed data, or repurposed data

All other digital and non-digital content have the potential to become research data

What are research data?

Why should I manage my research data?



RESEARCH DATA MANAGEMENT PRIMER

Research Data Management (RDM)

- RDM refers to the processes applied throughout the lifecycle of a research project to guide the collection, documentation, storage, sharing, and preservation of research data.
- RDM practices are integral to conducting responsible research and can help researchers save resources by ensuring their data is complete, understandable, and secure.
- RDM practices also follow institutional and funding agency guidelines that protect their investments.
- The broader research community can derive maximum value from research data that can be accessed, shared, reused and repurposed.

The Research Data Lifecycle

Plan **Create** **Process** **Analyze** **Disseminate** **Preserve** **Reuse**

* Life cycle model developed by the Leadership Council for Digital Research Infrastructure. For more information visit http://digitalleadership.ca

Defining Research Data

- Primary sources supporting research, scholarship or artistic endeavours
- Can be used as evidence to validate findings and results
- May take the form of experimental data, observational data, operational data, third party data, public sector data, monitoring data, processed data, or repurposed data
- All other digital and non-digital content have the potential to become research data

*Research data. (n.d.) In CASRAI's Dictionary. Retrieved from dictionary.casrai.org/Research_data

portage

https://portagenetwork.ca/training-resources/

Plan

Supporting Canadian innovation

through shared expertise and stewardship of research data

Launched in 2014 by the Canadian Association of Research Libraries, **Portage** works within the library community to coordinate expertise, services, and technology in research data management, seeking to collaborate with other research data management stakeholders.

How do I manage my research data?

Write a data management plan

Use the **Portage** DMP Assistant tool to develop your own data management plan. It is freely available to all researchers in Canada and takes you step-by-step through a series of questions based on a general template for research data stewardship.

portage

plans de gestion des données

ASSISTANT

PGD

data management plans

portage

Shared stewardship of research data

Version française

Home About Help

https://assistant.portagenetwork.ca/

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.

Sign in +

If you have an existing account with DMP Assistant or previous

+

gle sign-in authentication. For
nt. You will have the option to
s ID when that feature

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 answers.

DMP Assistant

R4R Demo

Plan details | Portage Data Management Questions | Share | Export

1. Create a DMP using a template
2. Answer data management questions
3. Share your DMP
4. Export your DMP

"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

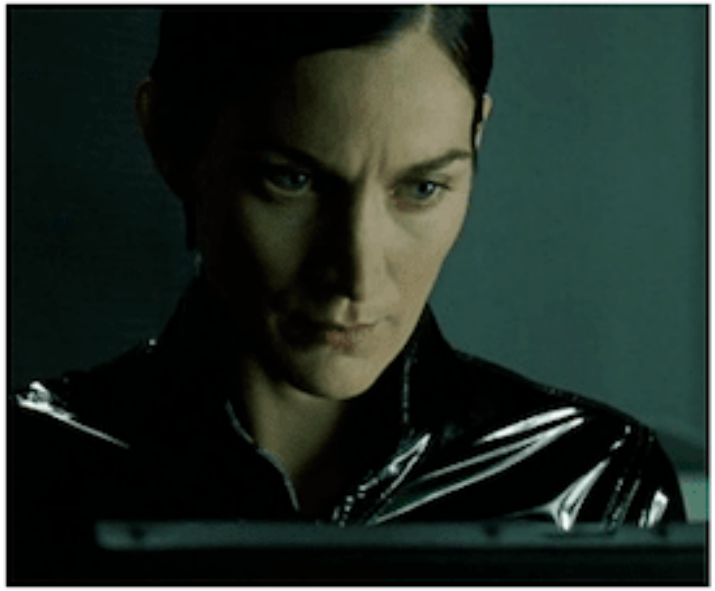
5.1 Submit final data files to a repository assigning a persistent identifier (e.g. handles or DOIs).

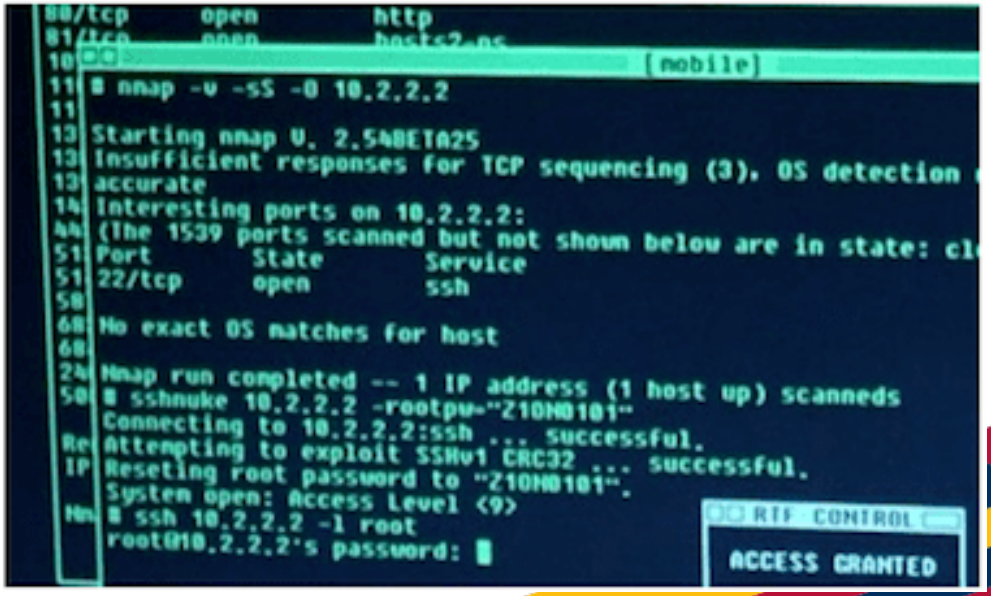
5.2 Provide good metadata for your study so others could find it (use your discipline's metadata standard, e.g. Darwin Core, DDI, etc.).

https://portagenetwork.ca/training-resources/

Create/Process/Analyze

Think Big: Big Data, AI, Blockchain, …





Research IT Platforms

The unification and integration of computers, storage, enterprise software, middleware, training, visualization, and analytics which enable researchers to transparently access, store, and analyze information.

a.k.a.

Simplifying Research

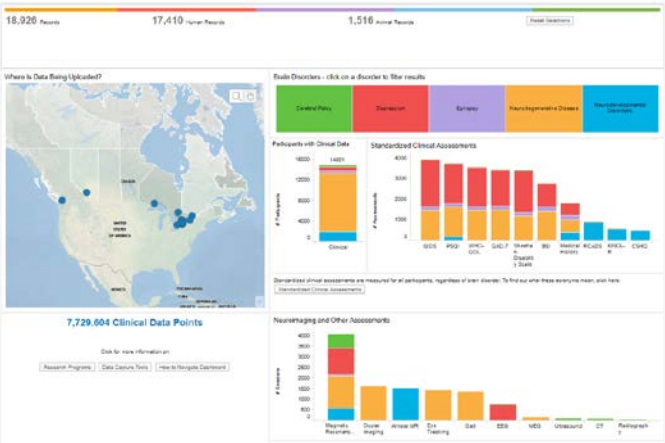
or

You don’t need a new degree to play

Platform Examples

ONTARIO BRAIN INSTITUTE
INSTITUT ONTARIEN DU CERVEAU

Brain-CODE



CAC Collaborative Teaching Platform for Big Data

jupyterhub
sas
hadoop
R Studio
Ambari

Canada Wide Resources

COMPUTE ONTARIO

Queen’s active data storage resources

Secure file storage

- 5 TBs cloud storage through OneDrive
- Teams collaborative platform

External open resources

OSF

Free, open source **project management tool**

- Keep all files, data, protocols in one centralized location
- Control access, including which parts of project are public or private
- Connect to 3rd party services (e.g., Dropbox, Google Drive, Mendeley, GitHub)

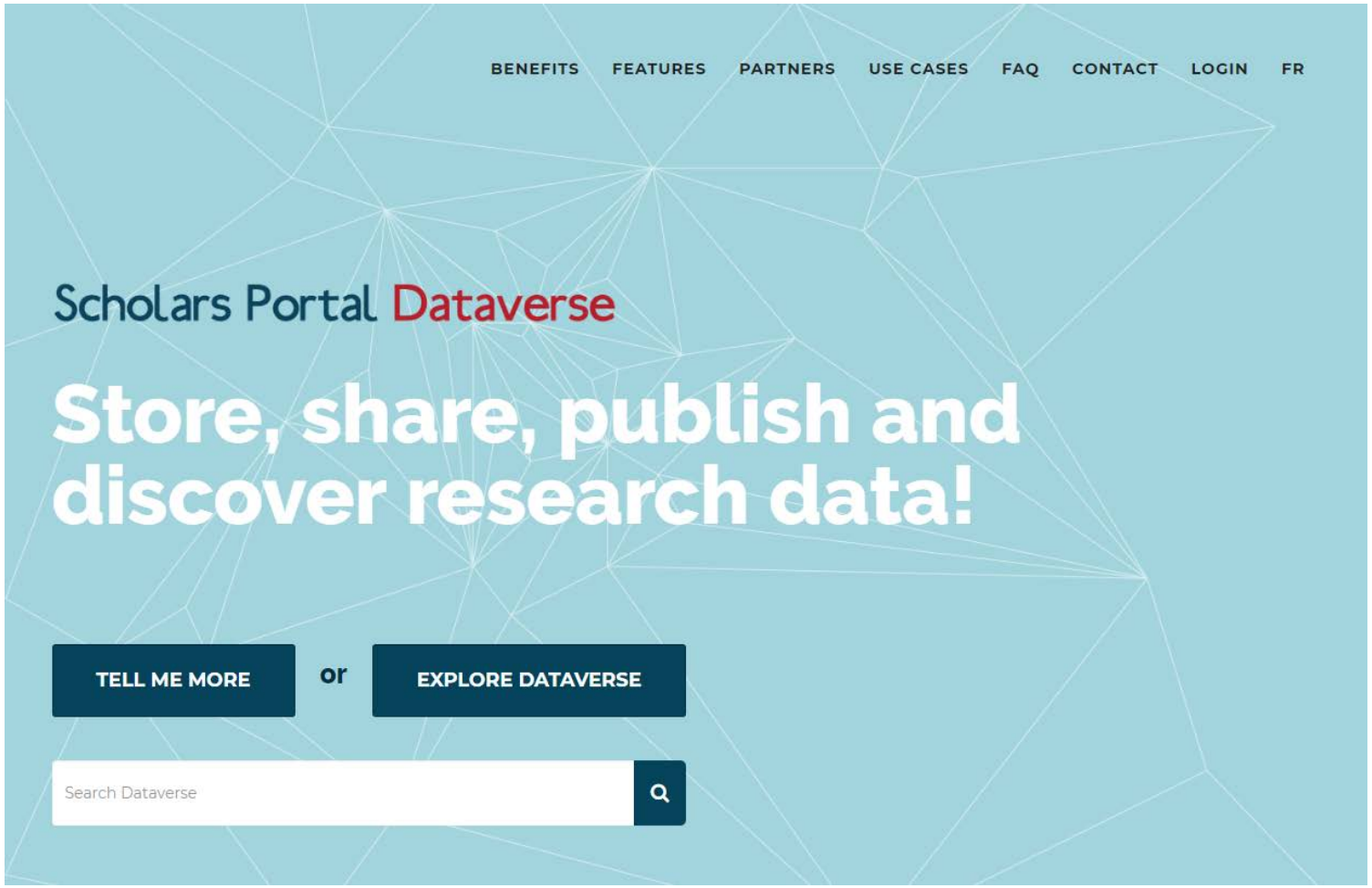
Disseminate/Preserve

How do I share my research data?

Deposit your data in to a data repository

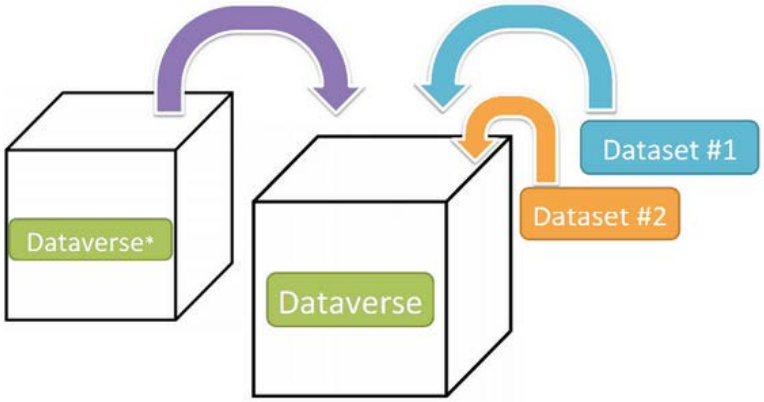
Review the Library's Research Data Management Guide and fill out the Data Deposit Request form to begin a library mediated deposit of your data in to Scholars Portal Dataverse.

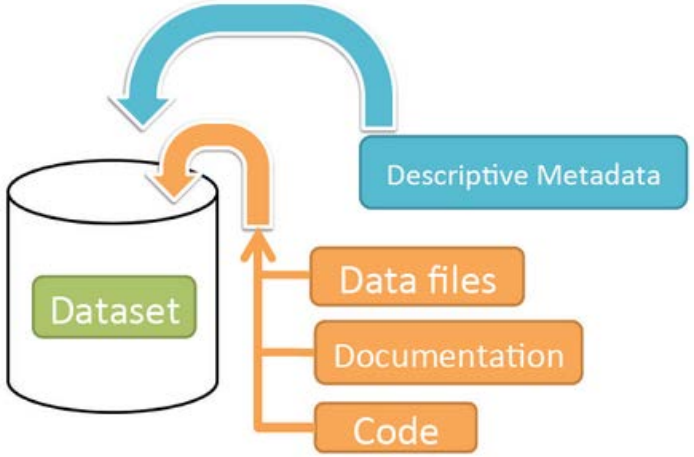
Scholars Portal Dataverse



https://dataverse.scholarsportal.info

Dataverses, datasets, files





Dataverse = Container for datasets and/or dataverses

Dataset = Container for your data, documentation, and code

Dataverse

- Files and Metadata:
  - All file formats accepted
  - 2GB maximum file upload

- Data citation:
  - Digital Object Identifiers
  - Establish links between scholarly outputs and data

- Data Licensing:
  - Default Creative Commons Zero “No rights reserved”
  - Custom data use agreement

Dataverse

- Discoverability
  - Integrated with SHARE notification system, Google dataset search, and ORCID identifiers that link to your scholarly record

- Access:
  - Ability to restrict individual files and assign specific permissions for collaborators
  - IP Group based permissions

Scholars Portal Dataverse

Demo: demodv.scholarsportal.info
Production: dataverse.scholarsportal.info

"Good Enough" Dataverse

(a brief guide for busy people)

This brief guide presents key information for researchers to understand Dataverse, its features, and how to access and use it.

1 What is Dataverse?

1.1 A free, open source web application developed at Harvard University.

1.2 Stores, shares, cites, explores, and analyzes research data.

1.3 Facilitates making your data available to others, as appropriate.

1.4 A robust repository environment for depositing your research data.

2 Why should I use Dataverse?

2.1 Helps you meet both funding and journal requirements.

2.2 Stores many types of data, including:
- Tabular data (CSV, SPSS, R, etc)
- Documentation (pdf, doc, text, code)
- Geospatial data
- Multimedia file (audio-visual)

3 What are the benefits?

3.1 Creates a Digital Object Identifier (DOI) so your data is a citable research output for publications and CV.

3.2 Increases the impact and visibility of your research.

3.3 Stability and longevity of your data as created at time of deposit.

3.4 Statistics and metrics to track and understand how your data is used worldwide.

4 How open is my data? It's up to you!

4.1 Choose your access settings: modify access to your data files - ranging from fully open to completely restricted.

4.2 Customize the terms of use, specifying how your data can be used.

4.3 Modify access to your data files at any time.

4.4 Your metadata is publicly available. Therefore, your data is always discoverable!

5 How do I start?

5.1 Find out if your institution has a local Dataverse - you can check the Portage site to see if your institution is using Dataverse.

5.2 If needed, check with your local Data contact for help.

5.3 Create your Dataverse account.

6 What's the next step?

6.1 Ensure that your data follows best practices for formatting and metadata.

6.2 Upload your data and add your metadata.

6.3 Set any access restrictions or terms of use for your data.

6.4 Option to create a 'guestbook' to track the use of your data.

6.5 Publish your data!

https://portagenetwork.ca/training-resources/

How do I share my research data?

Deposit your data in a disciplinary repository

Consult the Registry of Research Data Repositories re3data.org.

re3data.org

REGISTRY OF RESEARCH DATA REPOSITORIES

Reuse

Queen's University

Queen's University Dataverse (Queen's University)

Metrics 10 Downloads

Contact Share

Chronic Pain Surveillance at Queen's, 2013-2017 Version 1.1

Wilson, Rosemary A.; VanDenKerkhof, Elizabeth G.; Duggan, Scott; Gilron, Ian; Good, Mary Anne; Henry, Richard; Carley, Meg, 2018, "Chronic Pain Surveillance at Queen's, 2013-2017", https://doi.org/10.5683/SP2/GAPNRM, Scholars Portal Dataverse, V1, UNF:6:d+jC5YYQZO7ERTS1Y37v0Q== [fileUNF]

Cite Dataset

Learn about Data Citation Standards.

Description	The purpose of this surveillance system is to develop a better understanding of the characteristics of patients with chronic pain and their response to treatment in a general clinic setting. This ongoing surveillance system will contribute valuable information for both clinical and research purposes. The primary objective of this research is to collect information about pain, activity limitation, and psychological factors in patients seen for chronic pain conditions.
Subject	Medicine, Health and Life Sciences
Keyword	Activity limitation, Chronic pain, Evidence-based care, Pain, Quality of life
Related Publication	Hamilton, A.N.J. (2017). An Observational Study Examining the Association Between Chronic Pain Coping Strategies and Unplanned Health Care Use by Individuals Attending an Interdisciplinary Pain Clinic (Master's Thesis). Queen's University, Kingston, Ontario, Canada. Retrieved from http://hdl.handle.net/1974/22959

Files Metadata Terms Versions

Search this dataset... Find

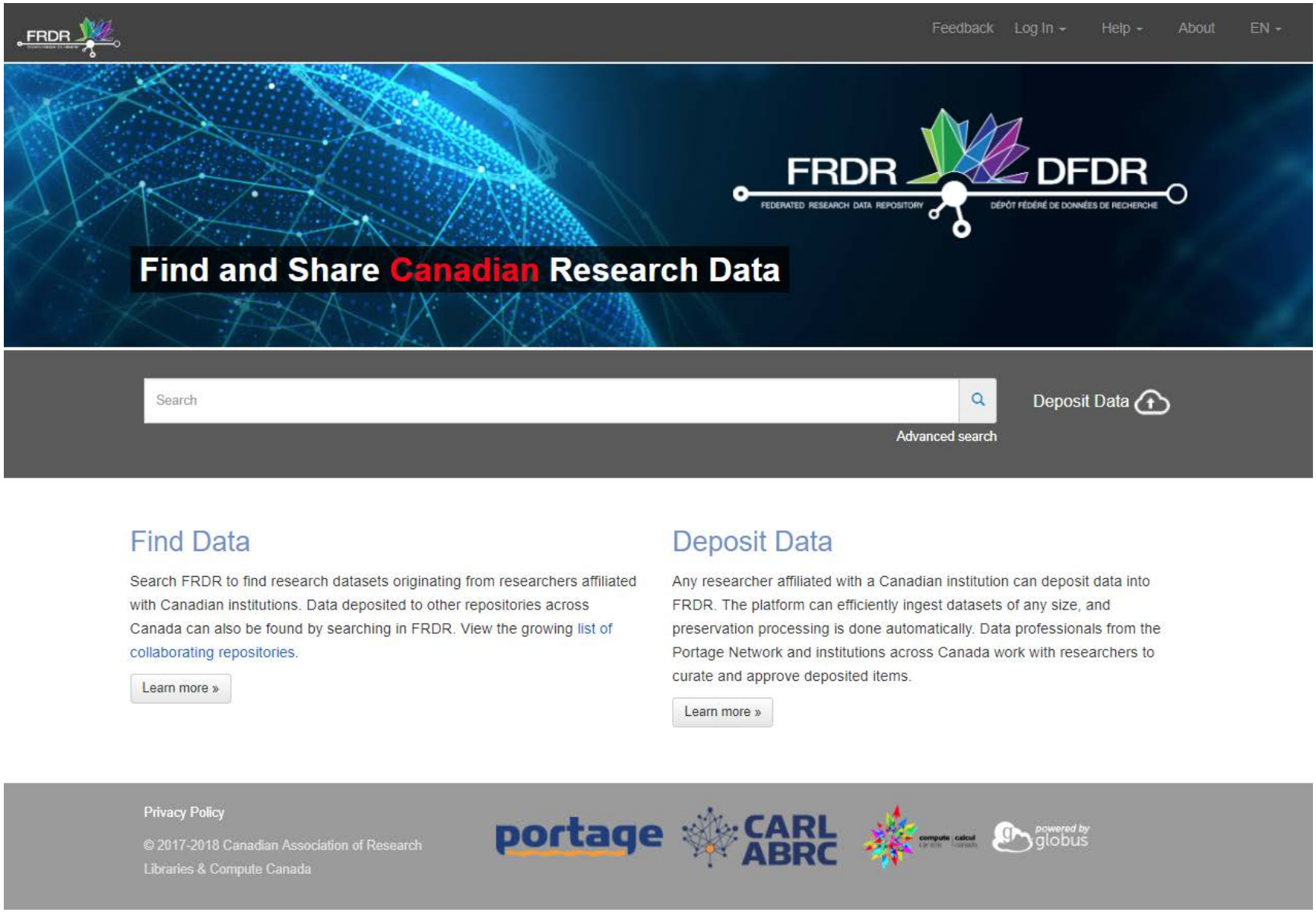
1 to 6 of 6 Files

Download ▾ Request Access

Codebook-CPSurveillanceArchive.pdf
Adobe PDF - 23.5 MB - Aug 21, 2018 - 0 Downloads
MD5: 4ca2f7bf5ceabc1c117cbbeb51347d8a
Codebook
Documentation

Request Access

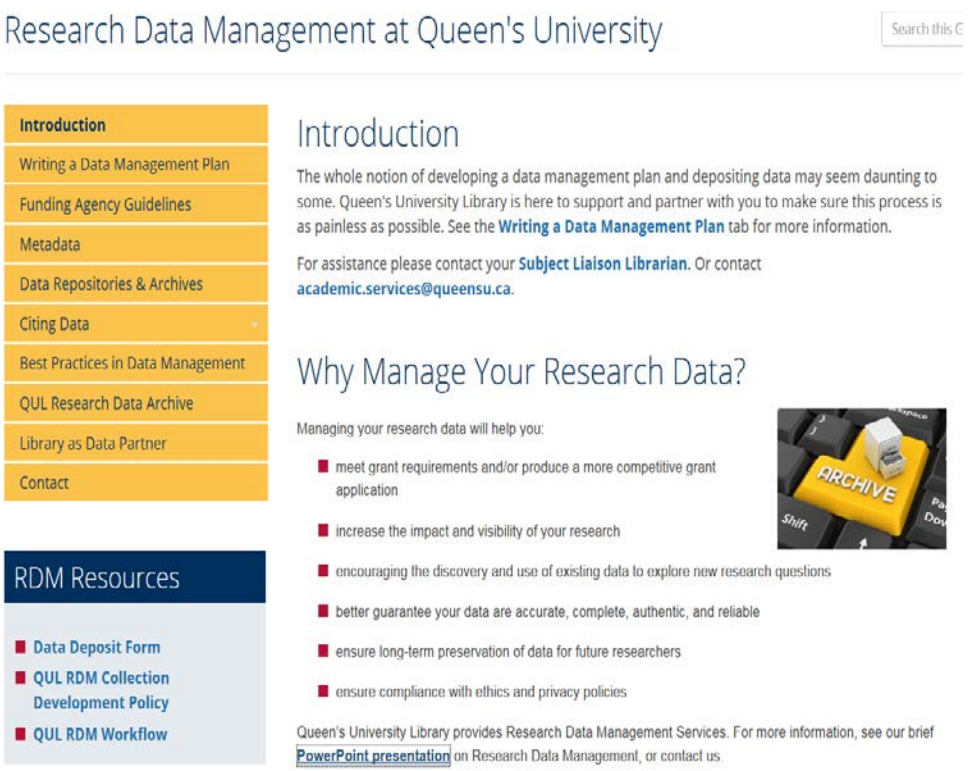
Federated Research Data Repository (FRDR)



Resources

QUL Research Data Management Services

- **Planning for data management**
  - Data management plans and grant applications
- **Preparing your data**
  - Coding, formatting, anonymizing, and more, conversion to non-proprietary formats
- **Documenting your data**
  - Describing your data so others can understand and use it (guide, methodology, codebook, data dictionary)
  - File naming
- **Archiving your data**
  - Finding an appropriate repository or archive
- **Making your data discoverable**
  - Promoting open access, indexing, assigning DOI's



Researcher Training Program

The Research Training Program is an expanding set of self-serve training modules that will help you to successfully manage your digital research and scholarship effectively throughout the research lifecycle and at your point of need:

- **Getting Published**
- **Open Publishing Support**
- **ORCID and your Research Profile**
- **Manage your Research Data**

Open Scholarship Services at Queen's University Library

Post Date: **Mar 5, 2019**

Queen's University Library is offering a brown bag lunch series to introduce researchers to the services and expertise that support them throughout the life of their research and in their teaching. The series will overview the supports provided by the Open Scholarship Services division including research data management, open access publishing, copyright and access to statistical, government and GIS information.

Sessions will be held once a month from March to May in Stauffer 121 from 12:30-1:30. The first series will include:

- March 27: **Open Scholarship Services Showcase**
- April 23: **Introduction to Data, Government Information and Maps and GIS**
- May 29: **Introduction to Research Data Management and Scholarly Publishing**

Refreshments will be provided.

For questions, contact **Courtney Matthews**. Please register via the individual session page as seating is limited.

Resources

- DRAFT Tri-Agency Research Data Management Policy For Consultation
- Portage Network
  - DMP Assistant
  - RDM Primer
  - 'Good Enough' Research Data Management
  - 'Good Enough' Dataverse
- Scholars Portal Dataverse
- QUL Research Data Management Guide
- CIHR Research Data Management learning module
- UBC Research Data Management DataGuide

Contact us

Library

Courtney Matthews, Head
Open Scholarship Services

Alexandra Cooper, Data Services Coordinator

Meghan Goodchild, Research Data Management Systems Librarian

Will Roy, Open Scholarship Services Librarian

open.scholarship.services@queensu.ca

Centre for Advanced Computing

Chris MacPhee, Assistant Director, Centre for Advanced Computing

chris.macphee@queensu.ca