Queen's University Biological Station Research Data Management Workshop

May 14, 2018

Francine Berish, Morag Coyne, Courtney Matthews

Learning outcomes

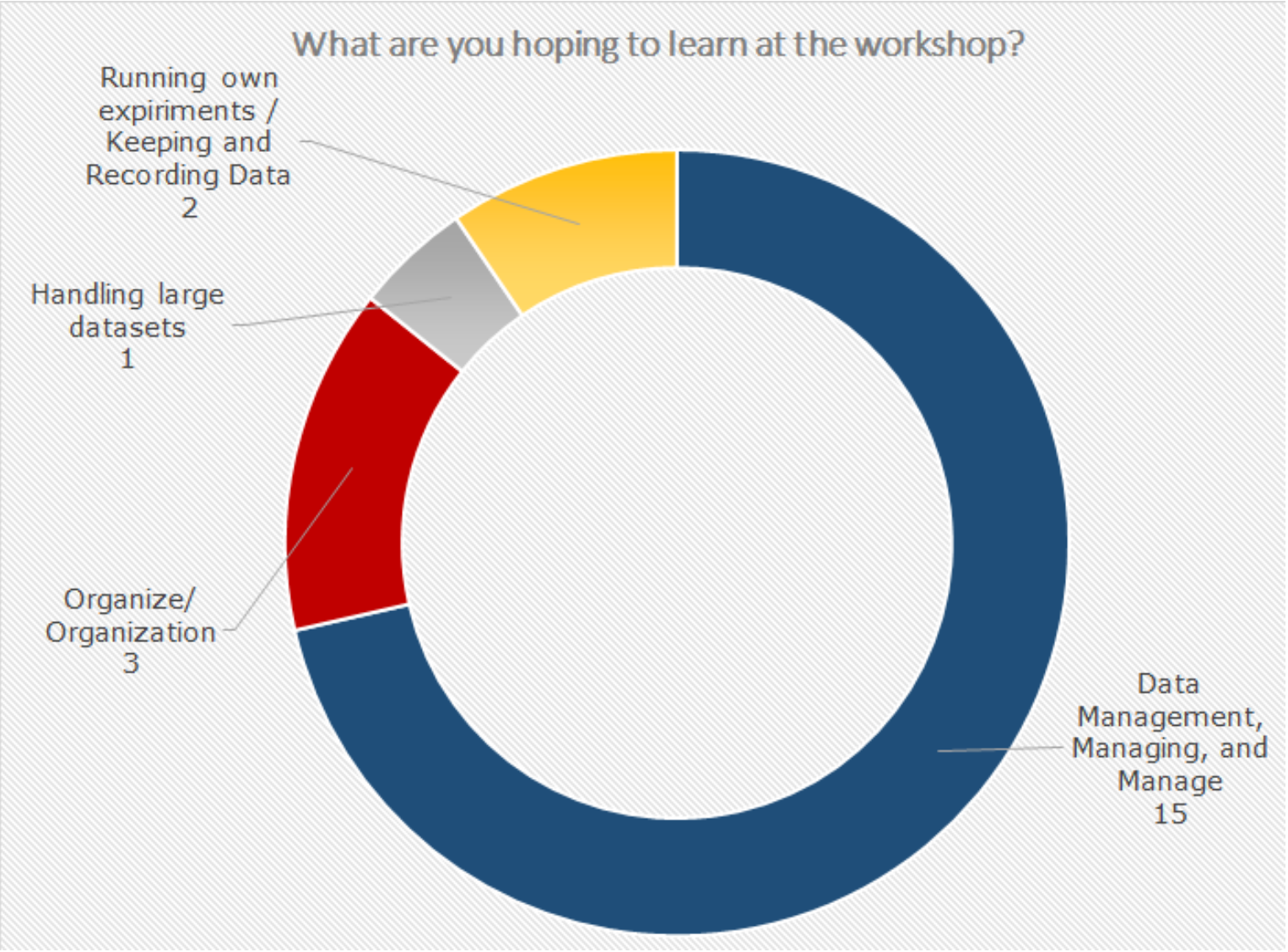
Part 1

- Pre-workshop feedback
- Introduction to RDM
- Gain a working knowledge of RDM

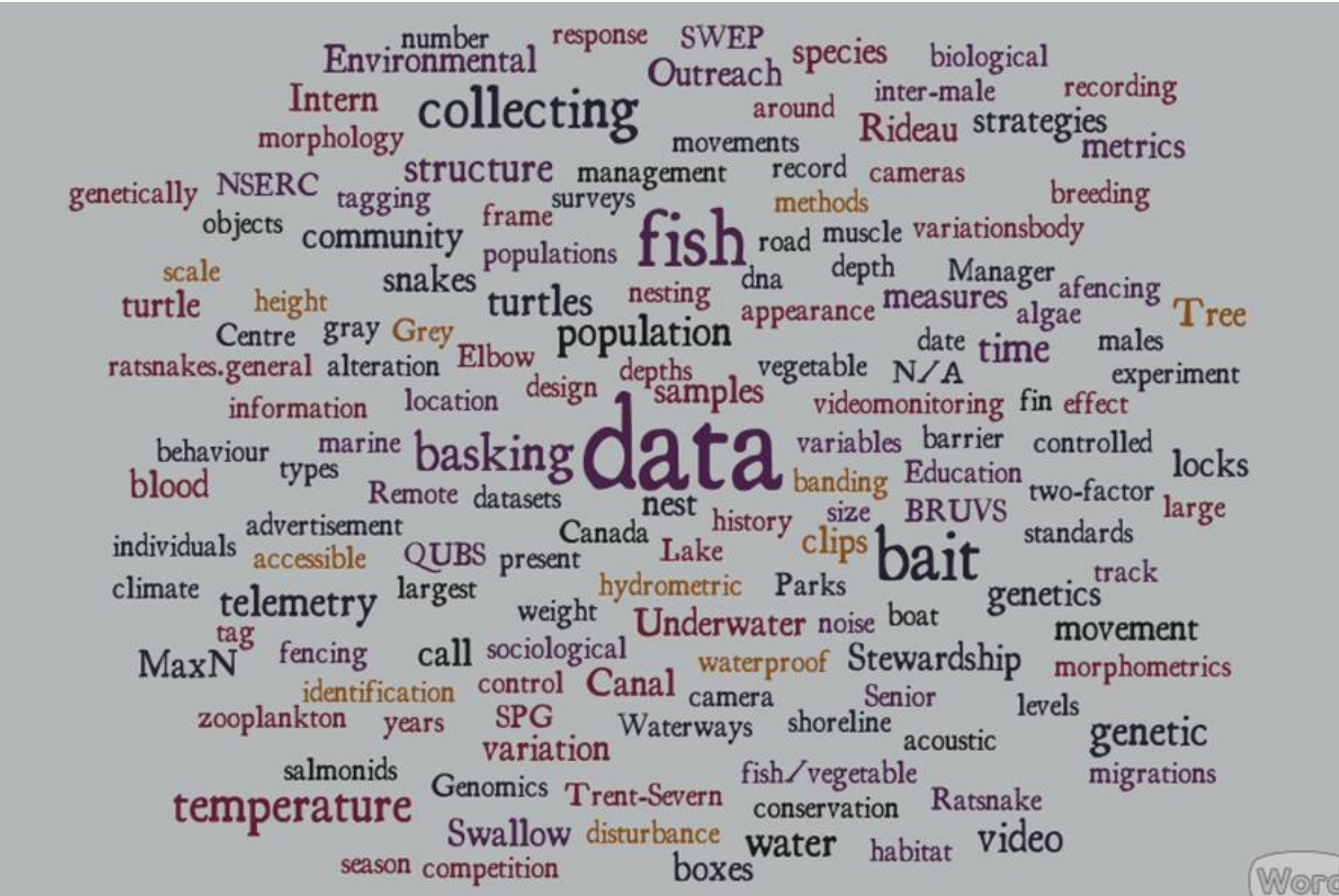
Part 2

- Learn how to write a DMP
- Learn about options for sharing your data
- Draft a DMP using the DMP Assistant
- Deposit draft data and metadata into Scholars Portal Dataverse

Results from feedback



Results from feedback: Please describe your research project, including the kinds of data you will need to collect or that you have collected.



So you're publishing in Ecology

Keywords: repositories, institutional repository, discipline-specific repositories, funded research, research data management (RDM), data management planning, research lifecycle, research translation , etc.

www.digitalbevaring.dk

Panda video

https://youtu.be/N2zK3sAtr-4?t=183

Research Data Management

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

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

What is 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

See the Portage RDM Primer

Examples of Research Data

Spreadsheets, images, output from sensors and instruments, transcripts, surveys, software source code and tools, video, and observation logs

What is 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.

See the Portage RDM Primer

What is Research Data Management (RDM)?

- 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.

See the Portage RDM Primer

What is the Research Data Lifecycle?

See the Portage RDM Primer

Why should I manage my research data?

Align with funding agency recommendations including the Tri-Agency Statement of Principles on Data Management

Why should I manage my research data?

- Meet funding agency requirements including the Tri-Agency's Open Access Policy on Publications
- Plan for your data's long-term preservation and usability
- Share your data for the advancement of research and the benefit of society

Why should I manage my research data?

- Get credit for your data and increase its impact and accessibility
- Ensure your data's long-term preservation
- Make your data easier to cite with a Digital Object Identifier (DOI)
- Meet publisher requirements: *PLOS, Springer Nature*

Open Researcher and Contributor ID (ORCID)

Queen's University

ORCID

- are unique persistent digital identifiers
- enable disambiguation of same-name authors
- accurately link authors to their achievements (education, employment, funding awards, publications)

Example:

- 0000-0001-5109-3700
- URL: http://orcid.org/0000-0001-5109-3700

orcid.org

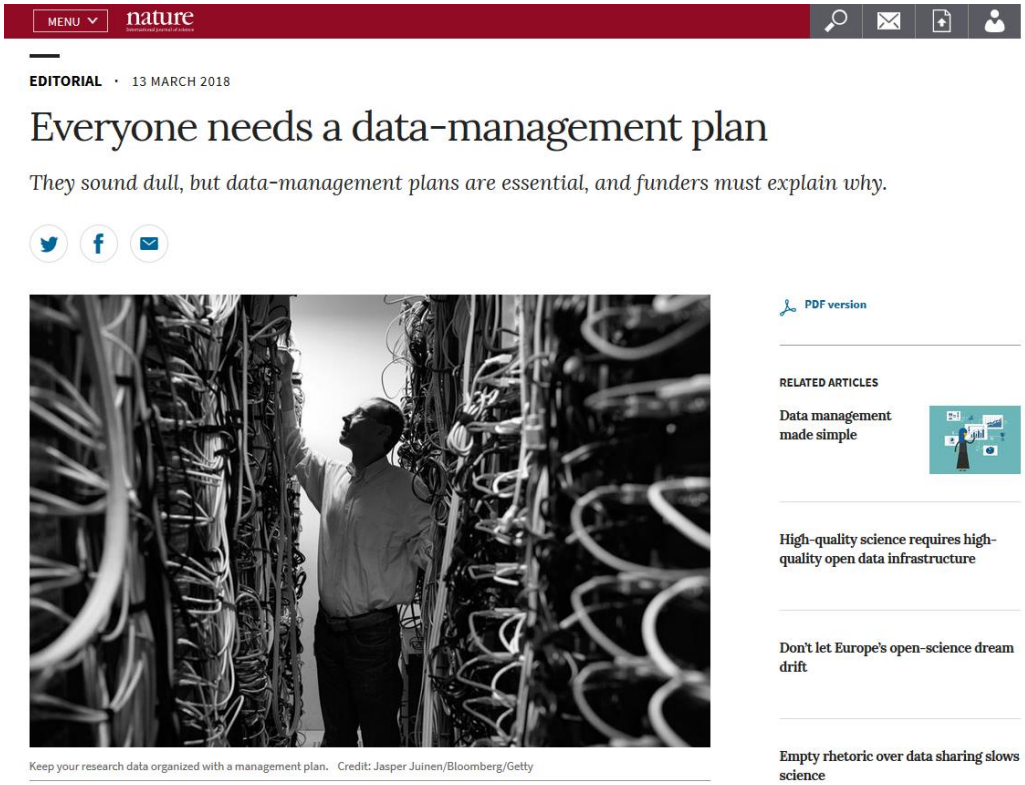
- is a non-profit, community driven organization platform/provider sustained by organizational memberships

Register for an ORCID at https://orcid.org/

Why does Portage care?

- The RDM landscape
  - Tri-Agency Draft Statement of Principles on Digital Data Management
  - Portage Network https://portagenetwork.ca/
  - Services at Queen's informed by Queen's Research Data Management Surveys (http://hdl.handle.net/10864/11651 and Summary of Findings Report http://dx.doi.org/10.5683/SP/E6LSVQ)

Why does Springer Nature care?



MENU nature

EDITORIAL · 13 MARCH 2018

Everyone needs a data-management plan

They sound dull, but data-management plans are essential, and funders must explain why.

Keep your research data organized with a management plan. Credit: Jasper Juinen/Bloomberg/Getty

PDF version

RELATED ARTICLES

Data management made simple

High-quality science requires high-quality open data infrastructure

Don't let Europe's open-science dream drift

Empty rhetoric over data sharing slows science

https://www.nature.com/articles/d41586-018-03065-z

SPRINGER NATURE springernature.com

Tracking glacier response to climate change

Research Data

PRACTICAL CHALLENGES FOR RESEARCHERS IN DATA SHARING

ADVANCING DISCOVERY

White paper

https://doi.org/10.6084/m9.figshare.5971387

"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.).

Queen's UNIVERSITY

DMP Assistant

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

How do journals want to you deposit data with them?

Adriana has collected a list of data requirements from different journals. See her list for more information.

Some journal examples:

Molecular Ecology.

It is mandatory for publication that data is publicly accessible in a repository. We require that authors include a 'Data Accessibility' section after the References (see 'Preparing the Submission' Section for details). This section must be present at initial submission, and data archiving must be completed before final acceptance.

Data are important products of the scientific enterprise, and they should be preserved and usable for decades in the future. As such, Molecular Ecology requires authors to archive the data supporting their results and conclusions along with sufficient details so that a third party can interpret them correctly. Papers with exemplary data and code archiving are more valuable for future research, and, all else being equal, will be given higher priority for publication.

Adriana's list continued...

Evolution requires, as a condition for publication, that data supporting the results in the paper should be archived in an appropriate public archive, such as Dryad, Figshare, GenBank, TreeBASE, the Knowledge Network for Biocomplexity or other suitable long-term and stable public repositories.

Data are important products of the scientific enterprise, and they should be preserved and usable into the future. Authors may elect to have the data publicly available at time of publication, or, if the technology of the archive allows, may opt to embargo access to the data for a period of up to a year after publication. Exceptions may be granted at the discretion of the Editor in Chief, especially for sensitive information such as a human subject data or the location of endangered species.

What is metadata?

Metadata is '*data about data*'. Put another way, it is the information necessary to make your data '**independently understandable**'. Using established metadata standards will help make your data **discoverable**, **citable**, and **ready-to-use** by others.

Basic Metadata Elements

- Title
- Creator
- Date Created
- Format
- Subject
- Unique Identifier (*ideally, a Digital Object Identifier, or DOI*)
- Description of the specific data resource
- Coverage (*spatial or temporal*)
- Publishing Organization
- Type of Resource
- Rights/Licensing/Ethics approval
- Funding/Granting Agency

From http://guides.library.queensu.ca/rdm/metadata

Metadata standards:

Examples:

- FGDC (Federal Geographic Data Committee)
- DDI (Data Documentation Initiative)
- Dublin Core
- Darwin Core (see example)
- ABCD (Access to Biological Collections Data)
- CSDGM (Content Standard for Digital Geospatial Metadata)
- EML (Ecological Metadata Language)

Advantages:

- Ensure you have a complete, standard set of information about each part of your data
- Enable your dataset to be organized with other datasets

Data Repositories and Archives

Generalist repositories

- Queen's Dataverse (contains QUBS Dataverse)
- Dryad Digital Repository

Specialist repositories

- For DNA and RNA sequences - Genbank
- For DNA and RNA sequencing data - NCBI Trace Archive

Check the journals you wish to publish with to see which repositories they recommend!

To find more repositories, check https://www.re3data.org/.

QUBS Dataverse

Scholars Portal Dataverse

Fr (Beta)

Search all dataverses... Find About User Guide Support Sign Up Log In

Queen's University Biological Station Data Archive Dataverse (Queen's University)

Scholars Portal Dataverse > Queen's University Dataverse > **Queen's University Biological Station Data Archive Dataverse**

Contact Share

Welcome to the Queen's University Biological Station (QUBS) Data Archive! This is a new resource that we are developing with Station staff and researchers to host and archive data from projects undertaken at QUBS. Research spans ecology, evolution, resource management and conservation, GIS, climate data, and environmental science. This is a newly developed resource, and we will undoubtedly experience growing pains as this unfolds.

Note that the data in this archive are of two types.

1. Data that are freely available for immediate download. We only ask that you acknowledge the station.
2. Data that will belong to an active researcher from whom you must request permission.

He or she will make the ultimate decision as to the availability of their data. We hope that the availability of these data fosters collaborations between QUBS researchers and those from other institutions, encourages new research directions, and raises the profile or our facility.

Search this dataverse... Find Advanced Search

Dataverses (0)
Datasets (13)
Files (91)

Publication Date

1 to 10 of 13 Results

Sort

QUBS Point Climate Data, 1993-2010

Jan 10, 2018

Queen's University Biological Station, 2013, "QUBS Point Climate Data, 1993-2010", hdl:10864/10402,

Resources

- Research Data Management at Queen's
- Portage
- "Good Enough" Research Data Management
- DMP Assistant
- Dataverse guide
- UBC DataGuide

- Questions? open.scholarship.services@queensu.ca