

---

# R4R@Q RDM Brown Bag Info Services - Developing a DMP

*A Data Management Plan created using DMP Assistant*

**Creator:** Alicia Cappello

**Principal Investigator:** Dr. Jasmine Riley

**Data Manager:** Grad Student 1, Grad Student 2

**Project Administrator:** Grad Student 1, Grad Student 2

**Contributor:** Lab Assistant 1, Lab Assistant 2, Undergrad Student 1

**Affiliation:** Queen's University

**Funder:** NSERC

**Template:** Portage Template

## **Project abstract:**

**WHAT:** Dr. Jasmine Riley is starting a new project to observe the birds in and around the oil sands in Fort McMurray, Alberta. Dr. Riley will be submitted a funding application to NSERC in 2023 for this project and is required to include a DMP.

**HOW:** Birds will be observed using two methods: (1) visual observations recorded via online form on a tablet, and (2) panoramic photographs taken with a digital SLR camera using specialized equipment. The online form (which has yet to be developed) will be created and maintained by grad students. Both types of observations will be conducted by undergrad and grad students.

**WHERE/WHEN:** There will be 10 locations within the oil sands where observations will take place over the 5-year project period. Each of the 10 locations will be observed by 2 students for a period of 15 minutes, twice per weekday (AM and PM) over the course of 4 months (May through August) for a total of 10 locations x 2 students x 2 obs/day x 10 weekdays x 8 bi-weekly periods = 3,200 observation **periods** per year, or 16,000 observation **periods** for the life of the project. *It is likely that dozens of bird sightings will be observed during each 15-minute period.*

**WHO:** The Principal Investigator is Dr. Riley. Her lab will have 2 grad students per year, 10 undergrad students per year, and 2 lab assistants for the life of the project. Some of the students may return for multiple years.

Dr. Riley, as the PI, is the overall decision-maker for the project. Grad students will develop the observation protocols, the online form, find the observation locations and create the schedule, review the data collected, and troubleshoot data-related problems. Undergrad students will conduct the majority of the visual and photographic observations and may need to make corrections to data they've entered. The lab assistants will be responsible for creating and maintaining access to the data.

**Identifier:** n/a

**Start date:** 01-01-2023

**End date:** 31-12-2027

**Last modified:** 22-11-2022

# R4R@Q RDM Brown Bag Info Services - Developing a DMP

---

## Data Collection

### What types of data will you collect, create, link to, acquire and/or record?

For this project, we will be creating and collecting the following **RESEARCH** data:

1. Survey responses - tabular data in CSV, TSV, or XLSX format
2. Survey response analysis - SPSS formats
3. Panoramic photographs - in JPG format

We will also be creating and collecting the following data/information that will be used as part of our project:

1. Qualtrics online survey form/questionnaire - only available within the Qualtrics application
2. Images or graphics in PNG or JPG format, exported from Excel or SPSS
3. Observation protocols/procedures - in both DOCX and PDF formats
4. Observation locations - marked by GPS location in Google Maps, may be downloaded as image files
5. Observation schedule - created and maintained in XLSX format
6. Other documentation as required, including literature - in either DOCX and PDF format

### What file formats will your data be collected in? Will these formats allow for data re-use, sharing and long-term access to the data?

As noted in the previous section, the following file types will be used in this project:

- CSV/TSV
- XLSX
- SPSS
- JPG
- PNG
- DOCX
- PDF

In general, CSV, TSV, JPG, PNG, and PDF file formats are easy to share and can be opened in free software. They will be used for preservation purposes.

On the other hand, XLSX, SPSS, and DOCX are formats that are easy to share, but require proprietary software to open and use. XLSX and SPSS files will be exported into CSV or TSV formats for preservation purposes, and DOCX files will be converted to either TXT or PDF format for preservation purposes.

More information about longer term file formats can be found in the [Digital Preservation Handbook](#) and on the [Government of Canada website](#).

### What conventions and procedures will you use to structure, name and version-control your files to help you and others better understand how your data are organized?

In order to ensure the proper organization and identification of our data files, we will use the following naming conventions for folders and files:

At the end of each day, that day's survey responses -- from all locations -- will be exported from Qualtrics into CSV or TSV format. These files will have the following naming convention:

**SurveyObs\_YYYY-MM-DD\_HH:MM.csv**

Also at the end of each day, the photos taken that day will be uploaded to our SharePoint folder (from the camera's SD card) using the following naming convention:

**Photo\_LocationCode\_YYYY-MM-DD\_HH:MM.jpg**

At the end of each year's observation period, all survey responses will be combined into one CSV or TSV file for annual analysis using the following naming convention:

**SurveyObs\_YYYY.csv**

Once imported into either Excel or SPSS, the file names will be as follows:

**SurveyObsAnalysis\_YYYY\_Versionxx-xx.xlsx**

More information about [file and folder naming conventions](#) can be found on the Queen's University Records Management and Privacy Office website.

## Documentation and Metadata

### What documentation will be needed for the data to be read and interpreted correctly in the future?

The following documentation will be developed for this project that will assist current and future project team members, as well as future users of the data, understand:

- **HOW** the data was collected, including a copy of the questionnaire and any changes that occurred to the questionnaire throughout the life of the project, and the exact technical specifications of the camera and equipment that took the photos.
  - This documentation will also include a copy of all observation protocols, training materials, reference materials, etc. used by the students throughout the project.
- **WHEN, WHERE, and WHO** collected the data, including a list of location codes with their conventional names and GPS coordinates, the dates and times when observations were made at each location, and the names of the specific students who conducted the observations.
  - This list will also include any comments or notes from the students regarding any unusual events or issues they encountered during the observations. For example, if they were unable to access a specific site due to a road closure, or if the camera malfunctions, etc.
- **WHAT** happened to the data throughout the life of the project, including the various formats in which the data was collected, how the data was cleaned or manipulated or reformatted, any file format conversions that may have been conducted and what software was used, and the specific names and version numbers of all the software used to view, clean, reformat, or convert the files.
- **LIST** of all files and an explanation of the naming convention of those files.
- **LIST** of any variable or column headings used within the files and their associated descriptions.

More information about how to create documentation for your data can be found at the following websites:

- [Document your data section](#) from the UK Data Service
- [Guide to writing "readme" style metadata](#) from Cornell University
- [Creating a README for your dataset](#) from the University of British Columbia
- [How to Make a Data Dictionary](#) from OSF Support

### How will you make sure that documentation is created or captured consistently throughout your project?

Initial documentation will be created through a collaboration of the two graduate students and the PI in the first year of the project, before observations begin.

In January of each year, all documentation will receive a thorough review by the PI and the current graduate students to ensure it's up-to-date and reflects reality.

Throughout the year, if needed, the graduate students will update the documents as problems arise and solutions are found.

All documents will be saved using a similar file naming convention as was mentioned earlier so the versions can be tracked throughout the project. Annual updates will be considered MAJOR versions (i.e., **XX.xx**) and mid-year updates will be considered MINOR versions (i.e., **xx.XX**). Changes made throughout the year will be done using track changes and explanations for the changes will be entered via Comments. These tracked changes and comments will be removed annually when a new major version of the document is created.

### If you are using a metadata standard and/or tools to document and describe your data, please list here.

We will review the various metadata standards and schemas available to use (using this [Disciplinary Metadata](#) website) to determine which standard/scheme works best for our project.

Once a specific metadata standard/schema is identified, we will use a TXT document to track the specific metadata fields we use.

Once it comes times to deposit our dataset into a data repository, we will use the [Dataverse North Metadata Best Practices Guide 3.0](#) to determine the best fields to use within Borealis.

## Storage and Backup

### What are the anticipated storage requirements for your project, in terms of storage space (in megabytes, gigabytes, terabytes, etc.) and the length of time you will be storing it?

Storage requirements for survey responses and their associated analysis files will be relatively small and can easily be saved on our project's OneDrive or SharePoint site.

Storage requirements for the panoramic photographs will be significant, we may have up to 1,600 photographs per study year\*, or 8,000

photographs by the end of the 5-year project.

*\*10 observation locations x 2 observations per day x 5 days of observations per week x 16 weeks of observations per year = 1,600*

Due to the large size/resolution of each panoramic photograph, we will assume each file will be at least 1GB in size. Therefore, 1,600 photographs will require 1,600GB or 1.6TB of storage space PER YEAR. Over the course of the project, this will mean a total storage requirement of 8TB. Some photographs may require multiple copies if some are annotated or edited in any way.

We will investigate storage options for these photographs to determine our best options. We may be able to obtain space via the [Centre for Advanced Computing](#) (CAC) or [ITS](#). We may even be able to purchase multiple external hard drives. (Because this is not sensitive data, storing the photos on an external hard drive won't have any privacy or confidentiality issues.)

### **How and where will your data be stored and backed up during your research project?**

As per [Queen's ITS File Sharing and storage information](#), we will start with a project SharePoint site as our main storage location. Access to this storage location for project team members will be controlled by our Lab Assistants. All project team members will be provided access so they can save and review any documents in our project's folder. Because this location is on a Queen's server, it is regularly backed up by ITS and lost information is recoverable.

Survey data entered via the online forms during visual observations will go directly into the project's Qualtrics account, where a copy will always be kept. On a regular basis, the data from Qualtrics will also be exported into CSV or TSV format and saved in our SharePoint site so it can be accessed without having to use Qualtrics.

Photographs, as they are taken, will be saved on the camera's SD card. We will have 10 SD cards, 1 for each day in a 2 week period. SD cards will then be cycled through so that every set of photos are stored on an SD for 10 days. On a daily basis, assuming a decent internet connection, the photos from each SD card will also be uploaded to our SharePoint site for more permanent storage. We will need to [request additional storage space on our SharePoint site](#) in order to have sufficient space to store the photos. If necessary, we will obtain an account and storage space via the [CAC](#) where we can store the photos regularly.

In addition to uploading the photos from the SD cards to SharePoint, they will also be regularly transferred to two external hard drives. This will help to ensure there are copies on a device other than the SD card if upload to the SharePoint site (or elsewhere) cannot happen quickly due to internet connectivity issues.

### **How will the research team and other collaborators access, modify, and contribute data throughout the project?**

Due to the non-sensitive nature of this data, all team members will be provided access to the SharePoint site and any other storage locations we use. Each team member will have the ability to upload/save files into those locations and read/review/edit items in those locations.

## **Preservation**

### **Where will you deposit your data for long-term preservation and access at the end of your research project?**

Assuming we will receive NSERC funding for this project, and in line with the [Tri-Agency RDM Policy](#), will we deposit all digital research data and metadata that supports our conclusions as published into a research data repository. We will deposit this data to coincide with the publication of any articles we submit.

After a review of [re3data.org](#), and in consultation with Queen's library staff, we have decided to deposit our research data in the [Queen's University Collection](#) of Borealis, the Canadian Dataverse Repository. As Queen's researchers, this repository is free for us to use and is fully supported by library staff.

If we need to preserve a large number of the photos taken for this project and sufficient space cannot be obtained in Borealis, we will deposit those photos in the [Federated Research Data Repository](#) (FRDR), which is also free to use for Queen's researchers and generally has a larger file storage limit.

**More information about selecting a research data repository and data deposits will be provided in our next R4R@Q RDM Brown Bag Info Series session on December 14th at 12pm. You can register for that event on the [VPR's website](#).**

### **Indicate how you will ensure your data is preservation ready. Consider preservation-friendly file formats, ensuring file integrity, anonymization and de-identification, inclusion of supporting documentation.**

As mentioned previously, we will try to convert all file formats to those that are not related to proprietary applications for the purpose of preservation and longer term storage.

Due to the non-sensitive nature of this data, we do not have to de-identify any of the data before it is preserved in a research data repository.

## Sharing and Reuse

### What data will you be sharing and in what form? (e.g. raw, processed, analyzed, final).

The exact nature of the data we will share for this project is still to be determined and may change throughout the life of the project. In general, we will likely share our cleaned and final survey responses and analyses in CSV or TSV format. We will also likely share some of our photos - specifically those referenced in publications - in JPEG format. Some of these photos may be annotated or edited in order to highlight specific items we reference in a publication.

### Have you considered what type of end-user license to include with your data?

Both Borealis and FRDR use [Creative Commons licenses](#) for research datasets. Due to the nature of the data we will be sharing, we will likely use a CC-BY license, which will allow users to distribute, remix, adapt, and build upon the data as they require, with attribution to the research team.

### What steps will be taken to help the research community know that your data exists?

Both Borealis and FRDR allow metadata harvesting of all of their datasets. For Borealis, this metadata is then available for searching on other platforms, such as FRDR, Google Dataset, Web of Science Data Citation Index, Mendeley Data, and Harvard Dataverse. Therefore, any data deposited in either research data repository will appear in search results on any of these platforms, with a link back to the repository for data access. The digital object identifier (DOI) for our dataset will also be included in any publications for which that data was used to draw conclusions. This will ensure that anyone who finds and reads our publications will also be able to find and use our data.

## Responsibilities and Resources

### Identify who will be responsible for managing this project's data during and after the project and the major data management tasks for which they will be responsible.

**Graduate students** working on this project will be the main data managers for the project's data and will be responsible for:

- requesting proper data access for new members of the project team
- ensuring undergrad students are properly trained on how to collect and store the observational data
- ensuring the data is being transferred/uploaded to the proper locations on a regular basis and that there is enough space in those locations for continued transfers/uploads
- regularly downloading survey data from Qualtrics in CSV/TSV format and saving it to SharePoint
- ensuring our data protocols, documentation, and data management plan are up-to-date and accurate
  - this includes the full annual review with the PI and mid-year reviews when things change or problems occur
- troubleshooting data-related problems as they occur throughout the project
- escalating unresolved data-related problems to the PI, if required

**Lab assistants** for this project will be responsible to control access to our data storage locations. This includes setting up new team members with access and removing outgoing team members from having access.

The **principal investigator** will be ultimately responsible for the research data, and will review the data management plans and data protocols on an annual basis with the graduate students. The PI will also work to fix any data-related problems when escalated by graduate students.

### How will responsibilities for managing data activities be handled if substantive changes happen in the personnel overseeing the project's data, including a change of Principal Investigator?

If something happens to the PI for this project, the project will likely be stopped or put on hold. The PI would then be responsible for determining what happens with the data collected to that point, if it will be kept, and where it will be kept.

It is expected that both graduate and undergraduate students will change throughout the life of this project. This data management plan, plus other documentation and protocols, are being put in place to ensure that any new team members can be quickly and easily brought up to speed on data-related tasks and requirements when they join.

The data management plan, documentation and protocols will be reviewed and updated annually by the PI and grad students to ensure they continue to reflect reality.

**What resources will you require to implement your data management plan? What do you estimate the overall cost for data management to be?**

We will require the following resources in order to implement our data management plan:

- Additional storage space on our SharePoint site or via the CAC.
- Two external hard drives (at least 2TB in size) per year to use as backup for photos while in the field - ~\$100 each.
- Ten SD cards (at least 128GB in size) to use for the digital camera while in the field - ~\$50 each.
- Cellular-enabled tablets to record observations via survey directly to Qualtrics - ~\$500 each (plus monthly fees).

We will need to assess the availability of an internet connection and its associated speed while in the field to determine if we also require a cellular-based hotspot device for daily data uploads. We will update this DMP when this assessment has been completed.

At this time, we are unsure how much it will cost to expand our overall storage space either via our SharePoint site or the CAC. We will update this DMP when this info is available.

For the other storage and connectivity items, we estimate it will cost the project \$1,700 in the first year, plus ongoing monthly fees for the tablets. We will also need to purchase additional hard drives every year for approximately \$200. And it is highly likely we'll need to purchase additional SD cards if any become corrupt, damaged or lost.

## **Ethics and Legal Compliance**

**If your research project includes sensitive data, how will you ensure that it is securely managed and accessible only to approved members of the project?**

Our research does not include sensitive data.

If it did include sensitive data, we would submit an ethics application via the Queen's [General Research Ethics Board](#) (GREB) for approval. We would also ensure access to sensitive data is controlled as required, based on the type of data. And we would not deposit any of our sensitive or confidential data into a research data repository at the end of the project, unless it could be appropriately de-identified. Finally, we would also review the [Sensitive Data Toolkit](#) provided by the Digital Research Alliance of Canada for additional information on how to handle sensitive data.

**If applicable, what strategies will you undertake to address secondary uses of sensitive data?**

We will not be using sensitive secondary data for this project.

If we were, we'd review the [Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans](#) (TCPS2) for more information on the handling and use of sensitive secondary data in research project.

**How will you manage legal, ethical, and intellectual property issues?**

**Ethical Issues** - we do not have any ethical issues associated with this project. And if we did, we'd following the statements made in the previous two responses.

**Legal Issues** - if we are unsure if we have any legal issues with this project, or we encounter legal issues, including the requirement of a data sharing agreement, we will contact Queen's [Research Legal Services](#) for assistance.

**IP Issues** - all project team members will be included as co-authors on all publications submitted for this project. And, as previously stated, we will share our data using the Creative Commons CC-BY license, which allows other researchers to use our data as long as they cite the project.

**Indigenous Data Sovereignty** - this project is not being conducted with any Indigenous communities. Should we opt to partner with an Indigenous community on this project in the future, we will review [Chapter 9 of the TCPS2](#) and the [First Nations Principles of OCAP](#), and we will contact the [Office of Indigenous Initiatives](#) at Queen's University for advice and guidance.