Data Management Plan for Ecohydrology Research Group (v. 1.0)

Abstract

This exemplar DMP was created by the Ecohydrology Research Group at the University of Waterloo, with the purpose of providing a standardized DMP model for its many multi- and cross-disciplinary research projects. Here, fundamental processes and procedures core to the ERG have been incorporated as a starting point for its faculty, staff, and HQPs (Highly Qualified Personnel) when developing their research projects. This exemplar demonstrates the utility of a model DMP being used by a research group or collective to maintain best practices in data management.

Principal Investigators:

Philippe Van Cappellen, Fereidoun Reza Nezhad

Administrative Details

Project Name:

Ecohydrology Research Group Data Management Plan

Principal Investigators / Researchers:

Philippe Van Cappellen (Orcid ID: https://orcid.org/0000-0001-5476-0820)

Fereidoun Reza Nezhad (Orcid ID: https://orcid.org/0000-0002-9608-8005)

Description:

The Ecohydrology Research Group (ERG) at the University of Waterloo carries out fundamental research in support of the wise use of water resources, that is, one that balances society's water needs with those of natural ecosystems. ERG's research activities cover a vast range of spatial and temporal scales, from molecular-level studies on the processes determining the chemical forms and bioavailability of nutrients and pollutants to global scale assessments of anthropogenic perturbations of hydrological and biogeochemical cycles.

Faculty, staff and all Highly Qualified Personnel (HQPs – students, postdoc, research scientist/associates) in ERG are committed to data management practices to ensure that ERG data, software, code, inputs and outputs ("the data"), are safely stored, preserved and easily accessible for future re-use. ERG is aligning its Data Management Plan (DMP) with the common approach being pursued at the University of Waterloo.

Data Management Plan for Ecohydrology Research Group (v. 1.0)

In consultation with supervisors and laboratory technicians, all HQPs are expected to adopt this DMP in the conduct of their research project. They may also be asked to modify certain parts of their proposed research plans using the guidelines provided by Portage Network or to meet funding agency or contractual requirements. If, for a given HQP, significant deviations from the ERG DMP are expected, the HQP will incorporate an individual DMP in their research proposal, which will be revisited, and if needed, updated as their research project progresses.

If you are a member of a research group using a shared, standardized DMP, it is important to mention this in the preamble and provide details on any deviations your individual project took from the shared plan.

Institution:

Water Institute, University of Waterloo

Data Management Plan for Ecohydrology Research Group (v. 1.0)

Data Collection

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

The Ecohydrology Research Group (ERG) generates data from existing documents, fieldwork, laboratory experiments, secondary data sources (including remote sensing), and numerical modelling. Data types include laboratory, modelling, hydrological, chemical and biological data (hourly, daily, weekly, seasonal, annual) related to ERG's research goals. Additional data include field site observations/images, sampling location through Global Positioning System (GPS) coordinates/maps, climate data and modelling input files, parameters and output files. Generated data are used in model development, calibration and validation. Model tasks are done concurrently with the generation of the data from lab experiments and field observation within ERG. This approach allows us to discard unacceptable data and, if needed, repeat/redesign experiments.

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?

The Ecohydrology Research Group (ERG) generates several types of primary data using various proprietary file formats from the instruments or numerical models. Highly Qualified Personnel (HQP) are expected to convert these files to non-proprietary formats, such as Comma-Separated Values (.csv) or Text (.txt) whenever possible to help facilitate future re-use.

Providing definitions for acronyms in each new section or on each page prevents readers unfamiliar with the terminology from having to revisit previous sections.

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?

Directory and file naming conventions are established at the beginning of a project and may incorporate discipline-specific conventions for the variables collected. Version control conventions will also be established. File names include the project name (in shortened form) and a brief description of the file's content. For example:

PX_WP1_WaterQuality_Thames_River

Document versions should be sequentially named (with file names ending in v1, v2 etc.). An example is the following:

WSP_WaterQuality_Thames_River_raw_200617_v1

All dates should follow ISO 8601 date format (YYYY-MM-DD) and study site names will follow the Canadian Geographical Names DataBase where applicable.

Providing examples of controlled file naming structures helps demonstrate more specifically how you'll keep your data organized and understandable.

What Data Quality Assurance and Quality Control methods will be used?

This section is not a part of the Portage DMP default template. Although templates are not adaptable within the DMP Assistant, downloading your DMP as a Word document will allow for further modifications to suit your project's needs. For more information on DMP Assistant templates, please contact your local DMP Assistant Administrator.

All incoming Ecohydrology Research Group (ERG) researchers are trained in standardized Quality Assurance/Quality Control (QA/QC) protocols by experienced technicians and staff.

They are expected to implement QA/QC during sampling and sample analyses. For chemical and biological analyses, the quality control will include replicates, blanks, certified reference materials and spikes. Whenever possible, a specific lab task is to be conducted by the same student, technician or researcher for quality assurance. Final data workbooks will be stored in a project repository file server and will contain data sheets corresponding to samples and standards, Certified Reference Materials (CRMs), preparation, raw instrument output, manipulated (i.e., digestion/dilution corrected) instrument output, results, and reported limits of detection and quantification.

Documentation and Metadata

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

In order for the data to be read and interpreted correctly, contextual documentation is provided, including laboratory Standard Operating Procedures (SOPs), instrument operating and software manuals. Data documentation is aligned with national and international best practices.

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

All Ecohydrology Research Group (ERG) researchers will collect metadata and develop appropriate documentation at project inception with regular updates throughout the life of their research projects in accordance with practices developed by the ERG data management team. For example, when collecting laboratory and field data, all chemical nomenclature used will follow the International Union of Pure and Applied Chemistry guidelines. Researchers are asked to keep meticulous records of sampling and experimental details such as sample preparation, sampling procedures, time, etc.

Embedding links in your DMP will direct readers to external resources that can provide more detailed information about discipline-related topics.

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

Metadata associated with all projects is collected to aid future discovery, access and re-use. Data are described according to the ISO 19115 metadata schema, with additional metadata collected per the specific needs of the Highly Qualified Personnel's (HQP) research project.

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?

Each Highly Qualified Personnel (HQP) is expected to generate an average 500 gigabytes (GB) of data during the data management lifecycle, and all data generated for their individual project are to be stored until the thesis/article is published and a Digital Object Identifier (DOI) is assigned to the dataset to support discovery and reproducible research.

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

All Highly Qualified Personnel (HQP) are expected to follow the 3-2-1 backup and storage best practices. This means that Ecohydrology Research Group (ERG) members will create three copies of all data files, to be stored on two different types of media, with one copy kept in an off-site location.

> *When following a specific practice or protocol (like 3-2-1 backup and storage), it is helpful to provide a brief explanation of how it will be carried out. Not all readers will be familiar with any given practice, protocol, or guideline.*

Digital data is stored in spreadsheets, databases, word processing files and lab notebooks. Any data recorded in field or lab books are digitally transcribed, and the hard copies will be archived for a minimum of five (5) years after the life of the project.

ERG HQP decide how to best back up their data, as long as they follow the 3-2-1 rule above. However, all ERG HQP are encouraged to provide copies of their files for back up on the ERG

server on a node of Science Computing at the University of Waterloo. This server is maintained and backed up regularly by Science Computing staff. These backups follow a 52-week retention schedule that is backed up weekly. Additional storage options are outlined on the University of Waterloo IST website.

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

Active research data are stored by each Highly Qualified Personnel (HQP) responsible for its collection until it is ready for publication/dissemination. If it is necessary to share data among researchers and collaborators, the Ecohydrology Research Group (ERG) management team, with approval from the Principal Investigator/Supervisor, mediates data access and usage during the active phase of the project work. HQPs use Microsoft Teams to access and share data with collaborators. Microsoft Teams is a collaboration and communication application within the Microsoft Office Suite. Large data files may be shared via a portable drive.

Preservation

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

Data are be made available within one month of publication in a trusted repository, such as the Federated Research Data repository (FRDR), that provides persistent identifiers such as Digital Object Identifiers (DOIs) or handles. Adhering to UW Policy 73, all Highly Qualified Personnel (HQP) within the Ecohydrology Research Group (ERG) are also expected to provide a digital copy of all research data to their respective supervisors and collaborators at the end of their project/thesis/publication. Data will be published on the University of Waterloo Water Institute Storage Group on the FRDR, and if applicable, the Gordon Foundation DataStream.

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.

To facilitate interoperability, data will be saved in non-proprietary software formats which are accessible to others. Examples of these include Comma-Separated Values (.csv) for

spreadsheets, Tagged Image File Format (.tiff) files for images, Georeferencing Tagged Image File Format (.geotiff) for spatial data, and Text (.txt) for text.

Sharing and Reuse

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

Final data used in publications and theses will be shared to demonstrate reproducibility of Ecohydrology Research Group (ERG) research.

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

Processed research data resulting in publications will be made openly available via a Creative Commons Attribution 1.0 Generic License (CC BY 1.0 License). If a more restrictive license is warranted, researchers will consult with the Ecohydrology Research Group (ERG) management team (or University of Waterloo Library Research Data Services).

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

Data deposited into the Federated Research Data Repository (FRDR) and the DataStream repository have unique Digital Object Identifiers (DOIs), searchable keywords and other searchable metadata. These metadata with data links will be made available via the Ecohydrology Research Group (ERG) website, social media and Waterloo Metadata Index. ERG Highly Qualified Personnel (HQP) will also cite the data in their peer review articles. Email announcements will be sent to appropriate relevant list-servers upon release.

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.

Each Highly Qualified Personnel (HQP) in the Ecohydrology Research Group (ERG) is responsible for the day-to-day management of their individual project data. It expected that their overall data management aligns with best data management practices and ERG's DMP. This includes but is not limited to standard Quality Assurance/Quality Control (QA/QC) procedures, compiling, sharing and backing up of data and metadata regularly onto secure servers. If two or more team members are working jointly on a specific research project, they will determine at the onset of their work which member is responsible for implementing the data management practices for their project. The Principal Investigators are responsible for ensuring team members follow the ERG data management plan.

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?

Data management practices will be reviewed each semester by the Ecohydrology Research Group (ERG) management team with a minimum of two people. All researchers will be reminded of this plan through various communication channels such as email communication, ERG weekly meetings, and a Microsoft Teams communication channel.

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

Given the size of the Ecohydrology Research Group (ERG), hiring a full time data manager would be necessary to support the management of ERG research data; hopefully this type of Research Data Management (RDM) support can be built into future grant applications. However, in the interim, Highly Qualified Personnel (HQP) in the ERG are expected to leverage online resources (such as Mantra and Portage Network Resources) or seek guidance from the data manager of the Global Water Future Program and the University of Waterloo Library Research Data Services to ensure data management best practices are followed.

Because this is a DMP designed for a research group, it is important that details are provided about what is expected of the researchers managing data throughout their project.

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?

If datasets are considered sensitive, then data management will be the responsibility of the lead researcher. Data will be subjected to restrictions according to the best practices and protocols outlined by University of Waterloo Office of Research Ethics. Only appropriate and approved metadata will be made available. Highly Qualified Personnel (HQP) with sensitive data are encouraged to ensure that files are encrypted.

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

Active research data, as well as sensitive data, may be subject to restriction according to the best practices and protocols outlined in UW Policy 73, the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS2) and the First Nations Principles of OCAP®. The availability of data collected from external stakeholders will be subject to agreements made with these stakeholders and the project's lead researcher or HQP supervisor.

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

Any legal, ethical, and intellectual properties issues will be guided by relevant policy(ies) of the University of Waterloo and the Laws of Canada.