



GLOBAL WATER FUTURES – DATA POLICY

Originally Adopted March 1, 2017

Revised March 4, 2019

Contents

1. Introduction	1
2. Definitions	2
3. Objectives.....	3
4. Data Management Planning and Reporting.....	3
5. Active Phase Data Management and Storage	4
6. Data Sharing in Pre-Publication	5
7. Data Archiving, Discovery, and Access.....	5
7.1 Special Circumstances.....	6
8. Restricted Data.....	6
9. GWF-Funded Software Development.....	7
10. Data and Software Use and Acknowledgement	8
Appendices.....	9
Appendix A: Data Management Plan Guiding Questions for GWF Projects.....	9
Appendix B: Data Documentation Guidelines	10
B.1 GWF Observational Data Documentation Guidelines	10
B.2 GWF Model Data Documentation Guidelines	12
Appendix C: Data Management Services List	13
Appendix D: Terms and Conditions of GWF Data Use	18

1. Introduction

"Global Water Futures: Solutions to Water Threats in an Era of Global Change" (hereafter, GWF) is a Canadian research programme that is funded in part by a \$77.8-million grant from the Canada First Research Excellence Fund (CFREF). The aim is to transform the way communities, governments and industries in Canada and other cold regions of the world prepare for and manage these increasing water-related threats. Further details can be found on the GWF website at <http://gwf.usask.ca/>.

The GWF Data Policy (hereafter, “the Policy”) has been established to promote and govern the management, retention, use, and dissemination of data collected as part of the GWF programme’s research activities, with the intention that the Policy is sufficiently flexible to accommodate the multi-disciplinary nature of GWF-funded research and the wide variety of data collected therein. GWF is committed to the establishment, long-term maintenance, validation, description, accessibility, and distribution of high-quality datasets, and fully adheres to the FAIR data principles (Findable, Accessible, Interoperable and Re-useable) that specific forms of research data collected with public funds should be publicly accessible, with a special obligation to openness and accountability.

GWF will develop and maintain a central data repository and archive to contain certain datasets, and will develop a central data indexing system and metadata catalogue to search and find datasets within the central archive and those stored elsewhere.

Note: Continuation of funding to individual GWF projects is contingent on provision of data in accordance with the GWF data policy and the development and updating of data management plans throughout each project. Any exemptions must be approved by the GWF Strategic Management Committee.

2. Definitions

In this Policy, unless otherwise indicated:

GWF Data (hereafter, “data”): GWF data refers to any and all observed quantitative data and some analytical or modelling results and associated metadata that have been collected and/or generated by GWF researchers through the GWF programme, including CFREF and other funding partners.

GWF Software (hereafter, “software”): GWF software refers to any and all modelling and/or analysis software (computer programs, applications or code) and associated metadata and documentation developed by GWF researchers through the GWF programme, including CFREF and other funding partners.

GWF Researchers (hereafter, “researchers”): GWF Researchers are researchers, including but not limited to principal investigators, co-investigators, research assistants, technicians, and Highly Qualified Personnel (HQP), working on projects funded in whole or in part by GWF.

GWF Data Management Core Team (hereafter, “DM Team”): GWF Data Management Core Team refers to personnel employed in whole or in part by GWF to provide data management support services and training to GWF managers and researchers.

GWF Data Owner/Originator: A researcher who is responsible for collecting and/or generating and/or processing of primary and secondary data in support of a GWF project.

GWF Software Owner/Developer: A researcher who is responsible for developing code in support of a GWF project.

3. Objectives

The objectives of this policy are to:

- Outline the roles and expectations of GWF researchers and the DM Team in the collection, management, dissemination, and preservation of (anonymised) quantitative research data.
- Respect the rights of the data originators and/or software developers who have invested considerable effort in obtaining and/or generating data and developing software by ensuring consumers of these products use them ethically and with proper attribution to the originators and/or developers.
- Benefit society and research communities by promoting the re-use of data collected and generated through GWF-funded projects.
- Improve the transparency, efficiency and reproducibility of research by ensuring data is managed appropriately and easily discoverable through a central metadata catalogue.
- Ensure that certain key hydrometeorological, hydrometric, groundwater, soil moisture, snowpack, and water quality data that is useful for model development, parameterisation and validation is available in a central data repository.
- Foster collaboration among GWF-funded researchers, and the general research community, through clear mechanisms and mandates for data sharing.
- Protect privacy and respect data ownership rights, especially data that falls under OCAP (Ownership, Control, Access, and Possession) principles or their equivalent.
- Encourage responsible data sharing by providing guidance to GWF researchers working with sensitive data.

4. Data Management Planning and Reporting

As each GWF project is proposed and developed, careful consideration must be given to the project data needs, including data collection, production and use. Researchers, in consultation with the DM Team (see Appendix C), must create a detailed data management plan that describes how data will be collected, documented, stored, shared, and ultimately preserved for re-use. Appendix A provides a set of key guiding questions that should be considered and addressed in the project data management plan. Researchers will submit the project data management plan to the DM Team for archiving and to serve as the base for annual updates on project data management activities as part of the regular scientific progress reports to the GWF Strategic Management Committee.

5. Active Phase Data Management and Storage

Data must be managed in accordance with established, discipline-related standards and best practices for data stewardship, including proper documentation and metadata application (see Appendix B)¹. Documentation should include, but is not limited to, records related to the collection, storage, and retrieval of data, as well as steps taken to process, analyze, and visualize data. At minimum, data must include clear supporting documentation and metadata (and, where possible and applicable, computer code) sufficient for re-use and replication of results by other researchers.

Throughout the active research phase of a project, data must be stored in a suitable location with appropriate backup, security, and privacy commensurate with the value and sensitivity of the data. The DM Team will also advise on suitable storage locations to ensure the data can be indexed for inclusion in the GWF central metadata catalogue. In many cases, researchers will continue to use their current storage locations. However, certain GWF data must be submitted and stored within the GWF central archive as soon as possible. These include hydrometeorological, hydrometric, groundwater², soil moisture, snowpack, water quality, and other relevant data that are required by GWF projects and core teams for the development, testing and evaluation, and application of Earth system models and for other potential network uses. Which data are required centrally and which are not can be more fully clarified through communication with the DM team and GWF secretariat, overseen by the GWF Strategic Management Committee. GWF projects will need to liaise with the DM Team about this.

The DM Team will provide technical assistance and support for various aspects of data management as outlined in the Data Management Services List (Appendix C). Unpublished data must be discoverable by the DM Team, GWF secretariat, and internal research groups. GWF will develop and maintain a central metadata catalogue that will contain information about the data holdings, including quality assessments, supporting ancillary information, all associated documentation and metadata, as well as researcher defined protocols for locating and accessing the data. Researchers will be required to register their data with the central metadata catalogue for discovery. With respect to restricted data as outlined in Section 8 of this policy, documentation and metadata is still required to be accessible for data discovery. The DM Team will provide support for data registration.

¹ Documentation guidelines for observational and model output datasets are provided in Appendix B. It is the responsibility of the researcher to ensure proper quality assurance and control and data documentation, while the GWF DM Team will be available to provide needed assistance.

² Some of these data, such as groundwater and water quality, may require consent of collaborating parties.

6. Data Sharing in Pre-Publication

With the goal of fostering collaboration, synergy and expedited sharing of data within the GWF programme, the GWF Strategic Management Committee may require that specific datasets be shared among certain internal research groups during the GWF programme. For example, this may be necessary during coordinated, multi-partner observation and/or modelling campaigns as part of the GWF science programme, or for other currently unforeseen special collaborative activities. In these circumstances, the data requirements and timelines will be discussed with the individual researcher/research teams and agreed upon prior to funding allocations, or alternatively at the earliest possible opportunity.

Datasets may also be obtained through direct request to the data originator(s). It is up to their discretion to release partial or complete datasets that they have collected, but the explicit permission of others who have had a significant role in the data collection and/or processing and who retain ownership rights of the data is required. In certain instances, where there may be foreseeable conflict or other issues with its early release, permission of the GWF Strategic Management Committee should be sought. Some potential issues are discussed in Section 7.1. An offer of co-authorship for data used that has not been previously published is required unless it is waived in writing by the data originator. Further, it is required that all data requests and sharing be communicated to the GWF DM Team for tracking purposes. This can be done by completing a form provided by the DM Team or simply forwarding the details of the request and fulfillment to the DM Team.

7. Data Archiving, Discovery, and Access

GWF adheres to the principle that research data collected with public funds should be publicly accessible, as appropriate, with a special obligation to openness and accountability. Researchers are therefore required to deposit finalized non-personal quantitative data into a recognized digital repository for publication, discovery, and access within a reasonable period of time, that is, **within two years of data collection or one year of the completion of the research project for which the data were collected or at the time of journal article publication**. This includes time for quality assurance and control, documentation, and for progress towards and/or completion of the research project(s) for which the data were collected. Digital Object Identifier(s) (DOI) for published datasets are required and will be indexed for inclusion in the GWF central metadata catalogue for future discovery and access. To ensure proper recognition and attribution for data originators, data will be published under the Terms and Conditions of GWF Data Use outlined in Appendix D. The DM Team will advise on suitable locations for data publication, assist researchers in obtaining DOIs, and work with publication repositories to obtain data usage metrics (e.g., views, downloads, and citation counts).

Special circumstances, such as those outlined in Section 7.1, may require data to be withheld for longer than the aforementioned embargo period or exempted from publication and access. In these instances, researchers will provide an exemption request documenting the justification to the DM Team. After review, the DM Team will submit the request to the GWF Strategic Management Committee for approval, and to determine the appropriate course of action. All data, including restricted access data, must have an associated metadata entry in the GWF metadata catalogue (and potentially other appropriate catalogues) to facilitate data discovery.

7.1 Special Circumstances

A researcher may determine if a request for data, and the intended use identified, significantly overlaps with their own work, or would jeopardize the work of a student, including, for example, a graduate student thesis or dissertation in progress. In such cases, data may be justifiably withheld until such a time that its use does not jeopardize their own work or that of a student. The decision to withhold data that are past the embargo period and are not otherwise restricted by the circumstances detailed in Section 8 of this policy, will ultimately rest with the GWF Strategic Management Committee.

8. Restricted Data

The GWF data policy is two-tiered. Where most datasets will be made public, some data such as copyright-protected, third-party data (even if collected with GWF funds), environmentally sensitive and select human-subject data cannot be public for various reasons. Certain types of data (e.g., data containing personally identifiable information [PII], Traditional Knowledge, or data related to commercially valuable or endangered species) may be considered sensitive, with the release of such data resulting in potential harms. Copyright, or other considerations related to data ownership, may also be grounds for restriction of the data. Sensitive data will be considered on a case-by-case basis, but may be deemed safe for release if allowed by the terms of the collection agreements and if certain conditions are met such as de-identification (e.g., the locations and nests of endangered species or locations of sacred sites are omitted, or personal identifying information)³. Where possible, the DM Team should be consulted during the planning phase to assist researchers in the compliance with data restrictions imposed by Ethics Board requirements or consent granted by participants. It is expected that researchers and/or the DM Team will preserve and safeguard the original data, and be aware of/adhere to the ethical requirements pertaining to the retention or disposal of data obtained in a research context, as

³ Shearer, K. (2015). *Comprehensive brief on research data management policies*. Available: <http://docplayer.net/17594465-Comprehensive-brief-on-research-data-management-policies.html>

specified in the 2nd edition of the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans.

Per the Statement of Principles and Practices for Arctic Data Management, research projects working with Traditional Knowledge must not compromise the rights of the knowledge holders. Research involving Traditional or Indigenous knowledge must adhere to the concepts of respect, reciprocity, and responsibility. This includes the “appropriate engagement of Indigenous People, communities or organizations throughout the entire data life cycle, formal attribution of contributed knowledge, establishment of informed consent for use of knowledge and derived products, and the maintenance of contributor control of data and information resources.”⁴ Restrictions to data access may be made on the basis of ethical concerns arising from violating the rights of Traditional Knowledge holders.

Other guiding principles for ethics and data management for Indigenous-led projects or projects that work with Indigenous communities include Chapter 9 (Research Involving First Nations, Inuit and Métis Peoples of Canada) of the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans and the First Nations Principles of OCAP® which “are a set of standards that establish how [Indigenous] data should be collected, protected, used or shared”⁵ described in the document Ownership, Control, Access and Possession (OCAP™): The Path to First Nations Information Governance. Community research agreements between researchers and Indigenous communities including decisions regarding data sharing should be respected. **Traditional knowledge and sensitive Indigenous community data will remain under the ownership and control of the community or organization, and will only be released publically or to other GWF researchers or other parties with their expressed consent.**

9. GWF-Funded Software Development

GWF is committed to open source software development. Where GWF funds are used to develop modelling and/or analysis software, that software will be made available to the research community and the general public within a reasonable period of time. We consider this reasonable period of time to be the shorter of either **within two years of the development of the software, or within one year of the completion of the research project for which the software was developed or at the time of journal article publication**. In some circumstances, as above, it may be necessary to release the software earlier, and discussion and agreement will be reached prior to funding allocation or at the earliest possible opportunity. Software may also be

⁴ International Arctic Science Committee (IASC). (2013). The state of principles and practices for Arctic data management, p. 3. Available: http://iasc.info/images/data/IASC_data_statement.pdf

⁵ First Nations Information Governance Centre (FNIGC). (2019) <https://fnigc.ca/ocapr.html>

restricted under the special circumstances outlined in Section 7.1. In these instances, it is the responsibility of the researcher to explain and justify to the GWF Strategic Management Committee any delay in the release of GWF-funded software, who will then decide on an appropriate course of action. It is recommended that software be licensed under an appropriate open source licence, such as those found on <https://opensource.org/licenses>. In some cases, commercialization of software may be appropriate, in which case appropriate licensing arrangements should be discussed with the SMC, and GWF researchers should have free access for research purposes.

10. Data and Software Use and Acknowledgement

This policy is based on the principle of giving data originators and/or software developers fair credit and due recognition for their efforts. All users of GWF datasets must agree to the terms and conditions set out in this policy (Appendix D). Use of the data and/or software resulting in publications or reports, or for various forms of public presentation (including, for example, conferences, media engagement, etc.) will require, at the least, formal acknowledgement of the GWF programme and the data originator(s) and/or software developer(s). Any publications or reports should also contain citation of relevant earlier studies by the data originator(s) and/or software developer(s) describing or using the dataset and/or software in question (appropriate citations are provided with many GWF datasets as per the documentation guidelines in Appendix B). Data users are strongly encouraged to liaise with the data originator(s), who can advise on proper use of the data and other relevant issues (contact information for data originators will accompany the datasets as per the documentation guidelines in Appendix B). In some cases, it may be warranted to include data originator(s) and/or software developer(s) as co-authors on publications or reports, depending on the extent to which the data and/or software are used and the role of the data and/or software in supporting the research outcome(s). Where GWF datasets and/or software form an integral component of the user's research project, the users are required to liaise with the data originator(s) and/or software developers and discuss appropriate co-authorship or acknowledgement options. For clarification on any issue pertaining to use and acknowledgement of GWF datasets, please consult the GWF DM Team (contact information is on our website at <http://gwf.usask.ca/>).

As an example, acknowledgement of the use of GWF data should be made using a statement with the following format: *"The [***] data in this paper were provided by the Global Water Futures (GWF) Programme and by [###]."* where [***] refers to the actual GWF data used and [###] refers to the data originator(s). For the use of GWF software the following is an example: *"The [***] software in this paper was provided by the Global Water Futures (GWF) Programme*

and by [###].” where [***] refers to the actual GWF software used and [###] refers to the software developer(s).

Appendices

Appendix A: Data Management Plan Guiding Questions for GWF Projects

The following is a list of key questions that should be carefully considered and addressed in developing data management plans specific to the data production, needs, and use of individual GWF projects. (These questions originate from the Portage Network, <https://portagenetwork.ca/>.) Not all questions will necessarily apply to all GWF Projects. Researchers are strongly encouraged to liaise with the GWF DM Team in developing their plans.

Data Collection

1. What types of data will you collect, create, link to, acquire and/or record?
2. 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?
3. 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?

Documentation and Metadata

1. What documentation will be needed for the data to be read and interpreted correctly in the future?
2. How will you make sure that documentation is created or captured consistently throughout your project?
3. If you are using a metadata standard and/or tools to document and describe your data, please list here.

Storage and Backup

1. 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?
2. How and where will your data be stored and backed up during your research project?
3. How will the research team and other collaborators access, modify, and contribute data throughout the project?

Preservation

1. Where will you deposit your data for long-term preservation and access at the end of your research project?
2. Indicate how you will ensure your data is preservation ready. Consider preservation-friendly file formats, ensuring file integrity, anonymization and deidentification, inclusion of supporting documentation.

Sharing and Reuse

1. What data will you be sharing and in what form? (e.g. raw, processed, analyzed, final).
2. Have you considered what type of end-user license to include with your data?
3. What steps will be taken to help the research community know that your data exists?

Responsibilities and Resources

1. 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.
2. 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?
3. What resources will you require to implement your data management plan? What do you estimate the overall cost for data management to be?

Ethics and Legal Compliance

1. 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?
2. If applicable, what strategies will you undertake to address secondary uses of sensitive data?
3. How will you manage legal, ethical, and intellectual property issues?

Appendix B: Data Documentation Guidelines

To assist in meeting the objectives of GWF, and to provide a useful and lasting legacy for further research, it is necessary that all data collected for GWF, both observational and significant model results, be properly documented. It is the responsibility of the data originator to ensure that the data collected are properly documented using the established metadata standards in their field. The GWF Data Policy provides for a time period after data collection for the researcher to quality-control and to document their data. With respect to model output, the end of a data collection period is defined as the model run after a significant model revision (e.g. change of physics, improved routing, or coupling). Data documentation should be complete enough to allow unfamiliar researchers to replicate and use the data in the future.

Examples of best practice metadata format for field hydrometeorological data and model data follow (B.1, B.2).

B.1 GWF Observational Data Documentation Guidelines

The observational data documentation should contain the following headings:

1. Title;
2. Abstract – Name the dataset and describe why the measurement was undertaken and how it relates to GWF;

3. Contact Information – Give sufficient detail (name, affiliation, full address, telephone and fax numbers, e-mail, etc.) to contact those most knowledgeable about the dataset:
 - 3.1. Principal investigator (or person responsible for collecting the data)
 - 3.2. Associate or co-investigators
 - 3.3. Contact person for question
4. Site description – including the following:
 - 4.1. Data period(s) and location(s);
 - 4.2. Equipment used, including manufacturer and model numbers;
 - 4.3. Instrument heights and other relevant physical site characteristics;
 - 4.4. Methods/software used in acquiring the data;
 - 4.5. Data format, including examples;
5. Keywords used to describe the data topic
6. Funding information
7. Sharing and access information
 - 7.1. Licences placed on the data
 - 7.2. Links to related materials of the data, as well as the relationship (use, cite, ...)
 - 7.3. Recommended citation for the data
8. Data processing and quality control – including the following:
 - 8.1. Methods and software used in acquiring and processing the data;
 - 8.2. Post-collection data processing – description of any processing done on the data;
 - 8.3. Quality control methods – give an indication as to the degree of quality control;
 - 8.4. Datasets archived – original “raw” data should be one of the archived datasets in addition to any processed or quality controlled data (this will depend on the nature of the data, however – see section 8 of the data policy);
9. Data and file overview
 - 9.1. For each filename, a short description of what data it contains
 - 9.2. Format of the file
 - 9.3. If the data set includes multiple files that relate to one another, the relationship between the files or a description of the file structure that holds them
 - 9.4. Date that the file was created
 - 9.5. Date(s) that the file(s) was updated (versioned)
 - 9.6. Information about related data collected or related materials (software, paper, etc.) but that is not in the described dataset
 - 9.7. Data-specific information:
 - 9.7.1. number of variables, number of rows
 - 9.7.2. variable list: name, definition, abbreviation
 - 9.7.3. definitions for codes or symbols used to record missing/invalid data, and/or special records
 - 9.7.4. units of measurement

10. References

B.2 GWF Model Data Documentation Guidelines

The model data documentation should contain the following headings:

1. Title – Model name, version number;
2. Abstract – Briefly describe the model and its properties and describe why the model run was undertaken and how it relates to GWF;
3. Contact information - Name/institution/address/email/telephone/fax information for:
 - 3.1. Principal investigator (or person responsible for collecting the data)
 - 3.2. Associate or co-investigators
 - 3.3. Contact person for question
4. Keywords used to describe the data topic
5. Funding information
6. Sharing and access information
 - 6.1. Licenses placed on the data
 - 6.2. Links to related materials of the data, as well as the relationship (use, cite, ...)
 - 6.3. Recommended citation for the data
7. Run description – including the following (valid web links acceptable):
 - 7.1. Period(s) and location(s)/resolution/map projection;
 - 7.2. Initialization and boundary data used;
 - 7.3. Model used – complete description of the model, physics package, any coupling state, etc.;
 - 7.4. Data format – including examples;
 - 7.5. Archive location/media – online link or offline contact person;
8. Data and file overview
 - 8.1. For each filename, a short description of what data it contains
 - 8.2. Format of the file
 - 8.3. If the data set includes multiple files that relate to one another, the relationship between the files or a description of the file structure that holds them
 - 8.4. Date that the file was created
 - 8.5. Date(s) that the file(s) was updated (versioned)
 - 8.6. Information about related data collected or related materials (software, paper, etc.) but that is not in the described dataset
 - 8.7. Data-specific information:
 - 8.7.1. number of variables, number of rows
 - 8.7.2. variable list: name, definition, abbreviation
 - 8.7.3. definitions for codes or symbols used to record missing/invalid data, and/or special records

8.7.4. units of measurement

9. References.

Appendix C: Data Management Services List

GWF Data Management Services – How the DM Team can assist the Researchers?

This document outlines core data management services of the DM Team. It describes various data management services, the benefits they bring to the research program, and the expectations of the DM Team and researchers.

Service	Capabilities
1. Data Management Planning <ul style="list-style-type: none">Has the researcher developed a data management plan (DMP) to outline metadata, collection, storage, preservation, and security protocols?Have roles and responsibilities been defined for various data management activities?	<p>Data Management Planning is a core DM activity required by many funders and stakeholders. The DM Team offers full DMP consultation services.</p> <p><i>Benefit:</i></p> <p><i>Data Management Planning provides researchers and programs the opportunity to consider all their DM needs from collection and analysis, to storage, and backup, security and ethics, costs, and preservation. <u>DMPs are a requirement of Tri-Agency funding (2019 and forward).</u></i></p> <p>The DM Team will:</p> <ul style="list-style-type: none">Conduct DM needs assessments with researchersSupport or lead researchers through a DMP process (e.g. using Portage DMP Assistant)Link researchers to external DM supports and opportunities <p><i>The Researchers will:</i></p> <ul style="list-style-type: none"><i>Identify DM gaps and opportunities in their program</i><i>Create DMPs that meet their stakeholder and subject matter requirements</i><i>Consult with the DM Team on a timely basis</i>

<p>2. Data Sharing</p> <ul style="list-style-type: none"> • Does the research project have data sharing requirements beyond Global Water Futures? If so, what are these obligations and can workflows be streamlined? • Do the researchers have known or expected data deposit obligations for publication? 	<p>The DM Team helps researchers consolidate their data sharing obligations and provide platforms for sharing when needed.</p> <p><i>Benefit:</i></p> <ul style="list-style-type: none"> • <i>Outlining data sharing requirements prior to analysis and publication enables data synthesis and researcher collaboration, and helps DM Team and researchers produce outputs to meet stakeholder obligations</i> <p>The DM Team will:</p> <ul style="list-style-type: none"> • Consult on data sharing requirements with researchers • Advise on policy documentation and implementation • Provide data sharing services and platforms <p><i>The Researchers will:</i></p> <ul style="list-style-type: none"> • <i>Provide known data sharing requirements to the DM Team</i> • <i>When required, provide research data and associated metadata to the DM Team for ingest on data sharing platforms</i>
<p>3. Data Access and Security</p> <ul style="list-style-type: none"> • Who requires access to your research data? Are there different security requirements for different team members or groups? • Do the researchers collect any sensitive data regulated by external agreements or protocols? 	<p>The DM Team advises on data access and security requirements, and makes recommendations as needed.</p> <p><i>Benefit:</i></p> <ul style="list-style-type: none"> • <i>Developing access control provisions reduces the chance of inadvertent data leaks and increases the researchers' abilities to comply with its sensitive data stakeholder requirements</i> <p>The DM Team will:</p>

	<ul style="list-style-type: none"> ● <i>Advise researchers on security options on various platforms to meet their access needs and stakeholder obligations</i> ● <i>Help develop data access/security plans as required</i> <p><i>The Researchers will:</i></p> <ul style="list-style-type: none"> ● <i>Identify access/security obligations to the data management core</i> ● <i>Identify research data that might be sensitive or require security provisions</i> ● <i>As needed, identify user groups (if any) that might require particular access/security provisions</i>
<p>4. Data Storage and Backup</p> <ul style="list-style-type: none"> ● What are the anticipated storage requirements during the life of the project? Has storage been secured? ● What are the backup protocols (e.g., locations, copies, retention periods, versioning, etc.) for the project? 	<p>The DM Team will recommend storage and backup services for research data, including long-term storage and version control.</p> <p><i>Benefit:</i></p> <ul style="list-style-type: none"> ● <i>A storage and backup policy standardize practices for all researchers. It reduces the odds of data loss.</i> <p><i>The DM Team will:</i></p> <ul style="list-style-type: none"> ● <i>Suggest backup services for all GWF-associated research data</i> <p><i>If interested the Researchers will:</i></p> <ul style="list-style-type: none"> ● <i>Provide data, and a file manifest to the DM Team if storage and backup services are required</i> ● <i>Will house only project-related materials if storage is requested</i>
<p>5. File Formats and Organization</p> <ul style="list-style-type: none"> ● Are your files held in an open or commonly used format for long term access? 	<p>The DM Team advises researchers on best practices for file and data organization. It advocates for the use of open formats and standards whenever possible (i.e., when conversion would not risk data content or architecture).</p>

<ul style="list-style-type: none"> • Are similar data organized consistently across different files? • Are file naming conventions used consistently across files, folders, and records? 	<p><i>Benefit:</i></p> <ul style="list-style-type: none"> • <i>Standardized file naming conventions and data structure improve search, retrieval, and access to files, especially in team environments.</i> <p>The DM Team will:</p> <ul style="list-style-type: none"> • Consult with and advise on best practices for naming conventions and file organization • Advise on and recommend the use of preferred open file formats (when available) for long-term storage without compromising data integrity <p><i>The Researchers will:</i></p> <ul style="list-style-type: none"> • <i>Endeavour to develop or use established naming conventions</i> • <i>Endeavour to store data and content in open file formats when available for long-term storage without compromising data integrity</i>
<p>6. Documentation</p> <ul style="list-style-type: none"> • Has documentation summarizing field data collection, processing, QA/QC, analysis methods and techniques been produced? • Has necessary context for re-use or replication (e.g., instrumentation, calibration, etc.) been documented? • Are codebooks (e.g., readme files, data dictionaries, etc.) or technical manuals required and included with the data? 	<p>The DM Team helps collect and proofread documentation required for reuse or replication of data.</p> <p><i>Benefit:</i></p> <ul style="list-style-type: none"> • <i>Thorough documentation brings context to research data. It helps users understand the original purpose for the data collection, and how to correctly understand and use the data in the future.</i> <p>The DM Team will:</p> <ul style="list-style-type: none"> • Assist in the collection of existing documentation • Provide checklists to help researchers learn the kinds of documentation that should be stored with research data • Provide documentation QA/QC to ensure consistent formatting with data repositories <p><i>The Researchers will:</i></p>

	<ul style="list-style-type: none"> • <i>Lead in the collection or creation of documentation</i> • <i>Verify accuracy of text</i>
<p>7. Data Description (Metadata)</p> <ul style="list-style-type: none"> • <i>Object/Data-level data</i> <ul style="list-style-type: none"> • Are data in an accessible format (e.g. a spreadsheet or database) clearly marked with variable and value labels, code descriptions, etc.? Are these labels consistent? • Will data and metadata match documentation (described above)? • <i>Project-level Metadata</i> <ul style="list-style-type: none"> • Has structured project-level metadata been created for the dataset? • Do the researchers have any metadata standards and/or requirements to follow? 	<p>The DM Team consults and supports the development of rich metadata at both the project-level and object/data-level.</p> <p><i>Benefit:</i></p> <ul style="list-style-type: none"> • <i>Detailed data description improves data organization and architecture, increases synthesis opportunities, and mitigates backup and archiving issues.</i> <p>The DM Team will:</p> <ul style="list-style-type: none"> • <i>Provide metadata template to ensure standardized metadata collection within the GWF Program</i> • <i>Consult with researchers to confirm metadata standards in their fields and repositories</i> • <i>Provide metadata QA/QC to enable organization, analysis, and archiving</i> <p><i>The Researchers will:</i></p> <ul style="list-style-type: none"> • <i>Share metadata to the DM Team in a timely manner to enable metadata processing</i> • <i>Collaborate on metadata practices in their fields</i> • <i>Provide known metadata requirements</i>
<p>8. Archiving and Preservation</p> <ul style="list-style-type: none"> • What research data do the researchers plan to archive? • Where will data be deposited? What work is required to prepare data for preservation (e.g., open format conversions, variable coding, documentation, etc.)? 	<p>The DM Team provides data archiving services including preparation, curation, documentation, and preservation.</p> <p><i>Benefit:</i></p> <ul style="list-style-type: none"> • <i>Data preservation planning benefits researchers and their data by asking and answering what is required to access, use, and understand data in the future</i>

	<p><i>The DM Team will:</i></p> <ul style="list-style-type: none"> ● <i>Help researchers identify data and supporting materials for long-term preservation</i> ● <i>Identify appropriate repositories for preservation</i> ● <i>Aid in the preparation and ingest of data and metadata for preservation</i> <p><i>The Researchers will:</i></p> <ul style="list-style-type: none"> ● <i>Disclose their data retention, sharing, and archiving requirements early in the research phase</i> ● <i>Assist in the development of metadata, including data collector/author citation requirements</i> ● <i>Ensure research data is clear of copyright and intellectual property issues</i> ● <i>Consult with the DM Team in advance of data archiving deadlines</i>
--	--

Appendix D: Terms and Conditions of GWF Data Use

Permission is hereby granted by "Global Water Futures: Solutions to Water Threats in an Era of Global Change" (or GWF for short) to download and use the data, subject to the terms and conditions set forth in this license agreement.

1. GWF is pleased to grant you a perpetual, royalty-free, non-exclusive and non-assignable license to use the data, subject to the terms and restrictions set forth herein. If you are acting on behalf of an organization, you represent to GWF that you are duly authorized to act on behalf of this organization and that the organization hereby gives its assent to be bound to GWF by the terms and conditions of this license agreement. As used herein, "you" and "your" refers to you and to any organization on behalf of which you are acting;
2. You will have access to the data as offered at the time of access of the said data. GWF has no obligation whatsoever to keep on providing services and/or upgrades of the data;
3. You have the right to create derivative work with the data, but have the obligation, whether the data is used AS IS or in a derivative work, to visually and clearly acknowledge GWF and the data originator as the source of the said data. It must be easily identifiable for any individual given access to your services and/or products that the source of the data is the Global Water Futures Programme;

4. Redistribution to a third party is permitted only if that third party has been informed by you of the terms and conditions of use of the data and has agreed to said terms and conditions of use;
5. No additional rights or licenses other than as provided for herein, whether expressed or implied, are granted herein by GWF under or by virtue of this license agreement;
6. GWF does not warrant the quality, accuracy, or completeness of any of the data downloaded and/or retrieved from any of its servers. Such data are provided on an "AS IS" basis without warranty or condition of any nature. GWF disclaims all other warranties, expressed or implied, including but not limited to implied warranties of merchantability and fitness for a particular purpose, with respect to the information retrieved under this license agreement, and any accompanying materials, if any;
7. Furthermore, in no event shall GWF or any of its members be liable for any damages whatsoever in respect of any claim, action, demand or proceeding of any kind arising out or related to your use, inability to use or performance of the data retrieved;
8. This license and the rights granted hereunder will terminate automatically if you fail to comply with the terms herein and fail to cure such breach within 30 days of receiving written advice from GWF.