



GLOBAL WATER FUTURES -DATA POLICY

March 1, 2017

Contents

1. Introduction	1
2. Definitions	2
3. Objectives.....	2
4. Data Management, Submission, and Retention	3
5. Data Requests, Access, and Sharing	4
5.1 Withholding Periods and Embargo	4
5.2 Special Circumstances	5
6. GWF-funded software development	6
7. Data and Software Use and Acknowledgement	6
Appendices.....	7
Appendix A: Data Management Plan Guiding Questions for GWF Sub-Projects	7
Appendix B: Data Documentation Guidelines	8
Appendix C: Terms and Conditions of GWF Data Use	10

1. Introduction

"Global Water Futures: Solutions to Water Threats in an Era of Global Change" (or GWF for short) is a Canadian research programme that is funded in part by a \$77.8-million grant from the Canada First Excellence Research 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 principle that various forms of research data collected with public funds should be publicly accessible, with a special obligation to openness and accountability.

2. Definitions

GWF Data (hereafter, “data”): GWF data refers to any and all observed 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 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 postdocs, working on projects funded in whole or in part by GWF.

GWF Data Management Core Team (hereafter, “data management core”): 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.

3. Objectives

The objectives of this policy are to:

- Clearly outline the roles and expectations of GWF researchers in the collection, management, dissemination, and preservation of research data.
- Respect the rights of the data originators and/or software developers who have invested considerable effort in obtaining and/or generating data by ensuring that they are given proper recognition for these efforts by anyone using the datasets.
- Benefit society and research communities by promoting the re-use of data collected and generated through GWF-funded projects.
- Improve the efficiency and reproducibility of research by ensuring data is easily discoverable and accessible through a centralized repository.
- Foster collaboration among GWF-funded researchers, and the general research community, through clear mechanisms and mandates for data sharing.
- Encourage responsible data sharing by providing guidance to GWF Investigators working with sensitive data.

4. Data Management, Submission, and Retention

Data must be preserved and submitted internally to the data management core for storage in our central data repository as soon as possible. The data management core will provide technical assistance and support to various aspects of data processing, quality assurance and control, and preservation. Relevant metadata (see Appendix B) should also be contributed at the same time. Datasets will be safeguarded internally and will only be released publicly with the consent of the data originator or after the embargo period, as specified below in Section 5.1.

As part of this process, and under the terms and conditions of GWF funding, researchers are expected to:

- Adhere to national and/or international standards or best practices in data collection and management procedures (e.g., data processing, quality assurance and control, and documentation¹) where relevant and/or possible;
- Liaise with the data management core on a regular basis to ensure prompt submission of high-quality and properly documented datasets to the central repository;
 - The data management core may impose specific protocols to be followed in some circumstances, where feasible and/or necessary, to ensure compatibility and standardization of various datasets;
- Report on data management and submission activities as part of regular scientific and progress reports to the GWF Strategic Management Committee or other relevant overseeing body.

As GWF sub-projects are proposed and developed, careful consideration must be given to the data production, needs, and use of these projects, and a detailed data management plan is required for each. A set of key guiding questions to help develop these plans is provided in Appendix A, and should be referred to in this process and throughout the span of the individual projects. Researchers are strongly encouraged to liaise with the data management core in developing their plans and additional training may be provided as needed.

(Note: Continuation of funding to individual GWF sub-projects is contingent on provision of data in accordance with the GWF data policy and the data management plans developed from the outset of each project. Any exemptions must be fully justified and approved by the GWF Strategic Management Committee.)

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 data management team will be available to provide needed assistance.

5. Data Requests, Access, and Sharing

Upon submission and any follow-up data processing, quality assurance and control, and documentation, GWF-funded and other associated datasets will be stored in a long-term central repository and made discoverable and accessible through a data portal on our webpage, unless otherwise restricted by the special circumstances identified in Section 5.2. The central data repository will include easily accessible information about the data holdings, including quality assessments, supporting ancillary information, all associated documentation and metadata, and guidance for locating and obtaining the data. Data users will be required to register and agree to the terms and conditions set out in our data policy (see Appendix C), and downloads will be tracked to gather relevant information on data usage. Certain non-public datasets or embargoed data residing in our central archive may be obtained through request to the data management core, with the explicit permission of the data originator(s).

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 problematic issues with its early release, permission of the GWF Strategic Management Committee should be sought. We require that all data requests and provisions also be communicated to the GWF data management core for tracking purposes.

5.1 Withholding Periods and Embargo

Researchers have an obligation to make GWF-funded datasets freely and openly available to the research community and the general public within a reasonable period of time. 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. We consider this reasonable period of time to be the shorter of either **within two years of the collection of the data, or within one year of the completion of the research project for which the data were collected**. Special circumstances, such as those outlined in Section 5.2, may require data to be withheld for longer than the aforementioned embargo period. 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 datasets, who will then decide on an appropriate course of action. Datasets held in the central data repository will be automatically made publicly available after the specified embargo period without researcher intervention.

With the goal of fostering collaboration and expedited open sharing of data within the GWF programme, the GWF Strategic Management Committee may request or impose that some specific datasets be shared among certain internal research groups or more broadly ahead of the

time periods given above. 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 researchers or research teams and agreed upon prior to funding allocations, or alternatively at the earliest possible opportunity.

5.2 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 is past the embargo period and is not otherwise restricted by the circumstances detailed below will ultimately rest with the GWF Strategic Management Committee.

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. The GWF data policy is therefore two-tiered, where most datasets are public, but copyright-protected, third-party data, and select human-subject data are not public. Human-subject data may be made public after the removal of identifying information. Sensitive data will need to be considered on a case-by-case basis, but may be deemed safe for release if certain conditions are met (e.g., the locations and nests of endangered species or locations of sacred sites are omitted)². It is expected that researchers and/or the GWF data management core will preserve and safeguard the original data, and that researchers will be aware of the ethical requirements pertaining to the retention or disposal of data obtained in a research context, as 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.”³

² Shearer, K. (2015). *Comprehensive brief on research data management policies*. Available: <http://docplayer.net/17594465-Comprehensive-brief-on-research-data-management-policies.html>

³ International Arctic Science Committee (IASC). (2013). *The state of principles and practices*

Restrictions to data access may be made on the basis of ethical concerns arising from violating the rights of Traditional Knowledge holders.

6. GWF-funded 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**. 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 restricted under the special circumstances outlined in Section 5.2. 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 licenced under the Creative Commons Attribution-NoDerivs 3.0 Unported Licence (CC BY-ND 3.0). An example licence is found on the creative Commons Web site: <http://creativecommons.org/licenses/by-nd/3.0/>. In some cases, commercialization of software may be appropriate, in which case appropriate licencing arrangements should be discussed with the SMC and GWF researchers should have free access for research purposes.

7. 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 C). 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

for Arctic data management, p. 3. Available: http://iasc.info/images/data/IASC_data_statement.pdf

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 data management core 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 was 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).

Appendices

Appendix A: Data Management Plan Guiding Questions for GWF Sub-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 research projects. Researchers are strongly encouraged to liaise with the GWF data management core 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

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

Storage and Backup

7. 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?
8. How and where will your data be stored and backed up during your research project?
9. How will the research team and other collaborators access, modify, and contribute data throughout the project?

Preservation

10. Where will you deposit your data for long-term preservation and access at the end of your research project?
11. 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

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

Responsibilities and Resources

15. 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.
16. 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?
17. 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

18. 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?
19. If applicable, what strategies will you undertake to address secondary uses of sensitive data?
20. 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 is 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;
4. Site description – including the following:
 - Data period(s) and location(s);
 - Equipment used, including manufacturer and model numbers;
 - Instrument heights and other relevant physical site characteristics;
 - Methods/software used in acquiring the data;
 - Data format, including examples;
5. Data processing and quality control – including the following:
 - Methods and software used in acquiring and processing the data;
 - Post-collection data processing – description of any processing done on the data;
 - Quality control methods – give an indication as to the degree of quality control;
 - 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 2.1 of the data policy);
6. 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 – Give sufficient detail (name, affiliation, full address, telephone and fax numbers, e-mail, etc.) to contact those most knowledgeable about the model run;
4. Run description – including the following (valid web links acceptable):
 - Period(s) and location(s)/resolution/map projection;
 - Initialization and boundary data used;
 - Model used – complete description of the model, physics package, any coupling state, etc.;
 - Data format – including examples;
 - Archive location/media – online link or offline contact person;
5. References.

Appendix C: 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.