UNIVERSITY OF SASKATCHEWAN
Global Water Futures
GWF.USASK.CA

National Hydrology Research Centre
11 Innovation Boulevard
Saskatoon, SK S7N 3H5 Canada
Tel: (306) 966-2021; Fax: (306) 966-1193
Email: gwf.project@usask.ca

GLOBAL WATER FUTURES – DATA MANAGEMENT FRAMEWORK AND TEAM

March 1, 2017

Scope

This document describes the data management framework for the "Global Water Futures: Solutions to Water Threats in an Era of Global Change" (or GWF for short) Research Programme. The scope covers the organizational framework and management, along with the database system. An accompanying data policy describes the submission, documentation, archiving, access, and acknowledgement policies adopted by the GWF Programme.

Central Data System and Management Team

To most effectively manage the vast array of data generated through GWF activities, all datasets will be processed and submitted to a central system with hardware and software resources provided and supported by University of Saskatchewan (UofS), Compute Canada, and major GWF stakeholders from the industry and academic domain. Resources for this centralized data system will be used by the GWF Programme's four key partner Universities (UofS, Wilfrid Laurier University (WLU), University of Waterloo (UofW), and McMaster University (MU)), as well as other partner institutions to process, archive, share, and disseminate the collected and produced records. Mechanisms will be provided for all data originators to upload data and update their archived records, as well as to securely share these data according to the terms and conditions of the GWF data policy.

Support to all GWF investigators and management of the system will be carried out by a data management team, with personnel located at UofS, WLU, UofW, and MU. The GWF Programme lead data manager will be Amber Peterson at UofS, with two data managers supported from core funds at MU (supervised by Mike Waddington) and WLU (supervised by Michael Steeleworthy). UofW are providing internal support for a data manager, who will liaise with Amber and the other data managers and facilitate contributions to the GWF Programme database. This team will work to support Amber and ensure that all aspects of the GWF data policy are effectively carried out in a timely manner, and they will meet by teleconference or other means on a regular basis (quarterly or more frequently) to coordinate activities. Following each meeting, the data management team will produce a status report outlining progress, concerns, and action plans for review by the GWF Strategic Management Committee. The members of this team will also liaise and coordinate closely with relevant members of other core support teams, facilitated by the GWF Science Manager, including the technical teams and technicians, core modelling teams,

knowledge mobilization team, and computing team. This will ensure the ready flow of data and information, coordinated use of best practices and protocols, and effective compliance with the GWF data policy.

Data Types and Management Subsystems

Data sources include real-time hydrometeorological data from the large-scale observatories, large datasets from drone and satellite remote sensing, data from chemical, cultural, economic and ecological observations, data from communities and First Nations, historical archives from multiple sources, very large output fields from high resolution atmospheric and Earth system models, and crowd-sourced data. These must all be handled according to our data policy. To support the management of different types and classes of data, the GWF Programme will employ different sub-units as integral parts of the centralized data management system, with the following units included in the infrastructure:

Unit	**Provider**
Water Information System Kisters (WISKI)	Kisters
Aquarius	Aquatic Informatics
Cluster with continuous access to 600+ cores	Compute Canada
Cluster(s) with smaller processing power	University of Saskatchewan
Data storage with up to 2PB storage space	Compute Canada, University of Saskatchewan

More resources may become available through IBM and/or other providers.

The following non-exhaustive list of data types and classes that will be collected through the GWF Programme has been identified, along with the appropriate unit to manage and host each.

Data Type / Class	**Unit**
Environmental time series data including weather, soil, and flux records	WISKI
Hydrometric time series data including stream flow, water level and water temperature	WISKI or Aquarius
Water Quality, isotope and laboratory sample data	Aquarius
Climate model data	Compute Canada cluster and storage
Processed drone obtained data	Compute Canada storage
Social and health related data	University of Saskatchewan data store
Satellite sensor data	University of Saskatchewan data store
Processed environmental DNA data	To be determined

All units will support back-up service as well as the protection from unauthorized access. Also, datasets hosted at each unit will be released to the public or internally in accordance to the GWF data policy.

Metadata records for each GWF-funded project will be used to build one comprehensive, centralized data catalogue that will be searchable, providing information on the data hosted in the system and originator contacts, as well as the links or references to the sets that are released for internal or external use. The data catalogue will have public and private mode of operation and not all metadata will be visible to the public, depending on the special requests coming from the originators. The central catalogue will be maintained by the data management team and support personnel.