



GLOBAL WATER FUTURES

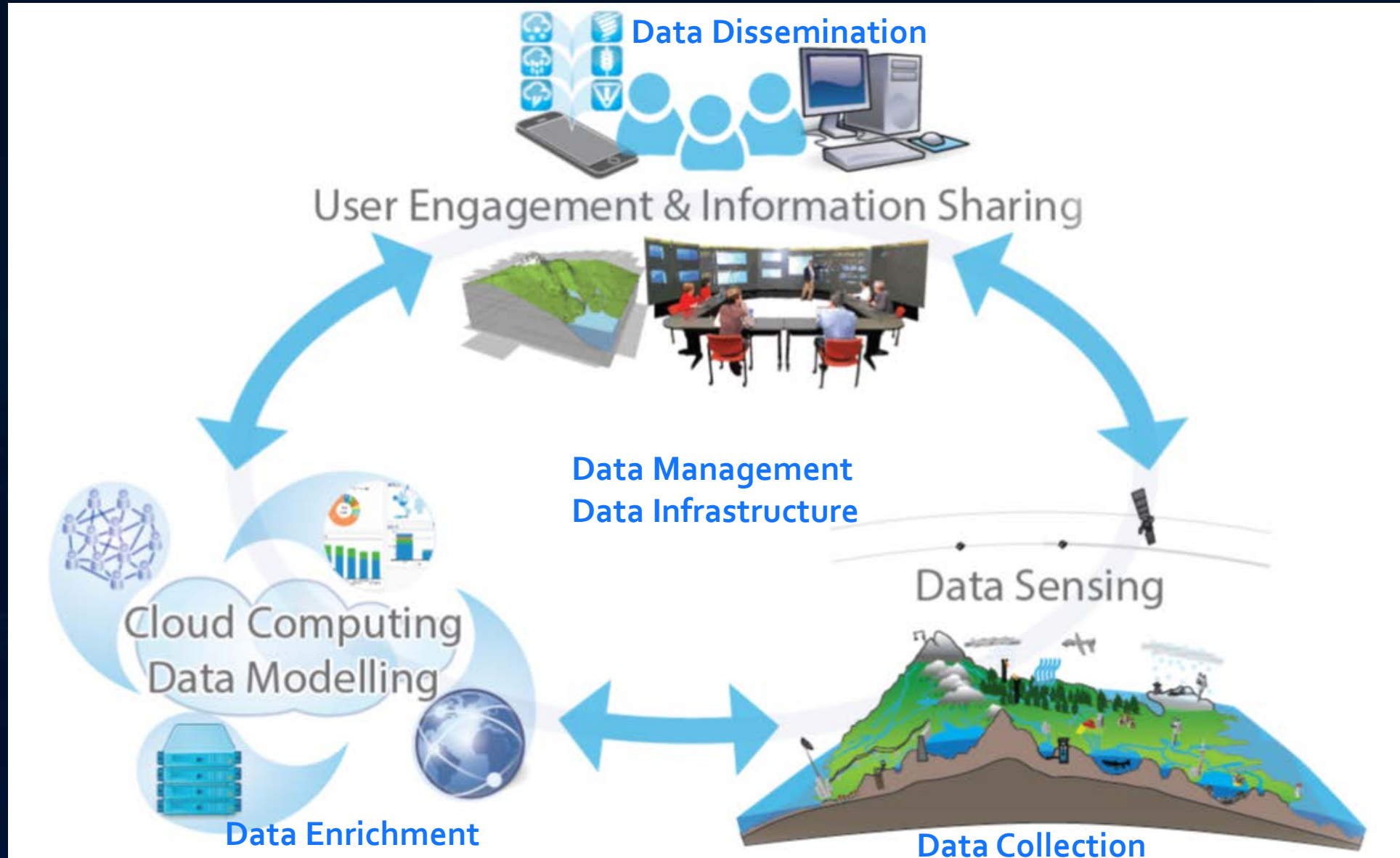
SOLUTIONS TO WATER THREATS
IN AN ERA OF GLOBAL CHANGE

WWW.GLOBALWATERFUTURES.CA

DATA COLLECTION | DATA ENRICHMENT | DATA MANAGEMENT | DATA INFRASTRUCTURE

Branko Zdravkovic Data Manager, GWF – GIWS – CCRN

GWF Inception Meeting, January 22-23





Overview

1

GWF Data Management

- Team Introduction
- Goals
- Project Organization

2

Data Infrastructure

- Systems
- Processes

3

GWF Project Requirements

- Storage
- Collaboration Data
- Projected Growth

4

Activities

- Workshops and Meetings



Data Management Teams

Wilfrid Laurier University

- Lead: Michael Steeleworthy
- DM: Gopal Saha

University of Waterloo

- Lead: Jimmy Lin
- DM: TBD

McMaster University

- Lead: Mike Waddington
- DM: Krysha Dukacz

University of Saskatchewan

- Lead: John Pomeroy
- DM: Branko Zdravkovic



Goals of the Data Management Team

PROJECT RELATED

Meet Funding Obligations

- Increase accountability
- Protect investment

Provide Governance

- Storage and discovery system
- Safeguard data, control access and ensure integrity

Meet Journal Requirements

- Data to support reproducibility
- Validation

Increase Impact

- Proper credit/citation
- Data Synergy



Goals of the Data Management Team

PROGRAM RELATED

Acquire

- Collect data from Observatories

Standardize

- Use uniform protocols for naming, organization, processing, quality assurance and dissemination
- Standardize metadata to adequately document and describe preserved datasets

Share

- Facilitate the efficient sharing of data and information among researchers
- Create a legacy data archive with stable and reliable services for the current and future users



Organization - Projects

DASHBOARD

EVERYTHING

PROJECTS

CALENDAR

STATUSES

PEOPLE

Global Water Futures

Categories

All Projects35

General1

Pillar 1 & 221

Pillar 312

No Category1

LIST

PORTFOLIO

CHART

Active Projects

Active12Current12

12 resultsA B C F I L N P

☆

▼

Agricultural Water Futures in Canada: Stressors and Solutions — (Global Water Futures)

USASK PI x

UW PI x

WLU PI x

— Updated: Tuesday Jan 16th 2018

☆

▼

Big Data Platform and "Smart" Watersheds — (Global Water Futures)

UW PI x

— Updated: Tuesday Jan 16th 2018

☆

▼

Canada's Boreal Wildland-Society-Water Nexus — (Global Water Futures)

Mac PI x

— Updated: Monday Jan 15th 2018

☆

▼

Climate-Related Precipitation Extremes — (Global Water Futures)

UM PI x

UVIC PI x

— Updated: Tuesday Jan 16th 2018

☆

▼

Co-creation of Indigenous Water Quality Tools — (Global Water Futures)

Mac PI x

— Updated: Monday Jan 15th 2018

☆

▼

FORecasting tools and Mitigation options for diverse BLOOM-affected lakes — (Global Water Futures)

USASK PI x

UW PI x

— Updated: Tuesday Jan 16th 2018

☆

▼

Future Water for the Mountain West — (Global Water Futures)

Mac PI x

— Updated: Tuesday Jan 16th 2018

⋮

☆

▼

Integrated Modelling Programme for Canada's Major River Basins — (Global Water Futures)

USASK PI x

— Updated: Tuesday Jan 16th 2018

☆

▼

Lake Futures: Enhancing Adaptive Capacity and Resilience of Lakes and Their Watersheds — (Global Water Futures)

UW PI x

— Updated: Tuesday Jan 16th 2018

☆

▼

Next generation solutions to ensure healthy water resources for future generations — (Global Water Futures)

USASK PI x

— Updated: Tuesday Jan 16th 2018

☆

▼

Northern Water Futures — (Global Water Futures)

WLU PI x

— Updated: Tuesday Jan 16th 2018

☆

▼

Prairie Water — (Global Water Futures)

USASK PI x

— Updated: Tuesday Jan 16th 2018



Organization - Tasks

DASHBOARD

EVERYTHING

PROJECTS

CALENDAR

STATUSES

PEOPLE

★

Future Water for the Mountain West

Global Water Futures

🔍

Quickly search this project...

KD

📅

Task Lists

All Lists

11

Data Management Plan Deve...

11

📊

Reports

Task Lists Report

Gantt Chart Export

☰

Views

Gantt Chart

Board View

OVERVIEW

TASKS

MILESTONES

MESSAGES

FILES

TIME

NOTEBOOKS

RISKS

LINKS

COMMENTS

PEOPLE

SETTINGS

Data Management Plan Development

+ Add a Task

Edit

✓

Krysha D.

Existing Data Inventory

[more...](#)

(Started Tue Jan 9th → Due Yesterday)

✓

Krysha D.

Overview of data to be collected over project life

[more...](#)

(Started Tue Jan 9th → Due Tue Feb 27th)

✓

Krysha D.

Assessment of current storage resources

[more...](#)

7

Krysha D.

Data Management Plan

(Starts Thu Jan 25th → Due Fri Feb 2nd)

✓

Krysha D.

List data to be collected

(Starts Thu Jan 25th → Due Fri Feb 2nd)

✓

Krysha D.

Determine format standards and naming conventions

(Starts Thu Jan 25th → Due Fri Feb 2nd)

✓

Krysha D.

Develop Metadata process to support consistent, robust documentation

(Starts Thu Jan 25th → Due Fri Feb 2nd)

✓

Krysha D.

Develop structure for data organization (file naming/version control)

(Starts Thu Jan 25th → Due Fri Feb 2nd)

✓

Krysha D.

Determine and create process to provide supporting documentation to fully describe data

(Starts Thu Jan 25th → Due Fri Feb 2nd)

✓

Krysha D.

Back-up Protocols - 3 copies - 2 media types - 1 off site

(Starts Thu Jan 25th → Due Fri Feb 2nd)

✓

Krysha D.

Determine access restrictions

(Starts Thu Jan 25th → Due Fri Feb 2nd)

Krysha Dukacz, GWF



Organization - Contacts

DASHBOARD

EVERYTHING

PROJECTS

CALENDAR

STATUSES

PEOPLE

★ **Future Water for the Mountain West**
Global Water Futures

OVERVIEWTASKSMILESTONESMESSAGESFILESTIMENOTEBOOKSRISKSLINKSCOMMENTSPEOPLESETTINGS

People on this Project

PeopleRoles

Quick Search

7 People: 4 users3 contacts

BMS✕ remove filter

McMaster University

SC

Sean Carey

PI

McMaster University

careysk@mcmaster.ca

University of Calgary

MH

Masaki Hayashi

PI

University of Calgary

hayashi@ucalgary.ca

University of Northern British Columbia

BM

Brian Menounos

PI

University of Northern British Columbia

menounos@unbc.ca



SOFTWARE

Water Information System Kisters (WISKI)

- Centralized database
- Processing of time series data
- Standardized QA/QC procedures
- Controlled sharing - web GUI and RESTful API

AQUARIUS – Aquatic Informatics

- Centralized database
- Processing of WQ observations and lab samples
- Cloud access
- Controlled sharing - web GUI

S-HA-RE

Compute Canada Graham Cluster

- Access to 1852 core-years
- Available to GWF researchers with CC accounts
- Processing of the GWF models at one location
- Planned 6000 core-years over 3 years

Centralized Repository

- Year 1: 650TB of project and 350TB nearline storage
- Over 200+ TB University of Saskatchewan storage
- Expected to increase over 2PB in total
- Globus supported transfer of data

HARDWARE

REPOSITORY

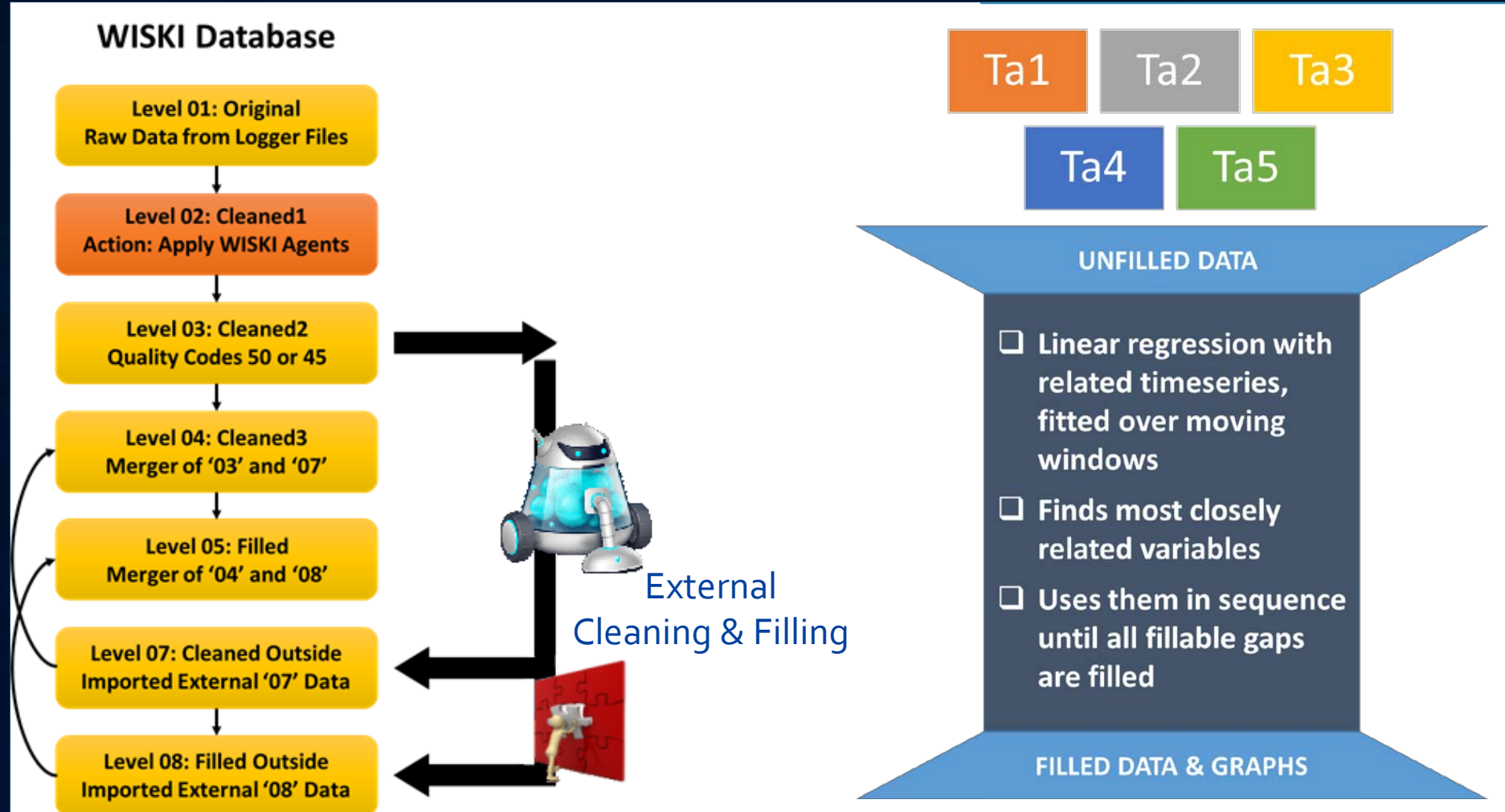


Standardized QA/QC Procedures

Quality Code		Processing Level								Quality
		01 raw	02	03	04 cln	05 fill	07	08	09	
255	M	●			●	●				Missing data
80	I					●			●	Infilled using WISKI agents
70	F					●		●		inFilled outside of WISKI
55	X		●							Auto-flagged by WISKI and eXcluded
50	R	●	●	●	●	●				original (R ecorded or logged) data
45	D		●							to be D ropped, flagged manually
40	E		●	●	●	●				manually E edited (in 02.Cleaned1)
30	C				●	●	●			Externally C orrected, imported as 07
25	Z						●			Flagged externally to be excluded

Imported Manual QC Populated automatically

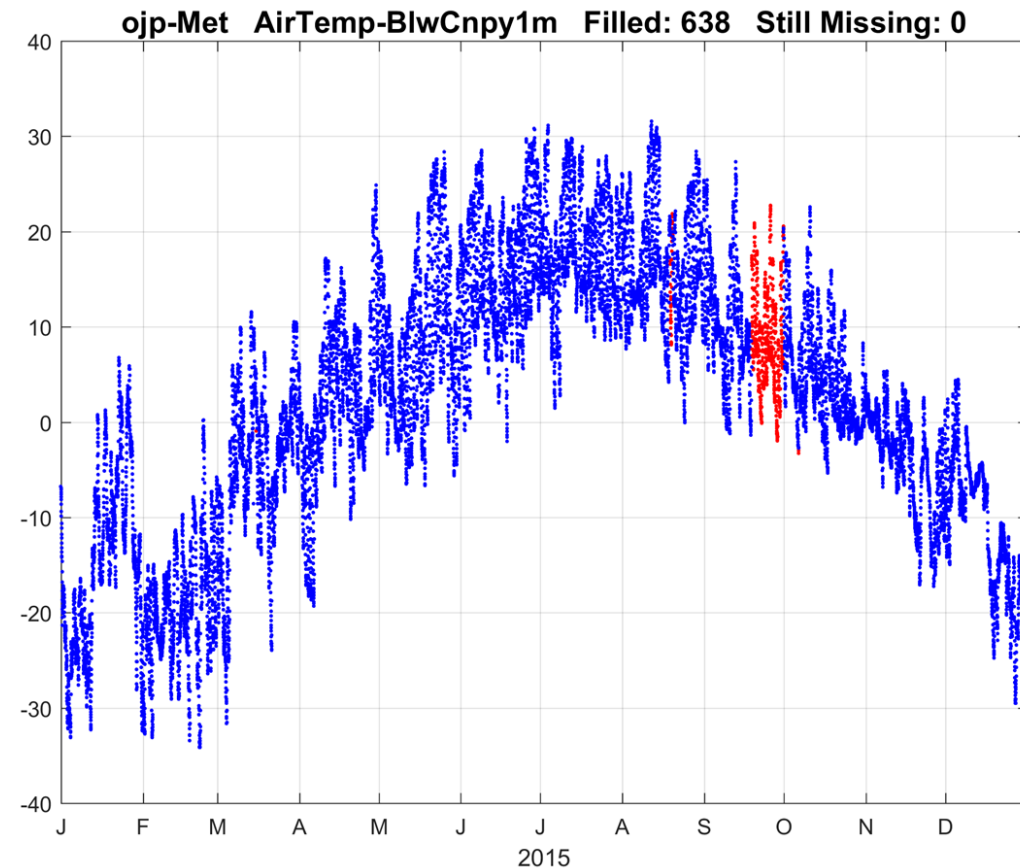
Automated Data Cleaning & Filling





Automated Data Cleaning & Filling

- Fully automated program
- Graphical output





Repository Allocation - Size (P-3)

Allocation per project excluding the airborne data

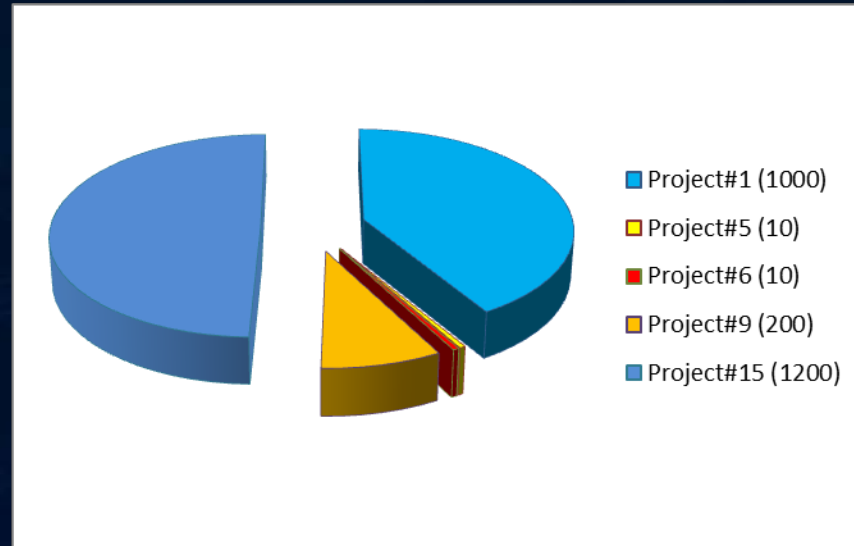


Figure: GWF Central Repository Allocation in TBs

- Over 2.5 PBs for observational datasets and model outputs in:
 - PROJECT 01: Climate-Related Precipitation Extremes
 - PROJECT 15: Big Data Platform and “Smart” Watersheds



Repository Allocation - Model Datasets (P-3)

The Weather Research & Forecasting Model (WRF)

Project
Project#1_Stewart-Zwiers_Climate-Related-Precipitation-Extremes_Manitoba-PCIC
Project#6_Macrae_Canadian-Ag-Water-Use_Waterloo
Project#7_Waddington_Canadas-Boreal-Wildland-Society-Water-Nexus_McMaster
Project#8_Spence_Prairie-WATER_UofS*
Project#9_Razavi_Integrated-Modelling_UofS
Project#11_Carey_Mountain-West_McMaster
Project#15_Duguay_Big-Data-Platform_Waterloo

* WRF in PGW mode

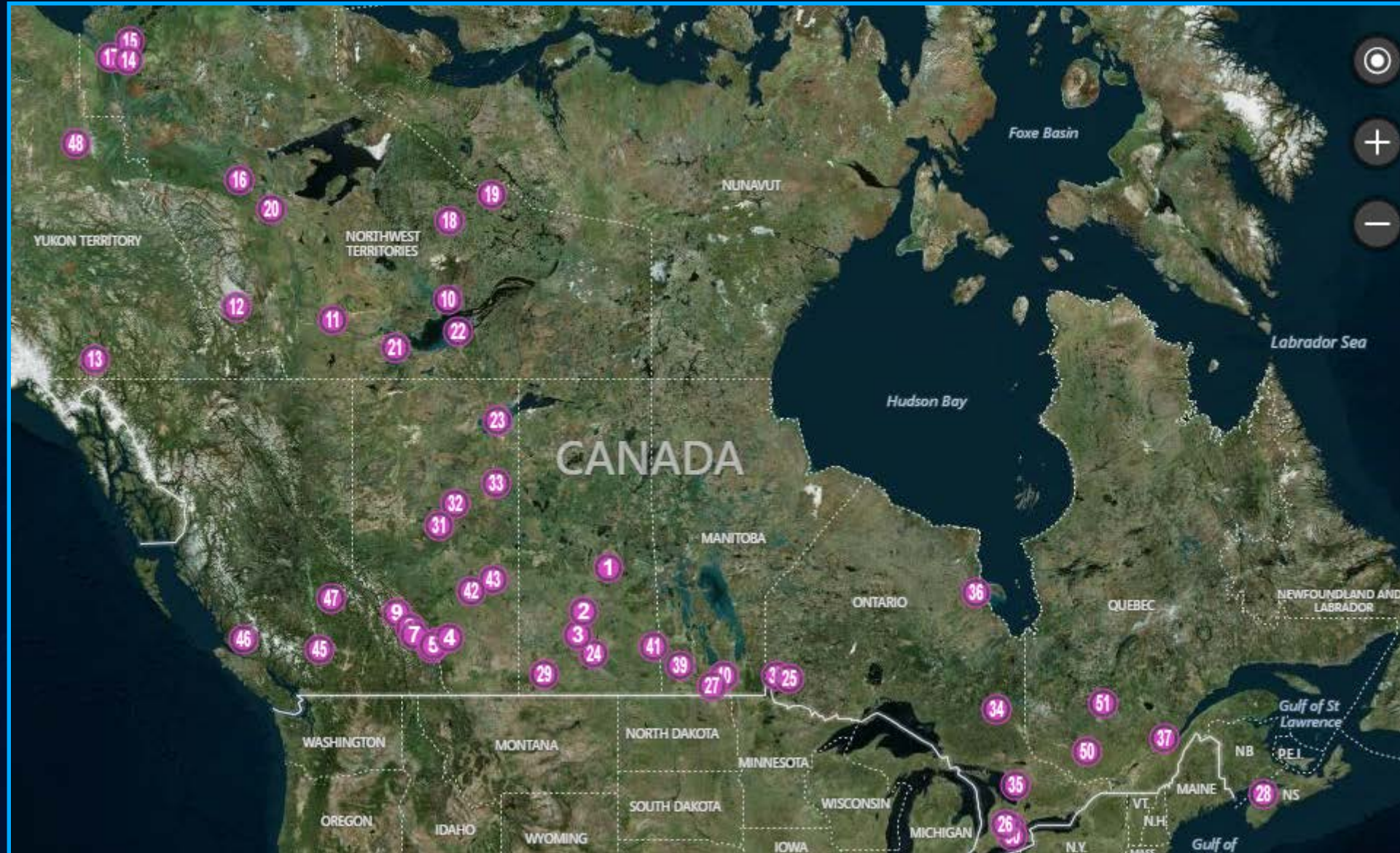
Growth of Observational Data





GLOBAL WATER FUTURES

SOLUTIONS TO WATER THREATS
IN AN ERA OF GLOBAL CHANGE





Current & Projected Data Acquisition

	CURRENT DATA	PROJECTED DATA (P-3)
Observatories	15	50
Monitoring Stations	50	150
Imported Variables	1800	10,000 – 20,000
Data Management Teams	University of Saskatchewan	University of Saskatchewan McMaster University Wilfrid Laurier University University of Waterloo
Collection Period	(1997) 2013 - 2018	2018 – 2020+
Location	Saskatchewan & Mackenzie River basins	Canada wide



Potential Meetings and Workshops

Workshops

- Research Data Management Overview
- Data Management Plan Development
- Data conversion
- Systems training – WISKI, Aquarius etc. Systems TBD

Meetings

- Meet with PIs or teams to:
 - Develop Format/Naming convention standards
 - Develop plans for data collection, consolidation, centralization etc.
 - Develop Metadata/Documentation procedures and strategies



Local Planning Details

WLU Data Management Activities, 2017-2018

- Data Management Needs Assessment
- Program and Project Data Management Plan Activities
- Metadata standardization and markup
- HQP and Investigator Training
- Outreach to CARL Portage
- In Development:
 - **Workflow and Best Practices for Data Capture and Consolidation**
 - **Data Management Protocols for Sensitive Data, Traditional Knowledge, and Indigenous Stakeholders**
 - **Further integration into projects**



GLOBAL WATER FUTURES
SOLUTIONS TO WATER THREATS
IN AN ERA OF GLOBAL CHANGE

QUESTIONS & COMMENTS



GLOBAL WATER FUTURES
SOLUTIONS TO WATER THREATS
IN AN ERA OF GLOBAL CHANGE

Telemetry Sites



UNIVERSITY OF SASKATCHEWAN

Global Institute for
Water Security



CCRN
Changing Cold Regions Network



UNIVERSITY OF
CALGARY



Environment
Canada



Telemetry Sites

Collection Map ▶

AB Fortress Mountain ▼

Fortress Ledge

Fortress Ridge

Fortress Ridge South

Bonsai

Canadian Ridge

Canadian Ridge North

Powerline

AB Marmot Creek ▶

AB Remote Stations ▶

AB GRIP Stations ▶

SK St Denis NWA ▶

SK BERMS ▶

BC Lake O'Hara ▶

YT Wolf Creek ▶

NT Nahanni NP ▶

Telemetry Sites

Canadian Rockies Hydrological Observatory

Centre for Hydrology
University of Saskatchewan

Station:	Fortress Ledge
Location:	Fortress Mountain
Altitude:	2565 m
TimeZone:	CST, UTC-6



Automated collection: most recent data not verified by provider

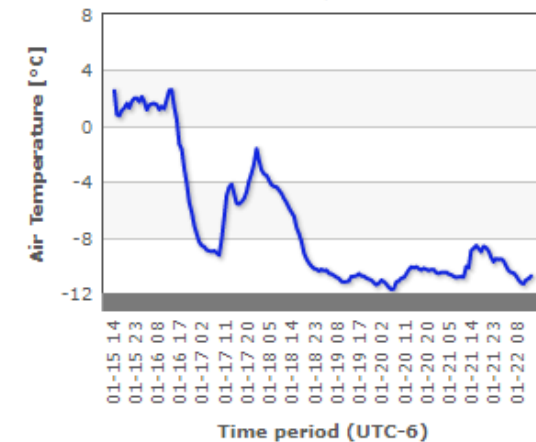
7-day

Recent

History

Air Temperature

Jan 22, 2018



<http://giws.usask.ca/telemetry>



GWF Core Data - Forcing

Name	Temporal resolution	Temporal extent	Spatial resolution	Spatial extent	Variables	Format
GEM (RDPS, HRDPS)	1-hour	2001/10- present	0.22 deg (~24 km): 2001/10- 2004/05/17 0.1375 deg (~15 km): 2004/05/18- 2012/10/02 10 km (likely 0.09 deg): 2012/10/03- present 0.0225 degree (~2.5 km): 2012/10/03- present	North America: 2016/09/08- present Canada, Mexico, conterminous US: 2010/11- present Canada, continental US: 2001/10-present	Precipitation, temperature, pressure, specific humidity, wind speed, downward SWR, downward LWR	fst (with rmnlib), ASCII, GRIB 2 (convertible to NetCDF)
CaPA (RDPA)	6-hour Daily	2002-present	10 km	Canada, Mexico, conterminous US: 2002/01- present	Precipitation	fst (with rmnlib), ASCII, GRIB 2 (convertible to NetCDF)
WFDEI	3-hour Daily	1979-2016	0.5 degree	Global	Rainfall-CRU, Snowfall-CRU, Rainfall-GPCC, Snowfall-GPCC, temperature, pressure, specific humidity, wind speed, downward SWR, downward LWR	NetCDF



GWF Core Data - Forcing

Name	Temporal resolution	Temporal extent	Spatial resolution	Spatial extent	Variables	Format
WFD	3-hour Daily	1901-2001 (updates?)	0.5 degree	Global	Rainfall-CRU, Snowfall-CRU, Rainfall-GPCC, Snowfall-GPCC, temperature, pressure, specific humidity, wind speed, downward SWR, downward LWR	NetCDF
Princeton V.1 and V.2	3-hour Daily Monthly	V1: 1948- 2008 (updates?) V2: 1901- 2012 (updates?)	V1: 0.25, 0.5, 1.0 degree V2: 0.5, 1.0 degree	Global	Precipitation, temperature, Tmin, Tmax, Specific humidity, Downward SWR, downward LWR, Wind Speed, Surface Pressure.	NetCDF
NARR	3-hour Daily Monthly	1979-2015	0.3 degree (32 km)	Lat: 15 to 90 Lon -50 to -170	Precipitation, temperature, pressure, specific humidity, wind speed, downward short wave radiation, downward long wave radiation (??)	NetCDF



GWF Core Data - Forcing

Name	Temporal resolution	Temporal extent	Spatial resolution	Spatial extent	Variables	Format
ANUSPLIN	Daily Monthly	1950-2015	10 km	Canada	Precipitation, Tmin, Tmax,	ASCII
CANGRD	Monthly seasonal yearly	1948- present (?) for entire Canada 1900-present (?) for southern Canada	50 Km	Canada	Temperature and precipitation anomalies	ASCII
AHCCD	Daily	Varies based on stations	N/A	Canada	maximum, minimum and mean temperature rainfall and snowfall and total precipitation	ASCII



GWF Core Data – Climate Projection

Name	Temporal resolution	Temporal extent	Spatial resolution	Spatial extent	Variables	Format
Can-RCM4_RCP8.5	3-hour	1979-2100 Just one run corrected so far for testing purposes. For now I have corrected the data against WFDEI but other surrogates would be used depending on GWF project needs. We plan to extend back to 1950 too.	0.125 degree	longitudes -142,-90, and latitude 45,75 which encompasses the MaCkenze and Saskatchewan River Basins	precipitation, temperature, pressure, specific humidity, wind speed, downward short wave radiation, downward long wave radiation	NetCDF
PCIC	Daily	1950-2100	0.0833 degree	Canada	precipitation, min and max temperature	NetCDF
NA-CORDEX	Daily	1950-2100	0.22 degree / 0.44 degree	North America	this simulation matrix gives more information as well including the RCPs https://na-cordex.org/simulation-matrix https://na-cordex.org/variable-list	NetCDF



Name	Spatial Resolution or Scale	Spatial extent	Land Cover Classes	Format
Circa 2000-Vector	1:250,000	Lat: 43 to 83 Lon: -141 to -52	42	shapefile
Agricultura 1 Circa 2000	30 m	UTM Zone 9 - 22	14 (??)	raster
2005 Land Cover of North America (Ed. 2)	250 m	Lat: 14 to 84 Lon: -170 to -50	19	raster
2010 Land Cover of North America (Ed. 1)	250 m	Lat: 14 to 84 Lon: -170 to -50	19	raster
2010 Land cover of Canada (North America) at 30 meters	30 m	Entire Canada	19	raster
National Atlas of Canada, 5th Edition	N/A	North America Lat: 47 to 84 Lon: -141 to -50	N/A	Image

GWF Core Data – Land Cover



GWF Core Data – Soil Data

Name	Spatial Resolution or Scale	Spatial extent	Format
Soil landscape of Canada, V2.2	1:1,000,000	Canada	shapefile
Soil landscape of Canada, V3.2	1:1,000,000	Agricultural regions of Canada	shapefile
Unified North American Soil Data (UNASM soil data)	250 m	Canada, USA	Raster
STATSGO2 USA	1:250,000 in US, HI, PR, VI and 1:1,000,000 in AK	USA	USA
Harmonized World soil data (FAO)	0.00833 degree	Global	Global



GWF Core Data – DEM

Name	Spatial Resolution or Scale	Spatial extent	Format
Canadian Digital Elevation Model (CDEM)	0.00028 degree (approx. 25 to 90 m)	Canada	Raster, many mosaics
High resolution Digital Elevation Model (HRDEM)	1-2 m	Few provinces, MN, MB, NS, QB (excluding Saskatchewan)	Raster
Conditioned DEM from Hydrosheds (SRTM)	0.0008333 degree (approx. 90 m)	LAT 40 to LAT 60	Raster
SRTM	0.0008333 degree (approx. 90 m)	Global (not available above 60 N)	Raster



GWF Core Data – Shape Files

Name	Spatial extent	Format
Basin Polygons	Canada	shapefile
Basin polygons for WSC gauges	Canada	shapefile
hydrobasin arctic level 0 to 12	arctic region	shapefile
hydrobasin arctic with lakes level 0 to 12	north America except arctic region	shapefile
hydrobasin north America level 0 to 12	arctic region	shapefile
hydrobasin north America with lakes level 0 to 12	north America except arctic region	shapefile
15 arc-second north American rives		shapefile
Non-contributing areas		Shapefile