Data Management Team Advances

Amber Peterson

Data Manager, Global Water Futures

Global Water Futures Operations Meeting
January 22, 2019

UNIVERSITY OF SASKATCHEWAN
Global Water Futures
GWF.USASK.CA

GWF

Overview

- Introduction

- Team

- Advances
  - Data Portal
  - WISKI Advantages
  - Radiam
  - Dataset Publication

- Next Steps

Introduction

- Data is an important legacy of GWF

- 2 important documents developed at the start of the program: GWF Data Policy, DM Framework

- The DM team tasked with ensuring that the data is preserved, findable and accessible in the long-term so that data can be used for future science and modelling generations after GWF

Structure of the DM team

	University of Saskatchewan	University of Waterloo	McMaster University	Laurier
Faculty Lead	John Pomeroy	Jimmy Lin	Mike Waddington	Michael Steeleworthy
Data Manager	Amber Peterson	Bhaleka Persaud	Krysha Dukacz	Gopal Saha
Support Staff	Laleh Moradi [1]	Julie Mai [2]		

[1] Research Data Analyst
[2] Core Modeller, Big Data Dissemination

Information on the Team

The data managers meet every two weeks

Fall 2018 Data Management Retreat (Oct 29-Nov 1)
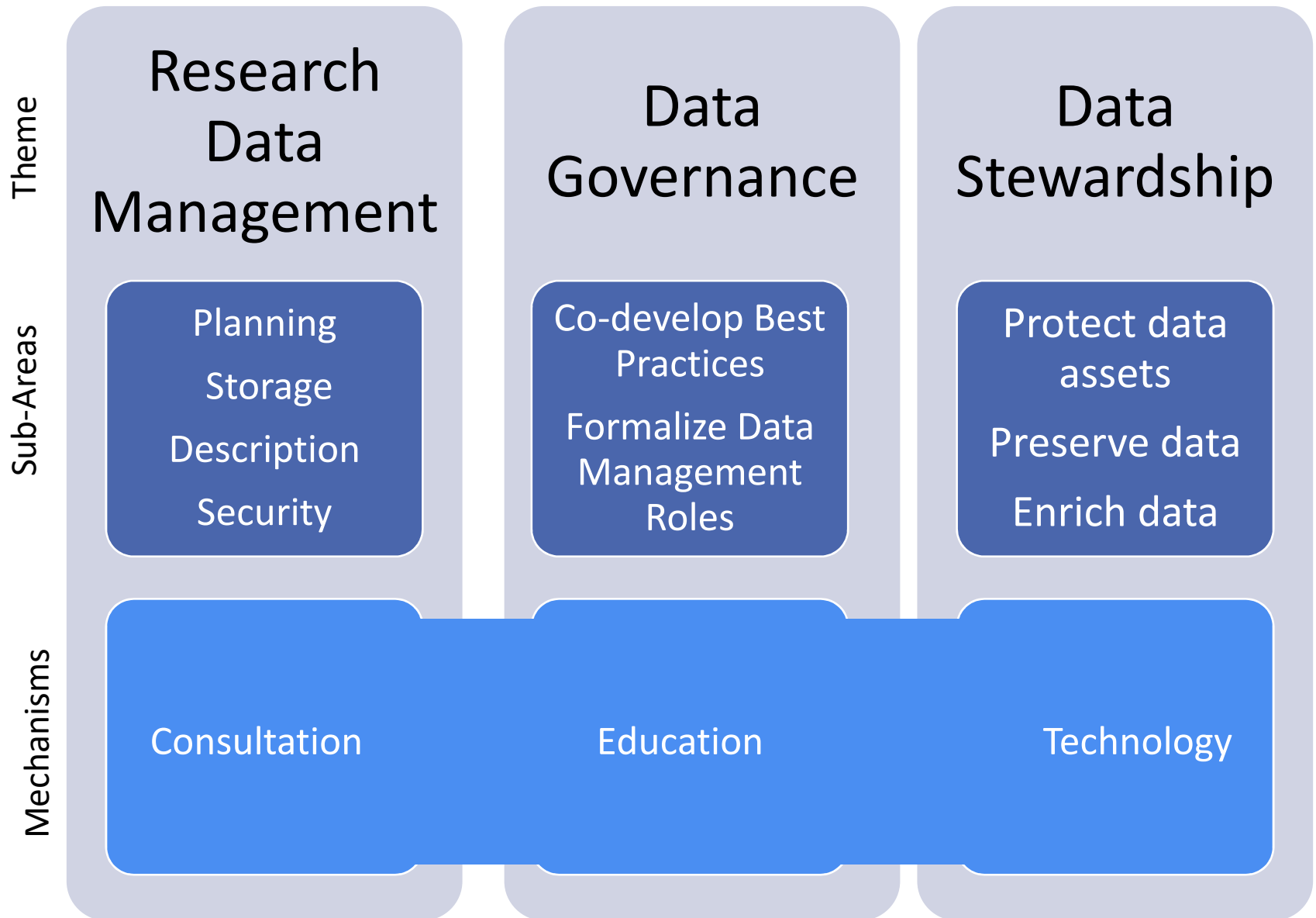
All project groups should be in contact with their local data manager.

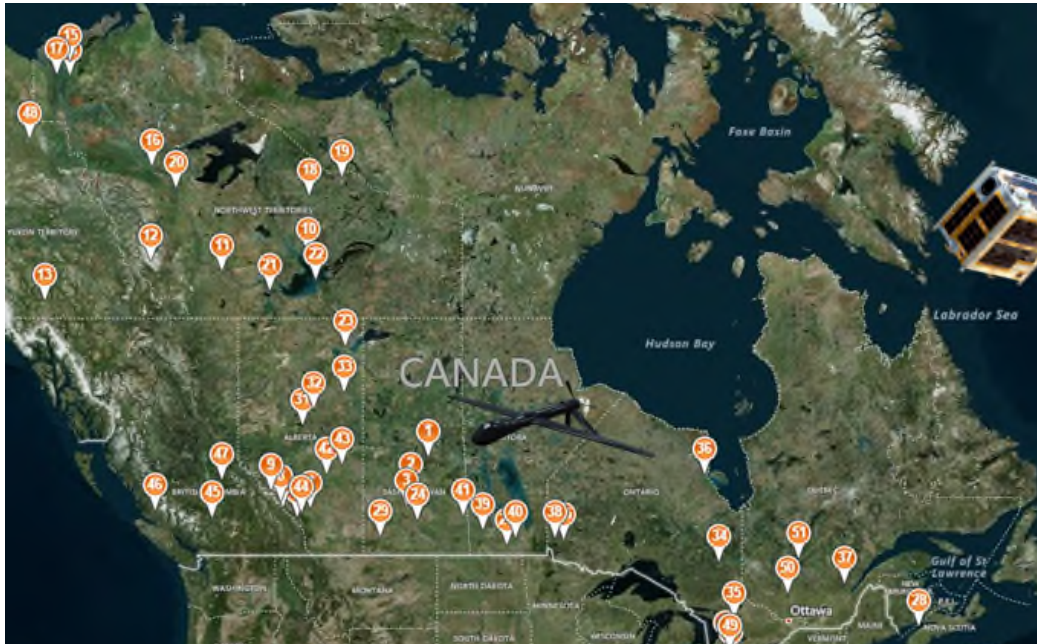**Each data manager should be able to provide similar data management services.**

Our members liase with library and technology organizations (Portage & ICT), Research Ethics Board
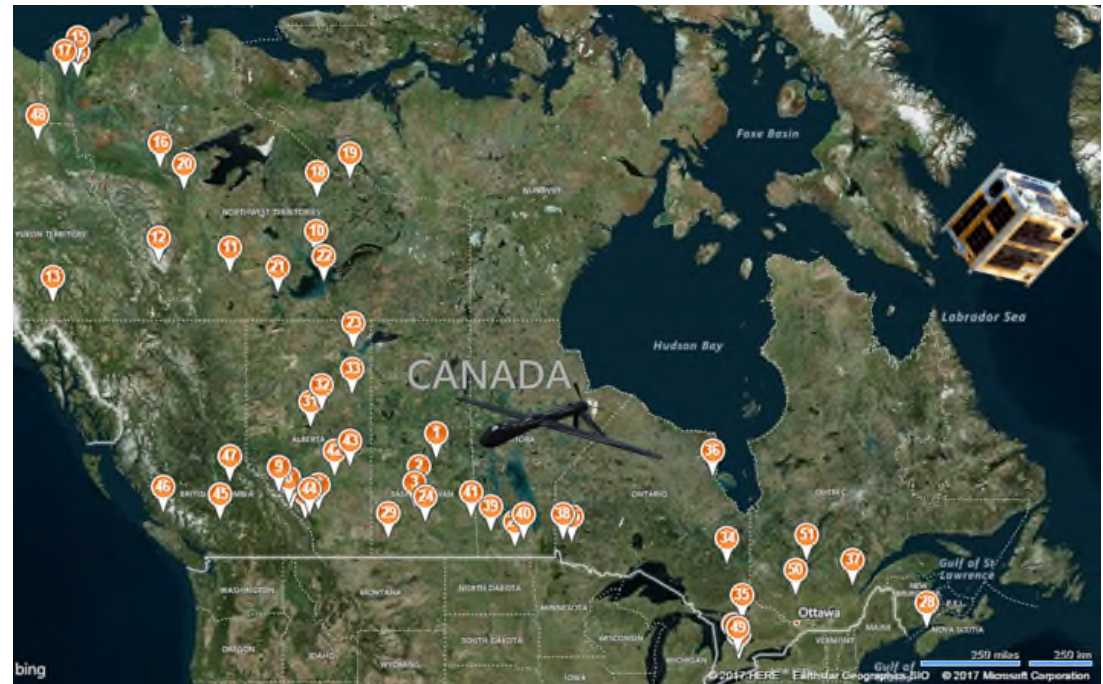
DM Team Objectives

Challenges of GWF Data

- 60+ observatories, 33 projects
- Many different data types
  - Wide variety of temporal and spatial scales, data formats, and frequencies

- Many partners
  - Data stored and processed at many locations
  - Communication
- Varying levels of data sensitivity
  - Data embargos and sharing restrictions

Data Management Activities

1) Create a data inventory (Metadata Form)
2) Continue to assist researchers with data management (Services List)
3) Create standards and best practices documents, including workflows
4) Develop a continued data management strategy
   - Identifying operational strategies – Short Term/Long Term
   - Evaluating components for GWF repository (Aquarius, KiBiD)
   - Identifying infrastructure for backup, sensitive data, and preservation
5) Provide workshops and create a website to facilitate communication
   - Research data management, metadata
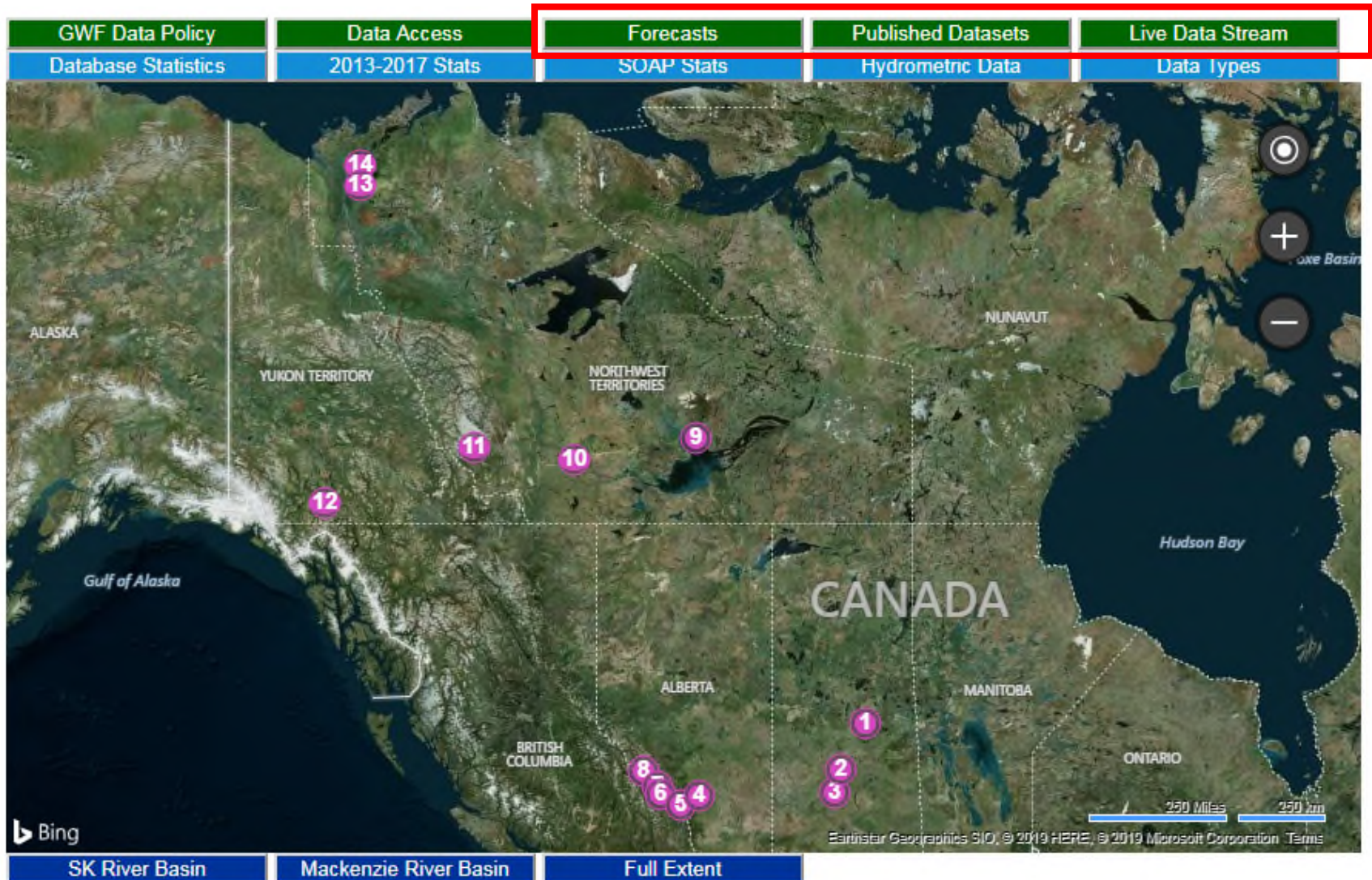
Assists Researchers with:

- Data management plans & annual reporting
- Development of rich metadata
- Standardized file formats and organization
- Accessing computing resources
- Data storage and backup
- Data sharing and sharing agreements
- Data archiving and preservation

Data Portal
http://giws.usask.ca/meta/

NEW TABS

Data Portal
http://giws.usask.ca/meta/

Forecasts

Name	Website	Developers	Description
SnowCast	http://www.snowcast.ca/	Chris Marsh, Nic Wayand	Snowpack estimates over the BowRiver Basin (centered over Banff). This is an experimental product that uses the GEM model forecasts to drive the Canadian Hydrological Model.
CaSPAr	http://www.caspar-data.ca/	Juliane Mai (backend), Kurt C. Kornelsen (frontend), Michael Leahy (frontend), and others	Canadian Surface Prediction Archive. The portal contains 10 products in total of which 5 are operational forecasts (GEPS,GDPS,REPS,RDPS,HRDPS), 4 are analyses (CaPA_coarse, CaPA_fine, CaPA_fine_exp, CaLDAS), and one is a re-analysis (RDRS). Find more information under https://github.com/kckornelsen/CaSPAr_Public/wiki .

Data Portal

UNIVERSITY OF SASKATCHEWAN
Global Water Futures
GWF.USASK.CA

http://giws.usask.ca/meta/

Library of Published Datasets for the GWF and CCRN Programmes

Filters: list only papers which...

...include this author (1 only):	
...were published in this year:	
...include in their title the following word, sub-string, or (exact) phrase:	

Sort by...   Pub. Year / youngest to oldest ▾

(comparisons are case-insensitive: empty field to remove corresponding filter)   Submit

- Asong Z.E., Wheater H.S., Pomeroy J.W., Pietroniro A., Elshamy M. (2018)
  A Bias-Corrected 3-hourly 0.125 Gridded Meteorological Forcing Data Set (1979-2016) for Land Surface Modeling in North America
  FRDR
  https://doi.org/10.20383/101.0111

- Bam E., Brannen R., Budhathoki S., Ireson A., Spence C., Van der Kamp G. (2018)
  Atmospheric, soil, surface and groundwater data from the St Denis National Wildlife Area, Saskatchewan, Canada
  FRDR
  https://doi.org/10.20383/101.0115

- Coles A., Russell M., Onclin C., Helgason W., Peterson A., Solohub M., McDonnell J. (2018)
  The Swift Current hillslopes, Saskatchewan: Digital elevation data
  FRDR
  https://doi.org/10.20383/101.0117

- Fang X., Pomeroy J.W., DeBeer C., Harder P., Siemens E. (2018)
  Hydrometeorological data from Marmot Creek Research Basin, Canadian Rockies
  FRDR
  https://doi.org/10.20383/101.09

- He J. and Hayashi M. (2018)
  Hydrological and meteorological dataset from the Lake O'Hara alpine hydrological observatory, 2004-2017
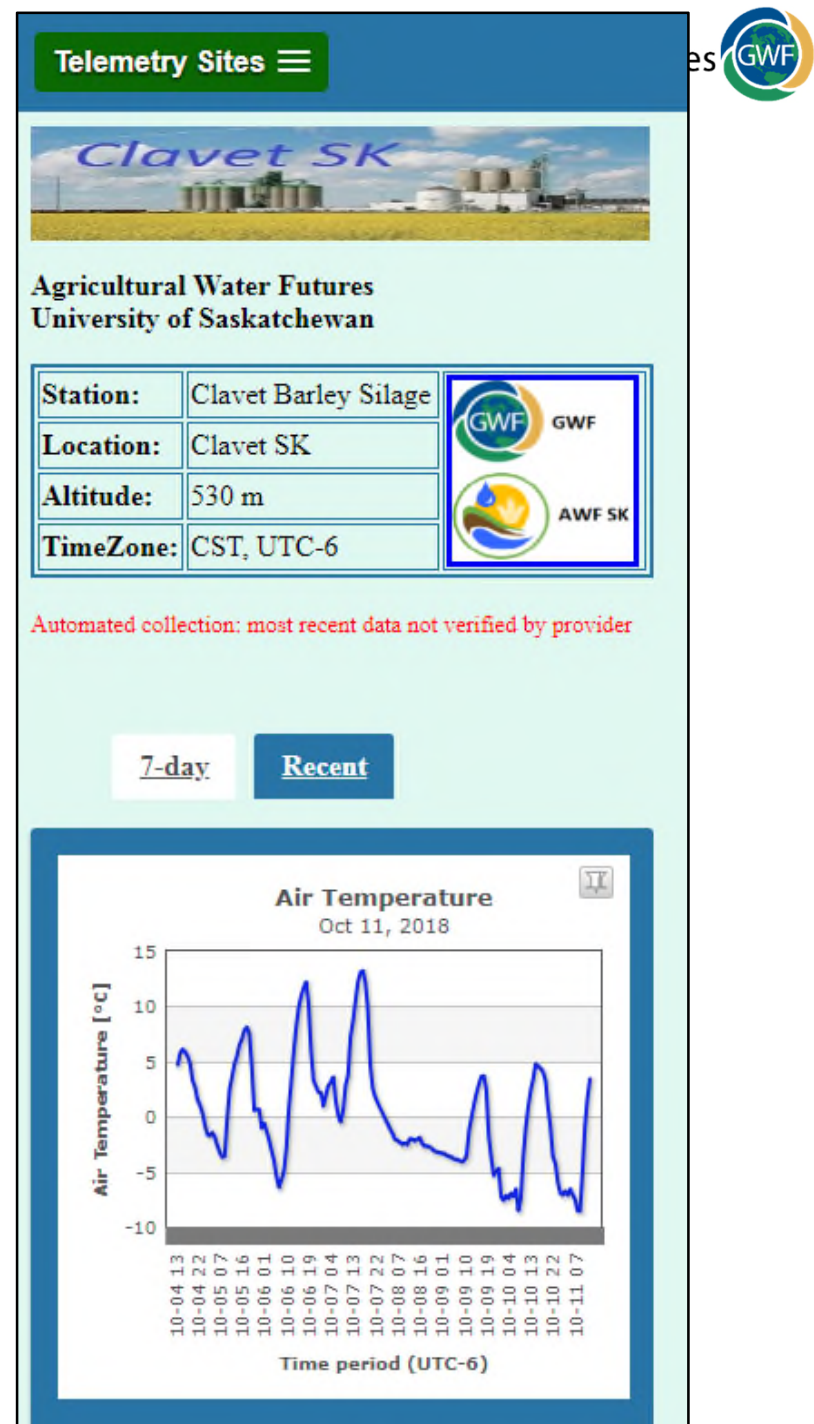 *FRDR*
  https://doi.org/10.20383/101.035

Advances

Telemetry Website
http://giws.usask.ca/telemetry/


AB | BC | SK | YT | Full

1. Boreal Ecosystem Research and Monitoring Sites
2. St. Denis National Wildlife Area
3. Marmot Creek Basin
4. Fortress Mountain
5. Rocky Mountain Remote Stations
6. Alberta GRIP Stations
7. Lake O'Hara Research Basin
8. Wolf Creek Research Basin
9. Nahanni National Park
10. SK Agriculture Sites

Telemetry Sites ≡

Clavet SK

Agricultural Water Futures
University of Saskatchewan

Station:	Clavet Barley Silage
Location:	Clavet SK
Altitude:	530 m
TimeZone:	CST, UTC-6

Automated collection; most recent data not verified by provider
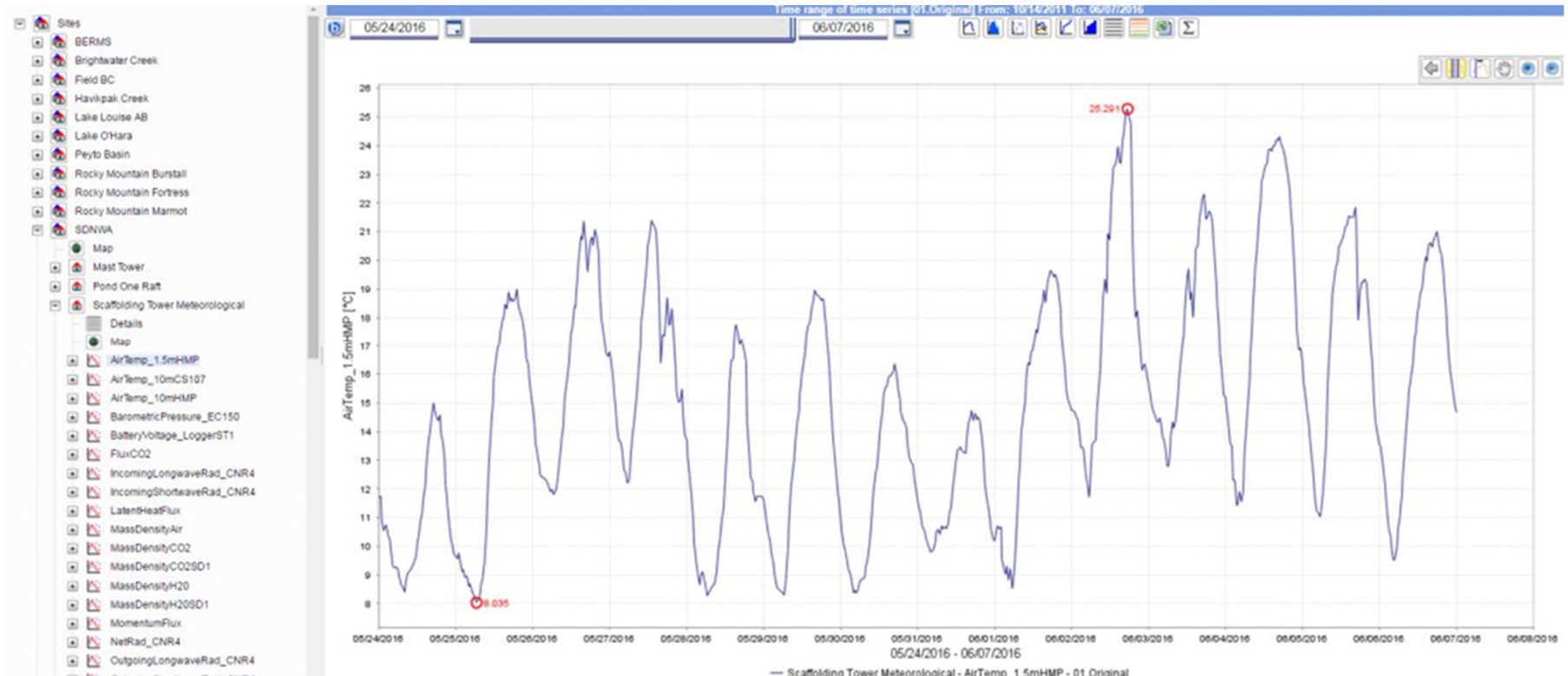
7-day | Recent

Air Temperature
Oct 11, 2018

Advances

For Technicians

- Weekly reports with comparison of like variables
- View instrument and telemetry failures

Fortress - Air Temperature
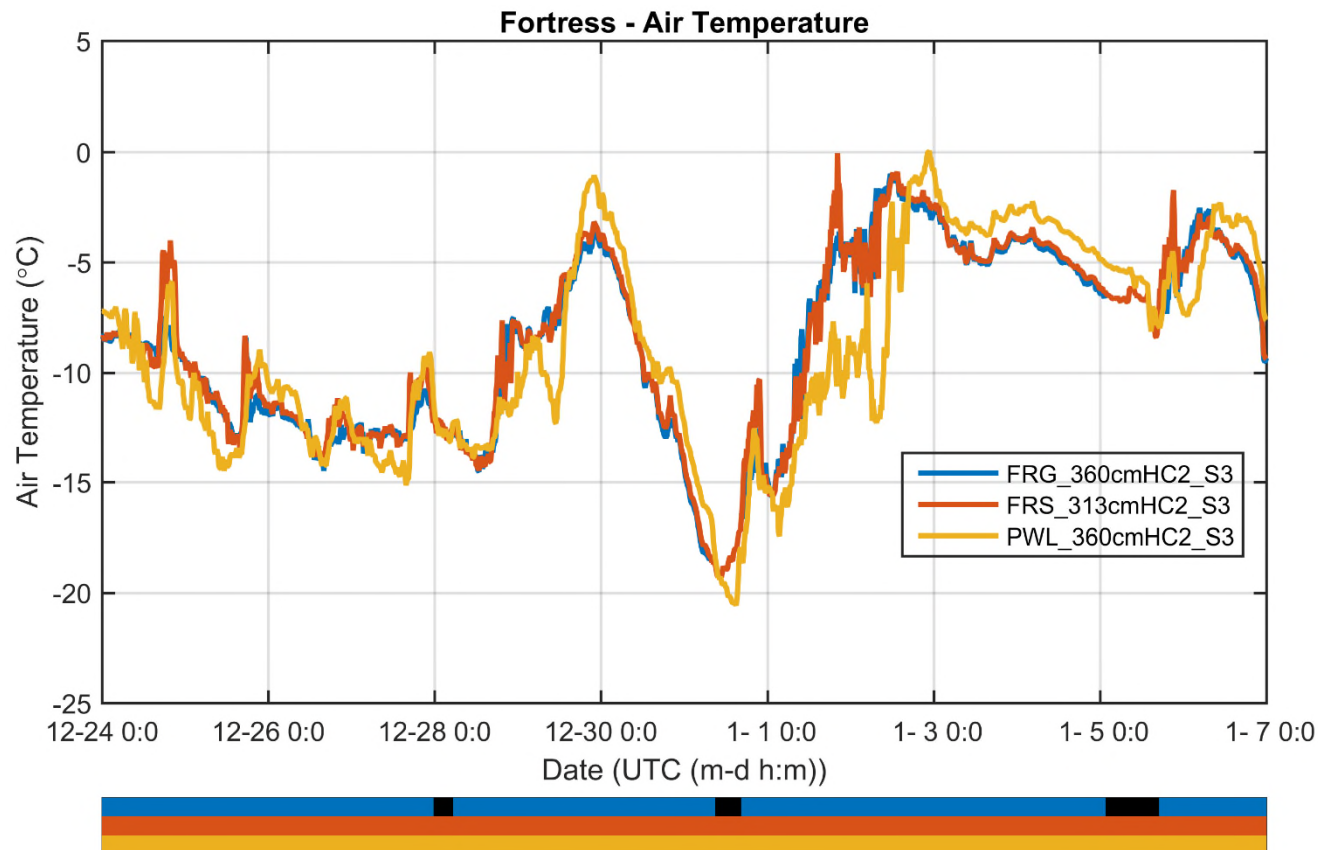
WISKI Advantages

- Searchable database, customizable exports with R, Python, and MATLAB scripts

- Visualization through Web Client

WISKI Advantages

- Preliminary cleaning of data
- Tracks data quality with codes
- If cleaning & filling externally, compare with original & assign codes

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

# Radiam (Active Research Data Management Platform)

- Provides a central point of discovery for data that lives in multiple storage locations

- Improves researchers' ability to discover vital project data from colleagues, regardless of where the data is physically stored

- March 2019 - Beta users

- June 2019 - Begin roll-out to other GWF groups

Radiam System Architecture
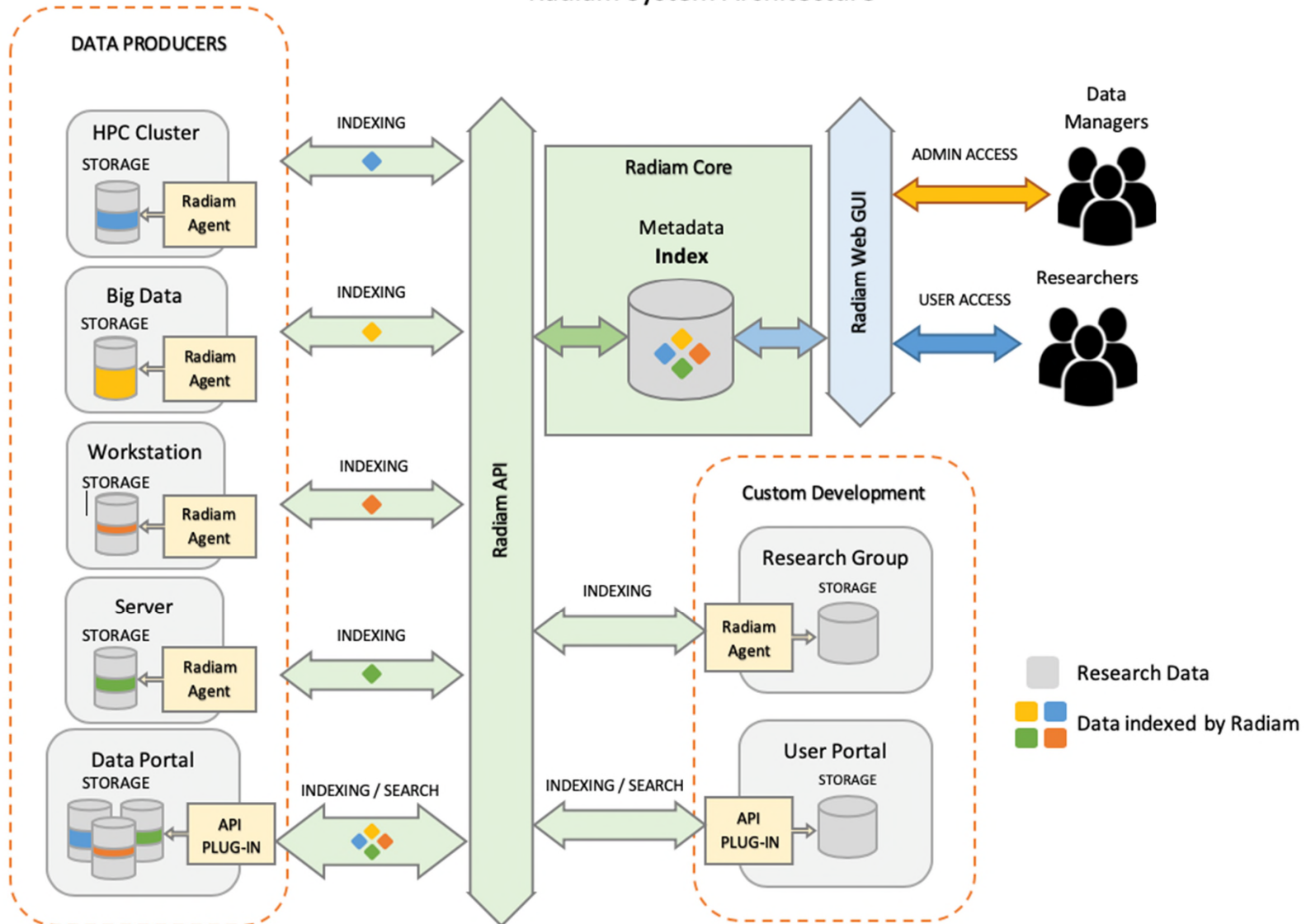
Dataset Publication

- FAIR data principles (published in Nature 2016)
  - Findable, accessible, interoperable, reusable

- Journals encourage or require dataset publication before the paper is accepted

- Types of data repositories
  - Federated repositories (e.g. FRDR)
  - Local repositories (e.g. Dataverse)
  - Discipline specific (list at re3data.org)

Next Steps

- Meet with HQP and researchers to collect metadata

- Update the data portal

- Radiam: provide feedback during beta-testing, migrate projects over during roll-out

- Compile data management resources GWF website

- Continue to provide data management support to projects

Thank You
Questions?