



Data Management

Operations Meeting 2021

Amber Peterson (Saskatchewan)

Laleh Moradi (Saskatchewan)

Bhaleka Persaud (Waterloo)

Gopal Saha (Wilfrid Laurier)

Krysha Dukacz (McMaster)

Stephen O'Hearn (Saskatchewan)



Discussed During This Presentation

Data Management Considerations

Current Data Management Activities

Data Management Team Highlights

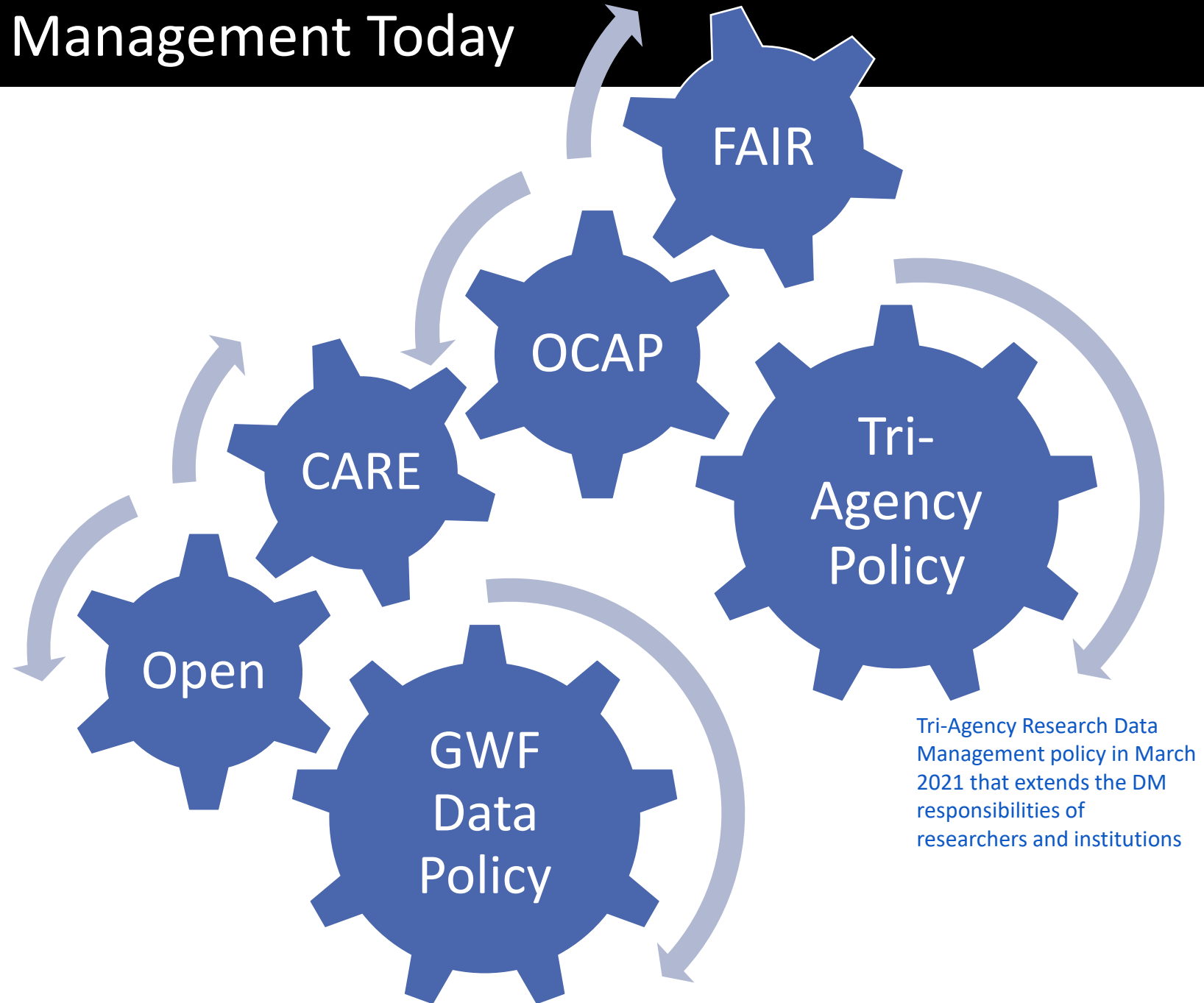
The Year Ahead

DM Considerations: Data Management Today

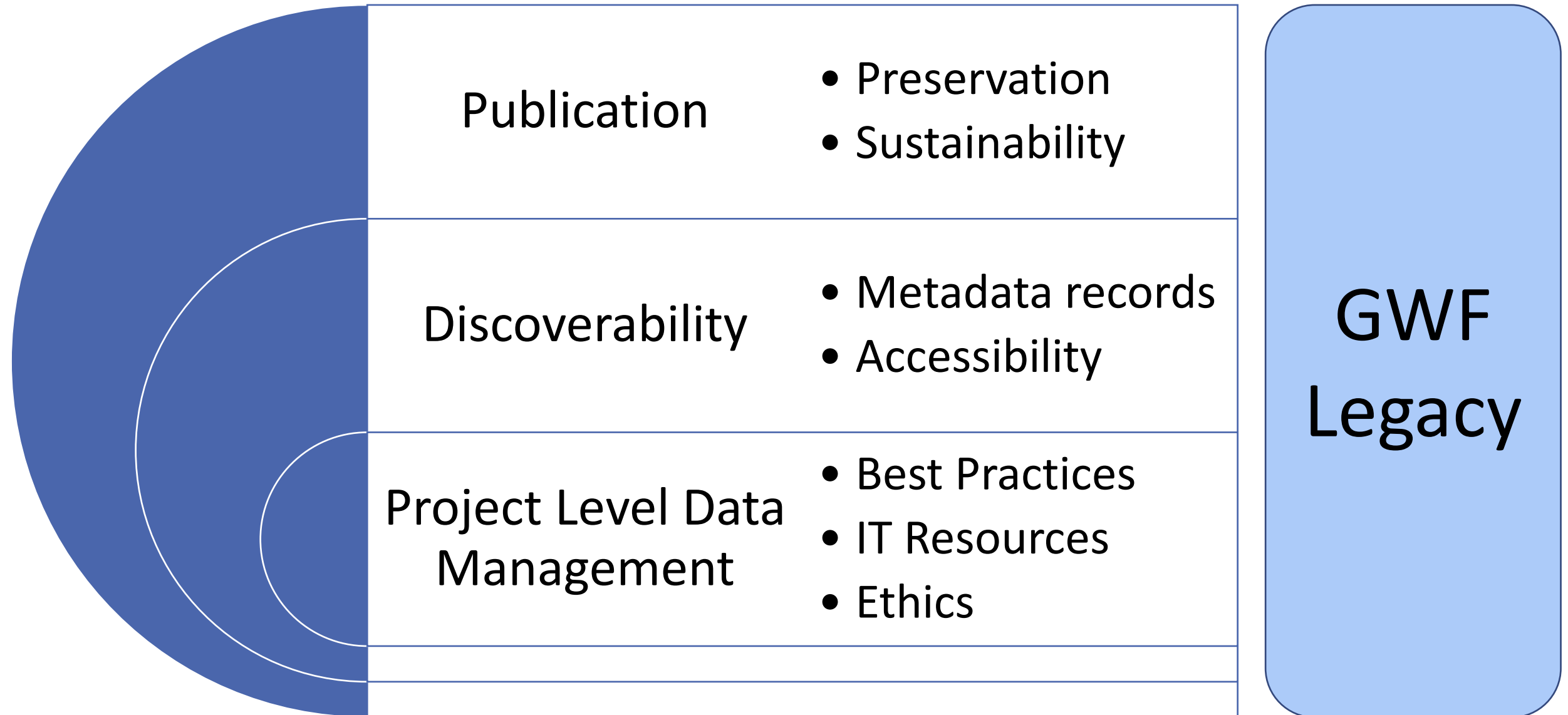
The Data Management landscape is evolving shaped by the directives of funders, publishers and stakeholders to enhance and improve the:

- Accountability of researchers to uphold ethics standards
- Accessibility of publicly funded data
- Publication of data on its own or accompanying journal articles

The GWF Data Policy closely aligns with the current landscape and supports its “special obligation to openness and accountability”



DM Considerations: Data Policy Goals

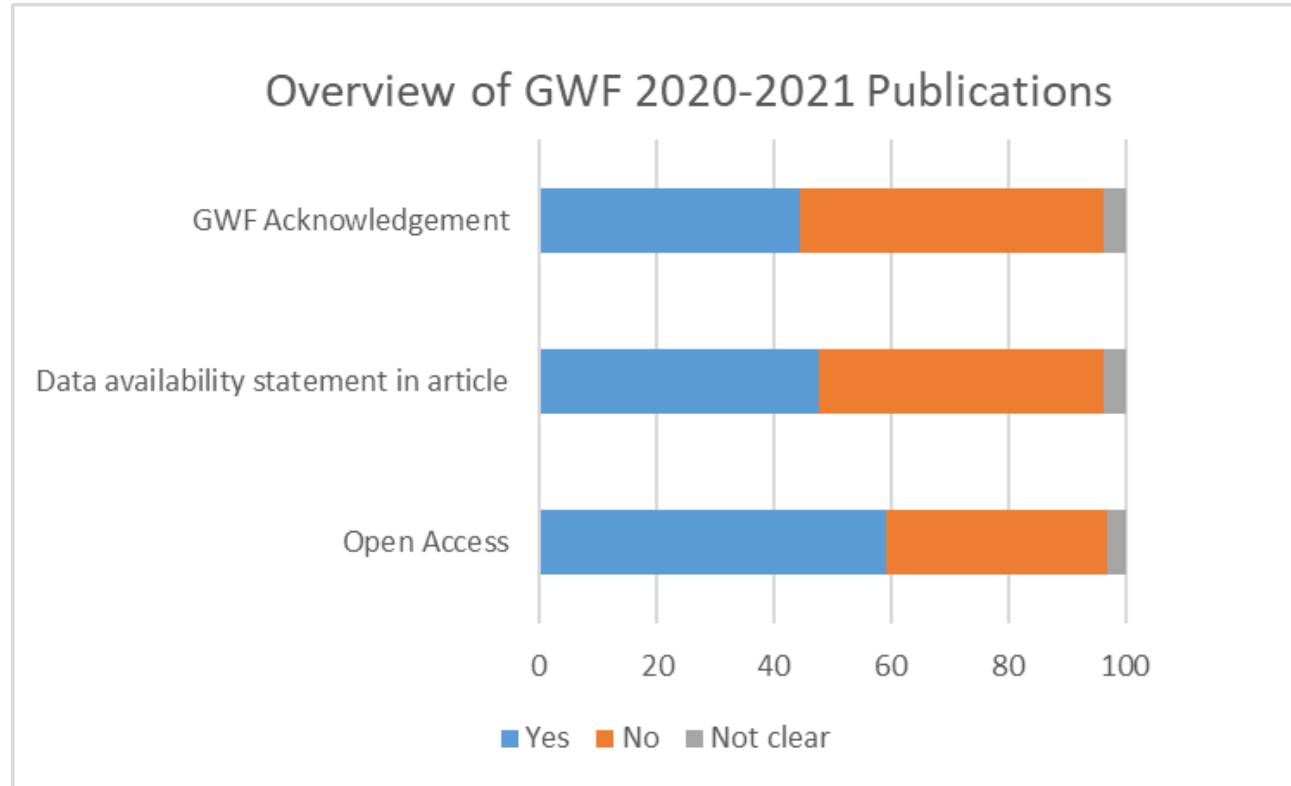


DM Considerations: Review of GWF Publications:

As the world moves toward FAIR, increased scrutiny is being placed on the accessibility of articles and data.

Articles from the 2020-2021 annual reports were reviewed to identify:

- GWF acknowledgement (per CFREF guidelines)
- Data availability statement inclusion
- Open Access



While we will work to analyze previous years to understand GWF progress, the DM finds these results encouraging.

Moving forward, the Data Management team will work with the SMC and projects to identify metrics to help us understand our progress in the Data Management Goals

DM Activities: Overview

Education

- Webinars
- Presentations
- DM Wiki
- Project consultation

Storage and Access

- Institutional
- Compute Canada
- Copernicus
- WISKI
- Cuizinart
- GWFNet metadata catalogue

Sharing and Preservation

- Repositories
 - FRDR, Dataverse, DataStream, Zenodo...
- Data Publication
- DOIs

DM Highlights: Paper on DM best practices:

Ten best practices to strengthen stewardship and sharing of water science data in Canada. Hydrological Processes

- The need to engage in data management in the context of Canadian water science research.
- Insights and practical advice for water science researchers on how to integrate data management best practices and tools into their research



Persaud, B. D., Dukacz, K. A., Saha, G. C., Peterson, A., Moradi, L., O'Hearn, S., Clary, E., Mai, J., Steeleworthy, M., Venkiteswaran, J. J., Kheyrollah Pour, H., Wolfe, B. B., Carey, S. K., Pomeroy, J. W., DeBeer, C. M., Waddington, J. M., Van Cappellen, P., & Lin, J. (2021). Ten best practices to strengthen stewardship and sharing of water science data in Canada. *Hydrological Processes*, 35(11), e14385. <https://doi.org/10.1002/hyp.14385>

DM Highlights: Updated DM Wiki

Data Management Resources for GWF Researchers

Created by Moradi, Laleh, last modified by Peterson, Amber on Jun 24, 2021



Learn about:

- The [GWF Data Management Core Team](#)
- The [services that our team provides](#)
- How to long-term preserve your data through [dataset publication](#)

[Contact us.](#) We would look forward to hearing from you. Share your questions, ideas, concerns, and data management successes with us.

Search this documentation

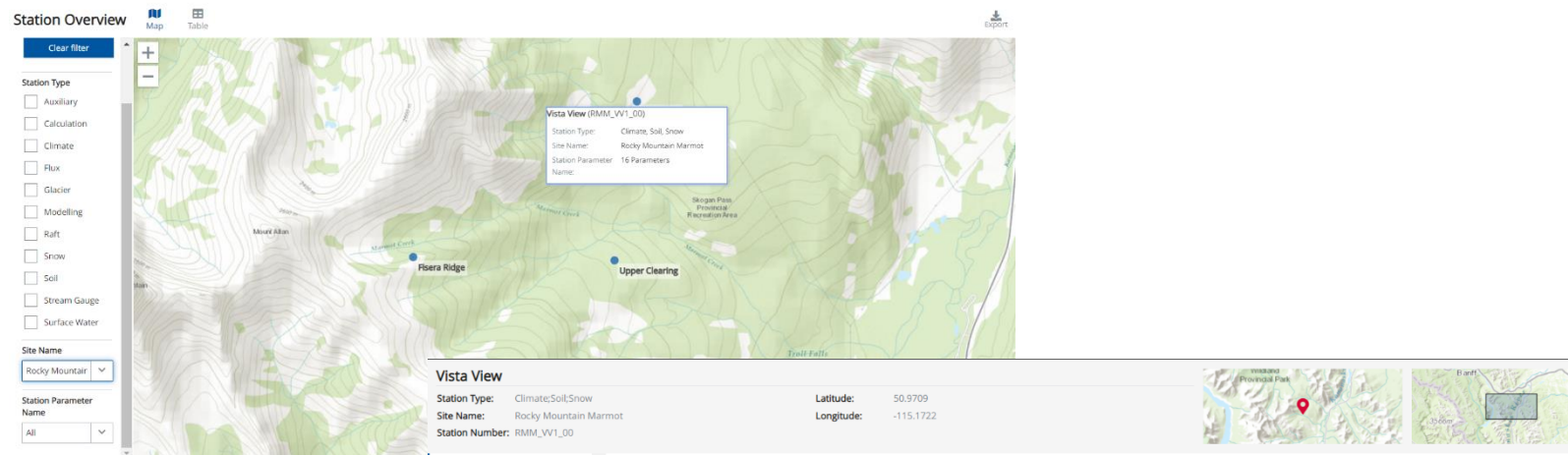


Featured Pages

- [Data Management Presentations](#)
- [Dataset Publication](#)

- How-To documentation on topics such as Dataset Publication (repositories, DOIs, curation, embargos, licensing etc.)
- Access Webinar Recordings on Data Management, Collaboration tools, Compute Canada resources and more!
- Access Data Management Presentations

DM Highlights: New WISKI web portal



Multiple Timeseries

Selection Graph

Search

Station Parameter Name	Time Series Name	Unit Name	From	To
<input checked="" type="checkbox"/> AirTemp_2.62m	04.Cleaned3	degree Celsius	09/01/2005 12:15:00 AM	12/07/2021 4:00:00 AM
<input type="checkbox"/> AirTemp_2.62m	05.Filled	degree Celsius	09/01/2005 12:15:00 AM	12/07/2021 4:00:00 AM
<input type="checkbox"/> IncomingLongwaveRad_ASP110	04.Cleaned3	watt per square meter	09/01/2005 12:15:00 AM	01/13/2006 1:45:00 PM
<input type="checkbox"/> IncomingLongwaveRad_ASP110	05.Filled	watt per square meter		
<input type="checkbox"/> IncomingShortwaveRad_ASP110	04.Cleaned3	watt per square meter		
<input type="checkbox"/> IncomingShortwaveRad_ASP110	05.Filled	watt per square meter		
<input type="checkbox"/> NetRad_1.71mNRL	04.Cleaned3	watt per square meter		
<input type="checkbox"/> NetRad_1.71mNRL	05.Filled	watt per square meter		
<input type="checkbox"/> OutgoingLongwaveRad_ASP110	04.Cleaned3	watt per square meter		
<input type="checkbox"/> OutgoingLongwaveRad_ASP110	05.Filled	watt per square meter		
<input type="checkbox"/> OutgoingShortwaveRad_ASP110	04.Cleaned3	watt per square meter		
<input type="checkbox"/> OutgoingShortwaveRad_ASP110	05.Filled	watt per square meter		
<input type="checkbox"/> Precip_TB4				
<input type="checkbox"/> RelHum_2.62m				
<input type="checkbox"/> SnowDepth_1.59m				
<input type="checkbox"/> SoilHeatFluxAtPlate				
<input type="checkbox"/> SoilTemp_05cm				
<input type="checkbox"/> SoilTemp_10cm				
<input type="checkbox"/> SoilTemp_20cm				
<input type="checkbox"/> WindDir_4.11mRMYoung				



Enhanced:

- Mapping
- Data access
- Visualization

DM Highlights: External Participation



Smart Great Lakes Initiative (GLOS)

(Bhaleka Persaud (Leadership Committee); Krysha Dukacz (Steering Committee, Data & Information WG))



Covidex AI enhanced search

AI enhanced search of Covid-19 literature created by Jimmy Lin's group at the University of Waterloo

A prototype search engine was build and available at: <https://covidex.ai>

Edwin Zhang, Nikhil Gupta, Raphael Tang, Xiao Han, Ronak Pradeep, Kuang Lu, Yue Zhang, Rodrigo Nogueira, Kyunghyun Cho, Hui Fang, and Jimmy Lin.

[Covidex: Neural Ranking Models and Keyword Search Infrastructure for the COVID-19 Open Research Dataset.](#) *Proceedings of the 1st Workshop on Scholarly Document Processing*, pages 31-41, November 2020.

DM Highlights: Modelling Dashboard – In Progress

Dashboard project with Core Modelling

Dashboards of research basins with long-term monitoring programs

Dashboards show basin characteristics and summarize the available data

Basin Name	Marmot Creek Research Basin
Location	Canadian Rocky Mountains
Size	9.4 km ²
Elevation	1590 – 2829 m
Description	Alpine-montane forest headwater catchment
Part of	Bow River Basin, Saskatchewan River Basin
Years of Data	1962-1987, 2005-present

Highlights

- High concentration of monitoring stations; decades of data
- Suitable for examining the effects of forest harvesting and climate change
- High resolution elevation data (Lidar)

Shortfalls

- 20 year gap between the historical and recent data
- No evapotranspiration data

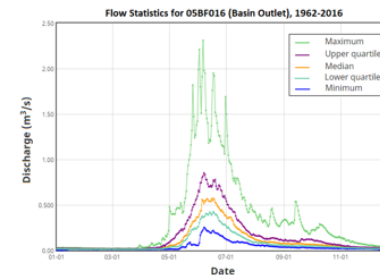
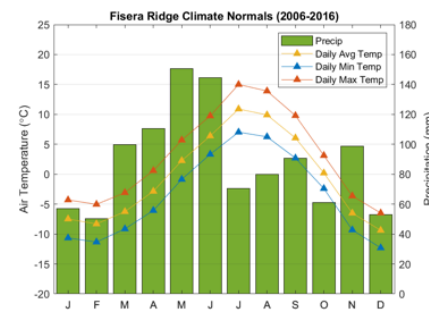
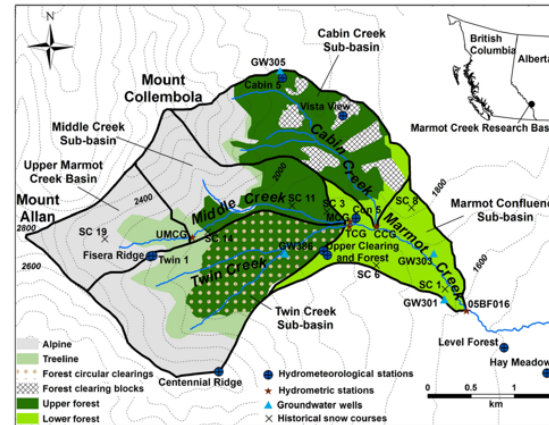
Geospatial Data	Notes
Elevation	Resampled 2007 Lidar 8m DEM (Hopkinson et al., 2012)
Landcover	Map from Alberta Forestry Service (1963) with recent updates from site visits
River Network	Delineated from terrain processing using Lidar DEM
Basin Delineation	Delineated from terrain processing using Lidar DEM

Meteorological Input (Gap-filled Observations)

Stations	Coverage	Notes
Cabin 5, Twin 1, Confluence 5	1969-1987	-Hourly air temp, relative humidity, and wind speed. -Daily precipitation amounts
Centennial Ridge (CR)	2005-2016	-Hourly air temp, relative humidity, wind speed, precipitation, incoming shortwave, and soil temperature (depth of 5 cm or 10 cm).
Fisera Ridge (FR)		-No precipitation measurements at CR, LF, UF, VV.
Hay Meadow (HM)		-No incoming shortwave measurements at VV
Level Forest (LF)		
Upper Clearing (UC)		
Upper Forest (UF)		
Vista View (VV)		

Additional Information

Fang, X., Pomeroy, J. W., DeBeer, C. M., Harder, P., and Siemens, E.: Hydrometeorological data from Marmot Creek Research Basin, Canadian Rockies, Earth Syst. Sci. Data, 11, 455–471, <https://doi.org/10.5194/essd-11-455-2019>, 2019.



Validation Data			
Type	Stations	Coverage	Notes
Soil Moisture	HM	2005-2007	-15 min volumetric water content at 15 cm
Soil Moisture	UF	2005-2013 2013-2016	-Daily volumetric water content at 25 cm -15 min volumetric water content at 25 cm
Soil Moisture	LF	2005-2016	-Daily volumetric water content at 25 cm
Snow (Point)	HM, LF, UC, UF, VV, FR, CR	2005-2016	-15 min snow depth measurements from SR50
Snow (Area)	HM, LF, UC, UF, VV, FR	2007-2016	-Monthly measurements during accumulation; higher frequency during melt -No snow data at HM in 2007 -Snow depth, density, and water equivalent measured -Snow depth measured at 5m intervals along transect; density measured every 5 depth measurements
Snow (Area)	SC: 1, 3, 4, 8, 11, 14, 19	1963-1986	-Monthly measurements Feb – June -Average snow depth and water equivalent from 10 staked points along each snow course
Streamflow	05BF016 (Outlet)	1962-2016	-Daily flow (1962-2016), hourly flow (2013-2016)
Streamflow	05BF017 (MCG), 05BF018 (TCG), 05BF019 (CCG)	1963-1986 2007-2012	-Daily flow (1963-1986), hourly flow (2007-2012) -Gauges destroyed in June 2013 flood
Streamflow	05BF020 (UMCG)	1964-1986 2007-2016	-Daily flow (1964-1986), hourly flow (2007-2016)
Groundwater	301, 303, 305, 386	1965-2020	-Daily groundwater level -386 installed in 1989 -All stations inactive 1997-2006 -303 & 305 inactive 1975-1981

Download the Data

Daily Streamflow from WSC: [link](#)
Groundwater from Alberta Environment: [link](#)
All other data: [link](#); Details in the [README file](#)
Reference:
Fang, X., Pomeroy, J., DeBeer, C., Harder, P., Siemens, E. (2018) Hydrometeorological data from Marmot Creek Research Basin, Canadian Rockies. Federated Research Data Repository. <https://doi.org/10.20383/101.09>

DM Highlights: Metadata Development GeoNetwork – In Progress

GWFNet catalogue

200+ metadata records

The image shows the left sidebar of the GWFNet catalogue, which contains various search filters. The filters are organized into sections: KEYWORDS (Environment, Canada, Inland waters, Climatology, meteorology, Boundaries), CONTACT FOR THE RESOURCE (University of Saskatchewan, Canadian Cryospheric Information Network, Wilfrid Laurier University, University of Waterloo, University of Montreal), PROVIDED BY (GWF Dataset Catalogue), YEARS (2021, 2020), FORMATS (Digital File (CSV), Digital File, Digital File (text/plain), Digital Files), REPRESENTATION TYPES (Vector, Grid, Text, table), UPDATE FREQUENCIES (As needed, Not planned, Irregular, Fortnightly, Annually), STATUS (Completed, On going, Under development, Planned, Not obsolete), and SCALE (1, 0.1).

The image shows the search results for the dataset 'A Bias-Corrected 3-hourly 0.125 Gridded Meteorological Forcing Data Set (1979 – 2016) for Land Surface Modeling in North America'. The results are displayed in a list format, with the first result highlighted. The highlighted result includes a title, a description, a link to the dataset, and a 'Download and links' section. The description states: 'Cold regions hydrology is very sensitive to the impacts of climate warming. More physically realistic and sophisticated hydrological models driven by reliable climate forcing can provide the capability to assess hydrologic responses to climate change. However, hydrological processes in cold regions involve complex phase changes and so are very sensitive to small biases in the driving meteorology, particularly temperature and precipitation. Cold regions often have sparse surface observations, particularly at high elevations that generate the major amount of runoff. The effects of mountain topography and high latitudes are not well reflected in the observational record. The best available gridded data in these regions is from the high resolution forecasts of the Global Environmental Multiscale (GEM) atmospheric model and the Canadian Precipitation Analysis (CaPA) reanalysis but this dataset has a short historical record. The EU WATCH ERA-Interim reanalysis (WFDEI) has a longer historical record, but has often been found to be biased relative to observations over Canada. The aim of this study, therefore, is to blend the strengths of both datasets (GEM-CaPA and WFDEI) to produce a less-biased long record product (WFDEI-GEM-CaPA). First, a multivariate generalization of the quantile mapping technique was implemented to bias-correct WFDEI against GEM-CaPA at 3h 0.125° resolution during the 2005-2016 period, followed by a hindcast of WFDEI-GEM-CaPA from 1979. The variables (units) available are specific humidity, precipitation (kg m-2 s-1), surface air pressure (Pa), Surface Downwelling Longwave Radiation (W m-2), surface_downwelling_shortwave_flux_in_air (W m-2), Surface Downwelling Shortwave Radiation (W m-2), wind speed (m/s) and Air Temperature (K). Note that this dataset was created using a 365-day calendar, hence leap years have a missing day. These data are in NetCDF format and can be downloaded via the Cuizart Platform (<http://cuizart.io>) by selecting dataset labelled wfdei-gem-capa.

The image shows the GWF Cuizart platform interface. The platform is a web-based tool for accessing and processing geospatial data. It features a search bar, a map view, and a list of datasets. The 'wfdei-gem-capa' dataset is selected, and its metadata is displayed. The metadata includes the title, date range (1979/01/01 – 2017/01/01), variables (Select variables...), and a 'PROCESS' button. The platform also includes a 'Download and links' section and an 'About this resource' section. The interface is designed to be user-friendly and accessible, with clear navigation and search options.

GWFNet catalogue

-
- The image is a collage of screenshots from the GWFNet website, illustrating various data portals and project information. The central focus is the 'CORE: Core Modelling and Forecasting Project' page, which lists datasets like 'Aggregated gridded bedrock depth dataset' and 'Aggregated gridded soil texture dataset'. Other visible pages include 'Section 1: Project Information' with a list of project websites and participants, and 'Section 2: Research Site' showing temporal and geographic bounding boxes. A green arrow points from the 'CORE: Core Modelling and Forecasting Project' page to the 'Section 2: Research Site' page.

The Year Ahead: Next Steps



Build the GWF Legacy

- Continue to build the GWF metadata catalogue with related metadata and publication information
- Expand review of GWF publications to all available and develop metrics to measure DM progress
- Enhance location data for projects
- Work with projects to publish data with articles or as standalone datasets.
- Extend GWFNet capabilities to enable reporting and enable direct content creation
- Potential link to AI

The Year Ahead: How You Can Help!

- Review content of the GWF metadata catalogue
- Contribute metadata to the GWF metadata catalogue
- Publish data as appropriate with journal articles or as stand alone datasets
- Provide feedback on GWFNet to help us serve your needs better
- Let us know how we can help you.