

DataStream

An open access hub for sharing water data



Carolyn DuBois

Patrick LeClair

The Gordon Foundation

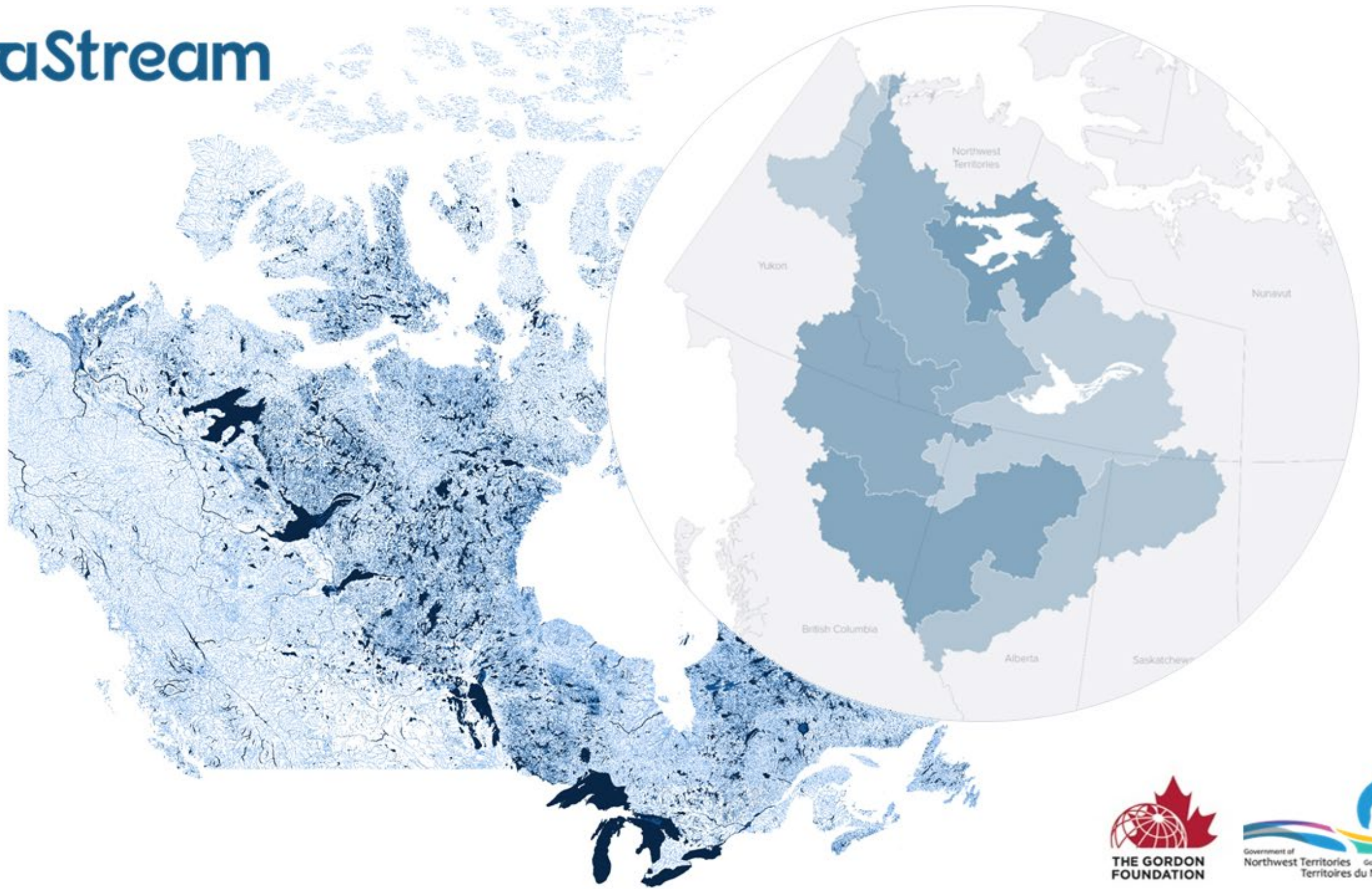
Our Common Water Future

November 18, 2020

Overview

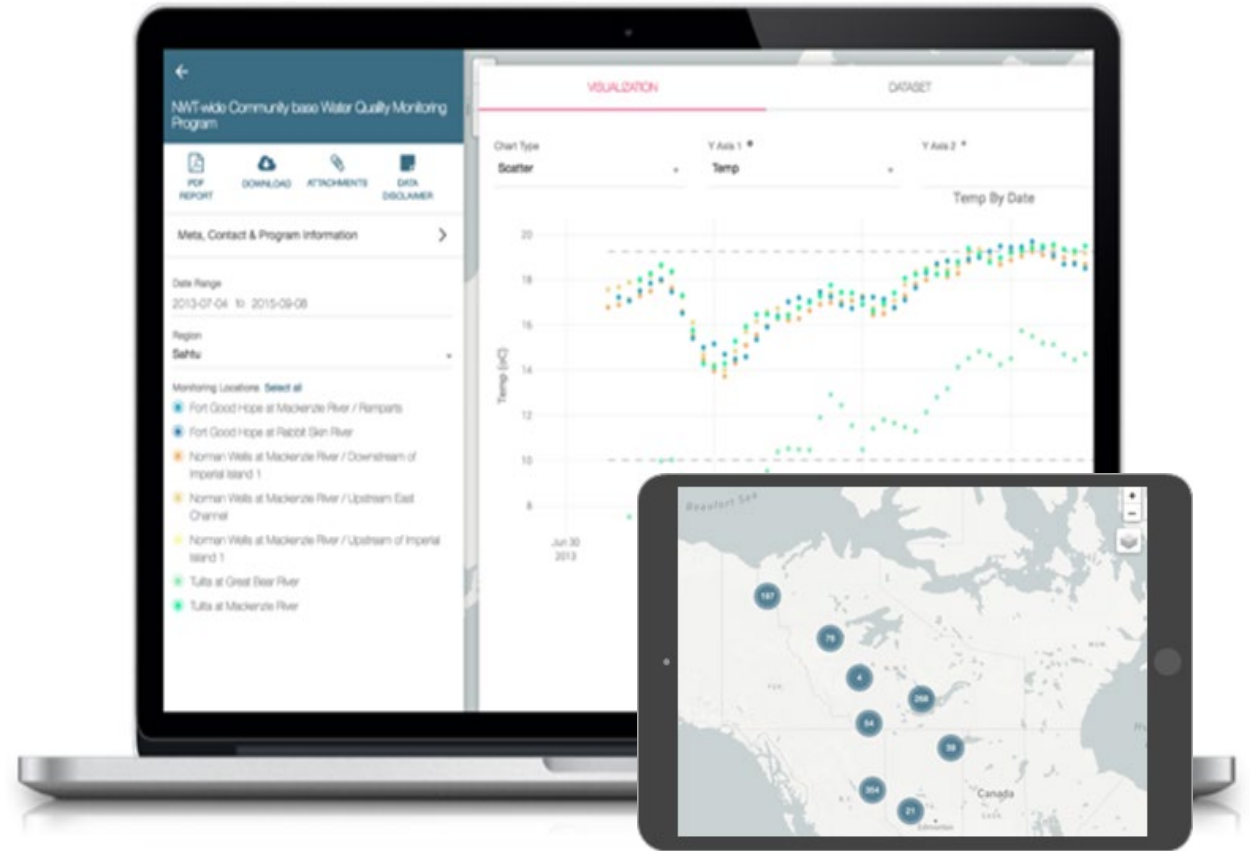
- Background: DataStream's origin story
- Demo
- Commitment to FAIR Principles
- What's next?
- Questions

DataStream



DataStream

- Open access site for **water quality data**
- Brings data together at basin level
- Free and open for anyone to use
- Minimizes barriers to accessing/sharing data

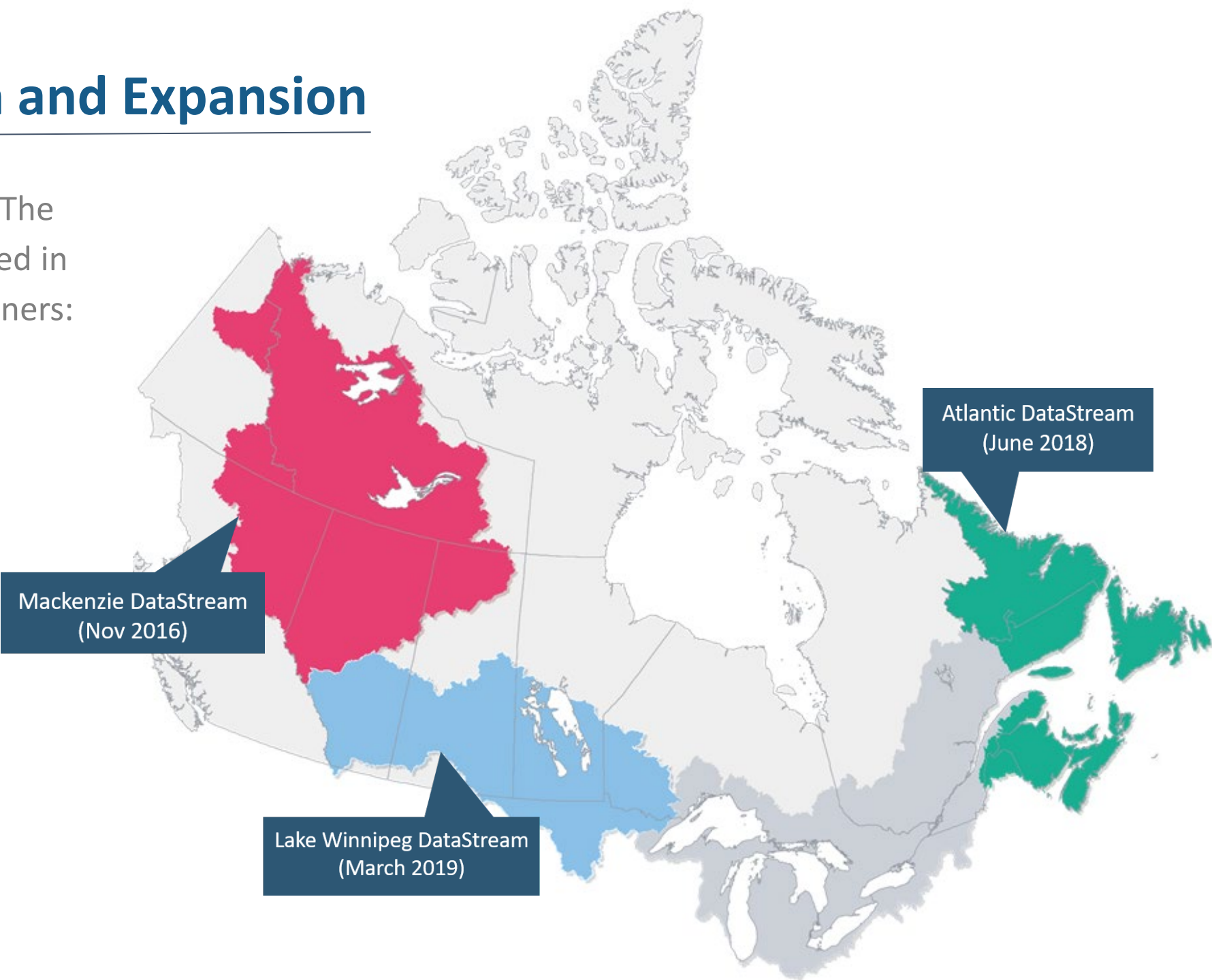


DataStream Growth and Expansion

DataStream is led nationally by The Gordon Foundation and delivered in collaboration with regional partners:

- Government of Northwest Territories
- Atlantic Water Network
- Lake Winnipeg Foundation

Data is contributed by over **100 different groups** across all three regions
9+ million unique obs.



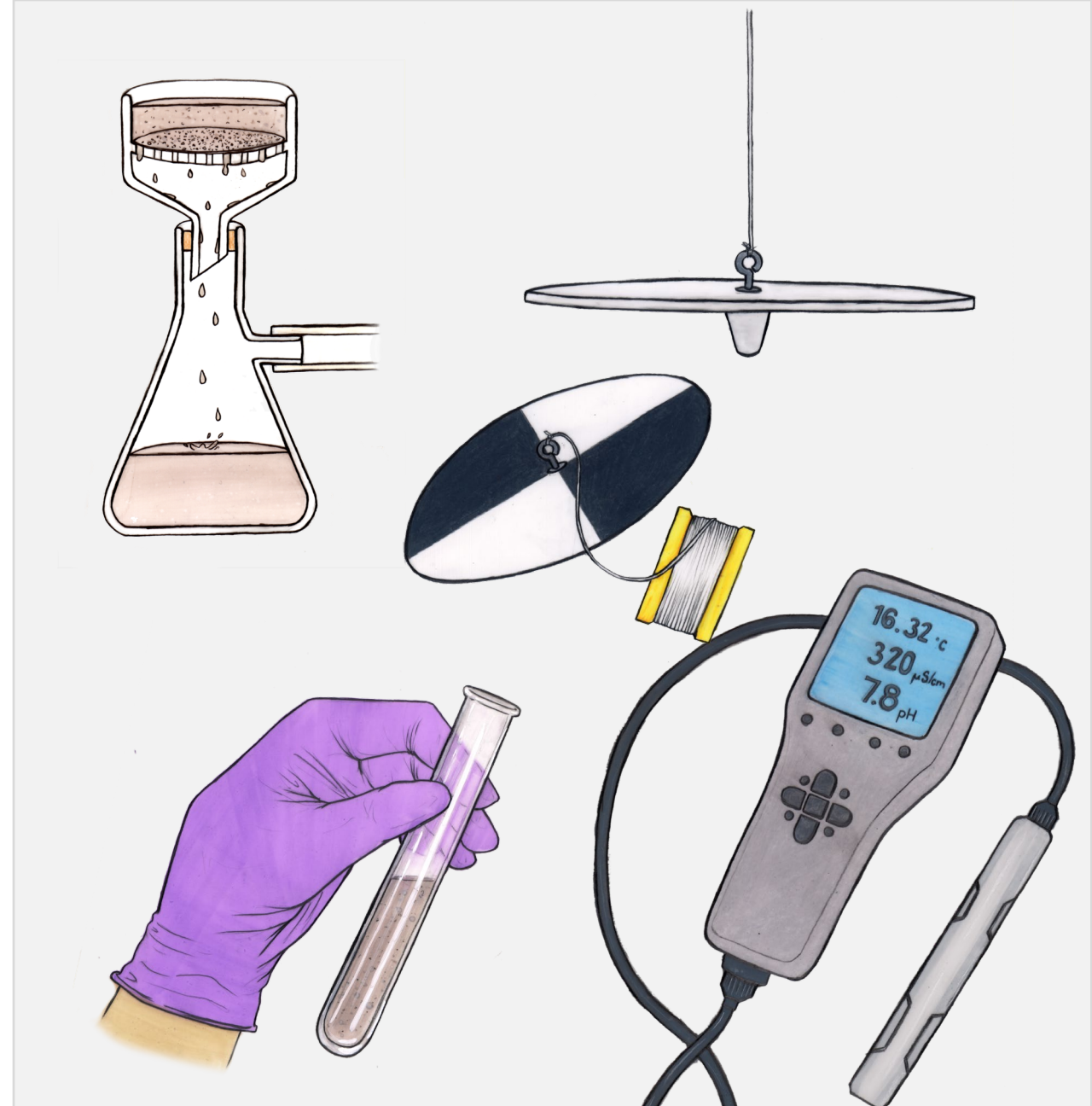


Who Contributes?

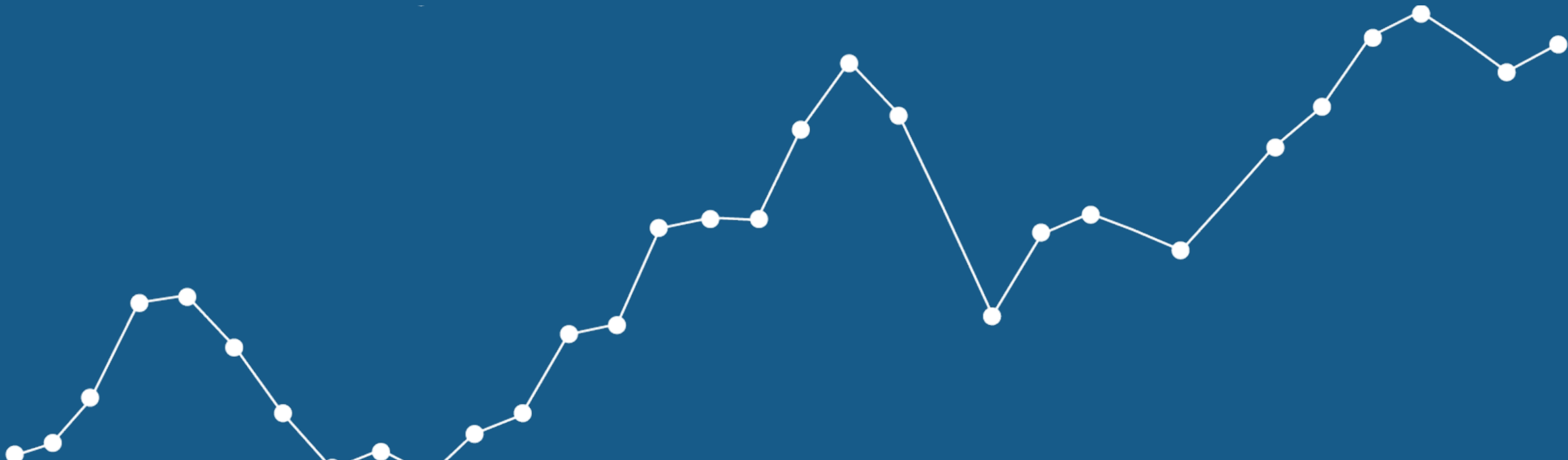
- Community groups
- Watershed organizations
- Academics
- Indigenous governments
- Provincial/Territorial Governments
- Federal Government

What Do They Contribute?

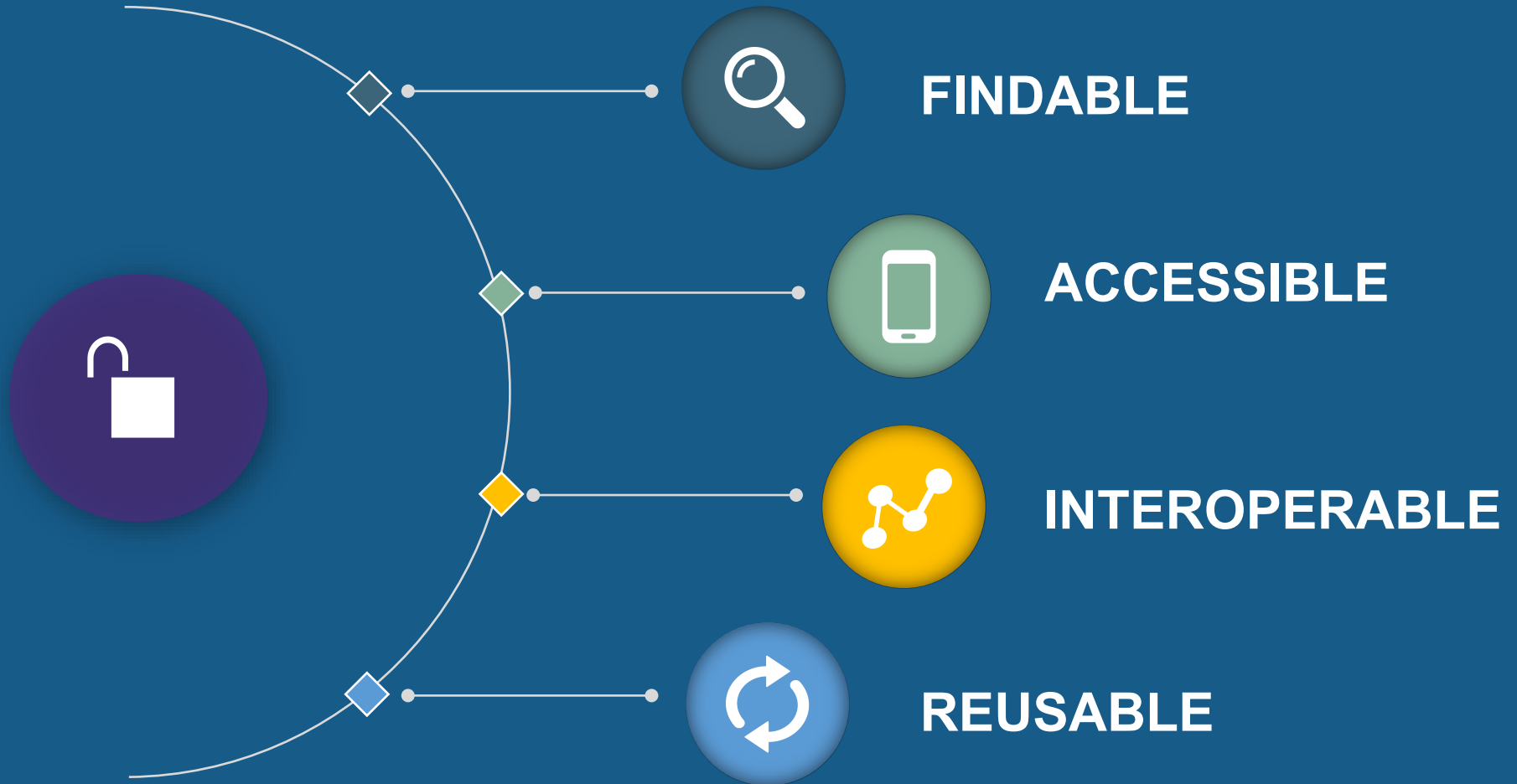
- Over 6,000 parameters possible
 - 200+ parameters currently stored
- Media Types:
 - Water Quality
 - Sediment
- Lab Data
 - Nutrients, metals, etc.
- Field Data
 - pH, temperature, DO, etc.



System Demo



'FAIR' Principles of Open Data



Secondary Use of Water Quality Data Study

- **Goal**
 - *Determine what challenges were faced when using water quality data for secondary use*
 - *'Secondary Use': "the use of data beyond the original intent determined by the organization that collected the data"*
- **Results**
 - USD\$12 billion determined ambiguous
 - USD\$8.2 billion unambiguous

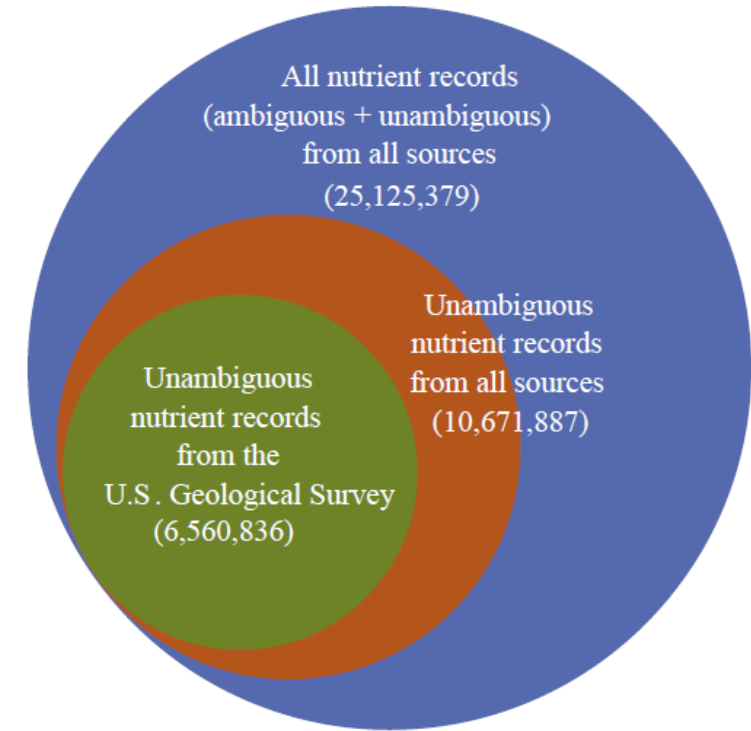


Fig. 3. Total number of nutrient records (unambiguous plus ambiguous) from all sources, number of unambiguous nutrient records from all sources, and number of unambiguous nutrient records from the U.S. Geological Survey.

FINDABLE



- Easy for both humans and computers to find the data

DataStream Features

- Map-based Search Tool and Visualizations
- Robust Machine-Readable metadata
 - Cross-sharing metadata (Google Dataset, FRDR)
- Digital Objects Identifiers (DOIs)
- Search Engine Optimization (SEO)

ACCESSIBLE

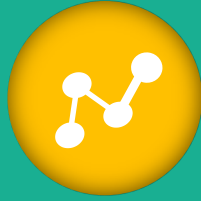


- User can easily access the required data

DataStream Features

- No Log-in Required and Free to Use
- Clear and Easy User Interface (UI)

INTEROPERABLE

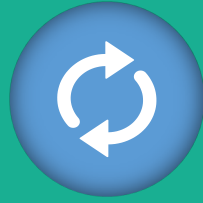


- Data can be integrated with other data and into existing applications or workflows

DataStream Features

- API
 - Allows for direct query to database
 - Ability to download large amounts of data programmatically
 - Uses OData Standard
 - Documentation shared via GitHub
- ETL
- Data Provided in .csv Format

REUSABLE



- Data can be replicated and or combined in different settings

DataStream Features

- Use Modified USEPA WQX Data Standard
 - Required data columns
 - Validation to ensure data follows standard
 - Data standard open on GitHub
- Data Licensing
 - Use open data license
- Data versioning

FAIR: Beyond The Platform

- Science Literacy

- Science Explainers

- Data Literacy

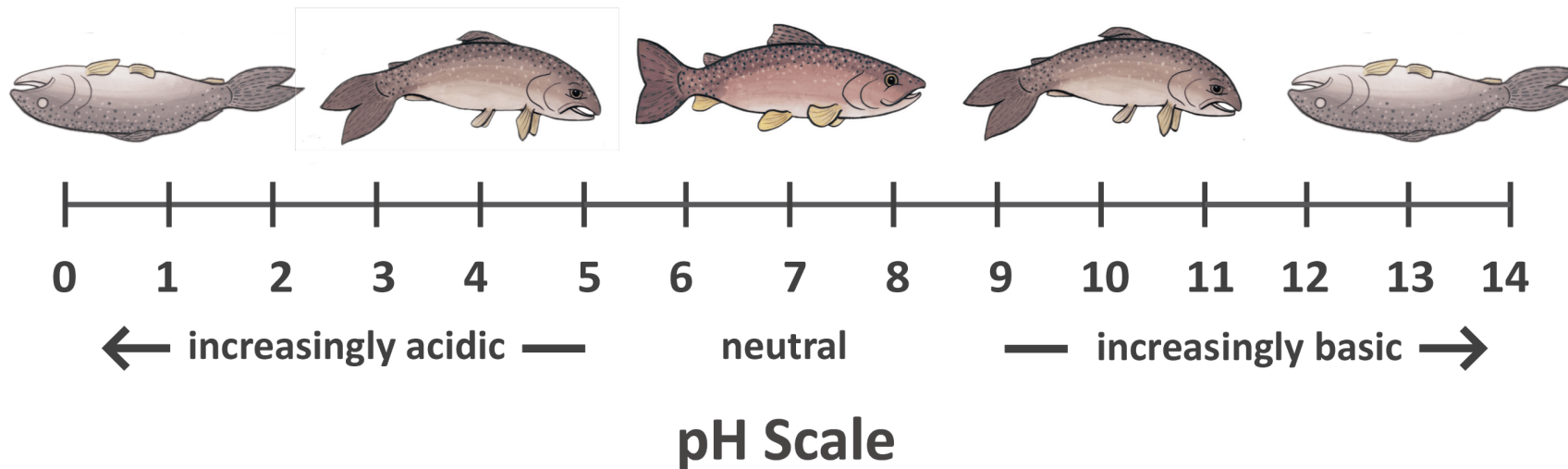
- Dive into Data Webinar Series

- Open Lines of Communication

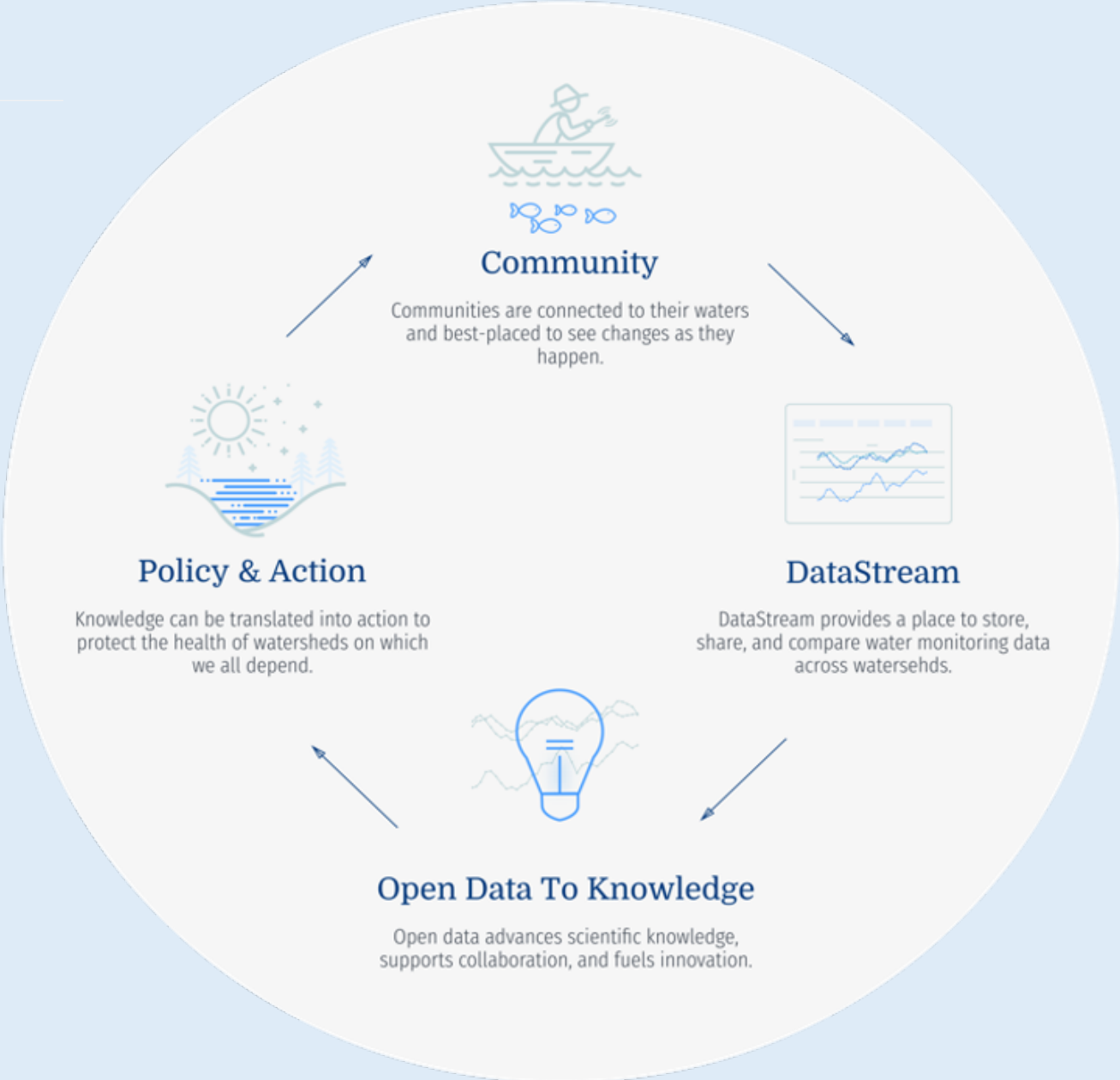
- Open Office Hours
- Fall-O-Up

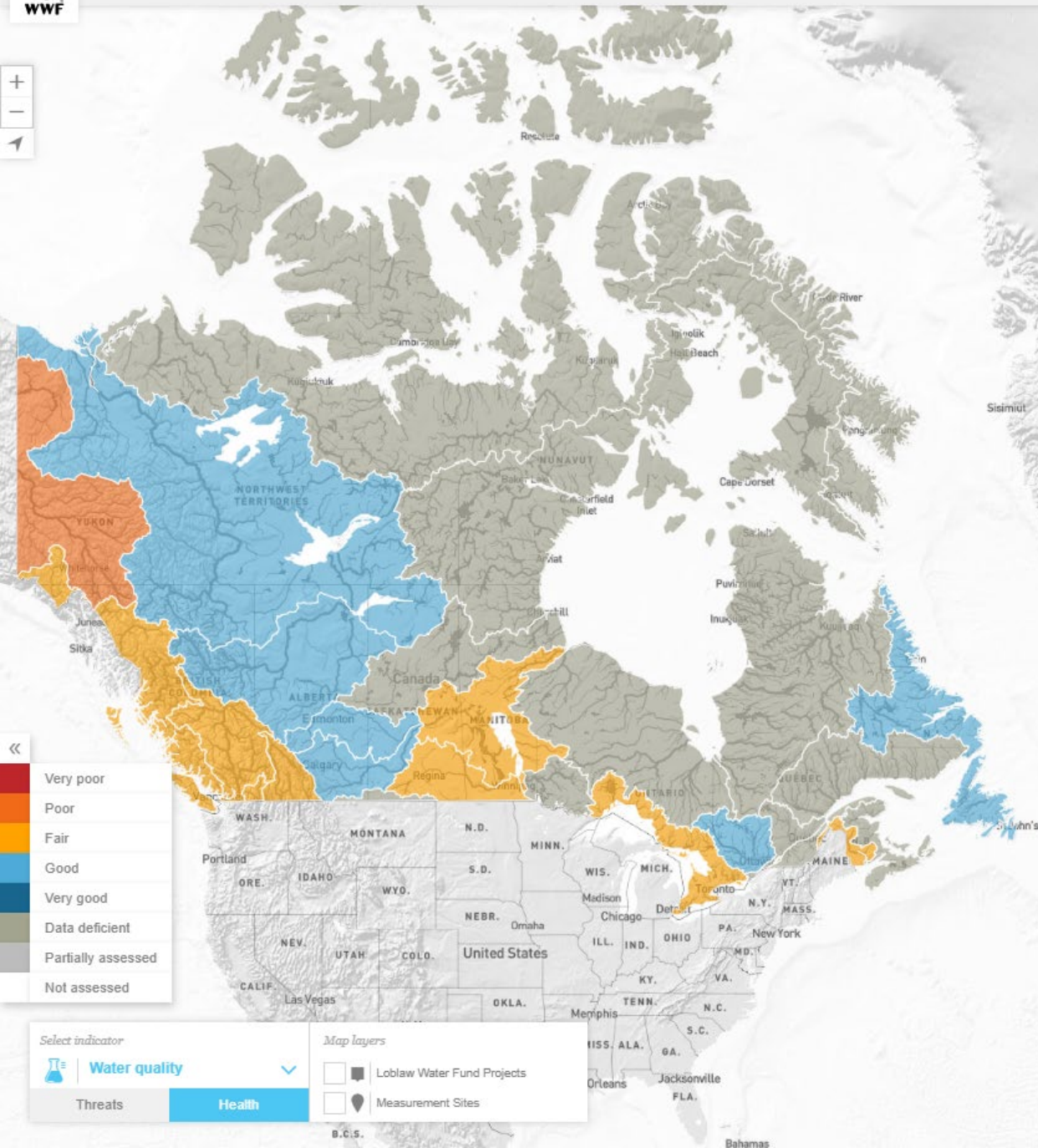
- Promoting Public Access

- Encouraging Use of Common Standards



Where DataStream fits in





Key Takeaways

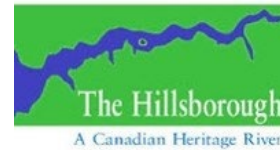
- Use *existing* standards whenever possible (e.g. WQX)
- Share data on open repositories (like DataStream)
- Get (and stay) in touch!



Partners



Sponsors/Contributors



Environment and
Climate Change Canada

Thank you



DataStream.org

- MackenzieDataStream.ca
- [Atlantic DataStream.ca](https://AtlanticDataStream.ca)
- LakeWinnipegDataStream.ca



@DataStreamH20



@DataStreamH20



Accessed through DataStream.org/webinars



datastream@gordonfn.org

Subscribe to our newsletter: bit.ly/TGFnewsletter

Data Schema

Dataset level metadata

Observation level (within dataset)

- Aligned with WQX
- Validation on upload



The screenshot shows a mobile application interface for the "LSWC tributary monitoring program". At the top, there is a dark blue header with a back arrow and the program name. Below the header are three icons: a cloud with a download arrow, a paperclip, and a document with a checkmark, labeled "DOWNLOAD", "ATTACHMENTS", and "DATA LICENSE" respectively. The main content area is titled "Dataset Information" and lists the following details: Dataset Name (LSWC tributary monitoring program), Data Collection Organization (The Lesser Slave Watershed Council), Data Steward Email (info@lswc.ca), Start Date (April 25, 2017), End Date (May 23, 2019), and DOI (10.25976/vmet-ct64). At the bottom, there is an "Abstract" section with a dropdown arrow, showing the beginning of a description: "The Lesser Slave Watershed Council Tributary Monitoring Program collects water quality data for the five major tributaries of Lesser Slave Lake at upstream, mid and". A red circular icon with a white book symbol is visible in the bottom right corner.

Data Schema

Dataset level metadata

Observation level (within dataset)

- Aligned with WQX
- Validation on upload



Monitoring location

Site name, ID
Coordinates
Reference datum
Waterbody type

Sample collection

Sample type (e.g. field, lab, blank)
Date, time
Collection equipment
Depth

Parameter/result

Characteristic name, speciation, sample fraction
Result value, units
Analysis method, date, time
Detection limits