

Smart Great Lakes Initiative’s Data and Information Survey Results

The Smart Great Lakes Initiative (SGLi) developed the survey to understand how Great Lakes data are discovered, accessed, and shared. The survey was sent to participants through SGLi partners by email and advertised through GLIN Announce. The 42 responses to the survey were accepted between August 2, 2023 and September 29, 2023.

Of the 42 respondents, 19 or 45.24% were from the government, 9 or 21.43% were from non-profit organizations, 6 or 14.29% were from academia, 2 or 4.76% were community members or citizen scientists, and 1 or 2.38% was an individual. There were zero respondents from Indigenous/First Nations/Metis/Tribal organizations and commercial/private companies filling out the survey. Five or 11.9% chose “other” as their sector. Those who chose “other” included a school teacher, a trade organization representative, Cooperative Institute of Great Lakes Research staff, and persons from two Conservation Authorities.

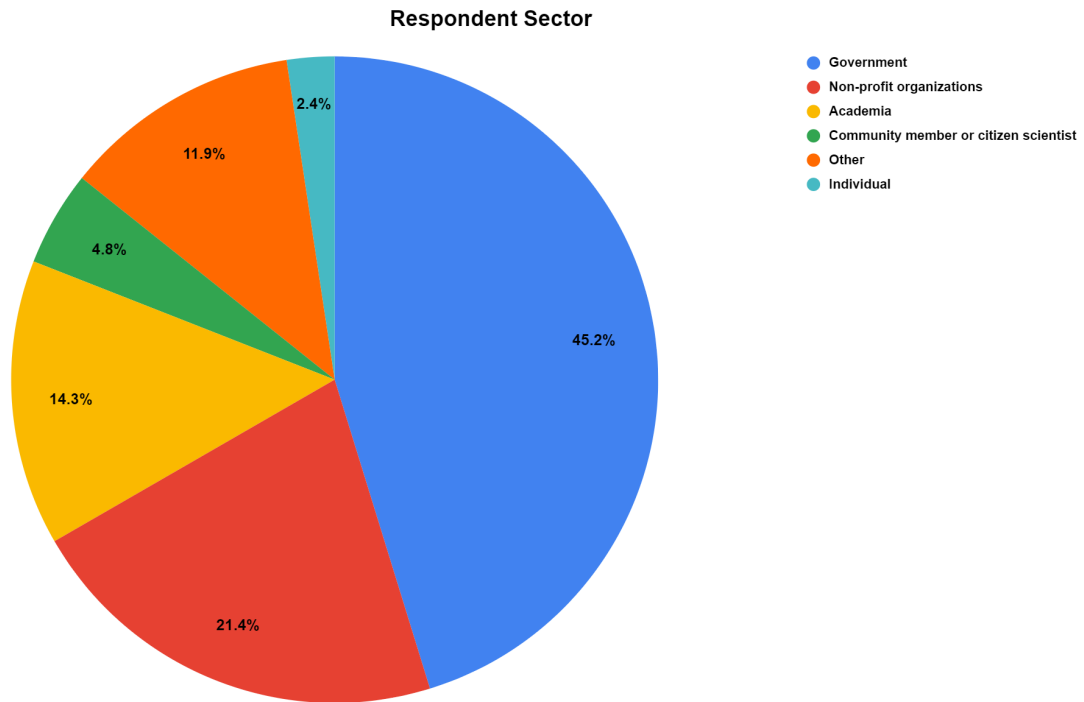


Figure 1. Respondents by sector

Respondents were asked about their geographic interest. The number of geographies is shown in Table 1 and Figure 2 below.

Area	Count
All Great Lakes/Watershed	9
Lake Erie	11
Lake Huron	5
Lake Michigan	10
Lake Ontario	3
Lake Superior	3
North Bay/Algonquin region	1

Table 1. Number of geographic areas selected

Geographic Areas of Interest

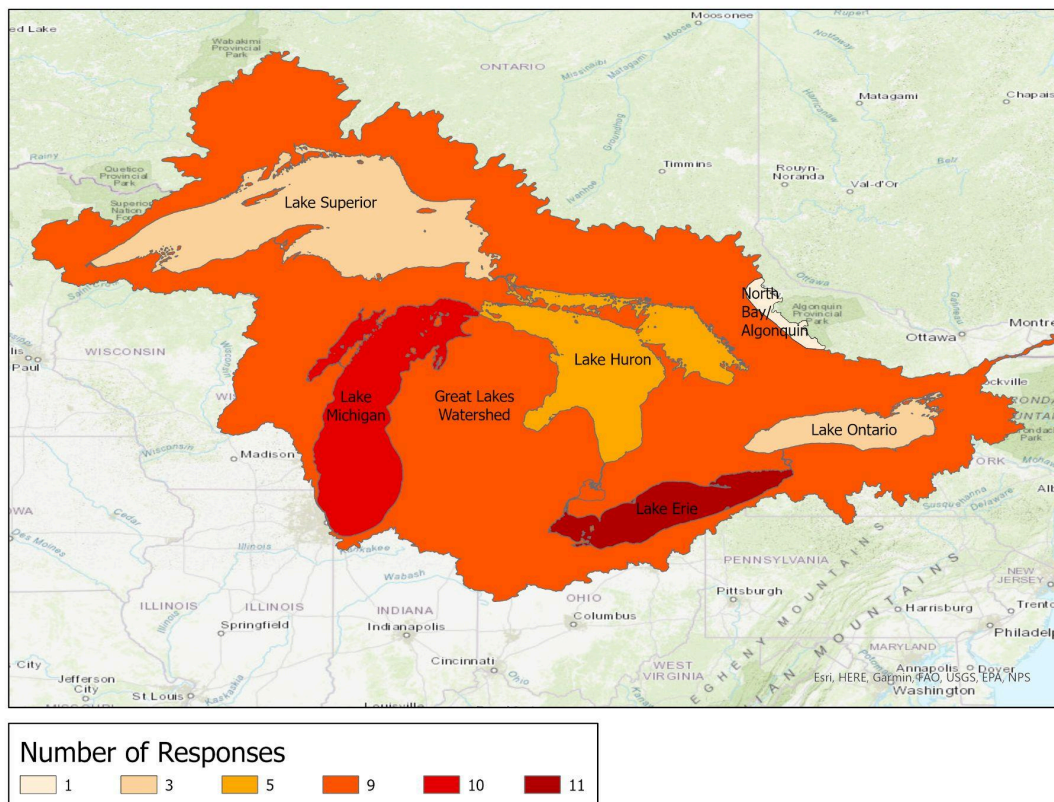


Figure 2. Map of geographic areas of interest

Respondents were then asked what questions are you trying to answer using Great Lakes data and what types of data are of interest to you. Responses ranged from interest in water quality generally to specific needs at a local site. Data types were extrapolated from the response and categorized into 15 groups and subgroups. The most frequent response was for

climate/meteorological data with 21 responses. Water quantity had 14 responses. Water quality in general was the next highest with 10 responses.

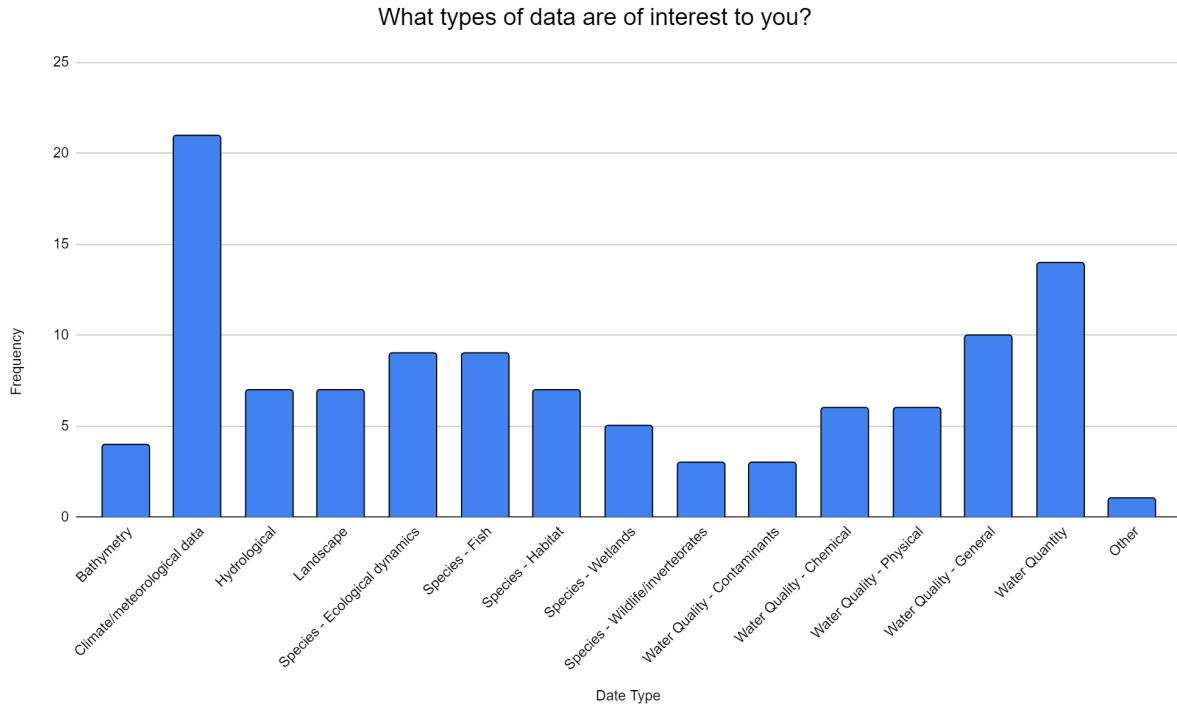


Figure 3. Data types of interest

The survey then asked if participants collected primary data. Of the 42 respondents, 31 or 73.81% collected primary data and 11 or 26.19% did not. The distribution is as follows:

Sector	Collects Data/Total	Collects Primary Data	Does Not Collect Primary Data
Government	15/19	79%	21%
Non-Profit Organization	7/9	78%	22%
Academia	5/6	83%	17%
Community Member or Citizen Scientist	1/2	50%	50%
Individual	0/1	0%	100%
Other	3/5	60%	40%

Table 2. Distribution of respondents who collect primary data by sector

For those who responded that they collected their own data, the survey asked what data they collect. The answers were put into the same broader categories for the question above on what types of data are of interest. Fish data was the most frequent response with 23 responses. Climate/meteorological data was the next frequent with 18 responses. Water quality - physical followed with 16 responses.

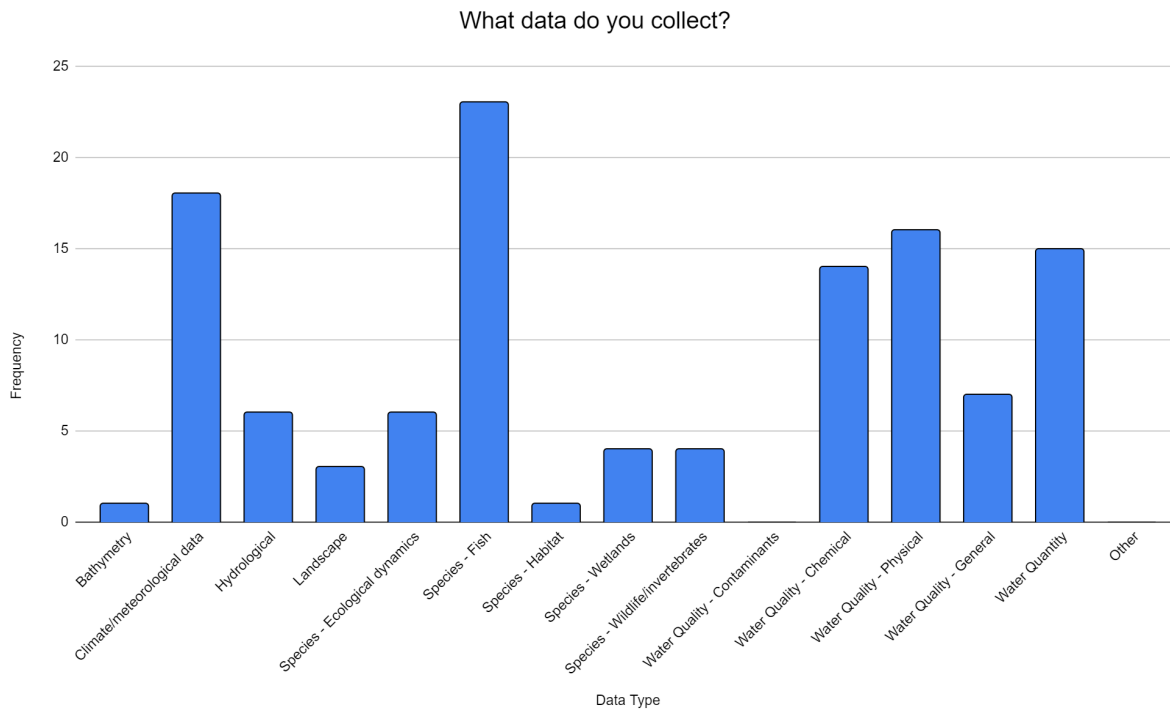


Figure 4. Types of primary data collected

Respondents were asked why they collect their own data, if they are required to collect data per regulations and if the data they require is not available. Since this was an open ended question, answers varied with the most common responses being organization mandates, data is not available and program support.

Reasons for Collecting Own Data	Count
Organizational mandate	15
Data is not available	11
Program support	8
Specific projects/Research	5
Control/Reliability	3
Scale/Higher resolution	3
Temporal specific	2

Completeness	1
Fun	1
Location specific	1
Metadata	1

Table 3. Reasons for collecting own data

Of the 31 respondents who collect primary data, 28 or 90.32% do share the data whereas 3 or 9.68% said they did not share the data.

Respondents were then asked to describe how data is shared if they are sharing it. For example, where it is shared and how it is accessed. Twenty-seven responses were received for this question. The responses were categorized using the following definitions:

Public website - Download: Data is available for download from a public website.

Public website - Download (subset): Parts of the data can be downloaded via a public website.

Public website - View: The data is available to view on a public website but is not downloadable.

Shared via reports: Data is contained in reports that are available.

Shared by request: Data must be requested via email or other means.

Shared with specific partners: Data is shared only with select research partners/stakeholders.

Data available by link to source: Third party data that is used is made available via link to original data source.

The distribution of data sharing categories is in Table 4.

Primary Sharing	Secondary Sharing	Count
Public website - Download		6
Public website - Download	Shared via reports	2
Public website - Download (subset)	Shared by request	3
Public website - Download (subset)	Shared via reports	2
Public website - View	Shared by request	3
Public website - View	Data available by link to source	1
Public website - View		1
Shared by request		2
Shared via reports	Shared with specific partners	1
Shared via reports		1
Shared with specific partners		4

Shared with specific partners	Shared by request	1
-------------------------------	-------------------	---

Table 4. Frequency of selected data sharing categories

The survey asked what the barriers or challenges are to sharing the data you collect? Three respondents shared specific barriers or challenges to sharing the data they collected which are categorized and listed below:

- Cost Prohibitive: “Time required for initial data auditing, maintenance, and archiving”
- Data Silos: Data is shared internally with project collaborators, but not to outside organizations
- Data Protection/First Right of Use: “Will likely share data after initial publications”
- Ambiguous Steps to Sharing: “No particular place to send it”
- No Known/Authoritative Repository: Central public Great Lakes repository to upload it to.

The survey then asked what secondary (third party) data sources they used. While all 42 respondents indicated that they used secondary data sources, 7 did not use any of the listed secondary sources. The top five sites for accessing secondary data were chosen by over 30% of respondents. These were the National Buoy Data Center (NDBC/NOAA), 47.62%, Great Lakes Dashboard (NOAA-GLERL), 47.62%, Great Lakes Coastal Forecasting System (NOAA-GLERL), 38.1%, Great Lakes Bathymetry (NOAA-NCEI), 35.71%, and Government of Canada Open Data (Federal government - Canada), 33.33%.

Secondary Data Sources	Count	Percentage
National Buoy Data Center (NDBC/NOAA)	20	47.62
Great Lakes Dashboard (NOAA-GLERL)	20	47.62
Great Lakes Coastal Forecasting System (NOAA-GLERL)	16	38.1
Great Lakes Bathymetry (NOAA-NCEI)	15	35.71
Government of Canada Open Data (Federal government - Canada)	14	33.33
US Army Corp of Engineers Water Levels (US Army Corps of Engineers)	12	28.57
USGS National Water Dashboard (USGS)	11	26.19
Lake Level Viewer (NOAA Office of Coastal Management)	10	23.81
Ontario Data Catalogue (Province of Ontario)	10	23.81
Great Lakes Aquatic Habitat Framework (Michigan Sea Grant)	8	19.05
Conservation Ontario Hub (Conservation Ontario/Conservation Authorities in Ontario)	8	19.05
National Center for Water Quality Research (Heidelberg University)	7	16.67
Seagull (seagull.glos.org)	7	16.67
Great Lakes Aquatic Nonindigenous Species Information System (GLANSIS) (NOAA-GLERL)	7	16.67

Science in the Great Lakes (SiGL) data viewer (USGS)	6	14.29
Grand River Information Network (Lake Erie watershed) (Grand River Conservation Authority)	5	11.9
National Centers for Environmental Information (NOAA-NCEI)	4	9.52
National Ballast Information Clearinghouse (Smithsonian Environmental Research Center)	3	7.14
Great Lakes Monitoring (IL-IN Sea Grant)	3	7.14
Toronto and Region Real-time Gauging Network (Toronto and Region Conservation Authority)	3	7.14
City of Chicago data viewer (City of Chicago)	1	2.38
Angler Archive (IL-IN Sea Grant)	0	0
Federated Research Data Repository (Canadian Association of Research Libraries & Compute Canada)	0	0
None	7	16.67
Other	12	28.57

Table 5. List of secondary data sources used

In response to choosing secondary data sources, 12 respondents provided additional data sources that they use that were not on the original list. They are listed in Table 6.

Additional secondary data sources
Canadian Wildlife Federation's Canadian Aquatic Barriers Database
Environment and Climate Change Canada (ECCC)
Fisheries and Oceans Canada (DFO) via research collaborations
Great Lakes Commission Blue Accounting
Great Lakes Fisheries Commission Commercial Fish Harvest database
Great Lakes National Program Office database
Illinois State Water Survey data (water suppliers, water supply networks, Illinois Water Inventory Program (IWIP) data)
International Joint Commission
Ministry of Natural Resources and Forestry WISKI portal
NOAA Center for Operational Oceanographic Products and Services (CO-OPS)
NOAA Weather Prediction Center
Ontario GeoHub (LIO)
Ontario Ministry of Natural Resources

USEPA GLENDA
Various individual Conservation Authority open data sites
Water Survey of Canada

Table 6. Suggested secondary data sources used by respondents

Respondents were asked what are the challenges they have in meeting their data needs. Ten respondents indicated one or more challenges in meeting their data needs. Table 7 below summarizes the issue and number reported.

Challenge	Count
The quality of available data is insufficient (Interpreted as lack of metadata/inconsistent units)	5
Required data is pre-digital	4
Data is raw/unprocessed	4
The data I need is not discoverable	4
Data is not updated frequently enough to meet needs	3
Ownership/IP constraints	2
Data is out of date/not collected for the right time period	2
Period of record is insufficient	2
Data is not available in the area of interest	2
The geographic extent does not meet data need	2
Accessible data is not provided at a sufficient level of detail (i.e. watershed level vs. state/province)	1
Clear time/date not specified	1

Table 7. List of challenges to meeting data needs with the number of times chosen by respondents.

Lastly, the respondents were asked what data they are interested in that are unavailable. Responses were categorized in the same manner as previous questions. Climate/meteorological data and water quantity had the most interest with seven counts each. Fish and “other” data followed in interest with five counts each. Bathymetry was also of interest with four counts.

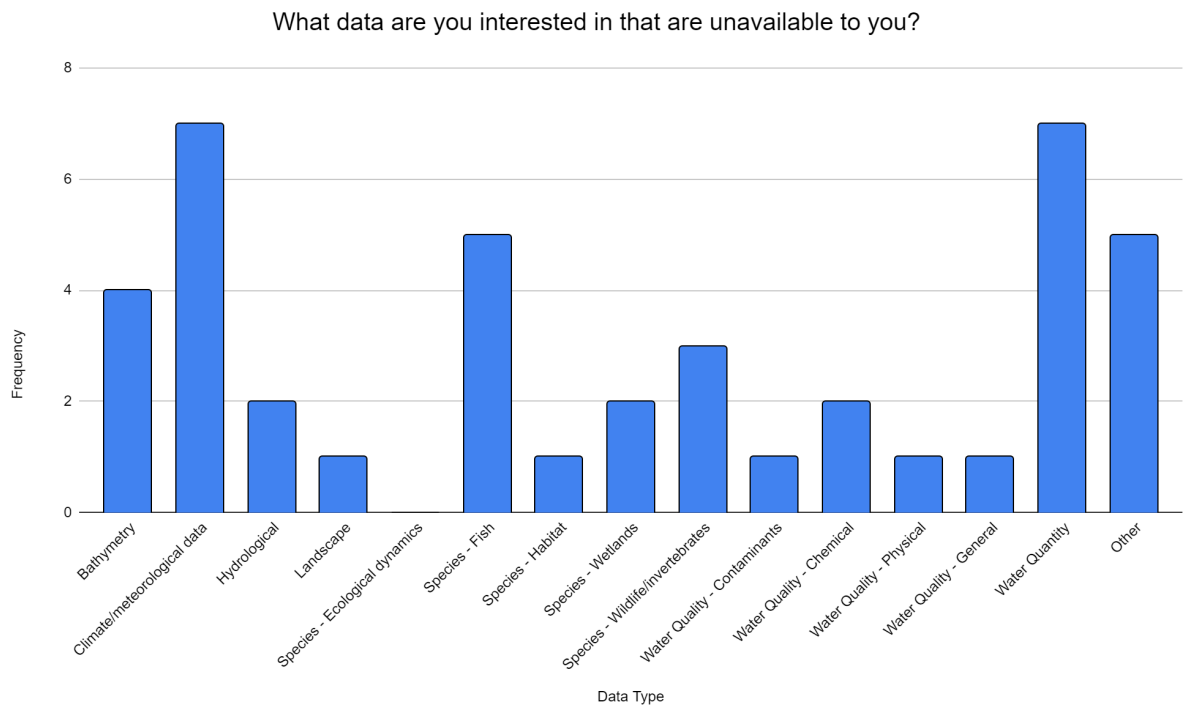


Figure 5. Data of interest that might not be available