Great Lakes Observing System Data Management and Communications
(GLOS-DMAC)

DMAC Background/Overview

The GLOS Data Management and Communications (DMAC) infrastructure supports ingestion, management, dissemination and archiving of "internal" data generated from GLOS-funded observation systems. GLOS DMAC also provides access to "external" data consumed from a variety of entities such as NOAA, MTRI, USGS, and the EPA-USGS Water Quality Portal.

A. Process

GLOS data products integrate data internally managed on the GLOS DMAC with external data harvested and displayed "on-the-fly". The internal datasets are either associated with GLOS partners, making the lifecycle a GLOS responsibility, or cloned for performance reasons, or both. Figure 1 depicts the principal data streams associated with internal GLOS holdings:

- Sensor data entering the GLOS DMAC Observations Database (OBS DB) (1a)
- Sensor data entering the GLOS 52 North SOS database (1b)
- Data entering the GLOS THREDDS collection (2)

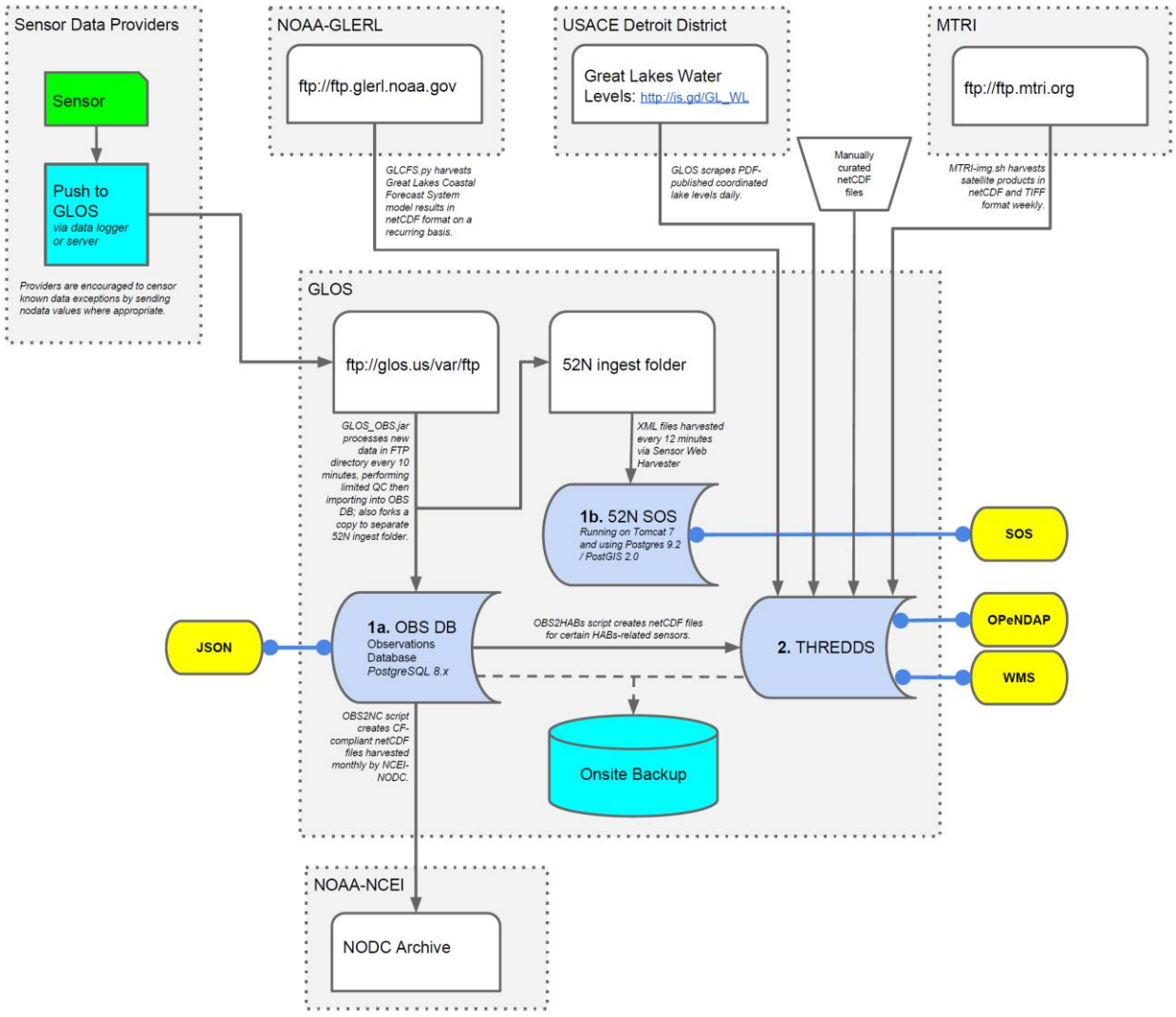


Figure 1. Principal GLOS DMAC data streams.

	1. Non-Federal NRT Sensor Data		2. THREDDS				
	1a. OBS DB	1b. 52N SOS	NOAA-GLERL GLCFS	USACE Lake Levels	HABs Sensors	MTRI Remote Sensing	Manually Curated
Overview	Legacy data store ingesting from buoys, other fixed platforms	IOOS-compliant data store ingesting NRT sensor data	Nowcast and 120-hour forecast models of the Great Lakes	Officially-coordinated water levels for each of the Great Lakes	Subset of NRT sensor data supporting GLOS WLEB HABs Viewer	Satellite products derived from MODIS-AQUA	Various datasets added to the GLOS THREDDS catalog manually
Data types	Air and water temperature, wind, currents, waves, nutrients, chlorophyll, pH, dissolved oxygen, conductivity, phycocyanin and turbidity		Water temperature, waves and currents, and wind and air temperature forcings	Daily coordinated lake level as published by USACE Detroit District	Same as OBS/52N SOS	Water surface temperature, natural color, and color-producing agents	Thermistor data; water quality and meteorological data from vessels of opportunity; water use data; optical properties.
Management	Data providers push to GLOS ftp directory in XML format parsed every 10 minutes, QC'd and injected into PostgreSQL	XML files from data providers copied to 52N ingest folder following QC, then brought into 52N using Sensor Web Harvester	GLOS harvests results and forcings every six hours (nowcast) or twelve hours (forecast) in netCDF format	GLOS scrapes water levels from PDFs published on USACE site and stores in netCDF format	NRT data in XML files for selected sensors are converted to netCDF for viewing within HABs portal	GLOS harvests satellite products from MTRI site in netCDF format	Datasets are added to the GLOS THREDDS catalog manually on an opportunistic basis.
QA/QC	Data providers expected to conform to GLOS Quality Management Plan and provide regular reports on calibration, etc.; automated QC performed by GLOS on incoming data to detect gross errors. QARTOD or similar by 10/31/2017		Federal source	Federal source	Same as OBS/52N SOS	QA/QC information, where available, has been referenced in metadata records. Going forward, GLOS will ensure that QA/QC information be included in or linked from the associated metadata record for all datasets	
Metadata	ISO 19119/19139-compliant metadata compiled by GLOS with assistance from data providers and stored in the GLOS GeoNetwork metadata catalog		ISO 19119/19139-compliant metadata compiled by GLOS with assistance from data providers and stored in the GLOS GeoNetwork metadata catalog				
Backups	OBS DB backed up weekly to external HDD	Not currently backed up (duplicative of OBS DB). Will be backed up weekly by 12/31/2017	Backed up weekly to external HDD				
Archiving	Archived to NCEI on a monthly basis	Not currently archived (duplicative of OBS DB). Will be archived monthly by 12/31/2017	Not archived - datasets are replicates of data maintained by the data providers				GLOS will archive all applicable and acceptable manually-curated data to NCEI
Data Access	JSON	OGC SOS 1.0	Accessible through any TDS service, including OPeNDAP, WMS, WCS, and ncISO				
Notes	The OBS DB will be phased out by December 31, 2017	52N ingestion will use GLOS ftp directory directly for ingest after OBS DB phase-out			The HABs viewer will be reconfigured to use 52N SOS, at which time these netCDF files will no longer be generated		Applicability and acceptability for archiving based on individual assessments in conjunction with NCEI

Table 1. Data Management Summary for Datasets Internally Managed by GLOS DMAC.

GLOS partners also collect data on glider missions, whose data flow is not represented in Figure 1 because (1) the glider data are submitted directly to the IOOS Glider DAC (Data Assembly Center) and (2) access to the glider data is not currently provided by GLOS. GLOS anticipates adding transient access to GLOS partner glider data available in the Glider DAC by December 31, 2016, relying on Glider DAC QC and data services.

Additional external data are currently presented in the GLOS data portal on a transient, on-the-fly basis. These data are accessed from federal sources and therefore do not require QA/QC measures; in addition, no data management, backup, or archiving are required, and data access requests are fulfilled using the sources' data services. The sources for these data are:

- NOAA CO-OPS
- NOAA NDBC
- NOAA NERRS
- NOAA NWS Forecasts and Warnings
- NOAA NWS Level III NEXRAD (accessed unaltered from Weather Underground using the Weather API)
- NOAA OFS
- USCG Marine Reporting
- USGS NWIS
- USGS/ EPA/ NWQMC WQ Portal

Data management processes for the datasets internally managed by the GLOS DMAC are summarized in Table 1, with more detail presented below.

1a. Sensor Data -> GLOS OBS DB

The GLOS Observations Database (OBS DB) is a legacy data store that ingests near real-time sensor data from non-federally owned buoys and other fixed platforms in the Great Lakes.

Data Types. Data ingested into the OBS DB may include air and water temperature, wind, currents, waves, nutrients, chlorophyll, pH, dissolved oxygen, conductivity, phycocyanin and turbidity.

Management. Data from GLOS-funded and other buoys and fixed platforms are pushed to a GLOS ftp directory by the data provider in an agreed XML format. The ftp directory is scanned every ten minutes to identify new data files, which are then parsed to extract the sensor observations. The observations are subjected to basic quality control (QC) tests such as range checks, and then injected into the PostgreSQL OBS DB. Harvesting status is continually updated on the GLOS Buoy and Model Status page (http://data.glos.us/status), which tabulates the time of the most recent ingestion for all buoys.

QA/QC. Data providers are expected to conform to the GLOS Quality Management Plan to ensure appropriate procedures are in place to ensure proper calibration and deployment of instruments. GLOS currently performs automated testing of incoming data to detect gross errors.

The current QC modules perform the following simple validation checks:

- water temperature (0-40 deg C)
- air temperature (0-50 deg C)
- dew point (-30-50 deg C)
- relative humidity (0-100%)
- significant wave height (0-10m)
- wave period (0-15 sec)
- maximum wind gust and wind speed (0-50 m/s)
- wind direction (0-360 deg)
- barometric pressure (700-1200 mmHg)
- turbidity (0-1000 NTU)
- chlorophyll (-5-120 RFU)
- phycocyanin (-5-120 RFU)
- specific conductivity (100-700 uS/CM)
- pH (6-10)
- dissolved oxygen (-1-25 mg/L)
- dissolved oxygen saturation (0-150%)
- thermal string (0-40 deg C)

Additional details are included in the Quality Documents portion of this application. GLOS anticipates a rolling implementation of all Required QARTOD tests to be completed by October 31, 2017. GLOS expects to also implement a subset of the Strongly Recommended and Suggested tests.

These tests will be implemented for all ingested parameters for which a relevant QARTOD manual exists. Currently relevant manuals are: Wind Speed and Direction; In-situ Temperature and Salinity; Dissolved Oxygen; Waves; Currents; Ocean Optics; and Dissolved Nutrients. GLOS will track progress of the QARTOD project, and will implement QARTOD tests from relevant new QARTOD manuals within six months of the manuals' publication. For non-QARTOD parameters, GLOS will, at a minimum, implement continuity, range check, spike, stuck sensor, and rate-of-change tests similar to the equivalent tests for QARTOD parameters and informed by input from subject matter experts such as the sensor owners.

Metadata. GLOS has worked with the data providers to develop ISO 19119/19139-compliant metadata records for each platform. The records are manually curated and stored in the GLOS GeoNetwork metadata catalog.

Backups. The OBS DB is backed up weekly to an external, portable hard drive.

Archiving. All non-federally owned near real-time data in the OBS DB are archived to NCEI in netCDF format. Additional details are available in the accompanying NODC Submission Information Form for GLOS initialized 8/24/2015.

Data Access: Sensor data in the OBS DB are exposed for discovery, query and access through a JSON service implemented via Java servlet.

Deprecation. GLOS is transitioning away from the OBS DB towards 52N SOS and anticipates decommissioning of this data store and associated processes by December 31, 2017.

1b. Sensor Data -> 52N SOS

GLOS has also deployed an instance of the IOOS-compliant 52 N SOS server for ingestion, management and communication of non-federally owned sensors.

Data Types. Same as OBS DB.

Management. Sensor Web Harvester is used to harvest observation data every 12 minutes from XML files placed in the 52N ingest folder. These XML files are currently copied to the 52N ingest folder by the same cron process used to manage ingestion of OBS DB data. As GLOS transitions away from use of the OBS DB, SWH will be pointed directly to the GLOS ftp directory receiving XML files from data buoys, and the Buoy and Model Status page updated to reflect 52N harvests instead of OBS DB.

QA/QC. Same as OBS DB.

Metadata. Same as OBS DB.

Backups. The 52N SOS data store is not currently backed up as all observations are duplicates of those in the OBS DB. As GLOS transitions away from the use of the OBS DB, the weekly backup process will be updated to drop the OBS DB and include the 52N SOS store. The backup strategy may be expanded to include daily Veeam snapshots of the virtual machine hosting 52N.

Archiving. The 52N SOS data store is not currently archived because all observations are duplicates of those already archived from the OBS DB. As GLOS completes its transition away from OBS DB, the archive process will adjust to stream non-federally owned near real-time data to NCEI through the 52N SOS store instead of OBS DB.

Data Access: Sensor data in the 52N SOS data store are exposed through the OGC SOS service implemented in the IOOS 52N SOS reference package.

2. Data -> THREDDS

GLOS uses THREDDS to provide data services for numeric model and remote sensing datasets that cover the entire Great Lakes Basin, certain observation datasets that do not match the near real-time data model for continuous data sensors, and a subset of sensor data used for a custom HABs-related data viewer.

Data Types. Varies by source.

Management. There are five different data sources feeding data into the GLOS THREDDS catalog:

- **NOAA-GLERL GLCFS.** GLOS harvests nowcast and 120-hour forecast model results (water temperature, waves and currents) together with meteorological forcing variables (wind, air temperature) from the Great Lakes Environmental Research Laboratory's Great Lakes Coast Forecast System on a regular basis in netCDF format. Harvest status is shown on the GLOS Buoy and Model Status page (http://data.glos.us/status)
- **MTRI Remote Sensing Products.** GLOS harvests derived satellite products from the Michigan Tech Research Institute in netCDF and TIFF format weekly. These products, which consist currently of water surface temperature, natural color, and three different color-producing agents for each of the Great Lakes, are delivered only when cloud-free MODIS-AQUA coverage is

available for a given lake. The creation of these products is partially supported by GLOS, which displays the MTRI-derived products as provided and perform no additional processing.

- **USACE Coordinated Lake Water Levels.** GLOS scrapes daily coordinated water levels published by the U.S. Army Corps of Engineers Detroit District on behalf of the Coordinating Committee on Great Lakes Basic Hydraulic and Hydrologic Data, and stores the scraped values in netCDF format.
- **HABs Sensors.** Observations from a subset of the near real-time sensors streaming into the OBS DB are streamed in turn into netCDF files stored in the THREDDS catalog. These files drive the initial version of the GLOS Western Lake Erie Basin HABs Viewer, a rapidly-deployed specialized portal that is architected around netCDF data. Data parameters are the same as for OBS DB.
- **Manually curated datasets.** A number of datasets have been added to the THREDDS catalog manually, including thermistor data downloaded seasonally from selected MTRI platforms; water quality and meteorological data from Ranger III, a vessel of opportunity; etc.

QA/QC. Data from NOAA and USACE are assumed to be covered for QA/QC. QA/QC information, where available, has been referenced in metadata records for the remaining datasets. Going forward, GLOS will ensure that QA/QC information be included in or linked from the associated metadata record for all datasets unless exceptions are agreed to with IOOS.

Metadata. GLOS has worked with the data providers to develop ISO 19119/19139-compliant metadata records for each platform. The records are manually curated and stored in the GLOS GeoNetwork metadata catalog.

Archiving. GLOS does not currently archive any files in the THREDDS catalog, as all datasets are replicates of datasets maintained by the data providers – with the exception of the manually curated datasets. GLOS anticipates adding – where applicable and acceptable to NCEI – all current manually curated datasets to the NCEI archive by December 31, 2016. GLOS will also plan to add all such future datasets – where applicable and acceptable to NCEI – to the NCEI archive on an ongoing basis.

Backups. The THREDDS catalog is backed up onsite weekly to a removable HDD.

Data Sharing. Data in the THREDDS catalog can be accessed through any TDS service, including OPeNDAP, WMS, WCS, and ncISO (WCS and ncISO are not depicted in Figure 1).

Deprecation. The HABs viewer is anticipated to be reconfigured to allow access to sensor data through SOS in the near future, at which time the HABs-related netCDF generation will cease.

B. Data Services

GLOS data services offer data discovery, data access and data archival through standardized protocols, services and applications from OGC and others that have been widely adopted by IOOS and other RAs. GLOS has also developed and maintains a legacy observation database and associated custom tools and data services that meet regional needs for adaptive data management and application development.

The THREDDS server plays a central role for data services at GLOS since NetCDF is the primary data format for the numeric models and remote sensing data that GLOS currently possesses, and data streaming through OPeNDAP protocol is the data access approach most frequently used by site visitors. Relational PostGRES/PostGIS databases are used for persistent storage of in-situ observation data managed by GLOS, with IOOS 52N SOS and the legacy GLOS observation web services running in tandem to provide access to near real-time observation data. GLOS has also deployed GeoServer to support mapping functionality within its GeoNetwork metadata catalog.

Sci-WMS and the GLCFS point query tool are also available from GLOS to address certain limitations of THREDDS. Sci-WMS, developed by RPS-ASA, was deployed to handle numeric model datasets in NetCDF format with an irregular grid. Sci-WMS provides data visualization as well as data access against irregular grids through OGC's Web Map Service. The GLOS-developed GLCFS point query tool acts as a complementary data access method on top of THREDDS for retrieving Great Lakes Coastal Forecasting System modeled nowcasts and forecasts in a more intuitive manner.

1. THREDDS

THREDDS is used at GLOS to provide data streaming services against numeric model and remote sensing datasets that cover the entire Great Lakes Basin. THREDDS is also used to store certain observation datasets that do not match the near real-time data model for continuous data sensors.

Due to the large amount of data it serves and its caching mechanism implementation, THREDDS is generally memory-bound. A dedicated virtual machine with 24 GB memory running CentOS 6 was therefore configured to host solely the GLOS THREDDS instance. The datasets exposed to THREDDS are stored on a disk array with open access to this virtual machine through NFS on a gigabit LAN network switch. At the application level, Tomcat 7 with Oracle JRE 1.7 was installed and certain optimizations were done at Tomcat level, such as activating the Tomcat Native APR library to boost performance, and configuring the jsvc daemon/watchdog for better control and reliability.

Tomcat 7, as the JEE server, only listens to tcp/80 on the localhost. Nginx was implemented as a reverse proxy server in front of Tomcat. With its asynchronous I/O capacity, nginx offers better performance for proxy purposes. Here is a snippet of the configuration in nginx for the THREDDS proxy:

```
location ~^/thredds/* {
 proxy_set_header X-Real-IP $remote_addr;
 proxy_set_header X-Forwarded-For $proxy_add_x_forwarded_for;
 proxy_set_header Host $host;
 proxy_pass http://localhost:8080;
 }
```

The THREDDS dataset inventory catalog (catalog.xml) contains the master configuration files for datasets that are served on THREDDS. This file serves as an index so that each entry in catalog.xml

provides a virtual directory of available data and its associated metadata. It then points to a sub-catalog which is a different xml file that implements the configuration for that particular dataset.

For example, the index entry below indicates that the specific configurations for the dataset "Cumulative Impact Assessment – Water Use Datasets" can be found in a file called wateruse.xml, which is located in the folder called "glc", and that the title of this dataset that is displayed on THREDDS web page is assigned in the attribute "xlink:title".

```
<dataset name="Cumulative Impact Assessment -- Water Use Datasets">
 <catalogRef xlink:title="GLC - Water Use Historical 2006 to 2011"
xlink:href="glc/wateruse.xml" name=""/>
 </dataset>
```

Both static and dynamic datasets are presented in the THREDDS catalog.

Static datasets are backed either by individual NetCDF files or a series of NetCDF files that will not be updated once they have been created. An example of the catalog configuration for the former case is listed below.

```
<dataset name="ER43 - Hyperpro II Optical Profile" ID="er43_hyperpro"
urlPath="glop/er43_hyperpro.nc">
<netcdf xmlns="http://www.unidata.ucar.edu/namespaces/netcdf/ncml-2.2"
location="/var/thredds/glop_datasets/er43_hyperpro.nc">
 <attribute name="title" value="ER43 - Hyperpro II Optical Profile" />
 <attribute name="summary" value="Great Lakes Optical Properties" />
 <attribute name="time" value="2007-09-12 13:00:00-04:00" />
 <attribute name="metadata_link" type="String"
value="http://data.glos.us/portal/" />
 <attribute name="Metadata_Conventions" type="String" value="Unidata
Dataset Discovery v1.0"/>
 <attribute name="standard_name_vocabulary" type="String"
value="http://www.cgd.ucar.edu/cms/eaton/cf-
metadata/standard_name.html" />
 </netcdf>
</dataset>
```

For the latter case, the virtual aggregation against netCDF files is performed using NetCDF Markup Language (NcML). A union of NetCDF files can be done as shown below:

```
<dataset name="GLOS Glider - unit_236 - Aggregation"
urlPath="gliders/unit_236-Agg" ID="unit_236-Agg" dataType="Point">
 <netcdf xmlns="http://www.unidata.ucar.edu/namespaces/netcdf/ncml-
2.2">
 <attribute name="title" value="GLOS Glider - Aggregation" />
 <attribute name="summary" value="Aggregation of Glider data from The
Great Lakes." />
 <attribute name="metadata_link" type="String"
value="http://data.glos.us/portal/" />
 <attribute name="geospatial_vertical_min" type="double" value="0.0" />
 <attribute name="geospatial_vertical_max" type="double" value="0.0" />
 <attribute name="geospatial_vertical_units" type="String"
value="meters" />
 <attribute name="geospatial_vertical_resolution" type="double"
value="0.0" />
```

```
 <attribute name="geospatial_vertical_positive" type="String"
value="up" />
 <aggregation dimName="time" type="joinExisting" recheckEvery="60 min">
 <scan location="/var/thredds/glider/unit_236/" suffix=".nc"/>
 </aggregation>
 </netcdf>
  </dataset>
```

Dynamic datasets, on the other hand, are served using Feature Collection, which is the TDS way to handle collections of CDM Feature Datasets. The feature collection is designed to be used for the collection of grid datasets; the virtual dataset can be created in THREDDS on top of the raw collection. An example of a feature collection dataset on GLOS THREDDS server is below:

```
<featureCollection name="Lake Michigan - Nowcast - 2D - 2009"
featureType="FMRC" harvest="true" path="glos/glcfs/archive2009/michigan/ncfmrc-
2d">
 <collection spec="/var/thredds/GLCFS/Archive/2009/m#yyyyDDDHH#\.out1\.nc$"
olderThan="5 min" />
 <protoDataset choice="Penultimate">
 <netcdf xmlns="http://www.unidata.ucar.edu/namespaces/netcdf/ncml-2.2">
 <remove type="dimension" name="validtime_length" />
 <remove type="variable" name="validtime"/>
 <attribute name="validtime" value="01-JAN-2009 00:00 GMT" />
 <attribute name="validtime_DOY" value="001, 2009 00:00 GMT" />
 <attribute name="comment2" value="1-hourly model 2D output starting at
validtime plus 1 hr" />
 <attribute name="Conventions" type="String" value="CF-1.6"/>
 </netcdf>
 </protoDataset>
 <fmrcConfig datasetTypes="Best Files" />
 <netcdf xmlns="http://www.unidata.ucar.edu/namespaces/netcdf/ncml-2.2">
 <variable name="depth" shape="ny nx" type="float">
 <attribute name="long_name" value="Bathymetry " />
 <attribute name="units" value="meters" />
 <attribute name="positive" value="down" />
 <attribute name="standard_name" value="depth" />
 <attribute name="coordinates" value="lat lon"/>
 </variable>
 <variable name="wvh" shape="time ny nx" type="float">
 <attribute name="long_name" value="Significant Wave Height" />
 <attribute name="units" value="meters" />
 <attribute name="missing_value" type="float" value="-99999.0" />
 <attribute name="standard_name" value="wave_height" />
 <attribute name="coordinates" value="time lat lon"/>
 </variable>
 </netcdf>
  </featureCollection>
```

Other than the elements for describing the dataset, the catalog file also contains the components to depict the metadata associated with the dataset. This information is not only displayed on the THREDDS web page, but can be retrieved through the ncISO metadata service as well. The metadata section is usually defined under the dataset tag:

```
<metadata inherited="true">
 <serviceName>all</serviceName>
 <keyword vocabulary="GCMD Science Keywords">GLOS, GLCFS, Nowcast, Great
Lakes</keyword>
 <date type="created">2012-01-01</date>
```

```
 <date type="modified">2012-01-01</date>
 <date type="issued">2012-01-01</date>
 <creator>
 <name vocabulary="DIF">Dr. Dave Schwab</name>
 <contact url="http://www.glerl.noaa.gov/" email="david.schwab@noaa.gov"/>
 </creator>
 <publisher>
 <name>GLOS DMAC</name>
 <contact url="http://glos.us" email="dmac@glos.us"/>
 </publisher>
 <documentation type="rights">No usage restrictions</documentation>
 <documentation xlink:href="http://www.glerl.noaa.gov/res/glcfs/"
xlink:title="Great Lakes Coastal Forecasting System"/>
 <documentation type="Summary"> Great Lakes Coastal Forecasting
System</documentation>
 <documentation type="Disclaimer"> NOAA GLERL is providing this data "as is,"
and NOAA GLERL and
 its partners cannot be held responsible, nor assume any liability for any
damages caused by
 inaccuracies in this data or documentation, or as a result of the failure of
the data or
 software to function in a particular manner. NOAA GLERL and its partners make
no warranty,
 expressed or implied, as to the accuracy, completeness, or utility of this
information, nor does
 the fact of distribution constitute a warranty. Real-time data have not been
subjected to
 quality control or quality assurance procedures. Timely delivery of data and
products through
 the Internet is not guaranteed. Before using information obtained from this
server, special
 attention should be given to the date and time of the data and products being
displayed.
 </documentation>
 <contributor role="distributor">GLOS DMAC</contributor>
 <contributor role="producer">GLERL</contributor>
 <property name="viewer" value="http://data.glos.us/portal/, GLOS Data Portal"
/>
 </metadata>
```

GLOS currently offers multiple web services through THREDDS: OPeNDAP, WMS, NCML, UDDC and ncISO. With these provided services, users can efficiently conduct data discovery and data retrieval in an automatic means.

Figure 2 lists the datasets that are available at GLOS THREDDS and can be found at http://tds.glos.us/thredds/catalog.html.

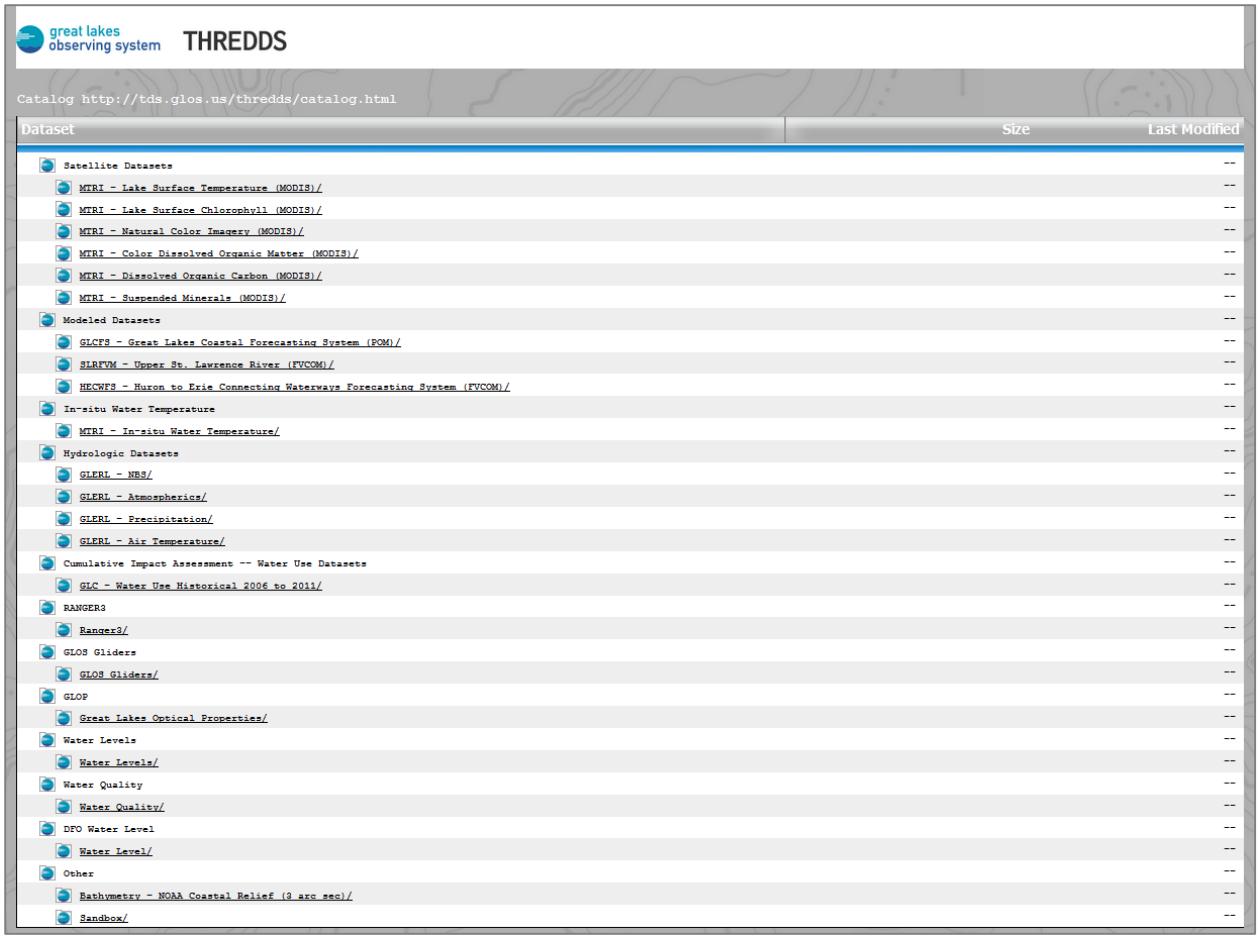


Figure 2. THREDDS Catalog

Due to the limitations of THREDDS on handling unstructured grid data, certain datasets, such as FVCOM model data, are provided on THREDDS as-is. GLOS only offers raw data access ability for these datasets through THREDDS.

2. SOS

SOS is an OGC standard used for interoperable management of sensor data. This web service interface provides a means for querying data and metadata, registering and removing sensors, and adding new sensor observations (OGC, 2015).

GLOS has implemented the IOOS customized build of the 52° North SOS, which has been extended with IOOS-specific encoding formats and test data as described in the IOOS 52N reference. The IOOS enhancements to 52N include extra metadata in the GetCapabilities response and extensions of SensorML to include extra metadata and incorporate the concept of hierarchies for networks, stations, and sensors. The GLOS SOS implementation complies with the IOOS SWE Milestone 1.0 templates and requires datasets to be in any of the CF 1.6 Discrete Sampling Geometries.

The GLOS 52 North instance for serving the in-situ observation data runs through Tomcat 7 and uses PostgreSQL 9.2 / PostGIS 2.0. The GLOS SOS data includes air pressure, air temperature, dew point,

humidity, significant wave height, wave period, water level, water temperature, wind speed and directional data from the GLOS network of sensors. Observation data are harvested every 12 minutes using Sensor Web Harvester from submitted XML and injected into 52N.

SOS is physically isolated from other JEE applications by taking a multiple instances approach with Tomcat: a crash of other applications and JVMs sharing the same Tomcat binary will not affect the SOS. The SOS Tomcat instance only listens on local ports and takes Apache web server as its proxy. The backend PostgreSQL 9.2 is installed on a separate database server.

In addition to the 44 buoys and stations in GLOS 52N SOS, GLOS applications consume SOS services from:

- Center for Operational Oceanographic Products and Services (COOPS) SOS – 53 stations
- National Data Buoy Center (NDBC) – 24 buoys and stations
- Lake Simcoe Resource Conservation Authority (LSCRA) – 2 stations

3. GeoServer

Although spatial data are not key datasets for GLOS, GeoServer is deployed to provide mapping services for the GLOS metadata catalog GeoNetwork instance. The GeoServer 2.6.2 instance was installed along with GeoNetwork in a multiple Tomcat instances environment, and provides a full stack of OGC-compliant services, such as WCS, WFS, WMS, WMS-T, etc. There are currently no GLOS datasets exposed externally through GeoServer.

4. sciWMS

SciWMS is an open source Python-based set of visualization services implementing the OGC WMS specification. Supporting both structured and unstructured grids, sciWMS was developed as existing WMS technologies did not support unstructured meshes in a format that preserved topology. SciWMS supports triangular unstructured meshes, curvilinear and rectangular grids. Features include styles that preserve unstructured mesh topology and topologically correct GetFeatureInfo requests for time-series data analysis. The services provide visualizations of scientific data stored as netCDF files via OPeNDAP.

GLOS has implemented sciWMS to serve irregularly-gridded model data ingested from the Great Lakes Coastal Forecasting System (GLCFS), the Huron to Erie Connecting Waterways Forecasting System (HECWFS), and the St. Lawrence River FVCOM forecasting system (SLRFVM). SciWMS is also used to serve MTRI's natural color satellite imagery.

5. EDS

The Environmental Data Server (EDS) provides homogenous access to meteorological and hydrodynamic data via web services. The EDS provides visualization services (WMS) as well as data access services so model data can be used in decision support applications such as Search & Rescue and Oil Spill response. It is designed to be scalable and modular and utilizes a Service Oriented Architecture (SOA). Users (clients) access the available data by using a web service, and the web service returns data based on the request. The web service may return data to the client from data products stored in distributed servers, or a more common approach is that the catalog server retrieves data to a central server. A TDS instance is layered on top of the EDS Data Store to provide catalog and DAP services. The EDS consists of three major sub-systems governing data acquisition (the Catalog Server), data storage (the Data Store), and data request (the EDS web services).

Several GLOS data products are currently served by EDS including the NOS Currents, NAM Winds, and GLERL Currents, Ice, Water Level, and Waves products. Data provided by GLOS partners is harvested by ASA's production EDS instance. The production EDS is currently used in the following manner:

- A source for gridded (models, radar, satellite) wind and current data from global, national and regional providers
- Model data are stored in its native format (generally NetCDF) and grid structure. See example of unstructured FVCOM for Gulf of Maine being displayed with EDS services. Model data are not regridded on the EDS servers
- Data are stored and maintained in its native format and can be delivered with custom data services
- Aggregation services are available to provide temporal and, in some cases, spatial aggregation products
- The EDS manages structured and unstructured gridded data and data are made available via visualization and data access web services
- Visualization services can be integrated into map-based web sites managed by different entities. New visualization techniques are being developed including high performance "on the fly" contouring techniques – see example below. Visualization is performed directly from the model grids, no data conversions are needed
- Staging and development versions of EDS are used to evaluate and distribute pre-operational model products. This provides a framework for easily sharing experimental model results.
- Model providers do not need to make any changes to their workflow – the EDS simply collects model results as soon as they are available in their native format.
- ASA is working on a new set of unstructured grid access services so unstructured gridded data may be subsetted and regridded
- ASA also maintains deep archives of the data. ASA archives most source data from the point that the new data source was added to the Coastmap EDS. The archived data are available for restoration to Coastmap Production servers with very little lead-time.
- Redundancy – ASA hosts 3 instances of the EDS, Development, Staging, and Production with the ability to quickly add instances. ASA also deploys an instance of the EDS on the Amazon Cloud.
- The EDS currently collects a wide variety of federal and regional data that can be made available to users and the infrastructure
- The EDS currently focuses on delivery of 2D time-varying data, but development is ongoing to efficiently deliver subsetted 3D data
- A new model skill assessment project is underway with the US Coast Guard to automate model evaluation based on surface drifter comparison

6. Non-SOS or SOS-proxy Sources

Data are harvested from a variety of other portals or data sources. This data follows the same path as the SOS data, flowing through a parser and then loaded into the Data Portal. These include:

- Water Quality Portal (EPA STORET) using an SOS proxy
- USGS data
- National Weather Service data

7. Radar and Satellite Data

NEXRAD data and products derived from remote sensing are ingested on-the-fly for display in the GLOS data portal:

a. Next Generation Radar (NEXRAD) – From Weather Underground, provides a layer depicting measures of precipitation and wind. This layer allows users to see current real-time weather as well as display of weather hazards.

b. Michigan Tech Research Institute (MTRI) satellite layers – For cloud-free dates during the ice-free season (typically April through October), MTRI provides satellite images for the following data:

- CPAs: The water color of inland and coastal zones results mainly from the concentrations of three different substances known as color-producing agents (CPAs).
    - Chlorophyll concentration (CHL) - Chlorophyll (Chl) is a green pigment found in plant and phytoplankton cells. Algal cells that are suspended in water produce a green color.
    - Colored dissolved organic matter (CDOM) - CDOM is the optically measurable component of the dissolved organic carbon in water and is naturally occurring due to decaying plant matter and micro-organism metabolism. CDOM is highly absorbent of blue light, which results in a yellow-brown color. CDOM is typically delivered to large aquatic ecosystems via river discharge.
    - Dissolved organic carbon (DOC) - DOC is a form of organic carbon that is produced during micro-organism metabolism and may also be transported from decaying vegetation via rivers and streams. DOC only absorbs light, it does not scatter light, and appears yellow to brown in color. A significant fraction of the total DOC concentration is not detectable from electro-optical satellite data and must be estimated empirically from the colored portion, known as Colored Dissolved Organic Carbon (CDOM).
    - Suspended minerals (SM) - Suspended minerals (SM) consist of inorganic particulate matter (such as sand, silt, and clay), which scatter and absorb light.
- Natural color - Collected by the MODIS satellite sensor, these natural color images are produced using the Red, Green, and Blue channels at a spatial resolution of 1 kilometer.
- Water surface temperature - Nighttime surface water temperature data are derived from MODIS satellite imagery, collected at the 4-micron wavelength when cloud cover is not significant. These data are accurate within approx. 0.5°C.

C. Technology Stack

See separate document > GLOS_Technology.docx

D. Complete listing of Assets

See separate document> GLOS_AssetsInventory.docx and http://data.glos.us/status

E. System Components (Tools / Access to Data)

GLOS has a variety of data management and services components. These include instruments from non-GLOS partners, models, satellite feeds, and servers:

a. Instruments and models

1. Glider DAC

2. Water Quality Buoys

3. Nearshore Sensors

4. Archived Data from non-GLOS assets

5. Numerical Models

6. Atmospheric Model feeds

7. Data derived from remote providers contains data GLOS either updates frequently or archives (e.g. University of Minnesota-Duluth / University of Michigan: Graham Sustainability Institute (GLISA) / Cornell University / University of Toledo Lake Erie Center (LEC) / Heidelberg University / Evaporation Network / Ranger3 (MTRI) / Optical Properties database (MTRI) / Army Corps Water Level data / Lake Simcoe / Historical Water Use (GLC) / GLERL Hydrologic data / Thermistor data (MTRI))

b. Web Pages

All GLOS tools are accessible from the main URL http://www.glos.us.

Additional information is accessible from this website such as current news and events and general information about the purpose of GLOS.

Users can access featured tools (described in more detail in section c below):

- Data Portal
- Boaters' Forecast
- Huron to Erie Connecting Waterways Forecasting System
- Point Query Tool for the Great Lakes Coastal Forecasting System
- Great Lakes Acoustic Telemetry Observation System (GLATOS)
- Data Catalog
- Western Lake Erie HABs Viewer
- Great Lakes Model Inventory

Partner data links are also provided to:

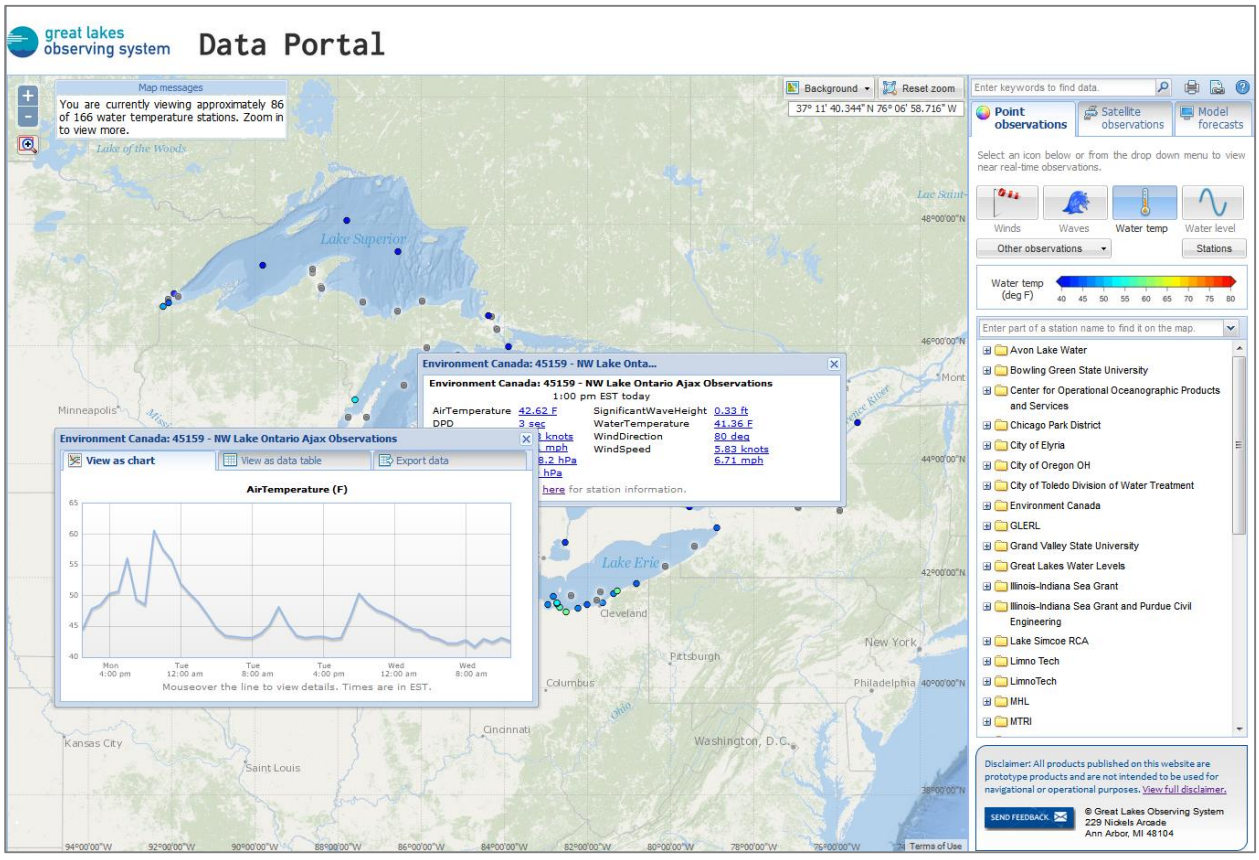
- Data on current lake conditions (http://www.glin.net/conditions/)
- Surface temperature data (http://www.coastwatch.msu.edu/)
- Water monitoring data (http://www.epa.gov/STORET/)
- Great Lakes Environmental Database (http://www.epa.gov/glnpo/monitoring/data_proj/glenda/)
- Michigan Tech Research Institute (MTRI) Cladophora/SAV mapping project results http://www.mtri.org/cladophora.html

c. Data browsing tools

The following tools allow users to access (view), browse (search), and optionally download data or datasets. The software used by GLOS may be open source or proprietary tools depending on the need. All tools are browser-based, requiring only that users have access to a modern web browser and the Internet.

1. GLOS Data Portal

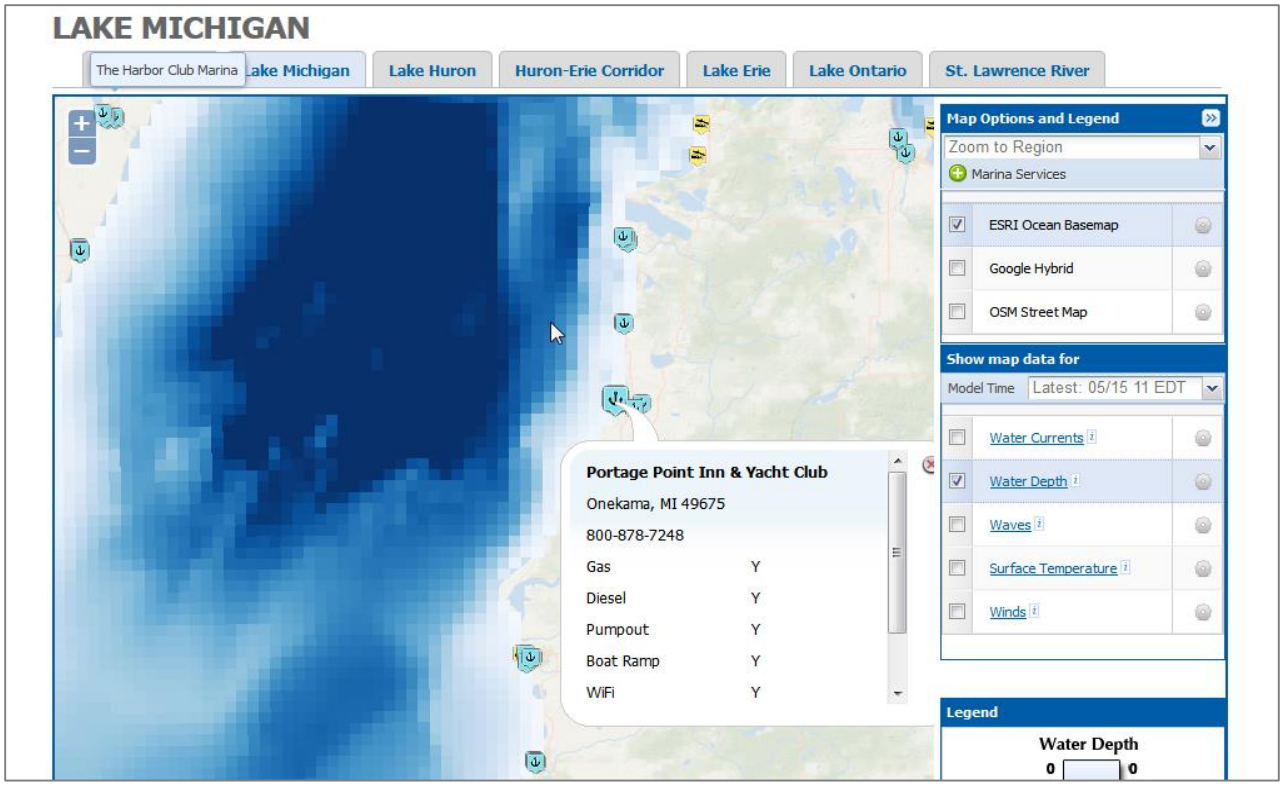
The primary user access tool is the GLOS Data Portal (http://www.data.glos.us/portal) The Data Portal provides browsing access to near real-time data for Point Observations (winds, waves, water temperature, water levels, air temperature, dissolved oxygen, streamflow and turbidity), Satellite Observations (base reflectivity including weather hazards, chlorophyll concentration, colored dissolved organic matter, dissolved organic carbon, natural color, suspended minerals and water surface temperature), Model Forecasts (currents, ice thickness, water level, waves, winds). It also provides search access to metadata from numerous partner stations and from our GeoNetwork instance. Some data are archived depending on agreements with data providers and may be made available upon request.



2. Boaters Forecast Tool

The Boaters' Forecast Tool (http://data.glos.us/boaters/) connects recreational boaters to information that helps them better understand the current conditions on the lakes and St. Lawrence River. The tool is customized to address information needs specific to boaters in the Great Lakes and includes data for water currents, water depth, waves, surface temperature and winds as well as information on marina and boat launch locations. Users are also able to save their "favorite" locations for ease of future access.

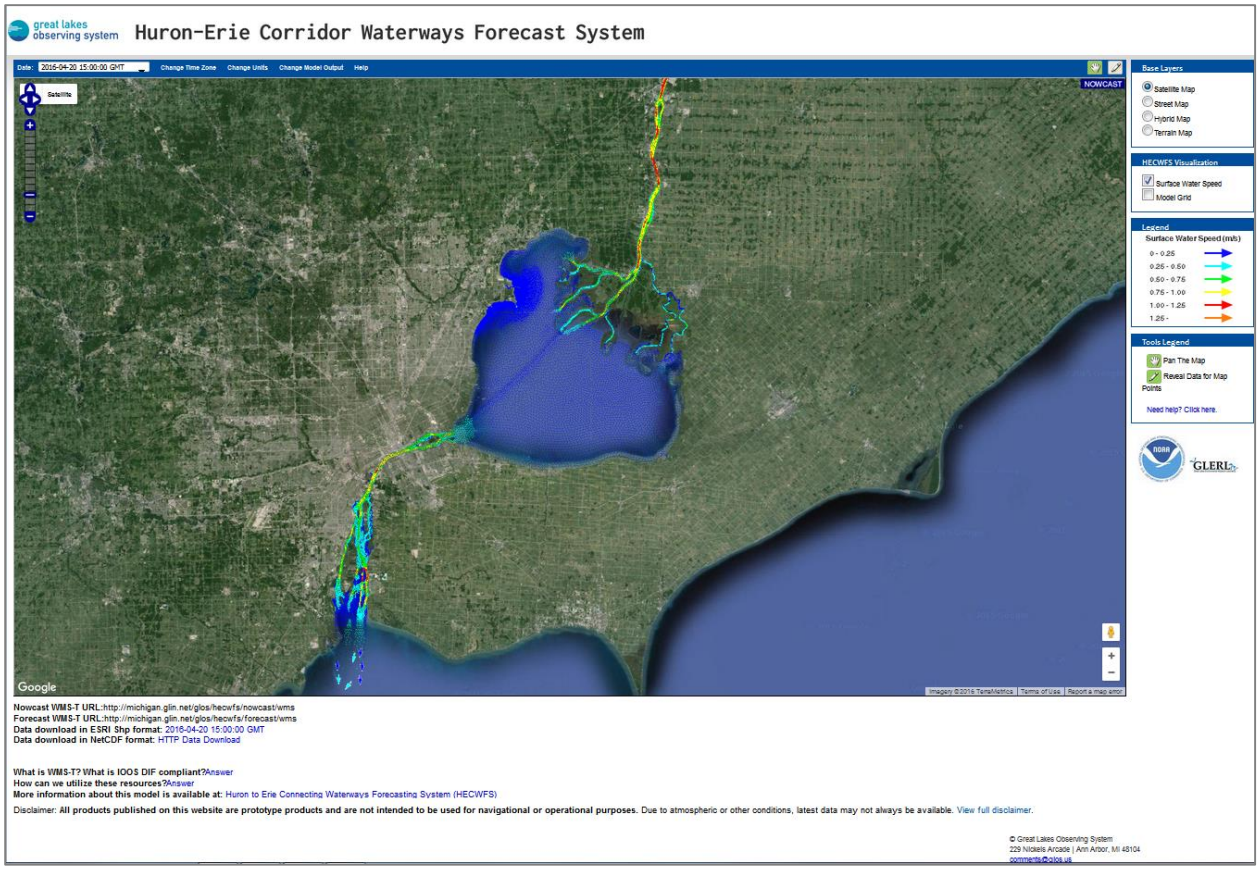
The tool was developed in partnership with New York Sea Grant and NOAA's Great Lakes Environmental Research Laboratory (GLERL).


LAKE MICHIGAN
The Harbor Club Marina
Lake Michigan
Lake Huron
Huron-Erie Corridor
Lake Erie
Lake Ontario
St. Lawrence River
Map Options and Legend
Zoom to Region
Marina Services
ESRI Ocean Basemap
Google Hybrid
OSM Street Map
Show map data for
Model Time
Latest: 05/15 11 EDT
Water Currents
Water Depth
Waves
Surface Temperature
Winds
Portage Point Inn & Yacht Club
Onekama, MI 49675
800-878-7248
Gas Y
Diesel Y
Pumpout Y
Boat Ramp Y
WiFi Y
Legend
Water Depth
0
0

3. Huron to Erie Connecting Waterways Forecasting System

The Huron to Erie Connecting Waterways Forecasting System (HECWFS) (http://data.glos.us/hecwfs/) provides predicted forecasts of water levels and currents through the Huron to Erie Corridor. The tool has varying applications including guiding spill response, search and rescue, beach health management, drinking water safety, finding the source of fish kills, and more.

Features include nowcasts every three hours and forecasts every 12 hours for currents and water level for the St. Clair River, Lake St. Clair and the Detroit River. Users are able to select visualizations for surface water speed and/or model grid.



4. Point Query Tool for the Great Lakes Coastal Forecasting System

The Point Query Tool (http://data.glos.us/glcfs/) provides access to the NOAA Great Lakes Coastal Forecasting System (GLCFS) input data and model output for a given location and time period. The GLCFS is a suite of numerical model that calculates waves, current and temperatures in each of the Great Lakes. The GLCFS Nowcast is run four times per day and provides estimates of conditions at the time the model was run. THE GLCFS Forecast runs twice a day and provides a prediction of conditions 60 hours into the future. Two sets of model results are created during each run: one defining conditions on the surface (two-dimensional) and one that defines circulation within the lake (three-dimensional).

Users are able to input specific Latitude/Longitude to query or can choose a point from a map view and then download the data in CSV format.

The GLCFS Nowcast and Forecast 3d model results include 10 variables:

Longitude in degrees north
Latitude in degrees east
Bathymetry depth in meters
Depth for each node in meters
Sigma value for each node
Time as UTC in seconds since 1/1/1970
Water temperature in degrees Celsius
Eastward Current velocity in meters/second
Northward current velocity in meters/second
Height above model sea level in meters

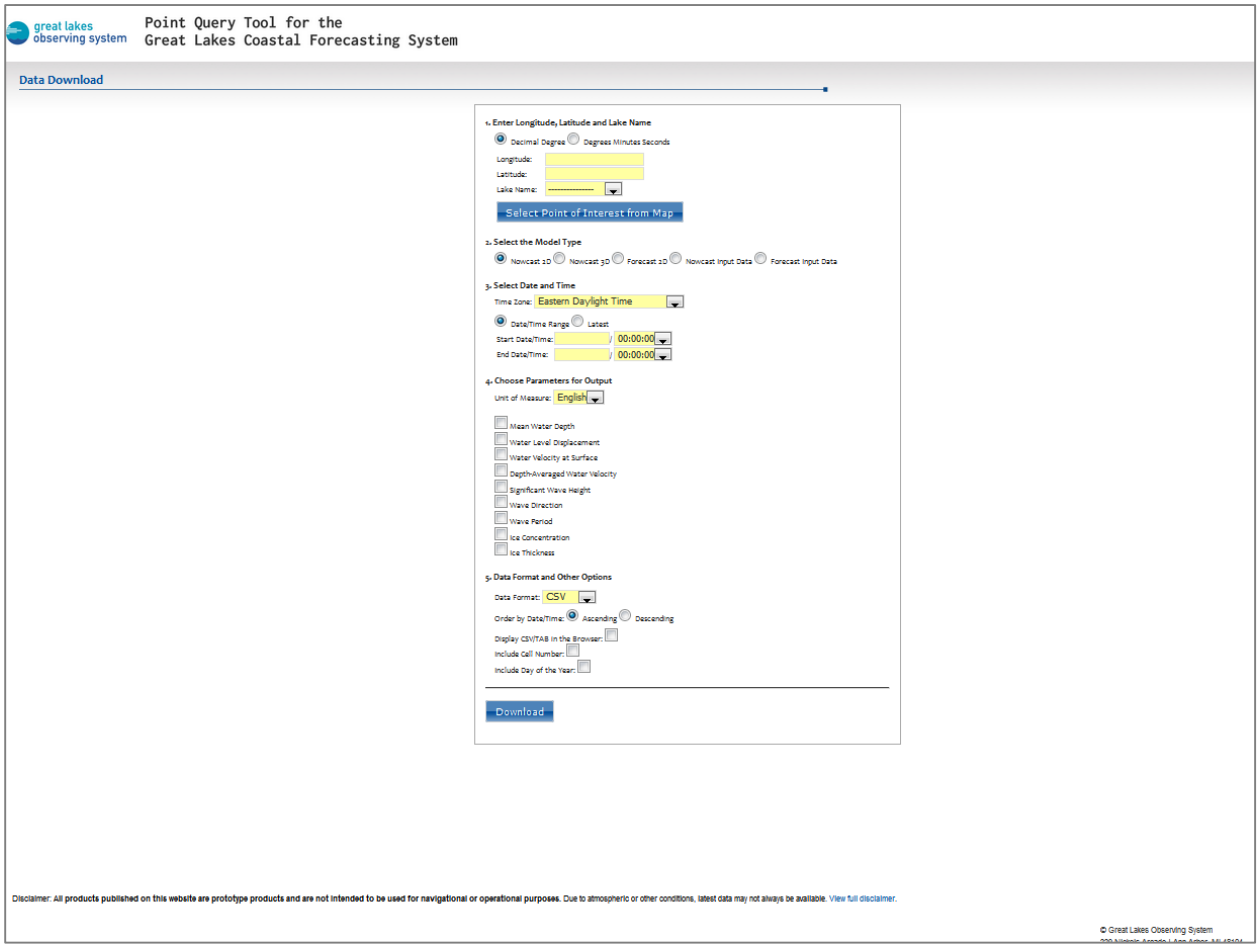
GLCFS data for the current calendar year is updated on the GLOS THREDDS server (TDS) after each run of the model. Archives of Nowcast results are created for each completed calendar year beginning with 2006. Use the links below to connect with the individual TDS dataset pages:

Overview
GLCFS Nowcast
GLCFS Forecast
GLCFS Forcing Data Current Year
GLCFS Forcing Data 2013 Archive
GLCFS Forcing Data 2012 Archive
GLCFS Forcing Data 2011 Archive
GLCFS Forcing Data 2010 Archive
GLCFS Forcing Data 2009 Archive
GLCFS Forcing Data 2008 Archive
GLCFS Forcing Data 2007 Archive
GLCFS Forcing Data 2006 Archive

GLCFS Nowcast 2013 Archive
GLCFS Nowcast 2012 Archive
GLCFS Nowcast 2011 Archive
GLCFS Nowcast 2010 Archive
GLCFS Nowcast 2009 Archive
GLCFS Nowcast 2008 Archive

GLCFS Nowcast 2007 Archive
GLCFS Nowcast 2006 Archive

http://www.glerl.noaa.gov/res/glcfs/

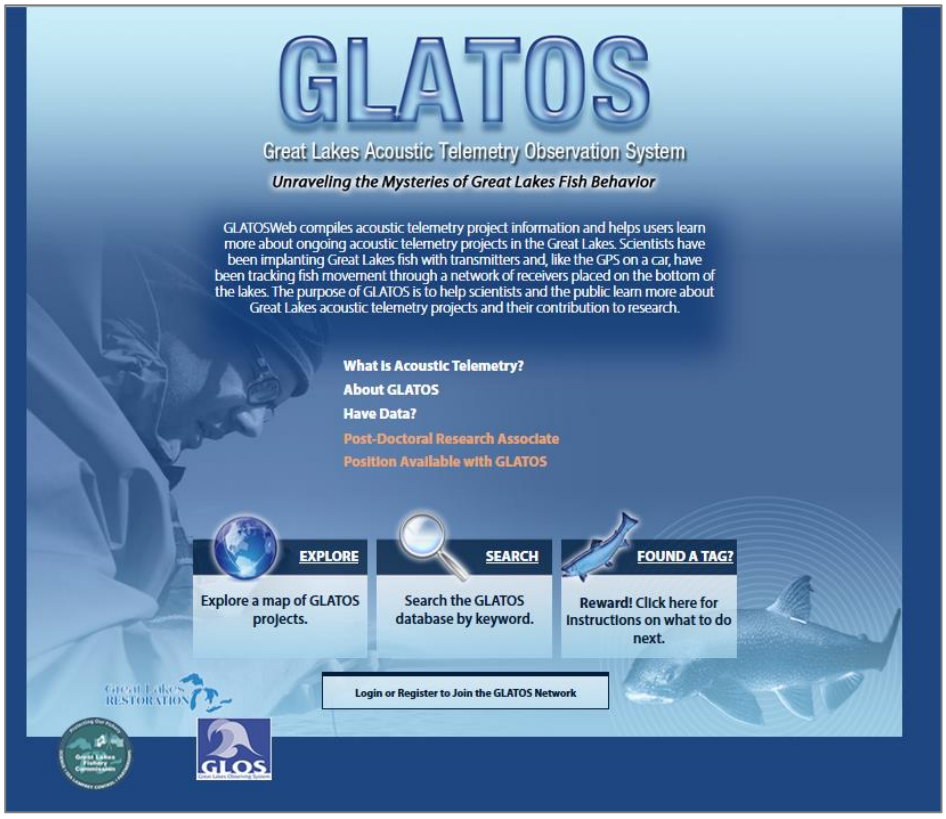


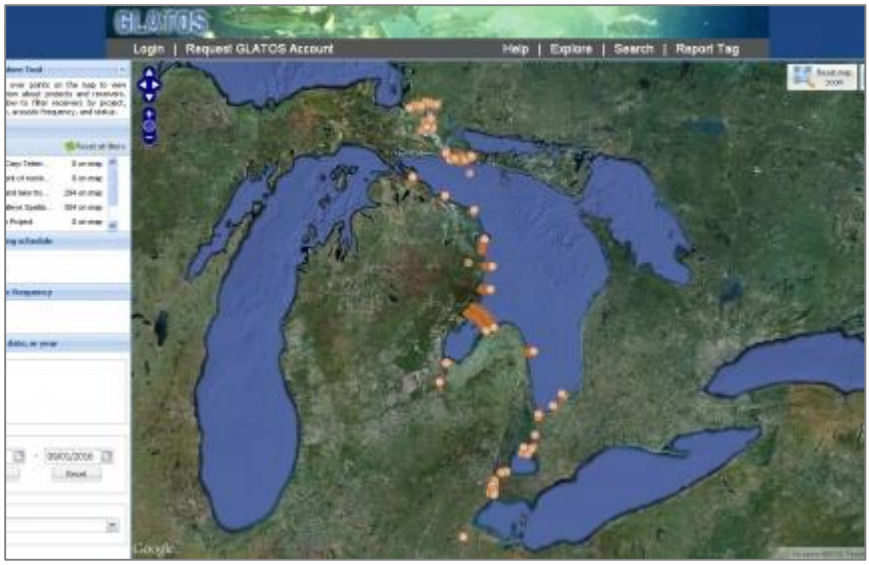
5. Great Lakes Acoustic Telemetry Observation System (GLATOS)

GLATOSWeb (http://data.glos.us/glatos) compiles acoustic telemetry project information and helps users learn more about ongoing acoustic telemetry projects in the Great Lakes. Scientists have been implanting Great Lakes fish with transmitters and tracking their movement through a network of receivers placed on the bottom of the lakes. The purpose of GLATOS is to help scientists and the public learn more about Great Lakes acoustic telemetry projects and their contribution to research.

The tool allows registered users to explore projects, search the database and report a tag.

The GLATOSWeb tool was developed in partnership with the Great Lakes Fishery Commission (GLFC) and the US Geological Survey (USGS). The GLFC acquired acoustic telemetry equipment in 2010, and fish tagged for tracking include lake trout, sea lamprey, walleye, and lake sturgeon.



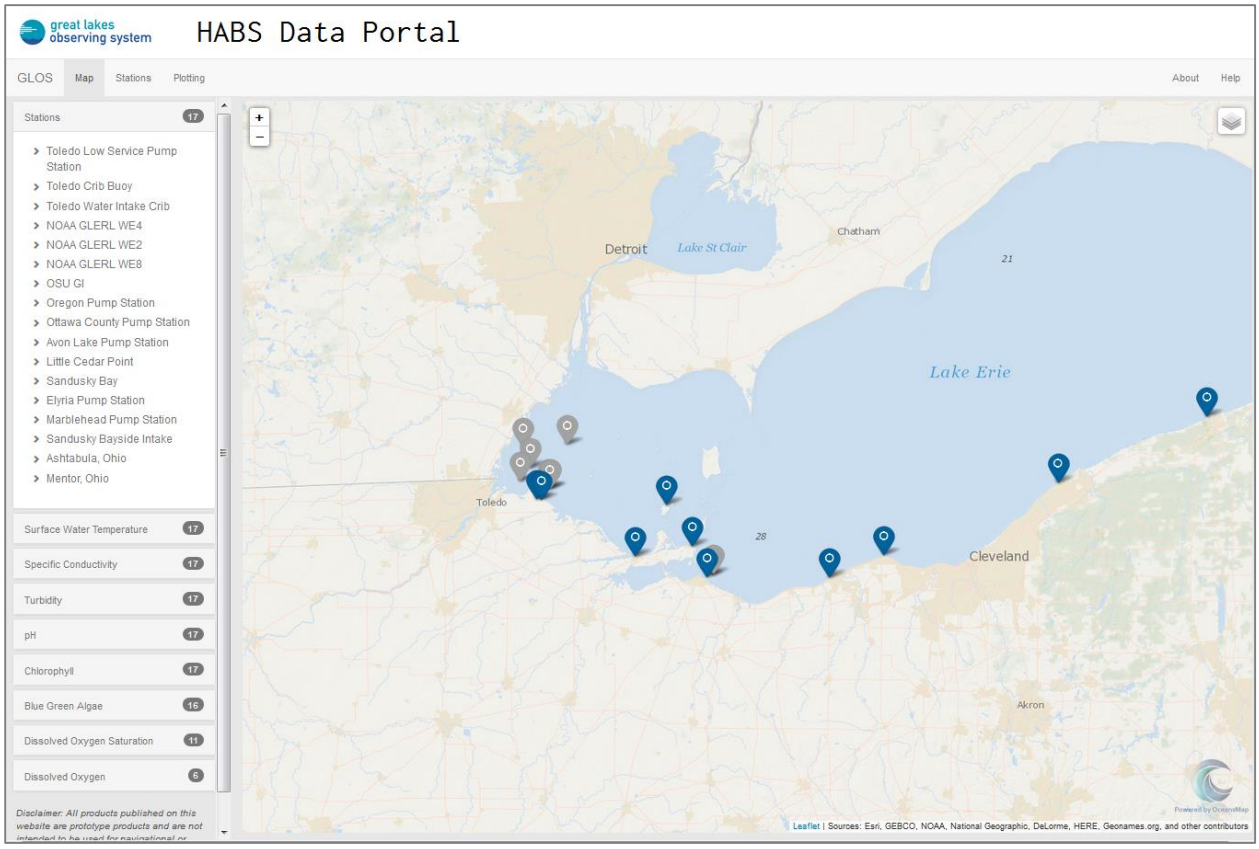

GLATOS
Login | Request GLATOS Account
Help | Explore | Search | Report Tag

6. Western Lake Erie Basin HABs Viewer

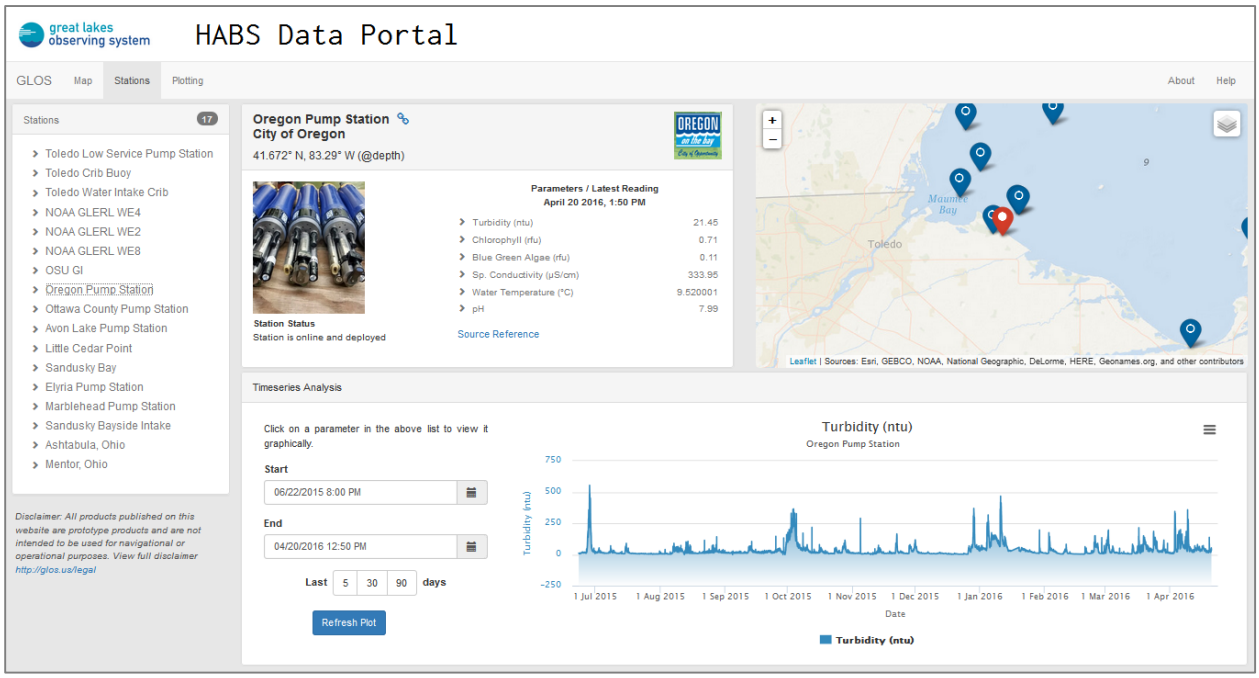
The Lake Erie Harmful Algal Bloom (HABs) Viewer (http://habs.glos.us) is a specialized instance of the GLOS data portal. In collaboration with numerous monitoring stations in the Western Lake Erie Basin, GLOS is providing access to critical real-time data needed to identify factors that influence harmful algal blooms. Measurements include temperature, specific conductivity, turbidity, pH, chlorophyll and blue green algae. Upcoming enhancements will add access to grab sample data, satellite imagery, hydrodynamic model results and meteorological observations.

Users can plot parameters from station(s) of interest as well as view current conditions on a map.

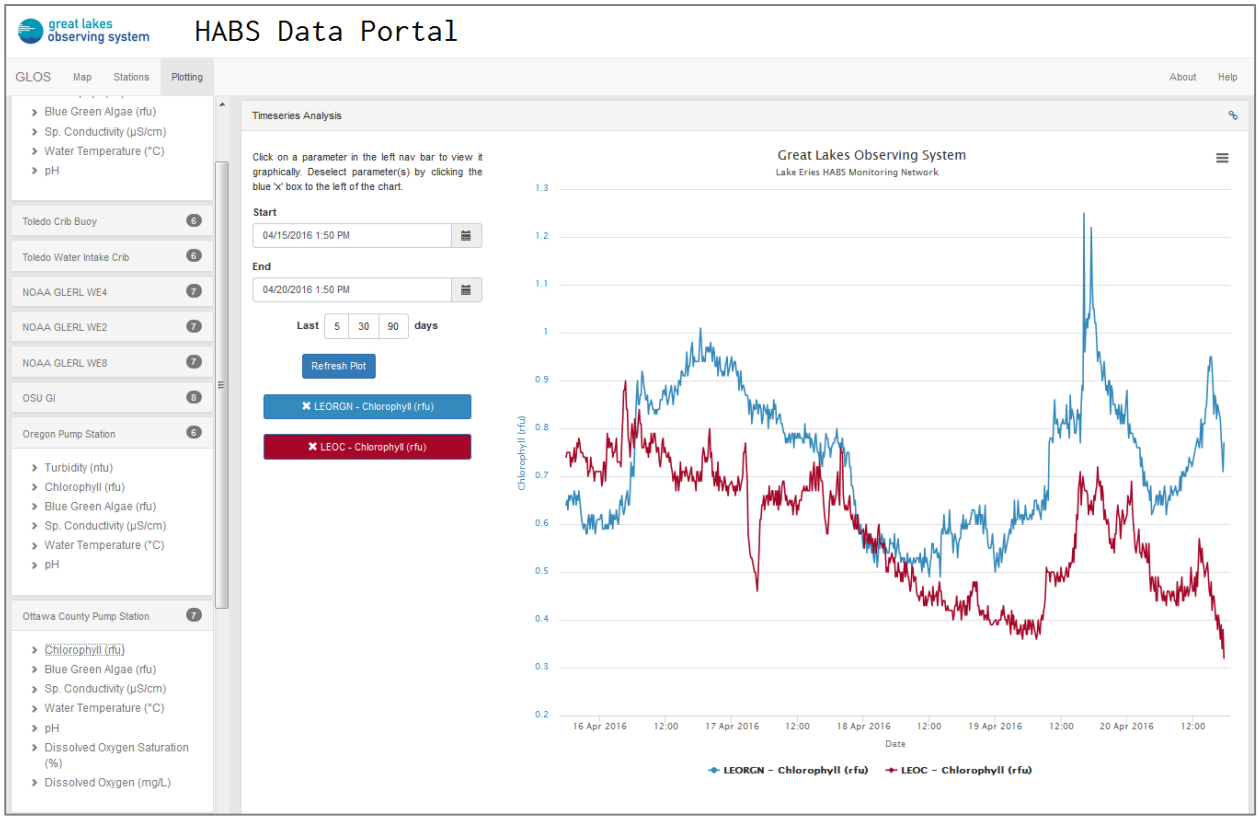
a. Map View



b) Stations View



c) Multi-Station or Parameter Plotting

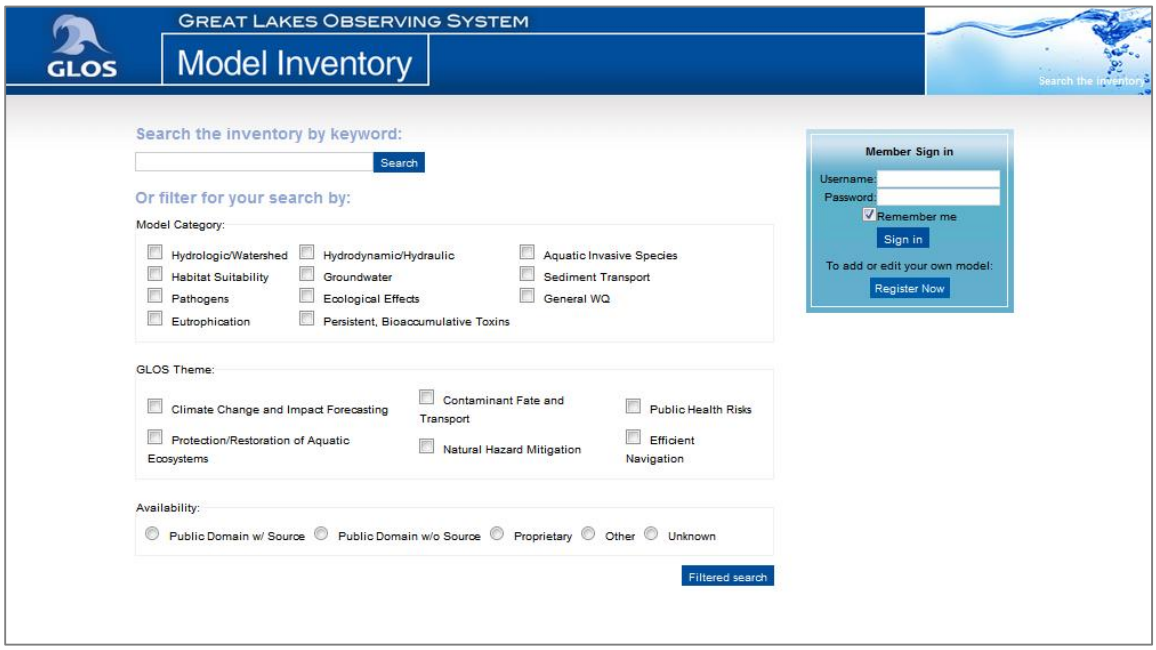


F. Discovery tools

1. Great Lakes Model Inventory

The GLOS Great Lakes Model Inventory (http://data.glos.us/glmi/pub/) allows users to search for models, applications, people, and organizations that address ecosystem health, marine operations, public health and water security, and measuring and adapting to climate change. Reviews of modeling and assessment tools are provided and some direct access to models. With content developed and managed by the Great Lakes modeling community, the inventory serves as a dynamic tool to facilitate information sharing and promote a regional modeling community of practice.

Users simply type a keyword into the search box. Additional features allow them to refine the search with categories using advanced search. A list of matching models and applications along with short descriptions are generated and then the user may click on a model to find out more information about it, including general characteristics, strengths and weaknesses, data requirements, contact information and references, as well as a list of applications that use the model.



2. GeoNetwork (GLOS Metadata Catalog)

The Great Lakes Metadata Catalog (http://data.glos.us/metadata) is a collaborative project with GEO-Great Lakes. This GeoNetwork 2.10.8 site contains a searchable catalog of information about geographically referenced data for the Great Lakes region, in particular GIS and remote sensing data. It features searchable descriptive metadata contributed by Catalog partners, allowing users to identify useful maps, datasets and other information products and find out how to obtain them. While users can go directly to this catalog via a web browser, the ISO19119/19139 curated data are also harvested by the GLOS Data Portal to allow users search access to pertinent metadata.

The GLOS Metadata Catalog is a service of the Great Lakes Observing System, the Great Lakes Information Network and partners throughout the region, including:

- International Upper Great Lakes Study
- MI DNR Institute for Fisheries Research
- Great Lakes Commission
- Environment Canada
- U.S. Geological Survey

G. Statistics

GLOS obtains metrics via Google Analytics and regularly monitors usage statistics to determine whether various tools are meeting their desired function and reaching the right population/user profile.