DMAC CONCEPT OF OPERATIONS
Near-Term Design of the Great Lakes Observing System Enterprise Architecture

DRAFT

June 30, 2011

Prepared for:
NOAA-GLERL
Contract Number:
WC133R-10-CN-0350

By:
Applied Science Associates
LimnoTech

LimnoTech
Water|Environment|Scientists|Engineers

DMAC Concept of Operations
Near-Term Design of the Great Lakes Observing System Enterprise Architecture

DRAFT
June 30, 2011

Prepared for:

NOAA-GLERL
(Contract Number: WC133R-10-CN-0350)

By:
Applied Science Associates
Narragansett, RI

LimnoTech
Ann Arbor, MI

TABLE OF CONTENTS

Appendix A. Details of Data Access

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

CZM – Coastal Zone Management

DIF – Directory Interchange Format (NASA GCMD)

DIF – Data Integration Framework (NOAA IOOS)

DMAC – Data Management and Communications

FGDC – Federal Geographic Data Committee

GCMD – Global Change Master Directory

GEOSS - Global Earth Observation System of Systems

GIS – Geographic Information System

GOOS – Global Ocean Observing System

GOS – Geospatial One Stop

IEOS – U.S. Integrated Earth Observation System

ISO – International Organization for Standardization

IOOS - Integrated Ocean Observing System

KML – Keyhole Markup Language

NASA – National Aeronautics and Space Administration

NMFS- National Marine Fisheries Service

NOAA – National Oceanic and Atmospheric Administration

OGC – Open Geospatial Consortium

OOI - Ocean Observatories Initiative

OPeNDAP - Open-source Project for a Network Data Access Protocol

SOA – Service Oriented Architecture

SOS – Sensor Observation Service

THREDDS - Thematic Real-time Environmental Distributed Data Services

TDS – THREDDS Data Server

SDE – Spatial Database Engine

USGS- United States Geological Survey

XML – eXtensible Mark-up Language

WMS – Web Mapping Service

WCS- Web Coverage Services

1. INTRODUCTION

1.1 BACKGROUND

This document describes the concept of operations for DMAC in the Great Lakes Basin. It reviews the current state of data management systems and focuses on the need to connect disparate data to decision makers. It is a follow up on the January 2011 technical memorandum titled "Current State of Data Management in Support of Observing Systems" that described at a high level a number of existing programs within federal agencies, universities, and private industry. These included:

- NOAA IOOS Data Integration Framework (DIF)—a framework to improve the management and delivery of an initial subset of ocean observations. The DIF will establish the technical infrastructure, standards, and protocols needed to improve delivery of the defined core variables.
- Ocean Observatories Initiative (OOI) — National Science Foundation's contribution to the U.S. Integrated Ocean Observing System (IOOS). This initiative looks to discoveries enabled by new technologies to construct a long-term network infrastructure of science-driven sensor systems to measure the physical, chemical, geological and biological variables in the ocean and seafloor.
- Lockheed Martin, Integrated Ocean Observing System (IOOS): Conceptual Design, 2006
- Raytheon, Integrated Ocean Observing System (IOOS): Conceptual Design, 2006
- U.S Navy Metoc Data Management
- U.S Coast Guard Environmental Data Server (EDS)
- Hydrology Community

This report is more technical and focuses on the actual operations of systems. The project team has analyzed existing data management and delivery systems where available to evaluate what data is discoverable, is it accessible and in what formats or services, and has considered user needs from all stakeholder perspectives.

Because of its centrality to the entire observing system (Figure 1), the DMAC concept of operations serves in many ways to define the concept of operations for the entire observing system. Data providers, data managers, and data consumers all must interact with the DMAC in order to make the observations useful; the success of their experience with the DMAC will determine the success of the overall enterprise.
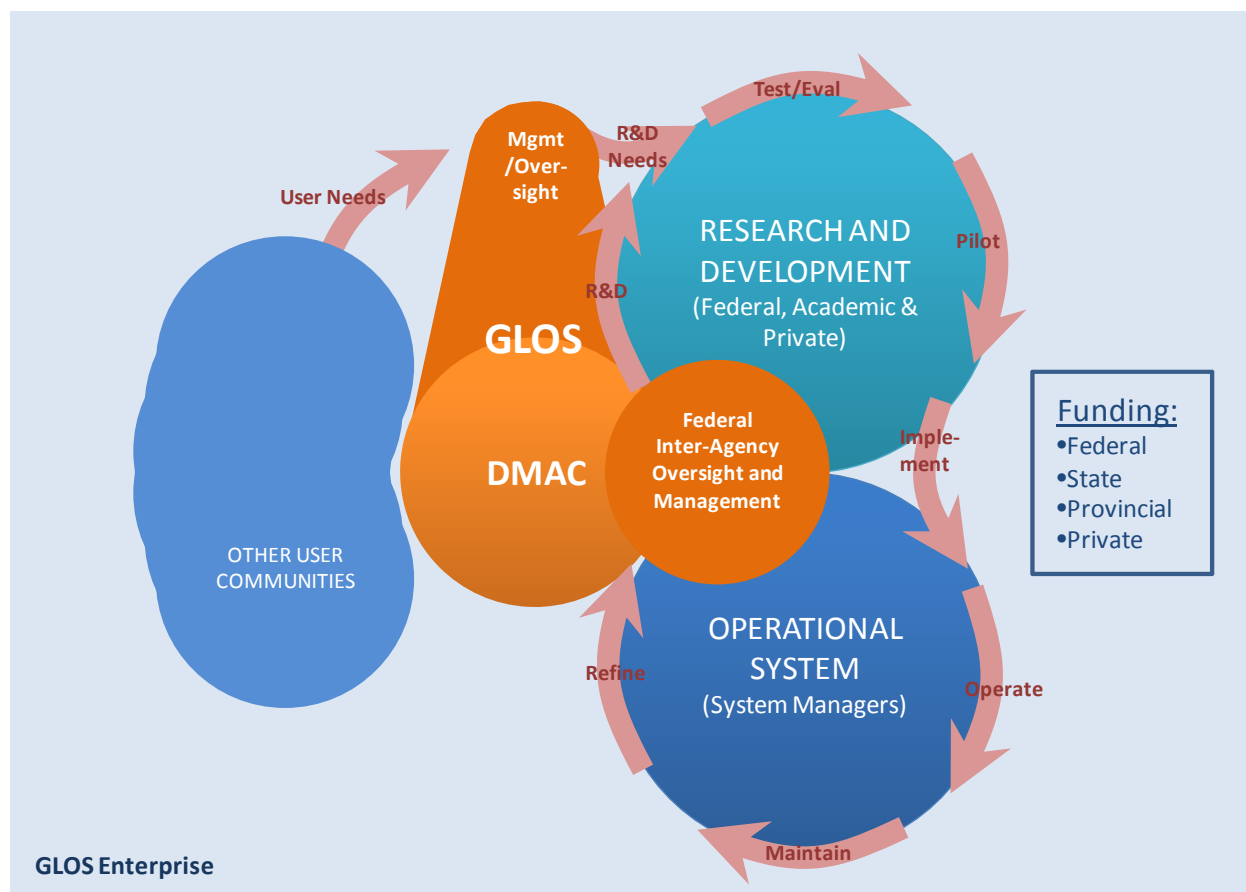
Figure 1. GLOS Enterprise System Management, Development and User Framework

1.2 PURPOSE

This report outlines details of data access for the numerous data sources that link to observations within the Great Lakes basin, and discusses the structure of particular database and modeling systems. The report also describes the structure of a GLOSEA web portal that was built to integrate existing databases and data streams in order to evaluate actual data availability and access. This practical application provides access to static GIS data, climatic or synthetic data converted to GIS, and real-time observation and model data.

The portal allowed the investigators to study the current challenges and data disparities in the current state, and to describe a target state along with a methodology that describes the research, development, testing, evaluation, and operational support steps necessary to arrive at a fully operational system in the near term (by 2015) that is consistent with the recommendations presented in the end-to-end and trade studies.

2. CURRENT STATE

2.1 OVERVIEW

An ocean data network or observing system is an infrastructure of data, systems, services, and tools that allow a variety of users including the public, coastal managers, and research scientists to access "live" and archived data related to coastal and ocean management. This may include maps, observations, and model data. The data management and communications (DMAC), often referred to in recent years as cyberinfrastructure, is a component of an integrated ocean observing system (IOOS). The DMAC system allows for integration of these disparate data feeds to provide data, information and analysis to the broad range of users.

Many reports have been prepared on this topic in recent years that suggest that the societal benefits of ocean observing are several hundreds of millions of dollars per year and recommend "Cost efficient and effective means of communicating the information derived from the ocean observations to users in a timely manner" and that users should be able to "effectively incorporate the information into their decisions."

2.2 USERS AND USER NEEDS

In the Great Lakes Basin, observing system needs are tied to a wide array of organizational goals, including those of the IOOS program, of the Great Lakes Observing System (GLOS), the Great Lakes Restoration Initiative (GLRI) and the State of the Lakes Ecosystem Conference (SOLEC):

- Improve predictions of climate change and weather and their effects on coastal communities and the nation (IOOS 1)
- Improve the safety and efficiency of maritime operations (IOOS 2)
- Mitigate the effects of natural hazards more effectively (IOOS 3)
- Improve national and homeland security (IOOS 4)
- Reduce public health risks (IOOS 5)
- Protect and restore healthy coastal ecosystems more effectively (IOOS 6)
- Enable the sustained use of ocean and coastal resources (Great Lakes) (IOOS 7)
- Improve early identification of climate change impacts on the thermal structure and chemistry of the Great Lakes (GLOS 1)
- Reduce risks of contaminated water supplies and improve predictive capabilities to protect public use of bathing beaches (GLOS 2)
- Enhance understanding of nutrient dynamics, algal blooms, and other factors adversely affecting a viable fishery (GLOS 3)

- Reduce loss of life and property damage to commercial navigation and recreational boating, while increasing economic efficiencies of commercial navigation operations (GLOS 4)
- Toxic Substances and Areas of Concern: Measuring Progress and Assessing New Toxic Threats (GLRI 1)
- Invasive Species: Establish early Detection and Rapid Response Capability (GLRI 2)
- Nearshore Health and Non-Point Source pollution: Generate critical information for protecting nearshore health (GLRI 3)
- Habitat and Wildlife Protection: Identify, inventory, and track progress on Great Lakes Habitats, including coastal wetlands restoration (GLRI 4)
- Accountability, Monitoring, Evaluation, Communication, and Partnerships: Measure and evaluate the health of the Great Lakes Ecosystem using the best available science (GLRI 5)
- Atmospheric Deposition of Toxics and General Air Quality (SOLEC Indicator 1);
- Nearshore and Offshore Nutrient Concentrations (SOLEC Indicator 2);
- Phytoplankton and Benthic Algae (SOLEC Indicator 3);
- Suspended Sediment available for coastal beach nourishment (SOLEC Indicator 4);
- Water Level Fluctuations (SOLEC Indicator 5);
- Climate Change Indicators (SOLEC Indicator 6); and
- Long-term Change Analysis (SOLEC Indicator 7).

These goals and indicators are discussed, along with attendant user needs, in other work products from this project, especially in the User Needs section (Section 3) of the Master Design Report and in Technical Memorandum 6, "Catalogue of User Needs."

Review of the goals and user needs clearly indicate that data collected from observation platforms will not address all important issues. In addition, users will be looking for sophisticated data products, including results from analytical and simulation models, to support their activities in the Great Lakes.

In this project, three primary classes of users have been identified:

- **Data providers** interact with the DMAC system manually and automatically. Manual interactions consist of communication with DMAC staff to negotiate data transfer protocols and implementation of the protocols. Automated interactions transfer observed data or model results to the DMAC infrastructure.
- **Data managers** are responsible for the administration and operation of the DMAC infrastructure. In addition to day-to-day activities maintaining the DMAC infrastructure and improving reliability and performance, the data managers interact with data providers and data consumers to negotiate data transfer protocols and to continually improve access and presentation of data.
- **Data consumers** use the DMAC to access data for a number of purposes, for example:
  - o Researchers access data for use in detailed analyses, including modeling of environmental processes

      o Managers use forecast data to make operational decisions about activities on the lake that may be impacted by extreme conditions, such as high winds or waves, or about policies that will affect the ecosystem on a long-term scale.

      o Members of the public use data to guide recreational decisions and plan lakeside/on-lake activities.

Within the Great Lakes community, there is often overlap between the classes of users. In particular, data providers are sometimes also data consumers who ingest observing data and synthesize new information via models that is then also made available for access. Also, in the highly distributed current state of the observing system, data providers often play the role of data managers as they handle and distribute their own data.

2.3 DMAC COMPONENTS

In the IOOS realm, a DMAC system is typically described by a series of components that can be summarized as:

- Storage and Data Formats
- Catalogs, Data Discovery, Metadata, and Vocabularies
- Quality Control
- Data Sharing/Delivery
- Data Products

For the most part, data managers within the IOOS community are focused on data once it has "reached shore" and this report, consistent with this view, looks at managing data from the base station or data center and does not consider the management of data and communication/telemetry protocols between the sensor and the receiving stations.

This section provides a high-level summary of the status of the DMAC components within GLOS both at the regional and national level. The later implementation section will describe suggested strategies for these components.

2.3.1 Storage and Data Formats

As described in the next section, there is a wide variety of formats with which data are stored, that can generally be defined by the category of data and the domain (science specialty) it originates from. Table 1 shows a preliminary breakout of storage formats for data. A related topic is archive options: within IOOS, archiving is generally discussed as a separate topic and is managed by federal archive data centers, although some regions do implement localized archiving activities.

Table 1. Data Format and Storage by Broad Category

Data Category	Data Format/Storage
Time-Static GIS Layers e.g. Coastline, bathymetry, topography, pipelines.	Dominated by relational databases, such as Postgres, SQL Server, and Oracle. This data is often served with map servers such as ESRI ArcGIS Server and Geoserver. SHP files and KML are also very common for storage as they are widely supported and easily shared for desktop applications.
Time-varying gridded data e.g. Model results, Satellite data, Radar data	Dominated by use of data formats from the earth science community. HDF is very common for satellite data, NetCDF for ocean and coastal models, and GRIB for meteorological grids.
Time-varying Observation data – stationary e.g static in-situ observations from a current meter or stream gauge	For the majority of data managers, relational databases are the storage mechanism for this data. There is a new trend to evaluate use of NetCDF data to manage this data, but it is immature in terms of the standards and tools available to share and consume this data category.
Time-varying Observation data – non-stationary e.g. Cruise samples, gliders	This data category has the largest scatter of data management techniques. The data is generally post processed and not available in real-time. The dominant storage mechanism appears to be Excel and text files for raw data, however, many regions do then process these files into relational databases. There are initiatives on-going to store and mange this data in NetCDF but it is not mature yet.

2.3.2 Catalogs, Data Discovery, Metadata, and Vocabularies

These categories are natural companions that are generally low on the data management priority list and can usefully be considered together. Despite sounding straightforward, successful implementation of these components is extremely difficult and time-consuming:

Catalogs – There remains a lack of a central national catalog system where ocean observing assets can be registered and data can be retrieved. The IOOS Obs Registry is in development and does have many in-situ assets in its catalog but it is used to access data only by a small community, Other catalogs such as NASA's Global Change Master Directory (GCMD), Geospatial One Stop, and GeoNetwork are used, especially for GIS data, but the catalogs remain splintered and there is difficulty in getting data into the catalogs and no central portal to search all catalogs. GLOS currently uses GeoNetwork for catalog management.

Data Discovery – Closely tied to catalogs, this encompasses how metadata (and data) are exposed so that they may be found. There are numerous domain catalogs such as GCMD, data.gov, and the Ocean Biogeographic Information System (OBIS), but no central "Google" style search where one portal can find all related data.

Metadata – There are recommended federal metadata standards, primarily the FGDC and ISO standards. Implementation of these metadata records is a challenge both in terms of the difficulty of persuading data providers to complete the records, and the lack of good tools to manage metadata. Tools such as ESRI's ArcCatalog , NOAA's Mermaid, and EPA Metadata Editor tools do make it easier for users to manually enter metadata records. Another challenge is that the different data storage formats described above use different methods; while the GIS community may use a tool like ArcCatalog to generate metadata files, users of NetCDF are asked to enter the data in the global attributes of the file.

Vocabularies – This is another topic that has been discussed at length within IOOS (e.g how does one refer to Sea Surface Temperature – SST, Ocean_Temp, etc.). There are a number of defined and shared vocabularies in place including the Marine Metadata Interoperability (MMI) project, but the challenge is similar to that of metadata compliance; it is often not considered as data is developed and is difficult to coordinate and enforce with multiple distributed data providers.

2.3.3 Quality Control

There are numerous quality control procedures available. The one most commonly referred to in the IOOS and ocean community is the Quality Assurance of Real Time Oceanographic Data (QARTOD) project, an inter-agency effort to provide quality assurance and control (qa/qc) procedures. The QARTOD project has primarily focused on wave data, with a parallel effort for HF radar data, so it does not provide mature guidelines or cookbooks to manage the wide variety of data the observing system is expected to handle. EPA have extensive QA/QC processes and GLOS are using those as a guideline.

In summary, a wide variety of implementation levels for QA/QC procedures is available, some manual and some automated. NDBC has implemented a variety of QA/QC protocols to manage their real-time data network; while the different IOOS regions implement different levels of QA/QC, it is in general not rigorous or "operational" at this stage.

2.3.4 Data Sharing/Delivery

Generally, effective designs for sharing data are based on a services oriented architecture (SOA), a suite of web services that provides data through different delivery methods that are based on standard REST protocols that define the communication between the clients and the server. The reality of the current status is that not all data is shared using web services. Table 2 summarizes the most common data sharing/delivery standards, many of which are maintained by the Open Geospatial Consortium (OGC).

Table 2. Common Data Sharing and Delivery Standards

SOS - Sensor Observation Service	A protocol for sharing primarily point observation data. A number of agencies including NOAA and USGS offer data via SOS. It still struggles to accept wide acceptance as a. the standard has been evolving with different payload protocols, b. performance issues related to delivering large datasets, c, lack of robust server and client software. It is the IOOS recommended standard for sharing point observation data.
WaterML – came from Cuahsi developments- developed by Hydrology community	WaterML 2.0 is evolving quite rapidly, it is yet to become a widely used standard for "operational" systems
WMS – Web Map Service	Very widely used for sharing map data, essentially images generated on a server. Broad server support from ESRI, Geoserver and many other opensource solutions as well as strong client-side support. ncWMS extension to THREDDS provides WMS support for gridded data.
WCS – Web Coverage Service	Provides "grid" services. Reasonable server support for WCS , but client software is less dominant, primarily due to the large variation in payload options
WFS – Web Feature Service	Provides access to "feature" data. Reasonable server support, but requires an intelligent client that can parse and interpret the payload.
ESRI Services	ESRI ArcGIS Server provides a wide range of services for delivering images, features and geoprocessing. ESRI client and server software supports their proprietary services as well as the OGC standards mentioned above.
OPeNDAP - Open-source Project for a Network Data Access Protocol	Represents a software kit as well as the DAP protocol for sharing data. Widely used in earth sciences for sharing and delivering large gridded datasets, including NetCDF, GRIB, and HDF. **Thredds Data Server (TDS)** is a server that delivers data using DAP and also provides catalog and other services. **HYRAX** is another implementation of an Opendap server. This is the IOOS recommended standard for sharing gridded data.
KML - Keyhole Markup Language	Originally developed by Google and now an OGC standard, it is very widely used as an XML-based file that is relatively simple to generate and is supported by Google Earth, Google Maps, and other GIS software including ESRI software.

Not all data in the current observing system is shared using web services, so migration towards a complete SOA-based design will improve data access and interoperability. However, when designing a data sharing system, it is important to remember that standards are of no use unless a user can consume them and retrieve necessary data. Ultimately, the same data will likely need to be shared using different methods for different clients (users). For example, a modeler using Matlab may be very comfortable retrieving wide-area data using OpenDAP, but a web developer building map-based data products may prefer a GIS-friendly feed of the same data. This required flexibility leads to a design based on a single data storage system (or systems – with a central storage system for each different data category) and a suite of data gateways to serve different users.

2.3.5 Data Products

The concept of a "product" differs within the IOOS community; it can cover a specific data package for a user (e.g. a specified data package for NDBC), to sophisticated data analysis tools often included in a web portal. The particular concept of "thematic" data products that serve a particular theme or user group (e.g Fisheries, Renewable Energy, Beach Users), is considered one of the biggest challenges in the IOOS community because of the subjective analysis of what the user needs really are. Despite the completion of numerous user needs studies, each IOOS region has implemented its own suite of data products built with different technologies and with little consideration of cross-region consistency or interoperability. A certain "build it and they will come" philosophy has led to the development of many individual web portals that provide different access methods, maps, and time series products.

2.4 DATA

The next sections describe existing data providers and data sharing techniques; a key issue in addressing the needs of the Great Lakes is to leverage the wealth of data that is being created by different regional and national agencies. The ability to do this effectively will have a significant impact on the costs for a Great Lakes observing system.

2.4.1 Data Flow

In the current observing system, data are available from different providers that maintain various data structures and services. Figure 2 depicts the multiplicity of these providers and variety of methods through which users may access the providers' data. These national or regional agencies generally operate independently of one another, although they store complementary data that could be integrated into a rationalized framework that provides access to all sources. The different communities, such as hydrology, meteorology, biology, oceanography, often use their own data management systems and protocols. A number of efforts, including this project, IOOS, and OOI, are attempting to provide infrastructure or designs that will allow these data to be merged and offer gateways to provide uniform access to these disparate data sets.
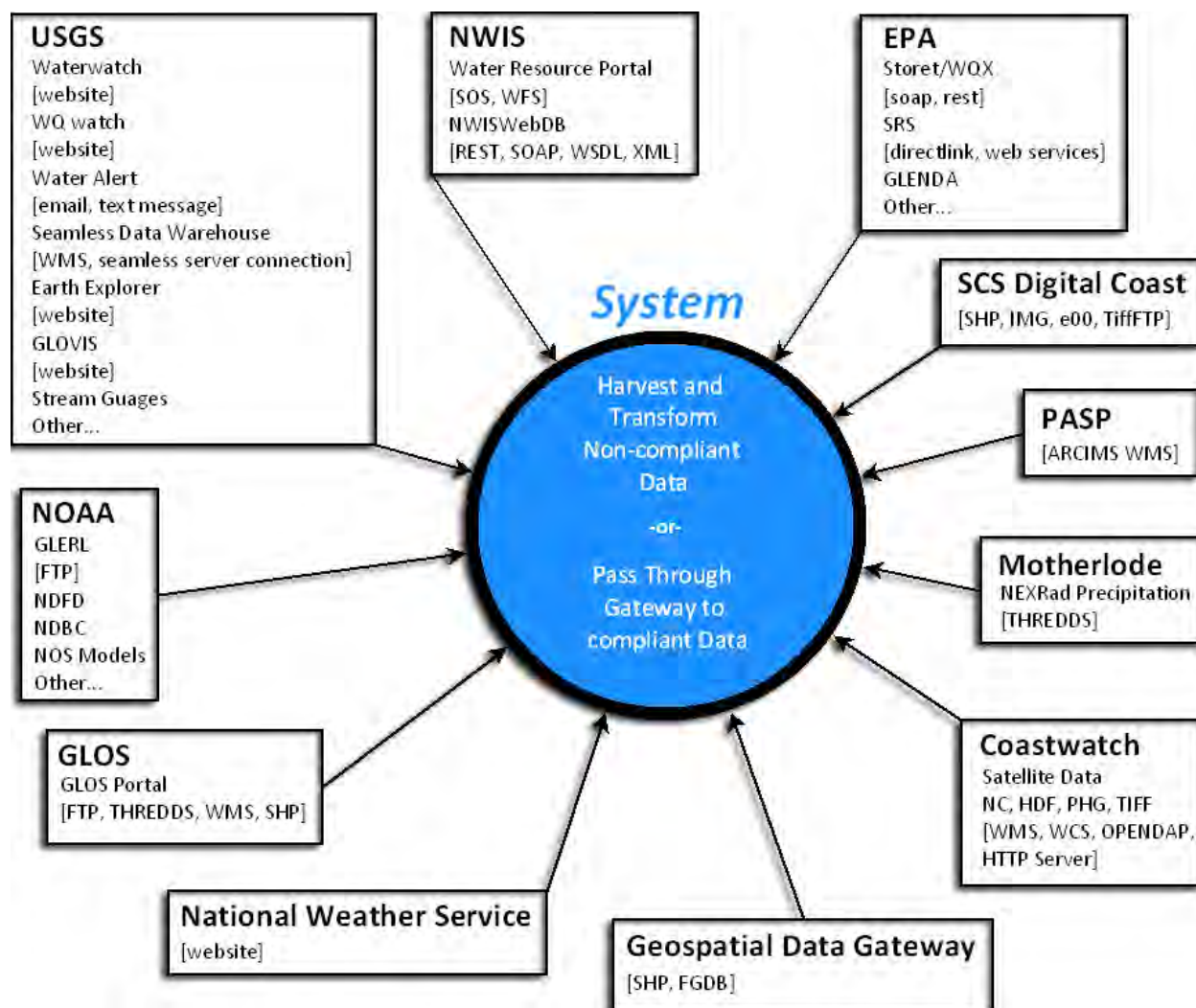
USGS
Waterwatch
[website]
WQ watch
[website]
Water Alert
[email, text message]
Seamless Data Warehouse
[WMS, seamless server connection]
Earth Explorer
[website]
GLOVIS
[website]
Stream Guages
Other...

NWIS
Water Resource Portal
[SOS, WFS]
NWISWebDB
[REST, SOAP, WSDL, XML]

EPA
Storet/WQX
[soap, rest]
SRS
[directlink, web services]
GLENDA
Other...

System

Harvest and Transform Non-compliant Data
-or-
Pass Through Gateway to compliant Data

SCS Digital Coast
[SHP, IMG, e00, TiffFTP]

PASP
[ARCIMS WMS]

Motherlode
NEXRad Precipitation
[THREDDS]

NOAA
GLERL
[FTP]
NDFD
NDBC
NOS Models
Other...

Coastwatch
Satellite Data
NC, HDF, PHG, TIFF
[WMS, WCS, OPENDAP, HTTP Server]

GLOS
GLOS Portal
[FTP, THREDDS, WMS, SHP]

National Weather Service
[website]

Geospatial Data Gateway
[SHP, FGDB]

Figure 2. Data Sources and Access Methods

An example of this is the Water Quality Data Exchange; the USGS and EPA have begun taking steps toward this goal by building a common suite of web services to share water quality and hydrological data.

2.4.2 Data Access by Providers

This section summarizes the various methods of data access and highlights the challenges of integrating data into one system. While many sources provide data access through on-site downloads, this access is not always readily obtainable by modelers and data users interested in real time or near real time products. Even where data sources do provide on-line access to data, there are often a variety of services that are not always readily exploitable due to the type of service and other issues that may inhibit data use.

Analysis of the current state of data storage and management identifies gaps where data may exist, but is not easily discoverable or accessible. Gathering information on the relevant sources for the Great Lakes Observing System and the data types, formats, systems, and services is

important to obtain a comprehensive understanding of how public users and the scientific community can incorporate the most applicable and accessible data for a variety of different needs, ranging from beach information for the public to vast amounts of data needed for modeling systems. These data may include in-situ observations (fixed moorings), static geospatial data, remote sensing, observation data from sampling programs, and model data.

The summarization of sources includes data provided from national, regional, and state agencies. For each source, the description outlines important elements regarding the data contents, acquisition and accessibility:

- the online address
- brief description of contents and purpose of data source
- data type
- service available, if applicable
- retrieval method

Where possible, we have attempted to access the data from these sources to identify the quality of data access and the practicality of automating data access. We have looked at the difference between accessing the data through a web site that requires manual user entry versus the ability to programmatically access the data with software so that the data may be linked to applications such as models.

The organization of data sources is based on data provider, and a single provider may contain multiple sources for different data types. Data providers discussed in this section include USGS, EPA, and NOAA, among others, and in some cases multiple data access points are detailed for a single provider's data source.

Data providers are summarized in Table 3, and full information is supplied in the Appendix. A table entry of "No" in the "Connect to Portal" column does not necessarily mean that the data was unavailable, but instead that the data stream was not immediately accessible by the Portal code. It is likely that it is possible to connect to these services with more custom code development.

Table 3. Summary Table of Data Providers

PROVIDER	NAME	WEB SITE	PARAMETERS	TEMPORAL SCALE	ACCESS METHODS	CONNECTED TO PORTAL
USGS	USGS Water Watch portal	http://waterwatch.usgs.gov/new/	Current Streamflow Flood Drought Past Flow/Runoff	Real time, Static	On website Download	No
USGS	USGS Water Quality Watch	http://waterwatch.usgs.gov/wqwatch/	Water Temperature, in °C Specific Conductance, in µS/cm pH, field, in standard units Dissolved Oxygen, in mg/L Turbidity, in NTU or FNU Discharge, in cfs	Real time	No download available	No
USGS	USGS WaterAlert	http://water.usgs.gov/wateralert/	Surface Water Groundwater Water Quality Precipitation	real time	USGS StreaMail - Alert via email, cell phone text message, or handheld device for gage height and streamflow	No
USGS	USGS Seamless Data Warehouse	http://seamless.usgs.gov/services.php	Orthoimagery Elevation Land Cover	Static - near real time	WMS	No
USGS	USGS Earth Explorer	http://edcsns17.cr.usgs.gov/NewEarthExplorer/	Aerial Photography AVHRR Cal/Val Reference Sites Commercial Declassified Data Digital Elevation Digital Line Graphs Digital Maps EO-1	Real time, sensor data	Geotiff; Export metadata as csv, shp, kml - choose the area of interest, address, or specific scene and click on "Data Sets." Select desired satellite products and click on "Results." This will prompt a user login and will lead to a download page.	No

Table 3. Summary Table of Data Providers - Continued

PROVIDER	NAME	WEB SITE	PARAMETERS	TEMPORAL SCALE	ACCESS METHODS	CONNECTED TO PORTAL
USGS	USGS Earth Explorer	http://edcsns17.cr.usgs.gov/NewEarthExplorer/	Forest Carbon Sites Global Fiducials Global Land Survey HCMM Land Cover Landsat Archive Landsat Legacy Landsat MRLC NASA LPDAAC Collections Radar	Real time, sensor data	Geotiff; Export metadata as csv, shp, kml - choose the area of interest, address, or specific scene and click on "Data Sets." Select desired satellite products and click on "Results." This will prompt a user login and will lead to a download page.	No
USGS	USGS Glovis	http://glovis.usgs.gov/	satellite data for individual scenes of imagery	Real time	Geotiff, Jpeg - must be registered user (Free) Select the satellite scene for the chosen sensor and click on "Add to Cart." This opens	No
EPA	EPA STORET/ WQX	http://storetnwis.epa.gov/storetqw/portal.html	Water quality data Biological data physical data	static	SOAP, REST	No
EPA	EPA Substance Registry Services	http://iaspub.epa.gov/sor_internet/registry/substreg/home/overview/home.do	Chemicals Biological organisms Physical properties Miscellaneous objects	static	Web Service in development	No
USGS	NWIS – SOS services for Discharge and Other Sensor Data	http://nwisvaws02.er.usgs.gov/ogc-swie/	Discharge GageHeight Temperature Precipitation Turbidity DO pH	Real time, Static	SOS, WFS	No

Table 3. Summary Table of Data Providers – Continued

PROVIDER	NAME	WEB SITE	PARAMETERS	TEMPORAL SCALE	ACCESS METHODS	CONNECTED TO PORTAL
USGS	NWIS Web Database	http://waterdata.usgs.gov/nwis/	Water Quality for USGS/EPA	Real time, Static	XML, REST, SOAP, WSDL	Available
NOAA	Great Lakes	www.glerl.noaa.gov/res/glcfs	Currents, Winds	Real time	KML, FTP	Available
NOAA	GLERL Great Lakes Advanced Hydrologic Prediction System (AHPS) Products (Experimental)	http://www.glerl.noaa.gov/wr/ahps/curfcst/	Plot of Water Surface Temperature Plot of Over-Lake Air Temperature Plot of Vapor Pressure Plot of Wind Speed Plot of Over-Lake Cloud Cover Plot of Incident Short-Wave Radiation Plot of Reflected Radiation Plot of Net Long-Wave Radiation Flux Plot of Latent Heat Flux Plot of Sensible Heat Flux Plot of Net Surface Flux Plot of Ice Area Plot of Lake Heat Storage Plot of Soil Moisture Storage	Daily updates	FTP	No
NOAA	GLERL Great Lakes Advanced Hydrologic Prediction System (AHPS) Products (Experimental)	http://www.glerl.noaa.gov/wr/ahps/curfcst/	Plot of Surface Moisture Storage Plot of Snowpack Moisture Storage Plot of Total Moisture Storage Total Basin Supply Outflow Mean Lake Level Forecast in tabular form Plot of Over-Lake Runoff Plot of Precipitation Plot of Over-Lake Evaporation Plot of Net Basin Supply	Daily updates	FTP	No

Table 3. Summary Table of Data Providers – Continued

PROVIDER	NAME	WEB SITE	PARAMETERS	TEMPORAL SCALE	ACCESS METHODS	CONNECTED TO PORTAL
GLOS	Great Lakes Observing system	http://glos.us/	Great Lakes Bathymetry Great Lakes CFS nowcast Great Lakes FC Forecast Greatlakes FC Archive from 2010-2006 Significant Observation data, not available via SOS yet, GLOS are working on it.	real time and historic	THREDDS, OPENDAP, WCS, NetcdfSubset, WMS	Available
EPA	Great Lakes Environmental Database (GLENDA)	http://www.epa.gov/glnpo/monitoring/data_proj/glenda/glenda_query_index.html	Sampled water quality survey data, sediment chemistry, Lake Michigan mass balance results, fish tissue chemistry, and fish sample details.	static	On-Site download	Available
Motherlode		http://motherlode.ucar.edu:8080/thredds/catalog.html	THREDDS catalog of modeled and satellite/radar data: NCEP model data, FNMOC model data, Satellite data, radar data	real time, near real time	WMS, WCS, OPENDAP, HTTPServer, NetCDFSub-set	No
NWS	National Weather Service Radar Image	http://radar.weather.gov/Conus/index_lite.php	NEXRad for precipitation and weather	regular updates	On website Download of KML, KMZ, GIF	No
NOAA	CoastWatch	http://coastwatch.noaa.gov/cwn/search/interface.html	Satellite sensor data on SST, reflectance, chlorophyll, true color, and ice coverage.	near real time	WMS, WCS, OPENDAP, HTTPServer	Available
NWS	National Weather Service Doppler Radar Sites	http://radar.weather.gov/	NEXRad for precipitation and weather	near-real time	ArcGIS services, WMS, WFS	No
NIC	National Ice Center	http://www.natice.noaa.gov/products/great_lakes.html	Daily ice concentrations, forecasts, ice analysis grids, charts, and satellite imagery	real time, near real time	query submission or media viewer	No

Table 3. Summary Table of Data Providers – Continued

PROVIDER	NAME	WEB SITE	PARAMETERS	TEMPORAL SCALE	ACCESS METHODS	CONNECTED TO PORTAL
NOAA	CSC Digital Coast	ftp://ftp.csc.noaa.gov/pub/	Rraster coastal data - Benthic, elevation, hydrography, land cover, marine boundaries, imagery, and socioeconomics	static	FTP or Data Access Viewer	No
NOAA	National Ocean Service	http://oceanservice.noaa.gov/	Currents, wind	Real time	OPeNDAP, WMS	Available
NDFD	National Digital Forecast Database	http://www.weather.gov/ndfd/technical.htm	Dewpoint, wind	Real time	XML SOAP	Available
USDA	Geospatial Data Gateway	http://datagateway.nrcs.usda.gov/	GIS data and orthoimagery for each state and county	static	On website Download	No

3. DATA MANAGEMENT TECHNIQUES

This section presents the data management techniques considered in development of a conceptual design for the DMAC system. The techniques are grouped by major category of data source:

- In-situ observation
- Satellite/Radar
- Models
- Sampling Data

The design and implementation of a prototype GLOSEA portal are also presented. The prototype was assembled to assess data gaps and general feasibility for construction of an integrated web portal offering access to disparate data sets.

3.1 IN-SITU OBSERVATIONS

In-situ instrumentation includes sensors and other elements of monitoring networks that are located in the place of and in contact with the phenomena they are measuring. The existing current capabilities that support investigations at basin- and lake-wide scales include: 1) water- and weather-monitoring buoys (NOAA National Data Buoy Center [NDBC] and GLOS, including University-run GLOS buoys), 2) National Weather Service (NWS) meteorological stations, 3) U.S. Geological Service (USGS) stream gauges and nutrient-monitoring sites, 4) NOAA water level stations, 5) the Integrated Atmospheric Deposition Network (IADN), and 10) Environmental Protection Agency (EPA) and Environment Canada (EC) research vessel monitoring. Current regional in-situ instrumentation includes use of the above as well as GLERL Real-time Environmental Coastal Observation Network (RECON) buoys, GLERL HABs sensors and data collection, state and local monitoring, and university observations.

In-situ data are very valuable but have limited geospatial and temporal coverage. National and regional agencies collect in-situ measurements for a broad range of phenomena, and each agency structures their monitoring programs and requirements differently. This section discusses approaches to integrate data by the NDBC, USGS, and EPA, which demonstrate different methods of providing data from in-situ instrumentation to users.

3.1.1 NDBC

The National Oceanic and Atmospheric Administration (NOAA) National Data Buoy Center (NDBC) is part of the National Weather Service (NWS) and NDBC designs, develops, operates, and maintains a network of data collecting buoys and coastal stations. NDBC manages atmospheric and oceanographic variables measured from monitoring systems. Moored buoys and C-MAN stations are transmitted hourly through NOAA Geostationary Operational

Environmental Satellites (GOES) to a ground receiving facility at Wallops Island, VA, operated by the NOAA National Environmental Satellite, Data, and Information Service (NESDIS). Some stations report via commercial low earth orbiting satellites. The satellite reports are immediately relayed to the NWS Telecommunications Gateway (NWSTG) in Silver Spring, MD.

NDBC also serves as a data assembly center for receiving, quality controlling, and disseminating measurement data from other stations owned and maintained by non-federal regional ocean observing systems, including the IOOS regional associations (RAs). NDBC play a key role in the IOOS community as a central data warehouse of in-situ physical oceanographic parameters. The vast majority of in-situ physical parameters collected by the IOOS regions is delivered to NDBC where it is centralized with data collected from other federal programs, (e.g National Estuarine Research Reserve System - NERRS).

The connection to NDBC is quite simple, NDBC define an XML-based data transfer file that is created by the regional data providers. NDBC then gather this data routinely and integrate the data into a centralized Oracle database. A few of the larger data providers to NDBC use an older methodology; a "modem kit" provided by NDBC that facilitates delivery of data to the center; the majority of regional providers use the XML-based file approach.

NDBC make the data available in a variety of formats/services from the data store including SOS, KML, ASCII, and NetCDF. A new initiative is underway at NDBC to store observation data as NetCDF4 and make available through a THREDDS server. One of the key advantages for data providers in IOOS is that data provided to NDBC is then delivered to the GTS system, one of the upcoming requirements for the IOOS regional data providers. NDBC do not generally manage chemical and biological parameters, but could expand their system to do that. They also do not generally manage gridded data, although they do handle gridded model that is used for data comparison analysis, although not in real-time.

3.1.2 USGS/EPA Data Exchange

The USGS maintains a broad system of monitoring sensors that measure hydrological conditions in water bodies throughout the U.S. These stations update daily stage and streamflow conditions and information on surface water flow, groundwater flow, and precipitation. Real time data are available from the National Water Information System (NWIS) web interface for individual monitoring stations via SOS feeds.

The EPA manages the Storage and Retrieval (STORET), which is an electronic data system for water quality monitoring, physical, and biological. Federal agencies, states, and tribes access data using the water Quality Exchange (WQX) framework, which is a standard set of data elements that partners map for the sake of sharing data.

The USGS and EPA have collaboratively developed the Water Quality Data Exchange to build a common suite of web services to improve sharing of water monitoring data via a common format and terminology. The EPA-managed water quality monitoring systems and USGS-managed hydrological monitoring systems have been integrated into one domain for storage and dissemination. The services provide the ability to combine data from USGS's NWIS and EPA's STORET systems. The USGS water quality web services are interoperable with corresponding

EPA web services and produce data formatted according to the Water Quality Exchange (WQX) Outbound XML schema. WQX uses the technology developed by the National Environmental Information Exchange Network.

3.2 SATELLITE/RADAR

Satellite remote sensing data are generally available free of charge in their basic, raw, moderate-resolution, and synoptic form. Many vendors and warehouses of remote sensing data provide download capabilities for individual image based on customer specified sensor, product, time period, and scene. These data are readily available, however there is little support for support services that provide automatic data updates.

The basic data require transformation through a set of algorithms (and models with algorithms at their core) to generate the derived products that are presented in the requirement tables for the regional, lake-wide and basin-wide scenarios. Prior to implementation, the products derived from remote sensing need to be prioritized. This prioritization should include consideration of algorithm complexity, cost of development as well as importance of observation. Various governments, academic and commercial entities are actively developing the algorithms identified in the table and this public-private partnership is working to advance the state of useful Great Lakes remote sensing science.

3.2.1 National Weather Service Doppler Radar Images

The National Weather Service uses a radar called the WSR-88D, which stands for Weather Surveillance Radar - 1988 Doppler. The WSR-88D is a Doppler radar, which means it detects motions toward or away from the radar and the location of precipitation areas. The National Weather Service provides several different images from the network of Doppler radars, including reflectivity, wind velocity, and precipitation.

NWS offers an enhanced view of NWS Doppler radar images, which allows interactivity or transparency with the display. Called RIDGE (Radar Integrated Display with Geospatial Elements), the radar image can be layered over other elements such as topography maps and weather warnings. For the RIDGE enhanced views, NWS Doppler radar images can be incorporated into GIS software. Each RIDGE radar image is created separately without backgrounds on the image and in a standard geographic projection referenced to North American Datum 1983 (NAD83).

All RIDGE radar images are downloadable via URL for different image types that include:

- Base reflectivity
- Storm relative motion
- Base velocity
- One hour precipitation
- Composite reflectivity
- Storm total precipitation

NWS uses the Integrated Radar Data Services (IRADS) system, which is a program started by the University of Oklahoma in 1999 to demonstrate the capability of transmitting Level II data over the Internet. Level II data refer to "base" data and consist of reflectivity, radial velocity, and spectrum width (velocity variance) "moment" data generated by the radar processor.

Radar data are transmitted from each radar via NWSNet to a server located at one of four associated NWS Regional Headquarters (RHQ) sites. Each RHQ server is linked to a central network and from there the data are transmitted directly to one of three top-tier sites (one of which is IRADS), and to the University of Maryland for subsequent transmission to the National Centers for Environmental Prediction (NCEP).

Files are available for download on the web in GIF format, and NWS offers national and regional mosaics, including for the Great Lakes region. NWS also provides ArcGIS, WMS, and WFS services for RIDGE radar watches and warnings, and ArcGIS and WMS services for radar base reflectivity.

3.2.2 NOAA CoastWatch

NOAA CoastWatch provides imagery and data products from a variety of satellites and sensors and in a variety of formats. It composed of central operations and six regional nodes that makes satellite and oceanographic data available, and which was established to monitor coastal waters using near real time satellite derived products. The central operations and regional manage the data distribution for the CoastWatch products.

Products distributed by CoastWatch include:

- Sea surface temperature (SSTs)
- Ocean color
- Winds
- Currents
- Chlorophyll
- Reflectance.

SSTs support meteorological weather predictions, while biologists can use ocean color radiometry data and derived chlorophyll-a and total suspended matter/turbidity products to identify runoff plumes and blooms and also predict HABs.

CoastWatch provides experimental, pre-operational, and operational products. Experimental products are not routinely processed or are being developed by an integrated product team. Pre-operation refers to products that undergo initial processing and validation, while operational products are fully transitioned to operations and continually functional. All products undergo calibration, validation, and QA/QC.

CoastWatch products are accessible to users through FTP, and HTTP, and THREDDS/OPENDAP server. Data are available for user download using a search criteria based on region, time period, satellite, sensor, and product. Supported formats include CoastWatch HDF, TIFF, PNG, ESRI Shapefiles, while data products on THREDDS server are accessible via

WMS and WCS services. Metadata are available in Plain Text, HTML, XML, DIF, and SGML format.

3.3 MODELS

Models play a key role in synthesizing diverse data sets and providing decision support information to meet a variety of user needs. These different needs may require data that spans very long periods (e.g climate change) to near real time data to support emergency situations such as oil spills and search and rescue. The models must be both site- and problem-specific to address Great Lakes issues. Models within the EA framework range in complexity from the elementary (e.g. aggregation of point measurements into a geospatial map) to the sophisticated, such as a 3D circulation model for a given region of one of the Great Lakes. Integrating different data types from a variety of sources is critical for the development and maintenance of models due to the importance of input parameters and assumptions in model design that influence outputs and in turn management decisions.

3.3.1 NOAA Great Lakes Research Environmental Laboratory (GLERL)

The NOAA Great Lakes Environmental Research Laboratory (GLERL) was formed to provide a focus for NOAA's environmental and ecosystem research in the Great Lakes and coastal marine environments.

GLERL sponsors the Great Lakes Coastal Forecasting System (GLCFS), which is a system of computerized models that can simulate and predict the three-dimensional structure of currents, temperatures, water level fluctuations, wind, waves, ice, and sediments in the Great Lakes. Developed by GLERL and the Ohio State University and supported by the National Weather Service, operational implementation of this utility is accomplished by NOAA's Coast Survey Development Laboratory and Center for Operational Oceanographic Products and Services (see sidebar, GLOFS). These predictions provide timely information to lake carriers, mariners, port managers, emergency response teams and recreational boaters.

GLERL nowcasts and forecasts are available on the website for each individual lake or for the region as a whole in KML files available for download. Wind and current data are provided to the US Coast Guard Operations Systems Center and are published to their FTP site as zipped current and wind files.

3.3.2 National Digital Forecast Database (NDFD)

The National Digital Forecast Database (NDFD) consists of gridded forecasts of various weather elements (e.g., cloud cover, maximum temperature, etc). NDFD contains a seamless mosaic of digital forecasts from NWS field offices working in collaboration with the National Centers for Environmental Prediction (NCEP). The database is available for members of the public to use in creating text, graphic, gridded and image products of their own. Over time, NWS will offer a wider array of gridded forecast elements and a larger set of graphical presentations.

You can access NDFD elements via ftp, http, XML, or web browser. The NDFD contains weather elements that include apparent temperature, dewpoint, maximum temperature, minimum

temperature, sky cover, wind gust, wind speed, wind direction, snow amount, relative humidity, precipitation, and wave height, among others. Data are in grid format and updated every 3 hours.

NDFD elements are available via XML web service. This service provides users and partners the ability to request data over the internet and receive the information back in an XML format. The request/response process is made possible by the NDFD XML SOAP server.

Figure 3. SOAP Access to NDFD Data

The first step to using the web service is to create a SOAP client. The client creates and sends the SOAP request to the server. The request sent by the client invokes one of the server functions, of which there are currently nine with associated parameters.

NDFD is an excellent source of multiple parameters available in relatively high resolution on a timely basis. Surprisingly, it is not used often by the regional associations, perhaps as it does not provide offshore data, but for the Great Lakes, it could be of great value to the observing system.

3.3.3 NOAA National Ocean Service (NOS)

The Great Lakes Operational Forecast System (GLOFS) is a NOAA automated model-based prediction system aimed at providing improved predictions (guidance) of water levels, water currents and water temperatures in the 5 Great Lakes (Erie, Michigan, Superior, Huron and Ontario) for the commercial, recreation, and emergency response communities. GLOFS generates hourly nowcast guidance (analyses) and four times daily forecast guidance (out to 30 hours) of total water level, current speed and direction, and water temperature for each of the Great Lakes. The GLOFS predictions will enable users to increase the margin of safety and maximize the efficiency of commerce throughout the Great Lakes.

NOAA sponsors the National Operational Coastal Modeling Program (NOCMP) to develop and operate a national network of Operational Nowcast and Forecast Hydrodynamic Model Systems (called OFS) to support NOAA's mission goals and priorities. An OFS consists of the automated integration of observing system data streams, hydrodynamic model predictions, product dissemination and continuous quality-control monitoring. State-of-the-art numerical hydrodynamic models driven by real-time data and meteorological, oceanographic, and/or river flow rate forecasts will form the core of these end-to-end systems. The OFS performs nowcast and short-term (0 hr. - 48 hr.) forecast predictions of pertinent parameters (e.g., water levels, currents, salinity, temperature, waves) and disseminate them to users.

NOS created the following systems for the Great Lakes region to provide the maritime community with improved short-term predictions of water levels and currents:

- Lake Erie Operational Forecast System (LEOFS)
- Lake Huron Operational Forecast System (LHOFS)
- Lake Michigan Operational Forecast System (LMOFS)
- Lake Ontario Operational Forecast System (LOOFS)
- Lake Superior Operational Forecast System (LSOFS)

These systems provide real time modeled information for water levels, water temperature, winds, and currents. Data are available via THREDDS Data Server and CO-Ops OPeNDAP server. Data are available in NC formats.

3.3.3.a Nowcasts

The hourly nowcast guidance cycles of GLOFS use surface meteorological data from a wide variety of observing platforms on and around the lakes. The different observing platforms include airway stations at airports, Coastal-Marine Automated Network (C- MAN) stations, NOAA/GLERL's mesonet stations, NOS' NWLON stations, and U.S. and Canadian fixed meteorological buoys. The real-time observations are provided courtesy of the National Weather Service's National Centers for Environmental Prediction (NCEP). Data used from these stations includes surface wind speed and direction, cloud cover, air temperature, dew point temperature and surface water temperature, if available. Wind observations taken at heights greater than 33 ft (10 meters) are adjusted to 33 ft based on the stability of the atmosphere over the lakes, then spatially interpolated to the hydrodynamic model grids on one hour increments.

3.3.3.b Forecasts

The four times per day forecast guidance cycles of GLOFS are forced by surface wind (10 m above ground level) and surface air temperature (2 m above ground level) forecast guidance from the National Weather Service's operational North American Mesoscale (NAM) weather forecast model. The original spatial and temporal resolutions of the NAM forecast guidance is 7.5 miles (12 km) and 3 hours, respectively. The NAM forecast guidance is interpolated to the grid resolution of the POM grid for each lake and time resolution of 1 hour intervals. These interpolated surface wind and air temperature fields are used by GLOFS' hydrodynamic models to generate oceanographic forecast guidance for each lake out to 30 hours. The forecast guidance cycle of GLOFS does not use or calculate surface heat flux from NAM output.

All CO-OPS official real-time products, including forecast guidance from the Ocean Service's Operational Forecast System for the Great Lakes and estuaries are monitored by the CO-OPS's Continuous Operational Real-Time Monitoring System (CORMS). CORMS provides data communications, data analysis, system monitoring and notification support to a variety of users, using input from real-time tide, current, water level, other marine environmental sensors, and forecast guidance model systems, which are deployed nationwide in many U.S. ports and waterways. CORMS provides 24 hour a day, 365 days a year monitoring and quality control of sensors and data in order to insure the availability, accuracy, and quality of tide, water level, current, and other marine environmental information.

3.3.4 Large Basin Runoff Model (LBRM)

LBRM estimates runoff within watersheds with specific variables that estimate snowpack, upper soil moisture zone, lower soil moisture zone, groundwater zone moisture, and surface storage. LBRM was developed by the Great Lakes Environmental Research Laboratory (GLERL) to model river and stream systems in the Great Lakes basin and estimate rainfall/runoff relationships.

It is physically based to model good representations of hydrologic processes and is a lumped-parameter model of basin outflow that has a cascade of moisture storages or "tanks," where tank outflows are proportional to tank storage. The model uses precipitation, temperature, and insolation as inputs to determine snow pack (accumulation), snow melt (loss), and net supply. This value of net supply is modeled to represent surface runoff, and infiltration of snow melt into upper and lower soil zones.
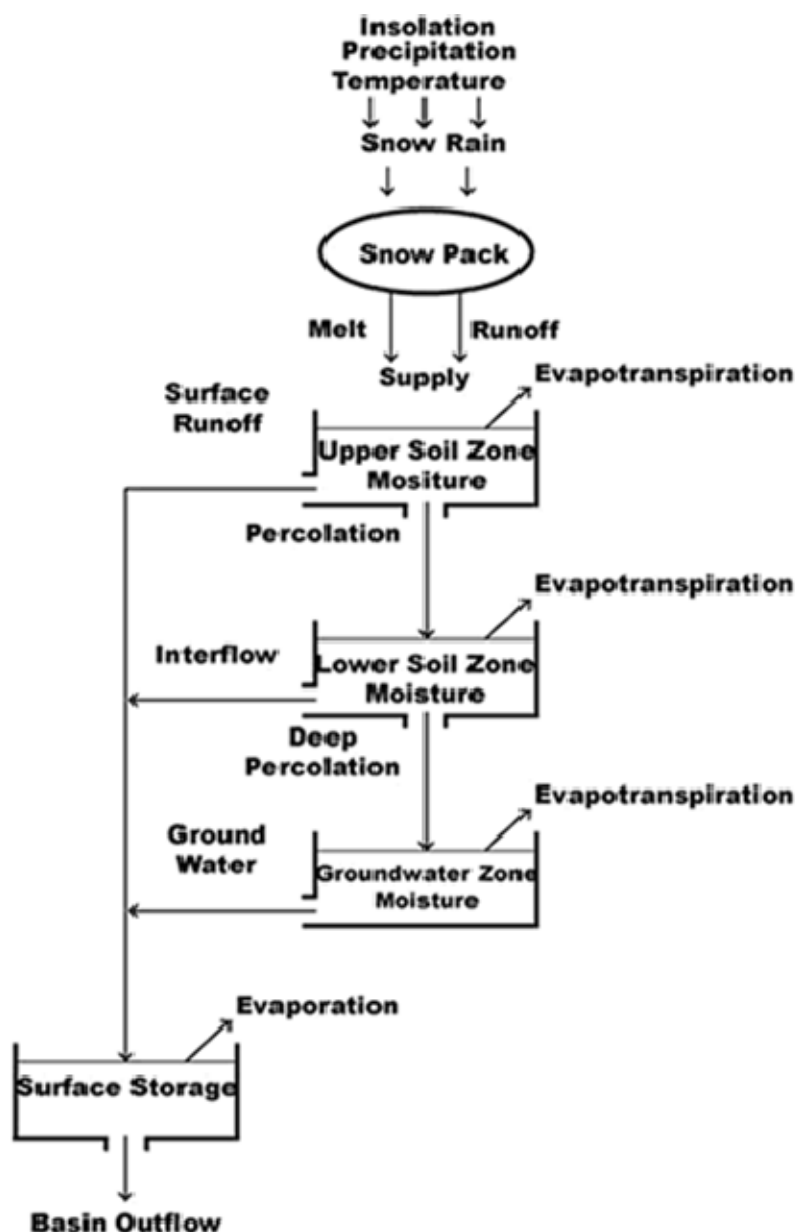
Figure 4. LBRM Conceptual Model

LBRM has relatively simple requirements and provides accurate runoff volumes in the Great Lakes watersheds. LBRM software is available for download on the LBRM website. Sample LRBM data are available on GLERL's FTP site and contain text files of daily LBRM observations. This system will be updated to manage LBRM model results via a server that offers the most recent LBRM data available to modelers. Meteorological data inputs in the US are available the National Climate Data Center (NCDC) and National Environmental Satellite Data and Information Service. Model results are available on a daily basis from the 1948 to 1995.

3.3.5 SPARROW

SPARROW (Spatially Referenced Regressions on Watershed attributes) is a watershed model that was developed by the USGS and estimates contaminant sources and transport in surface water and watersheds. The model estimates a relationship between water quality measurements made from in-situ observations and properties of the watersheds which contain the monitoring stations. SPARROW estimates contaminant flux, concentration, and yield in water bodies and enables decision makers to hypothesize about important contaminant sources and transport processes across large spatial scales.
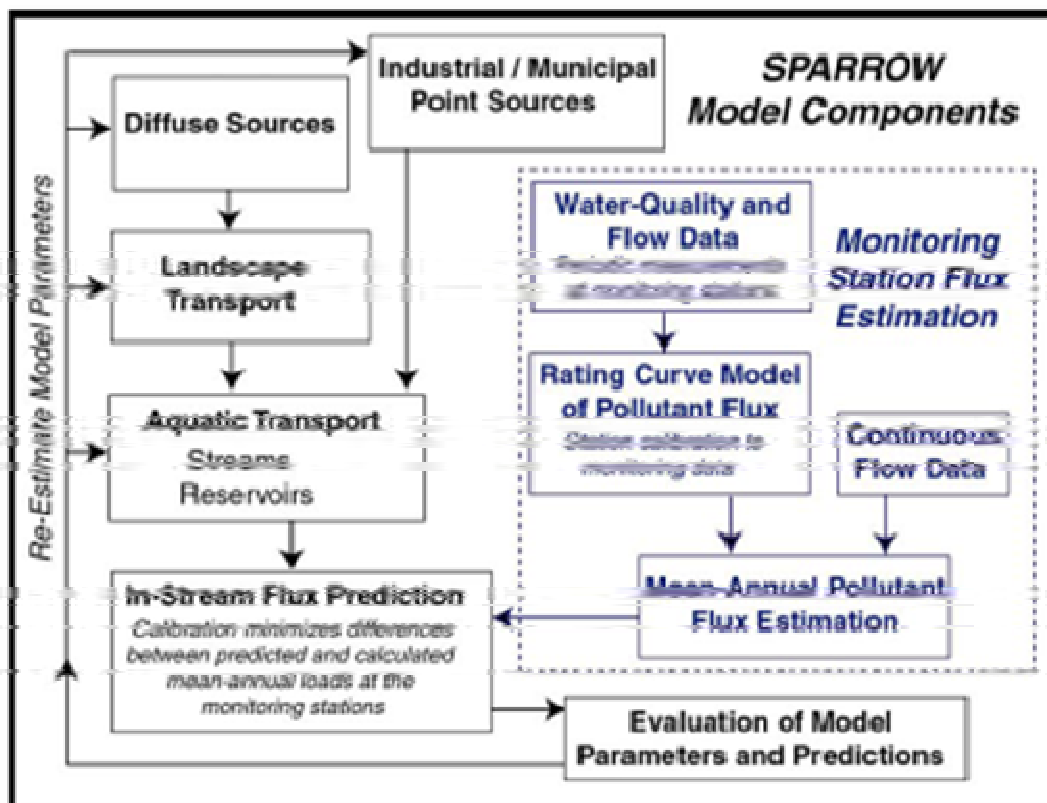
Figure 5. SPARROW Model Components

River reach data available on the USGS SPARROW website. SPARROW's complete spatial coverage is coupled with incomplete temporal coverage since it is limited to estimating annual contaminant loads. The model output is in SAS format and code is written in SAS Macro Language. The USGS are developing a web service interface to SPARROW which will make integration easier.

3.4 SAMPLING DATA

Observational data from non-permanent sampling regimes provide a challenge to data users and the scientific community due to inconsistencies in data collection among regional sampling programs and among different sponsor organizations which hinders data integration. Addressing this inconsistency is part of the broader framework of assessing end to end data connections for data providers and modelers. Attempting to incorporate information from various databases that maintain different structures and require different parameters engenders a need for an integrated database approach where users can provide non-permanent observational data that adhere to certain guidelines. This is particularly the case for sampling programs of environmental data which have specific geospatial, temporal, and physical, chemical, or biological components.

3.4.1 EPA STORET

The EPA hosts the STOrage and RETrieval (STORET) data system to monitor water quality data developed by the EPA. Water resource managers from federal and state agencies, tribes, universities, and volunteer organizations use the STORET Warehouse to submit their water quality data and make it available to the public. STORET requires a certain level of documentation detailing where, when, how, and what was collected for data submissions. All samples contain information on geographic location for latitude, longitude, state, county, and Hydrological Unit Code where data were sample, the time period, data type sampled, and sponsoring organization. Other information is available regarding data sampling, sample analysis, and contacts.

The EPA created the STORET Legacy Data Center to archive historic data from the early 20th century to 1998. All current data uploads are stored in STORET, although users may access and query data contained in STORET Legacy.

STORET uses the Water Quality Exchange (WQX) framework, which uses XML to define schema that enable different organizations to submit data to the warehouse. WQX is based on the Environmental Information Exchange Network, which is a partnership among the EPA, states, and tribes that enables efficient and secure data sharing via the Internet. Partners that use the Exchange Network maintain Network Node servers, which automatically submit requests and update data from other information partners. The Network uses XML to describe data and provide a standard through which anyone can partake in the data exchange by translating data to a common structure and format. Partners in the Exchange use Data Exchange Templates to standardize data sharing.
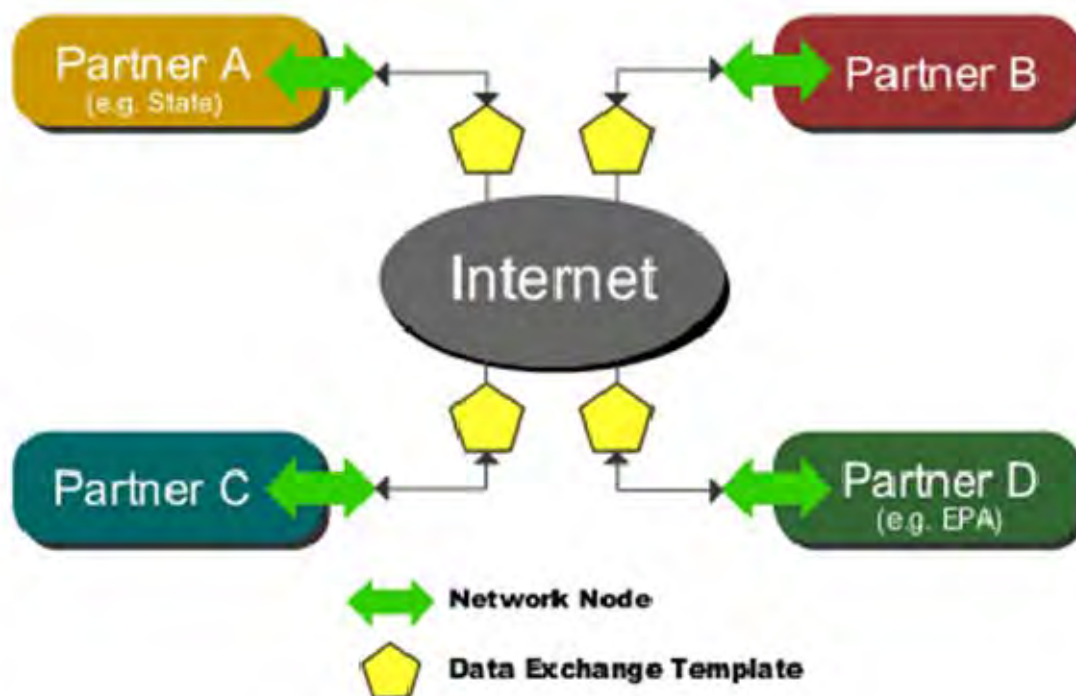
Figure 6. Environmental Information Exchange Network

The Exchange Network supports different communities for air, water, waste, health, natural resources, as well as cross-programs that integrate these networks. Organizations or individuals may request to join specific data exchange networks, or existing partners may build a new network.

Once data is uploaded by a provider, the Central Data Exchange (CDX) compresses it before storage in the EPA STORET warehouse. The CDX supports fast and efficient submissions of sampling data from EPA data providers through one central point of access. Once uploaded to the warehouse, it is accessible by different query or web services and available for public use.
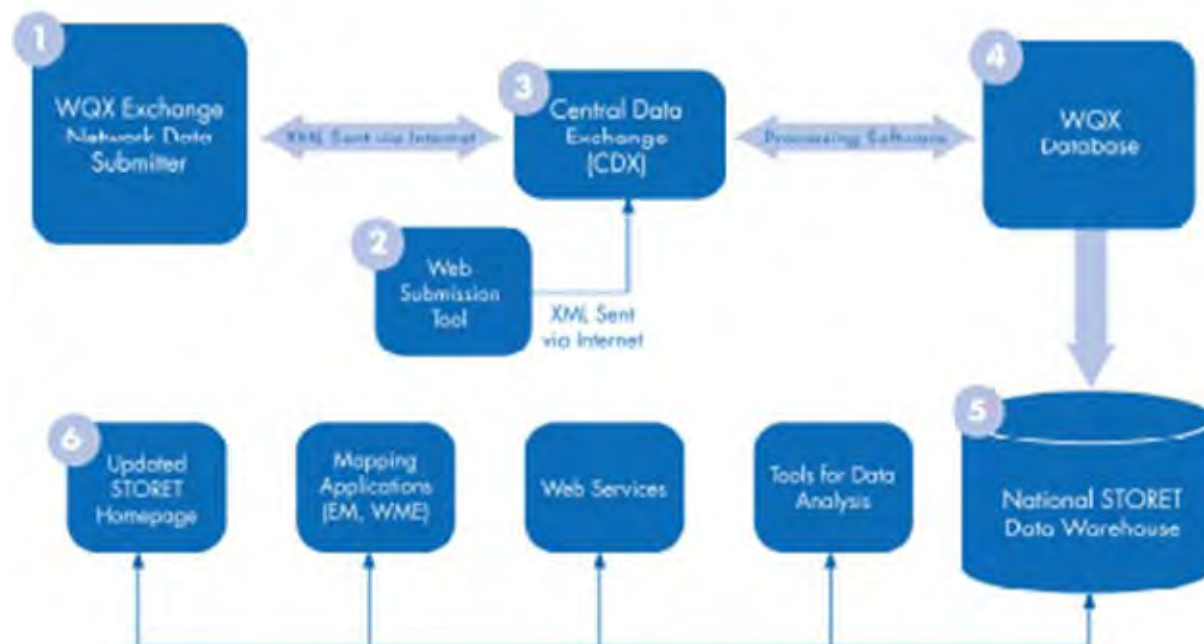
Figure 7. USEPA STORET/WQX Components

3.4.1.a STORET Query Application

Users can access STORET data using the query application. Supported file formats for data are WQX-XML, CSV, Tab separated, MS Excel, and KML. Users search for data based on the required information that providers must enter in order to upload the data to STORET. These include a bounding box of geographic coordinates that contain the area of interest, a certain distance radius from a known geographic location, state, county, site type, sponsor organization ID, site ID, Hydrologic Unit Code, sample media, characteristic group, characteristic, activity ID, as well as start and end dates. Simply entering one parameter may either return no results or may contain results that are too large to process in one request.

3.4.1.b Web Services

The EPA provides STORET Warehouse web services for specific elements in the database with service endpoints and output schema of the web service, which may be queried using REST or SOAP services. STORET also provides ArcGIS map service endpoints for water monitoring locations.

WQX web services are available with domains for WQX submissions and retrievals to clarify nomenclature for public use. A "domain-values" web service can be queried to determine the EPA-supplied values of the domains used for both submission and retrievals. These domain value web services directly access the WQX Domain Tables. Domain list is extensive and should be accessed via the STORET WQX Domain Value Services and Downloads site.

3.4.2 Great Lakes Environmental Database (GLENDA)

GLENDA is a relational database system that stores results from the Lake Michigan Mass Balance and other environmental sampling projects. GLENDA supports data entry, storage, and access capabilities for modelers and other data users to incorporate sample data into their decision support systems. This system maintains field and lab reporting requirements, reference tables, and a physical database with software tools that enable database access.

Data uploaded to GLENDA contain information for the following four subject areas:

- Project/Organization
- Station/Location
- Field monitoring activities
- Results

The project/organization subject area includes data about the project, contract, cooperative agreement or grant that sponsored data collection, sponsoring organization, and contacts. Station/location refers to the actual sampling station or location where data were monitored, geopositional technique and accuracy, and political feature information such as counties and states. Field monitoring activities include data on station visits, observations or measurements taken for sampling, and comments from field or lab work. Lastly, the results include the analytical information, measurement units, quality assurance codes, and correction factors.

GLENDA is accessible through the EPA's Central Data Exchange (CDX), which is the point of entry on the Environmental Information Exchange Network (Exchange Network) and which is touched upon in the previous section. CDX allows users to submit data through a central access point and submit data in a variety of formats and exchange data via web services. GLENDA data are accessible by logging in to the CDX, which requires free registration for all users. Registration requires users to select account profiles for the region or program of interest. Once registered, users can search for data for the following query topics:

- Water quality survey data
- Sediment chemistry
- Lake Michigan Mass Balance results
- Fish tissue chemistry
- Fish sample details

Other search parameters include the lake, station, time period, and analyte code for the topic of interest. Analyte codes include parameters for phosphorous loads, temperature, pH, turbidity, and water clarity, among others. Users search a combination of these parameters to return results. Similar to the STORET system, users may enter a set of different parmaters that return no search results. When data are returned, they are available for download as an XLS file. However the process of searching for data is not transparent and provides a hurdle for users seeking available sampling data.

3.4.3 Scribe

The Scribe software tool was developed by the EPA's Environmental Response Team (ERT) to manage environmental sampling, observational and monitoring field data. Users collect data from the field or lab and incorporate it into the Scribe framework, from where data can be exported and integrated with models, GIS, and other software programs.
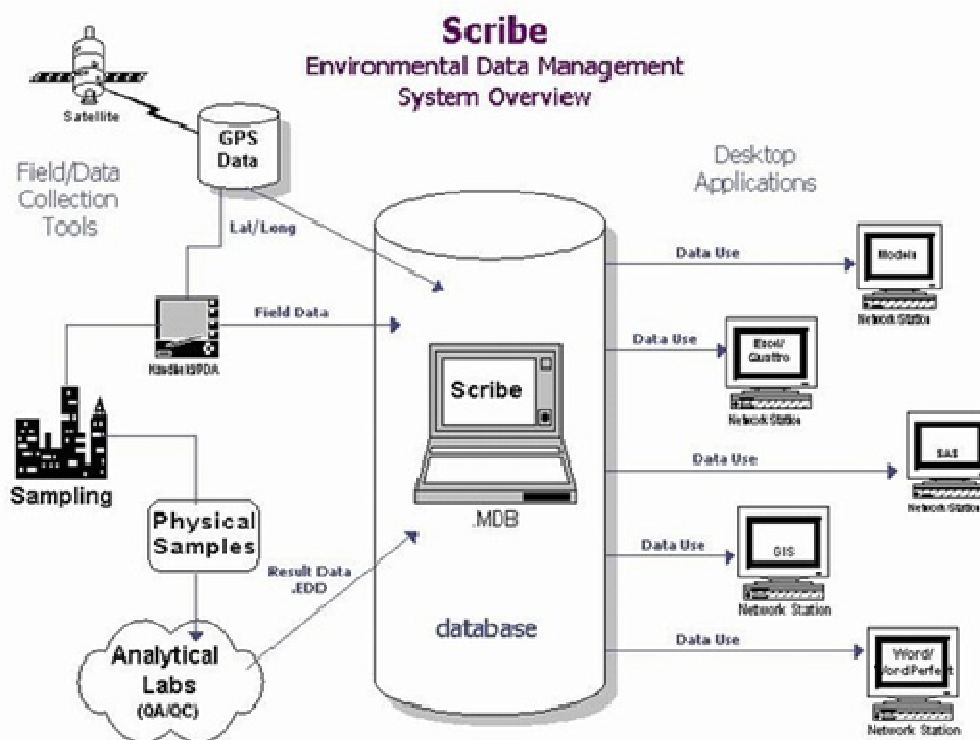
Figure 8. Scribe Overview

Data uploaded to Scribe is used to generate labels for collected samples, Chain of Custody, and Analytical Lab Results data reports, record field sampling, store spatial data (GPS), track field samples to labs, import lab results EDD, and query databases and produce report.

Scribe has a flexible user interface to manage and query data and contains an organized structure for Planning, Sampling, Sample Management, and Custom Data Views. Sampling programs include water, soil, air, and biota sampling. Custom Data Views have querying capabilities which provide analysis of sampling data for inclusion in reports and deliverables. Scribe's import tool supports a variety of data formats, including SEDD (Staged Electronic Data Deliverable), FORMS II Lite, and Rapid Assessment Tool (RAT), and supports export of databases to GIS, Oracle Database Applications, and spreadsheets (EPA ERT 2011). End users wishing to access the databases can download XLS, CSV, SHP, or KML files.

Scribe.Net is a web based publishing subscription tool that serves an important function of sharing Scribe projects across Scribe desktop or Oracle/SQL clients. Scribe projects are available to other authorized desktop/enterprise user subscribers once databases are published to a web

server (Figure 7). Publishing a project requires a user to enter an identification and password, and upon publishing the user releases ownership and publishing privileges. These projects are available to other authorized users via a subscription service. Enterprise clients that subscribe to multiple projects continually search the Scribe.NET server for updates, which are therefore automatically received. Desktop clients must re-subscribe to projects to receive updates.
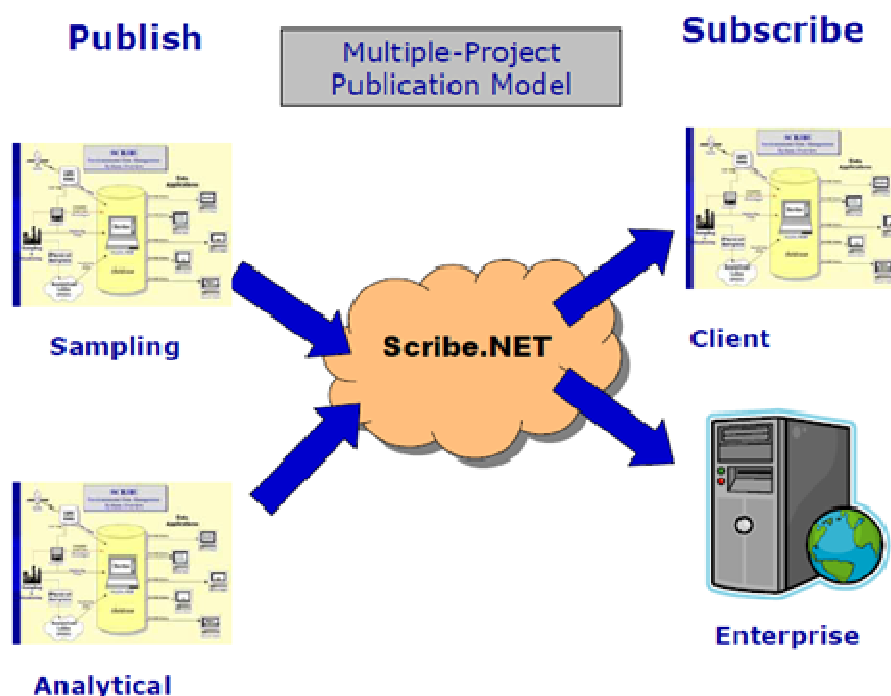
Figure 9. Scribe.NET Publish/Subscribe Model

The Scribe.NET website, which is available from the Scribe Help tab, lists published Scribe projects that detail the project ID, project name, US EPA region, and published data, as well as projects that other users subscribe to. Accessing Scribe.NET requires the user login and password that publishing user must set up upon first using the system.

The Scribe desktop application and Scribe.NET tool provide data users with a method of accessing sampling data from a single server using a structured database that promotes consistency in naming conventions, formatting, and definitions. Users publish and subscribe to projects using the Scribe application interface.

3.4.4 Summary of Sampling Data Programs

These methods of managing sample data demonstrate the need to maintain centralized database and server structures so diverse users can upload and access datasets. EPA STORET and GLENDA use the Exchange Network and CDX to provide data in a common format using a required set of key information. By using XML, the Network overcomes system incompatibilities so that government, tribal, academic, and private users can contribute to and

query database contents. However user accessibility is limited due to a large number of service query parameters and the legacy system of the query application. Scribe approaches the issue of system incompatibility by requiring all users to download the Scribe desktop application, which uses the Scribe.NET server to enable users to search data based on certain key indicators that identify individual projects.

Reviewing these programs identifies several items that can increase the efficiency of data flow between users and providers.

1) Sample data must be managed from a centralized database system to ensure consistent formatting and accessibility
2) there must be easy methods, e.g. upload an XLS file, to allow data managers to get data into the system ,
3) User requested data must be downloadable in an efficient manner and in a common format, such as XLS or CSV, and
4) Data accessibility must be transparent throughout the search and download process.

One possibility is to apply a spatial component that enables users to visually comprehend what data are available based on geographic location. Smaller data warehouses enable users to identify a geographic location by a bounding rectangle that triggers a list of available data within the selected region. While this may not be possible for large national repositories, the search process requires simplification to maximize effective use of the system.

3.5 END TO END INTEGRATION

Presenting environmental data through an online forum often requires users to search amongst a series of internal or external web links in pursuit of data. Providing practically and visually accessible data is important to increase the efficiency of database maintenance and support a creditable data source. A system should effectively demonstrate what data are available and how data are accessible.

ASA has developed a Great Lakes Observing System Enterprise Architecture (GLOS-EA) web client that seeks to demonstrate end to end data connections by integrating in-situ observations, GIS, remote sensing, and modeled data into a viewer for Great Lakes region. The client is purely a prototype to evaluate the ability to access disparate data and services. It was built using Flex and incorporates ArcGIS, WMS, and SOS services to represent real time, near real time, and static data. The client integrates data from satellite sensors, models, static and real time online services, and decision support systems for three domains: atmosphere, tributaries, and open water. A full summary of GLOS-EA portal data is available in the appendix.
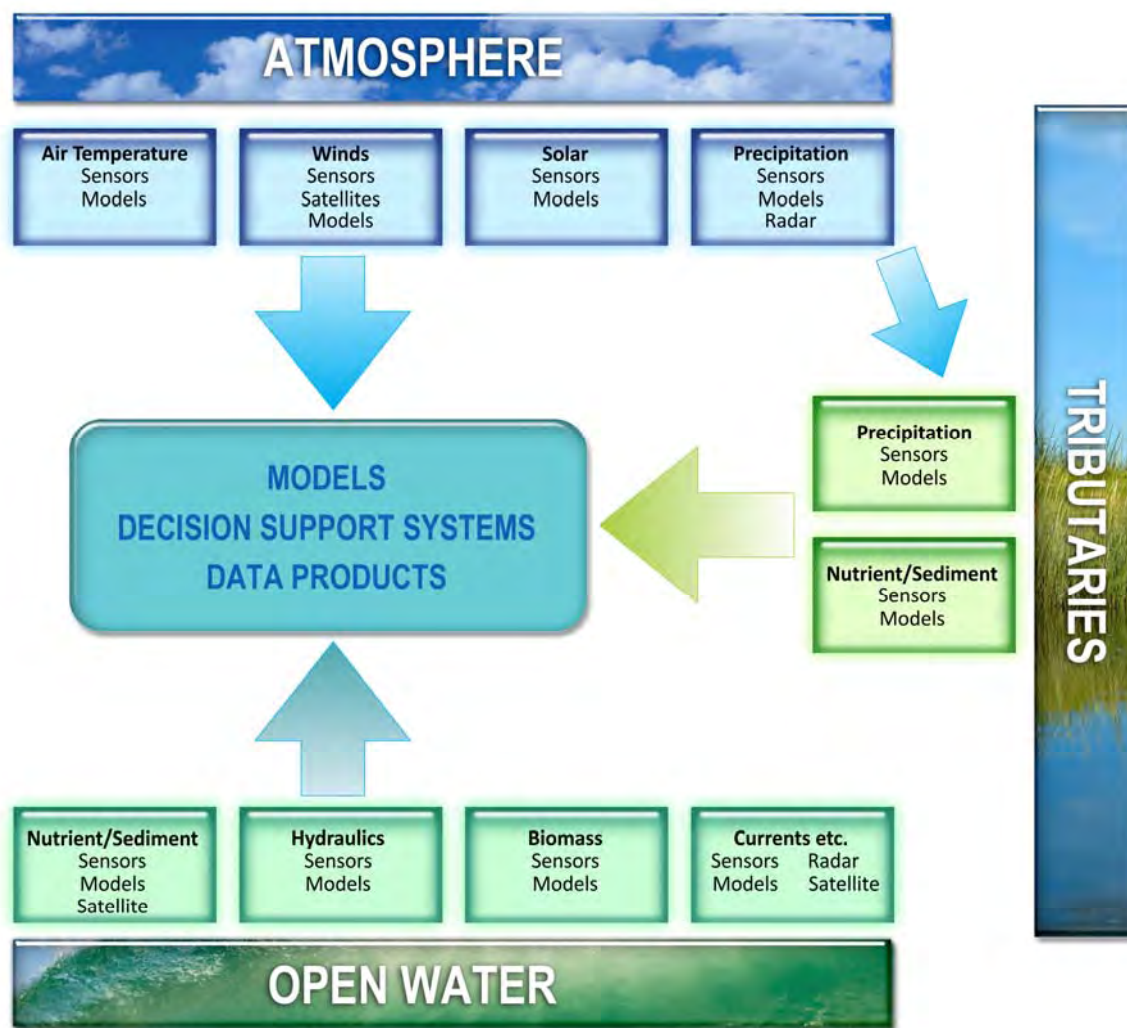
Figure 10. GLOS-EA Conceptual Schematic

The GLOS-EA portal contains tools that enable optimal data viewing and different querying capabilities.

These tools are presented using a collection of widgets that enable users to manage the viewer and connect to data services:

- General tools include Layer List, Identify, Legend, Bookmark, Draw, and Print tools.
- The Buoy Connector and Environmental Common Operational Picture Tool (ECOP) widgets are specialized tools that connect directly to modeled and observed meteorological and hydrodynamic data via WMS and SOS services.
- The Catalog Viewer contains information on the different data sources discussed in this report. It also contains datasets of sampled biological or water quality monitoring data processed from XLS or TXT files to demonstrate what output is available from sampling programs (e.g. GLENDA, EPA STORET).

GIS data were provided by LimnoTech and include hydrography, infrastructure, land cover, and political feature data. These static datasets are published as mapping services to ArcGIS Server and accessed in the client via the Layer List. The table does not include all GIS datasets available in the portal but outlines important layers that capture the nature and scope of the data. Three particular infrastructure map services of note are Water Quality Monitoring Stations, Watershed Water Level Gage stations, and Real Time Monitoring Stations, which include NDBC buoy stations.

The Layer List also currently houses a section for scientific data as an example of what is available to incorporate into the portal. These scientific datasets include modeled SPARROW and LBRM output, along with aggregated MODIS Aqua chlorophyll and reflectance products obtained from CoastWatch's THREDDS server. Modeled SPARROW output displays river reaches using a river reach shapefile obtained from the USGS SPARROW website with corresponding annual phosphorous loadings around the Great Lakes basin. The LBRM output is processed from sample text files on the GLERL ftp site and appended to watershed in the Great Lakes basin. Currently data are only presented for 10 U.S. watersheds around Lake Superior and represents daily LBRM results for a single year period.

The Buoy Connector widget is used to connect to in-situ and modeled data using SOS services. The Buoy Connector is an Adobe Flex implementation and visualization tool exposing SOS services through time and property filtering. Using SOS (Sensor Observation Service) protocol standards, the Connector widget provides both a spatial representation of real-time buoy locations and customizable query capabilities exposing station properties, time range and values. The tool enables a user to query a time series for observed phenomena at NDBC stations that monitor air temperature, wind speed, air pressure, water surface height, or currents.

Spatial points are requested in real-time using GetCapabilities request which returns station x,y locations. The user can select a station and time which then generates a time graph using the GetObservation request. For each available property, a graph is created using the specified date range.

Users may hover the cursor over the time series graph to obtain detailed information on buoy observations at a particular point in time.
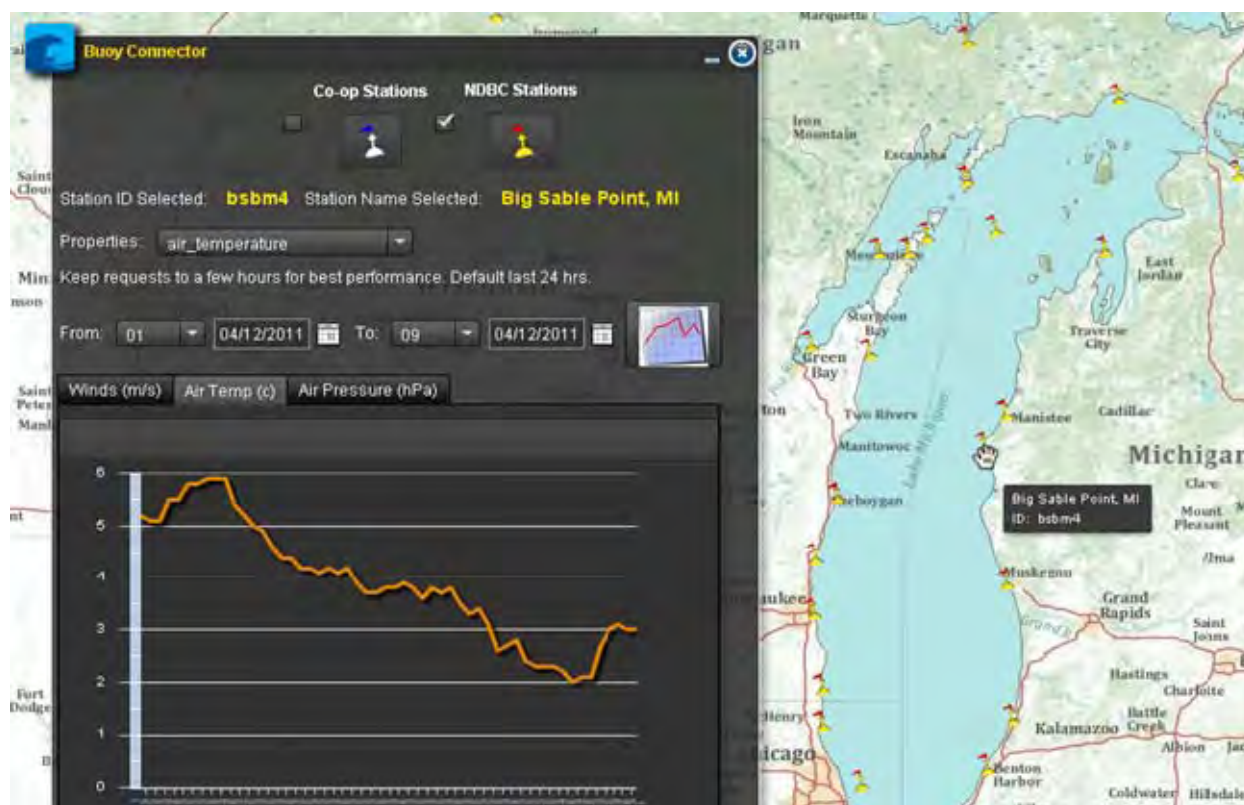
Figure 11. Buoy Connector Tool for GLOS-EA Portal

The ECOP (Environmental Common Operational) tool connects to services from the U.S. Coast Guard Environmental Data Server (EDS). The EDS provides live access to a variety of environmental data from models and observations. The EDS hosts a Catalog server, which is a series of data collection services governed by a timing service that automatically collects, processes and stores data from a number of disparate sources according to source availability schedules. The services use metadata stored in a SQL database that tell the catalog server when and from where to collect data. Data are processed daily into EDS and exposed as time-enabled WMS, which are accessed by ECOP.

The Environmental Common Operational Picture is a set of services that allows users through web sites such as this, and other applications, access the live data stored in the EDS. ECOP services are divided into Currents, Winds, and Other, and display remote sensing data and modeled nowcasts and forecasts of hydrodynamic and atmospheric conditions, with a time series capability. ECOP serves as a front-end interface for the user and sends data requests to a Virtual Windows Machine running the a suite of WMS services. This WMS service consumes these requests, pulling any necessary data from a separately hosted production server, returning the appropriate response to the user. These responses may include generating image files representing meteorological or hydrodynamic phenomena, or XML files which contain data for a particular location and time duration. Modeled forecast data are accessed for a three day period from the present. The ECOP is an example of a centralized architecture that harvests disparate model and observation data and makes it available via a set of services from a managed data warehouse.

Figure 12. ECOP Tool for GLOS-EA Portal

Sampling data from GLENDA are available in the Catalog Viewer in CSV format. GLENDA sample data contain multiple records at the same station location and therefore require non-spatial representation within the portal. GLENDA data are from the Lake Michigan Mass Balance lake nutrient results. The project is titled "Lake Nutrients" and each entry include fields for year, month, cruise, visit, station, geographic coordinates, date, sample depth, medium sampled, analyte code, and measurement (i.e. density of analyte measured). This dataset was downloaded and processed from XLS format into CSV, with modifications to improve presentation within the Catalog Viewer. Users may click and download the dataset directly from the Viewer.

The GLOSEA portal addresses basic issues for supporting and maintaining a DMAC system by using data assimilation techniques that integrate different data types and formats. It includes evaluation data connections to model data, gridded data, point observation data, real-time observations, and sampling data.

The findings of this demonstration effort are:

- Not all required data for relevant science projects are available programmatically – i.e a few data sources were available with open standard protocols, but many required the writing of custom code to demonstrate integration
- Although there is a lot of data available in a wide variety of formats, there is no central searchable catalog to find relevant data for the Great Lakes

- Near real-time data available from GLOS and the federal data providers is more likely to be provided in an open standard that can be consumed by software clients such as the demonstration portal
- Historical data, especially discrete data from sampling programs and cruises is especially difficult to integrate without manual data conversions

4. DMAC IMPLEMENTATION PLAN

There are numerous documents (see reference list) that discuss high level DMAC implementation recommendations including the IOOS DMAC Concept of Operations document (January 2009); these reports describe the need to use a services oriented architecture (SOA), open standards, focus on QA/QC methods, metadata management and archiving. For many regional associations, the budget is simply not available to tackle the complete array of DMAC challenges. This section presents a practical implementation plan along with associated staffing recommendations.

Although many regional DMAC groups do not strictly follow the recommended steps from the IOOS design documents and some sections may get skipped, this simply means that the regional DMAC focuses on priorities based on reasonable funding options. The core issue is a balance between a practical solution to meet the needs of the region versus the implementation of processes and standards that can support the "successful move of ocean research and modeling from a local to a national and then to a global capability" (IOOS DMAC CONOPS).

The core aspects of the conceptual design for the recommended DMAC structure are simple:

- The system must deliver data to end-users in a useful form.
- This useful form will vary based on the user, so multiple data formats/services must be made available to meet different user's needs
- Users must be able to easily find and access available data and data products
- Centralization of data management roles to facilitate successful data management and data product generation
- The use of a common data model (CDM) to store data and provide different access methods from the data store
- Scalability – able to increase capacity to address increased data volumes – and expandability – able to increase capabilities to handle new types of data and support new data products.
- Ability to work at regional, whole-lake, and basin-wide scales
- Focus on performance and robustness
- Coordination with federal/international programs
- A design that recognizes and preserves existing data management systems maintained by leading technical and research organizations in the basin
- A design that provides ways to manage and access data across different time scales, from real-time, immediate data to long-term sensing of annual and decadal trends.
- A system that allows for centralization of regional data, but not necessarily all data. The system needs to be able to connect to distributed data managed by mature data centers that deliver data using open standard protocols.

- IOOS DIF Compliance – although there is often discussion of a DMAC system being "IOOS compliant", there is no formal compliance or testing process to test compliance. The reality is that the IOOS guidelines are quite simple – ensure that model (gridded) data is available using the DAP/WCS protocols and in-situ observations are available using the Sensor Observation Service (SOS). Other data management tasks such as metadata, vocabularies, archive, and quality control are recommended but no detailed implementation guidelines are provided and no compliance testing is currently performed.

Review of the current observing system has demonstrated that the existing GLOS core infrastructure, which has shown considerable achievement to date in managing disparate data, developing data products and meeting IOOS standards requirements, is an appropriate foundation for near-term expansion.

We therefore recommend that the observing system should formally adopt the Community Managed System alternative presented in the project Technical Memorandum "Alternatives Development: Data Management and Communications (DMAC)." (Appendix A), and that GLOS should take the lead role within the Community Managed System in organizing and maintaining centralized servers and in providing a data management team. Also, the final design and implementation plan should consider offsite provisioning of necessary server resources in order to improve reliability and minimize potential bandwidth limits, and development of appropriate system redundancy to support high availability of proposed operational models and critical data feeds.
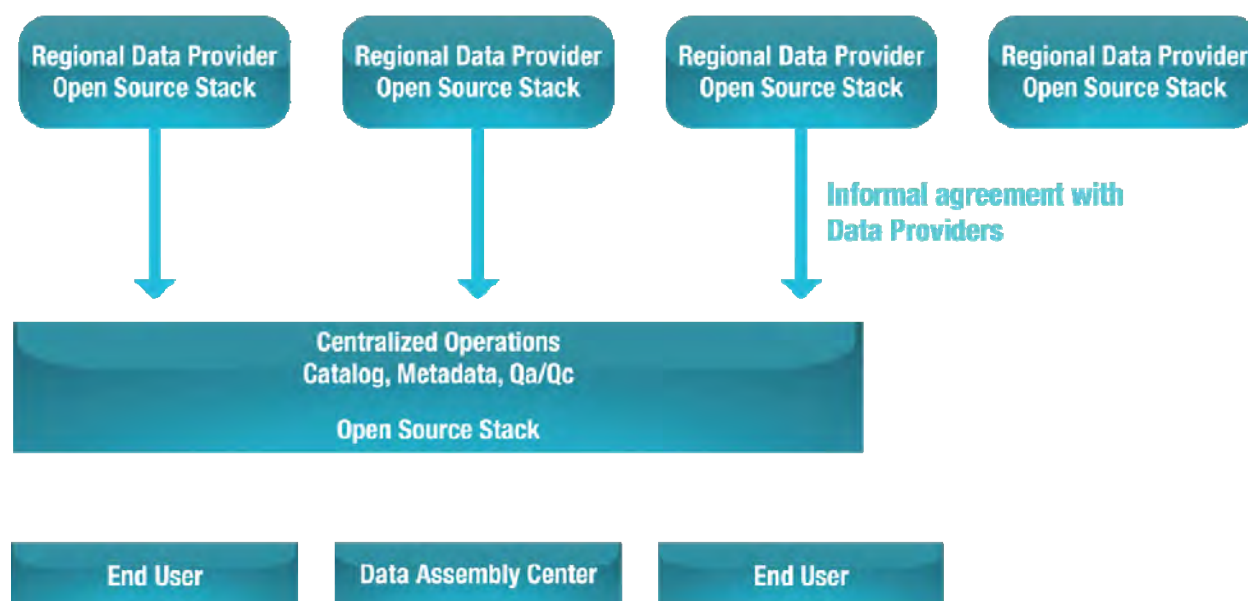
Figure 13. Conceptual design of a Community Managed System

As shown in Figure 13, in a Community Managed System, a central group organizes and maintains all of the data from all providers, and central servers typically containing an open

source technology(s) stack will be managed by a central data management team. The central servers will continually harvest data from all the data providers, who have the responsibility to make their data available to the central server using defined protocols. External data access will be performed through a central gateway. The agreement between the central data center and the data providers is informal with no formal SLA (Service Level Agreement) in place.

Not all data may need to be centralized; data that is readily available from national systems such as NDBC and the Data Exchange may not need to be duplicated based on the end user needs. The Community Managed System may therefore also act as a gateway to distributed systems.

This managed system **does not imply that all data are to be centralized**, but instead just those data that is not readily accessible in compliant formats from reliable providers – data that is not readily accessible and/or is not compliant will be harvested by the central server and transformed and then served using appropriate standards for the data class. The community managed system will also act as a gateway (pass through) to data providers that are compliant/reliable, and in these cases, data will not be duplicated.

4.1 RECOMMENDATIONS

As described in section 2, the DMAC system can be broken into the following components:

- Storage and Data formats
- Catalogs, Data Discovery, Metadata, and Vocabularies
- Quality Control
- Data Sharing/Delivery
- Data Products

Specific recommendations are presented for each of the components.

4.1.1 Storage and Data Formats

This section lists specific recommendations for data format and storage for different categories of data (Table 4) and presents a brief discussion of how cloud computing may be applicable to DMAC.

Cloud Workflow using Amazon EC2

There has been considerable discussion in the data management community about the use of the Cloud for remote storage and compute capabilities supplied by commercial providers such as Amazon. Very few of the regional associations have made use of the Cloud – Maracoos have implemented a THREDDS data server on the Amazon Cloud.

- Store original data as an S3 bucket. Use CloudFront to distribute data effectively. This does NOT provide OPeNDAP or other service access, it is raw file access. Allows anyone to access the raw data very efficiently.

Table 4. Storage and Data formats

Data Category	Data Format/Storage
Time-Static GIS Layers e.g. Coastline, bathymetry, topography, pipelines.	Recommend a central GIS storage system that stores core relevant GIS data and provides a portal connected to a map server. This project developed a prototype using ESRI software which is very efficient and robust and will be provided as a deliverable should GLOS want to implement it. A similar system could be built using an OpenSource technology stack such as PostgreSQL and GeoServer. This GIS data warehouse and portal can be a building block for other data products – as described below.
Time-varying gridded data e.g. Model results, Satellite data, Radar data	The GLOS/GLERL group have extensive modeling experience and have used NetCDF for many years – the continued use of the grid formats supported by the Unidata tools are recommended. This will also allow for integration with OOI-CI as it comes on-line.
Time-varying Observation data – stationary e.g static in-situ observations from a current meter or stream gauge	GLOS are already harvesting data from different sources into a relational database – we recommend this approach and this can be extended quite easily to add new data from other sources.
Time-varying Observation data – non-stationary e.g. Cruise samples, gliders	Recommend a relational database to store such data, data can be loaded from ASCII or Excel files. Explore use of existing framework such as GLENDA or SCRIBE.
Archiving	We recommend an agreement with NODC as a federal archiving facility. It should be decided based on the user needs what level of regional archiving should also take place, as there will be a combination of local/regional archiving as well as the long term archiving at a federal facility.

- Mount each S3 bucket on multiple EC2 instances using PersistantFS, each with READ-ONLY access. These EC2 instances can run anything (TDS/NcWMS/uWMS/Hyrax) all accessing the same core dataset. Load balance the EC2 instances providing data access.
- The EC2 instances can also do data processing (KML generation, Matlab plots, etc) on the datasets and upload back to S3 for distribution through CloudFront. Controlling which EC2 instance processes what and when can be controller through the Simple Queue Service, or through dedicated EC2 instances.

The concept is that no EC2 instance is actually storing any data. They act as both "service" hosts (providing access to the data through different protocols and tools) and processing nodes (various data processing tasks).

Example Costs (all per month)
A small EC2 instance is $62 ($0.08 / hour)
A large EC2 instance is $250 ($0.30 / hour)
S3 storage is $0.10 / GB for 99.99% reliability and $0.15 / GB for 99.999999999999%
S3 is $0.15 / GB transferred out and $0.10 / GB transferred in
Load balancer is $18 ($0.025 / hour)
CloudFront is $0.15 / GB transferred out (not including S3 transfer out costs)

Case 1 - $536.62 / month
10GB of new data per day
Public downloads **10GB** per day of raw output through **CloudFront**
4 small EC2 instances serve out 10GB of data per day via various interfaces (WMS, OPeNDAP)
1 small EC2 instances does processing on the data

Case 2 - $925.19 / month
10GB of new data per day
Public downloads **10GB** per day of raw output through **CloudFront**
3 small EC2 instances serve out 10GB of data per day via various interfaces (WMS, OPeNDAP)
2 large EC2 instances do intensive processing on the data

Case 3 - $519.48 / month
10GB of new data per day
Public downloads **30GB** per day of raw output through **CloudFront**
2 small EC2 instances serve out 10GB of data per day via various interfaces (WMS, OPeNDAP)
1 small EC2 instances does processing on the data

Case 4 - $535.12 / month
10GB of new data per day
Public downloads **30GB** per day of raw output through **S3 directly (no CloudFront)**
3 small EC2 instances serve out 10GB of data per day via various interfaces (WMS, OPeNDAP)
1 small EC2 instances does processing on the data

Advantages:
- Scalable: Add new EC2 instances as you need. S3 storage is "use-as-you-go".
- Safe: Backed up. Don't worry about losing data, worry about improving the data.
- Bandwidth: Access to data will be much improved for external parties
- No need for infrastructure (servers, power, air-conditioning etc.)
- All data in one place

Disadvantages:
- Speed of uploading raw model and observation results to S3.
- Linux system administration required
- Access to data may not be at "local" speeds if you are accustomed to accessing it directly from local disks
- NOAA has no formal IT security process for the use of the Cloud for operational uses – a number of NOAA non-operational projects using non-sensitive data are evaluating the use of the Cloud as part of the Federal GeoCloud Sandbox Initiative, so it is still an evolving process.

4.1.2 Catalogs, Data Discovery, Metadata, and Vocabularies

Catalogs and **Data Discovery** – GLOS currently hosts a GeoNetwork server. We also recommend evaluating ESRI Geoportal Server, which is in use at a some Federal data centers..

It is important that whatever catalog technology is implemented, that the catalog efforts are not in isolation. The IOOS DIF Assessment report (November 2010) suggests:

> *Enable GEOSS and Geodata.gov to harvest IOOS Registry and Catalog, and thereby enable IOOS data providers to meet the requirement to register servers at GEOSS and Geodata.gov simply by registering with IOOS.*

And as soon as clear guidelines are provided on this, it will be very important to comply. Automated harvesting by the IOOS Registry is not yet functional but it is important to monitor these developments so new data can be made available to this registry.

The same report also recommends:

> *Explore methods to enable commercial search engines to harvest IOOS Data Catalog such that users can find data simply by searching the web.*

We therefore recommend that metadata catalog information is made available in Web Accessible Folders (WAF) so that harvesters, such as the GoogleBot, can access the catalog records. An example of this is the COIN Atlantic Geocontent Generator (CGG) (http://www.marinebiodiversity.ca/COINAtlantic/reference/coinatlantic-geocontent-generator-cgg), which allows users to create KML-based metadata that is exposed to the Google Search engine.

Metadata – GeoNetwork provides metadata capabilities that should continue to be leveraged. We recommend all metadata be available using the ISO convention. Metadata may be stored in a single form and then represented in different formats, including ISO and FGDC. We also recommend utilization of the ISO metadata that is provided for datasets hosted through TDS. Most cataloging solutions support harvesting ISO metadata on a regular basis. Service metadata should be made available in ISO 19139 format using the OGC Catalog Service for Web (CS/W).

There are a number of metadata tools worth exploring for managing metadata including NOAA Mermaid (Metadata Enterprise Resource Management Aid) tool and the EPA Metadata Editor (EME)

Vocabularies – we recommend becoming a partner in the Marine Metadata Interoperability (MMI) project which provides a reference to numerous relevant vocabularies.

4.1.3 Quality Control

GLOS have already started work on implementation of the EPA quality control standards. The QARTOD processes, although still in development stage, should also be reviewed. Discussions with NDBC would be very useful to evaluate their QA/QC implementation plan which is very mature. A plan should then be designed to implement a formal QA/QC process for all data types.

4.1.4 Data Sharing/Delivery

The data storage recommendation is to harvest and store data using common data models in relational databases or gridded formats such as NetCDF. This allows the operational team to implement existing tools and services to be layered on top and provide different data sharing methods. We recommend at a minimum that the following should be considered

Table 5. Delivery Table Title

SOS - Sensor Observation Service	Add an SOS server. Note: GLOS have a very efficient JSON implementation, this approach is also used in other regions, and is very fast and robust – we recommend adding SOS as an option, but not to replace the JSON solution.
WaterML – came from Cuahsi developments- developed by Hydrology community	Not for immediate implementation, but the ability to harvest data via WaterML may be useful
WMS – Web Map Service	GLOS already have WMS services available. Explore new WMS technologies, including multi-core implementation for performance and ensure compliance with latest OGC standard, including support for the time specification and widely used projections.
WCS – Web Coverage Service	Optional
WFS – Web Feature Service	Optional
ESRI Services	Will depend on the technology stack used by the Products team
OPeNDAP - Open-source Project for a Network Data Access Protocol	GLOS have an operational TDS. Invest in optimal implementation for performance.
KML - Keyhole Markup Language	Some data in KML may be useful to serve the Products Team

4.1.5 Data Products

The GLOS Conceptual Plan (December 2007) described a large list of users and user needs. The task to define and build products to meet all of these needs is nearly too large to define, but the need for high priority data products is essential and a dedicated products team should be hired or contracted to build the products and tools that bridge the gap between the observation and model data with the end user needs.

This is a large, often not clearly defined effort and is driven by user needs assessments that include the products/analysis the users want and in what form they need it. Some of the decisions that will be made in this team are technology selections, including:

- Back end technology stack, OpenSource or Commercial, or both
- Relational database, OpenSource or Commercial, or both
- Web client tool options:
    - Javascript
    - Flex

- o Silverlight
- Mobile Apps
  - o iOS (iPad / iPhone / iTouch)
  - o Windows Mobile
  - o Android
- Desktop/Tool Product Support
  - o Matlab
  - o Python
  - o R
  - o ArcGIS
  - o Surfer
  - o SMS
  - o Others

The functionality of the data products must be defined, whether that product is created by a web tool or represents derived data products for download, by the user needs.

A mature and elegant public-facing web portal is essential to allow users, scientists and public, to discover and access data. Issues to be considered when determining the functionality of the web site tools include:

- Catalog interface to discover data
- Data/service end points accessible via catalog interface so sophisticated users can access raw data
- Data Visualization (observations and models)
- Surface and 3d data visualization
- Management of time-varying data – animations, time series
- Synthesized products (e.g temporal and spatial averages)
- Regridding tools
- Data conversion and download tools
- Subscription services and alert services
- Decision support tools (e.g. renewable energy siting analysis)
- Integration of remote data services
- Archive access

The decisions on these functionality requirements and the identification of user needs are likely to have a significant impact on DMAC development costs.

Table 6. Staffing Costs Table

Task	Description	Budget
Front Office	Marketing / PR person who deals with the external organizations about data and also interacts with the RA board, IOOS, and the public (what products, etc.).	0.5 FTE
DMAC Operations Manager	Technical FTE with at least 3 years experience - works closely with the Front Office person and understands the implementations to meet the RA needs. They coordinate the DMAC, are considered the regional "DMAC lead" and manage the efforts of the various teams described below.	1.0 FTE
Metadata/Catalogs/Registry, QA/QC/Federal Facility Liaison	Process-oriented person with technical skills to manage metadata, QA/QC processes and liaison with federal requirements and data centers such as NDBC and NODC. Ensures compliance with IOOS national and international standards.	1.0 FTE
Model Data Manager/Liaison	Staff working with the modelers/scientists to get their model output into compliant formats. They understand modeling and should have experience with Fortran. They need to be able to write code to convert historic datasets and manage archive and real-time model results. They also understand what the modelers need for data and provide support to modelers. Ideally, this person has good Fortran and Matlab skills.	1.0 to 2.0 FTEs
Observation Data Manager/Liaison	Manages all of the observation data and gets it into databases and other storage formats and provides data in multiple compliant services. Works with the Model team data format, variable naming (vocabularies) and ensures all data is collected and shared appropriately. Works closely with the data providers, the modelers, and the web site designer. This group knows where the data is coming from, where it is going, and what format it is in – the person is very technical. If existing tools don't exist to get the job done, they create them.	1.0 FTE
Systems Administrator	Systems administrator to manage installation and support of software and server hardware. Responsibilities include maintaining the servers and other computer systems including the required infrastructure to support the required performance, especially as data volumes and requirements expand. They must manage plans for redundancy, including ability to manage service outages and other problems. Should be familiar with virtualization and use of Cloud computing, and should have some light programming skills for scripting.	0.5-1.0 FTE
Web Site Designer UI/UX	Manages the public-facing web site. Tasks include: performance debugging, web browser compatibility, reviewing statistics and monitoring analytics, planning for new technologies, and maintaining content on site to reflect the latest developments and products available in the organization. Manages advertisements and e-commerce activities if appropriate. Skills should include web design, HTML, CSS, and Javascript.	0.5 – 1.0 FTE
Data Products	This group is in support of the web and mobile-based tools to address needs of different thematic user groups, e,g oceanographers, fishermen, water quality managers, bathers, CMSP, etc. They may also be tasked to build desktop products, perhaps Matlab or ArcGIS extensions to support specific user needs. This group creates stunning visual representations of the data as well as analysis tools. They work closely with the data managers, observation and model teams. They also support the outreach and proposal efforts of the region.	1.0 – 5.0 FTEs

4.2 STAFFING AND COSTS

Table 6 presents recommended roles and ranges of effort for the staffing of an observing system DMAC to address an aggressive implementation schedule and a continuing high level of support for scaling and expansion of the DMAC post-implementation.

The overall staffing for implementation over three years of a complete DMAC infrastructure is anticipated to require at least 6.5 FTEs/year, with the staffing level maintained post-implementation to support scaling and expansion. During design and implementation, staffing may include outside consultants with necessary specialized expertise, but in-house staff are expected to acquire this knowledge and take over most roles within one year. The roles shown in this table could be fulfilled solely within GLOS, but it is likely that efforts for certain roles, such as Model Data Manager/Liaison, Observation Data Manager/Liaison, and Data Products will be shared with organizations that provide data and data products.

4.3 IMPLEMENTATION OVERVIEW

Implementation of the DMAC is presented in detail in this project's Implementation Plan report. Selected highlights from the plan are presented below:

The DMAC architecture should be based on IOOS guiding principles to allow for regional, national and international interoperability, allowing various users to access data through a family of web services.

Multiple separate (physical or virtual) servers are recommended to provide optimized access to feature data, as well as gridded (raster) data through a variety of services including Web Coverage Service (WCS) and Web Mapping Service (WMS) as well as data delivery with OpenDAP protocols. Use of modern data models will allow for common access to in-situ observations, profiles and trajectories and the use of geospatially-enabled relational databases will allow for connectivity to GIS software.

The DMAC system will be able to ingest, manage, and disseminate a wide variety of data classes, including:

- Regular grids
- Unstructured grids
- Curvilinear grids
- Point time series
- Profile time series
- Collections of points or profiles
- 2D and 3D trajectories
- Collections of trajectories
- Swaths
- GIS features

The observation system data will be made available to dynamic web and mobile applications through a suite of standards-based protocols. Appropriate technology stacks for development and support of different dynamic client applications will depend on details of the application, but may include Adobe Flash/Flex, Silverlight, Java, Javascript, and HTML5 connected to the data through appropriate web services. The DMAC system should support asynchronous representative state transfer (ReST) communications for fast, reliable and powerful database access as well as data subscription and alert services.

Additional utility services can be provided within the DMAC system to support data registration and product development:

- Service registry (SOA convention)
- Data catalog services
- Metadata management
- Quality Control procedures
- Multiple mapping and visualization services
- Format conversion services
- Coordinate transformation services
- Product generation services
- Data integration services
- Workflows

Implementation for the DMAC will be realized in a different sequence from the general implementation of the observing system overall. Level A implementation of the DMAC will put in place necessary infrastructure to address observing system needs at all spatial scales (basin-wide, lake and regional) within the first three years, in contrast to the planned completion of full functionality only at the basin-wide scale for the system within five years. This reflects the expectation that deployment of new observing platforms and connection of observing data streams to the DMAC will be independent of the spatial scale of observing system applications and that the DMAC design is readily scalable to accept additional data sources.

This page is blank to facilitate double sided printing.

5. END-TO-END CAPABILITY ANALYSIS

This section presents an illustrative analysis of the capability of the DMAC to service data ingestion and dissemination needs for a likely end-to-end operational model. The example is for a fine-scale ecosystem model to inform the adaptive management process for addressing trophic gradient issues in Lake Michigan. In brief, nearshore eutrophication, believed to be driven by increases in watershed nonpoint source nutrient loads in conjunction with ecosystem re-engineering by Dreissenid mussels, has been contrasted with decreased productivity in offshore waters. This issue, discussed in detail in Section 5.3.3 of the Master Design Report, has led the water quality and fisheries management communities to express a need for quantitative understanding to support the development management strategies that address both nearshore and offshore concerns.

The lake-scale observing system presented in the Master Design Report will meet the data needs of the expected modeling framework, which will include NOAA-GLERL's existing POM code for hydrodynamics, a revised LM3-EUTRO for water quality, and appropriate watershed models (possibly including HSPF, LBRM, SPARROW, and/or SWAT) to characterize tributary inputs.

The observational requirements are expected to consist of the following elements:

- 8-15 buoys to measure meteorological data, currents, and water temperature; some buoys also equipped with multi parameter sondes to measure DO, Chl-a, PAR, turbidity, DOM.
- One or two cabled year-round platform to measure water temperature, waves, current, and ice cover
- 6-10 research vessel cruises along 6 transects (near buoys) with measurements for nutrients, phytoplankton and zooplankton biomass and speciation, and benthic algae and organism abundance
- Towed arrays as part of field campaigns to measure Chl-a, turbidity, PAR, conductivity, DOM, temperature, DO, side-scan sonar and lake-bottom video
- Multi-day glider deployments with same sensor payload as towed arrays
- Analysis for cladophora, chlorophyll, TSS, and DOM using existing free satellite imagery or purchased high-resolution satellite or airborne imagery.

All of these elements map directly to data categories in the DMAC design. The mapping of the observational elements to DMAC data categories is presented in Table 7 along with a brief description of anticipated DMAC processing to properly handle the observational data. Figure 14 portrays the resultant DMAC design in schematic format.

Table 7. DMAC Processing for End-to-End Observational Elements

Observation	Data Category	DMAC Process
Fixed Platforms Buoys and cabled platforms	Time-varying Observation data – stationary	The DMAC team will write processors to translate the data files from the shore receiving stations. These data files are generally created by the datalogger from the hardware manufacturers and are often ASCII. The processors will parse and store these data in the **Observations Database** – a relational database on the server at the DMAC center. QA/QC processes and vocabulary homogenization may be implemented at this step also. The data may also be made available from the database in XML for harvesting by NDBC. Metadata records (ISO-compliant) will be created for each sensor and stored in the **Regional Metadata Catalog**. SOS (and other) services will be installed and configured on top of the observations database so applications and users can access the data.
Field Campaigns Cruise data with nutrient and biological sensors	Time-varying Observation data – non-stationary	For the most part, these data are processed through laboratory analysis before being submitted to the DMAC system. There are no uniformly instituted guidelines for transmission of this data to the DMAC servers and Excel is often used. Development and implementation of an upload mechanism, similar to EPA Scribe, to allow Excel-based data with associated metadata records to be uploaded and stored in the DMAC **Observations Database** should be completed. QA/QC processes and vocabulary homogenization may be implemented during the upload process. GLENDA may also be used, perhaps in parallel, as data store, or mechanisms could be developed to make the DMAC Observations database accessible to systems such as Glenda, the Data Exchange, and Storet. Metadata records (ISO-compliant) will be created for each observation and stored in the **Regional Metadata Catalog**. A suite of services may be layered on top of the observations database (SOS, WaterML, other) for a variety of access to the data.
Mobile Platforms Towed arrays – physical parameters	Time-varying Observation data – non-stationary	For the most part, these data are processed through a datalogger from the instruments before being submitted to the DMAC system. Processors will parse and store these data in the **Observations Database** – a relational database on the server at the DMAC center. QA/QC processes and vocabulary homogenization may be implemented at this step also. Metadata records (ISO-compliant) will be created for each observation and stored in the **Regional Metadata Catalog**. A suite of services may be layered on top of the observations database (SOS, WaterML, other) for a variety of access to the data.

Table 7. DMAC Processing for End-to-End Observational Elements - Continued

Observation	Data Category	DMAC Process
Remote Sensing Analysis for cladophora, chlorophyll, TSS, and DOM using remote sensing	Time-varying gridded data e.g. Model results, Satellite data, Radar data	This data is generally gridded and will be stored as HDF/NetCDF and served with a THREDDS server. The data can be made available from TDS with Opendap, WMS, and WCS. Metadata records (ISO-compliant) will be created for each observation and stored in the TDS catalog as well the regional metadata catalog. ncISO can be used to harvest the TDS catalog and populate the **Regional Metadata Catalog** automatically.
Model Results	Time-varying gridded data e.g. Model results, Satellite data, Radar data	The end to end system will use a variety of models including POM, LM3-EUTRO and some watershed models such as SPARROW and LBRM. These results will be converted by processors to CF-compliant NetCDF and served by the TDS server. The model results can be made available from TDS with Opendap, WMS, WCS, and SOS. QA/QC processes and vocabulary homogenization may be implemented as model results are processed onto the TDS server. Metadata records (ISO-compliant) will be created for each model simulation and stored in the TDS catalog as well the **Regional Metadata Catalog**. ncISO can be used to harvest the TDS catalog and populate the regional metadata catalog automatically.
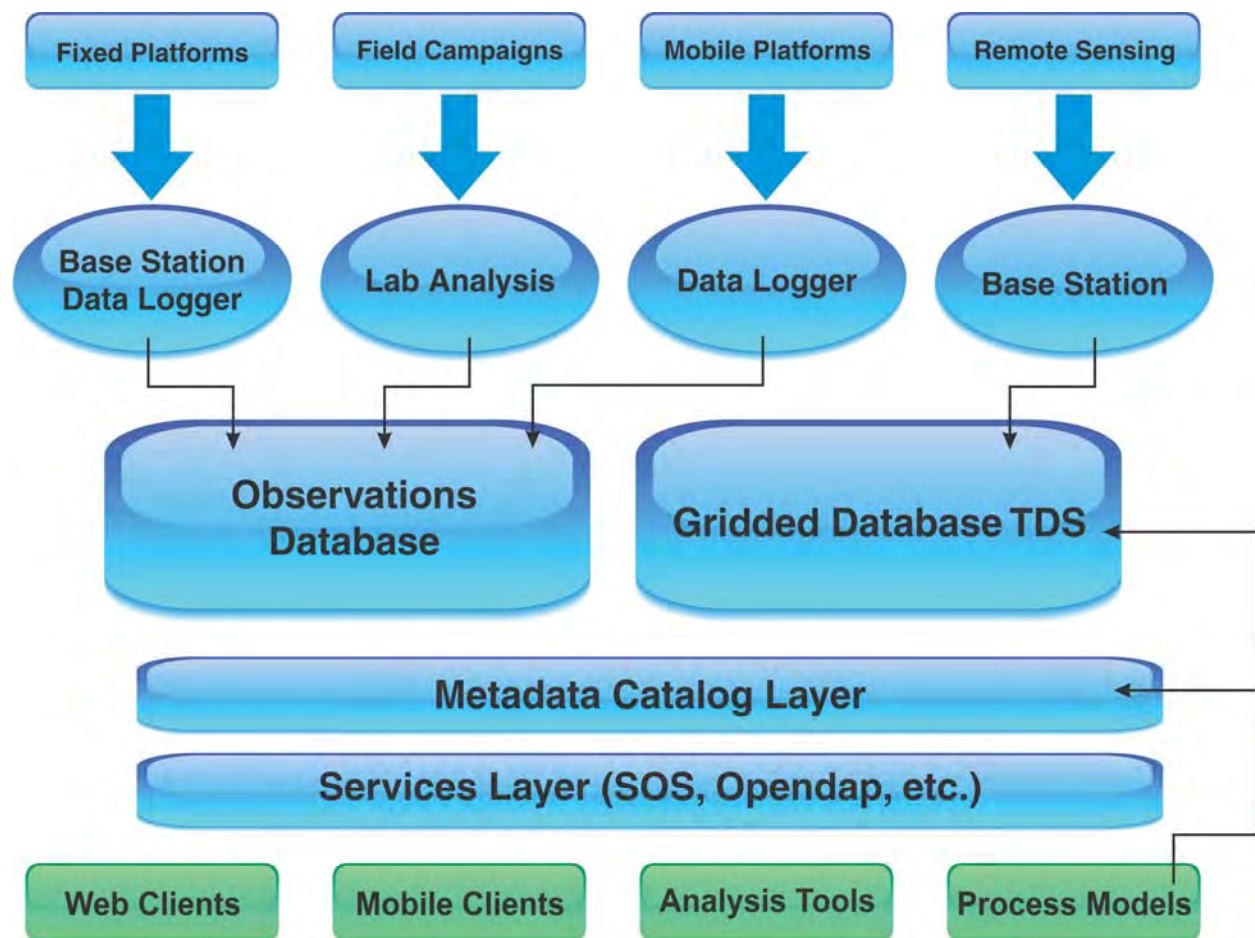
Figure 14. DMAC design schematic for Lake Michigan trophic gradient model

6. REFERENCES

Applied Science Associates, Inc. and the Massachusetts Ocean Partnership. March 2009. "An Integrated Statewide Ocean Data Network to Provide Data Analysis Necessary for Integrated Multi-use Ocean Management in Massachusetts." ASA Report 08-193.

Arrott, M., A.D. Chave, C. Farcas, E. Farcas, J.E. Kleinert, I. Krueger, M. Meisinger, J.A. Orcutt, C. Peach, O. Schofield, M.P. Singh, F.L. Vernon. March 1, 2010. "Integrating Marine Observatories into a System-of-Systems: Messaging in the US Ocean Observatories." Paper presented at Oceans 2009, MTS/IEEE Biloxi – Marine Technology for Our Future: Global and Local Challenges, Biloxi, MS. October 26 – 29, 2009.

Boeing Company. October 14, 2004. "System Design Specification Report." CDRAL A003. Draft. Joint Navy/NOAA NESIDIS IOOS Demonstration Project.

Committee on Archiving and Accessing Environmental and Geospatial Data at NOAA. 2007. "Environmental Data Management at NOAA: Archiving, Stewardship, and Access." National Research Council

Consortium of Ocean Leadership. October 2008. "Integrated Observatory Applications Architecture Document."

Consortium of Ocean Leadership. October 2008. "OOI Integrated Observatory Infrastructure Architecture Document."

Consortium of Ocean Leadership. September 2009. "OOI Annual Work Plan." Version 3-00 – CI_only. https://confluence.oceanobservatories.org/download/attachments/18907985/1001-00100_Annual_Work_Plan_OOI_2009-09-01_ver_3-00_CI_only.pdf

Consortium of Ocean Leadership. November 17, 2010. "OOI Operations and Maintenance Plan." Version 1-13

Department of Defense Chief Information Officer. 2003. "Department of Defense Net-Centric Data Strategy"

Great Lakes Observing System (GLOS) Regional Coastal Ocean Observing System (RCOOS). December 2007. "Conceptual Plan." Version 1.0 2008 - 2017

Gulf of Maine Ocean Data Partnership. November 2007. *Technical Guidance for Data Sharing*. Version 1.2,

International Oceanographic Data and Information Exchange (IODE). January 2007. "Marine Data Management: We Can Do More, but Can We Do Better?" International Oceanographic Data and Information Exchange (IODE), Intergovernmental Oceanographic Commission of UNESCO.

IOOS. 2009. "IOOS High Level Functional Requirements v 1.5." U.S. IOOS Office.

IOOS. November 2009 "U.S. Integrated Ocean Observing System Road Map to Full Operational Capability." Version 0.3. U.S. IOOS Office

IOOS. November 2010. "Data Integration Framework (DIF) Final Assessment Report." U.S. IOOS Office

IOOS. 2010. "IOOS Registry/Catalog/Viewer Functional Requirements" U.S. IOOS Office.

IOOS. November 2010. "U.S. Integrated Ocean Observing System: A Blueprint for Full Capability." Version 1.0. U.S. IOOS Office.

IOOS. April 2011. "U.S IOOS Draft Implementation Plan 2011." U.S. IOOS Office.

Joint Oceanographic Institutions. November 11, 2005. *ORION Cyberinfrastructure: Concept of Operations*. Version 0r8.

Joint Subcommittee on Ocean Science and Technology. December 2006. "Charter of the Interagency Working Group on Ocean Observations." Issued by the Joint Subcommittee on Ocean Science and Technology, Committee on Environment and Natural Resources, National Science and Technology Council.

Ladner, R. 2009. "METOC Data Management for Net-Centric Operations."

Lockheed Martin. 2006. "Integrated Ocean Observing System (IOOS): Conceptual Design."

Meisinger, M. 2009. "OOI Cyberinfrastructure Overview." Calit2, UCSD, OOI Instrumentation Workshop, March 2009

Meisinger, M., C. Farcas, E. Farcas, C. Alexander, M. Arrott, J. de La Beaujardiere, P. Hubbard, R. Mendelssohn, R. Signell. March 1, 2010. "Serving Ocean Model Data on the Cloud." Paper presented at Oceans 2009, MTS/IEEE Biloxi – Marine Technology for Our Future: Global and Local Challenges, Biloxi, MS. October 26 – 29, 2009.

National Office for Integrated and Sustained Ocean Observations. January 2006. "The First U.S. Integrated Ocean Observing System (IOOS) Development Plan." *Ocean*.US Publication 9.

NOAA. November 19, 2007. "Data Integration Framework (DIF) Functional Requirements Document." Version 1.0. IOOS Program Office,

NOAA. November 8, 2007. "Data Integration Framework (DIF) Master Project Plan." Version 1.0. IOOS Program Office.

NOAA. December 3, 2007. "Data Integration Framework Concept of Operations." Draft. Version 2.0. IOOS Program Office.

NOAA. October 2008. "Data Management and Communications Concept of Operations." IOOS Program Office

NOAA. March 12, 2010. "Guidance for Implementation of the Integrated Ocean Observing System (IOOS®) Data Management and Communications (DMAC) Subsystem." NOAA IOOS® Program Office White Paper (v1.)*

OGC. June 30, 2010. "Harmonising Standards for Water Observation Data - Discussion Paper." 09-124r2. Open Geospatial Consortium Inc.

Raytheon Corporation. August 31, 2006. "U.S. Integrated Ocean Observing System, Volume 1-IOOS Conceptual Design."

Taylor, A., C. Alexander, M. Arrot, J. deLaBeaujardiere, J. Graybeal, S. Hankin, R. Mendelssohn, J. Orcutt, B. Pritchet, R. Signell, D. Stuebe, S. Walker and J. Wilkin. 2010. "Improve the Process by which Researchers Get Data to Drive Models." OOI-CI & IOOS-DMAC Intersection Development Project-FY2010

U.S. Commission on Ocean Policy. July 2004. "An Ocean Blueprint for the 21st Century, Final Report."

This page is blank to facilitate double sided printing.

APPENDIX
DETAILS OF DATA ACCESS

This page is blank to facilitate double sided printing.

1   USGS

A. USGS Water Watch portal

Link: http://waterwatch.usgs.gov/new/
Online View:
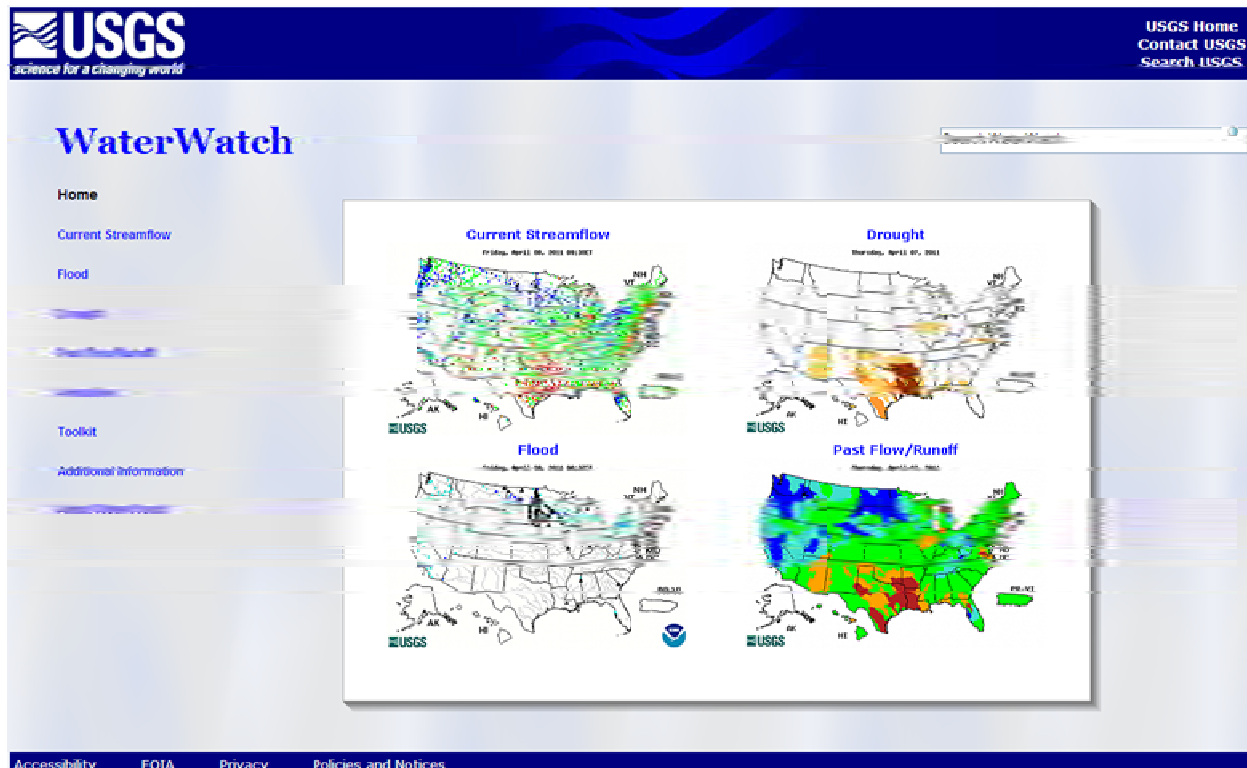
Description: USGS water data for the following:
- Current Streamflow
- Flood
- Drought
- Past Flow/Runoff

Data Type: Real time, Static
Data Format: Graphical, Tabular, Map Based
Retrieval Method: On site Download

B. USGS Water Quality Watch

Link: http://waterwatch.usgs.gov/wqwatch/
Online View:
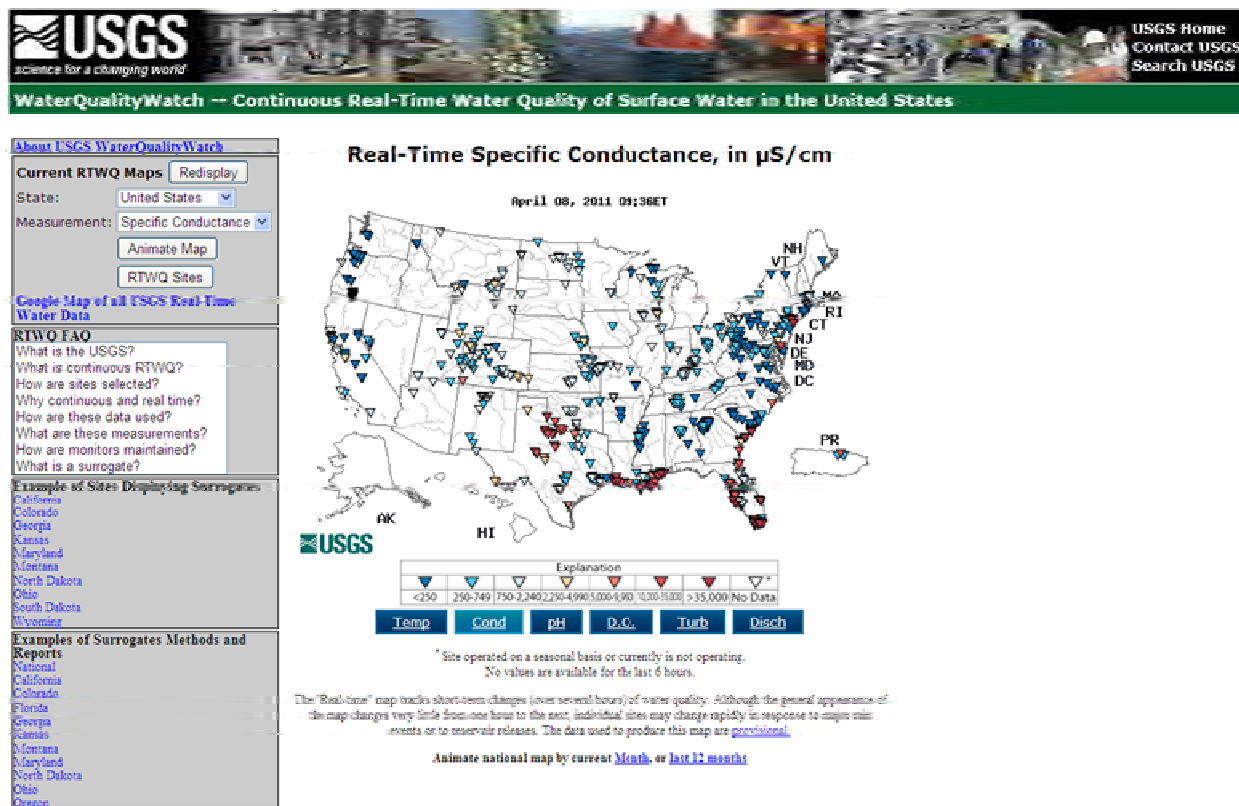
Description: Continuous Real-Time Water Quality of Surface Water in the United States
Data For:

- Real-Time Water Temperature, in °C
- Real-Time Specific Conductance, in µS/cm
- Real-Time pH, field, in standard units
- Real-Time Dissolved Oxygen, in mg/L
- Real-Time Turbidity, in NTU or FNU
- Real-Time Discharge, in cfs

Data Type: Real time

C. USGS WaterAlert

Link: http://storetnwis.epa.gov/storetqw/portal.html
Online View:
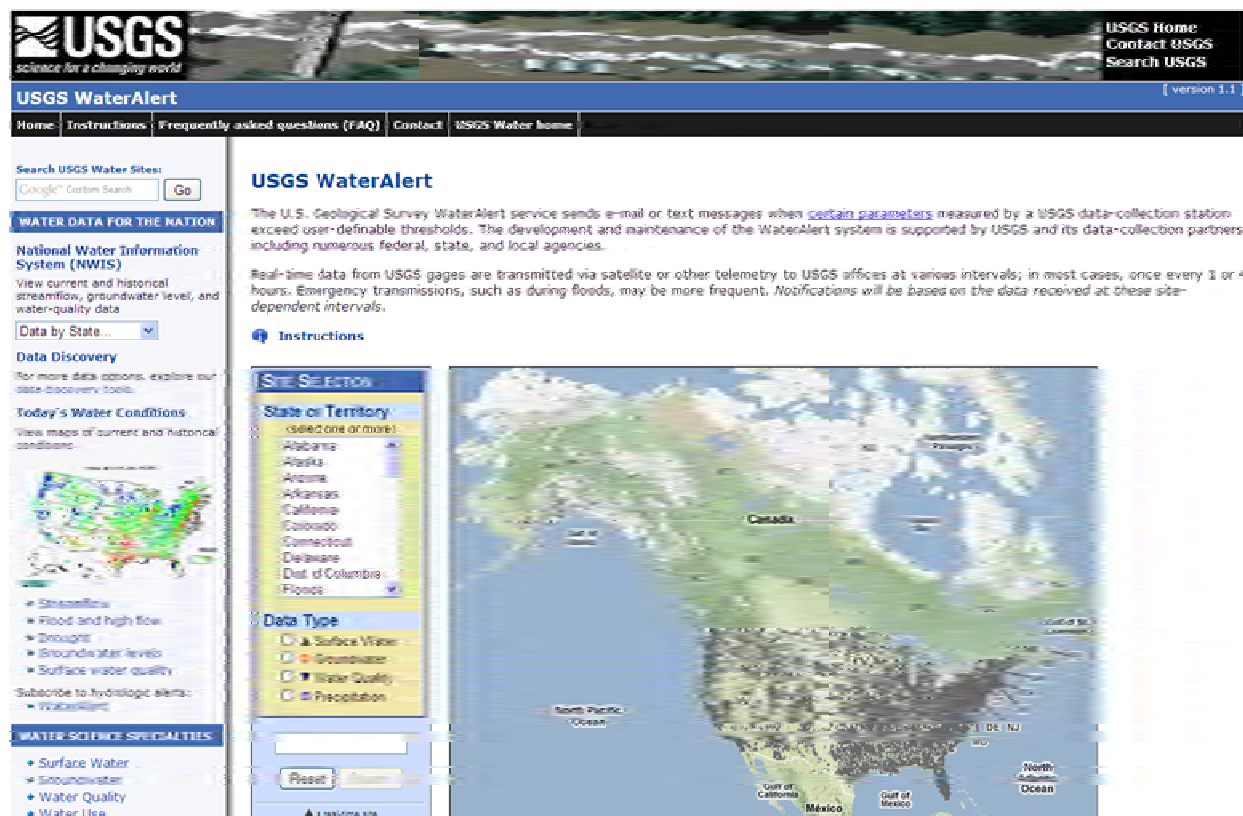
Description: The U.S. Geological Survey WaterAlert service sends e-mail or text messages when certain parameters measured by a USGS data-collection station exceed user-definable thresholds. The development and maintenance of the WaterAlert system is supported by USGS and its data-collection partners, including numerous federal, state, and local agencies.

Real-time data from USGS gages are transmitted via satellite or other telemetry to USGS offices at various intervals; in most cases, once every 1 or 4 hours. Emergency transmissions, such as during floods, may be more frequent. Notifications will be based on the data received at these site-dependent intervals.

Shows data for:
- Surface Water
- Groundwater
- Water Quality
- Precipitation

Data Type: Real time
Retrieval Method: email, cell phone text message or handheld device

D. USGS Seamless Data Warehouse

Link: http://seamless.usgs.gov/services.php

Online View:

Description: The seamless viewer provides different national geospatial datasets for orthoimagery, landcover, and elevation.

Services Available: WMS

Data Type: Static, near real time

Retrieval Method: Retrieve using WMS address

Download Service

- Initiate download http://extract.cr.usgs.gov/axis2/services/DownloadService/initiateDownload?
  - o Input parameter: string, Download URL obtained from the Request Validation Service
  - o Output: string contains unique job identifier
- getDownloadStatus
  http://extract.cr.usgs.gov/axis2/services/DownloadService/getDownloadStatus?downloadID=
  - o Input parameter: string, The job identifier obtained from the intitateDownload method.
  - o Output: string contains current identifier
- getData http://extract.cr.usgs.gov/axis2/services/DownloadService/getData?downloadID=
  - o Input parameter: job identifier
  - o Output: class="datasection"A finished download bundle.
- setDownloadComplete
  http://extract.cr.usgs.gov/axis2/services/DownloadService/setDownloadComplete?downloadID
  =

- o Input parameter: (Required) The job identifier obtained from the intitateDownload method.
- o Output: A string signifying a successful action

Elevation Query Web Service

- getElevation
  http://cumulus.cr.usgs.gov/XMLWebServices2/Elevation_Service.asmx?op=getElevation
  - o Input Parameters: X_Value(Required), Y _Value, (Required) , Elevation_Units, (default = FEET, or METERS), Source_Layer
  - o Output Results: String
- getAllElevations
  http://cumulus.cr.usgs.gov/XMLWebServices2/Elevation_Service.asmx?op=getAllElevations
  - o Input Parameter: X_Value, Y_Value, Elevation_Units
  - o Output: string

Inventory Service

- returnAttributeList: JSON, SOAP
  http://ags.cr.usgs.gov/index_service/Index_Service_JSON.ashx
  http://ags.cr.usgs.gov/index_service/Index_Service_SOAP.asmx?op=return_Attribute_List
  - o Inputs: None
  - o Outputs: A list of attributes that can be used as input to other methods in this webservice.
- returnAttributes: JSON, SOAP
  http://ags.cr.usgs.gov/index_service/Index_Service_JSON.ashx
  http://ags.cr.usgs.gov/index_service/Index_Service_SOAP.asmx?op=return_Attributes
  - o Inputs: Attribs, XMin, YMin, XMax, YMax, EPSG
  - o Outputs: A list of all of the requested attributes per dataset intersecting the requested area.
- returnAttributesElevation (JSON, SOAP)
  http://ags.cr.usgs.gov/index_service/Index_Service_JSON.ashx
  http://ags.cr.usgs.gov/index_service/Index_Service_SOAP.asmx?op=return_Attributes_Elevation
  - o Inputs: Attribs, XMin, YMin, XMax, YMax, EPSG
  - o Outputs: A list of all of the requested attributes per dataset intersecting the requested area.
- returnAttributesOrtho
- http://ags.cr.usgs.gov/index_service/Index_Service_JSON.ashx
  http://ags.cr.usgs.gov/index_service/Index_Service_SOAP.asmx?op=return_Attributes_Ortho
  - o Inputs: Attribs, XMin, YMin, XMax, YMax, EPSG
  - o Outputs: A list of all of the requested attributes per dataset intersecting the requested area.
- returnDownloadsOption
  http://ags.cr.usgs.gov/index_service/Index_Service_JSON.ashx
  http://ags.cr.usgs.gov/index_service/Index_Service_SOAP.asmx?op=return_Download_Options
  - o Inputs: ProductIDs
  - o Output: A list of productid, type, outputformat, compressionformat, metadataformat per dataset.
- return_Meta_URLs

Link: http://seamless.usgs.gov/website/seamless/viewer.htm
Online View:

Description: The seamless viewer provides different national geospatial datasets for orthoimagery, landcover, and elevation.

Data Type: Static, near real time

Retrieval Method: Download via Seamless Server. Select product and choose to download by Rectangle, area coordinates, rectangular area for tiled data, or area by multiple template section.

E. USGS Earth Explorer

Link: http://edcsns17.cr.usgs.gov/NewEarthExplorer/

Online View:

Description: Data warehouse for the following satellite products, requires user registration and log in (free)

Data Type: Real time, sensor data

Data Format: Geotiff; Export metadata as csv, shp, kml

Services Available: N/A

Retrieval Method: Choose the area of interest, address, or specific scene and click on "Data Sets." Select desired satellite products and click on "Results." This will prompt a user login and will lead to a download page.

F. USGS Glovis

Link: http://glovis.usgs.gov/

Online View:
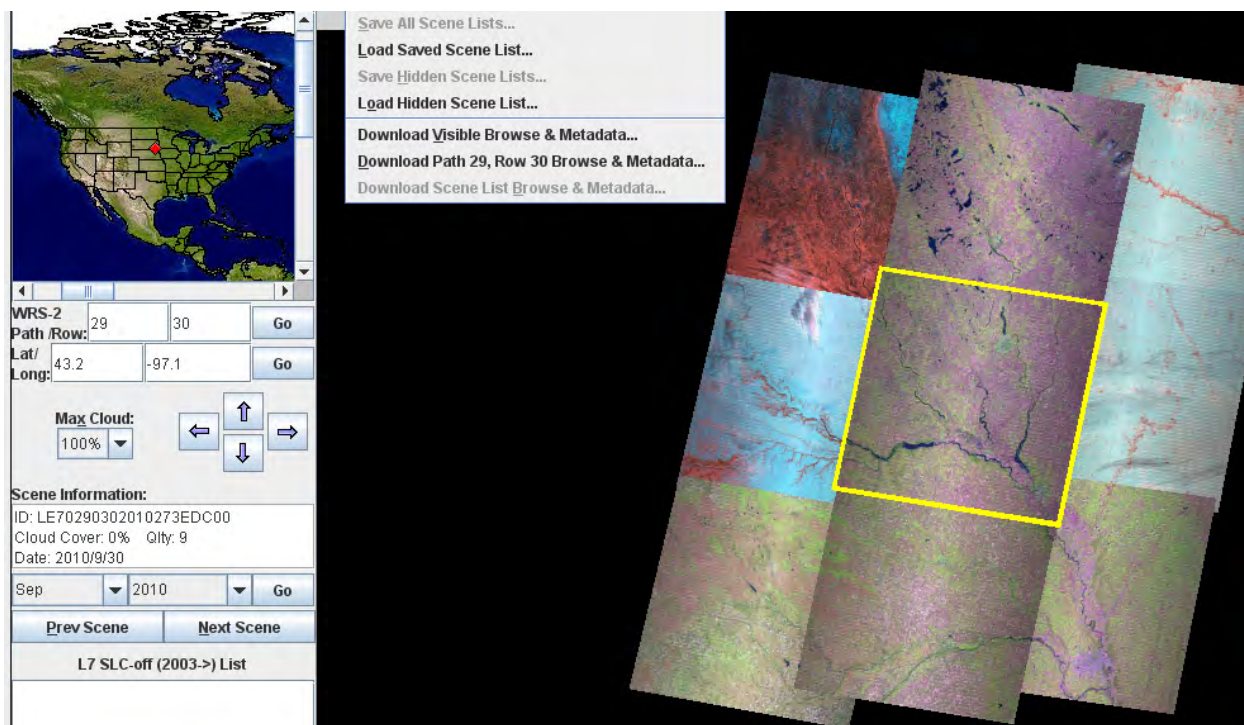
Select satellite sensor and choose location or specific scene (path/row)

Description: Warehouse of satellite data for individual scenes of imagery. Must be a registered user (free).

Data Type: Real time

Data Format: Geotiff, Jpeg

Services Available: N/A

Retrieval Method: Select the satellite scene for the chosen sensor and click on "Add to Cart." This opens up a new tab, which requests a user login and leads to a download page

2   EPA
A. EPA STORET/ WQX

Link: http://storetnwis.epa.gov/storetqw/portal.html

Online View:

Description: The STORET (short for STOrage and RETrieval) Data Warehouse is a repository for water quality, biological, and physical data and is used by state environmental agencies, EPA and other federal agencies, universities, private citizens, and many others. Take a minute to browse around our site or click on the water drop to retrieve monitoring data!

Data Type: Static

Data Format: Tabular, XLS

Services Available: SOAP, REST

Retrieval Method: Download using above webpage or can also use web services.

B. EPA STORET/WQX Central Warehouse

<u>Link</u>: http://www.epa.gov/storet/dw_home.html
<u>Online View:</u>

<u>Description</u>: Central warehouse for STORET data, enables users to search for data by Geographic location, Station, or Project. Searching by these criteria leads to different data query interfaces.

<u>Data Type:</u> Static

<u>Data Format</u>: TXT

<u>Services Available:</u> NA

<u>Retrieval Method:</u> Users select how they wish to query results (location, station, project) and then select the parameters for their search (station, date, activity, medium, characteristic, etc). STORET cannot return more than 600,000 records at once so modifications to search query may be necessary. Once results are returned, users select the information desired in the output file, the user email address, and then choose to submit for immediate or overnight processing. The EPA sends an email confirming the data request and a second email with a link to a zipped TXT file upon data processing.

C. EPA Substance Registry Services

Link: http://iaspub.epa.gov/sor_internet/registry/substreg/home/overview/home.do
Online View:

Description: Searchable database of substances. Downloadable substance details.
Substance Registry Services (SRS) is the Environmental Protection Agency's (EPA) central system for information about substances that are tracked or regulated by EPA or other sources. It is the authoritative resource for basic information about chemicals, biological organisms, and other substances of interest to EPA and its state and tribal partners.

The SRS makes it possible to identify which EPA data systems, environmental statutes, or other sources have information about a substance and which synonym is used by that system or statute. It becomes possible therefore to map substance data across EPA programs regardless of synonym.
The system provides a common basis for identification of, and information about:

- Chemicals

- Biological organisms

- Physical properties

- Miscellaneous objects

Data Type: Static
Data Format: PDF, CSV, Excel, XML
Services Available:
SRS Direct Link [System to System URL link]
Single Substance Request (CDX Query Function) [Web Service]
Substance List Request (CDX Solicit Function) [Web Service]
Add Substance to List Request (CDX Submit Function) [Web Service]
NOTE**:** Currently in production through the exchange network , not yet completed
(http://www.exchangenetwork.net/exchanges/cross/srs.htm)

D. Great Lakes Environmental Database (GLENDA)

Link: https://cdxnode64.epa.gov/glnpo/Query.do?method=start

Online View:

Description: Relational database for sample data that include queries water quality survey data, sediment chemistry, Lake Michigan mass balance results, fish tissue chemistry, and fish sample details. Depending on query type data are available from 1997 to 2010.

Data Type: Static

Data Format: XLS

Services Available: NA

Retrieval Method: GLENDA query system. Users must register for free on the EPA GLENDA website. Users must enter query type, year period, station, and analyte code parameters to retrieve data for desired location. Output data is downloadable in XLS format.

3 NWIS Web Database
  A. NWIS – SOS services for Discharge and Other Sensor Data

<u>Link</u>: http://nwisvaws02.er.usgs.gov/ogc-swie/
<u>Online View:</u>

<u>**Description**</u>: water Resources of the United States Data for Discharge, GageHeight, Temperature, Precipitation, Turbidity, DO, pH

<u>**Data Type:**</u> Real time, Static

<u>**Data Format**</u>: Tabular

<u>**Services Available:**</u> SOS, WFS

<u>**Retrieval Method:**</u>

  1. Sensor Observation Service (SOS)
        a. Daily Values

GetObservation - featureID(required), offering(required), observedProperty(required), beginPosition(optional), endPosition(optional), Interval(optional), Latest(optional)

observedProperty: Discharge, GageHeight, Temperature, Precipitation, Turbidity, DO, pH
offering: Mean, Maximum, Minimum, Variance, Mode, STD, SUM
beginPostion: YYYY-MM-DD, YYYY-MM, YYYY
endPostion: YYYY-MM-DD, YYYY-MM, YYYY

Interval: Today, ThisWeek, ThisYear *Future plan to implement ISO-8601 Duration option*

Latest: only the most recent data point is reported

Daily mean discharge observation by feature ID with begin and end time:

http://nwisvaws02.er.usgs.gov/ogc-swie/sos/dv?request=GetObservation&featureId=01446500&observedProperty=Discharge&beginPosition=1970-01-01&endPosition=1980-01-01&offering=Mean

Daily maximum temperature observations by feature ID and begin time:

http://nwisvaws02.er.usgs.gov/ogc-swie/sos/dv?request=GetObservation&featureId=05407000&observedProperty=Precipitation&beginPosition=2010-01-01&offering=Maximum
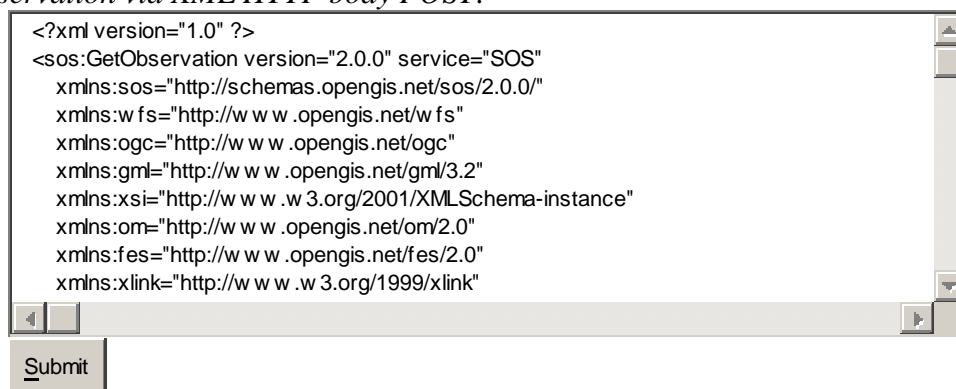
Daily mean discharge observations by feature ID this year:

http://nwisvaws02.er.usgs.gov/ogc-swie/sos/dv?request=GetObservation&featureId=05407000&observedProperty=Discharge&Interval=ThisYear&offering=Mean

Latest daily mean discharge observations by feature ID:

http://nwisvaws02.er.usgs.gov/ogc-swie/sos/dv?request=GetObservation&featureId=05407000&observedProperty=Discharge&offering=Mean&Latest

GetObservation via XML HTTP body POST:

```xml
<?xml version="1.0" ?>
<sos:GetObservation version="2.0.0" service="SOS"
    xmlns:sos="http://schemas.opengis.net/sos/2.0.0/"
    xmlns:wfs="http://www.opengis.net/wfs"
    xmlns:ogc="http://www.opengis.net/ogc"
    xmlns:gml="http://www.opengis.net/gml/3.2"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xmlns:om="http://www.opengis.net/om/2.0"
    xmlns:fes="http://www.opengis.net/fes/2.0"
    xmlns:xlink="http://www.w3.org/1999/xlink"
```

Submit

GetCapabilities

http://nwisvaws02.er.usgs.gov/ogc-swie/sos/dv?request=GetCapabilities

DescribeSensor

http://nwisvaws02.er.usgs.gov/ogc-swie/sos/dv?request=DescribeSensor

GetDataAvailablity - featureID, offering, observedProperty, beginPosition and endPostion are all optional. If not used, all the features/properties in the SWIE will be displayed

GetDataAvailablity by feature ID and offering:

http://nwisvaws02.er.usgs.gov/ogc-swie/sos/dv?request=GetDataAvailablity&featureID=05082500&offering=mean

GetDataAvailablity by observed property and feature ID:
> http://nwisvaws02.er.usgs.gov/ogc-swie/sos/dv?request=GetDataAvailablity&observedProperty=Discharge&featureID=05568500

General (very large file / long load time):
> http://nwisvaws02.er.usgs.gov/ogc-swie/sos/dv?request=GetDataAvailablity

GetDataAvailablity via XML HTTP body POST:

```
<?xml version="1.0" ?>
<sos:GetDataAvailablity version="2.0.0" service="SOS"
    maxFeatures="3"
    xmlns:sos="http://schemas.opengis.net/sos/2.0.0/"
    xmlns:wfs="http://www.opengis.net/wfs"
    xmlns:ogc="http://www.opengis.net/ogc"
    xmlns:gml="http://www.opengis.net/gml/3.2"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xmlns:om="http://www.opengis.net/om/2.0"
    xmlns:fes="http://www.opengis.net/fes/2.0"
```

Submit

b. Unit Values

GetObservation - featureID(required), observedProperty(required), beginPosition(optional), endPosition(optional), Interval(optional), Latest(optional)

> observedProperty: Discharge, GageHeight, Temperature, Precipitation, Turbidity, DO, pH
> beginPostion: YYYY-MM-DD, YYYY-MM, YYYY
> endPostion: YYYY-MM-DD, YYYY-MM, YYYY
> Interval: Today, ThisWeek *Future plan to implement ISO-8601 Duration option*
> Latest: only the most recent data point is reported

Gage height observation by feature ID and begin time:
> http://nwisvaws02.er.usgs.gov/ogc-swie/sos/uv?request=GetObservation&featureId=01446500&observedProperty=GageHeight&beginPosition=2011-03-31

Discharge observation by feature ID and begin time:
> http://nwisvaws02.er.usgs.gov/ogc-swie/sos/uv?request=GetObservation&featureId=01446500&observedProperty=Discharge&beginPosition=2011-03-31
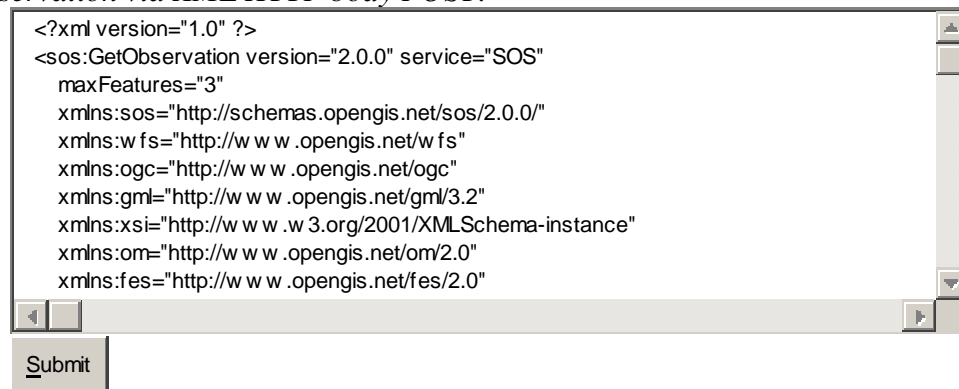
Latest discharge observation by feature ID:
> http://nwisvaws02.er.usgs.gov/ogc-swie/sos/uv?request=GetObservation&featureId=05407000&observedProperty=Discharge&Latest

Temperature observation by feature ID for this week:
> http://nwisvaws02.er.usgs.gov/ogc-swie/sos/uv?request=GetObservation&featureId=05407000&observedProperty=Temperature&Interval=ThisWeek

GetObservation via XML HTTP body POST:

```
<?xml version="1.0" ?>
<sos:GetObservation version="2.0.0" service="SOS"
    maxFeatures="3"
    xmlns:sos="http://schemas.opengis.net/sos/2.0.0/"
    xmlns:wfs="http://www.opengis.net/wfs"
    xmlns:ogc="http://www.opengis.net/ogc"
    xmlns:gml="http://www.opengis.net/gml/3.2"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xmlns:om="http://www.opengis.net/om/2.0"
    xmlns:fes="http://www.opengis.net/fes/2.0"
```

Submit

GetCapabilities
> http://nwisvaws02.er.usgs.gov/ogc-swie/sos/uv?request=GetCapabilities

DescribeSensor
> http://nwisvaws02.er.usgs.gov/ogc-swie/sos/uv?request=DescribeSensor

GetDataAvailablity - featureID, observedProperty, beginPosition and endPostion are all optional. If not used, all the features/properties in the SWIE will be displayed

GetDataAvailablity by feature ID:
> http://nwisvaws02.er.usgs.gov/ogc-swie/sos/uv?request=GetDataAvailablity&featureID=05568500

GetDataAvailablity by observed property and feature ID:
> http://nwisvaws02.er.usgs.gov/ogc-swie/sos/uv?request=GetDataAvailablity&observedProperty=Discharge&featureID=05568500

General (very large file / long load time):
> http://nwisvaws02.er.usgs.gov/ogc-swie/sos/uv?request=GetDataAvailablity

GetDataAvailablity via XML HTTP body POST:

```
<?xml version="1.0" ?>
<sos:GetDataAvailablity version="2.0.0" service="SOS"
    maxFeatures="3"
    xmlns:sos="http://schemas.opengis.net/sos/2.0.0/"
    xmlns:wfs="http://www.opengis.net/wfs"
    xmlns:ogc="http://www.opengis.net/ogc"
    xmlns:gml="http://www.opengis.net/gml/3.2"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xmlns:om="http://www.opengis.net/om/2.0"
    xmlns:fes="http://www.opengis.net/fes/2.0"
```

Submit

2. Web Feature Service

GetCapabilities

http://nwisvaws02.er.usgs.gov/ogc-swie/wfs?request=GetCapabilities

DescribeFeatureType

http://nwisvaws02.er.usgs.gov/ogc-swie/wfs?request=DescribeFeatureType

GetFeature
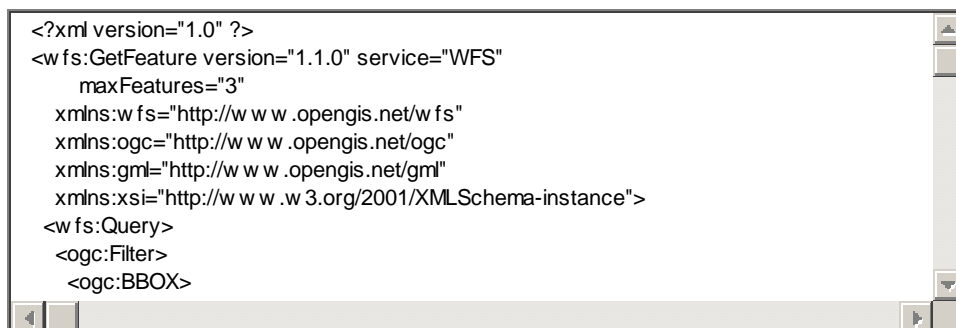
All features related to SWIE:

http://nwisvaws02.er.usgs.gov/ogc-swie/wfs?request=GetFeature

GetFeature by feature ID:

http://nwisvaws02.er.usgs.gov/ogc-swie/wfs?request=GetFeature&featureId=01446500

GetFeature via XML HTTP body POST:

```
<?xml version="1.0" ?>
<wfs:GetFeature version="1.1.0" service="WFS"
    maxFeatures="3"
    xmlns:wfs="http://www.opengis.net/wfs"
    xmlns:ogc="http://www.opengis.net/ogc"
    xmlns:gml="http://www.opengis.net/gml"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <wfs:Query>
   <ogc:Filter>
    <ogc:BBOX>
```

3.　Map
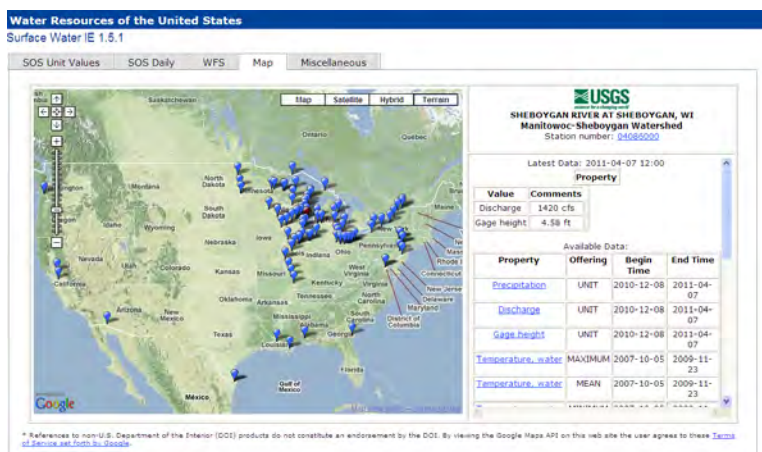
When you click on a link for a station available data is shown. You can then click on the data reuired and link to the data. For instance Precipitation

You can then link to the plot data from this webpage.

B. NWIS Web Database

Link: http://waterdata.usgs.gov/nwis/
Online View:
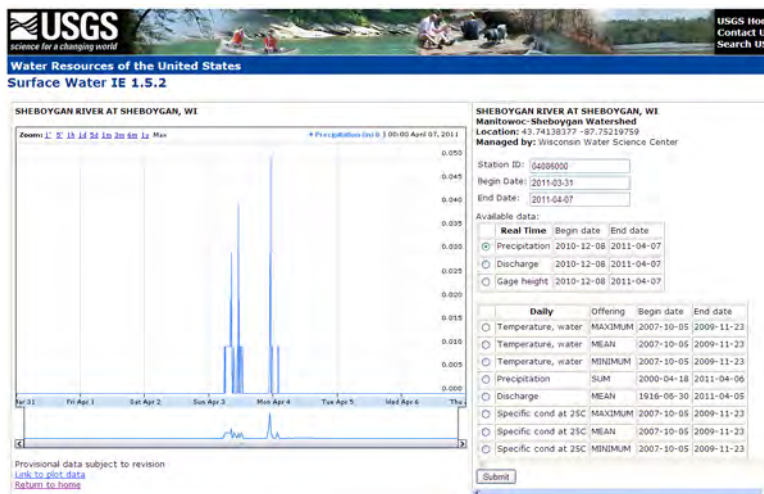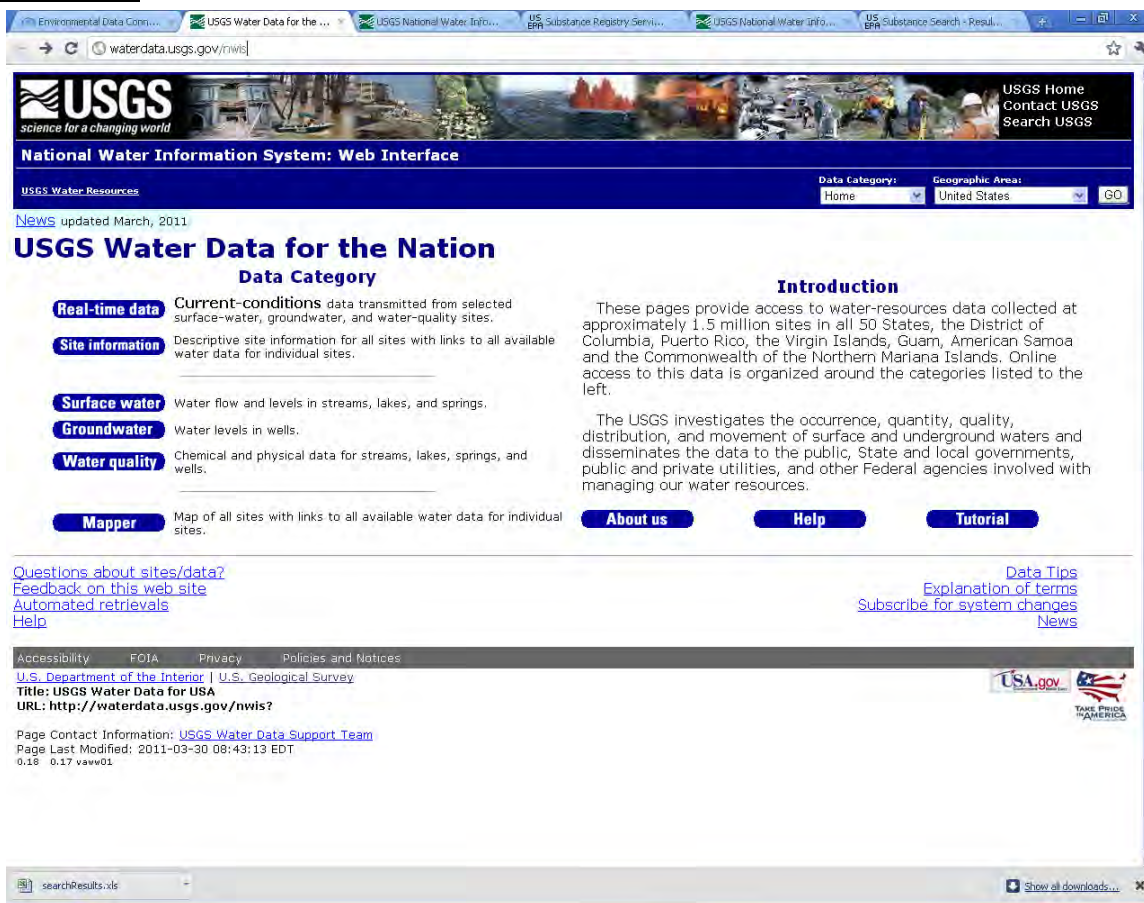
Description: Water Quality data store for USGS and EPA.
Data Type: Real time, Static

Data Format: graphical, tabular, tab-delimited (rdb) data files, XML, REST, SOAP, WSDL

Services Available: Web Services

Retrieval Method:

1. Real Time Data

Allows you to look at real time data through NWIS

2. Site Information

System for finding information on sites available. Can Download a XLS of the data with Lat Long etc.

3. Surface Water

System for finding information on data available for Water Data.

- Real-time Data
- Daily Data
  - o Stats
  - o Daily
  - o Monthly
- Annual
- Peak-flow data
- Field measurements

4. Groundwater

System for finding information on data available for Groundwater Data.

- Real-time Data
- Daily Data
  - o Stats
  - o Daily
  - o Monthly
- Annual

- Peak-flow data
- Field measurements

5. Water Quality

- Real-time Data
- Daily Data
- Stats
    - Daily
    - Monthly
    - Annual
- Field/lab samples

6. Mapper
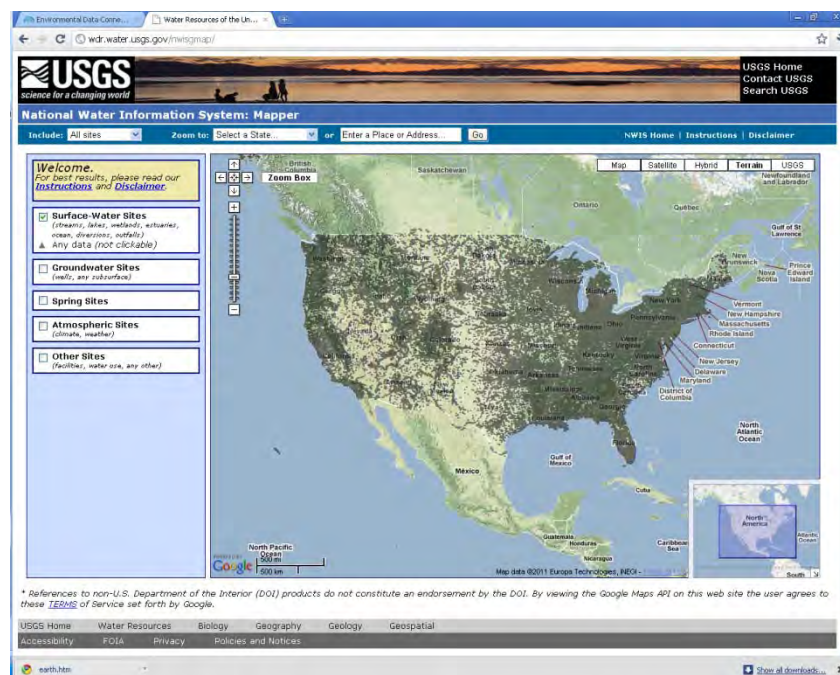
Map system to display:

- Surface Water
- Ground Water
- Spring Sites
- Atmospheric Sites
- Other Sites
- Allows you to view the data on a large map and choose between:
- All sites
- Active Sites
- Real Time sites
- Inactive Sites

- State
- Place Name or address

1.1 Automated retrieval

http://waterdata.usgs.gov/nwis/news/?automated_retrieval_info

4. NOAA

A. CoastWatch

Link: http://CoastWatch.noaa.gov/cwn/search/interface.html

Online View:

Description: Provides access to satellite sensor data in near real time for different geographic regions within the U.S. Products available may be aggregated or non aggregated. Sensors provide information on SST, reflectance, chlorophyll, true color, and ice coverage.

Data Type: near real time

Data Format: NC, HDF, PNG, TIFF

Services Available: WMS, WCS, OPENDAP, HTTPServer

Retrieval Method:

Thredds http://CoastWatch.noaa.gov/thredds/catalog.html

FTP ftp://ftpCoastWatch.noaa.gov/, ftp://ftpoceanwatch.noaa.gov/

HTTP http://CoastWatch.noaa.gov/data/

Recent Data Search: Search and Download:
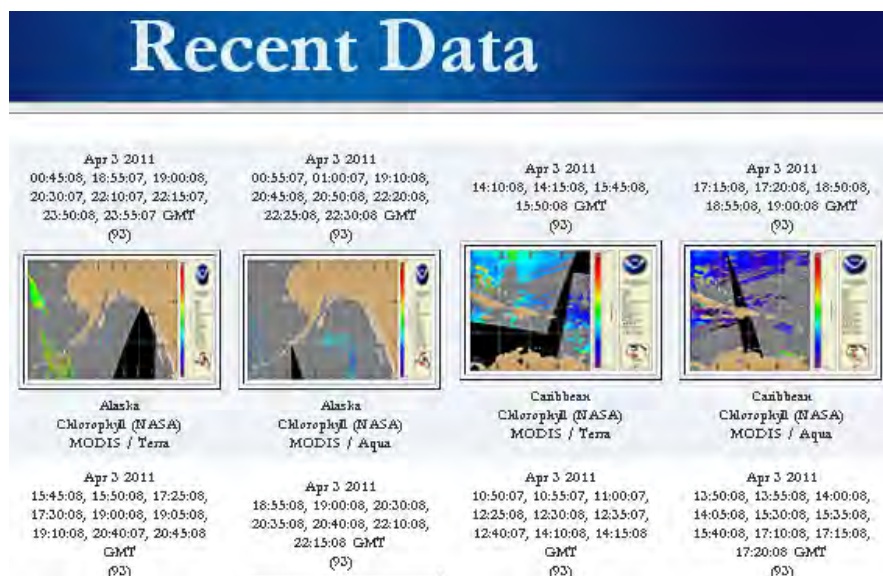
B. National Weather Service Doppler Radar Sites

<u>Source</u>: National Weather Service Doppler Radar Sites

<u>Link</u>: http://radar.weather.gov/

<u>Online View:</u>

<u>Description</u>: NEXRad for precipitation and weather

<u>Data Format</u>: KML, KMZ, GIF

<u>Services Available:</u> ArcGIS services, WMS, WFS

<u>Retrieval Method:</u>

Services

http://gis.srh.noaa.gov/ArcGIS/rest/services/Radar_warnings/MapServer

http://gis.srh.noaa.gov/ArcGIS/rest/services/RIDGERadar/MapServer

On site Download

http://www.srh.noaa.gov/jetstream/doppler/ridge_download.htm
http://radar.weather.gov/ridge/Conus/RadarImg/ (for cntgrtlakes)

C. National Data Buoy Center

Source: National Data Buoy Center (NDBC)

Link: http://sdf.ndbc.noaa.gov/sos/

Online View:
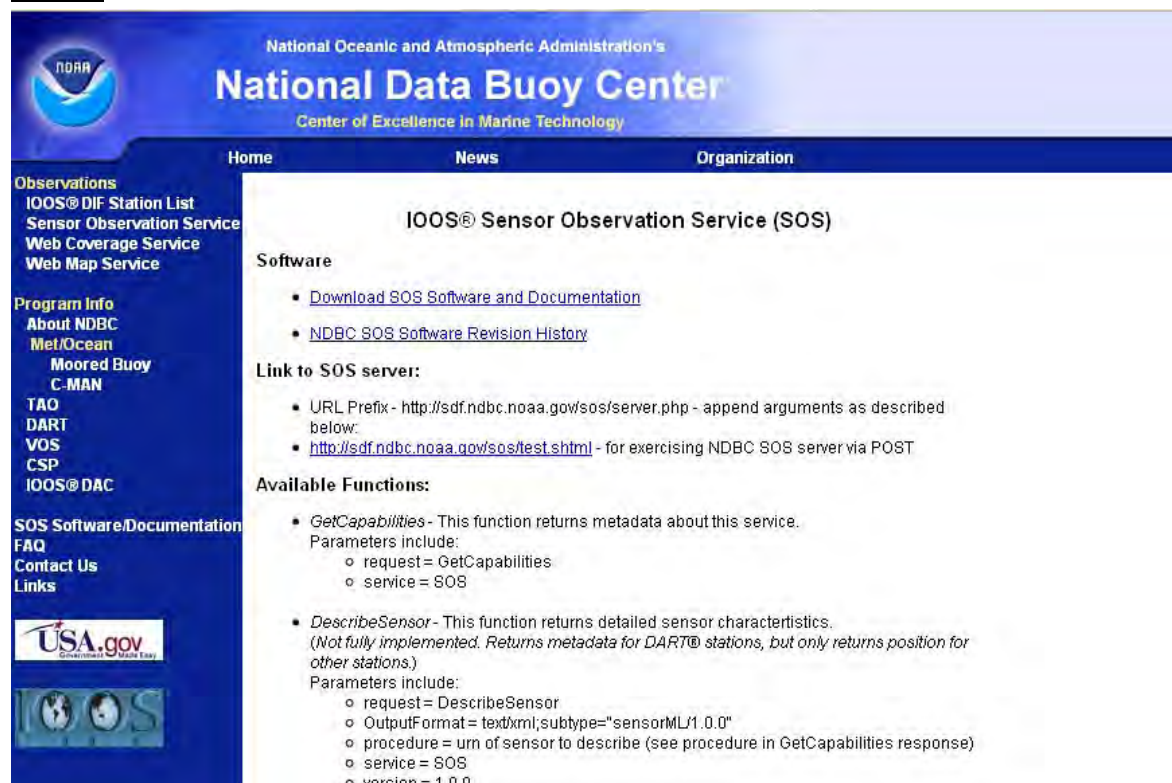
Description: Provides a link to the NDBC SOS server with available parameters for accessing data.

Data Type: real time

Data Format: GML, CSV, KML, TSV

Services Available: SOS

Retrieval Method: Connect using service parameters

D. National Ice Center

Link: http://www.natice.noaa.gov/products/great_lakes.html
Online View:

Great Lakes Ice Analysis Products

Use this page to select ice analysis products for a desired area over a selected time span. Remember that the more you include, the larger this page will become. Also keep in mind that larger files in the result listing will take longer to download.

These products are constructed by NIC analysts using various imagery sources with resolutions down to 50 meters per pixel. Imagery sources include but are not limited to ENVISAT, DMSP OLS, AVHRR, and RADARSAT. NIC ice analysts provide the necessary value-added interpretation of these imagery sources to properly identify the extent of the ice edge contours.

Format	File	Size (k)
Composite West	wl110321color.jpg	72
	wl110322bw.jpg	716
	wl110322color.jpg	605
	wl110323bw.jpg	700
	wl110323color.jpg	587
	wl110324bw.jpg	693
	wl110324color.jpg	604
	wl110325color.jpg	74
	wl110327color.jpg	73
	wl110328color.jpg	78
	wl110329bw.jpg	711
	wl110329color.jpg	622
	wl110330bw.jpg	740
	wl110330color.jpg	630
	wl110331bw.jpg	770
	wl110331color.jpg	676
	wl110403color.jpg	67

Description: Provides data for daily ice concentrations, forecasts, ice analysis grids, charts, and satellite imagery. Sensor data may not have complete coverage for entire Great Lakes region. Chart or sensor data display data in JPEG format.

Data Type: real time, near real time

Data Format: JPEG, ASC, SHP

Services Available: N/A

Retrieval Method: download data products via query submission or media viewer

Ice Analysis Products – Select desired time period, region, category, and format for data.

Media viewer http://espcgis.nesdis.noaa.gov/website/ssdsnow/viewer.htm

Source: National Ice Center

Link : http://www.natice.noaa.gov/products/grib.html
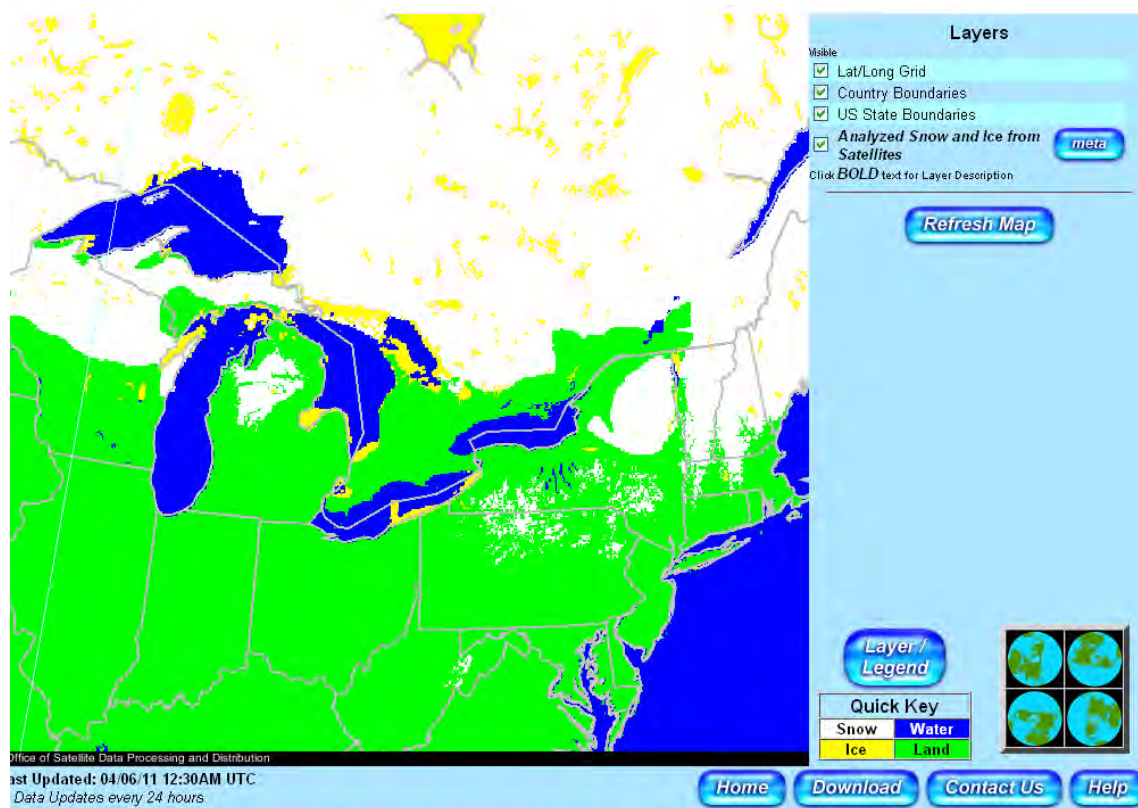
Online View:

Alaska and Great Lakes Ice Edges in GRIB Format

Recent files mapping the ice edges around Alaska and within the Great Lakes.

Alaska ice edges are mapped throughout the year. Great Lakes ice edge files are seasonal; they are discontinued when the Great Lakes are ice free.

Area	File	Size (K)
Great Lakes	T_OEBA88_C_KNWC_20110330120000.gr1	53
	T_OEBA88_C_KNWC_20110331120000.gr1	53
	T_OEBA88_C_KNWC_20110401120000.gr1	53
	T_OEBA88_C_KNWC_20110402120000.gr1	53
	T_OEBA88_C_KNWC_20110403120000.gr1	53
	T_OEBA88_C_KNWC_20110404120000.gr1	53
	T_OEBA88_C_KNWC_20110405120000.gr1	53

Description: Provides data for ice edges in the Great Lakes. These datasets are seasonal and discontinued when the lakes are ice free.

Data Type: real time, near real time

Data Format: GRIB

Services Available: N/A

Retrieval Method: Download

Source: National Ice Center, Great Lakes Products on Demand

Link: http://www.natice.noaa.gov/products/products_on_demand.html

Online View:
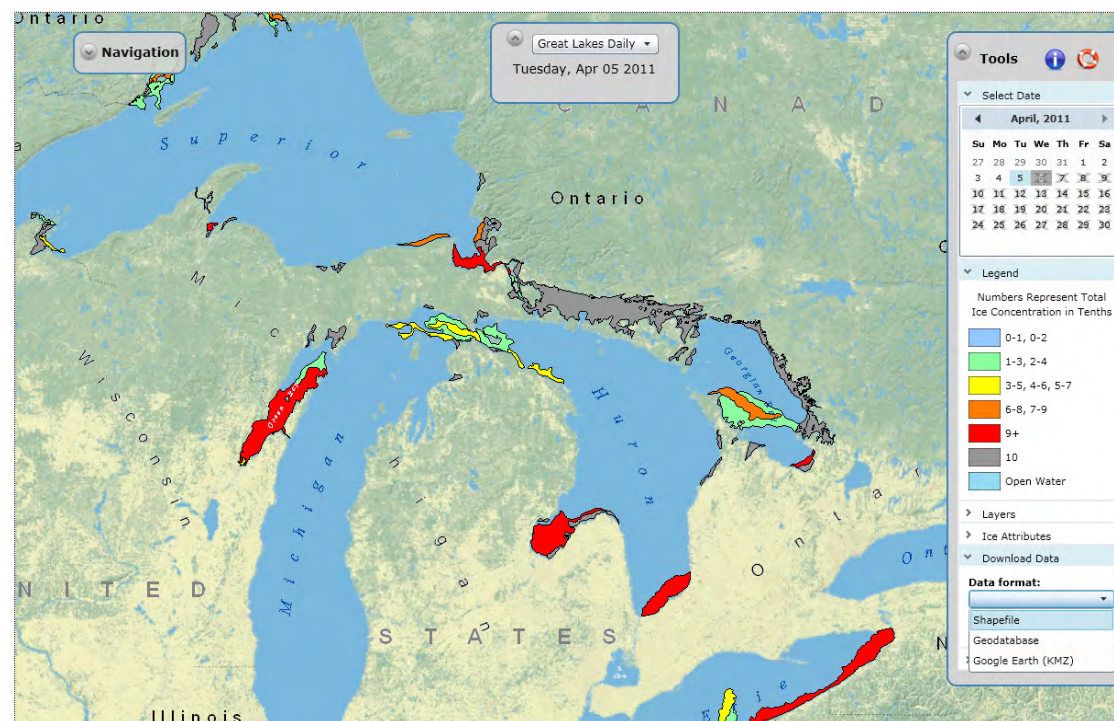
Description: Daily ice edge and concentrations (in tenths) for the Great Lakes. This data is based on imagery sources from Radarsat, Enivsat, DMSP OLS, AVHRR, and Quikscat.

Data Type: near real time

Data Format: SHP, GDB, KMZ

Services Available: N/A

Retrieval Method:

Select Great Lakes in the widget that displays region and date. Download data via Download tool in the Tools widget

E. CSC Digital Coast

Link: ftp://ftp.csc.noaa.gov/pub/

Online View:

Description: National repository for coastal data. Houses most requested GIS and raster coastal data as well as ESRI ArcGIS based tools for coastal management. Categories include Benthic, elevation, hydrography, land cover, marine boundaries, imagery, and socioeconomics. Data available directly through FTP or through Data Access Viewer.

Data Type: Static

Data Format: SHP, IMG, e00, TIFF, etc

Services Available: N/A

Retrieval Method:

FTP site: ftp://ftp.csc.noaa.gov/pub/

Media Viewer: http://csc-s-maps-q.csc.noaa.gov/dataviewer/viewer.html

- Select by area or keyword and add data to cart
- Enter email address and geographic information

F. AHPS

GLERL Great Lakes Advanced Hydrologic Prediction System (AHPS) Products (Experimental)

Link: http://www.glerl.noaa.gov/wr/ahps/curfcst/

Online View:

Description: GLERL's AHPS is a semi-automated system combining near real-time data with a suite of mathematical models developed to simulate the current state of basin hydrology

Data available for the lakes for the following:

- Total Basin Supply
- Outflow
- Mean Lake Level
- Forecast in tabular form
- Plot of Over-Lake Runoff
- Plot of Precipitation
- Plot of Over-Lake Evaporation
- Plot of Net Basin Supply
- Plot of Water Surface Temperature
- Plot of Over-Lake Air Temperature
- Plot of Vapor Pressure
- Plot of Wind Speed

- Plot of Over-Lake Cloud Cover
- Plot of Incident Short-Wave Radiation
- Plot of Reflected Radiation
- Plot of Net Long-Wave Radiation Flux
- Plot of Latent Heat Flux
- Plot of Sensible Heat Flux
- Plot of Net Surface Flux
- Plot of Ice Area
- Plot of Lake Heat Storage
- Plot of Soil Moisture Storage
- Plot of Surface Moisture Storage
- Plot of Snowpack Moisture Storage
- Plot of Total Moisture Storage

Data Type: Daily updates
Data Format: GIF, WMF
Services Available: FTP

5. Great Lakes Observing System (GLOS)

Link: http://storetnwis.epa.gov/storetqw/portal.html

Online View:

Description: The Great Lakes Observing System (GLOS) is an effort dedicated to providing wide internet access to real-time and historic data on the hydrology, biology, chemistry, geology and cultural resources of the Great Lakes, its interconnecting waterways and the St. Lawrence River. GLOS is a regional node of the U.S. Integrated Ocean Observing System (IOOS) initiative.

1. THREDDS Catalog

Catalogs available for the following:

- Great Lakes Bathymetry
- Great Lakes CFS nowcast
- Great Lakes FC Forecast
- Greatlakes FC Archive from 2010-2006

Services Available: OPENDAP, WCS, NetcdfSubset, WMS

2. Data Mapping

Observations Explorer

<u>Online View:</u>
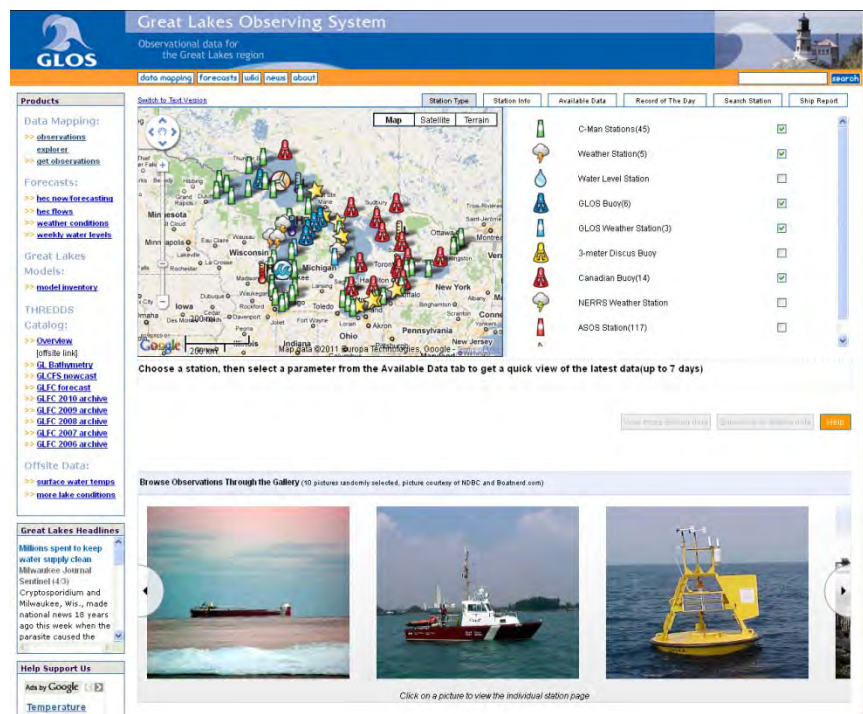
Allows users view observation data on a map for GL area.

3. *Forecasts*
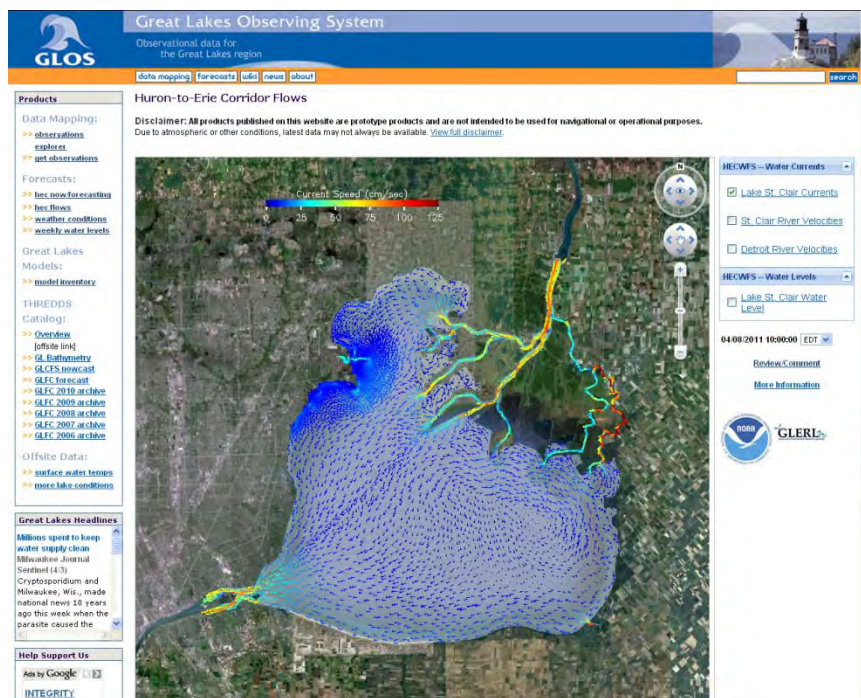
Nowcast WMS-T URL:http://michigan.glin.net/glos/hecwfs/nowcast/wms
**Forecast WMS-T URL:**http://michigan.glin.net/glos/hecwfs/forecast/wms
Data download : ESRI Shp format or NetCDF format via FTP

6. Motherlode THREDDS Catalog

Link: http://motherlode.ucar.edu:8080/thredds/catalog.html

Online View:

Catalog http://motherlode.ucar.edu:8080/thredds/catalog.html

Dataset	Size	Last Modifie
📁 Realtime data from IDD		
📁 NCEP Model Data/		
📁 FNMOC Model Data/		
📁 Radar Data/		
📁 Station Data/		
📁 Point Feature Collections/		
📁 Satellite Data/		
📁 Other Unidata Data		
📁 Unidata Real-time Regional Model/		
📁 Unidata GALEON Experimental Web Coverage Service (WCS) datasets/		
Test Restricted Dataset		

Motherlode TDS at Unidata
THREDDS Data Server [Version 4.2.5 - 20110302.2315] Documentation

Description: THREDDS catalog of modeled and satellite/radar data.

Data Type: real time, near real time

Data Format: GINI, NC

Services Available: WMS, WCS, OPENDAP, HTTPServer, NetCDFSubset

Retrieval Method: Connect using service addresses

7. Geospatial Data Gateway

Link: http://datagateway.nrcs.usda.gov/

Online View:

Description: GIS data and orthoimagery for each state and county
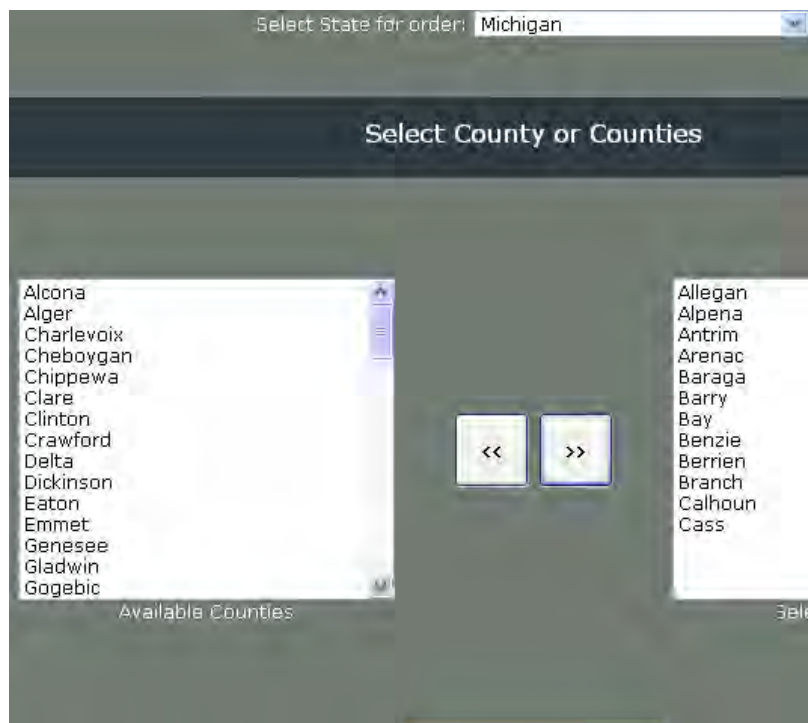
Data Type: static

Data Format: SHP, FGDB

Services Available: N/A

Retrieval Method:

Click on the Get Data button on the main page. Select State and Counties for interest to receive a list of available data for the specified area, and click continue to specify the data format and geographic information. User must fill in personal contact information in order to receive a confirmation email about the order and download details.

Summary of GLOS-EA Portal Data

Data Type	Domain	Data	Connection*	Source
GIS	Open Water	Erie Bathymetry	ArcGIS service	http://24.249.210.121/ArcGIS/rest/services/Glos/Hydrography/MapServer
		Ontario Bathymetry	ArcGIS service	http://24.249.210.121/ArcGIS/rest/services/Glos/Hydrography/MapServer
		Michigan Bathymetry	ArcGIS service	http://24.249.210.121/ArcGIS/rest/services/Glos/Hydrography/MapServer
		Huron Bathymetry	ArcGIS service	http://24.249.210.121/ArcGIS/rest/services/Glos/Hydrography/MapServer
		Superior Bathymetry	ArcGIS service	http://24.249.210.121/ArcGIS/rest/services/Glos/Hydrography/MapServer
		St. Clair Bathymetry	ArcGIS service	http://24.249.210.121/ArcGIS/rest/services/Glos/Hydrography/MapServer
		Great Lakes Depth	ArcGIS service	http://24.249.210.121/ArcGIS/rest/services/Glos/Hydrography/MapServer
	Tributary	8 digit HUCs	ArcGIS service	http://24.249.210.121/ArcGIS/rest/services/Glos/Hydrography/MapServer
		12 digit HUCs	ArcGIS service	http://24.249.210.121/ArcGIS/rest/services/Glos/Hydrography/MapServer
		US Land Cover	ArcGIS service	http://24.249.210.121/ArcGIS/rest/services/Glos/Landcover/MapServer
		Canada Land Cover	ArcGIS service	http://24.249.210.121/ArcGIS/rest/services/Glos/Landcover/MapServer
		US EPA Beaches	ArcGIS service	http://24.249.210.121/ArcGIS/rest/services/Glos/Landcover/MapServer
	Other	Port and Waterway Facilities	ArcGIS service	http://24.249.210.121/ArcGIS/rest/services/Glos/Infrastructure/MapServer
		National Water Network	ArcGIS service	http://24.249.210.121/ArcGIS/rest/services/Glos/Infrastructure/MapServer
		Real Time Monitoring Stations	ArcGIS service	http://24.249.210.121/ArcGIS/rest/services/Glos/Infrastructure/MapServer
		Water Quality Monitoring Stations	ArcGIS service	http://24.249.210.121/ArcGIS/rest/services/Glos/Infrastructure/MapServer
		Watershed Water Level Gages	ArcGIS service	http://24.249.210.121/ArcGIS/rest/services/Glos/Infrastructure/MapServer
		Beacon, Lateral	ArcGIS service	http://24.249.210.121/ArcGIS/rest/services/Glos/Infrastructure/MapServer
		Buoy, Lateral	ArcGIS service	http://24.249.210.121/ArcGIS/rest/services/Glos/Infrastructure/MapServer
		Daymark	ArcGIS service	http://24.249.210.121/ArcGIS/rest/services/Glos/Infrastructure/MapServer
		Fog Signal	ArcGIS service	http://24.249.210.121/ArcGIS/rest/services/Glos/Infrastructure/MapServer
		Radar Transponder Beacon	ArcGIS service	http://24.249.210.121/ArcGIS/rest/services/Glos/Infrastructure/MapServer
		Offshore Platform	ArcGIS service	http://24.249.210.121/ArcGIS/rest/services/Glos/Infrastructure/MapServer
In-situ				
	Atmosphere/			
Open Water	NDBC buoy observations	SOS via Buoy Connector	http://opendap.co-ops.nos.noaa.gov/ioos-dif-sos/	

Data Type	Domain	Data	Connection*	Source
	Atmosphere	NDBC Winds	WMS via ECOP	http://www.ndbc.noaa.gov/data/hourly2/
Model	Open Water	GLERL Erie Currents	WMS via ECOP	ftp://ftp.uscg.mil/
		GLERL Ontario Currents	WMS via ECOP	ftp://ftp.uscg.mil/
		GLERL Michigan Currents	WMS via ECOP	ftp://ftp.uscg.mil/
		GLERL Huron Currents	WMS via ECOP	ftp://ftp.uscg.mil/
		GLERL Superior Currents	WMS via ECOP	ftp://ftp.uscg.mil/
		NOS Erie Currents	WMS via ECOP	http://opendap.co-ops.nos.noaa.gov/index.php?dir=/netcdf/leofs/
		NOS Ontario Currents	WMS via ECOP	http://opendap.co-ops.nos.noaa.gov/index.php?dir=/netcdf/loofs/
		NOS Michigan Currents	WMS via ECOP	http://opendap.co-ops.nos.noaa.gov/index.php?dir=/netcdf/lmofs/
		NOS Huron Currents	WMS via ECOP	http://opendap.co-ops.nos.noaa.gov/index.php?dir=/netcdf/lhofs/
		NOS Superior Currents	WMS via ECOP	http://opendap.co-ops.nos.noaa.gov/index.php?dir=/netcdf/lsofs/
	Atmosphere	GLERL Erie Winds	WMS via ECOP	ftp://ftp.uscg.mil/
		GLERL Ontario Winds	WMS via ECOP	ftp://ftp.uscg.mil/
		GLERL Michigan Winds	WMS via ECOP	ftp://ftp.uscg.mil/
		GLERL Huron Winds	WMS via ECOP	ftp://ftp.uscg.mil/
		GLERL Superior Winds	WMS via ECOP	ftp://ftp.uscg.mil/
		NOS Erie Winds	WMS via ECOP	http://opendap.co-ops.nos.noaa.gov/index.php?dir=/netcdf/leofs/
		NOS Ontario Winds	WMS via ECOP	http://opendap.co-ops.nos.noaa.gov/index.php?dir=/netcdf/loofs/
		NOS Michigan Winds	WMS via ECOP	http://opendap.co-ops.nos.noaa.gov/index.php?dir=/netcdf/lmofs/
		NOS Huron Winds	WMS via ECOP	http://opendap.co-ops.nos.noaa.gov/index.php?dir=/netcdf/lhofs/
		NOS Superior Winds	WMS via ECOP	http://opendap.co-ops.nos.noaa.gov/index.php?dir=/netcdf/lsofs/
		NDFD Conus Winds	WMS via ECOP	ftp://tgftp.nws.noaa.gov/SL.us008001/ST.opnl/DF.gr2/DC.ndfd/AR.conus/VP.001-003/
		NOGAPS Winds	WMS via ECOP	https://weather.fnmoc.navy.mil/v2/rule_bot_ba
		North American Mesoscale Winds	WMS via ECOP	ftp://ftpprd.ncep.noaa.gov/pub/data/nccf/com/nam/prod/
		GFS Air Pressure	WMS via ECOP	https://weather.fnmoc.navy.mil/v2/rule_bot_ba
		GFS Ice Coverage	WMS via ECOP	https://weather.fnmoc.navy.mil/v2/rule_bot_ba
		NOGAPS Visibility	WMS via ECOP	https://weather.fnmoc.navy.mil/v2/rule_bot_ba
	Tributary	Sparrow – Phosphorous Yield	ArcGIS service	Limnotech/ASA prepared, http://24.249.210.121/ArcGIS/rest/services/Glos/Models/MapServer
		LRBM Output	AcrGIS service	ASA prepared, http://www.glerl.noaa.gov/ftp/ahps/LBRM/, http://24.249.210.121/ArcGIS/rest/services/Glos/Models/MapServer

Data Type	Domain	Data	Connection*	Source
Remote Sensing	Atmosphere	Nexrad Radar Coverage	WMS via ECOP	http://motherlode.ucar.edu:8080/thredds/dodsC/nexrad/composite/1km/files/N0R/
	Open Water	HF Radar	WMS via ECOP	ftp://sandbox.ucsd.edu/pub/CORDC/outgoing/HFRNet
		Chlorophyll – MODIS aggregated product	WMS via Layer List	http://CoastWatch.noaa.gov/thredds/wms/CoastWatch/MODSCW_P/closest_chlora/Daily Agg/GL05?
		Reflectance – MODIS aggregated product	WMS via Layer List	http://CoastWatch.noaa.gov/thredds/wms/CoastWatch/MODSCW_P/closest_Rrs667/Daily Agg/GL05?
Sample	Open Water	STORET Biological sampling	Catalog Viewer	ASA processed, http://storetnwis.epa.gov/storetqw/portal.html
	Tributary	GLENDA Lake Nutrients	Catalog Viewer	ASA processed https://cdx.epa.gov/SSL/CDX/login.asp, requires login

*All ArcGIS map services are available via the Layer List