Hydroclimatological Data Rescue, Data Inventory, Network Analysis & Data Dissemination

Greta Burkart, Jessica Cherry, Amy Jacobs, & Jenny March

University of Alaska Fairbanks

International Arctic Research Center and

Institute of Northern Engineering

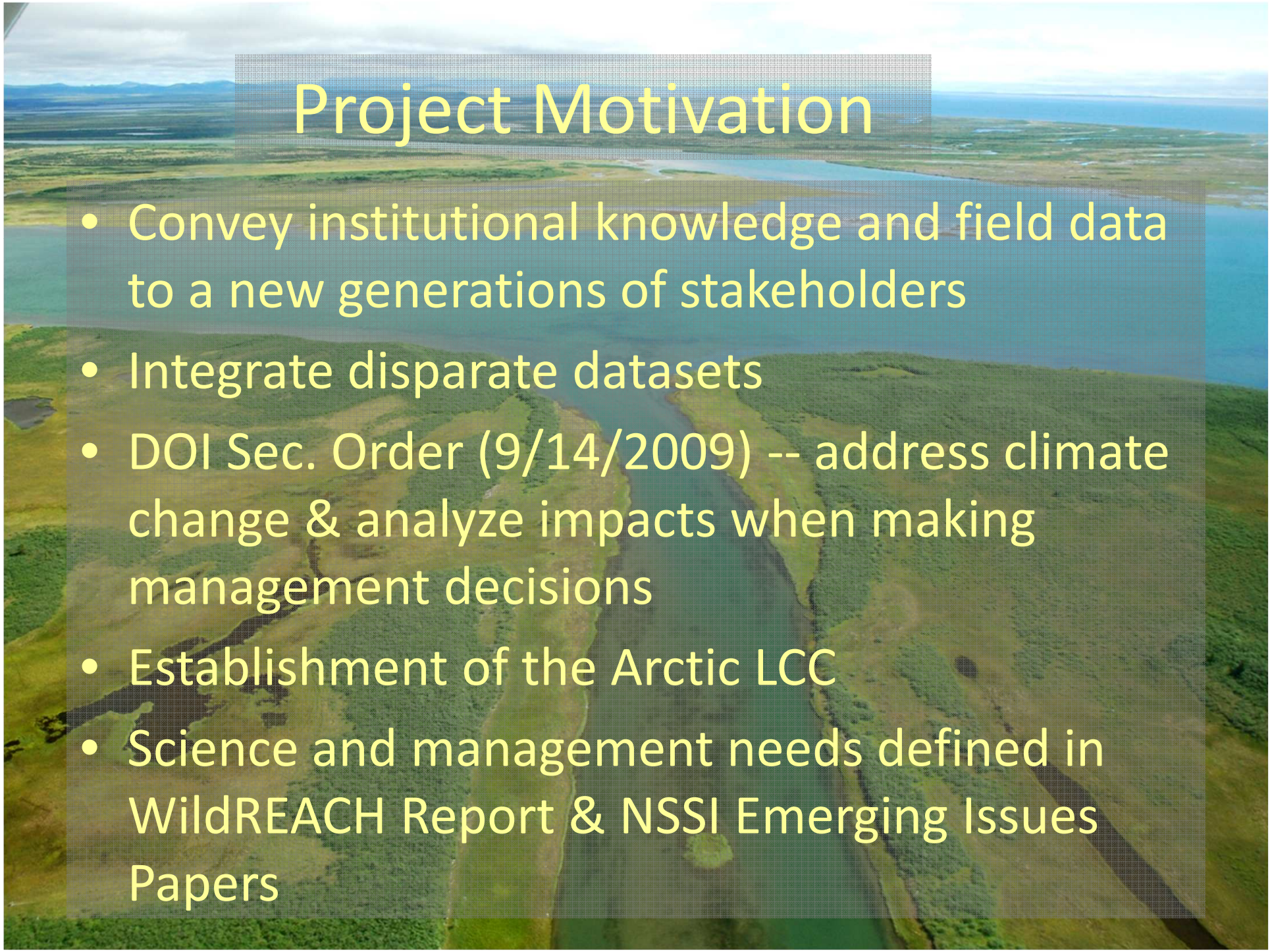
Year 2 Collaborators: Horacio Toniolo, Bill Schnabel

Ongoing Partners: Arctic Landscape Conservation Cooperative, BLM, USFWS

American Water Resource Association: Alaska Chapter, 4-6 April 2011, Chena Hotsprings, AK

Outline

- Motivation
- Objectives
- Phase 1
  - Create a database
  - Acquire and Inventory Metadata
  - Disseminate – interface for queries, visualizations, & downloading data
- Phase 2
  - Populate the database with data
  - Network Analysis



Project Motivation

- Convey institutional knowledge and field data to a new generations of stakeholders
- Integrate disparate datasets
- DOI Sec. Order (9/14/2009) -- address climate change & analyze impacts when making management decisions
- Establishment of the Arctic LCC
- Science and management needs defined in WildREACH Report & NSSI Emerging Issues Papers

Scientific Priorities: WildREACH/NSSI

- How much water is available at local & regional scales?
- Is the Arctic hydrologic cycle undergoing significant change?
- How will these changes affect:
  - water balance, phenology, snow pack, active layer depth, drainage patterns, connections between water bodies, flooding, rain on snow events, etc
- What kind of long-term hydrologic monitoring network is needed in the Arctic?

NSSI Emerging Issues Papers Highlights

- Data collection is currently ad hoc.
- Status of hydrological processes are poorly understood.
- Potential changes are difficult to predict and with the current level of information will be difficult to document.
- Need to:
  - inventory existing stations
  - undertake a gap analysis
  - invest in an improved well-coordinated monitoring network with consistent methods for collecting, storing, managing, and disseminating data

Objectives

Acquire Relevant Climate, Hydrology & Fisheries Metadata and Data

Data in field books, loggers, reports, etc

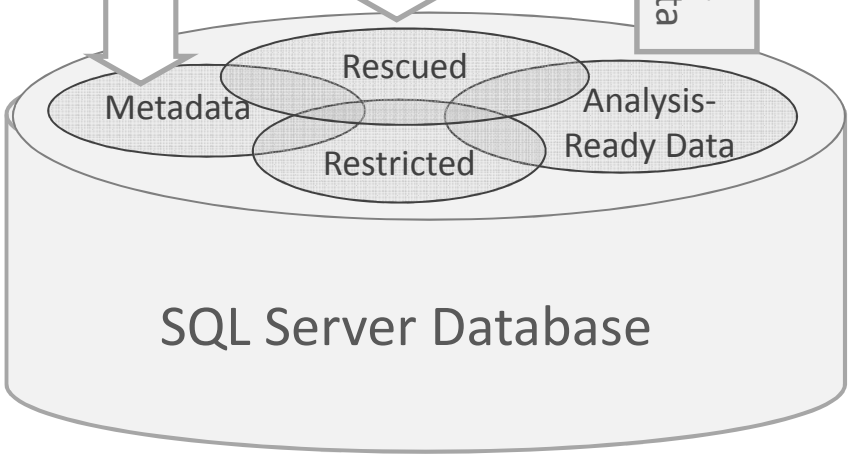
Data in online ODM databases

Metadata Inventory

Data Rescue & Ingestion

Link to Online Data

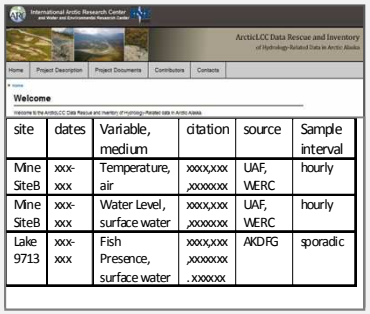
Data & Metadata



Website

I. Geo-Inventory

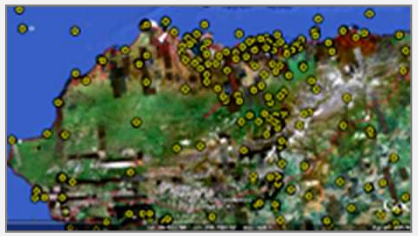
II. Geo-Database

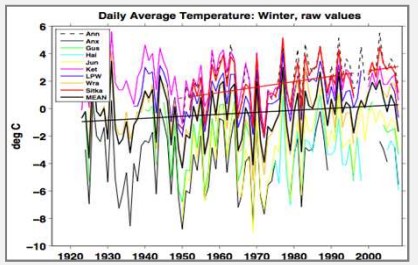


site	dates	Variable, medium	citation	source	Sample interval
Mine SiteB	xxx-xxx	Temperature, air	xxxx,xxx ,xxxxxxx	UAF, WERC	hourly
Mine SiteB	xxx-xxx	Water Level, surface water	xxxx,xxx ,xxxxxxx	UAF, WERC	hourly
Lake 9713	xxx-xxx	Fish Presence, surface water	xxxx,xxx ,xxxxxxx .xxxxxx	AKDFG	sporadic

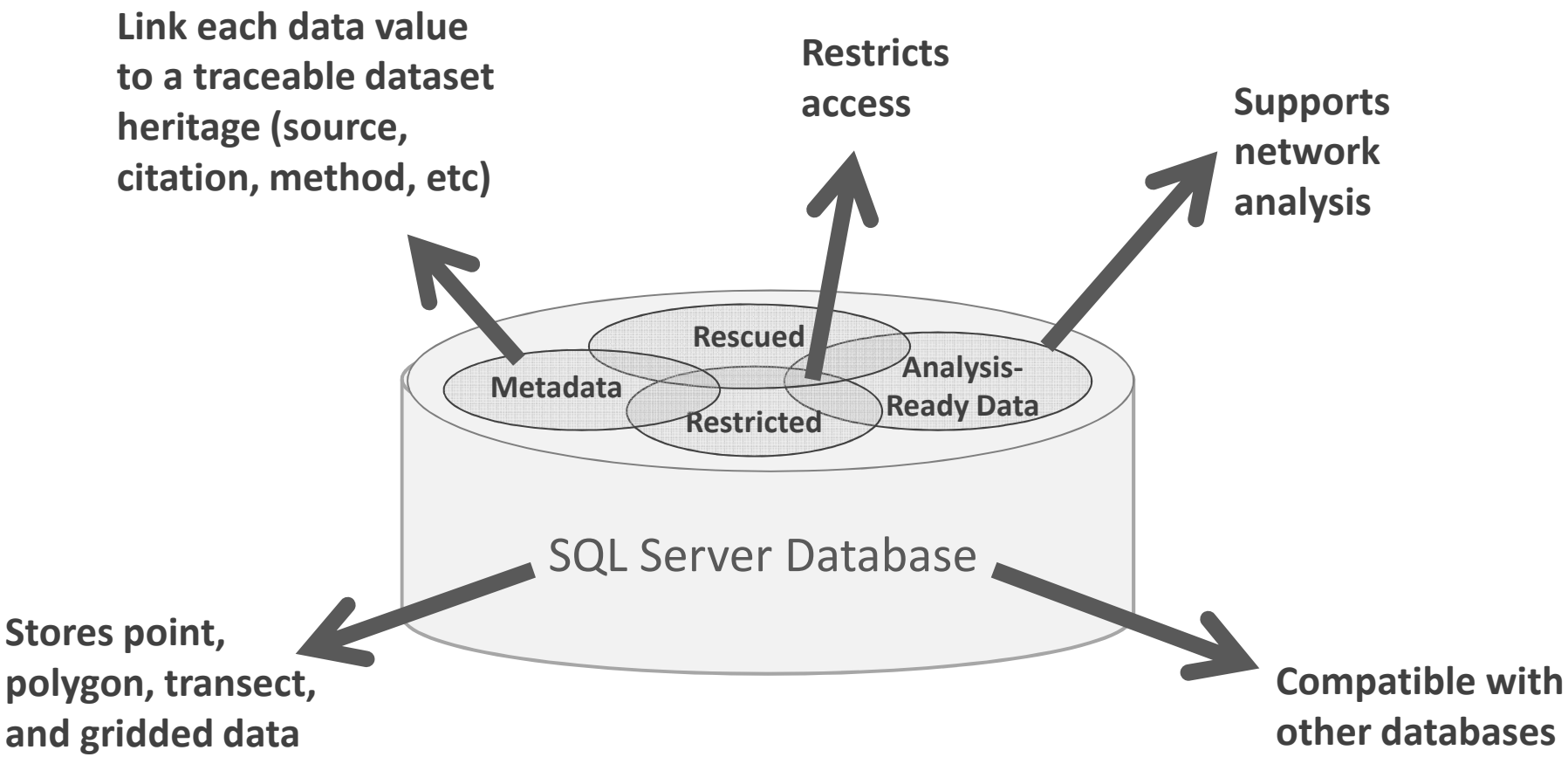
Network Analysis and Report

- Data Density
- Length of Record
- Trend Analysis
- Correlation
- Cost-surface





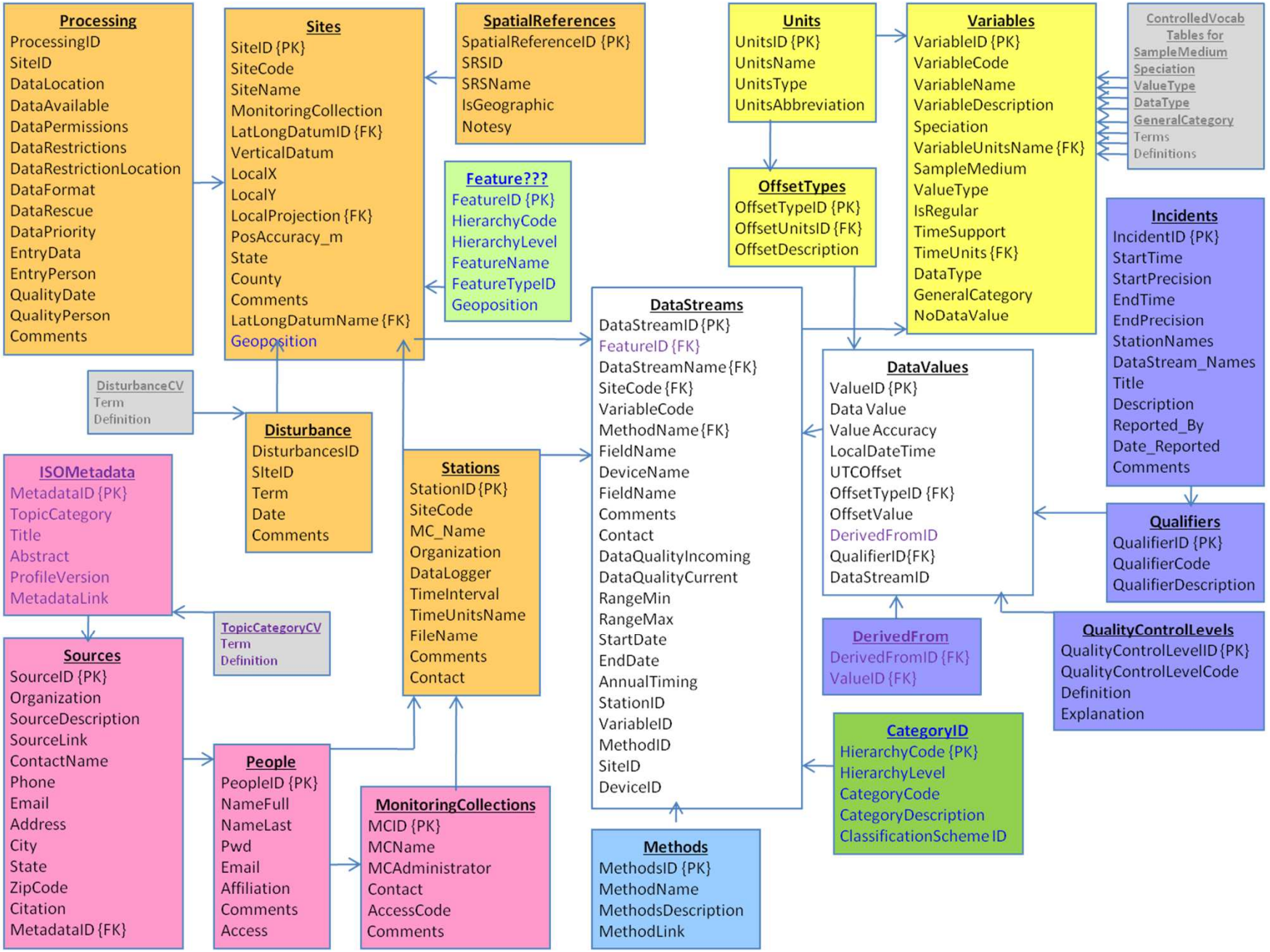
We need a container to hold the data



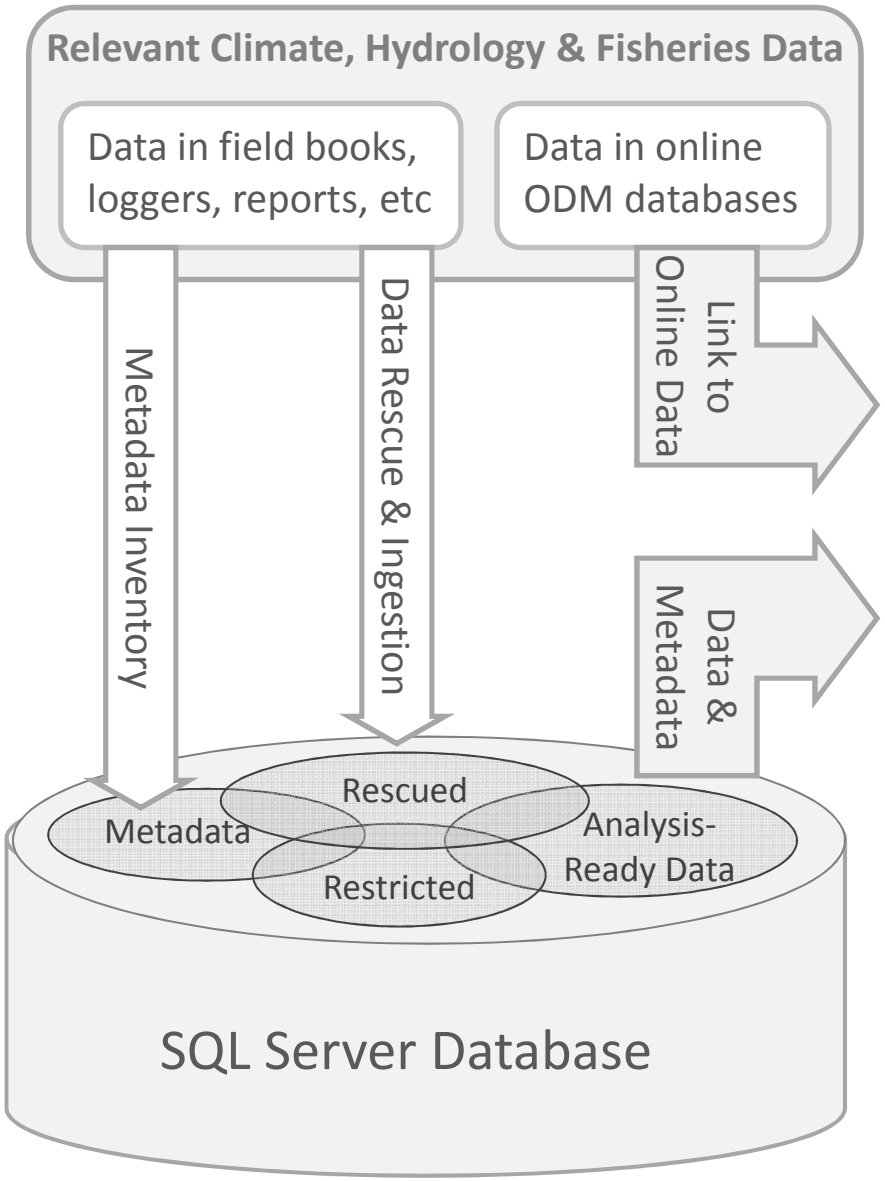
The Container –
Needs Versus Available Models

Needs	CUAHSI ODM	Berkeley Sensor	SciScope	Aquabase Hybrid
Support network analysis				X
Compatible with other ODM databases (WERC, NSDSS, etc)	X	X	X	X
Store point observations	X	X	X	X
Store polygon, gridded, and transect observations			X	X
Store hydrology, climate, and, potentially, biological data	P	P	X	X
Tables to prioritize processing of inventoried data				X
Link each observation to traceable dataset heritage, including appropriate citation and QA/QC information	X	X	X	X
Restrict access to select datasets		X	X	X

Evolving Aquabase Hybrid

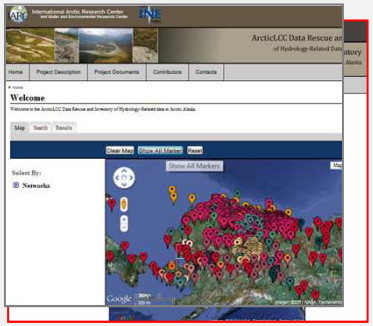


We need an interface for queries, visualizations, and downloading data

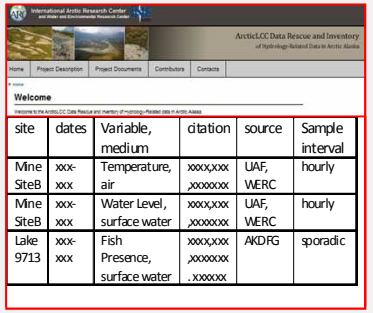


Website

I. Metadata Inventory



II. Geodatabase



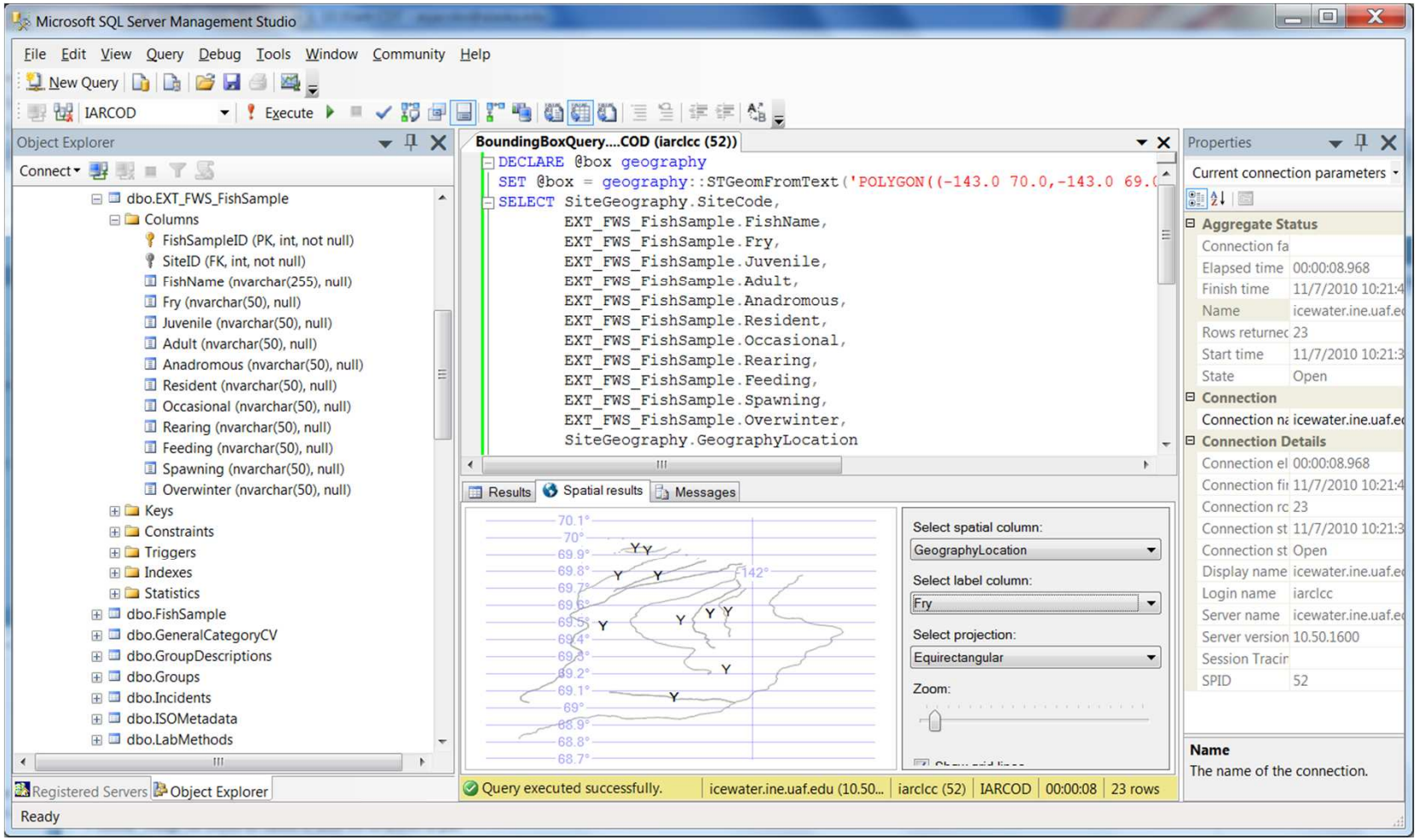
Welcome

site	dates	Variable, medium	citation	source	Sample interval
Mine SiteB	xxx-xxx	Temperature, air	xxxx,xxx ,xxxxxxx	UAF, WERC	hourly
Mine SiteB	xxx-xxx	Water Level, surface water	xxxx,xxx ,xxxxxxx	UAF, WERC	hourly
Lake 9713	xxx-xxx	Fish Presence, surface water	xxxx,xxx ,xxxxxxx .xxxxxx	AKDFG	sporadic

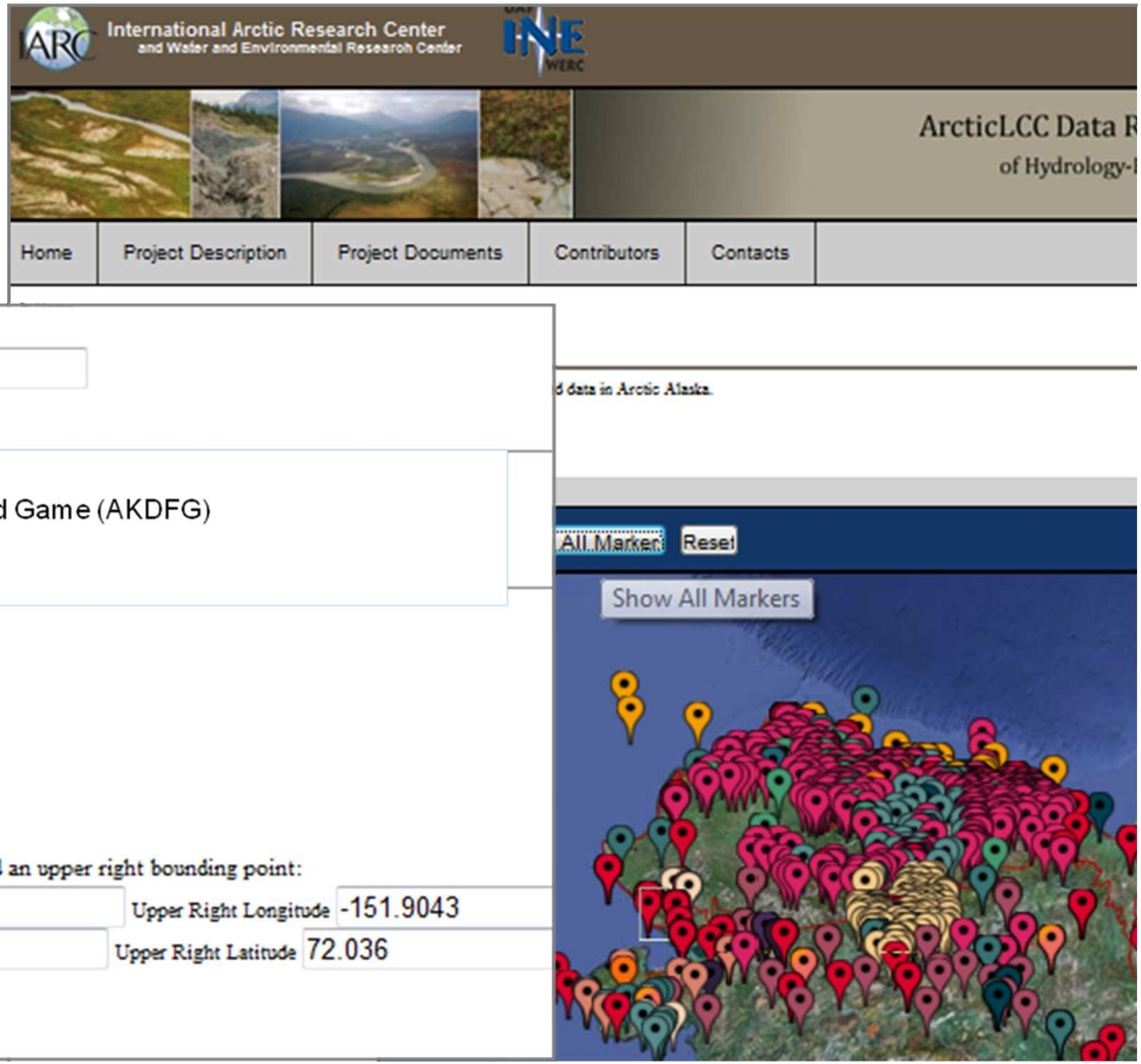
- **Geolocated metadata inventory with pop-ups and links to data sources and datasets**
- **Users will be able to search and download data in their favorite format**

Queries and Data Visualization

Trying to go with off-the-shelf freeware -- SQL Server Management Studio:

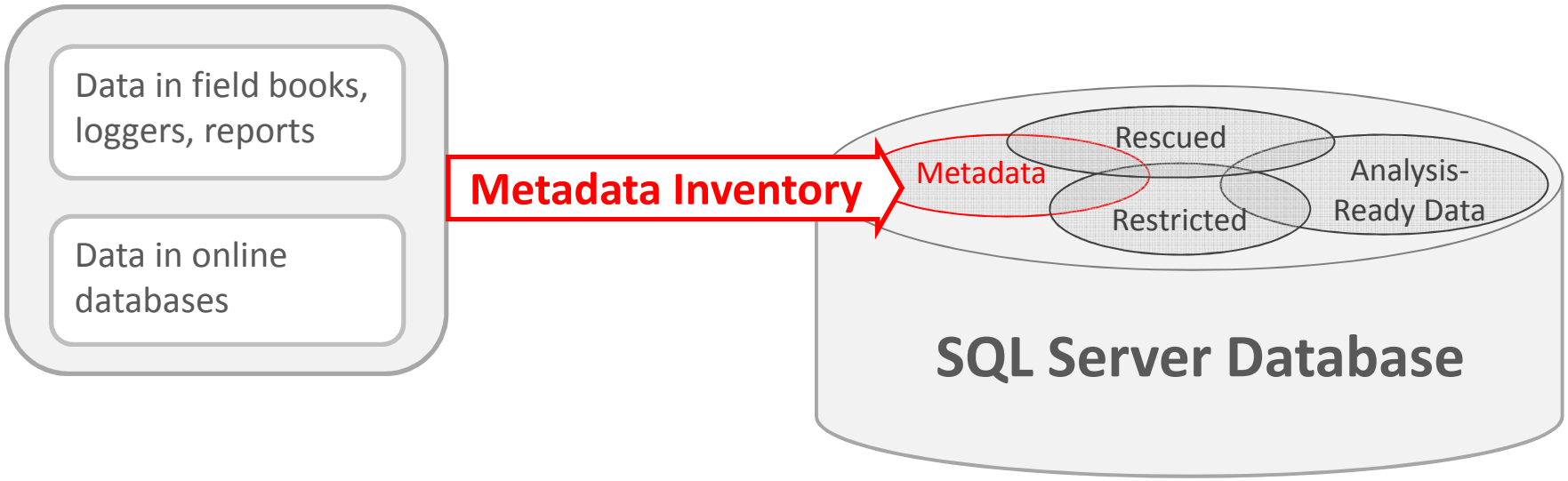


Project Web Site



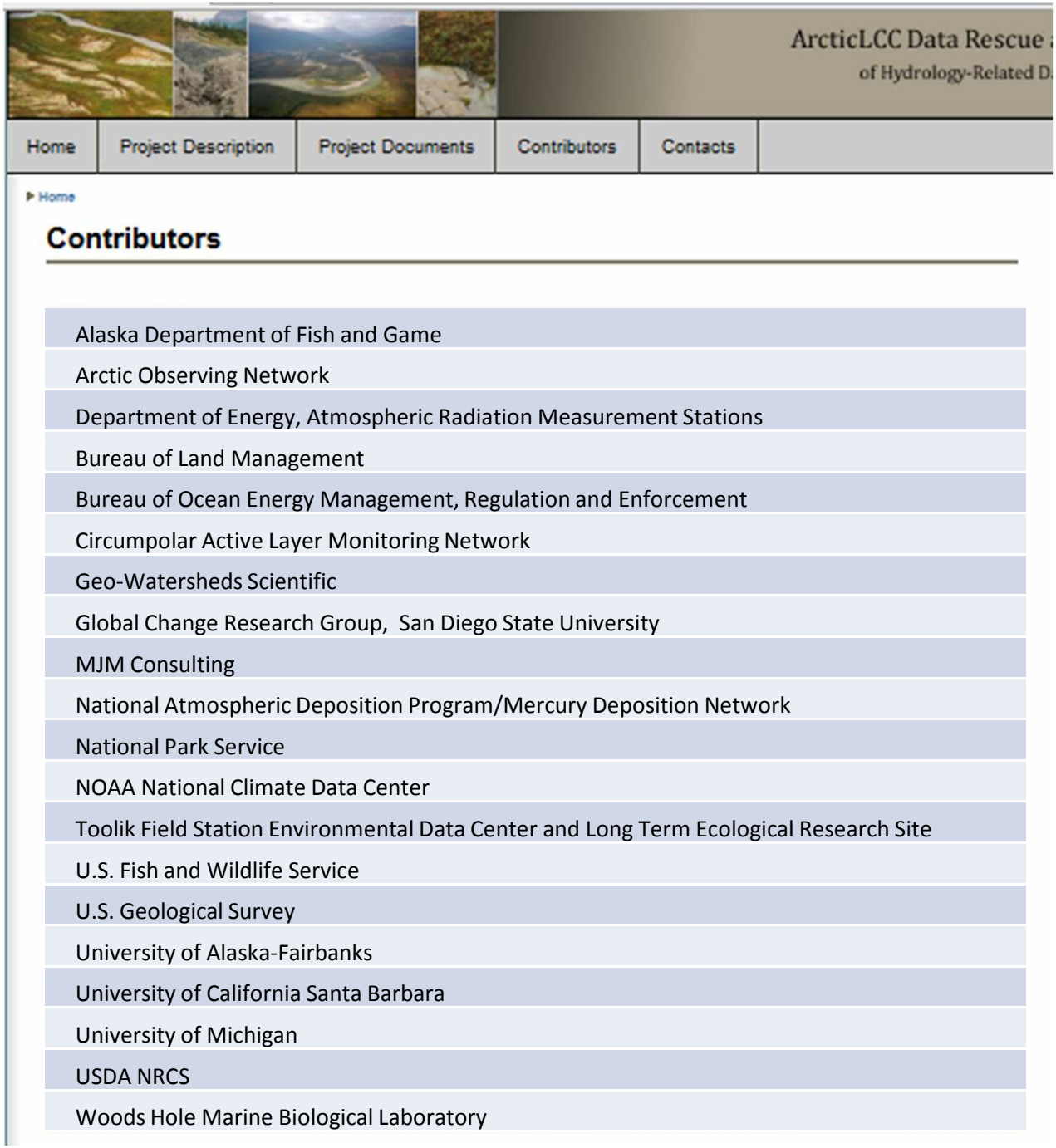
Metadata Inventory

- Identified data sources
- Acquired metadata
- Interpreted, reformatted, standardized, and entered metadata



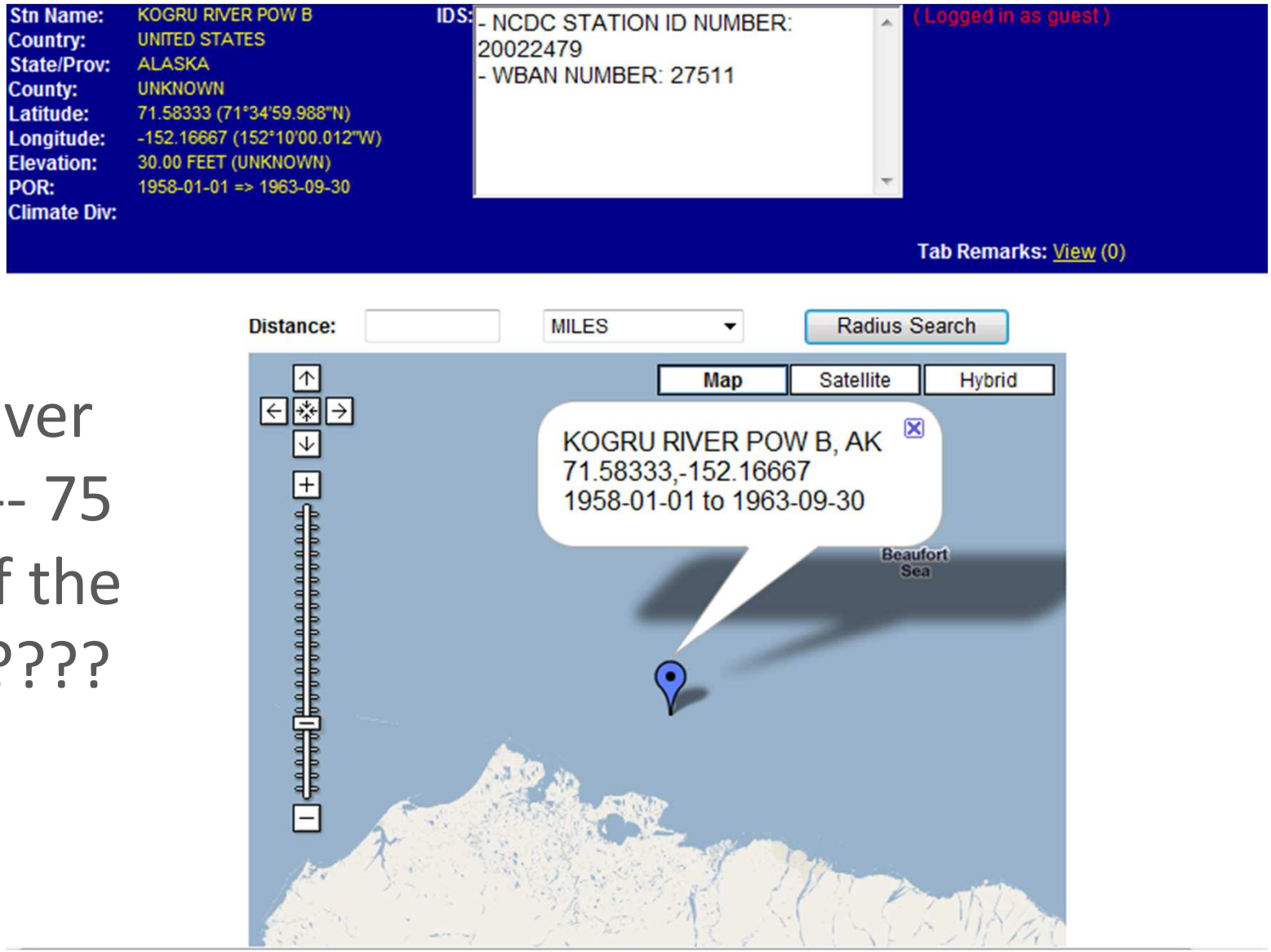
Metadata

- Requests ongoing
- Entered to date:
  - Site (5000)
  - Organizations (20)
  - Source (75)
  - Variables (74)
  - Data Streams (634)



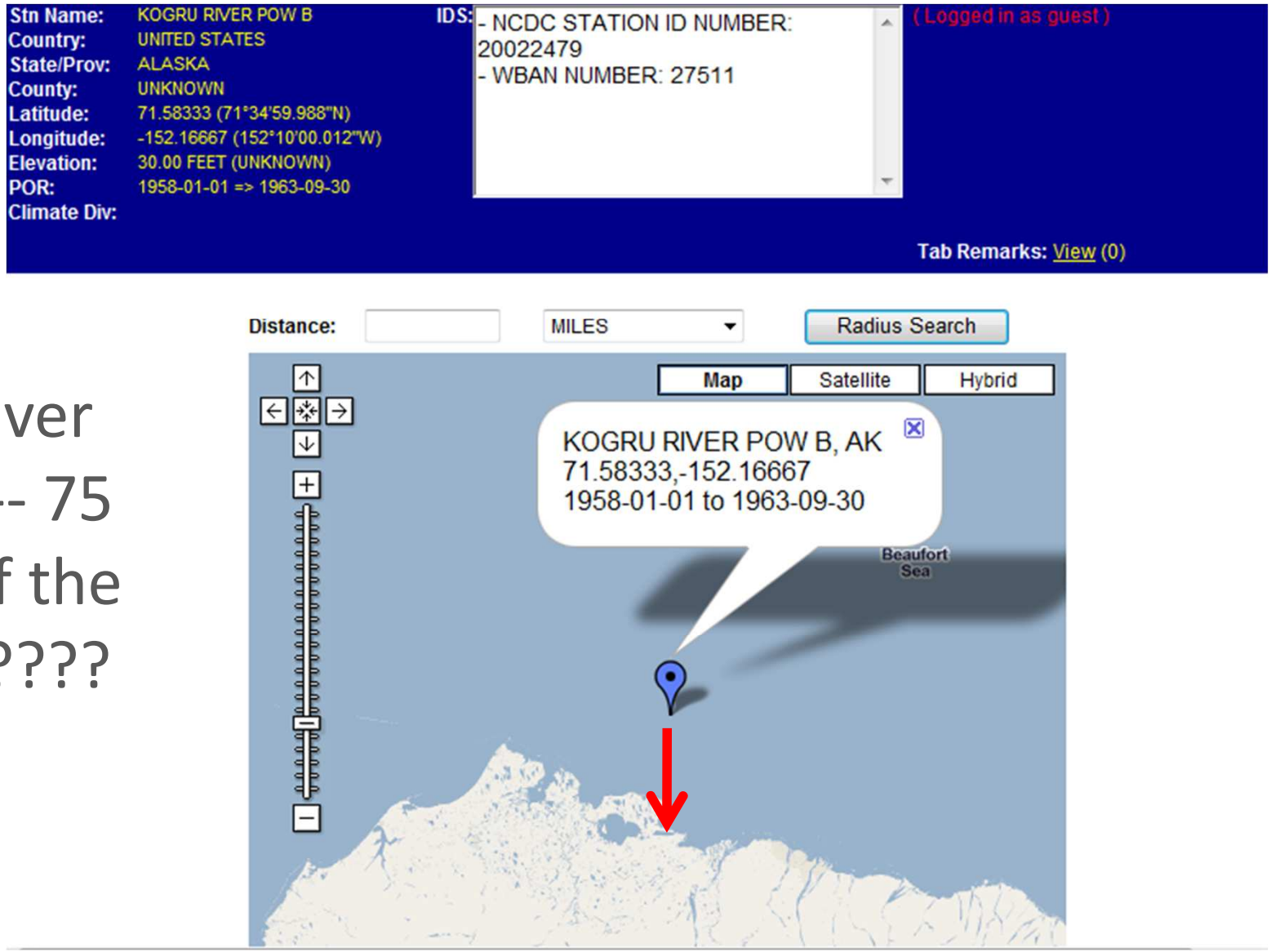
Station Location Issues

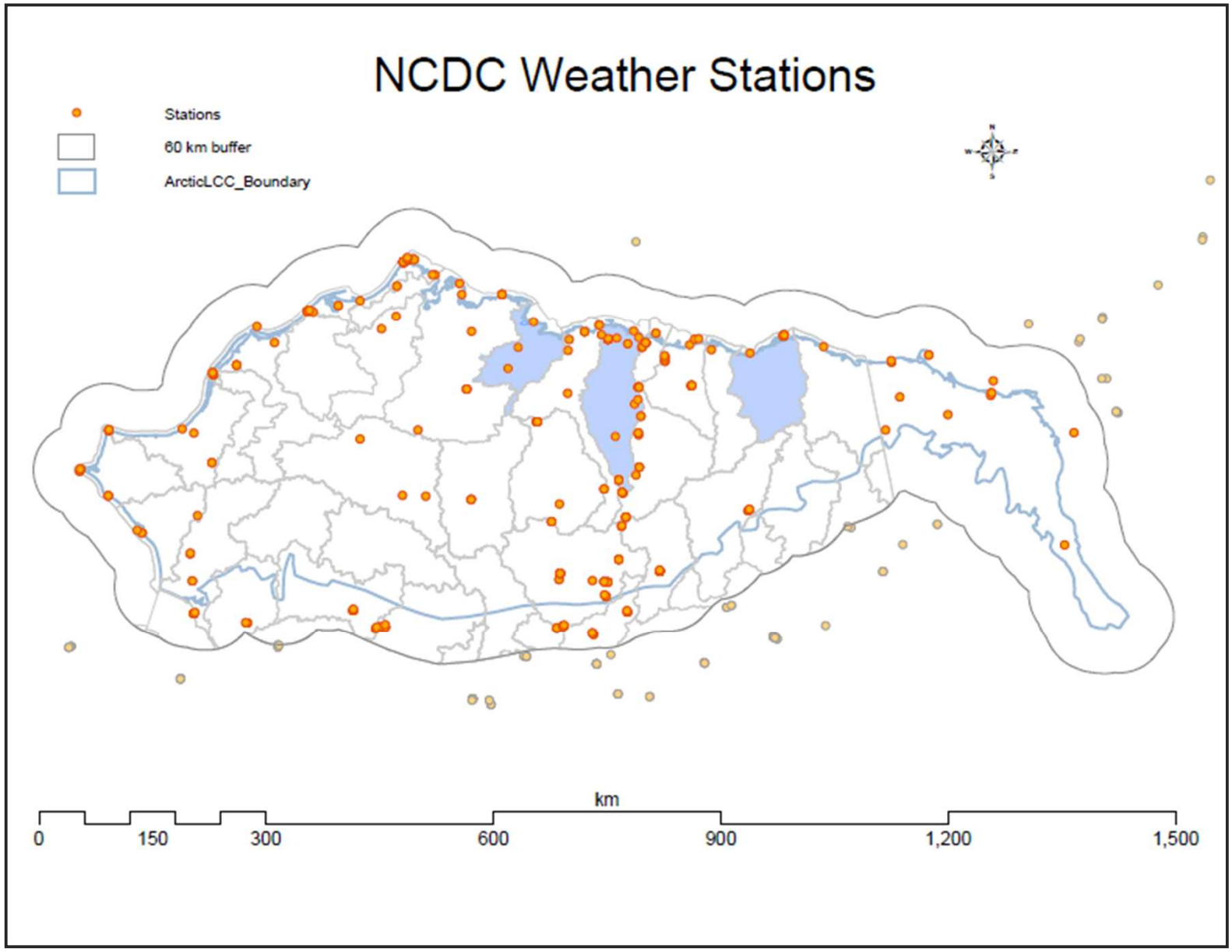
Kogru River Station -- 75 miles off the coast??????



Station Location Issues

Kogru River Station -- 75 miles off the coast??????



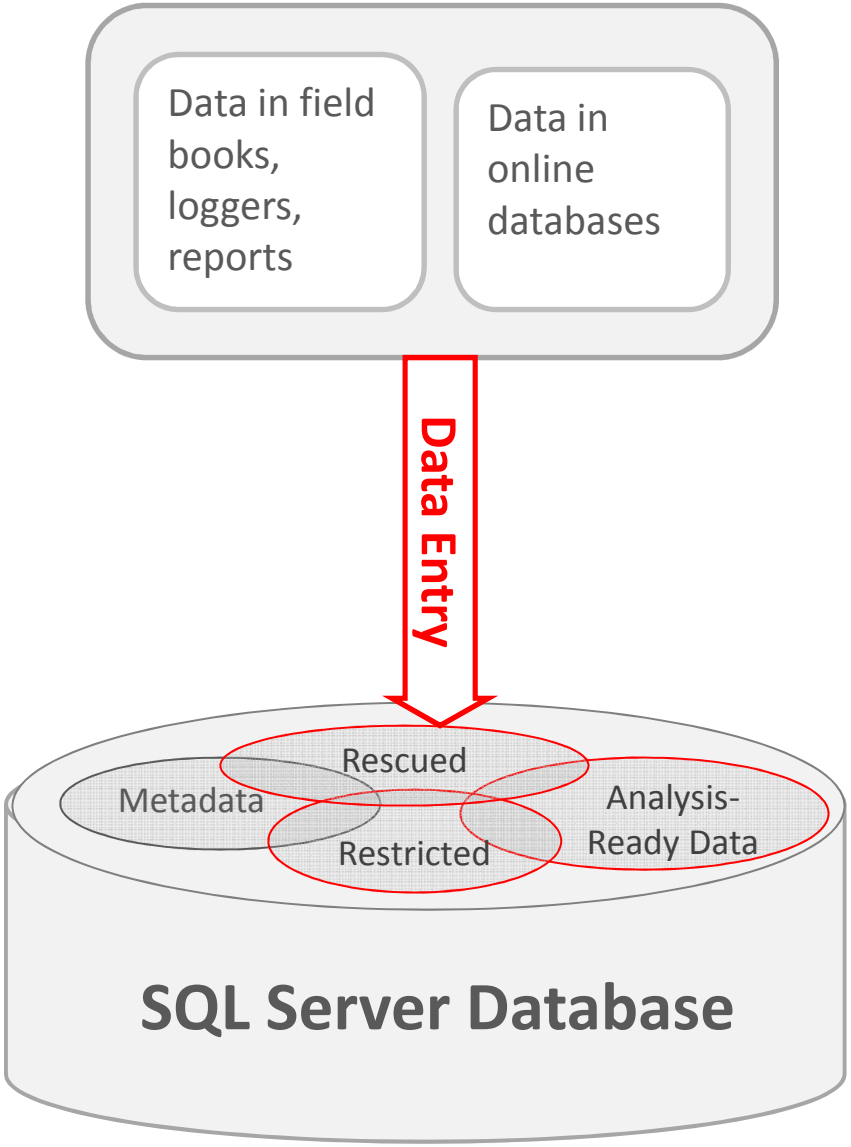

NCDC Weather Stations
Stations
60 km buffer
ArcticLCC_Boundary
km
0
150
300
600
900
1,200
1,500

Phase 2

- Populating the database
- Network analysis

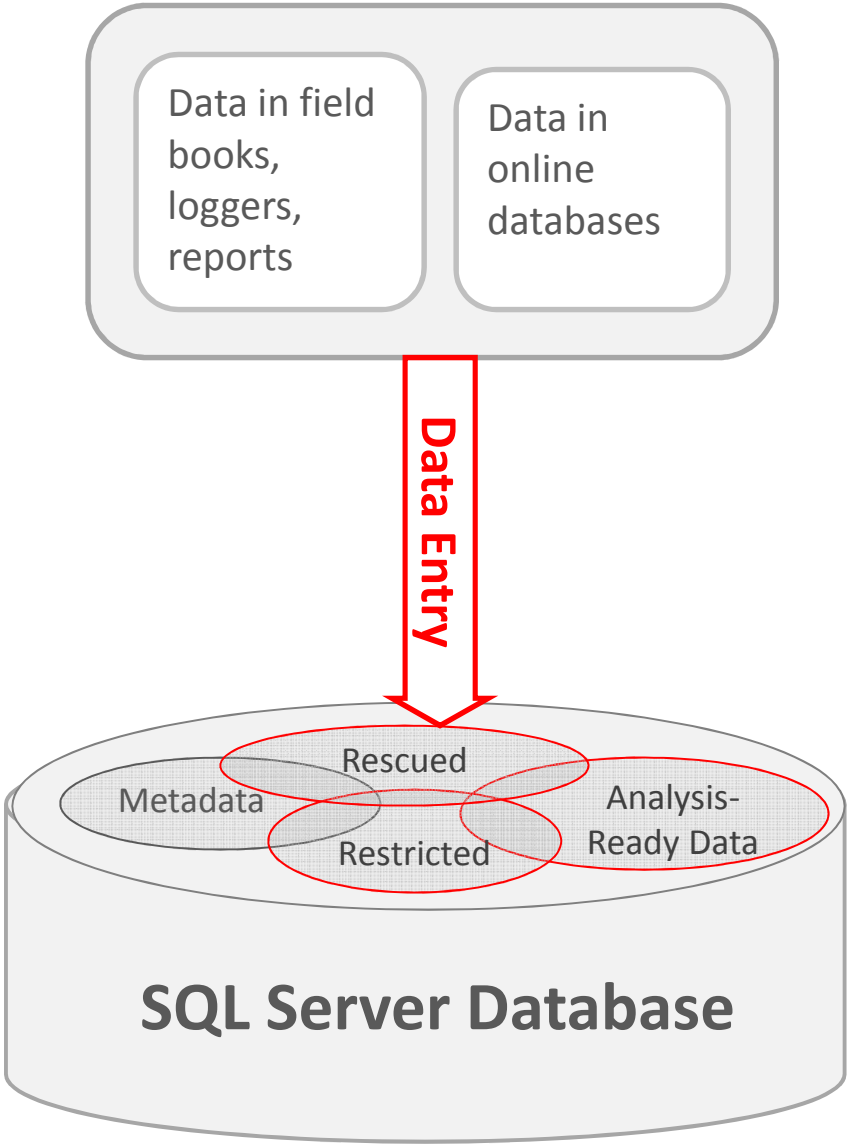
Populating the database

- Acquire and archive data
- Rescue as needed
- Reformat, standardize, and enter data

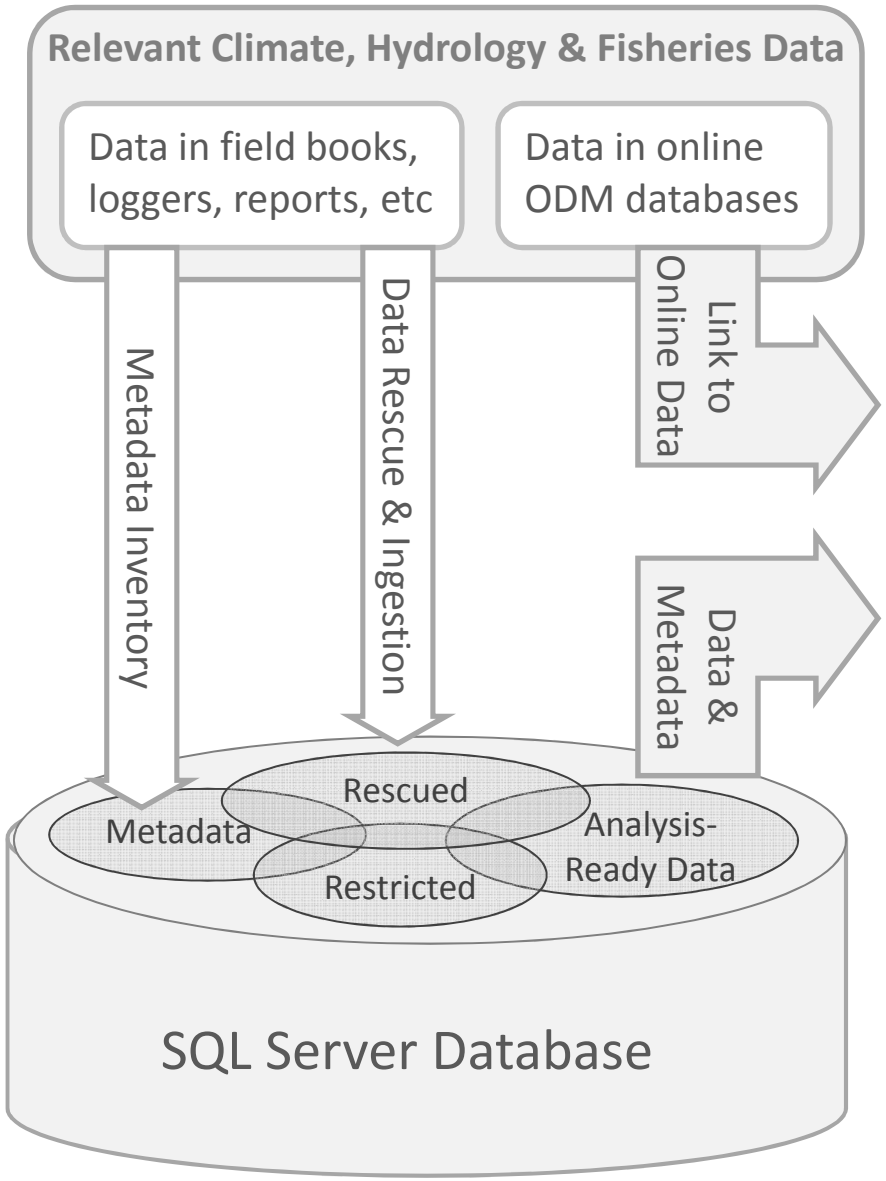


Populating the database

- Acquire and archive data
- Rescue as needed
- Reformat, standardize, and enter data
- To date:
  - 216,000 data values for multiple variables at 150 sites
  - Millions of raw data values (1 site)



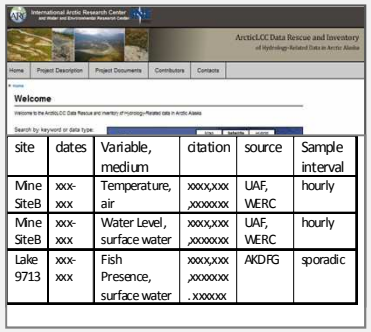
Network Analysis



Website

I. Geo-Inventory

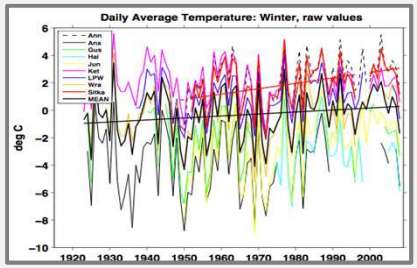
II. Geo-Database



site	dates	Variable, medium	citation	source	Sample interval
Mine SiteB	xxx-xxx	Temperature, air	xxxx,xxx ,xxxxxx	UAF, WERC	hourly
Mine SiteB	xxx-xxx	Water Level, surface water	xxxx,xxx ,xxxxxx	UAF, WERC	hourly
Lake 9713	xxx-xxx	Fish Presence, surface water	xxxx,xxx ,xxxxxx .xxxxx	AKDFG	sporadic

Network Analysis and Report

- **Data Density**
- **Length of Record**
- **Trend Analysis**
- **Correlation**
- **Cost-surface**



Network Analysis

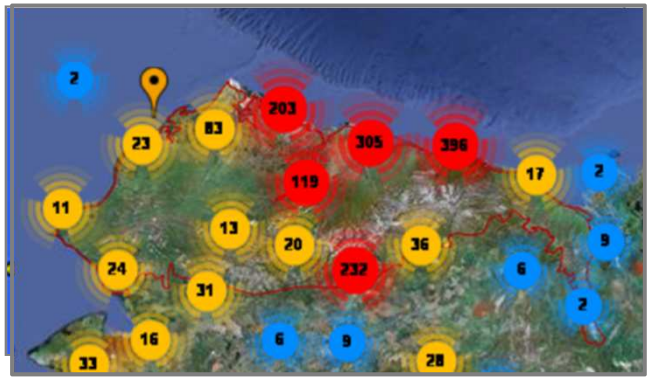
Address WildREACH/NSSI priorities

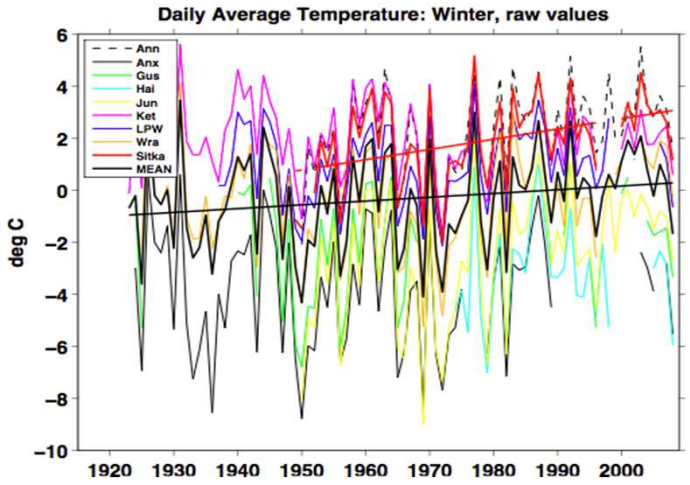
- Temporal and spatial trends
- Power to detect change
- Assess reliability of projections

Recommendations for monitoring networks will be based on

- Historic data density and period of record
- Major gaps
- Assess potential station modifications to increase data consistency
- Practicality -- cost surface analysis

Wrap up in February 2012





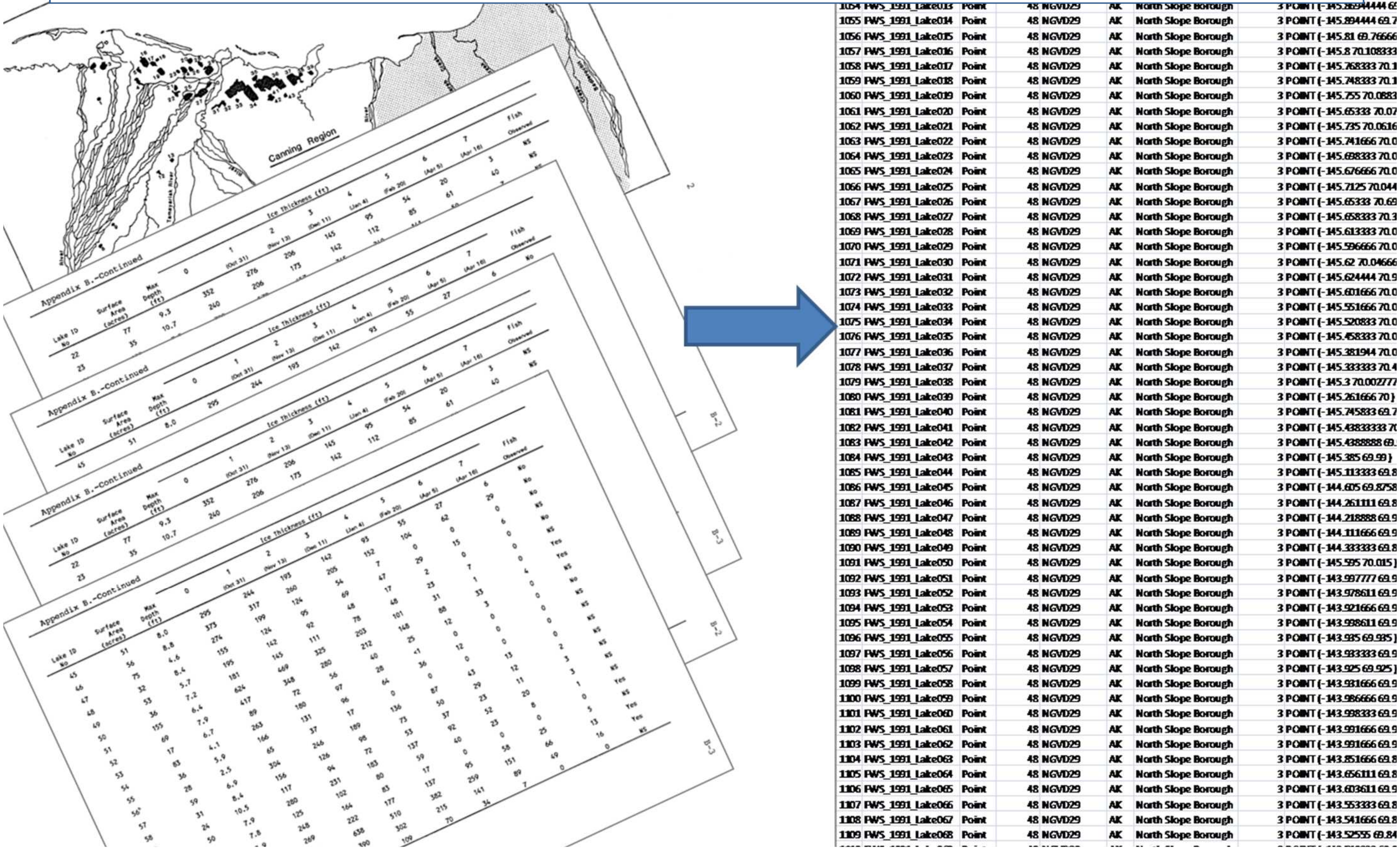
Acknowledgements

Amy Tidwell, Anja Kade, Anna Liljedahl, Ben Jones, Bill Schnabel, Bob Busey, Bruce Peterson, Chris Arp, Craig Stephenson, Dave Daum, Emily Youcha, Frank Urban, Gary Clow, Gaius Shaver, George Kling, Greta Myerchin, James Long, Jason Stuckey, Jennifer Jenkins, Joe Buckwalter, John Trawicki, Katrina Bennett, Kathleen Wedemeyer, Ken Hill, Lawrence Moulton, Matt Barkdull, Michael Lilly, Pam Sousanes, Peter Prokein, Philip Martin, Richard Kemnitz, River Gates, Rob Gieck, Scott Miller, Sveta Stuefer, Steve Frenzel, Sydonia Bret-Harte, Steve Kendall, Trevor Haynes, Walter Oechel, Matthew Whitman, Arctic LCC, NOAA, USDA-NRCS, WRCC, & many others.

gburkart@alaska.edu

	Science/Management Question	General Database Query
Network analysis, general management, and data for models	What variables were sampled on Lake 9817 or within bounding box (example: 70N 68N 150W 155W)?	All variables for all time for all space within a specified geographic extent (basin, reach, river, site)
	What snow water equivalent data are available on BLM land in the Arctic LCC?	All data for one variable for all time for a certain area
	Where has ice thickness been measured?	Sites (space), identified by variable.
	What variables under the general category of climate were measured at which sites within the bounding box, a particular land management unit, or entire Arctic LCC?	All variables under the General Category of Climate for all time for all space within a specified geographic extent (basin, reach, river, site)
	What reports or articles are available for Lake 9312?	All documents by specified geographic extent (basin, etc)
Management questions	What winter precipitation data are available for areas that had unusually low musk ox survivorship during 199X, 200X, and 200X?	All precipitation data for specific areas during selected years
	Which lakes in development area X are deep enough to support fish?	All data for variable X (water depth, maximum) in geographic extent X
	What data are available to understand fluctuations in water level in lakes where ice roads are planned?	All water level data under general category of inland waters for all

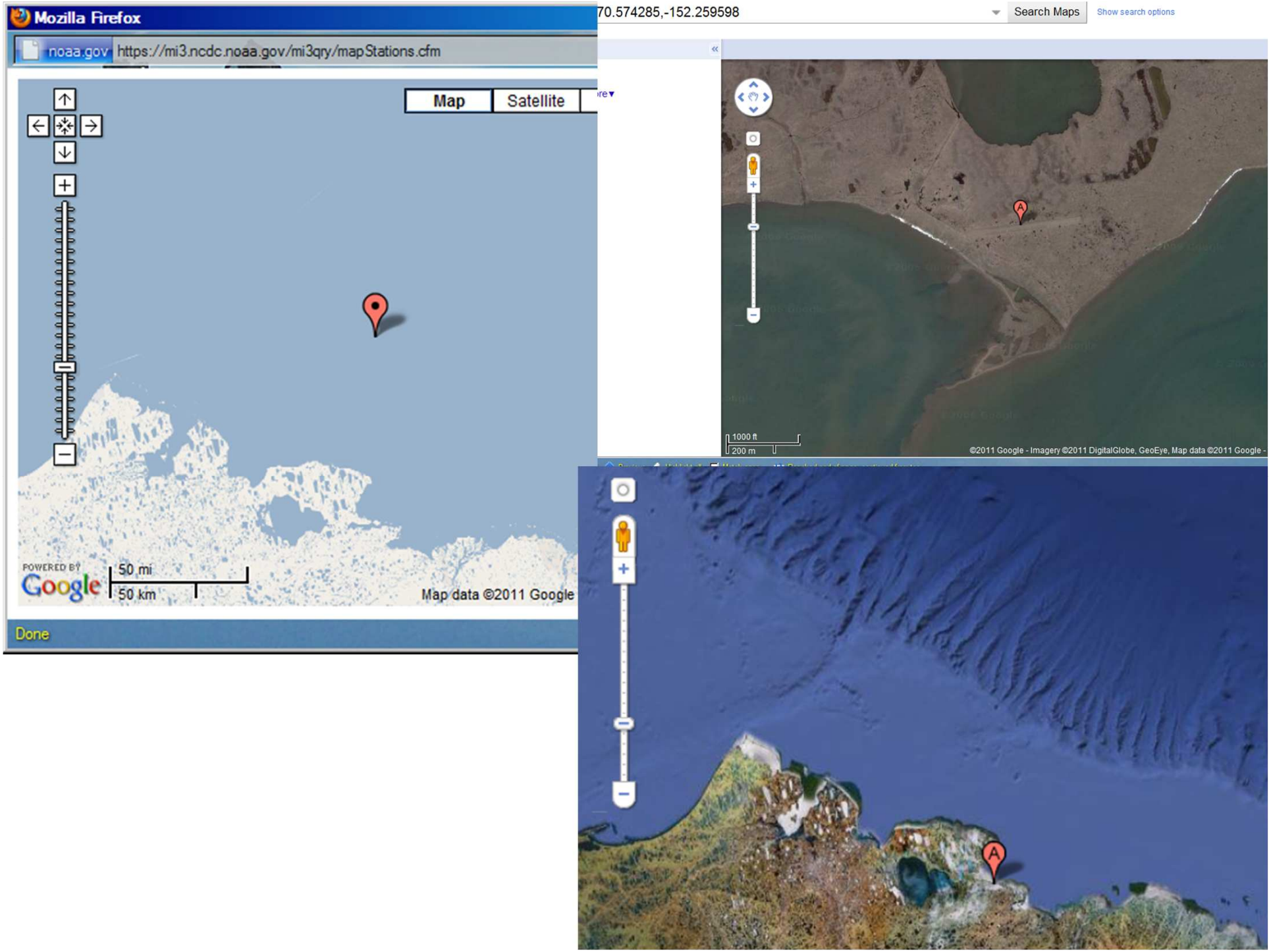
Data Rescue, Normalization, and Ingestion



Station Identity Issues????

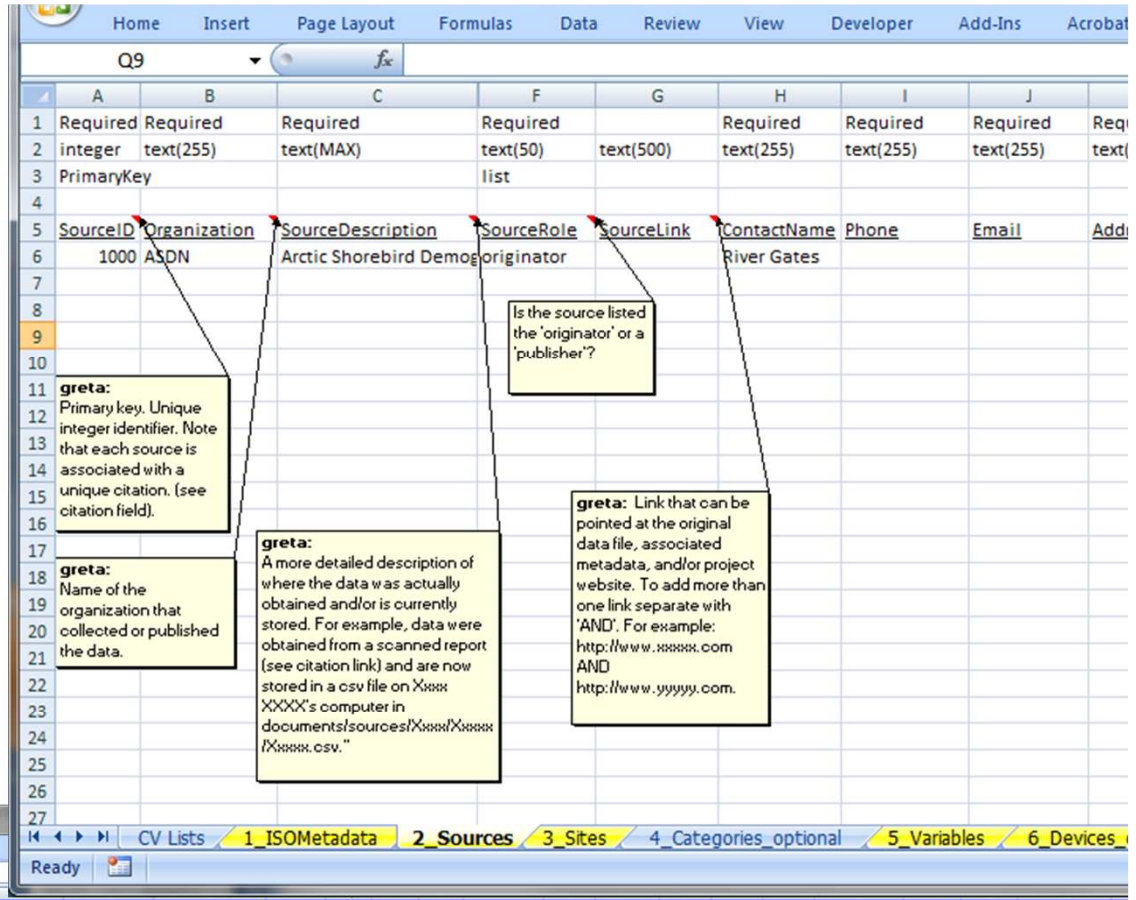
Kotzebue WSO AP, Kotzebue Ralph Wein Memorial, Kotzebue FAA AP, Ralph Wein Memorial AP

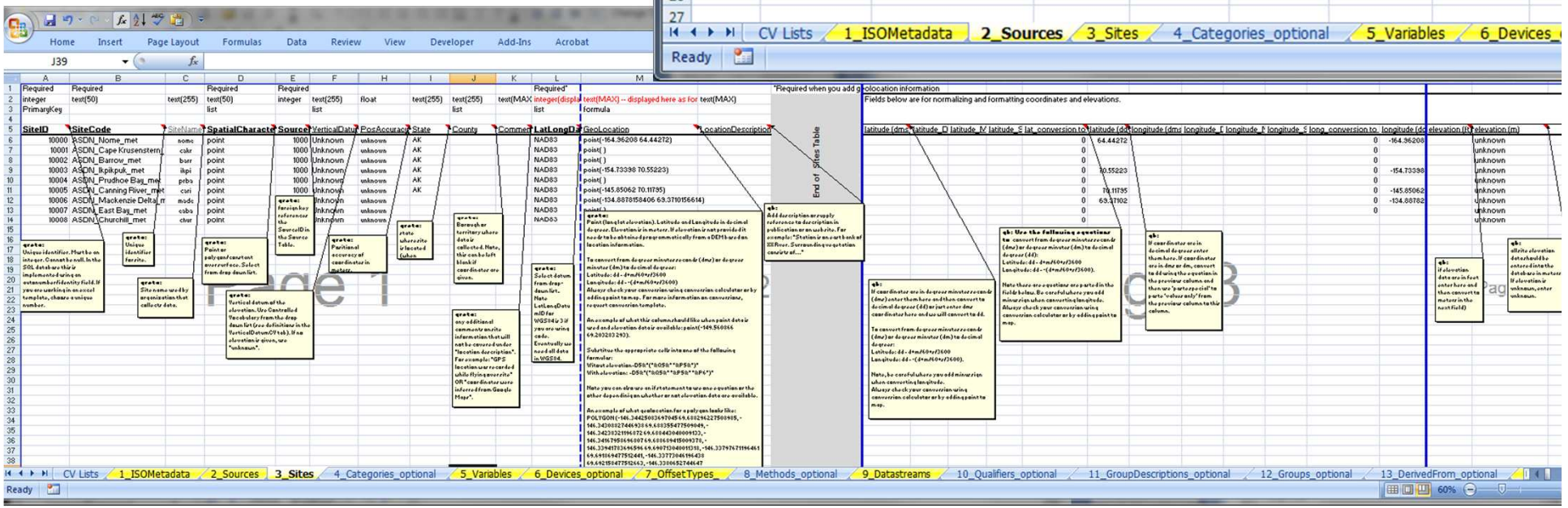
Begin Date	End Date	Stn Name	Name Type	Stn Type	COOP	WBAN
[2010-11-25]	Current	KOTZEBUE WSO AP	COOP NAME	ASOS, ASOS-NWS, COOP, COOP-A, COOPERATIVE SUB-NETWORK- A, LAND SURFACE	505076	26616
		KOTZEBUE RALPH WEIN MEMORIAL AP	PRINCIPAL NAME			
[1997-12-01]	2010-11-25	KOTZEBUE WSO AIRPORT	COOP NAME	ASOS, ASOS-NWS, COOP, LAND SURFACE	505076	26616
		KOTZEBUE RALPH WEIN MEMORIAL AP	PRINCIPAL NAME			
[1981-12-31]	1997-12-01	KOTZEBUE WSO AIRPORT	COOP NAME	COOP, LAND SURFACE	505076	26616
		KOTZEBUE RALPH WEIN MEMORIAL AP	PRINCIPAL NAME			
[1973-01-01]	1981-12-31	KOTZEBUE WSO AIRPORT	COOP NAME	COOP, LAND SURFACE, WSO	505076	26616
		KOTZEBUE RALPH WEIN MEMORIAL AP	PRINCIPAL NAME			
[1969-01-01]	1973-01-01	KOTZEBUE WSO AIRPORT	COOP NAME	COOP, LAND SURFACE, WBO	505076	26616
		KOTZEBUE RALPH WEIN MEMORIAL AP	PRINCIPAL NAME			
[1962-12-01]	1969-01-01	KOTZEBUE WSO AIRPORT	COOP NAME	COOP, LAND SURFACE, WBAS	505076	26616
		KOTZEBUE RALPH WEIN MEMORIAL AP	PRINCIPAL NAME			
[1946-09-01]	1962-12-01	KOTZEBUE WSO AIRPORT	COOP NAME	COOP, LAND SURFACE, WBAS	505076	26616
		KOTZEBUE FAA AP	PRINCIPAL NAME			
[1944-01-01]	1946-09-01	KOTZEBUE FAA AP	PRINCIPAL NAME	LAND SURFACE, WBAS	---	26616
[1943-01-01]	1944-01-01	KOTZEBUE FAA AP	PRINCIPAL NAME	LAND SURFACE, WBAS	---	26616
[1942-10-26]	1943-01-01	KOTZEBUE FAA AP	PRINCIPAL NAME	LAND SURFACE, SA	---	26616
[1930-09-01]	1942-10-26	RALPH WEIN MEMORIAL AP	PRINCIPAL NAME	LAND SURFACE, SA	---	26616


Mozilla Firefox
noaa.gov
https://mi3.ncdc.noaa.gov/mi3qry/mapStations.cfm
Map
Satellite
POWERED BY Google
50 mi
50 km
Map data ©2011 Google
Done
70.574285,-152.259598
Search Maps
Show search options
1000 ft
200 m
©2011 Google - Imagery ©2011 DigitalGlobe, GeoEye, Map data ©2011 Google -

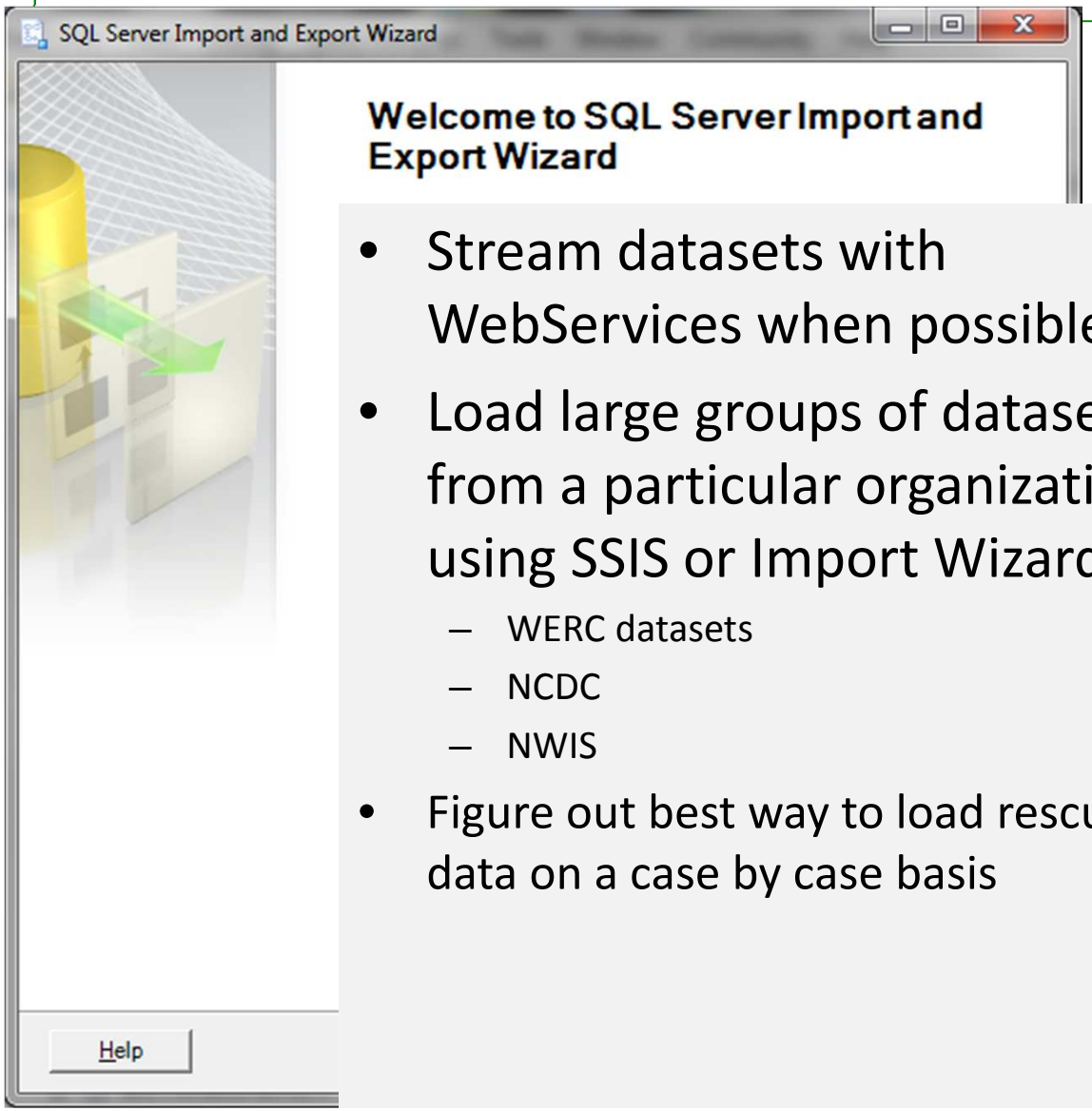
Template

- Supplemented by SOP for appending database
- 14 Tables (worksheets)
- 13 lists with controlled vocab
- Consolidate & simplify template → load with SSIS or import wizard

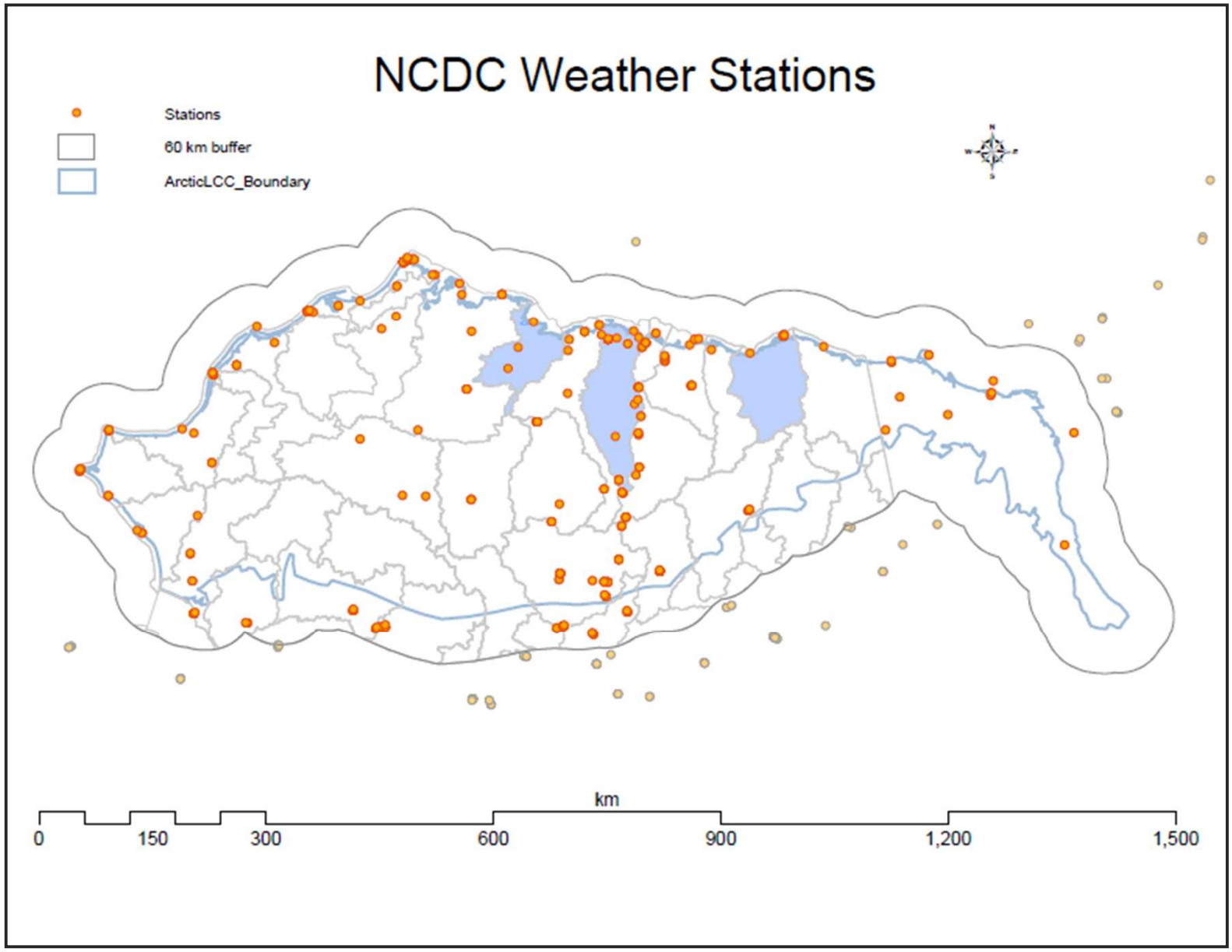




Loading Data



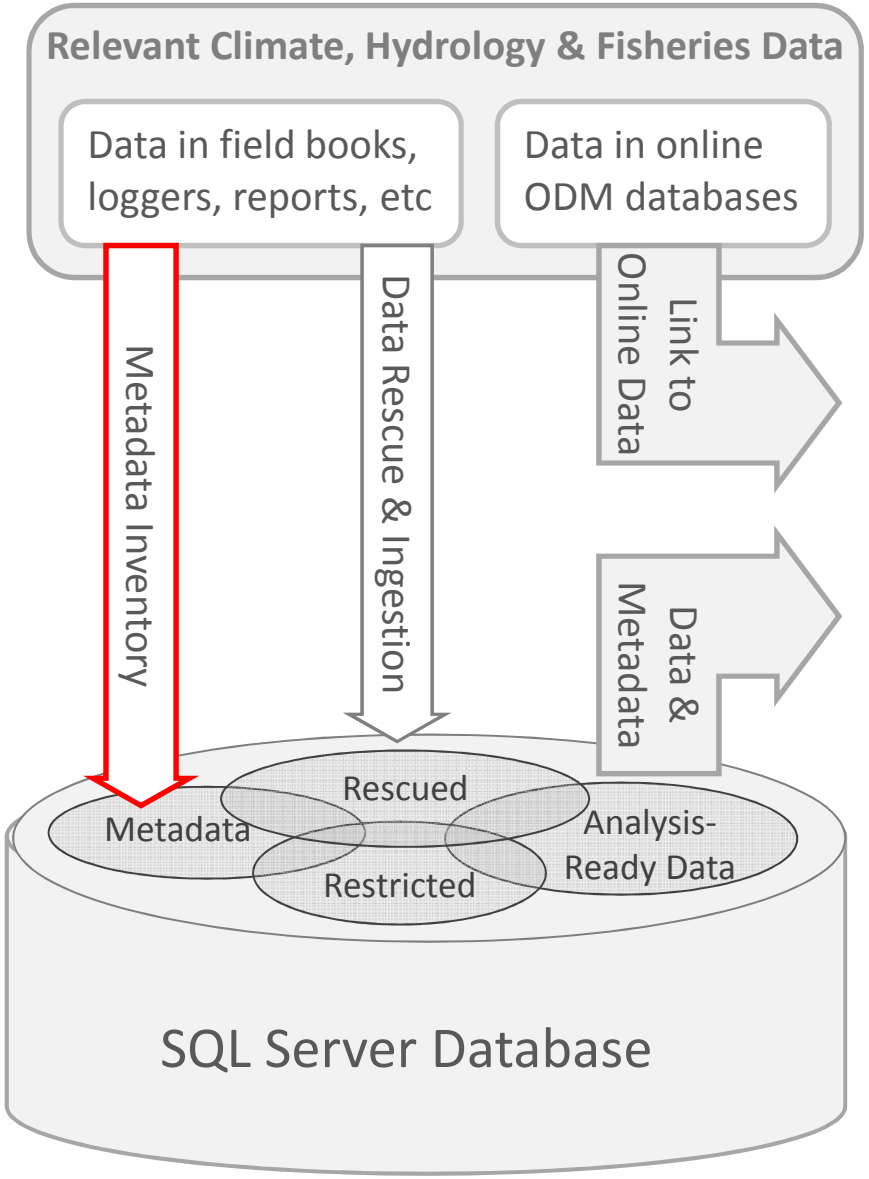
- Stream datasets with WebServices when possible
- Load large groups of datasets from a particular organization using SSIS or Import Wizard
  - WERC datasets
  - NCDC
  - NWIS
- Figure out best way to load rescued data on a case by case basis

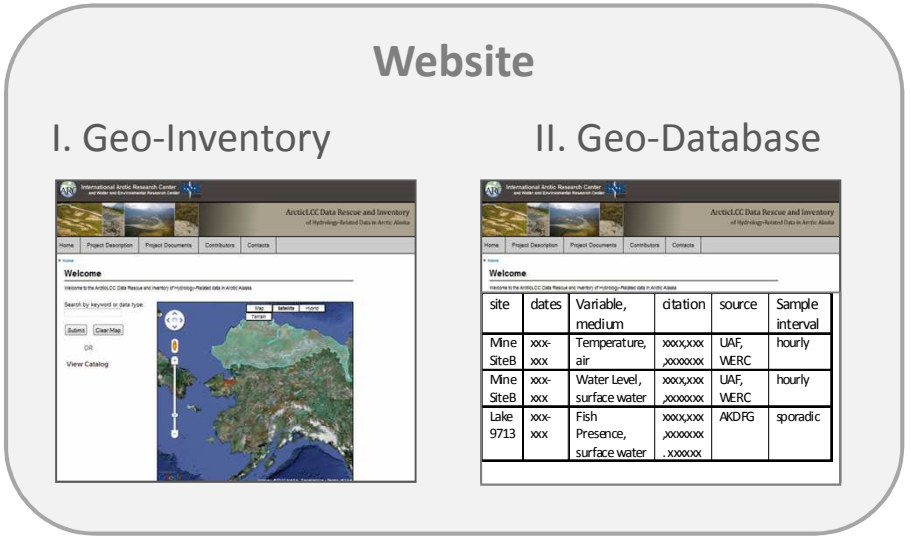

NCDC Weather Stations
Stations
60 km buffer
ArcticLCC_Boundary
km
0
150
300
600
900
1,200
1,500

Phase 1 -- So let's get started

- We need a container to hold the data
- We need an interface for queries, visualizations, and downloading data
- We need to acquire and ingest metadata

We need to acquire and ingest metadata

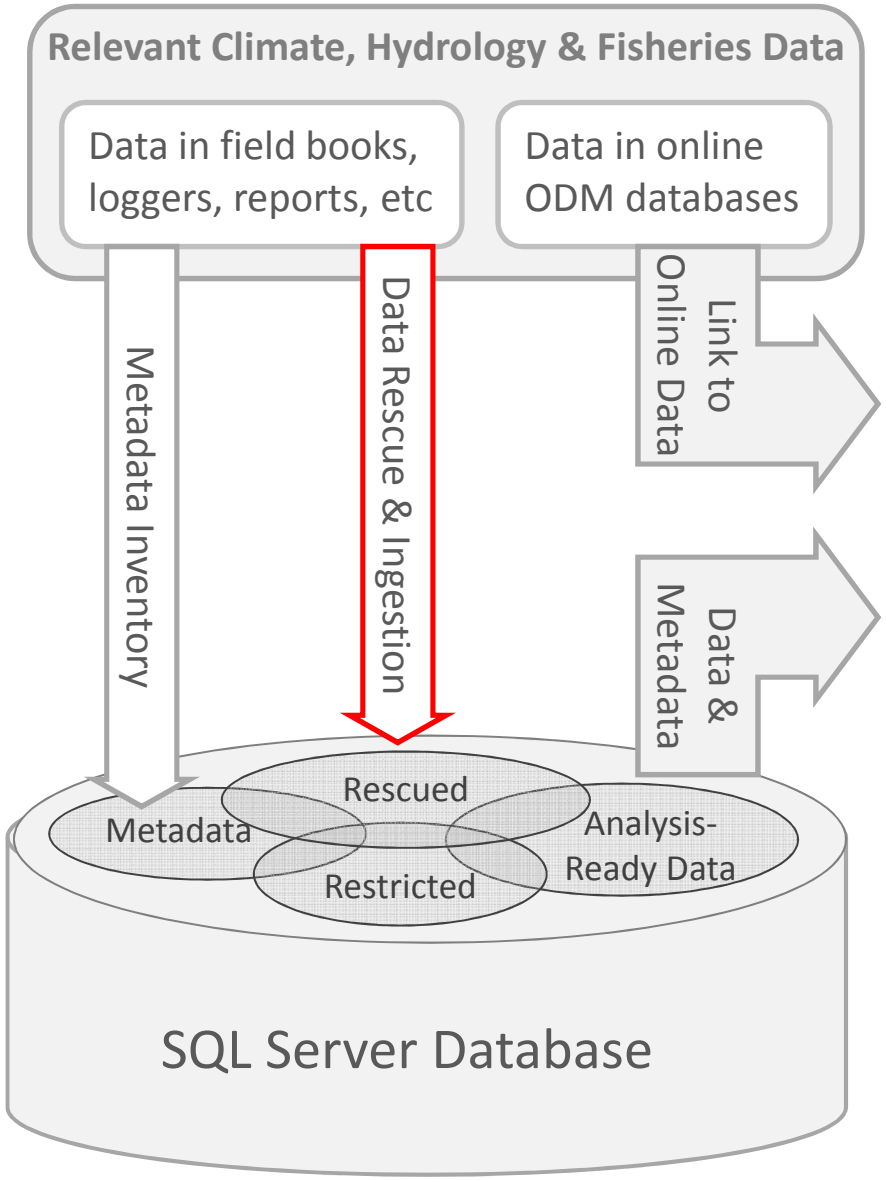


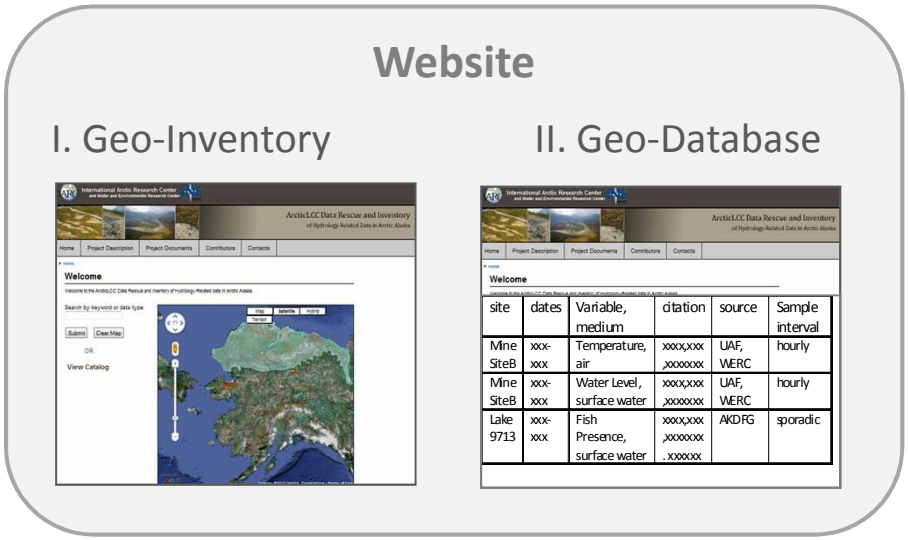


Metadata Inventory

- Identify data sources
- Acquiring metadata
- Interpret, format, standardize, create and upload metadata

Phase 2 – Populating the Database

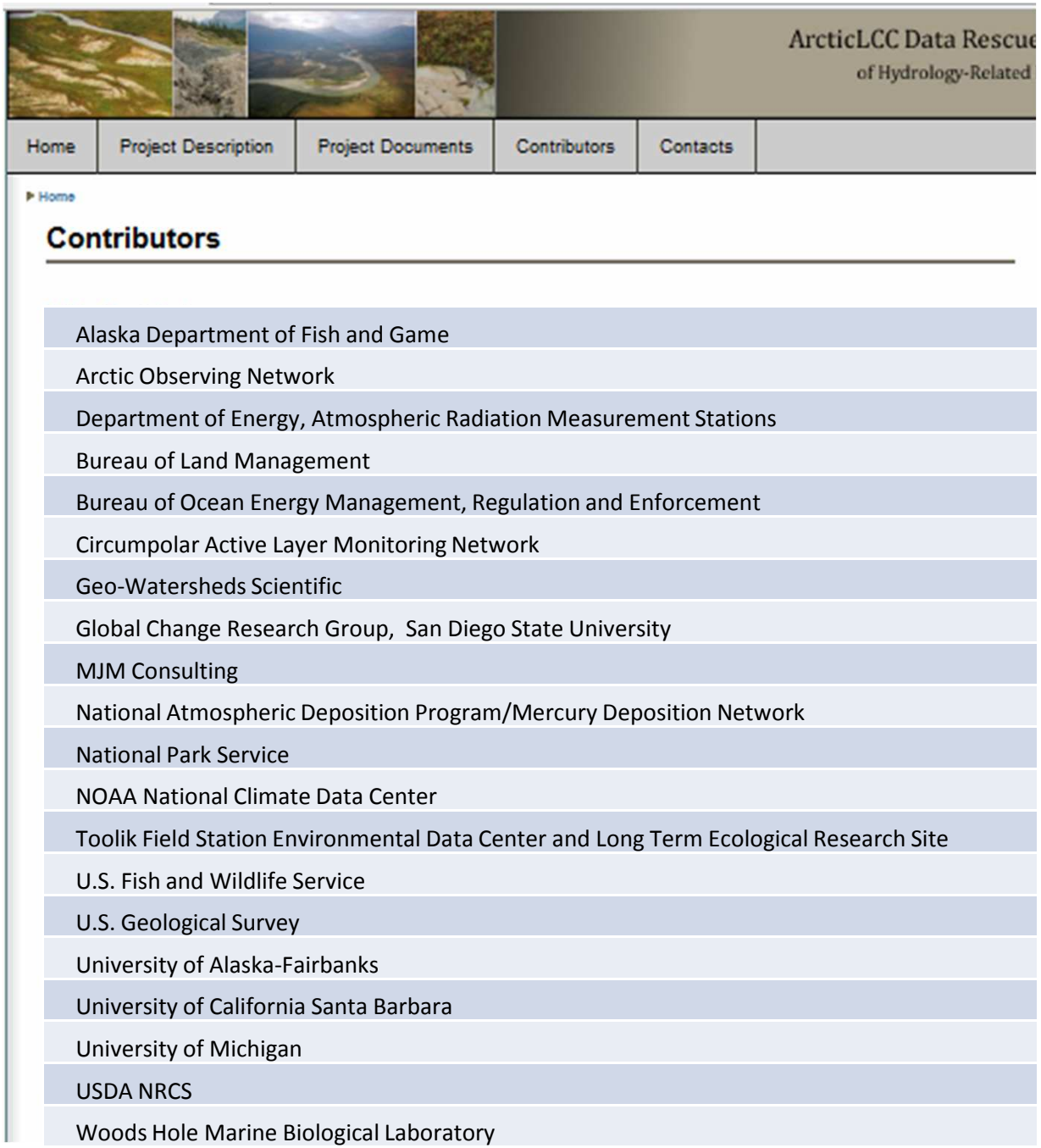




- **Acquire and archive data**
- **Rescue as needed**
- **Interpret, format, normalize, and ingest data**

Data Entry

- Interpreted, normalized and ingested into tables:
  - Isometadata (21)
  - Organizations (20)
  - Source (75)
  - Sites (5000+)
  - Variables (74)
  - Data Streams (634)
  - **Data Values -- 216,000 from 415 Data Streams (150 sites)**
  - **Raw Data Values – millions**



Metadata Inventory

- Identified data sources
- Acquired metadata
- Interpreted, reformatted, standardized, and entered metadata

- **Organizations (20)**
- **Source (70+)**
- **Site (5000)**
- **Variables (74)**
- **Data Streams (634)**