Progress Report: Data Rescue and Inventory of Hydrology-Related Data in Arctic Alaska

A collaborative project for the Arctic Landscape Conservation Cooperative
International Arctic Research Center (IARC) and
Water and Environmental Research Center (WERC) at
University of Alaska Fairbanks

CESU agreement #701817K403

23 September 2010

Greta Burkart
Jessica Cherry (PI)
Amy Jacobs
Jennifer March
International Arctic Research Center
University of Alaska-Fairbanks

William Schnabel
Horacio Toniolo
Peter Prokein
Water and Environmental Research Center
Institute of Northern Engineering
University of Alaska-Fairbanks

Submitted to

Jennifer Jenkins
Philip Martin
Fish and Wildlife Service
Arctic Landscape Conservation Cooperative
Fairbanks, AK

Progress Report:

Data Rescue and Inventory of Hydrology-Related Data in Arctic Alaska

The purpose of this report is to inform the Arctic Landscape Conservation Cooperative (LCC) staff of our progress, facilitate collaboration with Arctic LCC staff, express our needs, and guide future discussions that will ensure the success of this project.

Overall Project Objectives (Figure 1):

- Inventory and acquisition of hydrologic and related data held by entities such as the USFWS, BLM, USGS, NSF, DOE, MMS, UAF, ADNR, ADFG, ADEC, other state and federal agencies, and the private sector. The focus of this inventory will be datasets that can be used to model how hydrologic processes may change and potentially affect fish and wildlife habitat under different climate scenarios.
- Design a public database that will house data and metadata related to hydrology, water quality, climate and aquatic ecosystems in arctic Alaska and begin to populate with metadata (Figure 1).
- Design and host a simple webpage with metadata, project description, and access to a geodatabase with station locations from the inventory (Figure 1).
- Participate in FWS-or agency discussions on North Slope hydrology and hydrologic data.

UAF Progress Report:

1. Initiation Meeting with Arctic LCC staff and project PI and staff.
    a. We participated in a project meeting held May 2010 to discuss project direction and define responsibilities with Arctic LCC staff. Meeting notes including a list of action items, roles, and responsibilities were disseminated to participants.

2. Design a database that will house data and metadata related to hydrology and climate in Arctic Alaska.
    a. Before developing this database, we assessed database needs (Table 1), reviewed other database efforts, and, when relevant, contacted individuals to inquire about database schema and/or potential collaboration.
    b. We created a draft Arctic LCC relational geodatabase to store meteorological, hydrological, water quality and aquatic ecology data and metadata. This database includes a Structured Query Language (SQL) framework and was built

using the Observed Data Model (ODM) developed by the Consortium of Universities for the advancement of Hydrologic Science CUAHSI. A key feature of this database is that individual records are stored for each variable measured at each point in space and time. In combination with ODM's controlled vocabulary, this feature allows users to easily query and retrieve 'analysis-ready' datasets from disparate sources. All components of FGDC ISO (International Standard for Organization) 19115-compliant metadata, the best practice standard for geospatial metadata, are associated with each data value. This ensures that every value downloaded has a traceable dataset heritage and will allow the end user to reduce the ambiguity associated with interpreting results from analyses conducted on data from disparate sources. The database is also scalable allowing for individual ODM databases to be readily ingested into community-scale databases and mapped using many platforms. Even though the ODM was created to store hydrologic and associated climate data, the general model structure will accommodate data from a variety of fields. This database is capable of linking to data stored in EPA STORET and USGS NWIS databases and will be compatible with the database currently being developed by the UAF Water and Environmental Research Center.

c. Despite its several advantages this initial draft ODM database does not meet all of our needs (Table 1). The main disadvantage of the database is the lack of an efficient way to store or query non-point observations (transects, grids, polygons, lines) -- we expect non point observations to be important for modeling effects of climate change on fish and wildlife habitat. In addition, the draft database has no reliable mechanism for informing users of known natural disturbances, anthropogenic effects, and experimental manipulations that may affect how the user works with particular datasets. There is also no way to restrict access to data that will be available for network analyses and/or to contributing PI's, but not to the general public -- to increase the scope of data available for these analyses, we think it is necessary to provide a restricted option. Furthermore, the draft database did not allow meaningful queries to be conducted on datasets that had been inventoried but had not been acquired and entered – queries related to timing of measurements will be important for prioritizing which datasets should be acquired and entered into the database.
d. Because the initial ODM database that we started using does not meet all of our needs (Table 1), we are currently finalizing the development of a modified version of the ODM database (Table 1 and Figure 2). To date, we have modified the ODM database schema to support a data inventory, control access to data, and inform users of potential effects of disturbances, anthropogenic effects, and

experimental manipulations on data. We have also modified schema to work with non-point observations (transects, rasters, polygons). These modifications take advantage of new geography and geometry data types in SQL Server 2008; allow for more efficient storage and representation of gridded, transect, and polygon data; and will be less ambiguous to users, allowing them to take advantage of spatial information and avoid pseudo replication in analyses. We have recently developed procedures for importing and querying polygon files in SQL Server 2008 and expect to finalize working procedures for other non-point observations soon.

3. Design potential queries for the inventory and database.

   We have summarized potential queries (Figures 3 and 4, Table 1) and have used most of these queries to retrieve fisheries and meteorological data. We will refine potential queries for the public database with input from Arctic LCC staff.

4. Requesting and acquiring information about hydrologic and related data held by entities such as the USFWS, BLM, USGS, NSF, DOE, MMS, UAF, AKDNR, ADFG, ADEC, other state and federal agencies, and the private sector.

   a. We developed working procedures for requesting and acquiring datasets from various agencies and created a standardized data request letter (appendix A).
   b. We have contacted several individuals at Universities and state and federal agencies to acquire datasets (USGS, NOAA, AKDFG, NPS-ARCN, UAF-WERC). Most individuals have indicated they will contribute data this fall after fieldwork has ended.
   c. We obtained several climate and hydrology datasets from a variety of agencies and organizations; however, most of these are incomplete.

5. Inventory of hydrologic and related data.

   a. To date, we have inventoried a number of hydrologic and water quality datasets and more than 350 weather stations in and around the Arctic LCC. More than 100 of these weather stations are located on the North Slope and are/were operated by Universities, USGS, NRCS, and environmental consultants. More than 200 of these stations are located within and near the US and Canadian portions of Arctic LCC boundary and are/were operated by the National Weather Service, Canadian weather service agencies, and the Federal Aviation Administration. We have created ISO 19115 compliant metadata with a complete list of variables measured during different time periods for many of

these stations. We are still validating the inventory and acquiring datasets for these stations.

b. We have developed a template and working procedures for normalizing inventory information to comply with CUAHSI standards and to create ISO 19115-compliant metadata. These procedures are still being streamlined as we modify the database and deal with different data types. To date, normalization of inventory information has involved redefining variables using a controlled vocabulary and transforming dates, locations, and source information to adhere to guidelines defined by CUAHSI and FGDC. Thus, the inventory is not simply a list of datasets that could potentially be queried by topic category, organization, or general spatial and temporal coverage; datasets could potentially be queried by topic category, organization, variable, location, and start and end dates for specific variables. The extra effort associated with dataset normalization and standardization will be valuable as we prioritize which datasets are to be included in the final database. Furthermore, given the information available in the inventory, end users could, for example, potentially query locations where soil temperature has been measured at 50 centimeters or less for more than 5 years or locations where wind speed and direction were measured more than 9 meters above the ground for more than 10 consecutive years.

6. Begin populating the database with time-series datasets.
    a. We have tested uploading of various types of data using a variety of methods (ODM data loader, ODM streaming data loader, and SSIS) and have added over a million data observations for one climate station. The number of observations in this database increases hourly as new data is communicated from the field. As new data is communicated to the SQL database the appropriate geophysical units are assigned to raw data, any changes in sensor and logger configuration are noted, metadata is assigned, quality assurance/quality Control (QA/QC) checks are performed, data is flagged as necessary, and data quality levels are assigned. We have also successfully tested real-time use of the database to display weather station data online.

7. Tentative timeline for project completion (Phase 1):
    a. UAF will receive feedback on progress report. UAF and Arctic LCC will discuss progress, future directions, and prioritize what information will be added to the inventory during Phase 1 (October 2010)
    b. UAF will complete design of an initial draft database (October 2010)
    c. UAF will create project website (October 2010)

d. UAF will provide the ArcticLCC hydrology committee with a database and website demo. UAF will receive feedback on database demo and will discuss database needs with Arctic LCC staff and committee members. At this meeting we will also prioritize data rescue efforts and, if necessary, reprioritize what information will be added to the inventory during Phase 1, and which datasets will be added during Phase 2 (November 2010)
e. Revise database design based on Arctic LCC and hydrology committee input (December 2010)
f. Finalize dataset inventories and project website (February 2011)

Requested Input from Arctic LCC Staff:

1. Provide feedback on Progress Report.

2. Convey guidance from wildlife management community and land managers relative to project direction, database content, spatial extent of database relative to LCC boundaries, potential queries, etc.

3. Provide relevant data from Fish and Wildlife Service and Conoco Philips / MJM Research

    a. Relevant shape files, such as Arctic LCC boundary.
    b. Relevant hydrology data
    c. Relevant fisheries data

4. Review this summary report and provide feedback.

5. Coordinate between Arctic LCC cooperators working on overlapping projects.

Figure 1. Proposed project workflow and products for Phase 1 and 2.

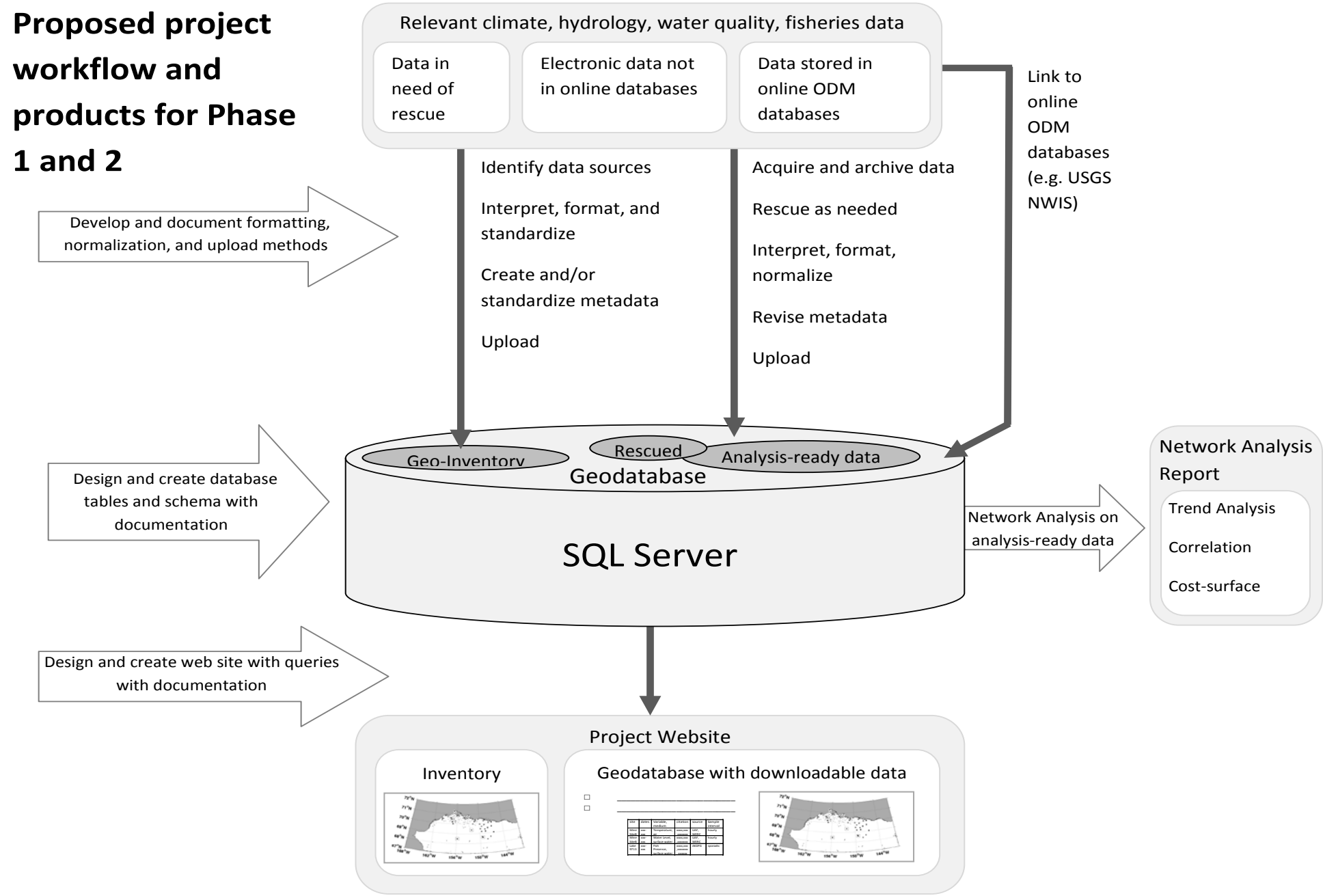

Table 1. Arctic Landscape Conservation Cooperative database needs compared to our original database (ODM) and the database we are currently developing (Modified ODM). Features of the ODM V2 database that are to be developed as part of Phase 1 (2010) are denoted by 'TBD' and those that are to be developed during phase 2

Geodatabase Need	ODM	Modified ODM
Supports point observations	X	X
Supports polygon, gridded and transect observations		X
Tables support processing and prioritizing of datasets		X
Supports ISO 19115 metadata standards	X	X
Stores complete and searchable ISO 19115 metadata without complete dataset entry		X
Traceable dataset heritage with appropriate citation and QA/QC information associated with each data value.	X	X
Links to US EPA STORET and USGS NWIS datasets	X	X
Supports biological datasets	X (limited)	X
Multiple datasets can be downloaded at once.	X	X
Has graphing tools that allow user to plot more than one variable and/or site at a time		X (TBD Phase 2)
Supports decision support tools and network analyses		X (TBD Phase 2)
Works well with different OS	X	X
Requires little to no software downloading by user	X	X

Figure 2. Diagram of modified ODM SQL Server database that is capable of handling non-point observations, controlling access, supporting relevant data inventory queries, and notifying users of potential disturbance related effects on datasets.

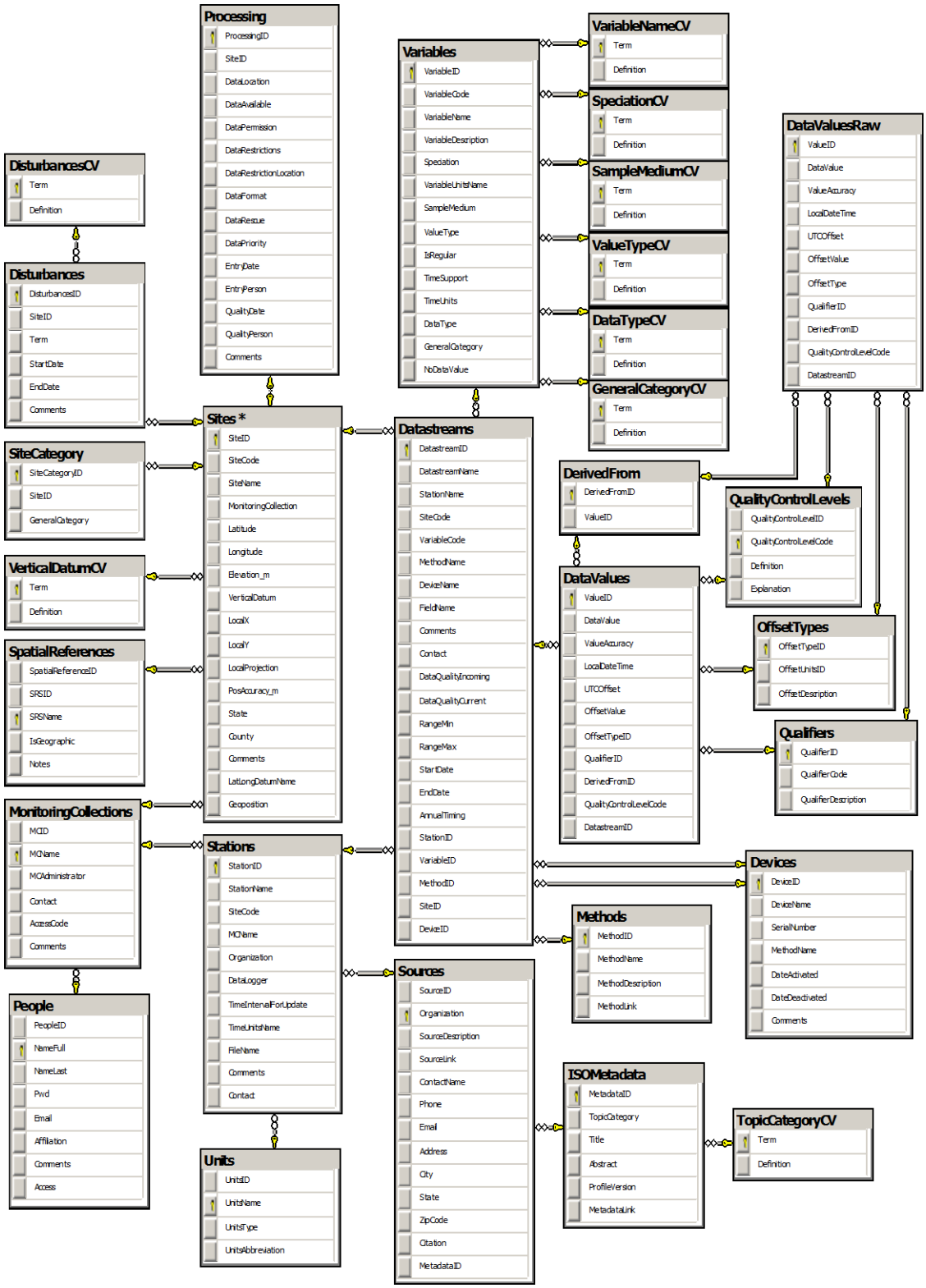

Figure 3. Potential queries and output for of the Arctic LCC inventory during Phase 1.

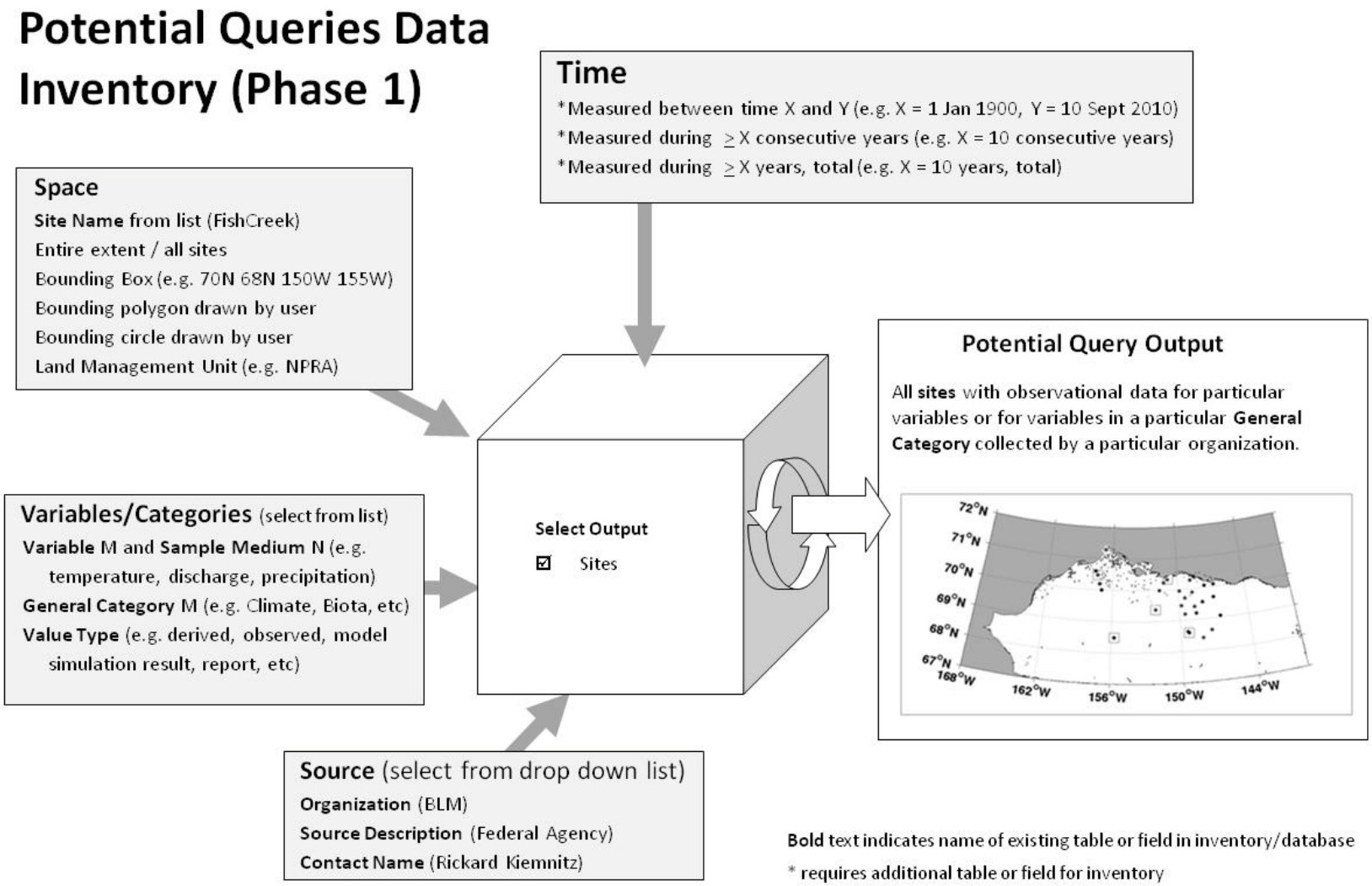

Figure 4. Potential queries and output of the Arctic LCC database during Phase 2.

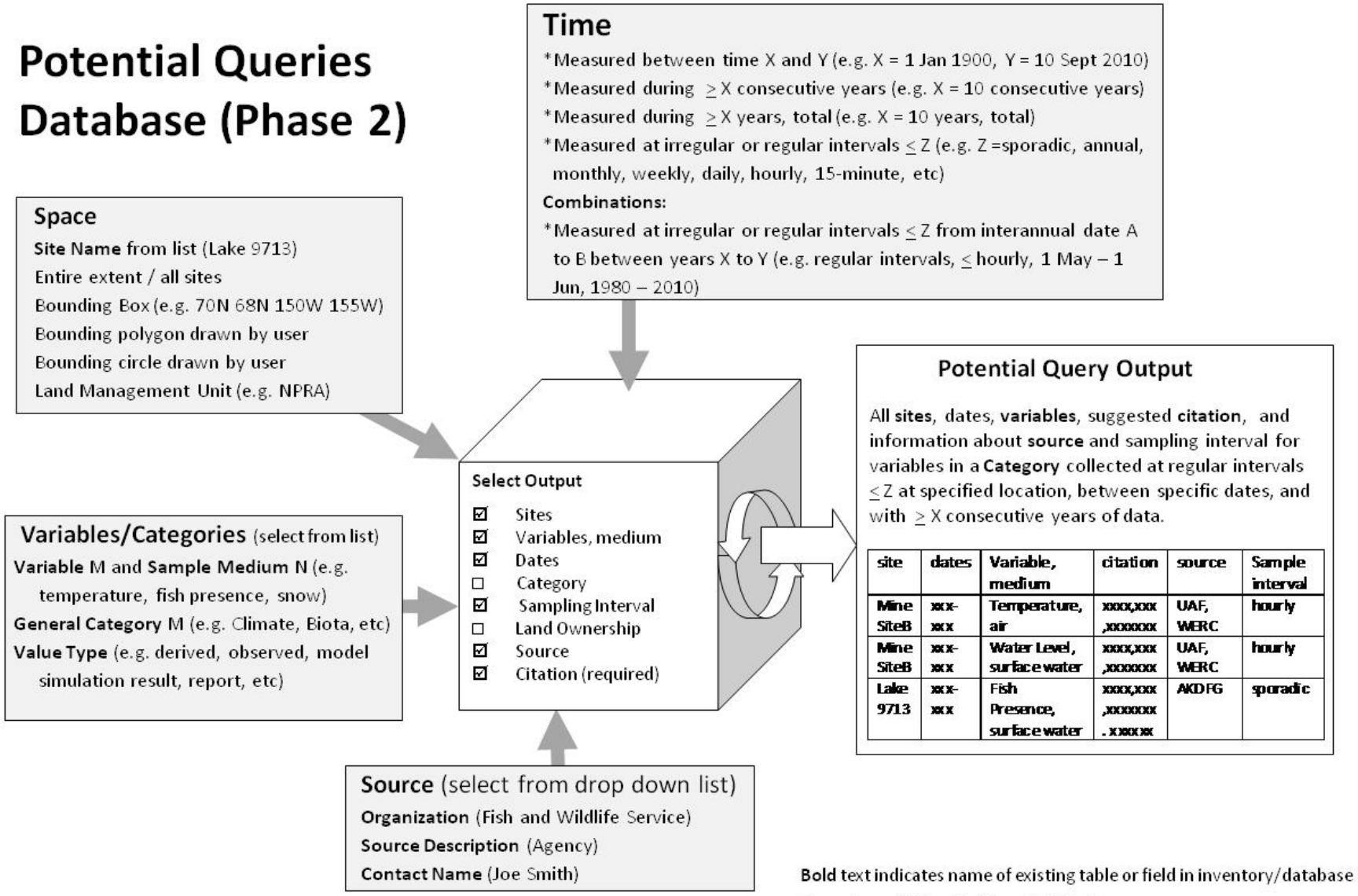

Table 2. Potential Queries for Web-Based Inventory

General Query	Example of relevant question	Priority
Sites identified by **general category** of variables observed	What **locations** were **variables** related to **climate observed**?	1
Sites identified by **general category** of variables observed and **area**	What **locations** were variables related to **climate** observed within a **specific area** (e.g. bounding box: 70N 68N 150W 155W)?	1
Sites identified by **variable**	**Locations** of **ice thickness** observations?	1
Sites identified by **variable** and **area**	**Locations** of **soil moisture** observations for **N Slope Borough**?	1
Sites identified by **organization**	**Locations** sampled by INE-WERC?	1
S**ites** identified by **organization** and **area**	**Locations** sampled by FWS in **Arctic National Wildlife Refuge**?	1
Sites identified by **variable**, **organization**, and **area**	**Locations** of **wind speed** observations by **NPS** in National Parks?	1
Sites identified by **general category**, **organization**, and **area**	**Locations** of **water quality** observations by **BLM** in the **NPRA**?	1
All **variables (or general categories)** identified by **area**	What **variables (or general categories)** were sampled on **L9817** or within a **specific area** (e.g. bounding box: 70N 68N 150W 155W)?	2
All **sites** identified by **multiple variables** (up to X variables)	All **sites** where net radiation, wind speed, snow depth, and/or air temperature were measured.	2
All **sites** identified by **variable(s), organization**, and/or **general category** observed between certain **years** and **interannual dates**	_____ between **10 May 1979 and 1 Jan 2010**. _____ between **1 Apr and 1 Jun** between **2001 and 2010**.	3
All **sites** identified by **variables** that have been measured for more than X years consecutively or X years total	Which locations have wind speed and direction been measured at for at least ten years in a row? Or ten years, total?	3
All **sites** identified by variables measured at interval less than X	Where has air temperature been measured at intervals less than or great to hourly?	3