Optimizing Arctic Observing Through Interoperable Information Sharing Across Networks

William F. Manley[1], Allison G. Gaylord[2], and Craig E. Tweedie[3]
[1] Institute of Arctic and Alpine Research, University of Colorado, USA, william.manley@colorado.edu
[2] Nuna Technologies, Homer AK, USA
[3] Department of Biological Sciences, University of Texas at El Paso, USA

Abstract: A fundamental challenge exists for assessment, planning, integration, and synthesis: the diverse and distributed nature of observing networks and observing systems. For example, which networks and monitoring sites are capturing measurements necessary for any given analysis? Are there disciplinary or geographic gaps? Overlaps? What are the current capabilities? At this time, it is impossible to strategically assess across relevant networks. True, recent efforts have made headway (e.g., by SAON CON, EU-PolarNet, INTAROS, AOV, AOOS, and others). However, most inventories and portals are limited in scope, and almost none share information in a way that can be harmonized and aggregated for a necessarily comprehensive perspective. A solution is that networks share vital observing-related details that go beyond the dataset level. To this end, we recommend formation of an "Observing Network Interoperability Working Group", established with broad representation and the goals of: 1) refining and promoting community-based standards for “network-level”, “project-level”, and “site-level” metadata; 2) adoption of controlled vocabularies for SBA’s, EV’s, and other metadata elements; and 3) establishment of compatible web service endpoints to make such metadata Findable, Accessible, Interoperable, and Reusable (FAIR). Indeed, instead of reinventing the wheel each time, many networks would appreciate taking advantage of established approaches to more efficiently populate databases, aggregate metadata for identified target audiences, or showcase organizational contributions. In these ways, coordination and collaboration can reduce effort while helping to integrate the summed contributions of multiple networks.

A diverse and distributed range of existing Arctic observing networks could be better utilized to address societally relevant scientific problems. For example, observations and associated datasets could be integrated through synthesis in new ways, focusing on a single societal benefit area of interest using one or more indicators or essential variables (EV’s).

Across many Arctic-related data repositories, the requisite scientific datasets for such an approach are becoming more Findable, Accessible, Interoperable, and Reusable (“FAIR”). The Arctic data community has come a long way in the last decade, coming together to better achieve interoperability through coordinated efforts on such things as crosswalks, federated search, and controlled vocabularies. Many, if not most, data repositories now publicly share vital metadata records — the documentation that makes individual scientific datasets findable and accessible — using interoperable metadata standards and web service protocols, sharing information across a fragmented landscape of entities in such a way that data discovery and integration are possible.

However, how would one assess whether networks were making the desired observations? For example, which networks and sites are capturing parameters related to food security through the ECV “Ocean - Physical - Sea Ice”. Are there gaps? Overlaps? What are the current

capabilities? Would it be helpful to know who is actually measuring "what", "where", and "when" to better plan and achieve observing goals?

At the current time, it is impossible to strategically assess Arctic observing networks – at least comprehensively. True, some efforts are making headway. For example, the Arctic Observing Viewer (AOV) now contains over 32,000 observing sites across 35 networks. INTAROS has recently compiled important details about numerous European **observing systems** across multiple domains. There are a few inventories of **networks and programs** (EU-PolarNet, SAON CON). There are many portals for **research projects** (e.g., Arctic LCC, ARMAP, ASDI, BOEM, ELOKA, Isaaffik, NASA ABoVE, NPDC, NPRB, NSSI, PDC, RiS, USGS ScienceBase). And there are numerous other portals with details about individual **observing sites** (ie. monitoring assets; e.g., AOOS Research Assets, ArcticConnect, BAID, GTN-P, JCOMMOPS, NCEI, NEON, NIPR, SIOS, WMO WIGOS). These resources exist to distribute information that goes beyond the level of scientific datasets.

However, almost none of these resources are "aggregators", compiling observing-related information across multiple networks for a comprehensive perspective, or for the specific purpose of addressing the synthesis of an essential variable requiring the lateral fusion of data across many observing systems or networks. A fundamental challenge is that most portals pertain only to individual networks.

And only a few networks share their observing-related details in a manner that permits cross-network integration (harvested and aggregated; ie. sharing "site-level metadata" through interoperable metadata standards and web service protocols; Table 1). Many share details ("fields", or "metadata elements") through custom metadata structures – reinventing the wheel each time, resulting in apples and oranges that are difficult to harmonize. Most do not share the information through machine-readable access. Many simply present the information as a table on a webpage or PDF. Finally, there are many networks that do not share any information.

Table 1: Example High-Order, Discovery-Level Details (Metadata Elements) to Share "Beyond the Dataset Level"

Network-Level	**Project-Level**	**Site-Level**
e.g. GTN-P	*e.g. TSP*	*e.g. borehole*
Network Name	Funding Country	Network Name
Network Description	Funding Agency	Project ID
Region	Funding Program	Observation Type
Time Range	Logistics Provider	GCOS ECV's
Observing Platform	Discipline	Other EV's
Contact Name	Region	GCMD Science and Services
Contact Info	Subregion	Keyword
Funding Agency	Location	GCMD Platform Keyword
Project ID	Project Title	GCMD Instrument Keyword
Links to network pages, data, etc.	Project Abstract	Site Name
	Project ID	Place
Societal Benefit Areas	Institution	Country

GCOS ECV's Other EV's	Contact Name Contact Info Project Start Date Project End Date Links to logistics reports, project summaries, data, and more	Latitude & Longitude Locational Accuracy Elevation Alternate ID1 Alternate ID2 Site Start Date Site End Date Links to logistics reports, project summaries, data, and more

Note: There is no existing metadata standard at the network level; the collection of elements shown above is just a quick start and could be refined and expanded. There are a few existing metadata standards at the project level, which could be refined or simply recommended. There are very few site-level metadata standards (including WIGOS and the implementation documented by AOV), which need to be considered and possibly harmonized or improved. At all levels, crosswalks might need to be developed across recommended standards.

AOV has partially circumvented these obstacles so far – for many of the 34 networks in AOV – by manually "data wrangling" or "screen scraping" the site-related details from network websites, documents, archived datasets, and interpersonal dialogue. This is a slow and laborious process prone to human error that will be difficult to sustain or keep up to date. Some catalogs "beyond the dataset level" fortunately provide API's, which nonetheless then necessitate custom scripting, harmonization, and backfilling. There are many more networks to add.

A solution is that networks compatibly share basic information about their networks and sites. What is needed are the same types of steps toward interoperability that have been taken by the broader Arctic data community for "dataset-level metadata", but in this case by the Arctic observing community for "network-level metadata", "project-level metadata", and "site-level metadata". Ie. that networks come together to agree on adoption of community-based standards and best practices toward a more comprehensive perspective of observing and monitoring activities (cf. Manley et al., 2015). This could be developed iteratively, starting with the most basic pathways for information sharing, with minimal "high-order" fields (Table 1), leading toward improvements by establishing compatible web service endpoints (for dynamic, live feeds with updated information), enabling federated search, and embracing semantics & controlled vocabularies.

Like in the dataset world, there are incentives for networks to do so:

- common scientific goals
- inclusivity in synthesis efforts
- to have their network be more fully utilized (and justified)
- to advertise and showcase their network
- and most important for many stakeholder networks, to take advantage of an established approach, saving them much time and effort (e.g., see this recent IARPC post).

In the broader dataset world, progress toward public release of scientific datasets and associated metadata has been incentivized or mandated via the funding of data repositories, the requirement

of data management plans, through annual or final reporting, and in announcements of opportunity. No such incentives or guidance has been applied at the network level for network-related information sharing, which limits larger scale efforts to assess, design, and implement improved capacities. In a related sense, many networks or observing systems might not be aware of recent advances toward interoperability.

AOV for its part has been trying to promote interoperability. We hired Ted Habermann, an ISO metadata specialist, to bring our implementations of project-level and site-level metadata appropriately into ISO 19115-1, part of a broadly adopted suite of standards. This approach modernizes observing-related metadata, in a way that is both hierarchical and distributed to solve the very challenge of cross-network integration (Habermann, 2018). We have shared documentation about our implementations online, along with a template spreadsheet, data dictionaries, use case examples, and annotated ISO XML's so that other entities could more easily adopt or adapt the approach for their own purposes. We share our metadata through interoperable web service protocols. And we have collaborated individually with numerous partners. However, AOV does not have the authority to establish network-level, project-level, or site-level metadata standards. And there needs to be some top-down guidance and community-engaged coordination in addition to bottom-up resourcing.

As an aside, as was discovered in the dataset world, the solution is not just one master portal or "one stop shop", but many portals that have enabled federated search or have aggregated metadata records for inclusivity and comprehensiveness, each often with their own funding, mandate, thematic or geographic scope, governance, or user audience.

In this way, by adopting "FAIR" principles beyond the dataset level, it would be much easier to assess status, identify gaps, and clarify directions across the diverse, distributed, and fragmented range of existing Arctic observing networks. Such an approach also allows for information and data to be maintained by those with the greatest domain knowledge, in a manner that facilitates the cross-network access needed for advancing scientific synthesis and coordination.

Furthermore, the adoption of community-based standards for observing networks, research projects, and monitoring sites would also enable users to more easily find and access related scientific datasets, which are appropriately hosted and maintained at established data repositories.

Recommendation:

We recommend that a "Network Interoperability Working Group" or task team be established (perhaps under the auspices of one of several active coordinating bodies, and preferably at an international level) with representatives from active networks or observing systems and related portals, with the goal of adopting community-based standards and best practices toward a more comprehensive perspective of Arctic observing activities. This group would potentially refine and then recommend & adopt standards for network-level, project-level, and site-level metadata. Relevant controlled vocabularies – and compatible web service protocols – could be identified and promoted, encouraging interoperable metadata sharing. Such metadata would include high-order, discovery-level elements for ease of implementation. Progress could be incremental,

starting with simple, static sharing of spreadsheets or compatible .csv files, perhaps in web accessible folders, followed preferably by establishment of service endpoints and API's. The work might also include developing a survey for other networks' participation. The team behind AOV could actively contribute to this group. Representatives would need to be technical specialists familiar with network assets, metadata management, and database services. Cross representation would be helpful with other coordinating initiatives.

Such an activity, if established, would help to reduce effort while showcasing and integrating the significant contributions of multiple networks.

Recognized Need:

A properly resourced, comprehensive effort is needed to identify strengths and gaps in the current set of systems, sensors, networks, and surveys used to observe the Arctic.

-- 2nd Arctic Science Ministerial, from the theme on Strengthening, Integrating And Sustaining Arctic Observations, Facilitating Access To Arctic Data, And Sharing Arctic Research Infrastructure

A knowledge map connecting these observations to societal benefits can then guide new observations, data management needs, and development of products and services, leading to a much-needed roadmap for Arctic system observing.

-- Arctic Observing Summit 2018

Performance Element: Advance system models of U.S. observing inventories and data centers to further understanding of these capacities so that informed, optimal, strategic decisions and design, and spending plans can be made.

-- IARPC'S Arctic Research Plan: FY2017-2021, from the research goal to Enhance Frameworks for Environmental Intelligence Gathering, Interpretation, and Application toward Decision Support

Objective 1.1: Conduct an inventory of national observational capacities ... Develop an on-going inventory of national capacities in terms of both long- and short-term observations and monitoring, science/implementation plans, and investment strategies.

-- SAON Strategy, July 2018

The value of creating a shared assessment system, building upon the Arctic Societal Benefit Areas, that availed itself of these existing inventories was clearly recognized. ... The need for a spatially "aware" Roadmap was also viewed as essential for interdisciplinary work. In addition, it was recognized that a Roadmap for observing would need to include a strong knowledge-based element to support data and information discovery across disciplines ... .

-- AOS Observing System Implementation and Optimization Working Meeting, December 2018

References:

Habermann, T., 2018, Metadata Life Cycles, Use Cases and Hierarchies: Geosciences, v. 8, 15 pp. http://dx.doi.org/10.3390/geosciences8050179

Manley, W.F., Gaylord, A.G., Kassin, A., Cody, R., Copenhaver, W.A., Dover, M., Escarzaga, S.M., Font, R., Garcia, A.E., Lin, D.H., Score, R., Villarreal, S., and Tweedie, C.E., 2015, The U.S. Arctic Observing Viewer: A Web-Mapping Application for Enhancing Environmental Observation of the Changing Arctic: Arctic, v. 69, , no. 5, 11 pp. http://dx.doi.org/10.14430/arctic4477