SIDEBAR > ARCTIC DATA MANAGEMENT AND SHARING

By Peter L. Pulsifer and Craig M. Lee

Established and emerging observing technologies provide the potential for expanding our view and understanding of the many dimensions of the Arctic, including its physical, biological, and social domains. New sensors, platforms, survey tools, and a community-driven monitoring program are generating what is referred to as "big data," a term used to describe not only the size of data resources but also the increasing speed of data collection and delivery, the many kinds of data, and the challenges of establishing the accuracy of these data streams. Without an appropriate system for managing data, observations are ephemeral, and their value is limited.

Data systems are advancing alongside and, in some cases, integrated with observing technologies, with the goal of establishing infrastructure that can support seamless data discovery, access, and usage across data providers and users. Building on decades of development, the current objective is to achieve findable, accessible, interoperable, reusable (FAIR) data (Wilkinson et al., 2016). Moreover, the Arctic is home to Indigenous people who have enduring, unique knowledge and observations of their homeland that are increasingly being documented and shared as part of an evolving integrated observing system. Protocols have been established to ensure that Indigenous people and their organizations are recognized as full partners who are actively engaged in the observing process. The FAIR principles exist alongside the CARE Principles of Indigenous Data Governance—Collective benefit, Authority to control, Responsibility, and Ethics (Russo Carroll et al., 2020)—and other regional and national protocols. These guiding protocols exist as part of an Arctic data "ecosystem" of interrelated and interdependent technologies, information objects, human actors, institutions, norms and practices (including standards), relationships, and the broader sociotechnical environment in which it exists (Parsons et al., 2011; Pulsifer et al., 2014, 2020).

The Arctic data management community is deploying and enhancing technologies and methods to ensure that the Arctic data ecosystem can serve all communities and achieve FAIR/CARE data. Underpinned by the collaboration fostered by the International Polar Year (2007–2009) and the resulting formation of bodies such as the Arctic Data Committee (https://arcticdc.org), a growing consortium of polar data stewards and coordinating bodies are collaborating through workshops, conferences, and working groups to make progress (e.g., Polar Data Forum, https://polar-data-forum.org/, and Polar to Global Hackathon, https://arcticdc.org/meetings/conference-calls-webinars/polar-to-global-online-interoperability-and-data-sharing-workshop-hackathon).

For example, the POLar Data discovery Enhancement Research (POLDER) working group has established a Pilot Federated Search tool (https://search-dev.polder.info/) that uses a shared metadata profile to connect the many different polar data catalogues hosted by data centers and other institutions. This tool dramatically improves the community's ability to find data and provides a gateway to access data. Conventional data download sites are still a common method for making data and associated metadata accessible; however, these sites are quickly being supplemented by the deployment of web services or web-accessible application programming interfaces (APIs). These dynamic, "live" services can support near-real-time access to data that does not require users to download data sets to their local environment. Data are streamed, and many different services can be used in combination to support complex modeling and research, while greatly reducing the time and resources required to manage and process data in a user's local environment. Web services are also contributing to enhanced interoperability—the ability of systems to readily share information and operations. Using standards and specifications such as those developed by the Open Geospatial Consortium (https://www.ogc.org/) and the International Organization for Standardization (https://www.isotc211.org/), data repositories can be connected to incoming data streams generated by new observing technologies, and to different end users including those who are mediating the data (e.g., modelers) or others who may want to simply aggregate data to create broader geographic coverage. Projects such as the Arctic Spatial Data Infrastructure, the Canadian Consortium for Arctic Data Management, and the Global Cryosphere Watch are deploying web services to make data FAIR (https://arctic-sdi.org/, https://ccadi.ca/, https://globalcryospherewatch.org/).

Web services and associated mediation methods are improving data interoperability; however, a major challenge remains—semantic interoperability. Simply transferring data does not guarantee that the exchanged data can be understood and used by the recipient. Data sets include various classes and attributes that have meaning to producers and users, for example, different ocean or atmospheric parameter names, feature classes on a classified satellite image or map, qualitative themes identified and named in a social science research study, and Indigenous knowledge concepts and place names. Interoperability and reuse can only be achieved if the meanings imbued in data elements are explicitly shared along with the data sets. To address this issue, the Vocabularies and Semantics Working Group is collaborating to develop and encourage the tools and methods needed
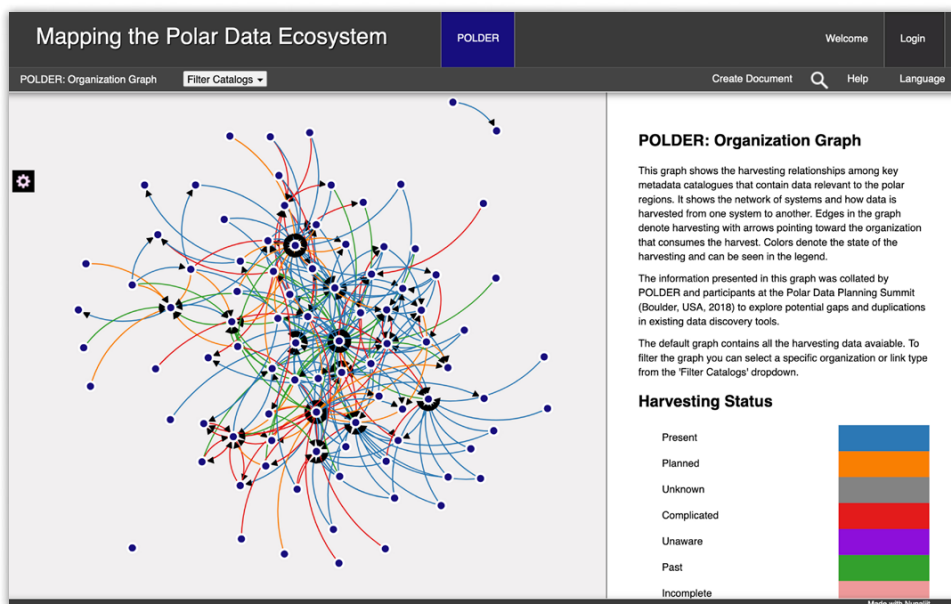
FIGURE 1. The Mapping the Polar Data Ecosystem project (https://develop.gcrc.carleton.ca/mdpe/) aims to use the established conceptual framework of information ecology as an analytical tool to help organize ideas and comprehend the complexity of the Arctic and polar data ecosystem. The associated website provides interactive visualizations of different elements of the Arctic and Antarctic data ecosystems.

to share data semantics (https://arcticdc.org/activities/core-projects/vocabularies-and-semantics-wg).

In the era of big data, accessing and using very large data sets can be challenging for users with limited storage and computational resources. New observing technologies can produce terabytes of data in a single day, and downloading and managing these data can be time-consuming and costly. New platforms that bring the user to the data rather than data to the user are now available to the Arctic community. For example, the Polar Thematic Exploitation Platform (https://portal.polartep.io/) provides a complete working environment where users can access algorithms and data remotely.

Emerging observing and data management technologies have the potential to revolutionize our ability to understand the Arctic and make informed decisions to meet current grand challenges. A key to realizing this potential is understanding and managing the system's complexity to improve collaboration and system integration. The Mapping the Polar Data Ecosystem project, currently a joint effort of the Arctic Data Committee, Arctic PASSION, and POLDER, is working to meet this meta challenge that is fundamental to realizing the FAIR and CARE principles (**Figure 1**).

REFERENCES
Parsons, M.A., Ø. Godøy, E. LeDrew, T.F. De Bruin, B. Danis, S. Tomlinson, and D. Carlson. 2011. A conceptual framework for managing very diverse data for complex, interdisciplinary science. *Journal of Information Science* 37(6):555–569, https://doi.org/10.1177/0165551511412705.

Pulsifer, P.L., L. Yarmey, Ø. Godøy, J. Friddell, M. Parsons, W.F. Vincent, T. de Bruin, W. Manley, A. Gaylord, A. Hayes, and others. 2014. Towards an international polar data coordination network. *Data Science Journal* 13:PDA94–PDA102, https://doi.org/10.2481/dsj.IFPDA-16.

Pulsifer, P.L., Y. Kontar, P.A. Berkman, and D.R. Taylor. 2020. Information ecology to map the Arctic information ecosystem. Pp. 269–291 in *Governing Arctic Seas: Regional Lessons from the Bering Strait and Barents Sea.* Springer, https://doi.org/10.1007/978-3-030-25674-6_12.

Russo Carroll, S., I. Garba, O.L. Figueroa-Rodríguez, J. Holbrook, R. Lovett, S. Materechera, M. Parsons, K. Raseroka, D. Rodriguez-Lonebear, R. Rowe, and others. 2020. The CARE Principles for Indigenous data governance. *Data Science Journal* 19(1):43, http://doi.org/10.5334/dsj-2020-043.

Wilkinson, M.D., M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L.B. da Silva Santos, P.E. Bourne, and others. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3:160018, https://doi.org/10.1038/sdata.2016.18.

AUTHORS
Peter L. Pulsifer (peter.pulsifer@carleton.ca) is Associate Professor, Department of Geography and Environmental Studies, Carleton University, Ottawa, Ontario, Canada. **Craig M. Lee** (craiglee@uw.edu) is Senior Principal Oceanographer and Professor, Applied Physics Laboratory, University of Washington, Seattle, WA, USA.