Polar Data and Platform Interoperability

Resource Requirements

Prepared by:

Standing Committee on
Antarctic Data Management

On behalf of the

Polar Data Community

January 31, 2017

Summary

The polar regions are of increasing interest to the whole world as a result of their linkage to global climate systems, opportunities for economic development, geo-political strategic importance, and their environmental importance as sensitive ecosystems and homes to Indigenous populations and other residents and sensitive ecosystems.

Polar data are required by the scientific community and residents to support research on topics such as climate, atmosphere, land, oceans, ecosystems, ice and snow, permafrost, and social systems; and by the operations community to support impact assessments, engineering design, safe navigation and operations, risk management, emergency response, weather forecasting, and climate change adaptation. These activities contribute to environmental protection, heritage preservation, economic development, safety of life and property, and national sovereignty.

The polar data community is well organized and is pursuing activities to improve data acquisition, access, and management for all of the diverse members of the polar community. In the Arctic, the Arctic Data Committee (ADC) of Sustaining Arctic Observing Networks (SAON) is taking a coordinating role. In the Antarctic, this role is being performed by the Standing Committee on Antarctic Data Management (SCADM) of the Scientific Committee on Antarctic Research (SCAR) and the Southern Ocean Observing System (SOOS)

Increasingly, the infrastructure associated with polar data is evolving from systems where data are discovered in data catalogues and downloaded to the local machines of users, to distributed platforms made interoperable using standards and providing users with storage and computational capacity close to large repositories of data.

There is still much to be done to move towards a new model for polar data management, and by working together, we believe the polar community can achieve significant improvements in polar data interoperability. However, making significant progress will require adequate financial, technical, and human resources. The first step in acquiring the necessary resources is to define what is required. The following table provides an initial estimate of the level or funding that will be needed over the next five years to tackle the most significant interoperability challenges facing the polar data community.

Theme	Five-Year Funding Requirements
Discovery and Documentation ▪ Development of common metadata elements for use in a “single window” search that produces results based on multiple catalogues	€ 300,000

▪ Enhancement of brokering technologies that can convert from multiple standards to the aforementioned common elements ▪ Promoting data publication and attribution so that researchers can make their data available through mainstream publication outlets, and receive credit for the publication ▪ Supporting the community in developing capacity to make users aware of existing data ▪ Consolidate, update and maintain observing platform inventories at a global level using a brokering approach.	
Data as a Service ▪ Supporting the community in developing capacity to engage as publishers and users of data services ▪ Working with communities of practice to develop terminology and knowledge models that enable users to fully understand data being shared (e.g. structure, classification systems etc.)("semantic interoperability") ▪ Develop tools that help to simplify the process of making data available as a service	€ 900,000
Indigenous Knowledge ▪ Support Indigenous communities in developing protocols that allow for ethical sharing of documented knowledge ▪ Support skills and knowledge development driven by Indigenous people and their representative organizations ▪ Promote an integrated approach to development to avoid silos – ensure that Indigenous people are "at the table" as systems are being collectively designed.	€ 600,000
Cloud Data and Computing Platform Interoperability ▪ Making connections between prominent cloud platforms for seamless integration of data and results ▪ Establish models for sharing algorithms and software within and between platforms ▪ Streaming user authentication across platforms and integrating resource accounting	€ 750,000

Community Building, Governance, Sustainability and Linkages Among Global Systems ▪ Enhance existing community building efforts to better link various initiatives. For example, resources to support a planned Third Polar Data Forum in late 2017 and a possible follow up meeting in conjunction with the Polar 2018 meeting in Davos. ▪ Identify an appropriate governance model for the international community – recognizing the "loosely coupled" nature of the community and the existence of the polar community within broader disciplinary and global communities. ▪ Develop a sustainability plan that directly engages funders ▪ Establish model for optimizing engagement with global data research, development and coordination initiatives	€ 450,000
Total	€ 3,000,000

Points of Contact

For further information on the contents of this paper, please contact any of:

- SAON, Christine Daae Olseng, Chair, cdo@forskningsradet.no
- ADC, Peter L. Pulsifer, Chair, peter.pulsifer@colorado.edu
- EU-PolarNet, Nicole Biebow, Project Manager, nicole.biebow@awi.de
- SCADM, Anton Van de Putte, Chair, antonarctica@gmail.com
- SOOS, Phillippa Bricher, Data Officer, data@soos.aq
- Polar View Earth Observation, David Arthurs, Managing Director, david.arthurs@polarview.org

Polar Data Context

There is currently intense interest in the polar regions, with an associated requirement for integrated information to support the research and operations of a growing range of user communities, including science, industry, government, and northern communities. A number of political, environmental, social, and technological trends are fueling this interest and activity, including:

- Political and Policy Trends – The interest in the polar regions of governments around the world is driven by perceived opportunities for economic development, more efficient shipping routes, and the regions' geo-political strategic and sovereignty importance. With these opportunities come concerns over their environmental impact and risks to life and property in an isolated and hostile environment which governments have a duty to mitigate through emergency response and search and rescue operations. The opportunities are also motivating countries to try to expand their jurisdictions and to better protect their borders.
- Economic Trends – Economic development opportunities include development of renewable resources such as fisheries and forests; non-renewable resources such as fossil energy resources and minerals; and other activities such as shipping and tourism. Closely associated with these opportunities is the need for related infrastructure development, such as offshore platforms, ice class ships, pipelines, railways, roads, sea ports, airports, and housing. There is also the potential for increased pollution and environmental accidents.
- Social and Cultural Trends – Concern about the impact of climate change is growing around the world and it is becoming evident that the impact is greatest in the polar regions. Of particular social relevance in the Arctic are the changes that are being imposed on Indigenous Peoples by climate change and increased economic activity. Such changes include impacts on hunting and fishing practices, impacts on infrastructure caused by coastal erosion and the melting of permafrost, and impacts on culture and social cohesion.
- Technological Trends – A number of technological trends are providing a flood of new data concerning the polar regions. Of particular relevance are space-based technologies such as earth observation, satellite telecommunications, global navigation satellite systems (GNSS), and the detection of ship-borne automatic identification systems (AIS), and a wide variety of in-situ observational networks. Each has a role to play in monitoring the vast and harsh polar regions and each is undergoing significant improvements in capabilities. However, ensuring the interoperability of these diverse data streams requires the development and implementation of appropriate data standards.

This broad range of scientific, operational, and societal imperatives is driving the requirements for information in the polar regions. The polar data community represented in this submission are endeavoring to fulfill these requirements by supplying both scientific and operational data.

Scientific enquires are taking place across many domains, including the climate, oceans, atmosphere, and ecosystems in the polar regions which, through complex earth system connections, have significant impacts worldwide. Operational data supports shipping and fisheries companies, offshore oil and gas operators, research organizations, coast guards, and local communities, who require access to reliable and often near real-time information to plan and undertake their activities. Drivers of information requirements include a range of regulations, standards, and policies (such as the new Polar Code[1]) aimed at ensuring safety of life and mitigating negative environmental impacts. More information on polar data requirements and sources of information can be found in the Polaris study of the European Space Agency. [2]

Polar Data Community

The polar data community consists of a wide variety of data producers, managers, and users in government, industry, academia and northern communities that need data for scientific research and to support operations and livelihoods in the polar regions. The recommendations made in this paper are supported by a variety of organizations which bring perspectives that include Arctic and Antarctic science, polar operations, and the interests of Europe, the United States, Canada, and Asia:

- **International Arctic Science** – The Arctic Data Committee (ADC) of the International Arctic Science Committee (IASC), Sustaining Arctic Observing Networks (SAON), and the Arctic Portal.
- **International Antarctic Science** – The Standing Committee on Antarctic Data Management (SCADM) of the Scientific Committee on Antarctic Research (SCAR); and the Southern Ocean Observing System (SOOS).
- **International Cryosphere Science** – WMO Global Cryosphere Watch (WMO GCW), Group on Earth Observation Cold Regions Initiative (GEOCRI), and Climate and Cryosphere (CliC)
- **International Polar Operations** – The International Ice Charting Working Group (IICWG) and Polar View Earth Observation (PVEO).

[1] To help address the risks of operating in the polar regions, the International Maritime Organization (IMO) Marine Environment Protection Committee approved the "Draft International Code for Ships Operating in Polar Waters" (known as the Polar Code) on 21 January, 2015. It will take effect on 1 January, 2017.

[2] The Polaris Study reviewed user requirements for polar environmental information, considered current and proposed sources of such information from space-based and in-situ sensors, evaluated the information gaps and the impact of filling those gaps with new integrated products and services, and provided a preliminary discussion of the considerations that will shape new satellite missions to fill the gaps. See: http://esamultimedia.esa.int/docs/spaceforearth/Final_Summary_Report_Polaris.pdf

- **United States** – The National Snow and Ice Data Center (NSIDC), the Interagency Arctic Research Policy Committee (IARPC) Arctic Data Coordination Team, the Alaska Data Integration Working Group (ADIwg), the NSF-funded Antarctic and Arctic Data Consortium (a^2dc), the Arctic Research Mapping Application, the Arctic Observing Viewer, and the Barrow Area Information Database.

- **Canada** – Polar Knowledge Canada, the Canadian Cryospheric Information Network (CCIN), the Geomatics and Cartographic Research Centre at Carleton University, and the Canadian Consortium for Arctic Data Interoperability (CCADI).

- **Europe** – EU-PolarNet and the European Space Agency (ESA).

- **Asia** – Japan's National Institute of Polar Research (NIPR)

All of these organizations represent wider assemblies of polar data stakeholders. Many function as the official international fora for the interests of national bodies with responsibility for polar science and operations. Together, these organizations represent over 50 countries.

The polar data 'ecosystem' is broad and complex. In an effort to manage this complexity, the polar data community has created organizations for the purpose of promoting and facilitating international collaboration towards improving data access and interoperability. In the Arctic, this is the Arctic Data Committee (ADC) of the International Arctic Science Committee (IASC) and the Sustaining Arctic Observing Networks (SAON); in the Antarctic, this is the Standing Committee on Antarctic Data Management (SCADM) of the Scientific Committee on Antarctic Research (SCAR) and the Southern Ocean Observing System (SOOS). These organizations work in collaboration with the Arctic Council and the Group on Earth Observation (GEO). The White House Arctic Science Ministerial that was held 28 September 2016 mentions SAON in its joint statement as a critical contributor to "Strengthening and Integrating Arctic Observations and Data Sharing".[3]

The roles of these organizations are to:

- Advise their communities on matters related to data management and data sharing.
- Contribute to the understanding of the nature and structure of the polar data system in the context of the global data system.
- Promote and enable:

[3] "We, the Ministers representing the eight Arctic States (Canada, the Kingdom of Denmark, Finland, Iceland, Norway, Russia, Sweden, and the United States), fourteen additional States (China, France, Germany, India, Italy, Japan, Republic of Korea, Netherlands, New Zealand, Poland, Singapore, Spain, Switzerland and the United Kingdom), and the European Union, in partnership with Arctic Indigenous representatives ... see a critical role for the Sustaining Arctic Observing Networks (SAON) initiative—a joint responsibility of the Arctic Council and the International Arctic Science Committee—and encourage continued cooperation in other international science organizations that contribute to Arctic observing and data-sharing, and building a network of community-based observation." See: https://www.whitehouse.gov/the-press-office/2016/09/28/joint-statement-ministers

  - Ethically open access to data
  - Norms of fair attribution and use of data
  - Long term preservation of data
- Facilitate the adoption, implementation and development (where necessary) of standards that will enable free, open and timely access to data.
- Facilitate interoperability of data and systems as needed to support the needs of researchers, Arctic residents, decision makers and others.
- Establish expert groups to examine specific questions or coordinate the implementation of data management and sharing solutions. Partnerships with existing or proposed initiatives driven by members of the polar science and data community and Northern communities will be encouraged.

Polar Data Interoperability

Interoperability can be defined as properties of cyberinfrastructure that allow it to work and share with other information products or systems, present or future, without unintended restrictions. Achieving interoperability is a multifaceted problem including technical (syntax and structure), semantic (how we define and label concepts), legal (intellectual property, etc.), and geopolitical (e.g. adherence to treaties) concerns, among others. The polar data community understands the importance of achieving interoperability between systems and has organized a number of initiatives to make progress towards that goal.

For example, in October 2015, more than 100 people gathered at the Second Polar Data Forum (PDF II) to address the challenge. Data managers, scientists, funding program managers, Indigenous people and their representatives, students, and others from eighteen nations shared their knowledge, experience, and ideas on how to make polar data more useful and valuable in solving global problems.[4]

In March 2016, the third Arctic Observing Summit was held to consider strategies for sustained support of long-term Arctic observing. The seventh recommendation from the Summit sets the stage for polar data interoperability efforts:

> "7. Work, through the SAON Arctic Data Committee, to develop a broad, globally connected Arctic observing data and information system of systems that is based on open access data and standards, in addition to recognizing and addressing ethical use and

[4] For a high-level overview of the results of the Forum see: polar-data-forum.org/programme/PDFII_Communique_FINAL.pdf.

proprietary rights of Indigenous Knowledge, and that delivers value to Arctic and global communities."[5]

Following up on the previous events, a workshop was hosted by the European Space Agency in November 2016 to bring together stakeholders and make tangible progress on establishing strategies and mechanisms for data interoperability.[6] The workshop focused on a set of key themes:

- Data discovery and metadata
- Data as a Service, including persistent identifiers, software as a service, brokering, and semantic interoperability
- Representing and sharing Indigenous Knowledge, community based observations, and the social sciences
- Cloud data and computing platforms
- Governance and establishing sustainable systems
- Connecting to the global data and information ecosystem

There are also major regional and national initiatives established and emerging. For example, in Europe the EU-PolarNet project includes a major component focused on data and related infrastructure[7], and INTAROS[8] and INTERACT[9] are making significant contributions to observing systems. In Canada, Polar Knowledge Canada is taking a lead in this area along with the Canadian Consortium on Arctic Data Interoperability, the Polar Data Catalogue, and other initiatives. In the United States, the Interagency Arctic Research Policy Committee (IARPC) Environmental Intelligence Collaboration Team is taking a coordinating lead in partnership with initiatives such as the NSF-funded Antarctic and Arctic Data Consortium (a2dc).

The community is working to link these efforts through strong cooperation across the groups. For example, ADC and SCADM have drafted a Memorandum of Cooperation to enhance collaboration and efficiency between these groups and other global and national initiatives such as the IARPC ADCT, NSF a2dc, Southern Ocean Observing System, Research Data Alliance, World Data System, Group on Earth Observations, and others.

[5] See http://www.arcticobservingsummit.org/aos-2016-conference-statement-0
[6] See: http://arcticdc.org/meetings/adc-meetings/interoperability-workshop
[7] See: http://www.eu-polarnet.eu/project-themes/infrastructures-facilities-and-data.html
[8] See: http://cordis.europa.eu/project/rcn/205992_en.html
[9] See: http://www.eu-interact.org/

Interoperability of Polar Data Platforms

There is already a considerable wealth of polar data available through portals that vary substantially in function, scope, capability, and content. However, the polar data community is aware that there are many opportunities for improvement in how polar data are stored, managed, discovered, and delivered to users across these platforms, and they are working collaboratively, with limited resources, to improve the situation.

The development of polar data platforms is occurring within a context of rapid growth in the provision of polar data and change in user expectations about access to and use of such data. The data available on the state of the planet is growing in precision, volume, velocity, variety, and value, increasing the complexity of scenarios for data exploitation, as well as the resources required by the communities using the data. A number of groups are developing innovative approaches to the creation of polar data platforms. These approaches share some common characteristics:

- Individual parameters by themselves are not nearly as valuable as integrated data sets. Therefore, the trend is to provide data platform users with access to a wide range of data types that they can be exploited together.

- With the explosion of the data that are available, data discovery and analysis is becoming increasingly challenging. As a result, the trend is to include sophisticated data visualization tools to enable data platform users to easily see and understand both the data they can utilize and the results of their analysis of that data.

- The quantity of data available, especially EO data, means that it is often not practical for each user to download the data they need to their local environment. Rather, the trend is to bring the algorithms to the data and only download the results of their calculations.

- Working with such large data sets is often computationally intensive. This means that modern data platforms need to provide users with highly capable ICT infrastructure for data processing, storage, and networking.

- Research is increasingly collaborative. Therefore, the trend is to combine data and computation capabilities with the tools required for such collaboration and the ensuing dissemination of research results.

- The increasing diversity of data sources and the need for scientific and operational communities to access data unfamiliar to them makes it essential that useable data quality information is available for all products.

- There is an aversion to lock-in with any one technology or supplier. Therefore, many data platforms use open source software where possible and are platform independent, often hosted in the cloud.

In summary, modern polar data platforms are going far beyond traditional data portals by combining multiple functionalities and making them available in the cloud. Examples of such platforms that are currently being developed include the Polar Thematic Exploitation Platform (Polar TEP) being sponsored by the European Space Agency, the INTAROS Integrated Arctic Observing System platform (iAOS) being sponsored by the European Commission, and the Arctic-Boreal Vulnerability Experiment (ABoVE) Science Cloud being sponsored by NASA. More work needs to be done to make these platforms truly interoperable.

Priority Activities

While the polar data community has collaborated to make significant advances in data management and availability, there is still much to be done. The current activities of polar data management organizations include:

- **Data Discovery and Description** – A foundational component of an interoperable data system is the ability to discover data and evaluate it in terms of quality, fitness for use etc.. There are many approaches available that can support the development of this component, including the establishment of a common set of metadata elements relevant across polar sciences, adoption of 'brokering' technologies that can help mediate metadata and other descriptions where different standards are in use, and the promotion of new paradigms for data publication and citation for polar researchers (e.g. using Digital Object Identifiers to publish data). One aspect of data discoverability is to have an overview of the platforms (stations, ships, satellites, aircrafts, buoys) that are producing the observational data. Many initiatives maintain inventories of these at the national or regional level. This is the case for initiatives like COMNAP, EU-PolarNet, EuroFleet, FARO, INTERACT, Polar Data Catalogue, SCAR, INTAROS (emerging) and WMO. There is, however, a need for consolidated, continuously updated inventory of observational platforms, including extension to lesser documented areas such as Russia.

- **Data and Systems Interoperability**: Interoperability, the ability to easily share data across systems and users, is one of the most important priorities identified by the polar data community. An interoperable system must enable data access that can support many different users. In 2017, this means moving towards a service-oriented model that makes data and processing services available as a live stream over the Internet, rather than a static download for local processing. This may include visualization or other mediation such as

translating vocabularies to make data usable by different communities. Achieving interoperability will require adequate resources, a certain level of standardization, and a connected community.

- **Standards and Specifications** – The overarching purpose of the polar data management community is to promote and facilitate international collaboration towards the goal of free, ethically open, sustained, and timely access to polar data through useful, usable, and interoperable systems. This includes facilitating the adoption, implementation and development (where necessary) of standards that will enable free, open and timely access to data. As indicated, establishing standards to support data exchange and interpretation is central to achieving interoperability.

- **Including Arctic Indigenous Perspectives, Knowledge and Information**: In this time of change, Indigenous knowledge and the underlying observations of Arctic peoples are more important than ever. Along with the knowledge of non-Indigenous local inhabitants, this knowledge is being increasingly documented and represented as digital data, but the nuances of these data are not well understood by the broader data management and science community. The perspectives of Indigenous people and other northern residents must be heard directly. This will enhance understanding of how Indigenous and local knowledge and observations can be used appropriately.

- **Community Building and Governance**: Improved polar data sharing that is part of a broader global system will require community building, collaboration, and coordination of efforts. To do this, a better understanding is needed across many scales of the nature of the polar data community (who is doing the work, where, what systems, etc.) and what it is collectively trying to achieve. Through the established bodies discussed above, improved communication, outreach, and coordination within the polar community is needed.

- **Data Preservation and Rescue**: Past observations must be continually re-used and re-purposed to increase current understanding. Therefore, data, Indigenous Knowledge (especially of Elders), and all the necessary descriptive information must be preserved. Too often, preservation is forgotten and data managers must pursue “data rescue” activities. Even current data are at risk of loss. Strategic data rescue programs must be developed, and preservation must be prioritized as a long-term investment and cost-saving measure.

Polar Data Resource Requirements

The most import conclusion from all of the studies and workshops to date is that making significant progress will require adequate financial, technical, and human resources. The first step in acquiring the necessary resources is to define what will be required. The following table provides an initial estimate of the level or funding that is needed over the next five years to tackle the most significant challenges facing the polar data community.

Theme	Five-Year Funding Requirements
Discovery and Documentation ▪ Development of common metadata elements for use in a "single window" search that produces results based on multiple catalogues ▪ Enhancement of brokering technologies that can convert from multiple standards to the aforementioned common elements ▪ Promoting data publication and attribution so that researchers can make their data available through mainstream publication outlets, and receive credit for the publication ▪ Supporting the community in developing capacity to make users aware of existing data (e.g training courses and materials) ▪ Consolidate, update and maintain observing platform inventories at a global level using a brokering approach.	€ 300,000
Data as a Service ▪ Supporting the community in developing capacity to engage as publishers and users of data services ▪ Working with communities of practice to develop terminology and knowledge models or "ontologies" that enable users to fully understand data being shared (e.g. structure, classification systems etc.)("semantic interoperability") ▪ Develop software tools that help to simplify the process of making data available as a service	€ 900,000
Indigenous Knowledge ▪ Support Indigenous communities in developing protocols that allow for ethical sharing of documented knowledge	€ 600,000

▪ Support skills and knowledge development driven by Indigenous people and their representative organizations ▪ Promote an integrated approach to development to avoid silos – ensure that Indigenous people are "at the table" as systems are being collectively designed.	
Cloud Data and Computing Platform Interoperability ▪ Making connections between prominent cloud platforms for seamless integration of data and results ▪ Establish models for sharing algorithms and software within and between platforms ▪ Streaming user authentication across platforms and integrating resource accounting	€ 750,000
Community Building, Governance, Sustainability and Linkages Among Global Systems ▪ Enhance existing community building efforts to better link various initiatives. For example, resources to support a planned Third Polar Data Forum in late 2017 and a possible follow up meeting in conjunction with the Polar 2018 meeting in Davos. ▪ Identify an appropriate governance model for the international community – recognizing the "loosely coupled" nature of the community and the existence of the polar community within broader disciplinary and global communities. ▪ Develop a sustainability plan that directly engages funders ▪ Establish model for optimizing engagement with global data research, development and coordination initiatives	€ 450,000
Total	€ 3,000,000