Data Management for Arctic Observing

A Community White Paper
Prepared for the Arctic Observing Summit 2013

Peter L. Pulsifer[1], Lynn Yarmey[1], Øystein Godøy[2], Julie Friddell[3], Warwick F. Vincent[4], Taco DeBruin[5], Mark A. Parsons[1]

1 National Snow and Ice Data Center, University of Colorado, Boulder, CO, USA
2 Norwegian Meteorological Institute, Oslo, Norway
3 Canadian Cryospheric Information Network, University of Waterloo, Waterloo, Canada
4 CEN: Centre d'Etudes Nordiques, Laval University, Quebec City, Canada
5 NIOZ Royal Netherlands Institute for Sea Research, Texel, The Netherlands

Corresponding author:

Peter L. Pulsifer
National Snow and Ice Data Center
Cooperative Institute for Research in Environmental Science (CIRES)
University of Colorado
449 UCB
University of Colorado
Boulder CO 80309
USA
e-mail: pulsifer@nsidc.org

Executive Summary

Well defined, efficient, and sustainable data management is a prerequisite to move Arctic observing initiatives from a loose collection of individual projects and missions to a unified observing system advancing a common vision. Besides interdisciplinary scientific questions, data management is the glue that links activities, projects, disciplines, and scientists, allowing them to leverage previous work while avoiding duplication of efforts.

Ambitions for data management during the International Polar Year (IPY) were high. Eventually, what worked in practise was the simple solution. Data management is a tool that, when used correctly, multiplies the investment in operational and scientific observations. It bridges operational and scientific communities and promotes interdisciplinary science. Many lessons were learned during IPY, the most important of which is that better coordination of both scientific and data management issues is vital to achieve progress.

IPY raised awareness of data management as an integrating tool. For many scientists it was the first encounter with metadata and proper documentation of data. It was the start of a educational process and cultural shift which should not stop. Through benefits to Arctic Science generally and Arctic Monitoring specifically, data management enables us to understand and address existing and upcoming challenges in the Arctic.

Moving forward, transnational funding and coordination must be in place to meet the transnational challenges in the Arctic through ongoing work to link existing infrastructure and systems, expand understanding of stakeholder needs, promote communication between scientists and data managers, and develop suitable and relevant tools in support of science.

Introduction

Science can be defined as the intellectual and practical activity encompassing the systematic study of the structure and behavior of the physical and natural world through observation and experiment. With the development of electronic instrumentation, remote sensing technologies, digital information, and communications technology (ICT, e.g., computers, the Internet), observations are now dominantly collected, stored, analysed, and shared as data in digital form (henceforth referred to as 'data'). In the case where observations are not collected in digital form they can be documented to create digital representations. The result is a rapidly increasing volume of data stored digitally and distributed around the globe. With proper data management and internet connections, these data resources can be accessed and used by integrated observation systems serving researchers within or across disciplines, policy makers, non-governmental organizations (NGOs), citizen scientists, or others with an interest in better understanding our world and meeting the grand challenges that face society.

Achieving the vision of open, discoverable, and broadly useful observational data presents a number of challenges. These include but are not limited to: promoting and adopting free and ethically open data sharing[1]; linking existing and emerging systems through standardization; making data resources discoverable by a wide range of users; articulating and supporting broad adoption of data management best practices; describing data with rich, usable and machine-readable documentation; providing researchers with access to professional data managers; and establishing the mechanisms and funding to ensure that data management systems and expertise can be sustained and data resources preserved over time.

The objective of this paper is to promote the value of data management to the Arctic observing and science communities, Arctic residents, and broader society. To meet this goal, we present an overview of data management issues, projects and programs, along with recommendations in support of the Arctic Observing Summit, April 30-May 2, 2013 in Vancouver, Canada. Additional background information on research data management is included in Appendix A.

1. Status of the current system: Existing Arctic data management coordination efforts

Observational efforts in the Arctic differ slightly from efforts in many other regions. In the Arctic, climatic conditions are harsh and monitoring programmes expensive, both in the sense of

---

[1] Open data accounts for the need to appropriately manage the sensitive nature of some forms of data including health records, sensitive cultural information, information on the location of vulnerable animal species in real time, and data constraints associated with ethics protocols.

deploying and maintaining instrumentation, and concerning data transfer. The harsh conditions impact deployment and maintenance of instrumentation and make data transfer difficult. When combined with unclear national jurisdiction, this has left much of the monitoring to scientific programmes. Consequently, Arctic data management is to a higher degree depending on the coexistence and linkage of operational and scientific efforts. In the context of an Arctic observing system, an ongoing challenge is to identify (a) what data are to be coordinated and (b) what systems are already established to manage data.

While it can be viewed as a regional effort decoupled from others, Arctic data management should be connected to broader efforts wherever possible. This section provides an overview of selected projects and programs. While we recognize that the list is not comprehensive, it does indicate that there are many existing data management initiatives that can provide a foundation for an integrated Arctic observing system. Despite the number of existing efforts, there are few well-established, interdisciplinary, international mechanisms for coordinating data management stakeholders, policies, activities, and systems.

Overview of selected data management initiatives

The International Polar Year program (IPY) has recently made progress in Arctic data management. The objectives of IPY were to promote international exchange and interdisciplinary science, and to generate a legacy of both observing systems and other infrastructure, focusing on the data collected during IPY 2007-2009. The International Polar Year Data and Information Service (IPYDIS) was formed as a distributed cooperation between data centres, archives, and networks. The main aim of IPYDIS was to track the data flow for IPY by sharing information on standards, data access mechanisms, archives, and services. Several countries, including Norway (through the Norwegian Polar Institute, the Institute of Marine Research, and the Norwegian Meteorological Institute), Canada (through the Canadian IPY Data Assembly Centre Network, including the Polar Data Catalogue of the Canadian Cryospheric Information Network), and the US, participated in the IPY data management efforts and contributed to the formation of the IPY Data Policy which promoted “fast” and “open” data access.

One of the programs most relevant to data management in the Arctic is the *Sustaining Arctic Observing Networks (SAON)* initiative. Developed before and during the IPY, SAON includes a number of tasks directly related to data management. For instance, Task 2: Polar Metadata Profile and Recommended Vocabularies aims to establish a minimum standard for documenting data for the purpose of discovery. At least nine of the twenty three active or proposed SAON tasks are focused on data management. Any data management efforts in support of an Arctic observing systems should consider these initiatives.

Additional international efforts include the International Arctic System for Observing the Atmosphere (IASOA), the International Ice Charting Working Group (IICWG), the International Network for Terrestrial Research and Monitoring in the Arctic (INTERACT), and the Scientific Committee on Antarctic Research (SCAR). Lessons learned from these efforts should be considered when designing data management for the Arctic.

Other national and international programs include the World Meteorological Organisation (WMO) Global Cryosphere Watch (GCW) which formed as a result of IPY, the Circumpolar Biodiversity Monitoring Program, and the Advanced Cooperative Arctic Data and Information System (ACADIS). There are also a wide variety of Community Based Monitoring (CBM) and Local and Traditional Knowledge (LTK) initiatives such as Exchange for Local Observations and Knowledge of the Arctic (ELOKA) and the Inuit Knowledge Centre at Inuit Tapiriit Kanatami. The Geomatics and Cartographic Research Centre at Carleton University works with northern community members to facilitate community based monitoring programs and collection and preservation of local and traditional knowledge.

In addition to regional Arctic efforts, the global science community is actively pursuing data management strategies and systems that offer opportunities for leverage, coordination, and learning. American funding agencies have turned to policy changes to inspire cultural shifts toward open data; requiring submission of Data Management Plans with research proposals is one such data management policy. The collaborative European SeaDataNet infrastructure includes data access and other software tools, a strong emphasis on standards, data products, and metadata services. The Global Earth Observation System of Systems (GEOSS) networks data providers through promoting standards, a single search portal, visualization tools, etc. over both the internet and telecommunication satellites. A contribution to GEOSS, the WMO Information System (WIS) is an attempt to create a unified data management system that can be utilised by all WMO programmes. WIS supports routine collection and dissemination of data and products as well as data discovery and access for all relevant WMO activities in weather, ocean, and climate. WIS is a global system, and will contain or link to relevant Arctic data. Members of the international oceanographic community, including participants of the International Oceanographic Data and Information Exchange (IODE) program, the OceanDataPortal, and the SCOR/IODE Data Publishing initiative, are deeply engaged with data management and associated challenges (e.g., Hankin et al., 2010) which overlap with Arctic data issues.

For links and additional program information, see Appendix B.

Summary/Recommendations

- There are many existing regional and global programs relevant to data management and Arctic observing.
- A comprehensive inventory of Arctic observing data efforts does not exist; however, programs such as SAON have identified this as a priority and some inventory projects are underway.

  RECOMMENDATION: Leverage investments made in past and current data management coordination efforts and operational systems to work towards linking various programs in ways that can more effectively support the needs of the Arctic observing community in meaningful ways.

RECOMMENDATION: Through the Arctic Observing Summit and other venues, make the Arctic observing community aware of, and encourage engagement in, observing data inventory initiatives. This can be driven by members of the data management community but will require participation by the broader community.

2. Observing System Design and Coordination

Well defined, efficient, and sustainable data management is a prerequisite to move Arctic observing initiatives from a loose collection of individual projects and missions to a unified observing system sharing a common vision. Besides interdisciplinary scientific questions, data management is the glue that links activities, projects, disciplines, and scientists, allowing them to share a common objective and avoiding duplication of efforts.

This section aims to provide a high level summary of lessons learned and recommendations related to data management system design and coordination. The body of research in this domain has grown significantly in recent years. Readers are directed to this corpus for a more detailed discussion (e.g. ICSU, 2004; NRC 2006a,b; ICSU, 2008; Lichota and Wilson, 2010; Parsons et al., 2011; ADCN, 2012; Fox and Harris, 2012; Pulsifer et al., 2011, 2012).

Design elements of a data management system in support of Arctic Observing

Community of Practice: At the core of sound data management in support of Arctic observing is a knowledgeable, engaged community of practice with representation from all stakeholders. Experience gained through recent local, national, and international data management efforts highlights the importance of an established community of practice in developing foundational elements of a effective data management system. These element include but are not limited to:

- A clearly defined and understood user community;
- Identification of data resources to be managed, curated, and/or shared;
- A collaboratively developed governance model that blends bottom-up and top-down decision making;
- Establishment of norms and collaborative approaches around data sharing, systems design, and funding. For example, many initiatives are currently focusing on establishing a culture of open data sharing (see http://sciencecommons.org, http://www.polarcommons.org);
- Mechanisms for acknowledging data contribution and use (e.g. data citation);
- Incentives for managing and sharing data;
- A community where members, both researchers and data managers, can be mutually supportive throughout the data life cycle

- Ability to collaboratively develop proposals or other approaches to secure the resources required to sustainably support management of observations-based data

Systems Development: Contemporary data management necessarily involves technology and the development of applied systems. A detailed discussion of technical elements of an observations-based data management system is beyond the scope of this document, however, key aspects must be considered as we move forward:

1) System interoperability: Interoperability can be defined as the ability for diverse systems and organizations to work together (inter-operate) on many levels. Establishing compatible formats provides some level of interoperability but does not guarantee effective sharing across disciplines or knowledge domains. Structural interoperability, including building, updating and maintaining networks of data systems across diverse technologies, standards, requirements, and funding schemes is even more challenging. Arguably the single biggest challenge to data management is semantic interoperability, or translating the knowledge embedded in data across contextual boundaries. Although semantic web technologies such as ontologies and Linked Open Data present possible approaches, we still need to develop good, well-documented knowledge models that are sustainable and representative of domains. This requires development of easy tools for involving the scientific community in the process. Interoperability is one of the core challenges that must be addressed and this must involve both human and technical systems.

2) Discovery tools: While seamless data access is the ideal for a data management system, a necessary and valuable first step is the establishment of effective and adequately populated data catalogues or other discovery mechanisms. Much progress has been made and we now have a number of well developed systems as outlined in Section 1 of this paper (e.g. GCMD, ACADIS, PDC etc.). However, to support an Arctic observing system requires that these catalogues be linked. This requires not only a physical connection between systems, but agreement on a core set o

3) Data access tools: While data discovery is necessary, data access is ultimately the most useful function of a data management system. As with data discovery tools, much progress has been made in this area in recent years, with many of the systems outlined in Appendix B including data repository and access functions. However, there is a need to develop integration tools that link distributed databases so that that they are useful and usable by the Arctic observing community.

4) Mechanisms for attribution: There is a legitimate concern by researchers that they may not receive due credit for the data in which they have hugely invested if they are made freely available to others before they have had the chance to publish. This presents a challenge for engagement with data management and data sharing. Mechanisms crediting the data provider must be developed and integrated in Arctic Observing Data Management systems (Parsons et al., 2011). Early work has focused on Digital Object Identifiers and associated data citations for

data sets. Several data journals are now in operation that provide a formal way of publishing data with attribution, for example Earth System Science Data (www.earth-system-science-data.net) and Nordicana-D (http://www.cen.ulaval.ca/nordicanad).

Collaboration and Coordination

Data management was a priority area for the IPY 2007-2008; this provided an unprecedented opportunity for global collaboration on polar data management systems. While not all IPY data management program goals were met, the increased recognition of the importance of data management set the stage for continued development (Parson et al., 2011). In 2010, members of SAON organized a meeting (SAON, 2010) that brought together more than twenty five international data managers (Lichota and Wilson, 2010). The meeting resulted in the development of a high level data strategy along with recommendations related to observation-based data management. In April of 2012, more than forty data managers and other stakeholders met to identify issues facing the Arctic data management communities and establish an action plan to move towards addressing issues and improving coordination among participants. Collaboration and coordination is happening as an organic, opportunistic process. Interested members of the Arctic observing community are engaging. However, there is no established forum to support such engagement. The Antarctic community has established the SCAR Standing Committee on Antarctic Data Management. No such organization exists for the Arctic. Initiatives such as SAON may provide a framework for such engagement; however, a coordinated approach and related resources are required to moved from the current ad-hoc approach to a more sustained, efficient, and consistent effort.

Summary/Recommendations

- Establishing a data management system requires a well developed community of practice.
- Progress has been made in establishing communities of practice and data systems, however, both need to be nurtured to continue moving forward.
- A data management system for Arctic observing data requires a number of design elements including interoperability, data discovery and access tools, and mechanisms for attribution.
- Ad-hoc collaboration and coordination is happening. More deliberate coordination and resources are required to move towards a sustained data management system.

  RECOMMENDATION: Establish an Arctic data management body to coordinate community building, system design, and general activity planning and execution. Build on existing communities, activities, projects, and other efforts across the Arctic.

  RECOMMENDATION: Leverage existing funding mechanisms that support community development (e.g. U.S. National Science Foundation's Research Collaboration Network program).

3. Stakeholder Perspectives and Integration in Observing System Design: How does data management address the universe of user needs?

Without data there is no science, and without science there is no data. Given the mutual dependencies between research and data, strong consideration should be given to stakeholder needs and requirements for data description, exchange, synthesis, and infrastructure when designing an integrated observing system. Stakeholders are recognized as including scientists, operational communities, Indigenous communities, and funders.

Scientific and operational stakeholder sub groups share similar data management motivations, but in many ways are also distinctly different. In the operational community, funding mainly goes to organisations to collect observations, while in the scientific community funding and observations are connected to individuals. Within science, funding depends on publications, a factor that effectively prevents fast sharing of data. This is a challenge not only for funding agencies but also for data management. Proper data attribution mechanisms will be crucial to addressing this challenge.

For operational data, data management and policies are highly disparate in various organisations, communities, nations, and frameworks. To benefit from all relevant activities and effectively realize the vision of a sustainable Arctic Observing System, data must be shared across communities. This will require national and international efforts to harmonise data policies. For example, in Norway, a national license (NLOD), compatible with Creative Commons Attribution License, has been developed, but scientific data has not been considered nor has the scientific community been involved in the development. By harmonizing policies across the heterogeneous Arctic observing community, establishment of an Arctic Data Management community could serve as a supporting pillar for sustainable Arctic Observing.

In the case where local and traditional knowledge is concerned and research is not driven from within a local community, the needs, rights, and management practices of Indigenous knowledge holders and stewards must be recognized. This includes appropriate engagement of Indigenous people, communities or organizations throughout the data life cycle. While required institutional ethics review processes may guide data management, Indigenous communities or organizations may have specific practices or requirements in place. It is the responsibility of researchers from outside the community to familiarize themselves with and adhere to these practices and requirements.

Successful design incorporates all aspects of usability into data system development by engaging and coordinating stakeholders early in the process and strategically throughout implementation. Usability should be considered broadly to incorporate stakeholder input in technical development, but also in tool selection as with attribution mechanisms, policy requirements, and ethical considerations.

Summary/Recommendations

- Data management system design must consider stakeholder needs and requirements.
- To understand stakeholder needs and requirements, we need better forums and mechanisms for communication and collaboration.

  RECOMMENDATION: Engage system users and other stakeholders early and often when designing systems.

  RECOMMENDATION: When establishing an Arctic data management body (see Recommendation from Section 2), include formal representation of key user groups. For example, if a committee is established, this committee should include representatives from the sciences, Indigenous groups, policy makers and possibly others groups.

4. Who are the funding organizations and other interested parties that support data management?

National

At present, funding for data management activities tends to be associated with national bodies. In the United States, the National Science Foundation (NSF), National Aeronautics and Space Administration (NASA), National Oceanic and Atmospheric Administration (NOAA), National Institutes of Health (NIH), and the US Geological Survey are high-profile agencies with data management goals.

In Canada, the federal government provides support for Arctic data management through departments and programs such as Aboriginal Affairs and Northern Development Canada (which funded/funds the International Polar Year, the Beaufort Regional Environmental Assessment, Northern Contaminants Program, and the Canadian High Arctic Research Station), the Department of Fisheries and Oceans, and GeoConnections. The ArcticNet Network of Centres of Excellence also provides funding to the Polar Data Catalogue for management and archiving of all ArcticNet data and metadata. The Canadian Tri-Council funding agencies, the Natural Sciences and Engineering Research Council (NSERC), the Social Sciences and Humanities Research Council (SSHRC), and the Canadian Institutes of Health Research (CIHR) are currently undergoing discussions regarding future plans for support of data management. At this time, SSHRC requires public access to data resulting from its funded projects.

In Norway, scientific data management is mainly supported through the Research Council of Norway (RCN) and to a lesser extent by the Norwegian Space Centre (NSC). However, most funding is allocated to development of systems, while operation and maintenance often lacks funding. Mandated by governmental agencies, operational data management usually has a long term mandate and functions quite differently from scientific data management, although harmonisation of data policies may improve coordination. The focus on Arctic data management is most prominent in RCN and NSC except for the new governmental initiative BarentsWatch (http://www.barentswatch.no/en/om/) which is more of a portal than a data management system. It is however also a funding mechanism as it supports development of products or systems that benefit the purpose of the portal.

The Netherlands Polar Programme is funded by 5 ministeries and organized by the Netherlands Organization for Scientific Research (NWO). More information can be found at: http://www.nwo.nl/en/research-and-results/programmes/Netherlands+Polar+Programme

The list presented is not a comprehensive list of national funding bodies. The key point is that there are many funding bodies at national levels. The science and data management communities need to understand their local context to support their efforts, however, these funding sources may or may not support engagement in regional or international activities.

Regional

Regional funding opportunities supporting the development of data management resources exist. For example in Europe, the European Union has a number of directives that effectively address sharing of data (e.g., INSPIRE, http://inspire.jrc.ec.europa.eu/; EU Water Framework Directive, http://ec.europa.eu/environment/water/water-framework/index_en.html). Funding is provided through Framework Programmes and COPERNICUS (http://copernicus.eu/). The main problem again is that funding is mainly allocated to development of systems, but operation and maintenance are still lacking appropriate support mechanisms as readily observed in the COPERNICUS evolvement.

International

International funding across continents remains a significant challenge. Although transnational funding exists in Europe, appropriate funding mechanisms to integrate Europe, North America, and Russia in Arctic data management are lacking. Neither coordination between funding bodies in the regions nor consideration of joint proposals are available.

Foundations with an interest in data management issues have established funding programs and may provide project funding opportunities (e.g. the Sloan foundation, http://www.sloan.org/major-program-areas/digital-information-technology/data-and-computational-research/). The Belmont Forum, another possible source, works to enhance cooperation and coordination of global environmental change research, including improving co-design, co-alignment, and co-funding of major research programs (http://igfagcr.org/index.php/about-us).

While many are interested in supporting data management as part of the implementation and operation of an Arctic observing system, there are few cross-cutting forums for coordinating the sustainability of these efforts. As a result, data management tends to be short-term, small-scale, locally-focused, and inconsistent. With longer-term, consistent support, data management efforts can mature to address gaps in vision, coordination, governance, and standardization. Well-managed data are an inherently necessary part of observational and system science. A long-term, efficient, international network of Arctic observing systems requires a long-term, efficient, international network of Arctic observation data.

Summary/Recommendations

- At present, funding for data management activities tends to be associated with national bodies though regional initiatives exist.
- Funds tend to go to system development, support for system operation and maintenance are limited.
- International funding opportunities are limited.

  RECOMMENDATION: The Arctic observing community, including data managers, must continue to work together with funders, legislators, and other stakeholders to provide international funding opportunities.

  RECOMMENDATION: When establishing an Arctic data management body (see Recommendation from Section 2), task this body with developing international, coordinate funding appeals.

  RECOMMENDATION: When establishing an Arctic data management body (see Recommendation from Section 2), include formal representation of funders.

Conclusions

While IPY provided a start, more needs to be done to expand coordination of data management efforts. Many observing initiatives and related data management projects are already in place, though the number of activities alone can present a barrier to data discovery, access, and further to the efficient advancement of science. Recommendations for data management include leveraging existing resources to develop a more integrated, widely accessible system that provides relevant functionality and is easy for researchers and other stakeholders to use. Do not reinvent data management systems for the Arctic, but utilise global or regional initiatives wherever possible.

A recognized, coordinated body is needed to promote and manage multi-level (lab, project, domain, and cross-domain communities as well as local, regional, and international communities) and multi-type (formats, structure, semantics) interoperability between projects and systems. Arctic data management is a transnational activity and requires transnational coordination and support; a formal, international structure or body is needed to advance observing system goals by addressing data management challenges. To avoid adding additional layers of management, embed this function in existing bodies such as SAON, IASC, and initiatives that are already connected with stakeholder communities. Directly engage funding agencies to ensure continuity of data management communities of practice and technologies.

The Arctic has the potential of being a truly interdisciplinary laboratory. A formal, representative body can best inventory existing efforts, coordinate and fund development of a common framework for linking and leveraging identified efforts, ensure all stakeholders are heard, and coordinate the ongoing communication and collaboration needed to advance understanding of this crucial region of the planet.

Appendix A - Background Information on Research Data Management

Research data management is not new - the challenges of discovering, accessing, and using data have existed for centuries. More that fifty years ago, the ICSU World Data Center system was developed to manage data resulting for the International Geophysical Year of 1957-58 (Ruttenburg, 1992). With the advent of digital computers and networked information systems, the nature of data has changed dramatically, particularly in the last 50 years. Data can be manipulated with new technologies that increase the speed and scope of existing analytical techniques and provide a platform for new techniques that were not previously practical or possible. The broad adoption of the Internet starting in the 1980s and the invention of the World Wide Web in 1992 provided the capacity to network data on a massive, global scale and increased data availability and access. By bringing together computational power with large volumes of accessible data, we are now entering a period of “data intensive science” or the 4th paradigm (Hey et al., 2009). For example, research in environmental microbiology has undergone a revolution in analytical power and interpretation as a result of the online, open access data bases of the National Centre for Biotechnology Information. This includes the integrated database retrieval system Entrez, which provides access to diverse, linked databases containing 690 million records (as at July 2012; NCBI Resource Coordinators, 2013). These databases are now routinely used in research in Arctic ecosystems (e.g. Galand et al., 2009), and are an essential resource for Arctic monitoring.

Advances in technology present opportunities for new approaches to science. Data can be more quickly and easily shared across the globe using Information and Communication Technology. A range of products from relatively low-priced, accessible handheld devices (Smartphones, tablets, etc.) to servers based in remote communities extends the reach of networks, while new computing environments and technologies, including supercomputers, parallel computing and the increasingly well-established cloud, offer more flexibility and lower the bar for individuals to conduct computation-heavy research. Data systems for discovery, access, and preservation are opening doors to broader and deeper data-driven science. Social media is changing our definition of a computing platform as Facebook and Twitter are being used to describe, collect and transmit data (ex. Takhteyev et al., 2012; Sea Ice for Walrus Outlook). These platforms not only support academic science, but also support professional and Citizen Science which “enlists the public in collecting large quantities of data across an array of habitats and locations over long spans of time" (Bonney et al., 2009). Along with the development of small, collaboratively developed microformats to support data transfer (eg., GeoJSON, http://geojson.org), these advances begin to bring diverse science communities together and open opportunities for transdisciplinary science.

However, with each new opportunity comes new data management challenges; effectively sharing data and information at a regional (i.e., Arctic) and global scale across disciplines and communities of practice remain an elusive goal. While there are many tools that can facilitate access, differences in data formats, systems, and semantics can still present a practical challenge to data sharing, interoperability and reuse.

Appendix B - Additional Information on Selected National and International Data Management Initiatives

Legacy of IPY Data Management Program: The objectives of IPY were to promote international exchange and interdisciplinary science, and to generate a legacy of both observing systems or other infrastructure, focusing on the data collected during IPY 2007-2009. To achieve the latter, data management was deemed important and a collaborative approach was attempted. During the IPY it was evident that operational data have a scientific perspective and scientific data have an operational monitoring perspective.

The International Polar Year Data and Information Service (IPYDIS) was a distributed cooperation between data centres, archives and networks. The main aim of IPYDIS was to track the data flow for IPY. To achieve this, information on standards, data access mechanisms, archives, and services had to be provided. IPYDIS activities were guided by the IPY Data Policy and Management Subcommittee. The most important data management document of IPY was the IPY data policy. To be recognised as an IPY project, projects had to adhere to the policy. In brief, the contents of the IPY data policy are best summarised by the words “fast” and “open” data access. The intention of this policy was to promote scientific cooperation and interdisciplinary science. This was achieved by providing searchable metadatabases and online access to data sets. IPYDIS lacked funding at the international level, however, a linkage of archives and systems at the metadata level was achieved utilising the standard metadata sharing protocols and descriptions. This was a low cost, easily implemented solution that has been used in support of WMO Global Cryosphere Watch (GCW) as well as national projects after IPY. NASAs Global Change Master Directory served as the central IPY catalog harvesting metadata for national sub nodes. The situation concerning IPY data management is summarised in the report State of Polar Data by Parsons et al. (2011).

Norwegian Program: In Norway, IPY data management was implemented as a distributed system to which the Norwegian Polar Institute, the Institute of Marine Research, and the Norwegian Meteorological Institute contributed. The linkage between catalogues at these institutions continues after IPY, both in support of the existing IPY datasets, but also in new national scientific infrastructure projects funded by the Research Council of Norway. In the development of the Svalbard Integrated Arctic Earth Observing System (SIOS) which is supported by the European Strategy Forum on Research Infrastructures (ESFRI), the IPY metadata exchange has also been used to set up a demonstrator including the previously mentioned Norwegian institutions and PANGAEA in Germany.

Canadian Program: The Canadian IPY Program leveraged existing government and academic resources to make significant progress in the area of polar data management (http://www.api-ipy.gc.ca/pg_ipyapi_016-eng.html). Specifically, data management capacity developed through ArcticNet (http://www.arcticnet.ulaval.ca) was further developed with support from the federally funded IPY program and the formation of the Canadian IPY Data Assembly

Centre Network (now the Canadian Polar Data Network). A flagship result is the Polar Data Catalogue (PDC, http://www.polardata.ca), co-founded by the Canadian Cryospheric Information Network and ArcticNet. The PDC is a database of metadata and data that describes, indexes and provides access to diverse data sets generated by Arctic and Antarctic researchers. The recent work has added data access and archival functions, and this resource is now being adopted by other programs within Canada (e.g. the Northern Contaminants Program) and internationally (e.g. the Circumpolar Biodiversity Monitoring Program).

Circumpolar Biodiversity Monitoring Program: The Circumpolar Biodiversity Monitoring Program has adopted the IPY data policy in developing their Arctic Biodiversity Data Service (ABDS). The ABDS portal aims to help users explore biodiversity data from across the Arctic. The system has started work with a pilot project looking at integrating data from the Seabird Information Network and the Seabird Expert Group of the Conservation of Arctic Flora and Fauna (CAFF). The portal focuses on connecting taxonomic and geospatial search elements and will be incorporating social science data including linguistic and population data (http://www.abds.is)

Community Based Monitoring (CBM) and Local and Traditional Knowledge (LTK) Initiatives: Community-based research, including community-based monitoring and local and traditional knowledge projects face unique data management challenges. The nature of data generated from these initiatives can be in the form of recorded narratives, qualitative observations, transcripts, various types of multimedia as well as tabular data and databases. Data resulting from CBM and LTK projects may have a need for stewardship that includes a consideration of confidentiality as well as explicit recognition of knowledge contributors, depending on the context. Communities and researchers are working together as part of the Exchange for Local Observations and Knowledge of the Arctic (ELOKA) program and continue to provide data management and user support to facilitate the collection, preservation, exchange, and use of local observations and knowledge of the Arctic. Other related programs include the Inuit Knowledge Centre at Inuit Tapiriit Kanatami (http://www.inuitknowledge.ca) and the Geomatics and Cartographic Research Centre (https://gcrc.carleton.ca) at Carleton University.

Advanced Cooperative Arctic Data and Information System (ACADIS) Metadata Brokering Effort: The ACADIS Arctic Data Explorer is an example of a metadata brokering system, meant to allow users to search across many different data repositories with a single query. Given the large number of Arctic data repositories, metadata brokering helps those looking for data by reducing the number of portals they need to know about, visit, and search. From an infrastructure perspective, metadata brokering offers the benefits of a data clearinghouse while allowing the participating repositories to maintain their specialized focus. Partnering with NSF's Earthcube and the Italian Earth and Space Science Informatics Lab (IESSIL), the initial launch of the Arctic Data Explorer offers keyword, temporal, and spatial search across multiple data catalogs including the ACADIS partners, National Snow and Ice Data Center (NSIDC), the National Center for Atmospheric Research (NCAR) Computational and Information Systems Laboratory's ACADIS gateway, and the NCAR Earth Observing Laboratory (EOL) Computing, Data and Software Facility. The Norwegian Meteorological Institute (NMI) and the US National

Oceanographic Data Center (NODC) are additional data catalogs accessible through the initial release of the brokered search (http://nsidc.org/acadis/search).

Relevant Global Programs

In addition to regional Arctic efforts, the global science community is actively pursuing data management strategies and systems that offer opportunities for leverage, coordination, and learning. The collaborative European SeaDataNet infrastructure includes data access and other software tools, a strong emphasis on standards, data products, and metadata services. The Global Earth Observation System of Systems (GEOSS) networks data providers through promoting standards, a single search portal, visualization tools, etc. over both the internet and telecommunication satellites. A contribution to GEOSS, the WMO Information System (WIS) is an attempt by the World Meteorological Organisation (WMO) to create a unified data management system that can be utilised by all WMO programmes. Powered by ISO 19115 standardized metadata, WIS supports routine collection and dissemination of data and products as well as data discovery and access for all relevant WMO activities in weather, ocean, and climate. WIS is a global system, and will contain or link to relevant Arctic data. Members of the international oceanographic community, including the International Oceanographic Data and Information Exchange (IODE) program, the OceanDataPortal, and the SCOR/IODE Data Publishing initiative, are deeply engaged with data management and challenges of overlapping Arctic data issues (e.g., Hankin et al., 2010).

The *WMO Global Cryosphere Watch* (GCW) is a direct consequence of IPY. It was endorsed by the WMO Congress in May 2011 and is focused on global monitoring of the cryosphere. GCW has dedicated data management activities which focus on integration with WIS and metadata synchronisation between data centres to establish a single point of entry for cryospheric data discovery. It does not replace existing archives but serves to improve the visibility of cryospheric data.

The *International Arctic System for Observing the Atmosphere* (IASOA, http://iasoa.org/) mission is to advance coordinated research objectives from independent pan-Arctic atmospheric observatories through (1) strategically developing comprehensive observational capacity, (2) facilitating data access and usability through a single gateway, and (3) mobilizing contributions to synergistic science and socially-relevant services derived from IASOA assets and expertise. IASOA is currently working to improve data management with special emphasis on metadata, controlled vocabularies, and linkage to WMO activities, especially WIS.

The goal of the *International Ice Charting Working Group* (IICWG, http://nsidc.org/noaa/iicwg/) is to promote coordination among ice centers on all topics surrounding sea ice and icebergs including data management. The IICWG Data, Information, and Customer Support Standing Committee has engaged specifically with data exchange, formats, mapping systems, and visualization.

The *International Network for Terrestrial Research and Monitoring in the Arctic* (INTERACT) brings together scientists from diverse domains to study and assess environmental change in the European Arctic and more broadly. With a focus on long-term monitoring stations, INTERACT recognizes the importance of quality data management for not only supporting research based on data reuse but also for maintaining the integrity of current research. The website for INTERACT Work Package 7 (http://www.eu-interact.org/joint-research-activities/data-management/) states "[Incompatible data archives impose] a risk that information gaps and redundancies remain undetected, that synergies across activities remain unexploited, and that new experiments are being performed on grounds that are affected by previous activities. In addition, public outreach is hampered by the lack of centralised information management, and data communication unnecessarily complicated."

Another global community relevant to Arctic efforts is the Antarctic research community as represented by the *Scientific Committee on Antarctic Research* (SCAR). The SCAR Data and Information Management Strategy Report provides recommendations on governance and policy, incentivising cultural change, standards, and interoperability (Finney, 2009). A subsequent report (SCAR, 2011) details a Data Policy for the community. Lessons learned from these efforts should be considered when approaching similar issues in the Arctic.

References

ADCN (2012). *Arctic Data Coordination Network (ADCN) Workshop report* http://www.arcticobserving.org/images/stories/Tasks/TN2/adcn_april_2012_minutes.doc

Bonney, R., Cooper, C.B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K.V., & Shirk, J. (2009). Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy. *BioScience*, 59 (11), 977-984. Retrieved from: http://www.bioone.org/doi/abs/10.1525/bio.2009.59.11.9

Finney, K. (2009), SCAR Report: Data and Information Management Strategy (DIMS) 2009 - 2013. *Scientific Committee on Antarctic Research Report #34.* Retrieved from: http://www.scar.org/publications/reports/Report_34.pdf

Fox, P. & Harris, R. (2012). ICSU and the Challenges of Data Information Management for International Science. *Data Science Journal*. Retrieved from: http://isds.nict.go.jp/wds-kyoto-2011.org/pdf/IS301.pdf

Galand, P.E., Casamayor, E.O., Kirchman, D.L., & Lovejoy, C. (2009). Ecology of the rare microbial biosphere of the Arctic Ocean. *Proc. Natl. Acad. Sci*. 52: 22427-22432. doi:10.1073/pnas.0908284106.

Hankin, S., Bermudez, L., Blower, J.D., Blumenthal, B., Casey, K.S., Fornwall, M., Graybeal, J., Guralnick, R.P., Habermann, T.,, Howlett, E., Keeley, R.,, Mendelssohn, R., Schlitzer, R., Signell, R., Snowden, D., & Woolf, A. (2010). Data Management for the Ocean Sciences - Perspectives for the Next Decade. *Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society (Vol. 1),* Venice, Italy, 21-25 September 2009, Hall, J., Harrison, D.E. & Stammer, D., Eds., ESA Publication WPP-306. doi:10.5270/OceanObs09.pp.21

Hey, T., Tansley, S., & Tolle, K. (2009). *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research. Retrieved from: http://www.amazon.com/The-Fourth-Paradigm-Data-Intensive-Scientific/dp/0982544200

ICSU (2004). Report of the CSPR Assessment Panel on Scientific Data and Information. ICSU, Paris. Retrieved from: www.icsu.org/Gestion/img/ICSU_DOC_DOWNLOAD/551_DD_FILE_PAA_Data_and_Information.pdf

ICSU (2008). Ad hoc Strategic Committee on Information and Data. Final Report to the Committee on Scientific Planning and Review. Retrieved from: http://www.icsu.org/publications/reports-and-reviews/scid-report/scid-report.pdf

Lichota, G., & Wilson, S. (2010). SAON Data Management Workshop report: Developing a Strategic Approach. *Sustaining Arctic Observing Network (SAON) Report*. Retrieved from: http://www.arcticobserving.org/images/stories/DRAFT_REPORT_-_SAON_Data_Management_Workshop_Report_FINAL_GBL0818101.pdf

NCBI Resource Coordinators (2013). Database resources of the National Center for Biotechnology Information. *Nucl. Acids Res.* 41(D1): D8-D20. doi:10.1093/nar/gks1189

NRC (2006a). Toward an Integrated Arctic Observing Network. Retrieved from: http://www.nap.edu/catalog/11607.html

NRC (2006b). Committee on Designing an Arctic Observing Network, National Research Council. ISBN: 0-309-65484-X, 128 pages. Retrieved from: http://www.nap.edu/catalog/11607.html

Parsons, M. A., de Bruin, T., Tomlinson, S., Campbell, H., Godøy, Ø and LeClert, J. (2011). *The State of Polar Data – the IPY Experience in Understanding Earth's Polar Challenges: Summary by the IPY Joint Committee, International Polar Year 2007–2008*. Editors: Igor Krupnik, Ian Allison, Robin Bell, Paul Cutler, David Hik, Jerónimo López-Martínez, Volker Rachold, Eduard Sarukhanian and Colin Summerhayes, ISSN 0068-0303 ; no. 69, ISSN 1795-7796 ; no. 4, ISBN 978-1-896445-55-7, ISBN-13: 978-952-484-403-1. Retrieved from: http://www.icsu.org/publications/reports-and-reviews/ipy-summary/ipy-jc-summary-part3.pdf

Ruttenburg, S. (1992). The ICSU World Data Centers. *Eos* Vol. 73, No. 46, November 17. pp 494-495. Retrieved from: http://tinyurl.com/btlphg4

SCAR (2011). SCAR Data Policy. Scientific Committee on Antarctic Research Report #39. Retrieved from: http://www.scar.org/publications/reports/Report39.pdf

Sea Ice for Walrus Outlook. Retrieved from: http://www.arcus.org/search/siwo

Takhteyev, Y., Gruzd, A., & Wellman, B. (2012). Geography of Twitter networks. *Social Networks*. Volume 34, Issue 1, January, pp 73–81. http://dx.doi.org/10.1016/j.socnet.2011.05.006

Pulsifer, P. L., Laidler, G. J., Taylor, D. R. F., & Hayes, A. (2011). Towards an Indigenist Data Management Program: Reflections on Experiences from the Inuit Sea Ice Use and Occupancy Project. *Cartographica*, 55(1), 108–124.

Pulsifer, P.L., Gearheard, S., Huntington, H., Parsons, M., McNeave, C., & McCann, H.S. (2012). The role of data management in engaging communities in Arctic research: Overview of the Exchange for Local Observations and Knowledge of the Arctic (ELOKA). In special issue of *Polar Geography* entitled 'Arctic Community Engagement During International Polar Year'. Alain Cuerrier, Nancy Doubleday, Bryan Grimwood Eds. Volume 35, 3:1-20. DOI:10.1080/1088937X.2012.708364