TOWARDS AN INTERNATIONAL POLAR DATA COORDINATION NETWORK

P L Pulsifer[1]*, L Yarmey[1], Ø Godøy[2], J Friddell[3], M Parsons[4], W F Vincent[5], T de Bruin[6], W Manley[7], A Gaylord[8], A Hayes[9], S Nickels[10], C Tweedie[11], J R Larsen[12], and J Huck[12]

[1]*National Snow and Ice Data Center, University of Colorado, 449 UCB, University of Colorado, Boulder, CO 80309-0449, USA*
**Email:* pulsifer@nsidc.org
[2]*Norwegian Meteorological Institute, Henrik Mohns plass 1, 0313 Oslo, Norway*
[3]*Canadian Cryospheric Information Network, University of Waterloo, 200 University Ave. W., Waterloo, ON, N2L 3G1, Canada*
[4]*Research Data Alliance, Rensselaer Polytechnic Institute, Troy, NY 12180, USA*
[5]*CEN: Centre d'Etudes Nordiques, Laval University, Quebec City, G1V 0A6, Canada*
[6]*NIOZ Royal Netherlands Institute for Sea Research, Texel, The Netherlands*
[7]*Institute of Alpine and Arctic Research, University of Colorado, Boulder, CO 80309-0450, USA*
[8]*Nuna Technologies, PO Box 1483, Homer, AK 99603, USA*
[9]*Geomatics and Cartographic Research Centre, Carleton University, 1125 Colonel By Dr., Ottawa, ON, K1S 5B6, Canada*
[10]*Inuit Quajisarvingat, Suite 1101, 75 Albert St., Ottawa, Ontario, K1P 5E7, Canada*
[11]*Biology and the Environmental Science and Engineering Program, University of Texas at El Paso, El Paso, TX 79968, USA*
[12]*University of Alberta Libraries, University of Alberta, Edmonton, Alberta, T6G 2J8, Canada*

ABSTRACT

Data management is integral to sound polar science. Through analysis of documents reporting on meetings of the Arctic data management community, a set of priorities and strategies are identified. These include the need to improve data sharing, make use of existing resources, and better engage stakeholders. Network theory is applied to a preliminary inventory of polar and global data management actors to improve understanding of the emerging community of practice. Under the name the Arctic Data Coordination Network, we propose a model network that can support the community in achieving their goals through improving connectivity between existing actors.

Keywords: Data management, Network, Arctic, Antarctic, International Polar Year, Interoperability, Data sharing

1 INTRODUCTION

Well defined, efficient, and sustainable data management is a prerequisite to moving 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 binds activities, projects, disciplines, and scientists, enabling them to leverage previous work while avoiding duplication of efforts. 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. 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 and, by extension, challenges faced by society as a whole.

Research data management in the polar regions is not new. The challenges of discovering, accessing, and using data have existed for centuries. More than fifty years ago, the ICSU World Data Centre system was developed to manage data resulting from the International Geophysical Year of 1957–1958 (Ruttenberg, 1992). The International Polar Year 2007–2009 (IPY) continued in this spirit and resulted in significant progress towards establishing an international polar data management network (Parsons, de Bruin, Tomlinson, Campbell, Godøy, & LeClert, 2011). However, the form and context of data collection and management have changed dramatically in recent decades and continue to rapidly evolve. On the basis of developments in information and communication technology, there are unprecedented opportunities to collaborate through Internet and data driven science (Hey, Tansley, & Tolle, 2009).

Among the many lessons learned during the recent IPY, one of particular importance is that maximizing the benefit of these new research and data management platforms requires better coordination of both scientific and data management activities. However, in the limited timeframe of IPY, no clear consensus on ‘how to do data management coordination’ (ADCN, 2012, page 2) was established. In this paper, we aim first to contribute to a better understanding of the Arctic data management domain, including an analysis of community values and objectives. Secondly, using ideas from network science and elsewhere, we present a model for how we might ‘do’ Arctic data management coordination. While the focus of the paper is on Arctic data management coordination, we see this as part of a move towards polar data management coordination that is nested within an even broader global coordination effort.

Although an identifiable, fully functioning coordinating body for Arctic data management does not yet exist, a community of practice (CoP) and a related network of actors are emerging. A CoP can be defined as a group of people who share a craft and/or a profession. It can evolve naturally because of the members' common interest in a particular domain or area, or it can be created specifically with the goal of gaining knowledge related to their field (Wenger, McDermott, & Snyder, 2002). The Arctic data management CoP has engaged in activities that have initiated the process of identifying and articulating the purposes, activities, and form of a network (Table 1).

Table 1. Selected meetings and other activities supporting the development of the ADCN

Activity	Year(s)	Outcomes
International Polar Year	2007–2009	Increased profile of science data management, documentation of IPY data, and establishment of major repositories
Sustaining Arctic Observing Networks (SAON) Meeting (IPY Conference, Oslo)	2010	Focussed dialogue on Arctic data management needs, recommendations to SAON on priorities including data sharing; interoperability; preservation of data through sustainable, long-term archiving that has dedicated funding; and governance
Sustaining Arctic Observing Task Definition	2011	Formalized the need for, and intent to, develop an Arctic Data Coordination Network
Meeting on establishing an Arctic Data Coordination Network (ADCN, 2012, IPY Conference)	2012	Built on previous meeting. Further elaborated details of governance requirements, standardization, system development, and funding
Arctic Observing Summit	2013 (May)	A series of white papers outlining data coordination issues, needs, and existing system; a meeting held during the Summit to confirm additional participation
International Forum on Polar Data Activities in Global Data Systems	2013 (Oct)	Establishment of the need to include data coordination as part of international science planning activities. See: http://www.polar-data-forum.org/International_Polar_Data_Forum_Communique.pdf

Under the name the Arctic Data Coordination Network (ADCN), members of the CoP are working towards more clearly defining and formalizing mechanisms for network building, collaborating on practical activities of collective interest and value, and working on higher level principles and policies to guide the process. In Section 2, we summarize the results of an analysis of the key concepts, objectives, and goals that can be supported by ADCN. In Section 3, we present a partial map of the actors involved in, or seen as important potential nodes in, ADCN. Using network science as a theoretical framework, we put forward a network model that aims to address the multiscale, multidomain nature of Arctic data management coordination. The paper concludes with an overview of activities and results to date.

2 PRIORITIES FOR ARCTIC DATA COORDINATION NETWORK

In this section, we present the results of a high-level content analysis of two documents resulting from the above activities and related to ADCN development, specifically, Lichota and Wilson (2010) and ADCN (2012). Content analysis can be defined as the analysis of the manifest and latent content of a body of communicated material through classification, tabulation, and evaluation of its key symbols and themes in order to ascertain its meaning and probable effect. For methodological details see Krippendorff (2004).

The content analysis revealed core concepts across both documents:

- Arctic data should be shared 'Arctic-wide', across national, organizational, and disciplinary boundaries, in a way that is ethically open and free.
- A practical and time-boxed approach using existing resources as 'building blocks' should be employed for ADCN, rather than trying to recreates new entities,.
- Broadly adopted metadata (for discovery) and data-sharing standards are needed, emphasising the requirement for controlled vocabularies with sufficient detail for effective data management and use.
- Broad recognition and acceptance of data citation/attribution practices are required.
- Full engagement of stakeholders is critical, including primary data and metadata producers, data centre representatives, producers of data products, data users, science coordination bodies, funding agencies, Arctic indigenous peoples and other residents, and physical and social scientists.
- Operational and scientific communities must link together as both are important in monitoring the Arctic.

3 MAPPING ADCN

3.1 Theory and methods

The focus of this paper is on contributing to efforts to move ADCN forward. To do this, we can draw on network science to inform our approach. The objective was not to perform a comprehensive analysis but rather to establish a conceptual and methodological framework to help the emerging community move forward.

3.1.1 Theory

There are many definitions of the term *network*. In the context of ADCN, we refer to a group or system of interconnected people or things, including a group of people who exchange information, contacts, and experience for professional or social purposes. It is important to note that this definition corresponds closely with the definition of a community of practice. Thus, at some level we can see network building as community building. Humans are still the primary actors overseeing the management of Arctic data. Although the goal is for ADCN to facilitate, and perhaps evolve into, other types of networks (data, funding, etc.), at this nascent stage, we are focussing on the human and organizational components such as communication, information sharing, and collaboration.

Network science has the potential to contribute to network and community development from an Arctic perspective. Network science focuses on how networks emerge, what they look like, and how they evolve. It is being applied in many contexts, including biology, social movements, the Web, and others (Jeong, Tombor, Albert, Oltvai, & Barabási, 2000; Barabási, 2002; Newman, Barabási, & Watts, 2006). Network science has established important insights about how networks form, are sustained and fail (Albert, Jeong, & Barabási, 2000). Some key points are:

(i) Only a small number of links per node are needed to create a highly interconnected and robust network (the idea of a 'it's a small world' or 'six degrees of separation').
(ii) 'Weak ties' involving rare or occasional contact, i.e., individuals who are not necessarily part of the same organization and have a limited personal relationship, are important. Although the establishment of weak ties is not difficult or resource intensive, these ties provide the connectivity necessary to establish a robust network.
(iii) Robust networks are those able to withstand or overcome adverse conditions such as removal of a major node or hub (e.g., loss of funding for a major programme) and include multiple, highly connected hubs as well as less connected nodes.
(iv) 'Connectors' are vital. In a social setting, these are the 'people who know people' whereas in an organizational situation, they are organizations that are highly connected but may or may not engage in the practical activities of the community. Connectors may also act as mediators, where they do more than simply connect but also actively engage in the subject matter, perform acts of synthesis, abstraction, transformation, and so on that enable disparate actors to better communicate.

3.1.2 Methods

Here we used an adaptation of formal social network analysis. In the first phase, we used a sampling of actors (nodes) most closely resembling an *egocentric network with alter connections* approach (Hanneman & Riddle, 2005). In this method, we began with a selection of actors known to the authors and identified other actors to which they were connected. This was done through personal contact and Internet searching. Once a sample of actors was established, the actors were typed:

- GR - groups/organizations/institutions, for example, data centres, projects, programmes, government agencies, and so forth
- RC - coordination and advocacy bodies
- SD - systems, infrastructure, and technology developers and hosts, including standards bodies
- NG - civil society groups (Nongovernmental Organizations: NGOs)
- FR - funders

In Phase 2, using the identification of actors from Phase 1, a limited sample of existing relationships known to the authors was added to the map (e.g., ‘funded by’, ‘shares knowledge/data’, or ‘provides observations’). The completed social network map was used to develop preliminary observations and new hypotheses about the nature of the emerging and possible ADCN and related networks.

3.2 Results and discussion

3.2.1 Phase 1: actors analysis

The ‘actors’ analysis revealed a large number of possible actors relevant to the development of ADCN (Table 2). On the basis of the node-type analysis, we see a variety of different types of actors engaging in many activities. The number and diversity of nodes suggests, however, that there is significant data management capacity within the Arctic data management domain without suitable connectivity and information flow through an effective network. This presents the risk of significant fragmentation and duplication of efforts.

There are numerous global or multiscale research coordination and advocacy nodes. These are variably active in standards, policy or infrastructure development, or education activities. While some of the nodes may seem unimportant to those actors interested in the more technical aspects of data sharing, these organizations are important because they provide hubs that connect nodes. Because only weak ties are needed to promote connectivity, these organizations can (but do not need to) have strong ties to all other actors in the network. Moreover, they do not need expert-level data management knowledge, extensive resources, or to be actively engaged in data management activities. They do, however, need to be highly connected (strong or weak ties) with various actors in the community as is the case with the global/multiscale organizations identified in our analysis. These should play the role of ‘connector’, and we see this happening through organizations such as SAON, the International Arctic Science Committee (IASC), the Scientific Committee on Antarctic Research (SCAR) and its Standing Committee on Antarctic Data Management (SC-ADM), the Research Data Alliance (RDA), and others. These connector organizations are not necessarily directly involved in hard infrastructure or systems development, but they do facilitate the process through dissemination of information, linking actors, promoting standards, coordinating the development of strategies, establishing policy, and promoting education.

At the same time, we see actors working at a national, regional, or local scale that are involved in significant data management activities and that can benefit the larger network. Thus, the less connected nodes, working at more local levels of geography (from the ‘bottom up’) will also play an important role in achieving the goals of the community of practice because this is the level where much of the data management activity takes place.

While the previous two points may seem obvious, it is imperative to point out that if viewed through a network science lens, the development of a highly connected, durable international network that supports data sharing, including sensor and community-based data, requires the presence of multiple types of nodes and relationships. It cannot be a matter of a top–down command and control model, nor can the grass roots approach be expected to provide the same level of network connectivity as provided by a diverse network.

The analysis and model presented here does, nevertheless, provide a novel and current contribution to the ongoing discussions around coordinating Arctic data management activities. We recognize the need to more fully engage actors from an even broader range of countries and domains.

3.2.2 Phase 2: network map of model network structure

On the basis of an analysis of the actors identified in Phase 1, we propose a model network structure that through connectivity is robust and sustainable, enables pan-Arctic sharing of data, is primarily built on existing nodes, can facilitate collaborative activities such as standards adoption or development, can be a part of cultural shifts

Table 2. Selected Actors in the Arctic Data Management community by type and activities. Type code key: FR = Funder, GR = Group, NG = NGO, RC = Research Coordination, SD = Systems/Infrastructure Developer. Activity columns key: SO = *Primary* scale of activities, S = Standards activity, P = Data policy, I = Infrastructure/systems development, E = Data education, Y = Yes, N = No, ? = Unknown to the authors

SO	Actor	Link	Type	S	P	I	E
Global	Arctic Spatial Data Infrastructure	http://arctic-sdi.org/	RC, SD	Y	Y	Y	N
	Circumpolar Biodiversity Monitoring Program	http://www.abds.is	RC, SD	Y	Y	Y	Y
	Federation of Earth Science Information Partners	http://www.esipfed.org	CP	Y	N	N	Y
	Global Earth Observation System of Systems	http://www.earthobservations.org/geoss.shtml	RC, SD	Y	Y	Y	Y
	Global Cryosphere Watch	http://globalcryospherewatch.org/	GR, SD	Y	N	Y	Y
	ICSU World Data System (ICSU-WDS)	http://www.icsu-wds.org/	GR, RC, SD	Y	Y	Y	Y
	IASC	http://www.iasc.info/home/	RC	N	Y	N	N
	International Arctic Systems for Observing the Atmosphere	http://iasoa.org/	GR, SD	Y	N	Y	N
	International Network for Terrestrial Research and Monitoring in the Arctic	http://www.eu-interact.org/…	GR, SD	Y	N	Y	Y
	International Oceanographic Data and Information Exchange	http://www.iode.org/	GR, SD	Y	Y	Y	Y
	Polar Information Commons	http://www.polarcommons.org/	GR, SD	Y	Y	Y	N
	RDA	https://rd-alliance.org/	GR, RC	Y	Y	N	Y
	SeaDataNet	http://www.seadatanet.org/	GR, SD	Y	Y	Y	Y
	SC-ADM	http://www.scar.org/researchgroups/	GR	Y	Y	N	Y
	SAON	http://www.arcticobserving.org/	RC	Y	N	N	N
	World Meteorological Organization–Information System	http://www.wmo.int/pages/themes/wis/	GR, SD	Y	N	Y	N
Regional	Alaska Data Integration working group	http://adiwg.github.io	GR	Y	N	Y	Y
	Alaska Ocean Observing System	http://www.aoos.org/	GR, SD	Y	?	Y	Y
	Geographic Information Network of Alaska	http://www.gina.alaska.edu/	GR, SD	Y	N	Y	Y
	Oceans North	http://www.oceansnorth.org/	NG, FR	N	Y	N	Y
National	Advanced Cooperative Arctic Data and Information Service	https://www.aoncadis.org/home.htm	GR, SD	Y	Y	Y	N
	Arctic Observing Viewer	http://www.arcticobservingviewer.org/	GR, SD	Y	N	Y	N
	Arctic Data Centre	http://arcticdata.met.no/	GR, SD	Y	Y	Y	N
	Arctic Research Mapping Application	http://armap.org/	GR, SD	Y	N	Y	N
	Canadian Cryospheric Information Network/Polar Data Catalogue	http://www.polardata.ca/	GR, CP, SD	Y	Y	Y	Y
	Environmental Climate Data Sweden	http://ecds.se	SD	Y	Y	Y	Y
	National Institute of Polar Research (Japan)	http://www.nipr.ac.jp/	FR, RC, SD	Y	Y	Y	Y
	National Science Foundation (i.e., Division of Polar Programmes)	http://www.nsf.gov/div/index.jsp?div=PLR	FR	N	Y	N	Y
	Netherlands Organization for Scientific Research	http://www.nwo.nl	FR, RC	N	Y	N	N
	Norwegian Polar Institute	http://data.npolar.no/	GR, SD	N	N	Y	N
	Research Council of Norway	http://www.forskningsradet.no/prognett-polarforskning/	FR	N	Y	Y	Y
	Royal Netherlands Institute for Sea Research	http://www.nioz.nl/	RC, SD	Y	Y	Y	N

Local	Exchange for Local Observations and Knowledge of the Arctic	http://eloka-arctic.org	GR, SD	Y	Y	Y	Y
	Geomatics and Cartographic Research Centre	https://gcrc.carleton.ca	GR, SD	Y	Y	Y	Y
	Inuit Qaujisarvingat: Inuit Knowledge Centre, Inuit Tapiriit Kanatami	http://www.inuitknowledge.ca	GR, SD	Y	Y	N	Y

and may allow for much broader engagement of a range of stakeholders by providing a recognizable network for engagement. Figure 1 presents a map of this model network. The map indicates relationships (or lack thereof) between nodes.

Relationships are indicated by type: strong, weak, or non-existent. In the case of a non-existent relationship, the relationship implied is seen by the authors as an example of a desired relationship that may entail data sharing, collaboration, funds exchange, and so on. Given the breadth and depth of some organizations, multiple relationships can exist where there is an overlap of roles, exchanges, or affiliations in a relationship. This means that some actors can function as a connector and can serve other functions (e.g., funder or infrastructure hosting). To manage the visual complexity, Figure 1 is necessarily a reduced detail abstraction of the entire model.

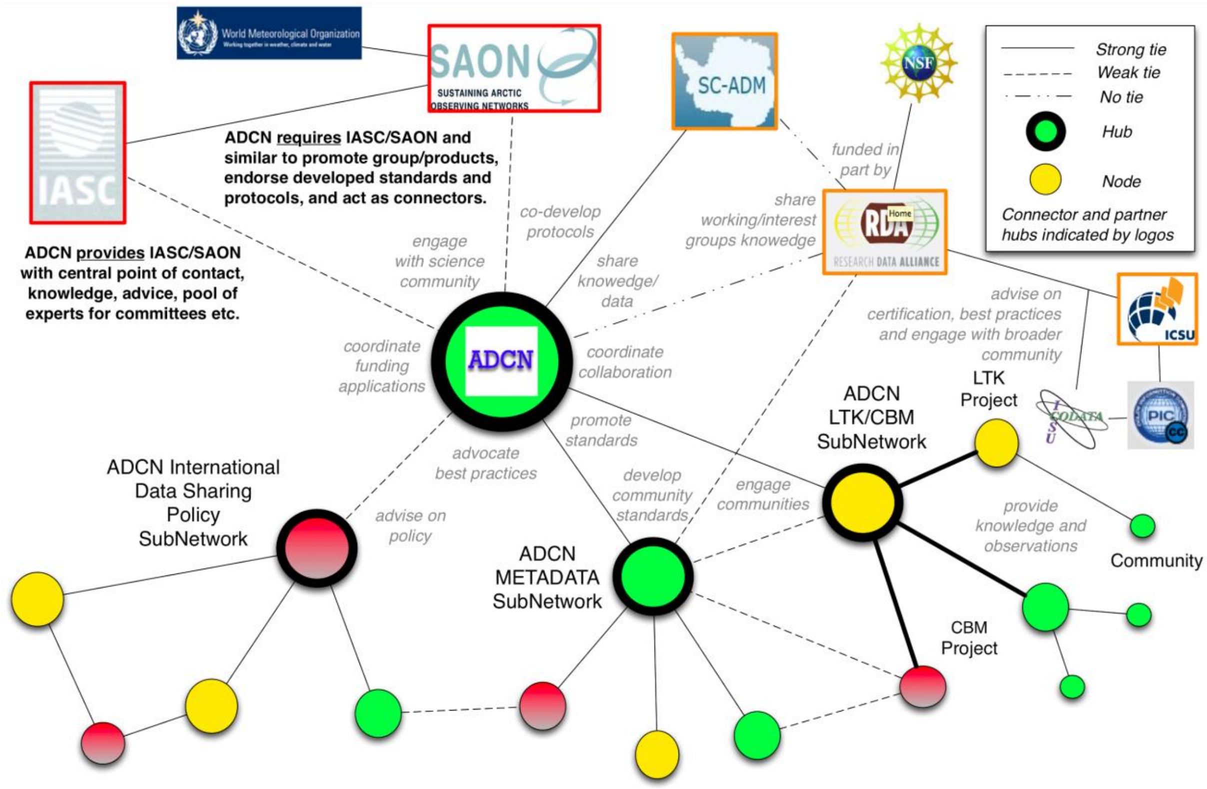

Figure 1. A potential network structure for the emerging community of practice. The different colours of the nodes represent the diversity of interest, activities, expertise, and other attributes of the actors involved.

ADCN can serve a number of functions through its members, ranging from technology development to advocacy. It will link well-connected hubs that in turn link to less-connected domain-specific nodes. The model sees ADCN connecting Arctic science and data management hubs and nodes to other Arctic, as well as global/multiscale, initiatives. In many cases, we see global initiatives that are highly relevant to Arctic data management and vice versa, and the evolving ADCN can promote the sharing of information about activities to a much broader community.

4 EXISTING AND PROPOSED NETWORK ACTIVITIES

Although ADCN is early in its development, a number of activities are underway. Networking tools have been established, such as a website and online collaborative platform on the Arctic Hub (https://arctichub.net/groups/adcn). To broaden our reach, a group has been established on the popular professional network site, LinkedIn. Social media tools such as Twitter are being used to disseminate information to the community.

A priority established by the community is the development of community-driven standards for sharing metadata and data. To move this forward, under the SAON programme, ADCN alongside SC-ADM is developing a community profile (template) of the International Organization for Standardization 19115 metadata standard, including a set of controlled vocabularies and translation to support semantic interoperability.

Discussions so far within this SAON task have revealed concern over both the usage of controlled vocabularies and how to effectively link and translate these in an interdisciplinary context. The goal is to establish a common profile that can be utilized by researchers, data managers, and others to readily exchange metadata.

During the International Forum on Polar Data Activities in Global Data Systems on 15–16 October 2013 in Tokyo, Japan, members of the community started formulating plans to engage with broader science and data management initiatives, including RDA, ICSU-WDS, the 3rd International Conference on Arctic Research Planning in 2015, and the SCAR Antarctic and Southern Ocean Horizon Scan.

5 CONCLUSION

The Arctic has the potential of being a truly interdisciplinary laboratory, with a manageable size of community that will enable advances in data sharing to be made. While IPY provided a start, more needs to be done. Many observing initiatives and related data management projects are already in place, but there is a lack of sufficient coordination to develop a more integrated, widely accessible, international system that provides relevant functionality and is easy and beneficial for researchers and other stakeholders to use.

Many of the interoperability efforts in the Arctic have been informally organized only recently as a legacy of IPY data management. A more formal, recognizable body, called ADCN in this paper, is needed to promote and cultivate a highly connected, robust network that can support open and free data sharing and the achievement of other goals established by the Arctic data and science communities of practice. To avoid adding unnecessary effort and cost, ADCN is emerging as a virtual organization without the establishment of new physical or extensive management infrastructure. These functions can be served by linking existing bodies through ADCN such as SAON, IASC, and other initiatives that already have some physical and management infrastructure and that are well connected with stakeholder communities. Additionally, efforts in the Arctic must be strongly linked to polar (i.e., SCAR) and global efforts because processes in the Arctic are an integral part of upstream processes.

ADCN and, by extension, this paper are an attempt to better understand and organize the informal activities undertaken to date. By using relevant network theory in tandem with our knowledge of existing polar data management resources, we can promote the development of a highly connected community of practice. With relatively small, targeted investments, this can significantly contribute to understanding and addressing existing and upcoming challenges in the Arctic and beyond.

6 ACKNOWLEDGEMENTS

This material is based in part upon work supported by the National Science Foundation under Grant Numbers ARC 0856634 and ARC 1231638.

7 REFERENCES

ADCN (2012) *Arctic Data Coordination Network (ADCN) Workshop report.* Retrieved January 7, 2014 from the World Wide Web: http://www.arcticobserving.org/images/stories/Tasks/TN2/adcn_april_2012_minutes.doc

Albert, R., Jeong, H., & Barabási, A.-L. (2000) Error and attack tolerance of complex networks. *Nature 406*, pp 378–381.

Barabási, A.-L. (2002) *Linked: The New Science of Networks*, Cambridge, MA: Perseus Publishing.

Hanneman, R.A. & Riddle, M. (2005) *Introduction to Social Network Methods,* Riverside, CA: University of California, Riverside. Retrieved October 1, 2014 from the World Wide Web: http://faculty.ucr.edu/~hanneman/

Hey, T., Tansley, S., & Tolle, K. (2009) *The Fourth Paradigm: Data Intensive Scientific Discovery*. Retrieved January 7, 2014 from the World Wide Web: http://www.amazon.com/TheFourthParadigmDataIntensiveScientific/dp/0982544200

Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N., & Barabási, A.-L. (2000) The large-scale organization of metabolic networks. *Nature 407*, pp 651–654.

Krippendorff, K. (2004) *Content Analysis*, Thousand Oaks, CA: Sage Publications.

Lichota, G. & Wilson, S. (2010) *SAON Data Management Workshop report: Developing a Strategic Approach*. Retrieved January 7, 2014 from the World Wide Web: http://www.arcticobserving.org/images/stories/DRAFT_REPORT__SAON_Data_Management_Workshop_Report_FINAL_GBL0818101.pdf

Newman, M., Barabási, A.L., &. Watts, D.J. (Eds.) (2006) *The Structure and Dynamics of Networks*, Princeton, NJ: Princeton University Press.

Parsons, M.A., de Bruin, T., Tomlinson, S., Campbell, H., Godøy, Ø, & LeClert, J. (2011) The State of Polar Data: the IPY Experience. In Krupnik, I., Allison, I., Bell, R., Cutler, P., Hik, D. López-Martínez, J., et al. (Eds.), *Understanding Earth's Polar Challenges: Summary by the IPY Joint Committee, International Polar Year 2007–2008*, Washington, DC: National Academies Press. .

Ruttenberg, S. (1992) The ICSU World Data Centers. *EOS 73*(46), pp 494–495. Retrieved October 1, 2014 from the World Wide Web: http://onlinelibrary.wiley.com/doi/10.1029/91EO00365/abstract

Wenger, E., McDermott, R., Snyder, W.M. (2002) *Cultivating Communities of Practice*, Cambridge, MA: Harvard Business Press.

(Article history:Available online 17 October 2014)