

Canadian Polar Data Workshop

26 and 27 May 2015

Albert at Bay Hotel, Ottawa, Ontario

**Report prepared in partial fulfillment of deliverables required by
NSERC Statement of Work:**

***Canadian and international polar/Arctic research data management –
Context and avenues to enhance collaboration***

March 2016

**Canadian Cryospheric Information Network/Polar Data Catalogue
Ellsworth F. LeDrew, Executive Director, CCIN/PDC
Julie E. Friddell, Associate Director, CCIN/PDC
Gabrielle Alix, Data Manager, CCIN/PDC**

TABLE OF CONTENTS

I.	Executive Summary.....	3
II.	Data Policy-Relevant Information and Outcomes.....	4
III.	Conclusions and Next Steps.....	5
 <i>The remaining sections are the detailed proceedings from the Workshop:</i>		
IV.	Welcome, opening remarks, and expectations for the Workshop.....	8
V.	Review of Questionnaire responses.....	8
VI.	Goals for the Position Paper.....	11
VII.	Introduction to the question “How can we coordinate polar data management activities in Canada, and what opportunities shall we pursue for Canadian leadership internationally?”.....	12
VIII.	Data Sharing.....	13
IX.	Data Access.....	18
X.	Data Preservation.....	21
XI.	Interoperability.....	24
XII.	Partnerships and Linkages.....	31
XIII.	Policy and Funding.....	35
XIV.	Coordination and Leadership.....	39

APPENDICES

1.	Invitation E-mail to Canadian Polar Data Workshop, April 2015.....	51
2.	Background document for a Canadian Polar Data Workshop, April 2015.....	53
3.	List of respondents to the questionnaire.....	58
4.	Open-ended questionnaire responses, edited to remove personal content (remaining questionnaire questions are represented in the Review of Questionnaire responses presentation by G. Alix, in Appendix 8).....	60
5.	Workshop Agenda.....	68
6.	Questions for breakout sessions.....	71
7.	Workshop participants and short bios (as available).....	74
8.	Presentations.....	77

Note: Audio recordings were made of all presentation and discussion sessions and can be provided upon request.

I. Executive Summary

As the importance of data and their proper management are increasingly recognized, the Canadian polar data community has grown in size, capability, and complexity. Concurrently, the need has grown to identify and connect the various organizations which produce, steward, and support polar data in Canada. In May 2015, a Canadian Polar Data Workshop was held in Ottawa to facilitate conversation between groups and present options for coordination of our activities. The Workshop agenda was crowd-sourced through a six-week national consultation on polar data, in the form of an online survey answered by individuals from 30 organizations. The Workshop was attended by 44 people, and 6 more individuals joined remotely. Presentations were made on data management topics and challenges, including data sharing, access, preservation, interoperability, policy, funding, and partnerships. In addition, national coordination, including the necessity to include all interested stakeholders at the very beginning of the process, and contribution to the international polar data community were discussed at length.



Canadian Polar Data Workshop participants, Ottawa, Ontario, Canada, 26-27 May 2015.

The primary outcomes of the Workshop are new opportunities for collaboration, creation of this comprehensive report on the proceedings of the Workshop and the landscape of polar data activities in Canada, and a position paper representing the views of the community and significant issues. The position paper, to be completed in mid-2016, will be drawn from the information in this report and will be actively shared with government agencies, academia,

northern Canadian organizations, funding bodies, and other stakeholder groups with connections to or responsibility for management of polar data in Canada.

The long-term goal of the Workshop and the coordination initiative is to design and implement, through consensus of the participants in the coordination exercise, a national management plan and structure for coordination of polar data activities and systems in Canada. Through this new structure, links within Canada and internationally will be established to create a robust technical infrastructure and human network for optimal stewardship of Canada's polar data resources. The model to be established will take lessons from the successes of other data communities and will ideally serve as an example of good practice for other sources and repositories of scientific data.

The Canadian Polar Data Workshop was co-hosted by the Canadian Cryospheric Information Network/Polar Data Catalogue (CCIN/PDC), the Polar Data Management Committee (PDMC), Exchange for Local Observations and Knowledge of the Arctic (ELOKA), the Canadian High Arctic Research Station (CHARS), the US National Snow and Ice Data Center (NSIDC), and the US National Science Foundation (NSF). Funding support was provided by the Natural Sciences and Engineering Research Council of Canada and the ArcticNet Network of Centres of Excellence at l'Université Laval. We thank all the speakers and the participants in both the meeting and the pre-Workshop consultation. The response and interest in this initiative have been significant and tell us that the time is right for coordination of the polar data community in Canada.

We look forward to working with NSERC, the Workshop participants, and all interested stakeholders to translate the excitement we felt at the Canadian Polar Data Workshop into actions for the polar data community in Canada. We are indebted to NSERC and to ArcticNet for their generous support of and enthusiasm for this initiative.

II. Data Policy-Relevant Information and Outcomes

As outlined in the detailed record of the Workshop discussions (provided in Sections IV to XIV below), meeting participants clearly felt that we need better data policy in Canada and that it is the federal government's place and obligation to lead and help unify policy development to drive improvements in data access and long-term preservation and to guide researchers on how to manage and archive data. Funding agencies should require formal data management plans for proposals and should work with the polar data community in Canada, as an initial partner, to build a system which can facilitate data archiving and access for future generations. The federal government should also provide leadership to develop an effective long-term funding strategy to support polar and other data management in Canada. These activities may be led by the Assistant Deputy Ministers of the science and technology agencies (S&T ADMs), including the Tri-Agencies, but would surely also include other relevant stakeholders such as universities and international organizations.

The cost of data management should be included in funding agencies' overall budgets, as incremental amounts to individual projects but also as baseline funding for development of data management infrastructure and systems. A funding program for data, similar to the Northern Supplements, was suggested as a way to begin support to individual researchers. Short-term funding of projects and data archives is a major concern, as data are so often lost and portals

go away when projects, or even major programs, come to an end. A detailed list of explicit funding needs is provided in the *Policy and Funding* section of this document.

Open data is considered almost universally to be a public good. Workshop participants felt that the default position of funding agencies and data collectors should be that data should be shared openly, if at all possible, and that exceptions to open sharing should be considered and approved on an individual basis. Notable exceptions to open sharing will almost certainly include traditional knowledge and sensitive data of Indigenous and northern communities and peoples. The appropriate approach to stewardship of sensitive northern and Indigenous data and information is currently under extensive consideration and will need to be a large part of the polar data community's future coordination and management plans. Participants felt that both encouragement ("carrot") and requirement ("stick") approaches would be necessary to achieve open sharing for data that are not determined to be legitimately exempt.

Other notable considerations from the Workshop deliberations included the following: (1) Employers and funding agencies should advocate for professional credit to researchers who archive data, to recognize their effort as a legitimate scholarly contribution; and (2) Participants expressed the desire that Canadian data, even if collected by people from other nations, should be archived in and available from Canadian repositories. Participants in the Canadian Polar Data Workshop were keen to coordinate their work together and are intent on having a follow-on Workshop in 2016 to continue the dialogue.

III. Conclusions and Next Steps

The overarching conclusions of the discussions at the Canadian Polar Data Workshop are summarized below:

- Canada has an obligation to provide easy access to polar data. It is the responsibility of our community to determine how to do that, and then to do it.
- Workshop participants felt the Canadian polar data management system has many strengths, but improvements are still required, including more tools for use and interpretation of the data, more training and support to educate and assist data contributors and users, and more coordination between various stakeholders, especially considering and including Indigenous and northern people and communities.
- The special challenge for preservation of polar data is the variety of our data. Archiving and providing access to varied data (the "long tail" of data) has different challenges than archiving data from massive projects, such as astronomy or genomics ("big data").
- Canada needs better data management policy to drive improvements in data access and long-term preservation, to guide researchers on how to manage and archive data, and to advocate for credit to data providers and support for infrastructure.
- Participants overwhelmingly consider open data to be a public good, and most people support sharing and free release into the public domain (without license conditions) as much as possible. Most participants feel like recent progress toward open data has been good, but the nuances of terms and timing of sharing data, including protecting sensitive data, particularly related to northern and Indigenous people and communities, need to be addressed and resolved so that our community can work together most effectively.

- Open data should be the default in all cases, because people want access to data as quickly as possible. Data producers should more seriously worry about data loss than about others stealing data and publishing them. By sharing data, we are saving data from permanent loss when data producers move or retire, when a computer fails, or other issues.
- If a data producer seeks to protect data, he/she should fully justify the exception and have the justification peer reviewed. The sharing policies and expectations should be defined in a formal data management plan for every project, required by the funding agencies, and funding for data management should be included in the project plan.
- Availability of funding, sustainability of archives and services, and data providers' and users' rights are viewed as the greatest challenges to polar data management. Interoperability of archives and data formats are viewed as the greatest technical challenges, but human interoperability, or the ability for all stakeholders in the Canadian polar data community to understand each other and work together effectively, is also considered to be a challenge. The success of this initiative will depend on improving human interoperability, especially regarding communication and collaboration related to participation and leadership by northern and Indigenous people in managing Canada's polar data.
- Long-term, sustainable funding for data management and for coordination of the polar data community is very constrained. The federal government should provide leadership to develop an effective long-term funding strategy. In the interim, we need to identify funding sources to support this coordination initiative so that we may be more efficient and productive in our work and support to polar research and monitoring in Canada.
- Participants seek a tangible, focused roadmap from the Workshop and related activities that can move polar data management in Canada forward and make a difference by creating a polar community to work together toward a common goal.
- The participants were extremely keen to collaborate and coordinate their activities on polar data management, to benefit Canada and to make our efforts more effective. The group developed a vision statement and agreed that a position paper would be drafted and circulated for community comment, then the final approved message, which would include a short list of policy and funding requests, should be taken in person to the pertinent decision makers.
- It is important to make sure all stakeholders are included and made to feel welcome in this effort, including Inuit and other northern Canadians. It is imperative to have all parties invited and engaged from the beginning as much as possible and to make it easy to join the effort. Special effort will need to be made to facilitate northern participation, keeping in mind the expense and logistics of traveling to and from northern Canada. Links should be made with international groups and with polar data communities in other countries, and the Antarctic community should not be overlooked.
- The Workshop participants agreed that we should form a committee, formal working group, or national secretariat to represent the polar data community in Canada. There may be a central coordinator or "polar secretary" to connect the different groups in Canada and in other countries along with an advisory committee formed of representatives from all sectors to provide guidance and leadership and to ensure proper communication throughout our large network.
- We must work with data contributors to find out what they need in terms of infrastructure, tools, and support for preparing and archiving data and with users to learn what they need

in terms of types of data, how data are presented and made available, and services for facilitating use and understanding of data.

- Our survey results suggest that the data most commonly managed by repositories are not the data most commonly used by people. Repositories and researchers need to recognize this and work together to archive and provide access to the data types that people use more often, to be as applicable as possible to users.
- Participants asked to continue the dialogue begun in the Workshop and are interested in joining another Workshop in 2016. CCIN/PDC staff have been in contact with Polar Knowledge Canada (POLAR, formed from the Canadian High Arctic Research Station and the Canadian Polar Commission in 2015) about co-hosting a 2016 Workshop, and funding support for annual meetings through 2020 has been included in a proposed federally-funded project related to sharing Arctic knowledge among stakeholders.

Specific activities were outlined for next steps:

- Draft the position paper and circulate to the community for comment prior to finalization and wider dissemination and advocacy activities. This should include options for consideration by the community regarding governance and activities for which the community and its advisory group would be responsible.
- Create a governance structure for the polar data community in Canada.
- Create an inventory of stakeholders (we can build on similar efforts of other organizations).
- Create a flow chart of the activities involved in polar data management in Canada, determine which organizations will work on the various parts of the process, and design a collaboration plan to facilitate efficient coordination between all participating stakeholder groups.
- Engage with fields and groups which already have good data management coordination and implement their best practices.
- Create an online communication forum.
- Plan a 2016 Canadian Polar Data Workshop to carry on the momentum of the 2015 meeting as well as the international Polar Data Forum II held in Waterloo in October 2015.

Many other specific activities and suggestions for coordination have been documented in the report below.

IV. Welcome, opening remarks, and expectations for the Workshop - Ellsworth LeDrew, CCIN/PDC Executive Director

Presentation - main points (slides in Appendix 8):

1. Good morning, and welcome!
2. Along with an invitation to this Workshop (Appendix 1), a *Background document for a Canadian Polar Data Workshop* (Appendix 2) on motivation for coordination of polar data management activities in Canada and Internationally, this Workshop, and the Polar Data Forum II was distributed to almost 200 people in mid-April. From the *Background document*:
 1. Canadian scientists and the public must have access to the data and information needed to facilitate their research activities and make informed decisions.
 2. National and global coordination of activities related to polar data management will benefit Canadian society by streamlining data management and access systems, making our work more time- and cost-efficient and putting data and information within easy reach of all stakeholders.
3. Our tasks during this Workshop:
 1. **To coordinate polar data management activities in Canada**
 1. Governance and management
 2. Interoperability: technical, human
 3. Long-term preservation of data and sustainability of archives
 4. Conversion of data to user-focused information and utilization in outreach
 2. **Prepare Canadian positions on issues in polar data management** in preparation for the second Polar Data Forum, the Tri-Agencies, and others.
4. An example of good data management is demonstrated, showing integration of a variety of datasets and effective display - *AirNow*, (<http://www.airnow.gov>) with air quality data from the US and Canada (managed by the EPA)
5. Logistics of the workshop - plans for breakout sessions today and tomorrow, note-takers and rapporteurs for sessions, using the microphone so that remote attendees can hear you, plans for group dinner tonight
 1. Participants, please let us know if you do not want to have your comments recorded or if you do not want to be photographed.
6. Thank you to all of our sponsors and participants in this initiative, including the Natural Sciences and Engineering Research Council of Canada (NSERC) which has contracted with us to report on the proceedings.

V. Review of Questionnaire Responses - Gabrielle Alix, PDC Data Manager

Presentation - summary (slides in Appendix 8):

1. An online questionnaire, on SurveyMonkey, was distributed in mid-April to individuals who responded with interest to the invitation for this Workshop (Appendix 1).
2. Goals of the questionnaire
 1. Gather information about the current landscape of polar data activities in Canada.
 2. Shape the agenda for this Workshop, based on participants' needs and interests.
3. Thirty-five people completed the questionnaire between 17 April and 25 May (see Appendix 3).

1. The process was very beneficial and we received a variety of perspectives and ideas.
2. Researchers and data repository staff comprised almost half of the respondents.
3. Few respondents identified themselves as northerners, but there were some northerners who identified themselves in other categories (e.g., researchers).
4. How do respondents use polar data?
 1. Majority use data for professional research, outreach and education, and professional data management activities.
 2. Three people use them for their daily activities, and seven use them for personal interest (research centre director, program managers, researchers, faculty and policy maker).
5. What category of polar data do you work with or require?
 1. Top three answers: Natural science, Satellite imagery, Ecological knowledge
 2. Traditional Knowledge (TK) data are required for about half of the respondents to do their work.
 3. Less frequently used categories - Political science, Health sciences, and Social sciences
 1. The same people use or require these three types of data - they are all working closely with Northerners and are policy makers, researchers, a research advisor, a faculty member, and a centre director.
 2. This is reasonable given that the three types of data are often sensitive and not as freely shared as natural science data.
6. What types of data do you **manage** or **use**?
 1. All nine categories of data were frequently managed and used, especially reports and publications.
 2. Top **used** data types - Scientific publications, written reports, natural history collections and associated data, and satellite imagery.
 3. Top **managed** data types - Raw data, maps derived from data, information derived from data, custom products derived from data, and processed data.
 1. People who manage data mostly manage raw data and derived information.
 4. **NOTE: There is no overlap between the top used data and the top managed data types!**
 1. Repositories and researchers need to recognize this and work together to archive and provide access to the data types that people use more often, to be as applicable as possible to users.
7. Why do you archive your data?
 1. We were surprised and happy to see that 94% of respondents said “data are a public good”! This is the most important reason why people archive their data.
 2. These results are encouraging and indicate that the culture is changing to be more favourable toward data sharing, ultimately leading to more collaboration.
 3. Lowest percentage responses - “access other people’s data” and “funding requirement” which was surprising as we anticipated a higher number.
 1. These responses came from researchers, a faculty member, students (Post-Doc and PhD), and data repository staff.
8. What are the technical challenges of polar data management?
 1. There are many challenges - **interoperability and data formats are viewed as the greatest challenges.**
 2. Moderate challenges - Long-term preservation and internet capacity and bandwidth.
 1. Internet capacity and bandwidth are good in southern Canada but are lacking in the North and on ships (particularly in the Arctic Ocean).
 3. Security was identified as the smallest challenge - this is surprising to us due to the significant security issues discussed in various fora that can threaten repositories.

1. Positive responses on security came mostly from data repository staff, researchers, policy makers, and a faculty member.
9. What are the ethical and political challenges of polar data management?
 1. **Long term funding comes in first with over 80% of the respondents saying this is an issue.**
 2. Providers' and users' rights are the next most popular answer with 77%.
 3. Almost all issues garnered over 50% response which means that most of the challenges listed are issues that will need to be addressed.
10. Which data repositories or archives do you use?
 1. Responses are located on a map - each star represent an archive or repository that was used, in North America, Europe, and Australia.
 2. Repositories listed by more than one respondent, in order of frequency of response:
 1. Polar Data Catalogue (most common response)
 2. National Snow and Ice Data Center, United States
 3. Environment Canada
 4. Arctic Science and Technology Information System, University of Calgary
 5. Canadian Geospatial Data Infrastructure
 6. Geomatics and Cartographic Research Centre, Carleton University
 7. National Oceanic and Atmospheric Administration, United States
 8. ICES Oceanographic Database
 9. Ocean Biogeographic Information System
 3. Access to international data is very important to respondents. It is assumed to be similarly important for other users and stakeholders.
11. What are the strengths and challenges of polar data management in Canada?
 1. Canadian polar data managers and systems have many strengths, but improvements are still required.
 2. Challenges have been covered in the questions before as well as in the six sub-topic presentations later this morning.
 3. All items in the table are viewed as strengths but many also have high levels of respondents indicating they are challenges.
 4. Top strengths
 1. IT and technology (challenge: It can be difficult to keep up with the latest technology)
 2. Data visualization (but we need more!). Users like displays of data and want more.
 5. Minor strengths
 1. Networking and partnerships (but we need more - this Workshop is a good opportunity to enhance those links!)
 2. Protection of sensitive data (but questions remain, so it is important to form strong policy that will help repositories protect data)
 3. Open data (but this is also seen as an ethical and political challenge)
 1. This pattern may be due to lack of a clear definition of open data - we should create a community definition of open data.
 4. Education and training (but again, we need more of this)
 1. People gave many suggestions for communication and outreach - meetings (like this one, plus larger conferences), newsletter, webinars, and others.
 1. We could use communication venues to present groups and organizations, share progress and highlight available datasets, learn about new initiatives and make connections, and provide resources, information, and technical help.

2. We need to share the functions and responsibilities for these activities among several participants.
5. Data portals (but archives are a challenge - we should strive to understand the difference and how we can address it).
6. Other strengths - these are elements that were not in the questionnaire list but emerged from other comments:
 1. Metadata management
 2. Ocean data
 3. Spatial data infrastructure
12. Final questions about coordinating Canadian polar data management and Canadian leadership on the international stage will be addressed next by Professor Warwick Vincent.

Discussion/questions from the audience:

1. Do you think the lack of use of social science data, as described in the questionnaire responses, is a reflection of the interests of the questionnaire participants rather than a generality?
 1. There were eight respondents who work in social sciences, but it is unclear what percentage of the polar research and data community these respondents represent.

VI. Goals for the Position Paper - Ellsworth LeDrew, CCIN/PDC Executive Director

Presentation - main points (slides in Appendix 8):

1. There are a number of prior summits and initiatives related to polar data and data management in general that can inform this workshop and the development of a position paper. Are there other references for previous work/documents on research data management, including Arctic or polar, which we should consult?
2. Audiences for the position paper are many and include the Tri-Agencies, the Canadian High Arctic Research Station (CHARS¹), the Canadian Foundation for Innovation (CFI), scientists, northerners, data portal managers, and others.
3. The process to develop the position paper started last year with meetings with key collaborators and stakeholders. Following the questionnaire and this workshop, we will write an initial draft and will request feedback. All perspectives, including dissenting opinions, will be documented.
4. The outcomes of this Workshop will form important input into Polar Data Forum II.
5. The idea of a second questionnaire has been raised, to augment the results from the first one and reach new audiences. We will discuss this more tomorrow afternoon.

Comments from the audience:

1. **We need a tangible, focused outcome or roadmap from this initiative that can move the polar data management agenda forward and make a difference by creating a polar community to work together toward a common goal.**
2. Who is our community? How many people are we talking about? It would be helpful to have a report which provides more information on the polar data management community:

¹ CHARS is now included in Polar Knowledge Canada, POLAR, which formed from combining CHARS and the Canadian Polar Commission in June 2015.

who are we, how many individuals and groups are involved, activities and future plans of each member.

1. The Geomatics and Cartographic Research Centre at Carleton University has done a scan of who is doing Arctic data management. This initiative listed eighty-five different groups which self-identify as creating data. However, it is not certain that we are actually a community due to our fractured work and relationships.
2. We should mine and report on who is getting funding for northern research and data management.
3. If a second questionnaire is conducted, send it to more than the 200 people from the original invitation list, if possible. The Tri-Agencies, ITK, and others could be surveyed for information on stakeholders and contributors to polar data management.
4. We should be careful about calling our plan a “position” as that can sometimes imply an official Canadian stance in the international sphere. We need to be sensitive to semantics and be clear that this is not the position of the country of Canada.

VII. Introduction to Workshop Question: *How can we coordinate polar data management in Canada and what opportunities should we pursue for Canadian leadership internationally?* Warwick Vincent, Scientific Director of Centre d'études nordiques (CEN), Université Laval

Presentation - main points (slides in Appendix 8):

1. Due to the recent focus on the importance of data and the formation of many relevant partnerships, the time is ripe to capitalize on opportunities for polar data management in Canada. I will list some pertinent programs and initiatives as examples and likely members in our work together.
2. ArcticNet Network of Centres of Excellence, 2004-2018
 1. Northerners want to know what is going on in their own backyard.
 2. Community consultations and outreach required for research permits are not necessarily reaching communities with data or information on research results.
 3. Inuit Research Advisors in each Inuit region are helping to bridge the gap.
 4. The ArcticNet community of young researchers is helping move the culture toward open data sharing.
3. Polar Data Catalogue (PDC)
 1. Initiated by ArcticNet, using metadata as a starting point.
 2. Has a geospatial focus because maps are the desired information sharing mechanism by northerners in Canada.
 3. Standards-based to facilitate interoperability and sharing with other polar data portals.
 4. With the addition of new partners, the PDC now houses data, along with metadata.
 5. In response to feedback from northerners who had difficulty with the PDC Geospatial Search application being slow due to limited bandwidth, the PDC Lite was created.
 6. The PDC has an oversight committee, the Polar Data Management Committee (PDMC), with representation from ArcticNet, northern organizations, government departments, and universities. You may wish to consider enhancing or extending this committee as a part of this current initiative.
4. Activities to coordinate data

1. Each research group and institute has its own data requirements, and strengths - the key is to connect these pieces together. Tools and advancements include:
 1. ArcticConnect - new suite of tools will be available soon
 2. *Nordicana-D* - journal for publishing datasets
 3. Journals are more frequently requiring publication of data.
 4. Inuit Knowledge Centre at Inuit Tapiriit Kanatami (ITK) is a model for data exchange and translation as required by users and stakeholders.
 5. Networks of stations and collaborators (e.g., CEN, Canadian Network of Northern Research Operators [CNNRO], International Network for Terrestrial Research and Monitoring in the Arctic [INTERACT], many others) are developing data policies and best practices, defining expectations and challenges for data management.
5. There is a need for data sharing across disciplines.
6. Increasingly, international players such as South Korea are interested in Arctic data collection and are even collecting in Canada. There is no formal mechanism in place by which we can access some of these data.
7. Questions that must be addressed regarding data coordination - these will be addressed in breakout groups over the next 1.5 days:
 1. Who should take the lead in this initiative, to improve Canadian data management for all stakeholders (northern communities, researchers, governments, industry, policy makers, public, etc.)?
 1. What is the role of CHARS?
 2. How can we connect and build on each other's initiatives?
 3. How would we fund this effort?
 4. How can researchers be encouraged to contribute?

Comments from the audience:

1. For International research, some data sharing is required through the Oceans Act for international researchers. Perhaps we should explore this as an example.
2. The oceanographic community has a culture of open data sharing in some sectors, but this may not be true in academic circles.
3. Efforts should be more closely aligned with the Arctic Spatial Data Infrastructure (ASDI) and the Canadian Geospatial Data Infrastructure.

VIII. Data Sharing - Peter Pulsifer, Research Scientist, US National Snow and Ice Data Center

Presentation - main points (slides in Appendix 8):

1. "Sharing" may have different meanings to different people, but the idea of sharing data being a "common good" is included in most definitions.
2. Sharing data openly is necessary for discovery, and data increase in value when shared and preserved. Data sharing supports reproducible research and is critical for transparency. The benefits of sharing and open data are generally considered to outweigh the risks.
3. The scientific community and society in general are in a transition, moving quickly from private data to sharing and open data, with a desired outcome of sharing being avoidance of conflict.
4. Governments are promoting the concept of open data. Organizations such as the International Arctic Science Committee (IASC) are releasing statements on data sharing which address the concept of open data while allowing ethically-legitimate restrictions.

5. There is competition in research to publish and be recognized. This is a sharing model, but it is one which is restrictive and is being challenged by newer, more open sharing models.
6. Trust is fundamental to sharing. Data sharing policies and clear information and expectations help to build trust, but trust also relies on human relationships and political and historical context. It takes time to earn trust.
7. Practical issues must be addressed in archiving data, such as which archive to use (if there are multiple appropriate options), how to maintain the archive, and how to build data management tools which are easy for both the data producers and data managers to use.
8. Questions for attendees (these questions may be addressed in the Data Sharing breakout session):
 - How can consideration for publication and credit to the data producer be reconciled with open and immediate sharing?
 - Many data producers expect an embargo period of exclusive access to data. How can we encourage early archiving? What timelines are realistic for the embargo period? Should sharing be immediate? Who decides on the timelines - the producers, funders, or users? Can the timeline be different for different types of data?
 - How do we reconcile the expectations of open data with the legal and ethical issues of providers' and users' rights or the potential for data sharing to cause harm?
 - How do we address considerations of intellectual property and "ownership" of data?
 - How do we promote trust?

Questions to be addressed during the breakout session:

1. The majority of participants indicated that "open data" is necessary and valuable. At the same time, there are notable exceptions to the fully open model (i.e., International Polar Year program Data Policy, IASC Data Statement which allow exceptions). What should the allowable exceptions to fully open data sharing be, and who should decide these exceptions?
2. In general, trust is a critical component of sharing data openly. What is required to establish trust between and among data producers, stewards, users, and other stakeholders?

Summary of Ideas Presented During the Data Sharing Breakout Session and Group Discussion

Chair: Peter Pulsifer

Rapporteur: Jeff Saarela

Participants: Marlene Doyle, Jesse Flowers, Ed Kennedy, Virginie Roy, Scott Tomlinson, Cameron Wilson

1. Open data

1. We should assume that open data is a good idea. Data should be collected and released, but there is substantial nuance in this, including consideration of timelines of release and legitimate permanent exceptions.
2. There are different kinds of data: scientific research data, monitoring data, potentially commercial data that may have proprietary restrictions, local and traditional knowledge, data from sensitive areas or species, others. Public domain is the desired outcome, but some types of data, like images, may have copyright holders who must agree to release the data to the public domain.
3. Data sharing expectations may be different when dealing with monitoring data, instead of research data. In general, there is no reason to not share monitoring data.

4. Monitoring data should be shared as quickly as possible - immediately/near real-time for sensor data, monthly or annually for analysed data.
5. To encourage open data sharing, it is important to communicate to researchers the value proposition in data sharing and explain what is in it for them.
6. Researchers and funders should be aware that there are costs to **not** making data open. If data are not shared or are not accompanied by high-quality metadata, they will be difficult to find. Sharing data increases the probability of wasted resources and time by duplicating data collection.
7. Where you work may be important to how you feel about the data you produce. For example, if you work for government and manage a long-term monitoring data project, you may not feel great ownership over the data. Your connection to the data may depend on your job function and what you are doing to generate the data.
8. Data sharing is perceived as a part of research, but data sharing can also happen with Elders who share knowledge and information with community members or researchers. In this case, it is the Elders who control how data are shared. The provider and the recipient of the sharing must both be defined, and the roles must be respected. This information may be designed and stated in a data management plan.
9. How shall we define data, information, knowledge, and analysis, and the expectations of sharing which go with each?
10. Do permits or licenses for research in northern Canada require open data? This should be explored and understood.

2. Exceptions to open data

1. There will always be exceptions to the general policy of open sharing, and they are context dependent. The exceptions make it difficult to develop a comprehensive data sharing policy that can be applied to all polar data sets.
2. Who should decide these exceptions to fully open data? Communities, individual researchers, journals, publishers, employers, funders, shareholders, and courts may all be involved.
3. The default should be fully open data, but if a data producer plans to deviate, he/she should justify the exception and have the justification peer reviewed. Data producers should not be allowed to just do whatever they like.
4. Currently, ethics procedures require that much social science data is destroyed. Do we, or does anyone, including ethics boards, have the authority to not destroy sensitive data?

3. Embargoes

1. Immediate access to data is what people, including northerners, want, but this is not happening due to a variety of reasons.
2. Some data can be made open immediately upon collection, such as most monitoring data, but other data sets will have different timelines for release.
3. Data may be placed into a repository but not made available until the embargo is over. This is good practice as it protects the data from loss (by archiving) as well as premature release (by not allowing public access).
4. The polar research data community needs to establish a reasonable expectation and maximum time for embargoes of certain types of data, such as 1 or 2 years.
5. Embargo periods should be shortened to reduce the risk of data loss.
6. If guidelines or requirements for embargo periods are enacted, then exceptions will need to be defined and approved.
7. When does the embargo clock start: When the data are collected, at the end of the project, or after publication? This should be clearly determined and stated.

8. The International Polar Year (IPY) data committee had a policy on embargo times. Should we set up the same kind of body which will set this kind of policy?
9. What are the pros and cons of embargoes? These should be explicitly stated.

4. Issues with sharing data

1. Some data are required to be shared, some even by law, but people are not sharing them. How may we work with data producers and others to address this?
2. The research agenda is driven by incentives in the research community such as publishing. This is a strong disincentive to releasing data too early. How can our community have any influence or control over this reality?
3. There is a sense that scooping (stealing data and publishing them as your own) does not happen very often in the real world. For example, a graduate student contributes her data to a large repository because her supervisor told her to do it. There is a perceived risk, but it may not be a very large risk at all.
4. People are part of different communities. The polar community supports fully open data, but other communities are not requiring fully open data. Some researchers are pulled by multiple sets of expectations.

5. Trust

1. It will take time to establish trust through mutually beneficial relationships.
2. Trust should be facilitated by funders and other stakeholders.
3. We should recognize the actions that are required to build trust and follow them.
 1. One such action is always providing data back to anyone who was promised they would receive it.
4. We should understand that it takes time to establish trust, and we should plan for it.
5. What is the mechanism for creating or facilitating trust? There should be a clear process resulting in written agreements and or statements of principles which stipulate the relationships between participating individuals and groups. Written agreements require and sometimes can engender a basic level of trust, but there is no magic formula to doing this. It took more than two years to get an international agreement for the ASDI signed, with much time spent building trust and relationships among countries and individuals. The level of work required in the polar community may be similar, due to the great variety in the size, level of experience, and sophistication of the players in our community.
6. We have a good chance of success if can get the key parties to the table to start the discussion and build trust. However, we need to know that people are prepared to provide the resources to work through a process for developing agreement.
7. When we talk about “community,” we are talking about the polar data community. However, there is also the community of Inuit in Canada who deal with trust all the time. Trust is complicated and difficult to establish but is very much a critical component. The path to establishing trust will be different depending on the types of data and different communities, and it must be based on relationships built over time. Unfortunately, information does not always get back to the community directly: Sometimes it happens indirectly, long after it should have happened, which erodes trust. Developing trust involves and depends on every stakeholder.

6. Examples related to data sharing

1. The Geogratis open data archive at Natural Resources Canada (NRCan, <http://www.geogratis.gc.ca>) was created because it is less expensive to have open data than to have and manage licenses for use. Also, different licenses complicate and hinder data interoperability.

2. For satellite data, different processing levels are required for different communities (level 0 raw data, level 1 processed data, etc.). Which data and which processing level is shared depends on the context and the community.
3. NRCan and MacDonald, Dettwiler and Associates Ltd. (MDA) partner to manage ground-receiving stations for satellite data. This is a public-private sector arrangement, part of which is a contract that limits what can be shared.
4. For ArcticNet, the embargo extends to three years after the end of the project. This allows time for appropriate analysis and quality control. Other polar programs in Canada, however, have embargo periods as short as one year.
5. In Yukon, there is a project to develop a flow of information for permits, licenses, and projects which presents information for decision making.
6. For the National Science Foundation in the United States, open data is a funding requirement. They have established the capacity and resources and provide archives for many communities.
 1. The NSF requires a mandatory two-page data management plan which is part of the proposal review process. The reviewer must determine if the budget supports the work proposed. Grants can be rejected if this is not justified. People have not felt victimized by this new requirement.
7. The geospatial community in Canada has gone through a five-year process of community building, developing a strategy and governance framework. The community is fragmented as well as overpopulated with organizations relative to its size. They have much in common, but all have discrete interests (cartography, surveying, science, industry, licencing) which need to be coordinated. A new organization, GeoAlliance Canada (<http://cgcr.ca/en/geoalliance-canada/>), represents all geospatial organizations in Canada, including corporations, universities, governments, and not-for-profit organizations.
 1. This process has led to success at building trust, which is key to making progress. It takes a long time, and it may be an example for us. We need a structure to work this out for the polar data community. Our community needs a champion, or a few champions.
 2. The process was supported by NRCan. We need funding to support the polar data community if we are to succeed in this effort.
 3. *A Case Study of the Canadian Geomatics Community Round Table* (<http://geoscan.nrcan.gc.ca/starweb/geoscan/servlet.starweb?path=geoscan/fulle.web&search1=R=296396>) may be an informative document.
8. In human health research, data must be taken to the community before release to the general public. Although this is different from the normal academic research process, it is the policy of the Northern Contaminants Program and perhaps other programs, too. This sharing is very important for the communities. Communities need more recognition as stakeholders in the research process.
 1. For health research, sometimes the community does not want the data to be published. These concerns must be considered and respected.
 2. This requires extra time to publish and may impact students' timing of graduation.
 3. As a community, we can help provide information on this, to make sure people know when this type of consultation is required and how it may impact a project, or a student's career.
9. What other communities are doing well in sharing data? The ocean research community, astronomy, genomics. We can learn from these and others which may have more similar sharing issues to the polar community.

7. Policy

1. How can we encourage the willingness to share? The sharing process needs to be defined early in the process of designing the research study, as an expectation and requirement.
2. The data management details for each dataset should rely on a formal data management plan that is part of the research proposal, and exceptions to sharing should be stated in the metadata.
3. If data licencing terms are required, they must be very clear.
4. The community itself may not be able to enforce firm policies about openness - the willingness and ability to share depends on community cultures, funding requirements, employer requirements, and others.
5. Some people are not optimistic about a community-wide consensus on a carrot or stick policy that applies to the entire polar data community. What may be best is documentation of existing and future required policies.
6. Journals and publishers are starting to require data to be archived. Once this is more common, it will force authors to archive their data, and this will quickly lead to a culture change.
7. Policy on intellectual property rights must be resolved before we can have clear definitions or expectations for sharing.
 1. Researchers operate in a world of pooled funding, e.g., NSERC and industry may fund a single project. In this case, who owns the intellectual property in the data?

IX. Data Access - Benoît Pirenne, Director of User Engagement, Ocean Networks Canada

Presentation - main points (slides in Appendix 8):

1. Data access can be defined in terms of expected functionality and qualities of the access system, including data, metadata, portals, sharing services, users, and other components.
2. Everybody has data, but few have well-managed data, and even fewer provide access to their data.
3. Creating a data access system can be confusing and difficult, but one way to start is to design the system according to user-defined requirements for access:
 1. Users want to be able to find data using online portals, select what is interesting, and download relevant data in a usable format.
 1. Users want data access and download to be fast. Access to data in Northern Canada is often slowed due to limited Internet capacity and speed.
 2. Users want access to real-time data whenever possible.
 3. Users want access to international data, not just Canadian data.
 4. Users would like to know whether data are of high quality. Archives should ideally provide quality indicators so users know whether to trust the data.
 2. Portals should be reliable, available, and intuitive, with good user interface design.
 3. If analysis functions or interactive access are available, that is even better.
4. There are a plethora of data portals, with more being built all the time. Should we have a centralized, national facility which would simplify access? This ties in to data preservation and interoperability.

5. If a user downloads data, can they or should they be expected to upload products or analyses which they make from the downloaded data?
6. When discussing open data, remember that a “proprietary period of access” and “privacy protection” are two different aspects.
 1. Perhaps there should be no proprietary period at all for data that do not require privacy protection (no embargoes).
 2. Privacy protection will almost always apply more to social science data than to natural science data.
7. Data access depends on good metadata which describes data observation conditions, data quality assessments, and other characteristics. Data without adequate metadata are not useful and are therefore not accessible.
 1. Metadata should be public, even if the data themselves are not.
 2. Should there be metadata quality enforcement/acceptance criteria for funding agencies?
8. All topics addressed during this Workshop (sharing, access, preservation, interoperability, partnerships and linkages, and policy and funding) are tightly coupled. Decisions in one will influence the others.

Questions to be addressed during the breakout session:

1. What does “Open Data” mean to you?
2. Who are the users, and what are their needs for data access? Is everything covered? What is important?

Summary of Ideas Presented During the Data Access Breakout Session and Group Discussion

Chair: Benoît Pirenne

Rapporteurs: Laura Petrunka and Karen Lauer

Participants: Shannon Asencio, Mathieu Ouellet, Ellsworth LeDrew

1. Open access

1. In the questionnaire responses, numerous people indicated that Canada has an obligation to provide easy access to polar data. It is the responsibility of our community to determine how to do that, and then to do it.
2. We must improve awareness by increasing education about the benefits and risks of open data. This will address some of the concerns of data producers and will encourage data sharing.
3. “Open data” means freely available, where the information is put online for anyone to use, without any proprietary period or restrictions.
4. Instead of worrying about others stealing your data and publishing it, it is perhaps more important to worry about data being lost, especially when data producers move or retire or when a computer fails. By sharing data, we are saving data from permanent loss.
5. Some people will want to have data ownership. This will be an impediment to open data.

2. Portals

1. Having a centralized location for data may make access easier, but having more than one portal is useful if the different portals provide different views to the same data or different ways to access the data.

2. Different portals can be adapted to specific audiences and designed for specific uses or specific types of data, and customized applications can be built which provide different ways to access data.
3. The Polar Data Catalogue and many other portals in Canada are working on this, to provide different types of access to different communities and users.
4. Too many portals can be confusing for users, so portals should be linked so that data are available across portals.
5. What kinds of portals and tools should be built? Who should build them?

3. Data reuse

1. Providing access does not guarantee reuse of data. Reuse only occurs if others find value in the data, if the access process is easy, and if others have adequate funding support to explore the shared data.
2. This community should focus more on analysis and reuse, in addition to simplifying the access process. We need to go from data to knowledge!
3. Universities and the private sector should be a part of improving access and facilitating data reuse, to add value to the data. As an example, the Group on Earth Observations asked for “user engagement proposals” to encourage reuse of shared data.
4. Download tracking is helpful for funding, as it provides information on what data are being reused.
5. For data which are not publicly available but are instead available only via password, even temporarily, systems can be designed to track who downloads the data and provide this information to the data producer. This may be welcomed, especially by the social science community.
6. In environmental monitoring, progress may be made by archiving and making raw datasets available, then converting datasets into a consistent format which may be easily reused.
7. Enforcing use of controlled vocabularies/keywords facilitates discovery and reuse (and interoperability).

4. User requirements for data

1. The end user must be defined. There are many different users and communities.
2. For geographic data, it is important to always relate the data to time and space. This is the concept of a data cube, and it facilitates searching and discovery of data.
3. Being able to view and download sub-sets of large datasets is very valuable.
4. Raw data are usually the data that are shared, but products derived from raw data may also be useful and should be considered as a valuable dataset which should be made available.
5. The community would like to have access to Canadian-based data that are collected by other countries. For example, there is a Korean ice breaker that has been operating in Canadian waters. How shall we obtain access to these types of data?
6. Reports and publications are very commonly used, perhaps even more than the data itself, but this may be changing as data are more available.
7. We need to serve the needs of the data producers and users and make data access as simple as possible for both groups.

5. Data licenses

1. The Canadian Open Government Action Plan has an open data license. Should the polar community follow this?
 1. The license allows a user to modify the data. There was some concern with this as it may cause data producers to not want to archive their data in a system which allows modification of contributed data. However, the government’s understanding of

modification of data may be different from how data producers view it. This should be explored to ensure that data are respected and that data producers are comfortable with the terms of the open data license.

2. It is important to make sure that the terms of use or license go along with the data. How can this be done?
3. There are many open data licenses which have differences which actually render them non-interoperable.
4. There is a move in the data world towards no licenses at all - all data would be available in the public domain. This would dramatically simplify archiving and access.

6. Limiting access

1. Protection of sites and species, including humans, may justify limiting access to data.
2. Raw data are provided only if privacy protection is not violated.
3. Data may be protected at the researcher/producer level (not archived and only accessible to the producer), at the portal level (archived but only available to the producer and data repository staff), or at the community level (archived but available only via password).
4. Data may be archived without providing open access. The data may remain hidden temporarily (until the end of an embargo) or permanently (for sensitive data). Is it likely that data producers will archive data that they do not want released either temporarily or permanently? Is limiting data access acceptable? Is it useful to archive data without making them available?

7. Metadata and documentation

1. Documentation is required for data to be useful. Metadata should be publicly available, even if the data are not. This provides a record of the existence of the data.
2. There are different types and levels of metadata. Descriptive metadata (ISO 19115 or FGDC standard format) are critical for discoverability, but there are additional levels of metadata which give further details and aid in reuse.
3. Documentation should include some indication of quality.

8. Policy

1. The community needs to formulate policies related to data access. This will take substantial effort.

X. Data Preservation - Scott Tomlinson, Environmental Scientist, Northern Science and Contaminants Research Directorate, Aboriginal Affairs and Northern Development Canada²

Presentation - main points (slides in Appendix 8):

1. Preservation protects data against loss, ideally through reliable and redundant storage systems.
2. We focus here on preservation of digital content, defined by Wikipedia as a formal endeavour to ensure that digital information of continuing value remains accessible and usable, through planning, resource allocation, and application of policies, strategies, and actions to ensure access and accurate rendering of authenticated content over time.

² AANDC is now Indigenous and Northern Affairs Canada, since November 2015.

3. Preservation should be facilitated through a national effort to create and support long-term archives and repositories.
4. Data producers and users should have confidence that the repositories will endure via long-term and stable support, so that users can rely on their existence and producers will feel it is worth the effort to archive their data.
5. Old data can be extremely valuable for climate and environmental change observations. Non-digital content such as paper records, books, and photographs should be located, identified, and rescued by digitizing them and making them available in public portals.
6. Data quality and documentation are critical and are primarily the responsibility of the data producers (researchers).
7. In the case that repositories are performing quality control for certain types of data, data producers and future users must have confidence that control processes are not introducing errors and are well documented.
8. Data should be provided in internationally-accepted formats when possible, to promote long-term access as well as interoperability.
9. Questions for attendees:
 1. Who should lead the effort to coordinate polar data rescue and preservation?
 2. What data should be rescued first? How do we prioritize data for rescue?
 3. Who should pay for data rescue?
 4. What should be expected of researchers with respect to data quality and documentation?
 5. Strict adherence to accepted formats may hinder data archiving. To what degree should producers be expected to provide data in specific formats?

Questions to be addressed during the breakout session:

1. Data documentation is a critical part of the preservation system. What are the best methods, including enforcement or encouragement, to ensure this essential element of the data life-cycle is completed?
2. Data rescue has been a hot topic in recent years. What are the key challenges to overcome when addressing the issue of data rescue? What triage methods can or should be used to prioritize datasets at risk of being lost?

Summary of Ideas Presented During the Data Preservation Breakout Session and Group Discussion

Chair: Scott Tomlinson

Rapporteur: Paula Tozer

Participants: Gabrielle Alix, Claire Austin, Elizabeth Griffin, Helen Halbert, Laura Petrunka

1. Motivations and challenges of preservation

1. Motivations for preservation include community expectations, funding requirements (return on investment), producer credit and citation (formal or personal, for a sense of well-being by contributing to the greater good), extra publications, and assurance that data will live forever.
2. This community should consider the actions which we can take to simplify archiving, to encourage preservation by streamlining the process for the data producer.
 1. We should use or design specialized applications or tools for data preservation, such as a special software stack for running data through the process. The tool should have processes which check for fixity (avoiding bit rot), produce archival packages

for incoming data, check formats every few years to ensure files are still accessible, and other characteristics.

3. The special challenge for preservation of polar data is the variety of our data. Archiving varied data (the “long tail of data”) has different challenges than archiving data from massive projects, such as astronomy or genomics.
4. Communities such as astronomy have a strong record of data management, but it is easier to manage astronomy data because there are fewer types. However, we can learn a lesson from astronomy in particular in that they make plans for data management and preservation even before the telescope is built! We should be similarly proactive.
5. There are many disciplinary portals available, but some communities do not have repositories. There must be a home for data sets before they can be archived, so focus should be on identifying gaps in availability of appropriate repositories for all polar researchers in Canada.
 1. Policy and infrastructure must go hand in hand to make it easy for researchers to comply.
6. For social science research, it is critical to ensure that consent forms are completed properly. If there are mistakes on the form, this can result in loss/destruction of data.
7. Security of archives can be a challenge, to ensure that data are safe for the long term and will not be compromised or corrupted by intrusion.
8. Much social science data simply can not be preserved, due to privacy considerations and other restrictions. Is there any way to save these data, instead of destroying them?
 1. Perhaps we need to think about social science data differently, to acknowledge that the “data” are already being “archived” in the publications and reports that are coming out of the work.
9. The issue of preserving physical samples must be addressed, if not by this community, then by someone appropriate.

2. Data rescue

1. We should, as a polar data community, set out policies for identifying data at risk and prioritizing data for rescue.
2. The idea of data rescue applies to Traditional Knowledge (TK) as well as digital data. Transfer of knowledge from Elders to youth is critical.
3. It may be difficult to provide complete documentation or assess the quality of rescued data. This may help determine whether a particular dataset should be rescued or not.
4. Some data may not be worth rescuing. Criteria for abandoning data should be documented in policy.
5. Data rescue success stories should be identified and highlighted for outreach as well as to justify funding rescue activities.
6. Climate data from the Polar Continental Shelf Program have been compiled and made available on the University of Ottawa website. These data may be considered high priority and should be entered into a permanent repository to ensure long-term access.
7. Related to rescue is data mining, to make archived data more useful.

3. Documentation

1. Documentation should be a mandatory component of data management policies and plans.
2. At the same time, the value of documentation should be emphasized to the data producer.
3. Documentation should itself be of high quality so that future users can fully understand the history of the data.

4. README files are very useful to provide documentation which goes beyond the initial discovery metadata, but standardized machine-readable metadata is even more useful. Requiring a standardized README file may be a good start to obtaining the information that is needed, rather than allowing the README to be completely free-form. The polar data community should define the contents of a standard README file to facilitate conversion to machine-readable metadata in the long-term.
5. It may be useful to have some fields that are not controlled, to allow additional comments and text that are useful for using the data. This information may be targeted in a keyword search.

4. Data formats

1. When translating information into different languages, the original records in the original language should also be preserved, since translation can result in a change of meaning.
 1. Alternatively, if source data may lose meaning when translated or converted to a new format or a different language, then it may be preferable to leave data in its own original format or language.
2. Data sets that support research publications should be archived in the format that supports the published research results.
3. Who should be responsible for data format conversion? Should it be done by the data producer or by the data managers or other repository staff? Should there be a separate function for data format conversion and long-term preservation?
 1. Optimal solution would be for producer to provide data in non-proprietary formats for long-term archiving, but if that can not be done, then the repository should do the conversion and confirm accuracy.
4. Who will pay for conversion, including development of tools and personnel time?
5. Proprietary software usually outputs data files in proprietary formats. This is discouraged for preservation as it hinders future access to data. It is preferable to output and preserve non-proprietary formats for data files if at all possible, along with the original proprietary files, or to provide software for opening the proprietary data files.

XI. Interoperability - Julie Friddell, Associate Director, CCIN/PDC

Presentation - main points (slides in Appendix 8):

1. Interoperability was identified in the questionnaires as one of the primary challenges for polar data management in Canada.
2. Interoperability may be defined in terms of *technical* as well as *human* and *legal* interoperability.
3. Technical interoperability is defined as linking portals for metadata sharing and data discoverability.
 1. Impediments to technical interoperability include the variety of standards and formats in use for both metadata and data, the lack of use of controlled vocabularies/keyword libraries in metadata, and the variety of sharing protocols available for both metadata and data.
4. Human interoperability is hindered by the lack of interaction between data producers and repositories and the fact that different people have different experiences and views on sharing data, thus it is sometimes difficult to work together due to a lack of familiarity and trust.

5. As with other topics discussed today, limited funding impacts our ability to achieve interoperability.
6. Benefits of interoperability include more visibility for Canadian data, enhanced knowledge of and access to Canadian and international data, reduction of the burden of archiving on the data producer, and reduction of duplication of data.
7. Interoperability can lead to new collaborations with Canadian and international polar data initiatives, including possibilities for co-hosting or ingesting other nations' data.
8. Interoperability can help link data for large-scale studies by providing online tools where researchers can upload and combine their data with others' data.
9. Online data management tools are required to aid interoperability. There are many tools which may be used, such as brokering to aggregate metadata and web services to share data files and metadata. The largest challenge is deciding which tools we want to use.
10. Manual creation of interoperable links is labour intensive. Can we make a "Google for data" to automate interoperability? This should be explored.
11. Canada is contributing to polar data interoperability through a variety of data portals and archives, the Canadian federal geospatial platform and data infrastructure, Research Data Canada (RDC) and other initiatives, and expertise and leadership in metadata formats and web services.
12. Canada is also a leader in making interoperable links with international data portals and organizations. We should use these connections to enhance and strengthen the polar data community in Canada.

Questions to be addressed during the breakout sessions:

1. What are the challenges for interoperability (technical, human, legal)?
2. What specific actions can your group/Canada take to improve linkages between polar data portals within Canada and internationally?

Summary of Ideas Presented During the Interoperability Breakout Sessions and Group Discussions

Note: Since interoperability ranked as one of the top challenges for polar data management, there were two breakout sessions for this topic. The discussions and notes from both sessions have been combined below.

Chairs: Julie Friddell (1) and Shannon Vossepoel (2)

Rapporteurs: Jennifer Sokol (1) and Shannon Vossepoel and Jeff Saarela (2)

Participants: Claire Austin, Tulio Crescuolo, Shirley Mills, Jenn Parrott, Robert Way (1) and Colline Gombault, Mathieu Ouellet (2)

1. Definition of interoperability

1. Interoperability must be clearly defined, to facilitate clear communication. This community should establish definitions of human, legal, and technical interoperability.
2. Technical interoperability may be defined as the ability to interact with data sets in a flexible way.
3. There are different levels of interoperability, from bare minimum to optimal/maximal. It may be important to consider what level to strive for in different situations.
4. Even though it may not be considered true interoperability, providing a link to an external source enhances discoverability and thus meets the goals of interoperability.

5. Both confederated searching for discoverability of data and true interoperability of portals and data are important.

2. Benefits of interoperability

1. Open data and interoperable systems simplify provision of information to various interested parties. For example, rather than having to answer the same question about a project or dataset from multiple people or provide data multiple times, the information can be prepared once, then subsequently provided around Canada and the world for discovery and use.
2. If a data producer can contribute data and metadata to one repository which then disseminates those resources to other polar data portals or systems which also require archiving of the target data, that saves work for the researcher.
3. If researchers can discover and access other data through effective interoperable networks, they do not have to collect those data for themselves. This saves money and is a huge motivation for interoperability.

3. Challenges and barriers to interoperability

1. Controlled vocabularies and standard keyword libraries are required for true metadata interoperability. Use of different keyword libraries/controlled vocabularies is a huge impediment to metadata sharing which can only be solved by manually equating keywords in different libraries. This is labour-intensive and is currently one of the largest issues in metadata interoperability.
2. Interoperability may be hindered if some groups do not have the ability to use or implement favoured or required technology. The community should be sensitive to the fact that not all organizations have the resources or capacity to be interoperable.
3. In the government, there are many barriers to interoperability such as firewall rules, security clearances, organizational policies, and others. These will result in delays or complete prevention of making interoperability a reality.
4. Sharing data is more complicated than sharing metadata: There are many formats and many possible sharing protocols.
5. It is difficult for different types of data (for example, bibliographic data and metadata which do not have a location vs. geospatial metadata and data) to be technically interoperable, just due to the nature of the data.

4. Requirements and expectations of interoperability

1. Interoperability relies on repositories having the expertise and support for setting up and maintaining the interoperable links.
2. Providing data and metadata via different portals requires data sharing, usually through web services, from the portal which holds the master copy.
3. The ideal interoperable system would have the following characteristics:
 1. Ingest information from the beginning of a project (preferably at the time of the proposal) and disseminate to other parts of the system, as appropriate, so that it does not have to be entered multiple times.
 2. Assign a Digital Object Identifier (DOI) to the project upon entry of data into the central project database, to enhance citation and credit to the data provider.
 3. Assign additional DOIs as additional metadata and data files are entered for complicated projects which produce a variety of datasets.
4. If we do create distributed networks for sharing data and metadata, we should ensure that there is still one portal which is always clearly responsible for the data, so that we do not find in the future that critical datasets are gone because no one person was looking after them.

5. We need to ensure protection of sensitive data when setting up interoperable links. It is imperative to not provide access to private data by mistake.
6. Location/geography is a means of linking data, but political, social, and other linkages also need to be made.

5. Human interoperability

1. Factors in human interoperability include personalities, jobs/mandates, backgrounds, experiences, agendas, goals, etc.
 1. One example of potentially conflicting mandates is provincial and territorial jurisdictions which have different systems and different feelings about their responsibilities.
2. We want to communicate, and we often want the same things, but if we do not use the same words, we may have difficulty understanding each other and finding common ground.
 1. Development of a common language about polar data management will facilitate human interoperability. This can only be accomplished by talking with and understanding each other.
3. There is some concern with the concept of a “one-stop shop” for polar data in Canada. This does not mean, however, that a variety of archives/portals should not be linked.
 1. For example, there are many different organizations and jurisdictions which want to steward their own data.
 2. There also may be other issues with centralizing data management. For example, making a central group responsible may take responsibility away from the researchers, who are critical to the process.
4. Challenges of human interoperability may be addressed via targeted partnerships, especially with regard to getting TK to researchers.
5. Better technology to keep in touch with each other allows us to increase capacity by sharing resources and group knowledge. This helps us to become humanly interoperable.

6. Choosing standards

1. Technical interoperability is not suffering from a lack of standards but from clarity on which standard to choose. There are standards for metadata, for data files, for transfer protocols, and many more.
2. Currently, our community does not have a set of recommended standards and guiding principles to follow.
3. Many groups, however, are already making choices of which standards to use.
4. There are many standards available. It is important to think about what is available and what is currently in use. We do not have to create a new “polar” standard. Use available standards as much as possible, and do new work only to fill actual gaps.
5. What level of standardization is required for true interoperability? Can we ensure effective discoverability and access without extensive standardization?
6. To simplify the work, perhaps we should think in terms of transfer formats instead of context transfer.
7. Popular standards change frequently, so we need to choose standards which are not losing popularity.
8. How is a standard defined? Is it defined by which information or fields are required or allowed?
9. When standards are followed, it is important to document which standard is being followed and the level of compliance with the standard.

10. Standards themselves may evolve, making it necessary to accommodate requirements for compatibility in the future. When the portal or the metadata or data standards change, the new system should be designed in such a way that users are not impacted and old ways of linking or accessing are left intact.
11. If a portal or repository does not require a particular standard, users should archive their data using existing standards and should try to conform to standards which are common within their field of research.

7. Examples of existing standards

1. Internationally, OGC (the Open Geospatial Consortium) is used for maps, and ISO 19115 is used for metadata. They have not changed much over the years and are adaptable.
2. ISO 19115 is quite extensive and complicated to use, so we may benefit from a definition of minimum elements which should be used to facilitate interoperability.
3. FGDC-STD-001-1998 (Federal Geographic Data Committee) and GCMD/DIF (Global Change Master Directory/Directory Interchange Format) are also popular metadata standards.
4. Darwin Core is the usual standard for biodiversity data.
5. Shapefiles and NetCDF files are not standards, but they are ubiquitous and could be considered “standardized” file formats. Sometimes standards are not decided from the top but are instead simply what everyone is using.

8. Proposed activities

1. We should create a central, comprehensive database of publicly funded polar research projects which are producing data or a summary of the research (where, when, who?) so that northern community members can know what is happening near them.
 1. Who can provide leadership on creating this database?
 1. ITK may be able to guide development of the system to facilitate use by Inuit.
 2. The Canadian Polar Commission’s (CPC³) research project map website may be useful.
 2. Research licenses are a huge potential source of information for this. The licenses may be linked to the data.
 1. Research license granting agencies may update their license application forms to require information on where data will be archived.
 2. If license or project information can be imported automatically into a central system, it would be easier than manually entering information.
 3. All sources of polar data in Canada should be included in this database.
 3. Perhaps we could plan an app-a-thon on creating a polar research project database!
2. We should recommend standards, formats, protocols, and policies which would facilitate interoperability of both metadata and data, without complicating data deposit and curation.
 1. We should produce or contribute to a list(s) of relevant standards that currently exist.
 2. It would be preferable to inventory the community and choose standards and formats which are already in use. If we choose a standard because it appears to be the best one, but it is not in common use, there may be lower adoption rates.
 1. We should consider NRCan’s geospatial standards and operational policies when deciding on standards to recommend to the community.
 2. Libraries may provide useful guidance in choosing standards.

³ CPC and CHARS became Polar Knowledge Canada in June 2015.

3. The private sector, including OGC, should be involved in deciding on standards for polar data and metadata.
3. We should capture in the documentation the users' intent for sharing the data. Would it be possible to augment the standard with this type of information?
4. We should draft a brief document which lays out recommendations for data and metadata standards to the community, and give people a chance to respond to the recommendations and perhaps vote on them.
 1. If we pursue this path, it will be critical to have a formalized process to ensure everyone's engagement.
3. We should bring together key portals and archives to implement recommended standards to facilitate metadata interoperability.
 1. If we start with the large groups or key portals which have the most experience, then other groups can be brought in afterwards as they have the chance to follow the standards.
4. We should link groups which have different levels of readiness for interoperability, to share knowledge and enhance progress.
5. We should provide "smart searching" for keywords in our portals, to facilitate discovery.
6. We should continue the efforts of linking repositories within Canada and expanding international links, too.
7. We should link data to the publications and articles which report on the data.
8. We should investigate options for automating creation of interoperable links.
9. Can we build the equivalent of Google search for data? What would this look like? Data files would need to be automatically available online and tagged as a data resource.
10. Interoperability is complex to coordinate. Perhaps we should implement an iterative approach of baby steps to reach the goal.
11. We should look to successful case studies to guide our activities:
 1. The Aurora Research Institute (ARI) may be a good example for setting up an integrated project database system.
 2. The Australian Antarctic Division (AAD) may be a good example for setting up a comprehensive data management system.
 1. AAD has an excellent system for automating management of project information, metadata, and data. All projects must have a data management plan, and creation and submission of metadata and data are incremental, with deadlines starting at the very beginning of a funded project and continuing throughout the project's life.
 2. This type of structure makes it easier on researchers and encourages their compliance by providing infrastructure and support to researchers and very clearly laying out expectations.
12. We need to develop a plan to foster and maintain linkages and clear communication between groups, to improve human interoperability.
13. Can we make an app which publishes data?
- 9. Existing Canadian and international activities which may contribute to the proposed tasks above**
 1. Canada is leading the IASC/SAON (Sustaining Arctic Observing Networks) Arctic Data Committee's metadata elements project to define a minimum set of metadata elements required for effective metadata interoperability.
 1. It will be important for the Canadian polar data community to adopt the recommendations from the metadata elements project, once they are published.

2. Canada has a large variety of portals and archives and is connected to polar data portals in other countries through interoperable links.
 1. The Polar Data Catalogue (PDC) metadata are currently being harvested by other polar data portals, and PDC is harvesting metadata from other portals.
 2. PDC and the Arctic Science and Technology Information System (ASTIS) have been discussing options for linking data with publications.
3. GeoGratis at NRCan provides publications, datasets, and maps as a complete package.
 1. This may not be ideal, however, for real-time, time-sensitive, or time-variable data.
4. The Northwest Territories and Nunavut provide basic information about projects online. The information is collected as part of the research permit system.
5. ARI has a well-established system for issuing research licenses.
6. ASTIS at the Arctic Institute of North America (AINA) is working on a licensing and permit record system. These include research project descriptions and translations of the licenses.
7. The CPC had a website of research projects on a map. The project has slowed down recently as it requires a significant effort to maintain.
8. NRCan publishes recommendations on geospatial standards and operational policies (<http://www.nrcan.gc.ca/earth-sciences/geomatics/canadas-spatial-data-infrastructure/8902>).
9. Research Data Canada's Standards and Interoperability Committee (RDC - SINC) works on standards and interoperability for all data, not just polar. They are in contact with the international Research Data Alliance (RDA) to advocate for generic and flexible use of standards. SINC welcomes new participants.
10. Research Data Canada has produced a list of relevant standards (available on the RDC website).
11. The Scott Polar Research Institute in the UK has a list of Arctic data producers, but it is not comprehensive.

10. Working together to achieve interoperability

1. We need to develop a plan to keep the conversation going so that human interoperability may improve and our enhanced collaboration can result in increased access to polar data.
 1. We should develop a plan to foster linkages between groups that need tools or resources and organizations which already have built tools or have resources.
 1. For example, Canadensys has an integrated publishing toolkit, developed by the Global Biodiversity Information Facility, that can be used by others to publish biodiversity occurrence data.
2. We should, as a community, share resources and technological capabilities.
 1. It will be necessary for organizations with more experience to provide guidance and assistance to groups which are just starting out.
3. The discussion of standards is difficult but we should work as a community on group agreement of where we want to go together. If we choose a path, we can slowly shift our implementations to be compliant with the chosen standards in the long term, which will facilitate interoperability.
 1. We need a formalized process to identify and choose standards which involves the entire community. This is part of building trust. People may not embrace the new standard if they were not involved at the beginning.
 1. Any recommended standards should be flexible and simple to implement, to enhance compliance. The approach to define or recommend standards for our

community may not work if groups are not able to follow or implement the recommended standards.

2. Adopting a minimum set of metadata elements to be applied to all metadata records may be valuable and should facilitate interoperability. However, it may be difficult to find a common set of elements to support a wide variety of stakeholders.
3. Several participants eagerly indicated they would be willing to use and support a standard once it was chosen by and for the community.
4. If we can not achieve the desired interoperability in the short term, can we link, mirror, or host other datasets to enhance access? This will require trust and cooperation.

XII. Partnerships and Linkages - David Arthurs, Managing Director, Polar View Earth Observation Limited

Presentation - main points (slides in Appendix 8):

1. We need to develop a framework for communication between participants in this effort.
2. There are too many organizations to list them all, so perhaps we should focus on “communities.”
3. There are many ways to define “community”:
 1. As different communities of data managers (polar, geospatial, scientific),
 2. By sector (citizens, government, academia, industry),
 3. By type of data (in situ, cultural, socio-economic, earth observation),
 4. By research interest (marine, atmosphere, land),
 5. By geography (Arctic, Antarctic, nations, international),
 6. Or in other ways.
4. These groups overlap in many ways, and one person may be a member of numerous communities. People and communities work together to collaborate and learn from each other.
5. Canada is only one facet of international data. In the polar regions, nationality is less important than circumpolar linkages.
6. Communication issues result from a variety of logistical difficulties:
 1. There seem to be more organizations than people!
 2. Communities may do a good job of working internally but not at interacting with other communities.
 3. Technical barriers such as standards get in the way of working together.
 4. Social barriers such as languages or cultures can result in a lack of common ways of communicating.
 5. Requirements for privacy of data or information can cause difficulties in even talking about the data or activities required for stewardship.
 6. Limited funding limits resources available for these activities.
7. Many communication pathways are informal. Formalization of communication channels may improve the links.
8. Questions:
 1. What communities should be involved in the “Canadian polar data community”?
 2. How can communication be facilitated between communities and, by extension, with organizations?

3. What does each community or organization have that can be shared, and how can it be shared?
4. What does each community or organization need?

Questions to be addressed during the breakout session:

1. There are too many organizations to establish linkages with every one. Which communities, organizations, groups are a priority for engaging?
 1. Should we consider influence, goals, impact, potential for success, or other criteria in answering this question?
2. How shall we link with industry?

Summary of Ideas Presented During the Partnerships and Linkages Breakout Session and Group Discussion

Chair: David Arthurs

Rapporteur: Ed Kennedy

Participants: Jesse Flowers, Jenn Parrott, Jennifer Sokol, Fraser Taylor

1. High priority groups

1. The polar data community in Canada needs to ensure the inclusion of Inuit and northerners. They are integral to decisions on data management in Canada, and their participation is fundamental to designing and implementing a successful leadership and coordination model.
 1. Target organizations include ITK, the National Metis Council, and the Assembly of First Nations.
 2. Northern Canadians should be represented not only as stakeholders in polar data management but also as producers of data.
2. Intra-Canadian linkages are very important. Past linkages and effective interactions have been limited, but now that this group is forming, there is more potential for effective partnership.
 1. We should engage with CNNRO.
 2. The Canadian Association of Research Libraries (CARL) is currently running a federated pilot on long-term data preservation.
 3. Canada's Department of Fisheries and Oceans has a strong research group in British Columbia to which we should link.
 4. Canada is not heavily involved in the Antarctic (funding for research in Antarctica is very limited), but we need to remember that the Antarctic is part of this process.
 1. We should make sure that CCAR, the Canadian Committee on Antarctic Research, is included.
3. IASC should be an international partner.
 1. The IASC Arctic Data Committee, with the input of the Arctic countries, is creating an international ontology of data centres, institutes, policies, metadata catalogs, and other components of polar data. This will be a semantic database that provides information on how to connect to data, for example, through web services. Effort is required within Canada to ensure that the polar data management system is well represented.
4. The Inuit Circumpolar Council (ICC) is a key international organization. It is a permanent member of the Arctic Council.

1. ITK links with the Arctic Council through ICC (ICC sits on the ITK Board) and is forming links with IASC.
5. OGC deals with open standards which facilitate interoperability and, additionally, links with most of the spatial data community's major players, including industry.
6. The CODATA Data at Risk Working Group will be important for data preservation.
7. RDA is involved in the Belmont Forum which has multimillion-dollar funding for Arctic research.
8. Libraries, including the Polar Libraries Colloquy, university libraries, and the federal library and archives, may inform our work. We need to study what libraries do well so that we can benefit from or join their activities.
9. Links to the United States are important.
 1. The US National Snow and Ice Data Center (NSIDC) is a successful case study for good polar data management.
 2. We should engage with the National Science Foundation.
 3. Since the Arctic Council chairmanship has been handed from Canada to the US, there is more awareness in the US about Canadian activity in the Arctic, particularly the focus on local residents, so there is an opportunity to build on that awareness.
 4. Under the US Arctic Council chairmanship, the US State Department is a new player. There is some political discussion between the US and the other strong Arctic nations about activities in and concerning the Arctic.
 5. Changes in the science funding models in the US present some new opportunities, with more focus on diversification and international partnerships, presenting opportunities for the Canadian polar data community to play a stronger role in helping drive the agenda.
10. Links should also be made with Google Scholar
- 2. International linkages**
 1. International partnerships are very important, but they must be done carefully. We must make sure we have the capacity to sustain the relationships once they are made - we do not want to reach out and then be unavailable.
 1. Receptor capacity in Canada is limited. Effective international engagement requires a minimum level of funding, facilities to attract partners, and recognized experts who are aware of the international situation and can work with others to generate world-leading knowledge.
 2. Some international links are successful, some are not, and some which should be made have not been made yet.
 1. The Inuvialuit Regional Corporation's Joint Secretariat has international linkages.
 2. NRCan is formally linked internationally with the Arctic Council and working groups such as the ASDI.
 3. How well positioned is Canada to maintain a dominant role in the Arctic, given other countries' activities?
 1. Canada has a strong presence and reputation as a leader in the Arctic, and we are well respected in polar data management. We can make a great contribution if this expertise is built upon.
 2. Japan has openly stated their interest in Arctic resources and is spending a lot of money on Arctic science and technology. Japan should work with Canada.
- 3. Industry**
 1. For industry linkages, both user and supplier communities should be considered.

2. The World Ocean Council includes industry and is active in polar science initiatives, funding scientific meetings, organizing proposals, and coordinating funding support for the Belmont Forum.
 3. The International Oil and Gas Association would be a good partner for this effort.
 1. We should reach out to individual oil companies. The Alaska office was very engaged at the ELOKA (Exchange for Local Observations and Knowledge of the Arctic) workshop a few years ago. The ConocoPhillips Calgary office is interested in forming relationships with northern communities.
 4. The International Association of Antarctica Tour Operators (IAATO) and Association of Arctic Expedition Cruise Operators (AECO) run ecotourism trips.
- 4. Examples of partnerships related to polar data**
1. The Inuit atlas project helps Inuit communities spatially enable records from local projects.
 2. The Beaufort Sea Online Platform supports contributions from 53 organizations involved in the Beaufort (e.g., industry, community groups, government, etc.) and includes a partnership with the Alaska Ocean Observing System.
 3. The Polar Geospatial Center at the University of Minnesota provides data to US scientists and has recently partnered with Google to help task Arctic imaging satellites with capturing imagery over communities of interest. Google has interest in longer-term location-based services (LBS) to these communities.
 1. The remote sensing community is mostly involved in provision rather than use of polar data.
 2. O'Reilly Group's *Where 2.0* is a forum for interfacing with the LBS industry. LBS involves a much broader community than traditional GIS.
- 5. Role of CHARS**
1. CHARS is trying to leverage polar science and polar knowledge.
 1. CHARS is receiving a lot of interest from the international community with respect to research in Canada's north.
 2. CHARS has international connections such as polar science groups in the European Union and individual countries (e.g., UK, China, and the Korean Polar Research Institute) and other countries that are establishing Arctic observatories in Canada.
 2. CHARS' partners in Canada include the scientific community and northern Canadians.
 1. These groups are working together to coordinate dissemination of information to northerners.
 3. A new organization, Polar Knowledge Canada (POLAR), was created 1 June 2015 to replace CHARS and the CPC. POLAR's three main programs will be (a) the research station, (b) northern science and technology, and (c) knowledge management (outreach, partnerships, and knowledge dissemination).
 1. David Scott has been named President of POLAR, and the Board of Directors, Steering Committee, and working groups are being formed.
 2. POLAR can act as a broker for linking international scientists with Canadian scientists or agencies and can broker a role to connect them with specific programs in Canada.
 3. Future linkages between POLAR and INAC and Environment Canada are not known yet.
 4. POLAR has facilities to share with the international community, but they need to set policies now for data management and other topics to ensure benefit to Canada.

XIII. Policy and Funding - Philippe-Olivier Giroux, Policy Analyst, Natural Science and Engineering Research Council

Presentation - main points (slides in Appendix 8):

1. There are two topics for this presentation: (a) What policies are needed, and (b) what are the needs for funding? The presentation today reports on the responses to the questionnaire and does not necessarily represent the views of NSERC.
2. Policy
 1. We need better data policy in Canada to drive improvements in access and long-term preservation and to guide researchers on how to manage and archive data.
 2. What happens to the data when short-term projects end? We need to avoid data loss.
 3. Labs and researchers need resources to manage data. Should this cost be integrated into grant proposals? Should data management be required in proposals?
 4. Graduate and undergraduate program curricula should include data stewardship.
 5. Data providers need to have professional recognition and credit for publication of data, which should be recognized as a legitimate scholarly contribution.
 6. What should be the role of each stakeholder, including universities, government departments, and funding agencies, in developing and implementing policy for managing polar data in Canada?
3. Funding
 1. Long-term, sustainable funding is very constrained.
 2. The federal government should provide leadership to develop an effective long-term funding strategy.
 3. We should expand the scope of this coordination initiative to include other groups that can provide tools, resources, and funding for our efforts.
 4. We need funding to engage with international groups, so that the expertise and data of Canada are known and shared around the world.
 5. Funding applicants should ensure that the cost of data management is properly represented in proposals.
 6. Where there are common interests, the polar data community should work together to develop joint proposals to fund data management activities.
 7. Who needs funding, to support what activities and expenses? We need to establish a long-term plan which includes tasks and estimated costs for polar data management in Canada.
4. Current work related to policy and funding for data management in Canada
 1. The Tri-Agencies⁴ are working on data policy. This group will be helpful to the Tri-Agency effort by helping them understand the needs of the polar data community.
 2. The Open Government Action Plan includes a focus on open science and open data.
 3. A 2014 report, *Seizing Canada's Moment: Moving Forward in Science, Technology and Innovation*, provides information on big data and open science.
 4. There is a variety of groups (including the Leadership Council for Digital Infrastructure, Universities Canada, and CANARIE, among others previously mentioned) dealing with data issues with whom this group should be linked.

⁴ NSERC (the Natural Sciences and Engineering Research Council of Canada), SSHRC (the Social Sciences and Humanities Research Council of Canada), and CIHR (the Canadian Institutes of Health Research). These are the primary scientific research funding bodies in Canada.

Questions to be addressed during the breakout session:

1. What policies are needed? Who should be responsible for them?
2. What are the explicit needs for funding? Who needs funding, for what activities, and what types of expenses?

Summary of Ideas Presented During the Policy and Funding Breakout Sessions (2) and Group Discussions

Note: Since policy and funding were identified as primary challenges for polar data management, there were two breakout sessions for this topic. The discussions and notes from both sessions have been combined below.

Chair: Philippe-Olivier Giroux (1) and Scot Nickels (2)

Rapporteur: David Arthurs (1) and Karen Lauer (2)

Participants: Gabrielle Alix, Colline Gombault, Amos Hayes, Dave Moorman, Fraser Taylor, Scott Tomlinson, Shannon Vossepoel, Warwick Vincent (1) and Julie Friddell, Ellsworth LeDrew, Simon Riopel (2)

1. Cross-linking comments

1. Requirements for new policy and funding models are being driven by the change in culture to new attitudes about data sharing and expectations of instant data access. Both researchers and the public are growing more amenable to sharing. This new perspective applies to all data, not just polar. It is important to build on this momentum.
2. For both policy and funding, we should build on what is already available. There are so many groups working on data, and they have already addressed many of the issues and questions that we face.
3. At the federal level, the Assistant Deputy Ministers of the science and technology agencies (S&T ADMs) should make a plan together for data policy and funding. This happened during IPY and was successful.
4. Northerners are fatigued with researchers coming with numerous requests for help and information. There are benefits to all the attention on the north, and the northern people benefit by more information being available, but it is also a burden. We need to make sure we clearly define the Canadian process for northern research before people from other countries start coming to northern Canada to do their research.
5. Because of limited funding, our scientists are not driving the polar science agenda, especially in Antarctica, and Canadian interests are not being served.
 1. There is great concern that we may become a landlord in the Arctic but international partners with more money will drive the science agenda. If they do not like Canadian policies and rules, then those investors will go to other Arctic nations. While the location of CHARS can provide some control on what happens in the Arctic, science funding is a more significant factor.
 2. NRCAN is trying to set up an international receiving centre for earth observation in Inuvik, but because Canadian funding is limited, others will drive agenda and Canada will obtain limited benefits.

2. Policy

1. Policy development is not just for funding agencies. Universities, international organizations, and others will also need to be involved.

2. Policy, infrastructure, and monitoring/enforcement of compliance must all come together for policy to be effective.
 1. An example of successful data policy was IPY. The IPY policy was simple and clear: it required data management planning and archiving, and both infrastructure and support were available to back up the policy.
 1. This success story should be more clearly communicated to funding agencies and others who could support future polar data management activities.
 1. Sometimes it is difficult because there is no one group to communicate to, and few people or organizations have the power to make changes.
 2. If we have a clear, simple message, then our probability of success with targeted communication is higher.
 3. We need to make sure we include all stakeholders when planning policies, especially when it comes to making policy which may impact international projects or partnerships.
 1. Industry Canada is forming a new working group on big data.
 4. Make sure that those who have data policies (e.g., Tri-Agencies, CHARS, IPY, etc.) coordinate to ensure that their policies align as much as possible.
 5. Federal agencies need to look internationally for guidance on developing policies, should take the lead for the country, and should make policy and enforce it.
 1. The Tri-Agencies are very interested in using the outcomes of this Workshop and national consultation to advance their data policy work.
 2. CHARS is looking for direction and advice on forming their policies.
 3. The Northern Contaminants Program (NCP) may be a good case study of how to make policy, of being flexible while enforcing requirements.
 4. If the polar community makes policy itself and tries to impose it on others, then others are likely not to follow.
 6. There could be problems if you try to centralize data management. The onus should be partially on the researchers.
 7. When is it better to use encouragement (“carrots”), and when is it better to use requirements (“sticks”)?
 1. The Tri-Agencies will probably need to enforce any data policy they produce (“stick”).
 2. However, it is critical to provide support to data producers to help them provide the data (“carrot”).
 3. Some people do not want to share or collaborate, and they do not even provide enough information or assistance to write metadata for knowing about their data. Some researchers will not respond to e-mail or answer their telephone. What can be done about that?
 8. This will not be an easy process, but we should be more afraid of inaction on data policy than of moving forward, even if we are uncertain.
- 3. Specific policies**
1. Having a clear and enforced data policy for specific research initiatives or projects is critical to the effective stewardship of the resulting data.
 2. Users of data need to be providers of data, too: Once someone downloads data, if they make changes to them, they should be expected to provide the data back to the archive.
 3. Funding applications should require data management plans, and review of new proposals should include an evaluation of the applicant’s past data management efforts.
 4. A set amount of funding (to be defined) should be used to support data management, either explicitly in each proposal or as a percentage of the national research budget.

5. We need a linked funding program, like the Northern Supplements, which provides targeted funding for data management for polar research proposals. Everybody should use it, but perhaps it could be voluntary at first.
 1. SSHRC already has a supplemental pool of funding for data management. Others should copy the SSHRC program or provide a “global” fund to which anyone could apply.
6. We need to make sure we have Canadian data in a Canadian archive, even if it is collected by other countries.
 1. One issue with this is to make sure that Canada has the capacity to deal with and make use of the data, once they are collected and deposited: If we do not have the correct researchers who have time to engage with the incoming data, then what use are the data?
 2. We need to be effective at reaching out to international groups which are working and doing research in Canada. Think strategically about making sure that Canadian knowledge, data, and expertise benefit Canada and are not just taken back to other countries.
 3. For example, for a foreign ship sailing through the Northwest Passage, should we require as a part of their license to sail that they deposit the data in a Canadian archive?
7. Metadata about all datasets relative to the Arctic could be registered in the Polar Data Catalogue or searchable in the PDC via interoperable links when the data are hosted elsewhere.
8. NCP plans to implement a data policy soon. Recommendations from this group can inform the NCP data policy.
9. We should have some meetings in northern Canada, to more successfully involve northerners.
10. Server farms should be in cold places, to take advantage of natural cooling. Facebook in Sweden has been successful and may be used as a good case study.
 1. Is Compute Canada environmentally friendly?

4. Funding

1. Support should be in the form of long-term, ongoing, baseline funding, as opposed to the short-term contracts that currently support most Canadian data activities.
2. The need for “resources” may be a better way to look at the question of funding.
 1. Resources do not only equal funding - there are many other resources which are available, such as people with expertise, infrastructure, etc.
 1. How can we make sure that these resources are leveraged optimally to accomplish our goals?
 2. We should make use of the resources already at our disposal, but we will need new resources for connectivity and forming links.
3. We need to consider funding with respect to the entire research lifecycle. There are many dimensions and levels of responsibility, with universities receiving funding from the provinces, but research is mostly funded by the federal government, and there are hidden costs such as support for buildings and infrastructure.
4. Given the size and complexity of our community, how shall we divide the funding and give it to the right people to get the work done, but minimize the effort to distribute the money?
 1. We need to make sure the researcher and the user do not have to pay for each service individually. This applies to all aspects (repositories, data management offices at universities, training, etc.).

5. We need new funding models and should consider innovative funding sources. The ones we have are not sufficient to facilitate collaboration and connection.
 1. As an example, CFI has a multi-source model for funding projects: 40% from the province, 40% CFI, and 20% from other sources. This makes CFI proposals and projects quite complicated, though, which is not efficient.
 2. Should we consider commercialization or monetization of data?
- 5. What activities are required for successful data management? How much do they cost?**
 1. Explicit funding needs
 1. Physical infrastructure, including repositories and networks
 2. Support of the entire data lifecycle: ingestion, archiving, maintenance, user interface, outreach, online access and analysis tools
 3. Funds for the researcher, to prepare data to get it into the repository
 1. A major disincentive is the time, cost, and effort to make data open. There are not enough incentives or support for researchers to prepare the data properly for sharing. Funding should be provided for making the data open.
 4. Long-term stewardship of data from short-term projects
 1. For example, IPY is over, but we are still ingesting data from IPY projects, and we are expected to preserve and provide them “forever,” theoretically for free (since IPY funding has ended).
 5. Data management offices at universities, to spread the responsibilities by supporting both researchers and archives. This is starting to happen.
 6. Education, on why data management is important
 7. Training, on proper data management procedures
 8. Travel for northerners, plus salary for time to participate (because many northerners do not have jobs which will support them to attend meetings)
 9. Open access publications
 10. Data rescue
 11. A national coordinator or secretariat for polar data management in Canada
 12. Managing northern community knowledge
 2. We should conduct a needs assessment from all community members, including northern communities, to determine additional needs and capacities.
 1. The private sector is driving some of the requirements for data needs, in addition to researchers and the public.
 2. Canadian funding of the Belmont Forum is very limited. There is so little Canadian money that potential Canadian participation in these exciting breakthrough technology programs is limited.

XIV. Coordination and Leadership - all participants

The following questions were given to the participants:

1. How can we coordinate polar data management activities in Canada?
 1. How can we connect and build on the various initiatives and programs (and meet all expectations)?
 2. Who could/should take the lead?
 3. How would such coordination activities be funded and supported?

4. How can researchers be encouraged to contribute to data management and coordination activities?
2. How can we improve Canadian data management for all stakeholders: northern communities, researchers, industry, governments, policy-makers, the public, and others?
3. What opportunities shall we pursue for Canadian leadership internationally?

All breakout groups addressed the questions above independently during their dedicated sessions. Below is a compilation of the responses from the individual breakout groups along with points raised during the subsequent full group discussion:

1. Why are we not coordinated?

1. There are many different groups doing polar data work, with funding coming from different areas. There is a lack of knowledge of data activities going on in different groups, and there are many conversations between various groups. This leads to duplication of effort, fractured progress, detrimental competition, a lack of clarity in organizational roles, and gaps in meeting needs.
2. This “bottom up” approach leads to individual groups trying to coordinate with others, instead of the entire community working together.
3. There is currently no coordinating body to set priorities and define the focus for polar data management activities in Canada.

2. Do we want to coordinate our community?

1. Yes. We need to organize ourselves and coordinate our activities to bring the desired benefits to Canada and success to our work.
2. To be effective for the long term, we will need a formalized structure with clear guidelines for reporting and responsibilities.
3. We need a theme and set of guidelines which will take us in the same direction.
4. We need overall coordination to provide outreach and awareness, both among ourselves as well as to others.
5. We need to determine who we are (who is involved in polar data management in Canada), our strategic objectives, the scope of our issues, and a long-term plan for where we want to be in 5-10 years.
6. We need to put data management at the forefront of polar research in Canada.

3. Ok, so we want to self-organize. What will be the end goals of this coordination?

1. We need a message, outlining the vision, mission, objectives, future goals, strategy, scope, and needs of the polar data community in Canada, which will be taken to our target audience. This should include a short list of policy and funding requests from our community.
 1. Once we define the goals, then we can break the tasks down into small steps.
2. We need to identify our target audience(s) - who should receive the message?
 1. We need to make sure we take the message to the right people and speak to them directly and personally, to increase understanding of our work and needs.
 2. The list should include the S&T ADMs and the research granting councils, but we may need to subsequently take the message even higher in government (ministers, etc.).
 3. We should go to university presidents, vice presidents of research, territorial and provincial governments, other levels of government, Inuit, other northern groups, the private sector, and others.
 4. Often, the people who have the power to make policy and provide funding do not have a very good understanding of the issue, and vice versa.

5. We need to coordinate with key stake holders to take information and feed it to other groups.
3. We need to identify our spokesperson(s) - who takes the message?
 1. There may be different spokespeople for different recipients/audiences.
4. We need to ensure that we have a group voice and that the message represents the polar data community in Canada.
 1. We need to involve all members of the community, as much as we can, and ensure that all groups are able to contribute to the internal and external communications as well as coordination of our activities together.
 2. We need to ensure that the views of all participants are represented and that there is support and buy-in from all members of our community.
 3. What we do must be identified as Canadian - of Canada, and about Canada.
5. We need to build trust and credibility, both within our community and with others.
 1. This includes support to researchers, so that we make it easier for them to manage their data.
6. We shall define the functions of the new organization:
 1. By creating a community of practice for polar data management, we shall establish best practices and guidelines which we should ourselves follow.
 2. We shall create a list of organizational activities and responsibilities and determine who is best fit to lead specific tasks.
 3. We shall choose a coordinator to strengthen connections within the community (including with researchers) and act as a broker for projects and activities. The coordinator may be a member of the community or someone outside the ecosystem of the group.

4. The Message

1. Background:

1. Changes in the polar regions compel us to record observations of all types so that change can be tracked and understood.
2. Research data in general, and polar data in particular, are difficult and expensive to produce and are thus very valuable.
3. We acknowledge that data have traditionally not been shared in Canada and have been used largely for publication.
4. Polar data are unique due to the diversity of types which include traditional environmental observations such as weather and sea ice records but also information on human health and wildlife, physical samples, oral records from Indigenous Elders, and many more. These data are valuable due to their diversity and the effort and expense incurred in collecting them.
5. At the same time, the polar data community has many commonalities with other research data communities, including challenges of developing policy and infrastructure for managing the data being produced.
6. In Canada, numerous groups are researching and managing data for the polar region, so coordination between groups is needed. Resources need to be pooled together, and a forum for communicating is required to improve knowledge of others' activities.

2. Vision:

1. Recognizing that Canada has great potential and an obligation for sharing polar data with the rest of the world, the Canadian polar data community will work together and with others to develop policies and systems for treating research data as a national resource, including polar data as a subset.

2. To promote the emerging national vision for sharing and preserving data, the polar data community in Canada will be a leader and will work with other Canadian and international groups to improve polar data management in Canada, contribute to improved data management across disciplines, and strengthen the international data community. This will increase transparency and inclusion and will reduce duplication of effort, including research as well as data management activities.
 1. We will organize our community and leverage individual resources to coordinate management of the entire data life cycle, from the genesis of the research idea, to planning for data management at the proposal stage, generating and preserving metadata and data, discovering and publishing knowledge, and rescuing data at risk of loss.
 2. We will ensure that stakeholders, including researchers, northern communities, governments at all levels, industry, and the public, have access to the data and resources that are required to meet the final goal, which is gaining knowledge from the data.
 3. We will ensure preservation of the data and unhindered availability on the Internet so that they and their knowledge are accessible for future generations.
 4. We will ensure that the data are accompanied by proper documentation and provenance information to guarantee that they are usable and understandable.
 5. We will ensure that sensitive data are protected and that producers' and users' rights are respected.
3. **Mission:**
 1. To strengthen the connections between the groups involved in polar data management in Canada as well as with groups in other countries.
4. **Strategy:**
 1. Polar change is a popular topic now. We should capitalize on the attention that the Arctic is receiving to take our message to receptive audiences while the interest is high.
5. **Goal:**
 1. To be effective leaders in data management.
6. **Objective:**
 1. Long-term management of polar information, primarily in the Arctic but also encompassing other polar areas, and coordination to ensure preservation.
7. **Benefits of participation** - why should busy people join this effort?
 1. Increased access to polar data for all stakeholders, including contributors who also need access to others' data.
 1. Streamlining the data stewardship life cycle simplifies the work of data management organizations, accelerates progress in academic research, promotes business, provides critical information to northern communities and decision makers at all levels of society, and informs the public.
 2. Formal data management is increasingly expected of researchers. Coordination of data-related tasks simplifies the process and takes work off data contributors by providing clear and unified instructions on how they should manage their data and providing resources and repositories for their use.
8. **Activities:**
 1. It will be our responsibility to do the work to achieve progress.
 2. We should establish best practices and guidelines for data management by others as well as for ourselves.

3. We should collaborate with researchers, northern communities, and decision makers to increase the information they can receive from the data, including participating in the formulation of research questions and projects and in data collection activities.
4. We should provide data support services to scientists, to simplify their work and make data management easier for them.
5. We need to enhance the preservation of data, to guarantee its future accessibility.
- 9. Value proposition:**
 1. Our work connects to government, industry, university, research, and northern community priorities. How will our efforts contribute to help them?
 2. The message is not only to add value to research and discovery but also how to enhance the value provided by cooperation between a variety of Canadian organizations.
- 10. Future plans:**
 1. Where will we be/where do we want to be in 5-10 years?
- 11. Scope:**
 1. Management of data and information created in Canada, but also data about Canada's polar region created outside of the nation.
 2. Advisory function only, not operational.
- 12. Title:**
 1. Should specify that it is Canadian, but perhaps First Nations and Inuit would prefer to not have Canada in the title as this may be considered to not be inclusive
 2. "Canadian Polar Data Advisory Council"?
 3. Should use "information" instead of "data"?
 4. Perhaps "Polar Information Management" or "Polar Knowledge Management" or "Polar Data Management Network" may be better titles?
- 5. Potential models for coordination**
 1. The most important aspect is for all groups to be involved. The group must be inclusive and accessible to all who are interested, including not only academia and government but also northern communities and industry. This may not be easy to accomplish, but it is critical to ensure inclusion and effectiveness.
 1. To ensure inclusion of Inuit and First Nations communities and voices, we may want to have dedicated seats on any committee or board and not just rely on the outcomes of an election. We need to project the message of inclusivity from the start.
 1. We may want to do the same with industry representation.
 2. Meetings and other communication venues must be easily accessible to all members.
 3. We should endeavour to engage people and organizations who may not be supportive of data management, to educate and inform as many people as possible of the benefits.
 2. Preference seems to be for creating a committee, formal working group, or national secretariat to represent the polar data community in Canada.
 1. There may be a central coordinator or "polar secretary," or perhaps two people to share the job (some other organizations have multi-person Secretariats), to connect the different groups in Canada and in other countries.
 2. There should be a national advisory committee or management council, as with IASC and the Scientific Committee on Antarctic Research (SCAR), to provide credibility to the effort.

1. This advisory committee may either provide direction to the secretariat or actually do the work. We will need to decide on the role of the advisory committee.
2. It may be possible to have two groups, one advisory and one operational.
3. There is some concern with providing an operational role as it may not be advantageous to have external people deciding how everyone should manage their data.
3. To be successful, this group will have to be relevant and useful and will need to be able to maintain the momentum and the connections.
4. The network will feed information to the secretariat which will liaise with the international community and share expertise. The secretariat will then distribute information back to the network.
5. The group will need to be made of representatives from all sectors, including government, university, and private members, and representatives who can speak for proper stewardship of TK.
 1. Group should perhaps have representatives from each major organization that manages polar data.
 2. Tri-Agencies would need to be on the advisory committee. They should be involved in creation of data management policy for the country, hopefully being able to use the polar community as a successful example.
6. Resources will be needed to support the group's work.
7. Such a group would need to define a clear governance structure, but the structure may need to evolve over time as roles become more clear over time.
8. Concerns:
 1. Forming a new committee. We should build on existing committees if at all possible.
 2. Secretariat being formed within the federal government (for example, if CHARS or another federal body were to lead this), as it would be subject to government mandate and may not be able to effectively manage competing priorities or respond to the needs of all community members. The Open Government initiative has not been a good example of coordination.
 3. Secretariat being formed at a university, as there may be issues of competition between groups and institutions as to who becomes the host, as this could influence where funding goes.
 4. One person or group which speaks for all - this may not work well. We may need a truly distributed model.
9. Should this committee or secretariat be given the authority to coordinate polar research in Canada so that data management could be coordinated and mandated uniformly? This could be very useful but would take a great deal of work to arrange.
10. We need an "honest broker" who can effectively and fairly speak for all.
3. Local or regional chapters which can get together more easily than a full national organization, so that there are multiple levels of governance.
4. Other options
 1. An institute that handles polar data itself
5. Examples of good data management
 1. The Australian system, which has built the national data management policy and infrastructure in increments.
 1. This requires coordination as well as funding to create the organization and system.
6. Examples of successful coordination efforts

1. The Canadian Federal Geospatial Platform, in developing a North American standard for geospatial data
 1. The group held a series of meetings with stakeholders from various countries. It took five years to accomplish the original goal!
 2. Progress was made by developing recommendations and providing options to the community to consider and select which ones to implement.
2. The Canadian Advisory Council for Remote Sensing began as a grass roots movement with community input. It had a flat structure and successful function.
3. The Alberta Library consortium
 1. They make joint purchases and have one person who coordinates the group, but smaller groups are tasked with performing specific functions.
4. DataOne in the US holds annual showcases of their activities. We could attend their events and learn from them.
5. The World Wide Web Consortium (W3C) - how is it coordinated?
7. Challenges
 1. It may be difficult to coordinate between the different groups engaged in polar data activities in Canada, based on the different goals and interests.
 2. Polar data management is complicated by the variety of disciplines, geographic regions, and types of data.

6. Potential activities

1. Create a chart of the activities involved in the different steps of the data life cycle, from research project formulation to planning for data collection, to sharing, publication, preservation, and data rescue.
2. Determine which organizations will work on the various parts, taking into account individual preferences and expertise and aiming to meet the needs of all contributors and users, from novice to expert.
 1. It will be important to use tools and portals which already exist and not re-create system components which have already been built.
3. A governance structure and organizational policies or charter would need to be drafted in a form which is acceptable to all.
4. Ask one or two people to research potential governance models, including the ones cited above, and provide suggestions and options for our community.
 1. Could the private sector lead the coordination function, as they have a vested interest in having access to the data?
5. Demonstrate the value proposition and make the case for resources.
 1. Create an inventory of requests for data and information across portals to establish the need for data access and services.
 2. Develop case studies to demonstrate the value of increased data access to present successes in research, policy development, industry, northern communities, and the public.
6. Reach out to fields which already have good data management coordination, and learn and implement their successful actions.
7. Make a list of groups which have recently had meetings like the Canadian Polar Data Workshop or who have written position papers on related topics, and talk with the people who led these efforts. These people are the ones who want to get things done, and we should engage with them to hear the recommendations and actions from their activities so that we can work together to coordinate in Canada and internationally.
8. Create an inventory of stakeholders, so we will know who should be involved, the expertise of each group, and who we are missing.

1. Others have been working on this already. The Arctic Data Committee map of organizations may be a good place to start. Data producers should also be included in the inventory.
2. Use the semantic web framework and visualization capacity to build a concept and geolocation map of the entities of the polar data ecosystem.
9. Assess the needs of data users and contributors, including tools, data and information requirements, data-related services, infrastructure, connections within the network, etc.
10. Work with all stakeholders in Canada to create and unify data policies, to simplify data management for researchers and programs.
11. Secretariat/committee would organize teleconferences, webinars, meetings, and workshops of the community, maintain a list of stakeholders, write position papers on our activities and goals, and advocate for support, best practices, and optimal coordination within our community.
12. Create an online communication forum, such as a LinkedIn group for polar data management, in which all stakeholders and interested parties could participate.
13. Organize projects and funding for the community.
14. Arrange for support for data management assistance, ideally as an individual or team of people who have been identified and can work with researchers or other data contributors to understand the requirements and help them prepare their data for archiving.
15. Work with journal publishers to advocate that authors archive the data supporting their published results (as increasingly required by publishers such as PLOS ONE, etc.). This will encourage the culture change toward open data and sharing.
16. Produce a guide on how to navigate the Canadian polar research system, including northern research licenses.
17. Produce a guide on how to effectively participate in the polar data management community in Canada.
18. Seek additional input from the community as we move forward.
- 7. Who would be involved?**
 1. We need to incorporate, leverage, and build on existing relationships.
 2. Regional Inuit must be present and engaged so that they bring their ideas and needs as well as take the message back to northern communities.
 3. First Nations must also be involved.
 4. We should partner with Arctic College. They are teaching a course on the research process.
 5. The role of the PDC, PDMC, ARI, Canadian Polar Data Network, POLAR/CHARS, etc. will need to be defined.
 6. We need to coordinate with national and international groups such as IASC and SCAR, to link to, learn from, and contribute to the needs and activities of the polar data community at the international level.
 1. The list of international stakeholders is very large.
 2. The Polar Data Forum II, in Waterloo, Ontario, will be an excellent opportunity for international and Canadian interaction.
 7. We should determine if we can work with international or industry groups who have collected Canadian data, such as the Great Lakes Commission, to obtain data for archiving.
 8. Stakeholders in Canada include government of all levels, the Tri-Agencies, academia, non-governmental organizations, and others.
 1. NRCan should be engaged.

9. Deputy Ministers and the ADM Board should be involved.
10. Libraries have long experience in preserving information and other critical aspects of data management. Hiring librarians into the data management world will help ensure good data management.
11. Research Data Canada engages with many stakeholders and is very interested in coordinating data management in Canada. They may help disseminate information as well as facilitate development of a national infrastructure and policy for data management.
12. What is the role of the private sector? Could they be involved in providing resources, data, or in driving the coordination function, because they have a vested interest in having access to the data? Could the community provide specialized data products for private sector use? Would private industry be interested in working on data preservation?

8. Involvement of Northerners

1. There is a full ecosystem of regional and community Inuit and First Nations organizations concerned with TK and data and ensuring they are well represented.
2. The First Nations have developed the principles of ownership, control, access, and possession (OCAP). These should be respected in funding agency agreements. OCAP may also be useful for Inuit and Metis.
3. Proper care of TK enhances northern communities and researchers.
4. Northern agencies such as ARI and the Nunavut Research Institute should define the metadata and other information that they need for research licenses.
 1. Each region currently handles this differently. Collaboration may simplify the process for everyone.
 2. The licensing process should obtain the information that the community requires for making sound decisions.
 3. The license should be linked to outputs and results of the licensed research project.
 4. A suggestion was made that ongoing research may only require one license, instead of annual renewals, to simplify the process.
 5. Licenses may be used to enforce data management and community involvement in the process of data archiving and stewardship, including feedback and involvement in the community, so that community members benefit from direct access to data and understanding of the research results, rather than just a research article which may not be informative.
 6. Licenses could require data management plans. Compliance with the plan should be required before granting of subsequent licenses.
5. Communities would like to benefit from the data and information that is collected in their areas. More involvement in the research process and knowledge gained from research provides an opportunity to enable more self-direction and empowerment in northern communities.
6. We should use this process to build capacity in the north, to train northern people to do polar data management. Northerners must be equal partners in the process. CHARS is trying to address this by planning for data management facilities in northern Canada.
7. Communities would also like to benefit from some of the funding for research that takes place in their areas.
8. ITK indicated it is very difficult, and actually not appropriate, for them to act as the sole representative for Inuit. To properly represent Canada's Inuit, ITK needs regional representatives who develop and produce data and who make decisions using data to be involved in this coordination process and ensuing structure.

9. Role of CHARS

1. CHARS should be a partner, and perhaps a leader, in this coordination effort. They have international partners and clear linkages with the government structure, including the CHARS President sitting on the ADM committee, thus they may be a good conduit to reach a wide variety of stakeholders.
2. CHARS/CPC personnel suggested they may host the secretariat, as knowledge mobilization and coordination are part of their mandate. They have also offered to provide in-kind resources toward this effort.
 1. Could CHARS fund a position to help with coordination and the advisory committee?
3. CHARS is dedicated to proper data and information management and is currently developing a data policy, thus POLAR would be a natural advocate for and contributor to coordination of polar data management in Canada.
 1. To do this, POLAR would need to have the capacity and resources, and they would need to learn about and develop the necessary linkages with the organizations and stakeholders in the science and data management communities, to receive input from the community, promote community preferences, and facilitate integration of stakeholders.
 1. Some progress is already being made on making linkages through a funding proposal for Arctic data interoperability to the emerging CFI cyberinfrastructure program, between CHARS, ITK, and several universities.
4. CHARS can help advance the Canadian Arctic scientific research agenda. POLAR should take advantage of its program and location to attract not only international researchers but also resident scientists who live in Cambridge Bay.
5. There were some concerns with a government organization taking on the role of leadership for this community, as there may be too much bureaucracy, potential interference with academic freedom, instability due to changing government priorities, and insufficient capacity to handle the role.
 1. Most people were more comfortable with CHARS playing a support role but not directing it.
 2. POLAR may play the honest broker role by hosting the secretariat, but they should not handle everything. Other members of the network should lead it and take actions to accomplish joint activities.
 3. POLAR may not be able to take a leadership role in technical issues such as interoperability.
6. The opening of CHARS in 2017 can give us a goal for making progress.

10. Role of the Polar Data Management Committee (PDMC)

1. The PDMC may be able to form the core of the new national committee or may be the national organization that officially represents and speaks for the polar data community in Canada.
 1. There is general consensus in the room that this may be appropriate. The mandate of the Committee would of course change and be broadened, and new members would be added.
2. PDMC could perhaps play the leadership role but will still need an advisory function for guidance and direction.
 1. The PDMC has recently discussed splitting its advisory and operational functions, so this may fit well into a dual-role management committee model.
3. The relationship between PDMC and POLAR will need to be determined.
4. The PDMC will need to be consulted on this possibility, to confirm that they are open to the idea of expanding their mandate.

5. A new name would likely be needed, and the new mandate would need to be defined. As always, resources would be required to guarantee sustainability and usefulness to the community.
6. We may also choose to just create a new structure, instead of using the PDMC as the base. In this case, there may be overlapping membership, but the mandates would be distinct to serve both the polar data community in Canada as well as the PDC itself.

11. Funding

1. Long-term sustainability of both data management and coordination are critical.
2. We need to ensure that the users find us indispensable so that they will support our continuation.
3. Success of the Canadian polar data management system may be facilitated by international coordination of activities and funding, in addition to coordination within Canada.
4. Data and information management is not prioritized in Canada as it is in some other countries. We should engage foundations and companies to consider taking on the Arctic as a priority issue. For example, Coca Cola may be willing to fund polar bear research and would be willing to support proper stewardship of the resulting data.
5. We should set up a strategic plan for acquiring funding to support secretariat/committee activities.
6. There are some universities which are working on including data management in the university infrastructure. We should make sure to work with these universities to understand and provide guidance on their plans, if applicable, so that they may optimize funding expenditures.
7. In addition to data managers, institutions and researchers need funding for data management, too.
8. One option would be for all of us, the polar data community, to self-fund a central body to coordinate ourselves. Such a grass-roots effort would ensure that we could direct it as we wished. Would this be worthwhile? Could we afford it? We would have to guarantee that the expenditure and effort would be worthwhile and well executed.
9. Alternatively, perhaps researchers or programs could use a percentage of their research funding to support coordination. If the researchers found this useful, then it is possible that they would be willing to support it. However, it is unlikely that it would happen without the funding programs requiring it and providing guidelines on amounts and timing of contributions.
10. This group will need institutional support from one or more organizations.

12. Plans for the position paper and a follow-up survey

1. The position paper should be short. It should include the message and the requests, but it will be supported by the long report which has the details.
2. A few people should be tasked with writing the position paper which should include options that can be discussed within the community.
3. Need to ensure that federal employees can access the document online for review and editing.
4. Once the position paper has been vetted fully by the community, it will be very important to take it directly to the decision makers to share the information and ask for partnership in accomplishing our community's goals.
5. Second questionnaire - should we do it?
 1. After discussion, it was determined that a second questionnaire may be useful but is probably not necessary at this point. It may be better to wait until the position paper

is ready or perhaps run a new survey before the next Workshop, to conduct a stakeholder assessment or environmental scan.

2. Need to ensure it is circulated to all interested and involved stakeholders, including industry.
3. The current participants can help form the next questionnaire!