Polar Data Management for Northern Science

Project Leader

Ellsworth LeDrew (University of Waterloo)

Network Investigators

Warwick Vincent (Université Laval)

Collaborators

Scott Tomlinson (Aboriginal Affairs and Northern Development Canada - Northern Contaminants Program); Yves Crevier (Canadian Space Agency); Leah Braithwaite, Michael Svoboda (Environment Canada - Canadian Ice Service); Mathieu Ouellet (Fisheries and Oceans Canada - Integrated Science Data Management); Scot Nickels, Peter Pulsifer (Inuit Tapiriit Kanatami); Josée Michaud, Julie Veillette, Marie-Ève Garneau (Université Laval); David Friddell, Julie Friddell, Kumari Gurusamy, Frank Lauritzen (University of Waterloo)

PhD Students

Virginie Roy (Institut des sciences de la mer de Rimouski); Yunwei Dong (University of Waterloo)

MSc Students

Maren Pauly, Freddy Paya, Brittany Renken (University of Waterloo)

Undergraduate Students

Amyn Aung, Claire Elliott (University of Waterloo)

Abstract

The progress of science towards interdisciplinary exchange and integration of information, such as in the ArcticNet Integrated Regional Impact Studies (IRISes), demonstrates the need for more efficient use of data resources using well-structured systems of data deposition and access. The central objective of ArcticNet's "Polar Data Management for Northern Science" is thus to facilitate exchange of information and data about the polar regions among researchers and other user groups, including northern communities and international programs. During Phase 1 of ArcticNet, members of Theme 2 initiated this data management project. ArcticNet then partnered with the Canadian Cryospheric Information Network (CCIN) at the University of Waterloo and the Department of Fisheries and Oceans Canada (DFO) to develop a database of metadata, the "meta-database" describing ArcticNet datasets. After several modifications and the joining of new partners such as the Canadian federal government program for International Polar Year (IPY), the Northern Contaminants Program (NCP), and Environment Canada, the ArcticNet meta-database, renamed the Polar Data Catalogue (PDC, https://www.polardata.ca), was launched in July 2007. In 2010 the RADARSAT Polar Science Dataset was added to the collection, and in 2011 archiving and access to data files became functional. The Polar Data Catalogue is now Canada's primary on-line source for data and information on research in the polar regions. The scope of research covers a range of disciplines, from natural sciences to policy and health and social sciences. Research projects presented in the PDC are conducted under the auspices of a wide variety of programs, including ArcticNet, NCP, the Canadian IPY (2007-2008), the Circumpolar Biodiversity Monitoring Program (CBMP), the Beaufort Regional Environmental Assessment (BREA), and others.

The Polar Data Management team is also working with other relevant projects (both national and international) toward integrated data management systems to ensure (1) preservation of polar metadata and datasets for the long term, (2) public accessibility of the metadata and datasets on the PDC in a timely and user-friendly format, and (3) responsive development of PDC data tools for use by various stakeholders, especially northern communities.

Key Messages

- The ArcticNet data policy and the development of the Polar Data Catalogue (PDC) were initially motivated by the need for data sharing among ArcticNet researchers and with Northern stakeholders.

- The PDC team works on the continuous development of tools that will help researchers share and preserve their data and that will help users properly find, visualize, and use the data archived in the PDC.

- Each year, the focus on sharing data and proper data management increases in the international research community. The PDC is a leader in Canada in building reliable infrastructure for archiving, serving, and sharing Arctic and Antarctic data and contributing to polar data management policies and best practices.

- Data management is rapidly becoming a formal requirement for research granting and coordination agencies around the world. In 2013, the International Arctic Science Committee (IASC) released its Data Policy (http://www.iasc.info/images/pdf/IASC_data_statement.pdf), stating that all IASC-endorsed projects require a formal data management plan including identification of the national or international repository in which the data will be archived and available in the public domain. In Canada, the Tri-Council released a digital scholarship policy framework for community consideration and response. The Polar Data Catalogue, co-developed by ArcticNet, CCIN, and many other contributing partners, places ArcticNet researchers at the forefront of meeting these evolving national and international expectations.

- In 2013, the online PDC Data/Metadata Input and PDC Search applications have been updated, including addition of over 20,000 new RADARSAT-1 scenes of northern Canada as well as a new GIS Viewer which allows visualization and focused querying of selected geospatial datasets.

- The PDCLite low-bandwidth search tool is being evaluated by Northerners for their needs and modes of interaction. This Lite version of the PDC Search is better adapted to search for projects around a specific community, and it was developed in direct response to feedback and requests from Northern stakeholders in ArcticNet.

- DOIs (Digital Object Identifiers) are now being attributed to ArcticNet and IPY data sets archived in the PDC. DOIs provide several advantages, including higher visibility, better access to datasets, and greater recognition via data citations.

- Progress has been made this year in sharing metadata with other polar data portals via internationally-standardized web services protocols. Metadata and data interoperability will continue to be a significant focus of effort in 2014.

- The PDC and the Polar Data Management Committee have been actively expanding partnerships and collaborations at both the national and international level. Linkages across Canada and the polar world will help ensure the success and sustainability of the PDC for the long term.

- The PDC server and hardware configurations, Java application code, and data not duplicated outside the PDC archive are stored offsite for disaster recovery and protection against data loss. The hardware configuration includes failsafe redundancy and continuous backup of data and software.

Objectives

The central objective of ArcticNet's Polar Data Management for Northern Science, as described in the ArcticNet Data Policy, is to facilitate exchange of information about the Canadian Arctic among researchers and other user groups including northern communities and international programs. To implement its Data Policy, ArcticNet was required to develop a system of information exchange among user groups and to create and maintain a database of metadata compatible with international standards. The creation of the ArcticNet metadata database and then of the Polar Data Catalogue has allowed ArcticNet to meet these objectives.

The specific objectives of data management for 2013-2014 have been:

- To work with other relevant projects (both national and international) toward management of their programmatic data as well as establishment of integrated data management systems. Partners include the Beaufort Regional Environmental Assessment, the Northern Contaminants Program, the Canadian Space Agency, Environment Canada, the Global Cryosphere Watch of the World Meteorological Organization, the Canadian High Arctic Research Station, the Circumpolar Biodiversity Monitoring Program, and the Nunavut General Monitoring Plan.

- To ensure that polar metadata and datasets are preserved for the long term and are publicly accessible on the PDC in a timely and user-friendly format.

- To make links with other polar data organizations for sharing metadata through internationally-standardized metadata interoperability protocols.

- To rebuild the PDC Search application using updated technology and enhance the portal's functionality by adding new data visualization and download capabilities for both research datasets and RADARSAT imagery.

- To archive the Amundsen data in the PDC.

- To assign Digital Object Identifiers (DOIs) to ArcticNet and Centre d'études nordiques datasets.

- To complete construction and configuration of the new CCIN/PDC hardware and server infrastructure.

- To build and release a new Help Manual for PDC users.

- To document the components and processes of the PDC and CCIN for ease of future use and maintenance.

Introduction

The Polar Data Catalogue (PDC, http://www.polardata.ca) is Canada's primary on-line source for data and information on research in the polar regions. Research projects presented in the Catalogue are conducted under the auspices of a wide variety of programs, including ArcticNet, the Northern Contaminants Program, the Arctic Council's Circumpolar Biodiversity Research Program (CBMP), the Government of Canada Program for the International Polar Year (IPY) 2007-2008, Environment Canada, and others. The scope of research covers a range of disciplines, from natural sciences to policy to health and social sciences. RADARSAT-1 imagery from the Canadian Space Agency and the Canadian Ice Service of Environment Canada are also available for viewing and download through the PDC.

The ArcticNet data management system was initiated during Phase 1 of ArcticNet by members of ArcticNet Theme 2 (Food, Water and Resources in the Shifting North-South Thermal Gradient of the Terrestrial Eastern Canadian Arctic), in response to the need to facilitate data access for collaboration and synthesis and to meet the information needs of ArcticNet Network Investigators (NIs), partners, and stakeholders (Vincent et al., 2010). To put in place this data management system, ArcticNet partnered with the Canadian Cryospheric Information Network

(CCIN, at the University of Waterloo) and with the Department of Fisheries and Oceans Canada (DFO) to develop a "meta-database," a database of metadata records describing research projects and datasets. After several modifications and updates, major development of online input and search engine tools by CCIN, and the addition of new partners such as the Canadian Federal IPY program, the Northern Contaminants Program, and Environment Canada, the ArcticNet meta-database, renamed the Polar Data Catalogue (PDC), was launched in July 2007. In 2010 the RADARSAT Polar Science Dataset from the Canadian Space Agency was added to the collection, and in 2011 the data archiving and access capability became functional. In 2012, the co-chairs of the Polar Data Management Committee were invited to participate in an Environment Canada Science and Technology Branch Northern Data Stewardship Workshop where the PDC was demonstrated as a model approach and success story to emulate.

The growing national and international profile of the PDC and the emphasis on data management by the ArcticNet international review panel underscores the quality of work of the ArcticNet data management team. This commitment is paralleled by the growing perception of government, funding agencies, publishers, and many scientists that data sharing and deposition should become a norm and a priority. This is particularly true in times where advances in information technologies for gathering and accessing large volumes of data not only increase the access to information but also allow data combination and synthesis in non-traditional or unexpected ways, leading to important new insights (Anonymous, 2009; Parsons et al., 2011). The progress of science towards interdisciplinary exchange and integration of information, such as in the ArcticNet Integrated Regional Impact Studies (IRISes), demonstrates the need for more efficient use of data resources using well-structured systems of data deposition and access. With its active participation in the PDC, ArcticNet is thus in a world-leading position for integrated data management systems and their ongoing advancement.

Activities

Data Management and Stewardship

ArcticNet

Engagement with ArcticNet scientists and students has continued as in past years, with 29 new metadata records and 14 new datasets being added this year, for a total of 659 total approved metadata and 28 total datasets, comprising 542 datafiles (further details are in the Results and Metadata sections of this report). In addition, the ArcticNet Data Manager allocated considerable time to managing the diverse ship datasets collected on the CCGS Amundsen since 2003 to ensure their security and to facilitate long-term archiving. A portion of the data was sent to Waterloo for storage in the PDC system. The next steps are to complete the data organization and transfer to the PDC and to produce metadata records for each type of data collected so that the data can be made publicly available (as appropriate) in the PDC. With the departure of the ArcticNet Data Manager this year, the PDC Data Manager at CCIN in Waterloo has assumed responsibility for review and approval of ArcticNet metadata submissions which have been extensive, as usual, in the lead-up to the winter ArcticNet annual reporting deadline. Discussions are ongoing to determine the optimal personnel configuration and process to most effectively complete the work on the Amundsen and CASES data and to move forward on the large task of working with the ArcticNet community to prepare and submit their collected datasets to the PDC.

Other programs

PDC, which is a member of the Canadian Polar Data Network (CPDN, http://polardatanetwork.ca), is the chosen repository for the Beaufort Regional Environmental Assessment (BREA) program, the Northern Contaminants Program (NCP), and, new this year, the Nunavut General Monitoring Program, all of Aboriginal Affairs and Northern Development Canada (AANDC). In support of these programs,

we are working with program staff to inventory and understand individual projects, researchers, and datasets, then we are subsequently working with researchers to ingest, archive, and serve their metadata and data. Also, we are writing "project-level" metadata for each of the individual projects within each umbrella program, starting with BREA and moving to NCP recently. Both of these are extremely time-consuming tasks as numerous contacts with researchers are required, especially as data preparation and submission begin. Additional activities with NCP this year involve analysis of possible methods for ingesting monitoring data streams and establishing interoperability with other data portals in which NCP scientists are obliged (such as through international agreements on reporting contaminants) to archive their data. Efforts to complete the metadata and data collections from the International Polar Year program have been ongoing through the PDC Data Manager at Waterloo, with many new and updated metadata records as well as new datasets available over the past year.

We have been working for a number of years with the Circumpolar Biodiversity Monitoring Program (CBMP) of CAFF, the Conservation of Arctic Flora and Fauna biodiversity working group of the Arctic Council, to create and enter metadata records in the PDC. Past efforts, managed largely by the ArcticNet Data Manager, have been on entering metadata records for the CBMP-Marine projects, with a final push by the Marine working group to complete a small number of outstanding records this summer. During 2013, efforts were focused on working with the CBMP-Terrestrial and CBMP-Freshwater working groups to organize and complete their respective monitoring project lists, in anticipation of uploading the new metadata into the PDC. It is expected that these new collections will add several hundred additional metadata records to the PDC during 2014. At this time, an official agreement is being drafted which will more clearly define the working relationship between the PDC and CBMP.

The partnership with CHARS is ongoing this year. Through our partnership with the CPDN, we have

produced a major report entitled Data and Information Management for Arctic Science and Technology: A Proposed Approach for the Canadian High Arctic Research Station. This 81-page report contains chapters on data management principles, a policy framework, information management services, and data and information technology infrastructure that are recommended for CHARS. The final chapter recommends a series of next steps with respect to principles and policies to support data management services, preliminary IT infrastructure design and initial planning, and initial contact with user communities. We have followed on this year with CHARS personnel with a number of discussions of server and storage hardware requirements, configuration, and cost, and we have been contracted to define their initial datasets so that appropriate data management policies and processes can be designed.

PDC Online Application Development

PDC PDC Geospatial Search tool

Previously known as "whitesnow," what is now the "PDC Search" application (accessible at http://www.polardata.ca/pdcsearch/) has undergone significant change over the past year. The code has been completely converted from raw Java to the Java Spring framework, dramatically improving security (such as preventing SQL injection and other potential hacking tricks), speeding up performance, and modularizing functions to streamline future development tasks. At the same time, a number of new features were added and some functions and database queries were optimized to provide faster searching and better performance for users. Through this exercise, the code base was cleaned, with legacy or extraneous code removed and non-standard usage addressed, so that the resulting code is streamlined and much easier to read and understand. This code rewrite represents a simplification of the web technologies used in the PDC (all three PDC web applications are now in the Java Spring framework) and a reduction in time and effort required to respond to user needs and requests, maintain and enhance the applications, and train new co-op students and developers.

In addition to the full code rewrite, a new PDC GIS Viewer has been added to the PDC Search tool. This tool gives users the ability to perform limited queries of selected PDC datasets which have been converted into geospatial format (shapefiles). This new feature displays the dataset(s) on a map and allows users to click on the data points to view targeted portions of large datasets. This facility was developed with support from BREA at AANDC and uses the new GeoServer function that has been installed this year (see the Metadata Interoperability section below for more information on GeoServer) which allows the PDC Search tool to interact with the newly-converted geospatial datasets. This new feature dramatically enhances access to PDC datasets by giving users a way to visualize a dataset that otherwise would only be accessible as a table of numbers. We have been in touch with scientists to enter additional datasets in GeoServer for access with the GIS Viewer (see figure 1).

Another major upgrade to the PDC Search this year is enhancement of the RADARSAT imagery delivery mechanism. Previously, the only imagery format that
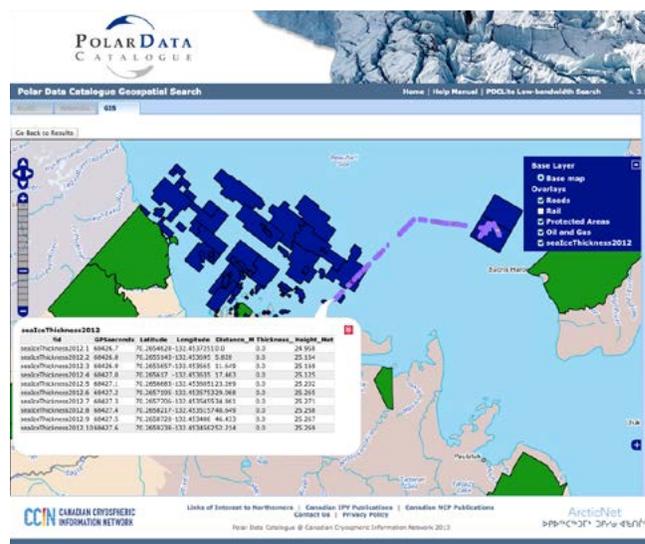

Figure 1. PDC GIS Viewer. The dataset, Sea Ice Thickness 2012, is in purple circles. Protected areas are green polygons, oil and gas leases in the Beaufort Sea are blue polygons, and roads are grey.

users could access and download was geotiff, but users had expressed interest in having access to the imagery in other formats to provide additional flexibility for sophisticated analysis. Through support from the Canadian Space Agency, the PDC Search interface was upgraded and new functions were provided through GeoServer to also provide image data in the original high-resolution CEOS format as well as low-resolution jpeg, pdf, png, tiff, and gif formats. New functionality to allow batch downloading of up to 25 images at a time has also been incorporated to meet needs expressed by users. Additionally, users may overlay multiple images and mosaic them together to cover larger areas with linked scenes. Along with this new functionality, over 20,000 new RADARSAT-1 images of Canada's north were added to the repository this year, bringing the total number of scenes available for download to over 27,000. Further work to improve the appearance of the mosaic output and add more imagery, particularly from Antarctica, are planned for the future (see figure 2).

Additional development in the PDC Search application has been done to support the BREA CanICE project with the IT services and Operations departments of the Canadian Ice Service (CIS, Lina Assad and Darlene Langlois), the geomatics research group at the Université Laval (Yvan Bédard), and geomatics

Figure 2. New RADARSAT satellite image view and download function in the online PDC Search application.

engineering at Ryerson University (Songnian Li). CIS has configured and provided feeds of their regional analysis charts and ice polygons through WFS web services, and the PDC Search tool is being enhanced to access these new web services and allow users to query the CIS database and download data as desired. The Laval SOLAP (Spatial On-Line Analytical Processing) team has designed a datacube which will summarize decades of CIS data and provide quick analysis and results for PDC users. We have worked closely with the Laval and CIS teams to design the interaction methodology between the web data feeds from CIS, the resulting SOLAP datacubes, and the PDC Search tool. Work will continue to expand the results of this project beyond their original intent.

PDC Metadata/Data Input tool

Enhancements to the PDC Input application (https://www.polardata.ca/pdcinput/login.ccin) include encryption of passwords in the database to protect user privacy, simplification of the ability to change user e-mail addresses, and correction of a number of small bugs. These improvements provide additional usability and dramatically enhance security for PDC users. More flexibility has been incorporated to allow PDC administrators (the CCIN Manager and PDC Data Managers) to upload data files on behalf of researchers and to review and approve metadata and data. Also, processes for automatic delivery of notification e-mails following metadata and data submission and approval have been standardized and optimized for reliability.

PDCLite low-bandwidth search tool

Plans have been made with our partners at Inuit Tapiriit Kanatami to engage Northerners in early 2014 for a full review of the PDCLite tool (https://www.polardata.ca/pdclite/), including its speed relative to the full-featured PDC Search (how "lite" it is in a variety of communities and situations) and how it is meeting the needs of northern users.
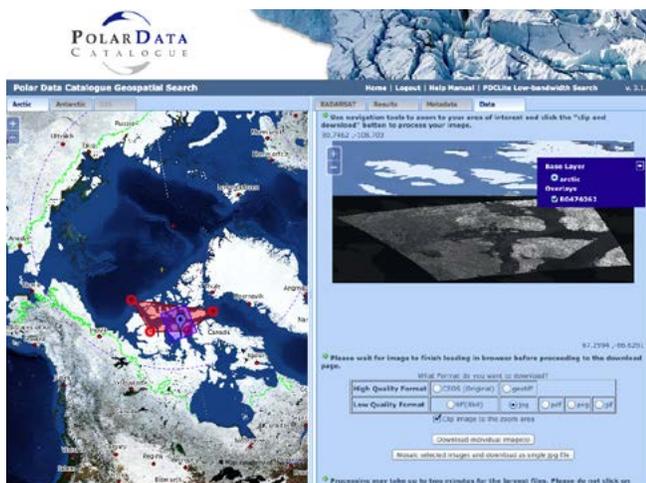
PDC Help Manual

A new PDC Help Manual (https://www.polardata.ca/pdcinput/public/PDCHelpManual.ccin) was released

earlier this year, with sections on entering metadata and data in the PDC Input tool, searching for metadata using both the PDC Search and the PDCLite tools, the approval process for data and metadata, and Frequently Asked Questions. This new Help Manual is dramatically improved over previous versions, with copious screenshots and detailed instructions on how to use the new GIS Viewer and RADARSAT download features.

Metadata Interoperability with Other Polar Data Portals

In addition to the OAI-PMH (Open Archives Initiative - Protocol for Metadata Harvesting) metadata interoperability that was previously established with Canadian and international partners, we added metadata sharing capabilities this year via WMS/WFS (Web Map Service/Web Feature Service) protocols through GeoServer and CSW (Catalog Service for the Web) protocol through GeoNetwork. Connection information through the three methods is available on the CCIN site at http://www.ccin.ca/home/webservices. Current interoperability partners include the US National Snow and Ice Data Center (NSIDC), the Norwegian Meteorological Institute (NMI), the Yukon Research Centre, Scholars Portal of the Ontario Council of University Libraries, and Inuit Qaujisarvingat of Inuit Tapiriit Kanatami. Following participation in the International Forum on Polar Data Activities in Global Data Systems, additional explorations are underway to form partnerships for sharing PDC metadata with polar data portals in Japan, Sweden, the United Kingdom, Finland, and a number of other countries. Initial discussions have revealed great interest in forming an international polar metadata sharing network, thus this will be a major focus over the next couple years.

CCIN has officially joined the SAON (Sustaining Arctic Observing Networks) Polar Metadata Profile task to enable and facilitate this international metadata sharing network. Another motivation for participating in the SAON metadata task is that it is helping guide our internal conversion of PDC metadata from the FGDC (Federal Geographic Data Committee) metadata standard to the North American Profile of the ISO 19115 international metadata standard. This conversion from FGDC to the ISO 19115 standard is a critical step in establishing metadata interoperability with European and other partners, including Canadian government archives which host Arctic data, particularly Environment Canada with whom we are working to share metadata. Since it is a simpler standard and still in use at some portals, the FGDC format will still be available to users who prefer to use it. Because of this, we have been making changes to clean up the FGDC-formatted files to correct some minor validation errors that have caused small issues for others who harvest our FGDC metadata.

An interoperable linkage was established to share the PDC metadata with the Alaska Ocean Observing System and the CBMP. Through joint effort with Axiom Consulting and Design, the PDC's CWS metadata feed was incorporated into the AOOS Data Explorer, providing an attractive graphical format for all our records at the following link: http://data.aoos.org/maps/search/arctic.php#metadata/71c798e0-972c-11e2-a7bd-00219bfe5678

Discussions are ongoing to explore metadata sharing linkages with the Northwest Territories Discovery Portal and the NWT Cumulative Impacts Monitoring Program and with ASTIS, the Arctic Science and Technology Information System of the Arctic Institute of North America at the University of Calgary. We have begun a case study to link publications related to PDC datasets with their entries in ASTIS and the PDC, using the IPY-MERGE project on Arctic microbial ecosystems.

Hardware, Infrastructure, and Security

New RAM (memory) has been added to our production and disaster recovery servers to enhance performance and to increase flexibility in configuration and capability of our virtualized server and networking infrastructure. Several additional backup functions have been added this year, including a new networked

storage array for tertiary backup of the database, servers, and documentation as well as off-site storage of backup hard drives in a safety deposit box at a local credit union. New storage space has been added to the archive to accommodate continuing incoming data files, and a new demonstration virtual server has been built and used for partners and advisors to test new applications and functions before full public release. This new server was used for testing by the Polar Data Management Committee members during the final check-out of the new PDC Search application.

Additional virtual servers have been built to complete our fundamental infrastructure (database servers for the development and testing environments were built early in the year) as well as to support project work, particularly for the BREA CanICE project which is wrapping up in 2014, as well as for more efficient management of the hardware and networking infrastructure. For BREA in general, we built a self-contained virtual server which fits on a laptop, with the goal of demonstrating and using the entire PDC and CCIN system (except for the full 50 TB archive of datafiles) without connection to the Internet. This new concept was developed to accommodate Julie Friddell's needs to get directed feedback from participants on the PDC tools and the CCIN website at the annual BREA Results Forum in Inuvik where it was not certain that Internet access would be available. There were some minor issues that prevented comprehensive Internet-free use of the system, but in the future, this "PDC on a laptop" functionality should be available for use in remote locations to give users the ability to test and query the entire PDC and CCIN infrastructure.

Firmware and software updates on servers and office workstations are applied at regular intervals, and routine security audits, including port scans of our infrastructure, are conducted to ensure all hardware and functions are secured. Along with regular modifications to the database required to optimize speed and structure as well as to support the numerous enhancements to the online applications, an audit of the Oracle code was undertaken to proactively

find and correct errors and to streamline operation and performance. Also, Oracle security patches are applied quarterly. An investigation has been initiated into replacement of Oracle by the open-source database PostgreSQL for use in potential future offsite implementations of the PDC where Oracle may not be available. Additional activities have been undertaken to secure the server and storage infrastructure for the PDC database, metadata, and data files. Advanced intrusion detection tools (AIDE and Deny Hosts) have been configured to stop attempts at brute-force intrusion and hacking. Firewall rules have been tightened, and the web applications have been further secured to more completely harden our system for protection against data loss and corruption. A password manager tool has been set up to simplify use and regular updates (such as when students leave the organization at the end of their term) to passwords for our many server and software logins. The web server configuration files have been audited and optimized for accuracy and efficiency of operation. The SSL (Secure Sockets Layer) certificate has been updated and streamlined on our production servers, ensuring end-to-end encrypted communication for entry of user passwords and other sensitive information in the PDC Metadata/Data Input application.

A four-month effort was undertaken during Fall 2013 to document all systems and components. We now have a new internal documentation website available to CCIN staff, and the office development wiki and activities management/bug tracker system (ToDoYu) have been organized and simplified for ease of use. Use of the code versioning system, SVN accessed through Cornerstone, has been standardized so that we now have a very clear process for tracking development in the PDC code base.

Canadian Cyospheric Information Network website

With support from Environment Canada, CCIN staff have continued to improve the Canadian Cryosphere Watch (CCW) and other components of the CCIN website which was completely redeveloped in 2012.

The Canadian Prairie Snow Water Equivalent maps from Environment Canada have been incorporated into the site in two different formats, the interactive data visualization that we built in 2012 as well as the classic-style static map that Environment Canada has been providing to water managers and other parties since the 1990's. The scientific content on snow, sea ice, lake ice, permafrost, glaciers, river ice, and cryospheric and climate modelling has been further updated with contributions from the CCW's Scientific Advisory Council and with new information and graphics from the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Social media tools have been configured, with CCIN and the PDC now having Facebook (https://www.facebook.com/CCIN.PDC) and LinkedIn (http://www.linkedin.com/groups?home=&gid=7437515&trk=anet_ug_hm) accounts, in addition to the Twitter (https://twitter.com/polardata) account that was set up previously. The Facebook page in particular is important to reach out to northerners who have very high use of FaceBook. The CCIN Graduate Web Developer, Maren Pauly, is posting to these social media tools during the term of her contract, then the responsibility for new material will shift to other staff at CCIN. In addition, a new Northern Hemisphere Snow Anomaly Tracker has been built by scientists at Environment Canada and posted on the site (http://www.ccin.ca/home/ccw/snow/current). This daily, near-real time map-based product is a significant Canadian contribution to the Global Cryosphere Watch and is a model for the type of data visualizations that are planned for the GCW in the future.

Additional spam-blocking features have been added to the blog section of the site to prevent large numbers of spurious entries from being added. Also, due to a potential security compromise of the site, extra effort has concentrated on protecting the site components against intrusion and ensuring that security updates are applied to the site's Drupal Content Management System very soon after they are made available online. These updates result in occasional planned downtime of one hour or so which is announced a day or two in advance.

Nordicana D

The Nordicana D data publication series (http://www.cen.ulaval.ca/nordicanad/), managed by Centre d'études nordiques (CEN), is developing rapidly. One example dataset is 25 years of permafrost monitoring data published online in 2013 (Allard et al., 2013). Data sets from ADAPT (Arctic Development and Adaptation to Permafrost in Transition, an NSERC Discovery Frontiers project that is closely linked to ArcticNet) are also being published via Nordicana D (Vincent et al., 2013). The PDC staff at CCIN are assisting in the development of this data resource, specifically by ongoing work to cross-reference Nordicana D data records in PDC so that they may be found and accessed by way of the PDC Search tool.

GeoWiki

We have been working with scientists at IIASA (International Institute for Applied Systems Analysis, Austria), on their Geo-Wiki project (Wiki http://www.geo-wiki.org/) to enhance the user experience of the PDC/CCIN web pages through incorporation of crowd sourcing tools using mobile platform applications and a Google Earth interface. Dr. Steffen Fritz, Dr. Linda See, and Mr. Christoph Perger of IIASA have been working with Google Earth to create a program such that other scientists and citizens can contribute knowledge and data to public archives for specific applications. We plan to develop this platform to permit Northerners to contribute photographs, observational data, and annotation to the archive within a specific protocol that will address their specific information needs and encourage their participation in the scientific process. This past year a version of the Geo-Wiki was developed and tested with university students as a proof of concept, and future work will be based upon test scenarios with Northerners to evaluate the potential for the PDC/CCIN. The most recent newsletter for the Geo-Wiki project can be found at http://geo-wiki.org/news-outreach/#Newsletters.

Outreach, Communication, and Service

Local presentations to the University of Waterloo community include a "lightning talk" and demo of the new GIS Viewer at the University's GIS Day (November 2013), presentation and discussions of Best Practices in Data Management to the Department of Geography's GEOG 600 seminar course in spatial data handling (February 2013 and January 2014), and a presentation to the University's second annual Data Management Day during Open Access Week (October 2013). This year, CCIN/PDC Manager Julie Friddell gave an invited presentation during the Data Management Day plenary session which was recorded and posted on the CCIN website (http://www.ccin.ca/home/videos). Additional presentations and panel discussions concerning the PDC were given at the CASRAI (Consortia Advancing Standards in Research Administration Information) ReConnect conference (October 2013) by our colleagues in the former IPY Data Assembly Centre Network.

Additional community presentations were given by Julie Friddell to two school groups. "How do we know about Polar Change?" was presented to the Waterloo Region District School Board elementary enrichment class in Waterloo, and "Arctic Change: Observations, the Past, and the Future" was presented to the University of Waterloo's Unlimited High School Enrichment program.

Presentations were also made at international and Canadian conferences, as follows:

- Canadian Cryospheric Information Network and the Polar Data Catalogue: Past, Present, and Future, oral presentation by Julie Friddell, National Institute of Polar Research, Tokyo, Japan (October 2013).

- The Canadian Polar Data Network (CPDN): Data Management for the NCP, poster presentation by Julie Friddell with co-authors S. Marks, C. Humphrey, and K. Morgenroth, Northern Contaminants Program 20th Annual Results Workshop, Ottawa, ON (September 2013).

- The CanICE Project, oral presentation by Julie Friddell, Beaufort Regional Environment Assessment program Results Forum, Inuvik, NWT (February 2013).

- The Polar Data Catalogue, oral presentation and live web demonstration by Julie Friddell, Beaufort Regional Environment Assessment program Results Forum, Inuvik, NWT (February 2013).

- The Canadian Cryospheric Information Network, oral presentation and live web demonstration by Julie Friddell, Global Cryosphere Watch SnowWatch Workshop, Toronto, ON (January 2013).

- Vincent, W., C. Barnard, and N. Bhiry, The CEN Network of Arctic Observing Stations and Nordicana D for Data Dissemination, poster presentation at the Arctic Observing Summit, Vancouver, Canada (April 2013).

Additionally, Julie Friddell served as the founding Chair of the Canadian Polar Data Network from its 2012 inception through the summer of 2013.

Polar Data Management Committee and other team meetings

- March 2013: Josée Michaud visited Julie Friddell in Waterloo to discuss numerous items regarding data management for the ArcticNet projects, particularly the current status and future plans for the Amundsen data collection, strategies for working with ArcticNet researchers to facilitate ingest of datasets into the PDC, upcoming enhancements to the PDC online applications, and year-end financial projections for the ArcticNet accounts at Waterloo and Laval.

- November 2013: The PDMC had a teleconference with a very high member turnout. We discussed the departure of the ArcticNet Data Manager and options for her replacement; a proposal for the PDC to hold more Antarctic metadata and data; national policies and funding support for

polar data management in Canada, particularly in regard to the Tri-Agency Big Data consultation and our interest to engage in the policy development process; requests from new partners to change the front-end of the PDC to provide organization-specific portals; and a request from the Canadian Polar Data Network for ArcticNet to consider their preservation services. For the last two items, Committee membership agreed that creation of multiple PDC inlets would create confusion, thus partners should use the current PDC interface and "look," and the CPDN was provided with a number of questions about their preservation services, for response and presentation at an upcoming PDMC meeting.

- December 2013: The PDMC has a face-to-face meeting at the ArcticNet Annual Scientific Meeting in Halifax. The agenda focused on recent discussions with the Tri-Council on future funding opportunities for polar data management in Canada, including potential support for the PDC for ongoing Arctic or other related programs. Other discussions involved news and updates on technical and strategic development at the PDC, progress on new features in the PDC applications, recent discussions about new partnerships, and publications and outreach for this year. Also, the lack of a presentation on data management at the ArcticNet Student Day during the ArcticNet ASM 2013 in Halifax was noted, and plans were made for making sure one is included next year.

- February 2014: A teleconference has been planned for follow-up on a number of outstanding action items, with a focus on plans for engaging the Tri-Council on future options for polar data management in Canada.

Network of Centres of Excellence, Knowledge Mobilization - ConnectNorth

CCIN and the PDC spear-headed a proposal this year for a new Network of Centres of Excellence on Knowledge Mobilization entitled "ConnectNorth."

A meeting was held in Ottawa in April 2013 with over a dozen stakeholders from CCIN, CEN, Inuit Qaujisarvingat of Inuit Tapiriit Kanatami, the Nasivvik Centre for Inuit Health and Changing Environments, the Geomatics and Cartographic Research Centre of Carleton University, Inuit Research Advisors from Nunavut and the Inuvialuit Settlement Region (ISR), community members and partners from the ISR, and the ARCTIConnexion group from Université du Québec à Rimouski. With these and over a dozen other partners and stakeholders (Figure 3), we submitted a proposal in August 2013 for a $1.6 million, 4-year project to "Make Research Work for Communities in Real-Time" and "Mobilize Knowledge from Regional Initiatives." At the request of northern partners, work was planned to coordinate the research project and licensing applications and processes that are currently in place in various forms throughout the northern territories and Inuit regions by building a new unified database and access portal for use by northern communities and researchers alike. A smaller project was planned to work with the ISR to guide them in development of a digital atlas which would

Figure 3. Map of NCE-KM ConnectNorth proposed partners. Red stars represent members included in the proposal, and yellow stars represent communities where network members have been conducting research and community visits/ outreach.

make pertinent data and information more available and useful for their organizations and communities by establishing mechanisms to share information; develop working relationships between regional organizations, representative communities, and southern organizations and experts; and configure interoperability between their system and others. We planned to have a variety of small and large community and regional workshops so that we could work together in person throughout the life of the Network as well as engage other interested stakeholders. Thirteen letters of support were provided by partner organizations.

After review, we received the unfortunate news that the proposal had not been chosen for funding. However, there is a great deal of momentum behind the ideas of the proposal, and the groups that had worked together are still intent on accomplishing the goals outlined in the proposal. We are looking for additional funding and partnership opportunities to further develop plans for building a unified research project and licensing database and application as well as other projects that have been brought to our attention by our northern partners. We look forward to investigating future options and to working together on these shared goals.

Cost Model for PDC and the CCIN

With ArcticNet support, the CCIN is engaging with the MBA program at Wilfrid Laurier University in Waterloo, Ontario to develop a business plan and cost model for the PDC. WLU students participating in the Applied Business Research Project will be working for a total of 1,000 hours to provide realistic and competitive costs per unit of service. Answers are expected to questions such as "How much does it cost to enter one metadata record in the PDC and preserve it for decades?" or "How much does it cost to ingest, serve, and archive 1 TB of data?". Information will also be provided on the costs of system sustainability and development of new features and functions. Additional considerations will be an assessment of and comparison to similar services elsewhere, costs

to cover infrastructure and portal development in a sustainable fashion, and possible developments in this field over the next decade, to help us anticipate and meet future challenges with the proposed business model. Students will be interviewing some of our colleagues, including possibly some ArcticNet members. The final report is expected April 2014.

Strategic Planning

Through a formal response to the Tri-Council's October 2013 community consultation on "Capitalizing on Big Data: Toward a Policy Framework for Advancing Digital Scholarship in Canada," the Polar Data Management Committee, chaired by Warwick Vincent, Ellsworth LeDrew, and Leah Braithwaite and responsible for guidance of the PDC, requested formal interaction with NSERC, SSHRC, and CIHR to initiate discussions on opportunities for federal support for polar data management. The PDMC also provided a formal response to the Tri-Council consultation on its "Draft Tri-Agency Open Access Policy" requesting that NSERC and CIHR follow SSHRC's lead in promoting a policy of open access to agency-supported data. We will follow up on these two communications in early 2014.

Results

Metadata in the PDC

By the end of January 2014, the number of metadata records in the PDC reached 1838, with 260 of those either in the SUBMITTED or SAVED states. Efforts will continue this year to address these incomplete metadata for ArcticNet and other programs. This year, ArcticNet scientists and students have contributed 29 new metadata records, for a total of 659 total approved metadata. Additionally, the number of CEN records is 57, and there are 204 metadata for the CBMP-Marine program and 111 records for the AANDC Adaptation Program for Aboriginals and Northerners. The new ADAPT program has entered 61 metadata. The IPY

program has 681 metadata, and the Canadian Arctic Shelf Exchange Study (CASES) has 82 entries. Finally, there are 70 records in the "Reference" category, usually for other polar-related data and information portals and websites from around the world.

Forty-six "project-level" metadata have been created in the PDC for the IPY projects, leaving 6 more to complete. Project level metadata have also been created for 11 of the 23 BREA projects, and records have been started for a small number of the NCP projects. These project-level metadata are high-level descriptions of each IPY project and represent an enhancement of PDC functionality. In the PDC database, project-level metadata are linked to each of the individual metadata that have been submitted for each project, which are called "dataset-level" metadata. Modifications will be made in the PDC Search application to allow hierarchical searching between project-level and dataset-level metadata records. Discussions are ongoing of adding project-level metadata for all of the ArcticNet projects.

Assigning DOIs to Metadata and Datasets

As a result of the new agreements that were signed by ArcticNet, CEN, and the Canadian Polar Data Network (for assignment of IPY DOIs) with the Canada Institute for Scientific and Technical Information (CISTI, the entity responsi- ble for DOI assignment for research data in Canada), DOIs have been assigned to metadata and datasets in the PDC as follows in 2013:

- ArcticNet: 9 DOIs, DOI prefix = 10.5884
- CEN: 18 DOIs, DOI prefix = 10.5885
- IPY: 115 DOIs, DOI prefix = 10.5443

Approximately 200 additional IPY metadata and datasets have been prepared and are waiting for DOI assignment.

Number of datasets

To date, 145 datasets have been submitted to the PDC archive, 28 of which are ArcticNet datasets. Fourteen

of these ArcticNet datasets have been added this year, for a total of 542 data files. Of all the archived data, 76 are available on-line, with the remainder either not available publicly due to privacy issues or temporary embargoes or due to ongoing efforts to properly organize and archive the files. There are 147,503 datafiles, approximately 20 TB in total volume.

Traffic on the PDC and CCIN websites

The PDC (http://www.polardata.ca) and CCIN (http://www.ccin.ca) website traffic is monitored via Google Analytics. Usage during January 2014 is as follows: 3,002 visits from 2,332 unique visitors, viewing 7,294 pages. This represents a small increase over average monthly usage in past years. Traffic generally increases in the winter, largely due to visitors to the CCIN website who are interested in snow and ice during the cold season as well as increased searching for data during the academic terms and increased submission of metadata during the winter and early spring funding reporting seasons. Most visits to all four sites (CCIN website, PDC Input tool, PDC Search tool, and PDCLite) are from Canada, but visitors also came from other countries such as the United States, Norway, United Kingdom, Ireland, Germany, France, Iceland, Russia, Australia, Switzerland, Romania, Bulgaria, Poland, Sweden, and Ukraine. The most popular browsers are Firefox, Chrome, Safari, and Internet Explorer.

Outreach and Publications

In addition to the many presentations and published conference abstracts, two formal publications have been written and submitted this year which will serve as important outreach materials for educating others on the functions of the PDC and CCIN.

Discussion

The PDC has reached a number of important milestones this year, including completion of the final details of the hardware and server infrastructure, a full

code update and addition of several user- and partner-requested functions in the PDC Search application, establishment and use of the DOI registration process for PDC datasets, and configuration of new web services for sharing the PDC metadata. Additionally, a huge new RADARSAT-1 image collection was added to the PDC archive, the scientific content on the CCIN website was completely renewed, the PDC Help Manual for users was completely rewritten and updated, and the PDC and CCIN operations were documented in a new internal website for project use. Partnerships were strengthened with Canadian and international supporters and collaborators. A total of nine proposals or Letters of Intent were submitted to expand and enhance the functionality of the PDC and its data management capabilities for Canada. Ingest of metadata and data files from ArcticNet and other partner projects continued in conjunction with significant efforts to educate and train researchers on data management principles and best practices.

Beyond the PDC, data management is rapidly becoming a formal requirement for research granting and coordination agencies around the world. The IASC released its Data Policy in 2013; this policy was approved after broad consultation, including with members of ArcticNet's Polar Data Management project and each of the IASC Working Groups. The policy (http://www.iasc.info/images/pdf/IASC_data_statement.pdf) states that all IASC-endorsed projects require a formal data management plan including identification of the national or international repository in which the data will be archived and available in the public domain. Two important concepts from the IASC Data Policy have been quoted below:

- "Preservation and Stewardship: Recognizing that the true value of scientific data is often realized long after these data have been collected, and to ensure the lasting legacy of IASC supported activities, it is essential to facilitate long-term preservation and sustained access to IASC data. All IASC data should be archived in their simplest, useful form and be accompanied by a complete metadata description in accordance with

the OAIS [Open Archival Information System] Reference Model. Projects should identify appropriate long-term archives and data centers and describe preservation processes in their data management plans."

- "IASC should not endorse future projects that do not adhere to the principles outlined in this document. IASC should actively encourage adherence to the principles and may withdraw project endorsement if necessary."

These statements are being evaluated internationally and are spurring discussion on the full application of preservation and access for polar research data.

In Canada, the Tri-Council released a policy framework on digital infrastructure and invited discussion. This ongoing process, to which the Polar Data Management Committee provided a formal response, is likely to eventually align with other countries such as the USA and United Kingdom. The Polar Data Catalogue, co-developed by ArcticNet, CCIN, and many other contributing partners, places ArcticNet researchers at the forefront of meeting these national and international expectations.

Conclusion

Continued development of the PDC is strengthening effective management of polar data in Canada. Support from ArcticNet and other partners provides the ability to build and maintain a capable and respected archive for long-term access to and stewardship of Canada's polar research data. Our commitment to incorporation of international standards for discovery and interoperability is positioning the PDC to be a contributor to the future of data management. Our progress is being increasingly recognized at the national and international levels. Linking the PDC to the rest of the world's polar data portals is anticipated to be one of our most important tasks for the next few years, including creating strong links to Canada's northern Aboriginal people to increase the service that we can provide as their communities experience environmental and social change.

Acknowledgements

References

Allard, M., D. Sarrazin, and E. L'Hérault, 2013. Borehole monitoring temperatures in northeastern Canada, v. 1.1 (1988-2013). Nordicana D8, doi: 10.5885/45291SL-34F28A9491014AFD.

Anonymous, 2009, Data's shameful neglect, Nature Editorial, Nature 461(7261):145.

Parsons, M., Ø. Godøy, E. LeDrew, T. de Bruin, B. Danis, S. Tomlinson, and D. Carlson, published online 19 October 2011, Managing Very Diverse Data for Complex, Interdisciplinary Science, Journal of Information Science, pp. 1-21. DOI: 10.1177/0165551511412705.

Vincent, W., C. Barnard, J. Michaud, M.-È. Garneau, 2010, Data management, In: Vincent, W., M. Lemay, and C. Barnard (Eds)(2010), Impacts of Environmental change in the Canadian Coastal Arctic: a compendium of research during ArcticNet Phase I (2004-2008). ArcticNet Inc. Québec City, Canada.

Vincent, W., M. Lemay, M. Allard, and B. Wolfe, 2013, Adapting to Permafrost Change: A Science Framework, Eos, Transactions of the American Geophysical Union, v.91, 373-375.

Publications

(All ArcticNet refereed publications are available on the ASTIS website (http://www.aina.ucalgary.ca/arcticnet/).

Canadian Polar Data Network, 2013, Data and Information Management for Arctic Science and Technology: A Proposed Approach for the Canadian High Arctic Research Station, report, 81

Fortin, M., Bédard, Y., Friddell, J., Rivest, S. and Roy, T., 2013, New Technology for the Interactive Analysis of Ice Data in Beaufort Sea, ArcticNet Annual Scientific Meeting, Halifax, NS, p. 52

Friddell, J., LeDrew, E. and Vincent, W., 2013, The Polar Data Catalogue: Best Practices for Sharing and Archiving Canada's Polar Data, CODATA Data Science Journal,

Friddell, J., LeDrew, E. and Vincent, W., 2013, The Polar Data Catalogue: Data Management for Polar and Cryospheric Science, Proceedings of the 70th Eastern Snow Conference,

Friddell, J., LeDrew, E. Vincent, W. and Michaud, J., 2013, News from the Polar Data Catalogue: New Maps, Data, and Improved Access for Northern Communities, ArcticNet Annual Scientific Meeting, Halifax, NS, p. 143

Friddell, J., LeDrew, E., Vincent, W. and Michaud, J., 2013, The Polar Data Catalogue: Data Management for Polar and Cryospheric Science, Scientific Program and Abstracts for the 70th Eastern Snow Conference, Huntsville, ON, p. 45

Friddell, J., Michaud, J., Vincent, W. and LeDrew, E., 2013, The Polar Data Catalogue: Best Practices

for Sharing and Archiving Canada's Polar Data,
Programme and Abstracts from International Forum on
Polar Data Activities in Global Data Systems, Tokyo,
Japan, p. 6-8

Pulsifer, P., Yarmey, L., Godøy, Ø., Friddell, J.,
Parsons, M., Vincent, W., DeBruin, T., Manley,
W., Gaylord, A., Hayes, A., Nickels, S., Tweedie,
C., Larsen, J.R. and J. Huck, 2013, Towards an
International Polar Data Coordination Network,
CODATA Data Science Journal,

Pulsifer, P., Yarmey, L., Godøy, Ø., Friddell, J.,
Vincent, W., DeBruin, T. and Parsons, M., 2013, Data
Management for Arctic Observing, A Community
White Paper, Arctic Observing Summit, 21