Polar Data Management for Northern Science

Project Leader
Ellsworth LeDrew (University of Waterloo)

Network Investigators
Warwick Vincent (Université Laval)

Collaborators & Research Associates
Josée Michaud (ArcticNet Inc.); Scott Tomlinson (Aboriginal Affairs and Northern Development Canada - Northern Contaminants Program); Yves Crevier (Canadian Space Agency); Leah Braithwaite (Environment Canada - Canadian Ice Service); Michael Svoboda (Environment Canada - Canadian Wildlife Services); Mathieu Ouellet (Fisheries and Oceans Canada - Integrated Science Data Management); Scot Nickels, Peter Pulsifer (Inuit Tapiriit Kanatami); Julie Friddell (University of Waterloo)

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

MSc Students
Freddy Paya, Daniel Xia (University of Waterloo)

Undergraduate Students
Liang Diao, Claire Elliott (University of Waterloo)

Technical Staff
Julie Veillette (Université Laval); Kulunu De Zoysa, David Friddell, Kumari Gurusamy, Frank Lauritzen, Nobuki Matsui (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 International Polar Year program (IPY), the Northern Contaminants Program (NCP), and Environment Canada, the ArcticNet meta-database, renamed the Polar Data Catalogue (PDC, http://www.polardata.ca), was launched in July 2007. In 2010 the RADARSAT Polar Science Dataset was added to the collection, and in 2011 the data archiving and access capability 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 to 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), and the Beaufort Regional Environmental Assessment (BREA).

The Polar Data Management team is also working with other relevant projects (both national and international) toward integrated data management systems to ensure (1) 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, and (2) the 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.
- The online PDC Data/Metadata Input and PDC Search applications were updated and now allow advanced searching using several fields.
- The new PDCLite low-bandwidth search tool was brought online. 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. It is being further evaluated by Northerners for their needs and modes of interaction.
- DOIs (Digital Object Identifiers) can now be attributed to data sets archived in the PDC. DOIs provide several advantages, including higher visibility, better access to datasets, and greater recognition via data citations.
- A great deal of progress has been made this year in sharing metadata with other polar data portals. Metadata and data interoperability will continue to be a significant focus of effort in 2013.
- 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.

- 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.
- The PDC database has been externally audited, and all 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 2012-2013 have been:

- To work with other relevant projects (both national and international) toward integrated data management systems, including the Beaufort Regional Environmental Assessment, the Northern Contaminants Program, the Canadian Space Agency, the Global Cryosphere Watch, and the Canadian High Arctic Research Station.
- 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 implement the new PDCLite low-bandwidth search tool for northern communities.
- To develop the ability to assign Digital Object Identifiers (DOIs) to datasets.

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), and the Government of Canada Program for the International Polar Year (IPY) 2007-2008, Environment Canada, and the Canadian Space Agency. The scope of research covers a range of disciplines, from natural sciences to policy to health and social sciences.

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 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 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 PDC 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

**New hardware and infrastructure**

With funding from the Canadian International Polar Year (IPY) program, we have deployed all new server and storage infrastructure for the PDC database, metadata, and data files. The new hardware was purchased in 2011 and major portions of it were put into service in March 2012, including new production and disaster recovery environments. The production system (the one which faces the public) includes a pair of IBM servers running virtualized web, database, load-balancing, and file servers configured for semi-automatic replacement in case of failure. The paired servers are connected to a 50 TB storage archive which holds the new Oracle 11gR2 database of metadata and all the ArcticNet, IPY, and other data files. The new database has been designed and built from scratch and is much more streamlined, with better logic and fewer tables than the old database. One important change that has been made is addition of “history” to the metadata tables, so that all versions of a metadata record will be preserved as it is updated over time. The disaster recovery hardware is in a separate building on campus, providing fully redundant servers and storage which can be opened to public use in case of complete loss of the production system.

A variety of back-up functions have been implemented, with server back-ups occurring on-site and metadata and IPY data back-ups being sent to DFO in Ottawa and Scholars Portal in Toronto, respectively. Since March 2012, we have established, using a portion of the old CCIN/PDC hardware, development and testing environments which duplicate the production system (except for the 50 TB storage) for use in developing and testing new CCIN and PDC functions and features.

This new hardware and database represents a major improvement to the infrastructure of the PDC. We have also added new functionalities to improve workflow in the office: We have installed and configured a new SVN application code repository, new task
management software to track activities, and an office wiki for documentation. With these improvements, we are ready to accept many terabytes of new data and are in a much better position to develop, maintain, and upgrade the system over the next decade, than was previously the case.

**PDC Metadata/Data Input online application**

Enhancements to the PDC Input application include establishment of “https” capability (to provide secure, encrypted metadata and data upload); extension of the length of text used for the Abstract, Purpose, and other fields in the metadata submission form (to accommodate needs from the IPY program); and upgrade of the Java code from the Spring 2 framework to Spring 3 (to stay current with technology and tools). These improvements provide additional usability and dramatically enhanced security for PDC users.

**PDC Geospatial Search online application (whitesnow)**

Improvements to the PDC Search application include addition of an “Advanced Search” capability (to search for keywords in specific fields of the metadata, such as Title, Abstract, Contact Person, etc.); modification of the metadata/dataset bounding box display to show in polar coordinates (Figure 1) instead of a square box, which was spatially incorrect; configuration of a
unique URL for each metadata record and dataset in PDC, as required for creation of DOIs (Digital Object Identifiers) for datasets/metadata; and replacement of the problematic e-mail notification/ftp log-in procedure to download RADARSAT geotiff images with an instant-access download capability. A variety of modifications has been made to the Java code to improve security, including hiding in the browser interface the explicit path to data files on the server, and to simplify future development and maintenance, including an ongoing project to completely rewrite the application from “raw” Java into the Spring 3 framework, the same Java framework used in the PDC Input application. This code rewrite represents a simplification of the web technologies used in the PDC and a reduction in time and effort required to respond to user needs and requests, maintain the applications, and train new co-op students and developers.

“PDCLite” low-bandwidth online metadata search application

In response to the survey of Northerners conducted by ArcticNet, we released in December 2012, in conjunction with the ArcticNet Annual Scientific Meeting, a streamlined “PDCLite” search tool that functions at low bandwidth and includes a search interface that is more appropriate to searching for projects around a specific community. On-the-ground “liteness” testing has been undertaken by a number of ArcticNet’s partners in the north, with very positive results indicating that PDCLite is up to 90 % faster than the full-featured PDC Search application for users with slow internet connections. Users can search by community name or latitude-longitude position and can select a “search radius” of 10 to 500 km around the search point. Searches can also be for specific Research Programs or keywords, or combinations of the various fields.

Metadata Interoperability with other polar data portals

2012 was a year of reaching out to new and existing partners and setting up new connections to share metadata. The old OAI-PMH (Open Archives Initiative - Protocol for Metadata Harvesting) metadata interoperability with the US National Snow and Ice Data Center (NSIDC) and the Norwegian Meteorological Institute (NMI) were replicated on our new server hardware. Work has continued with Oystein Godoy of NMI to modify the XML files of PDC metadata to facilitate availability of the harvested metadata in the NMI public portal. Successful harvesting via OAI-PMH has been realized with the Inuit Qaujisarvingat: Inuit Knowledge Centre of Inuit Tapiriit Kanatami (ITK), the CBMP, and the Yukon Research Centre, and testing has been done with Scholars Portal of the Ontario Council of University Libraries. Additional work is underway with CBMP to investigate establishment of metadata sharing via Web Map Service/Web Feature Service (WMS/WFS) or Catalog Service for the Web (CSW), all interoperability standards of the Open Geospatial Consortium (OGC). Inventories of PDC metadata submitted by Environment Canada researchers have been sent to EC’s Science Policy Division, for pre-population of their developing EC metadata catalogue of northern science datasets. Discussions have also been underway with the NWT Discovery Portal and the NWT Cumulative Impacts Monitoring Program on options for sharing metadata and with ASTIS, the Arctic Science and Technology Information System of the Arctic Institute of North America at the University of Calgary. We are endeavouring to make a two-way link between ASTIS and the PDC, linking the publications with their associated record or dataset in the PDC, and vice-versa.

Outreach and Communication

We have made a number of presentations to the University of Waterloo community: GIS Day (CCIN/PDC Manager Julie Friddell presented data visualizations and mapping at CCIN and provided a demo of the PDC), the annual WatITis IT conference (Julie Friddell presented the current infrastructure of the PDC archive), and the Department of Geography’s CryoGroup (presentations by undergraduate co-op student Claire Elliott on her work on the RADARSAT
catalogue and by CCIN Web Developer Kumari Gurusamy on the data visualizations and maps that she has built for the new CCIN website. Julie Friddell and Ellsworth LeDrew also contributed heavily to planning for institutional data management services that was undertaken by the University of Waterloo Library in 2012. These discussions resulted in a presentation on data management and the PDC by Julie Friddell to the UW Library Electronic Data Services department staff, two presentations (which included discussion of the PDC) by UW Digital Librarian Pascal Calarco at the 2012 Canadian Association of Geographers Annual Meeting and the Canadian Higher Education and Information Technology (CANHEIT) conference, and Professor LeDrew’s participation, with the UW Library staff, in the Association of Research Libraries’ e-Science Institute in early 2012. Data Management Day at the University of Waterloo, an event which drew over 100 people from the University community who wanted to learn more about managing their research data, was also an outcome of our collaboration with the UW Library. Warwick Vincent gave a presentation on data management from the scientist’s perspective at Data Management Day, and Julie Friddell gave a demonstration of the PDC during the workshop session chaired by Ellsworth LeDrew.

Warwick Vincent presented an update on progress in ArcticNet data management to the ArcticNet Board of Directors at their meeting in Whapmagoostui-Kuujjuarapik (13 June 2012), and presented the Polar Data Catalogue at an NSERC event with senators at the Parliament of Canada (Ottawa, 16 October, 2012) and to senior government officials (Rideau Club, Ottawa, 17 October 2012).

Presentations were also made at international conferences, as follows:

- April 2012, IPY Conference, Montréal: Julie Veillette presented a poster entitled: Analyzing the Polar Data Catalogue: What research has been done and what is missing in the Canadian Arctic. Ellsworth LeDrew, Scott Tomlinson, and Julie Friddell made oral and poster presentations on the PDC and on data management during IPY. John Huck, member of the IPY Data Assembly Centre Network from the University of Alberta Libraries, gave a presentation on our experiences in creating metadata.


- December 2012, ArcticNet Annual Scientific Meeting, Vancouver: Josée Michaud was an invited plenary speaker at the ArcticNet Student Day. She also presented a poster on assignment of DOIs during the ASM. Ellsworth LeDrew provided an oral presentation on the future role of citizen science and social media in polar data management, and Julie Friddell provided a poster on the infrastructure of the PDC and CCIN, to give people an idea of what it takes to build a data portal and archive.

Polar Data Management Committee and other team meetings

April 2012: The PDMC had a face-to-face meeting at the IPY Conference in Montréal. With IPY wrapping up, discussions centred around the task of finding funds to sustain the PDC, including the possibility of approaching the Tri-Council for direct support of polar data management at the PDC.

July 2012: Julie Friddell met with Julie Veillette in Québec City to discuss CCGS Amundsen data, the PDCLite, and assignment of DOIs to ArcticNet and CEN (Centre d’études nordiques) datasets.

November 2012: The PDMC held a teleconference to provide updates on the different activities of the members and to discuss strategies to ensure long-term viability of the PDC. This includes the utilization of models to calculate the true cost of building, enhancing, and maintaining the PDC, including costs.
for submission, approval, public availability, and archiving of data files and metadata, trying to capture the true “end to end” costs.

December 2012: The PDMC had a face-to-face meeting at the ArcticNet ASM in Vancouver. The cost schedule for PDC activities was further discussed. Yves Crevier from the Canadian Space Agency and Michael Svoboda from the coordination office (Environment Canada) of the Circumpolar Biodiversity Monitoring Program were formally invited to join the PDMC. The CSA has been involved in the PDC by providing RADARSAT data. There is the potential to include RADARSAT 2 data of Antarctica into the PDC.

**Canadian Global Cryosphere Watch website**

With support from Environment Canada, CCIN staff have developed a Canadian Cryosphere Watch website, released to the public in September 2012 (Figure 2), to replace the old CCIN/State of the Canadian Cryosphere (SOCC) website. Scientific content on snow, sea ice, lake ice, permafrost, glaciers, river ice, and cryospheric and climate modelling has been updated with contributions from an Advisory Council of thirteen cryospheric experts. The new site uses a modern, open-source Content Management System (Drupal) which simplifies updates and provides a platform for integration of sophisticated data visualization and analysis tools. Dynamic, interactive visualization maps have been developed for snow water equivalent (SWE) data from the Canadian Prairies (from Environment Canada) and the full Northern Hemisphere (Finnish Meteorological Institute’s GlobSnow product) to augment the visualization tools previously developed by CCIN Co-op students for current and historical lake ice data for Canada. The SWE maps have long been one of the most popular items on the current SOCC website, generating thousands of website visits each month during the snow season.

**Beaufort Regional Environmental Assessment (BREA)**

PDC is the chosen repository for all BREA project data and metadata. We have begun to receive BREA metadata and are working with the BREA program office to contact researchers and initiate data submission activities. We are also undertaking a project to configure OGC-WMS/WFS web services at the PDC so that BREA datasets can be shared geospatially with other partners and users.

For the CanICE project with Canadian Ice Services, Université Laval, and Ryerson University, we have had several face-to-face and phone meetings to share progress and plan next steps. The PDC undertook an exercise to determine whether REST or SOAP standards should be used for the web services link between CIS and the PDC. Due to the applicability of the SOAP standard to OGC protocols, SOAP was the chosen standard. We are waiting for the server connection strings from CIS to establish the link with the CIS database and begin serving the historical and real-time sea ice data via the PDC. We have also begun work with the SOLAP team at Laval (Y. Bédard) to design a datacube for PDC usage.

**Northern Contaminants Program (NCP), Aboriginal Affairs and Northern Development Canada (AANDC)**

We are working with Scott Tomlinson to provide metadata management for NCP projects, many of which have been ongoing for many years. Discussions are ongoing with NCP scientists on preparing datasets for submission to the PDC archive.

**Amundsen datasets**

The ArcticNet Data Manager allocated considerable time (20 %) to managing the ArcticNet core-data collected on the CCGS Amundsen to ensure their stewardship and long term archiving. Work included identifying, locating, categorizing and reorganizing the data collected since 2003 from diverse instruments on the CCGS Amundsen. These data were copied at the Department of Information services (DTI) at U. Laval and a complete copy will also be sent to U. Waterloo when the reorganization is finished. The objective of this work is to produce metadata records for each type of data collected and to provide access to these...
Figure 2. The Canadian Cryosphere Watch website. www.ccin.ca
datasets, under the appropriate security level (limited or public). A new protocol for managing the data collected on the CCGS Amundsen is also currently being prepared. This protocol will ensure the long term archiving of the data and will help respond to data inquiries from ArcticNet NIs and collaborators, as most core-data requests will now go through the ArcticNet data manager.

Other activities

Ability to assign DOIs to metadata/datasets

An important development for 2012 was the addition of Digital Object Identifiers (DOIs) to metadata associated with archived datasets. DOIs are persistent identifiers that provide long-term links to data, improving the discoverability, citability, and accessibility to the data to which they are assigned. In May 2012, DOIs became an ISO standard. In 2012, we worked in close collaboration with the Canada Institute for Scientific and Technical Information (CISTI), the entity responsible for DOI assignment for research data in Canada, to ensure the best use of this tool, and to continue to make the PDC an attractive database for ArcticNet and other polar researchers. Data sets for ArcticNet and IPY now have their own DOIs prefix (10.5884 for ArcticNet and 10.5443 for IPY).

Circumpolar Biodiversity Monitoring Program (CBMP)

The ArcticNet Data Manager continued to closely work with the Canadian CBMP program to create and enter metadata records in the PDC. This year, efforts were concentrated on creating 41 metadata records from older reports from the Department of Fisheries and Oceans. Although J. Michaud will remain available to help create metadata records, a new person at DFO will now be entering metadata records for the Marine components of the CBMP, with the guidance of Mathieu Ouellet (member of the Polar Data Management Committee).

Nunavut General Monitoring Plan (NGMP)

Julie Friddell has had a number of e-mail and telephone exchanges with contractors for NGMP regarding their plans for data and information management in the territory. Information has been provided on the background and current capabilities of CCIN and the PDC, with the expectation that our involvement may better inform their data management planning process. We have communicated that we are willing to assist further in the future.

External Oracle database review

We worked with DesTech, a consulting company specializing in Oracle services, to perform an audit of our new Oracle 11gR2 database. The CCIN Database Developer/Administrator, Frank Lauritzen, provided details of the database and its configuration to a DesTech specialist, who reviewed the information and drafted a document with further questions and suggestions for improvement. Answers and clarifications were provided through a series of interactions, and the report, which provides very good documentation of our database for those unfamiliar with it, has been finalized. In summary, there were no problems found with the database, but the review provided external feedback that instigated a number of changes to improve the structure and completeness of the new database.

New initiatives

Canadian High Arctic Research Station (CHARS)

With the Canadian Polar Data Network (CPDN), CCIN is leading a project to scope out the data and information management requirements for the Station and is working with the other CPDN members to write data management policies for CHARS.

Results

Number of metadata

As of January 24th 2013, the number of metadata records in PDC reached 1666; 1438 are approved and thus published online, 18 are submitted and awaiting approval, and 210 are in the “SA VED” status. In the latter case, the metadata creators will soon be informed that they need to submit their metadata so they can be revised and approved.
Of all metadata records in PDC, 78% come from Natural Sciences (Figure 3), followed by Human health and contaminants (10%). We are glad to report that there has been a substantial increase in the number of metadata related to the field of Social Sciences, Economics and Policy, from 41 records in 2011 to 113 in 2013.

Of all the metadata records, 789 are related to IPY and 681 to ArcticNet (Table 1). The number of CBMP metadata has increased to 207 in 2012. A total of 329 records are from other programs such as the AANDC Adaptation Program for Aboriginals and Northerners and the Canadian Arctic Shelf Exchange Study (CASES). More detailed analysis of the ArcticNet metadata found in the PDC is provided in the Metadata section of this report.

As part of the IPY program, new “project-level” metadata are being created in the PDC for the 52 IPY projects. These project-level metadata are high-level descriptions of each IPY project and represent an enhancement of PDC functionality. In the PDC database, these project-level metadata are linked to each of the individual metadata that have been submitted for each IPY project, which we are calling “dataset-level” metadata. Modifications are planned for the PDC Search application to allow for hierarchical searching between project-level and dataset-level metadata records. There have been discussions of adding project-level metadata for all of the ArcticNet projects.

The PDC (http://www.polardata.ca) and CCIN (http://www.ccin.ca) website traffic is monitored via Google Analytics. Average monthly usage during 2012 is as follows: 2,382 visits from 1,623 unique visitors, viewing 7,547 pages. This represents an approximately 10% increase in pageviews over 2011, possibly due to the release of the new CCIN website and the PDCLite in 2012. During January 2013, when ArcticNet researchers were submitting and updating their PDC metadata, usage topped out on Tuesday, 22 January 2013 at 24 visits and 201 pageviews in a single day. 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.

Core-data collected by diverse instruments during the ArcticNet expedition on the CCGS Amundsen were divided into two broad types: 1) Ship’s data:
Table 2. Description of the data collected by instruments hull/deck mounted or deployed from the CCGS Amundsen since the beginning of ArcticNet

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Type of Data</th>
<th>Years</th>
<th>Total Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADCP Surveyor</td>
<td>Currents</td>
<td>2004, 2007-2011</td>
<td>63.3 GB</td>
</tr>
<tr>
<td>AVFMC</td>
<td>Metrological data (wind, temps, atm. Press. etc.)</td>
<td>2005-2012</td>
<td>31.2 MB</td>
</tr>
<tr>
<td>EK60 Echo Sounder</td>
<td>Bathymetric and backscatter data</td>
<td>2003-2004, 2007-2011</td>
<td>1.15 TB</td>
</tr>
<tr>
<td>GPS</td>
<td>Navigation Data</td>
<td>2004-2011</td>
<td>12.2 GB</td>
</tr>
<tr>
<td>M5CL-Multi-Sensor_Core_Logger</td>
<td>Geological cores properties</td>
<td>2010</td>
<td>16.6 MB</td>
</tr>
<tr>
<td>Moving Vessel Profiler (MVP)</td>
<td>Oceanographic data (Temp, Sal, Fluoro, Etc)</td>
<td>2007-2008, 2010-2012</td>
<td>3.75 GB</td>
</tr>
<tr>
<td>Rosette-CTD</td>
<td>Oceanographic data and water samples (for nutrients, Chla O2, contaminants, etc)</td>
<td>2002-2012</td>
<td>&gt; 9.5 GB</td>
</tr>
<tr>
<td>Remote operated vehicle (ROV)</td>
<td>Underwater video recordings</td>
<td>2008-2011</td>
<td>475 GB</td>
</tr>
<tr>
<td>Sami</td>
<td>Dissolved CO2</td>
<td>2006, 2009-2012</td>
<td>26.7 MB</td>
</tr>
<tr>
<td>SX90 Sonar</td>
<td>Sonar data</td>
<td>2011</td>
<td>21.5 GB</td>
</tr>
<tr>
<td>Thermosalinograph (TSG)</td>
<td>Continuous surface temperature and salinity data</td>
<td>2003-2012</td>
<td>4.21 GB</td>
</tr>
<tr>
<td>Videoplankton recorded (VPR)</td>
<td>Underwater plankton videos</td>
<td>2003-2004</td>
<td>111 GB</td>
</tr>
</tbody>
</table>

Data collected by instruments on the vessel (hull/deck mounted or deployed from deck); and 2) Mooring data: data collected by instruments moored in the water.

For the data from each ship’s instrument, the file structure was reorganized so each directory now contains the data and the accompanying documentation needed to understand and use the data for any instrument in any given year. Documentation includes calibration and data processing information, manuals, and softwares. File directories are structured so updates can easily be made every year with the collection of new data (one folder to add per instrument, per year). Data collected by the ships’ instruments represent 1.89 TB and include a wide variety of data, from ship’s GPS position to oceanographic data, to underwater videos (Table 2). Ocean currents and CTD-Rosette data are ArcticNet-Core data and are available exclusively to all ArcticNet members for a period of three years after collection. After this period, these datasets will be publicly available. Other datasets are exclusively used by one team and will be publicly available three years after the end of a project, in accordance to the ArcticNet Data Policy. Finally, other datasets were collected in collaboration with partners and their access will be limited for longer periods, according to the partnership agreements.

Moorings have been deployed since 2002, first during CASES and continued with ArcticNet. Moorings have been deployed at more than 30 different sites in the Beaufort Sea, Baffin Bay, and Hudson Bay. Data collected include oceanographic properties, acoustic
data, currents, and sedimentation fluxes. As for ship’s instrument data, file structure was reorganized so users can access data, calibration, and documentation for each mooring for any given year. The same accessibility rules as for the ship’s instrument also apply to the mooring datasets, i.e., determined by the nature of the data.

Discussion

The PDC has reached a number of important milestones this year, including deployment of all new hardware and a new database, dramatically increased connections to other data portals through metadata interoperability, organization and back-up of the Amundsen dataset, release of the new CCIN cryospheric information and outreach website, and strengthening of linkages with other Canadian and international supporters and collaborators. A total of ten 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.

Conclusion

Continued development of the PDC is laying the foundation for effective management of polar data in Canada. Support from ArcticNet and other partners has provided the ability to build a capable and respected archive which will provide access to and stewardship of Canada’s polar research data far into the future. 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 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

We are grateful to the Polar Data Management Committee for their involvement with the Polar Data Catalogue. We also thank Cyndie Found and Diane Carley at the Canada Institute for Scientific and Technical Information (NRC-CISTI) for their invaluable help in organizing the DOIs for the PDC’s data. We would like to thank Luc Michaud, Richard Marquis, Keith Levesque, Kathy Blasco, and Sylvain Tougas who helped with the reorganization and backup of the Amundsen Data. This project was funded by ArcticNet, Aboriginal Affairs and Northern Development Canada, and Environment Canada.

References


Publications

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


