Advanced Cooperative Arctic Data and Information Service (ACADIS): Meeting our Vision
Year 2 Project Summary (NCAR/UCAR)

National Snow and Ice Data Center (NSIDC)
National Center for Atmospheric Research (NCAR)
Earth Observing Laboratory (EOL)
Computational Information Systems Laboratory (CISL)
University Corporation for Atmospheric Research (UCAR)
Unidata Program (Unidata)

ACADIS Vision and Introduction

ACADIS provides sustainable data management, data stewardship services and leadership for the NSF Arctic research community through open data sharing, adherence to best practices and standards, capitalizing on appropriate evolving technologies, community support and community engagement. ACADIS leverages other pertinent projects, capitalizing on appropriate emerging technologies and participating in emerging cyberinfrastructure initiatives.

The roots of ACADIS come from the need expressed in the Arctic Observing Network (AON) design to develop and implement a comprehensive data management strategy. The Cooperative Arctic Data and information Service (CADIS), initiated in 2007 and involving NCAR, NSIDC and UCAR, was designed to facilitate data discovery and use by encouraging the standardization and exchange of metadata through an IT structure friendly to Web services and other applications. We realized early on that the community also needed a repository for data collected during AON. We responded to this need by implementing the CADIS system within one year of initial start to collect both AON data and metadata (more information is available in the CADIS Final Report). Realizing that the need addressed by CADIS extended beyond AON, the CADIS collaborators submitted a proposal to expand services to support all Arctic investigators submitting data to NSF/OPP/ARC. We are completing the 2nd year of Advanced CADIS (ACADIS) support to the community.

This document summarizes the current status of the project and challenges in meeting the ACADIS vision. We focus on the following key elements:

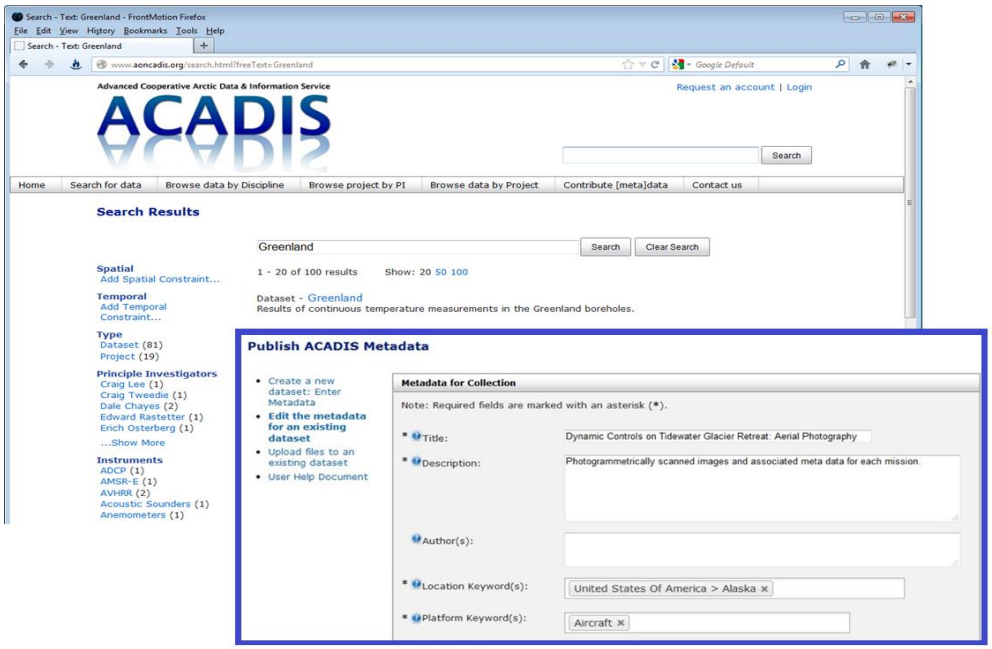
1) *Data Services:* Fundamental technical services offered through ACADIS;
2) *Support for Datasets with Special Requirements*: ACADIS handling of unique datasets;
3) *Metadata documentation, sharing, and usability:* All aspects of metadata generation, updating and utilization in ACADIS;
4) *Interoperability and initiatives:* Activities to enhance, develop and improve or data management practices and technology;

5) *Science Support:* Other community support activities as a result of expanding ACADIS acceptance and usage.

1) Data Services

The ACADIS Gateway has been developed to support NSF Arctic data coming from AON and now broadly across OPP/ARC and related archives. The Arctic Data Explorer (ADE) is an integral service of ACADIS bringing the rich archive from NSIDC together with catalogs from ACADIS and international partners in Arctic research. *Rosetta* and the Digital Object Identifier (DOI) generation scheme are tools available to the community to help publish and utilize datasets in integration and synthesis. The complementary and unique capabilities of the ACADIS Gateway and ADE are intended to offer the users choices for data discovery and access with the clear objective of increasing discovery and use of all Arctic data.

The ACADIS Gateway



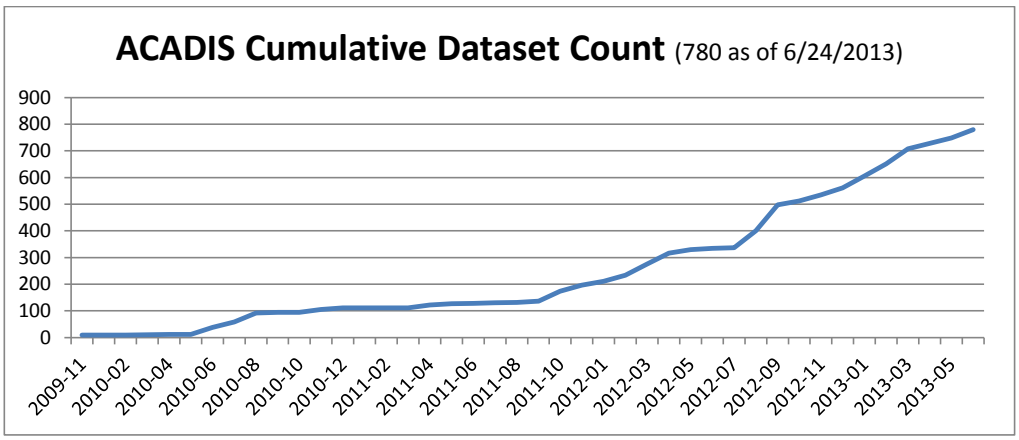
The ACADIS Gateway enables individual groups to manage their own data in a sustainable, cost effective way. The ACADIS Gateway offers full data life-cycle management services for data publication, discovery, access and preservation.

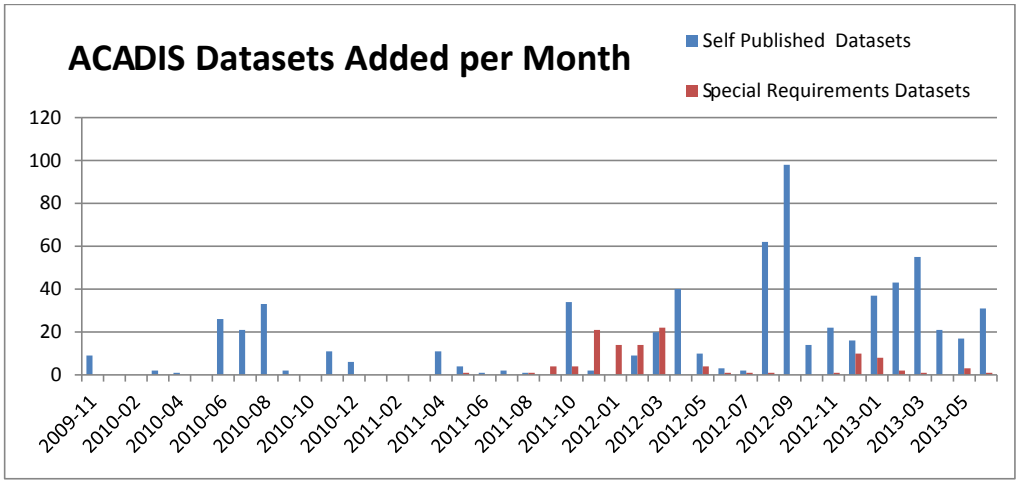
Status: The ACADIS Gateway, (http://aoncadis.org), was envisioned as a shared, central clearing house for NSF Arctic and related data from other sources for NSF researchers. It has evolved substantially over the last several years in terms of holdings, capabilities, and its ease of use. Current holdings represent over 664 self-published data collections, including 114 SRDC data collections brokered from the EOL archive. These contributed collections represent the efforts of over 120 Principal Investigators (PIs). At present, the ACADIS Gateway provides services for data-publishing research teams and individual PIs, data consumers, data curators and support personnel, and NSF or other staff interested in verifying publication of research results. These services currently include:

- A self-publishing capability that allows researchers to easily and effectively author and update rich metadata and upload data files of any type into ACADIS. This is an important element of sustainability in terms of dealing with many data-generating projects;
- Browse and search capabilities. We have improved search capabilities enormously over the course of the project, and now have a highly flexible, scalable, faceted search, built upon the industrial-class Apache SOLR component;
- User authentication and group authorization, with provisions for federated identity via the OpenID framework. While ACADIS data is generally open, publishing data into the system requires good security and, over time, we expect to support data collections with restricted access.
- Tools for ACADIS data curators and support staff;
- The federation of ACADIS catalogs with other data systems or aggregators;
    - The ACADIS Gateway has been harvested by the Arctic Data Explorer (ADE), NASA's Global Change Master Directory (GCMD), and the USA National Weather Service (NWS) Global Information System Center (GISC) nodes via the WMO Information System (UN/WMO WIS). WIS is considered to be a primary feeder for the Global Earth Observing System of Systems (GEOSS).
- Federation functionality that supports ingest of external catalogs, such as THREDDS and DIF (GCMD). We are currently harvesting EOL catalogs and are in progress of harvesting some of the NSIDC holdings. The ACADIS Gateway search capabilities index all of these catalogs, thus providing a powerful federated search capability;
- Basic support for digital data collection citations with a DOI;

- Digital Preservation support via periodic archiving to the Chronopolis Digital Preservation System;
- Community support resources, including an integrated help document, and guides for preparing NSF data management plans and publishing data into the system;
- An internal issue tracking system for managing end user technical problems.

Since the inception of ACADIS, efforts at CISL have focused on expanding and adapting the data provider interface for the ACADIS Gateway to handle submissions from all NSF Arctic projects. This included search by NSF proposal number and linking of consecutive grants. The focus in Year 2 has been on providing an easier and faster data submission process for data providers and ACADIS data curators. Key accomplishments include provider interface and workflow enhancements based on feedback from usability sessions, end-users and other stakeholder interactions. Feature highlights include improved keyword list selections, error handling, multi file upload, file delete, dataset copy options, curator tools, and font size updates to improve readability and appearance. The OAI based metadata services provided by the ACADIS Gateway have been upgraded to support broad scientific community search and discovery of ACADIS holdings.





Data provider activity has grown in project year two. In the first six months of 2013, an average of 40 new datasets have been added to the ACADIS archive, representing both self-published datasets created via the ACADIS Gateway and special requirements data collections created through curator assistance at EOL.

Challenges: Providing a self-service data publishing platform requiring minimal curation oversight while maintaining rich metadata for discovery, access and preservation is challenging. The Gateway has known limitations in dealing with the wide variety of datasets even within the current ACADIS scope. Efficiency is a challenge for both data providers and data curators, given low automation of processes such as notifications and reporting of collection change logs which affect provider confidence. Capturing necessary metadata is a challenge; keeping the barriers for data upload must be balanced with requiring enough metadata to make the data useful to other researchers (i.e. different requirements in different disciplines). Metadata records are currently captured for datasets, file collections or datasets, rather than for individual files themselves. Data discovery and re-use may require finer grained, file level metadata capture. Automated extraction of file metadata from highly varied formats may be required to sustainably support data publication. Further, some key elements of the metadata information are being stored as free text fields; these will need to be transitioned into structured fields for use with automated data citations, search, and metrics for data providers. Note that some of these metadata issues are being addressed in the *Rosetta* work, described below.

While efforts to date have focused on data provider needs, and that will continue to be a priority, we also need to shift some attention to streamlining data access and addressing usability concerns of accessing multiple data systems and varied file formats. Our primary

interface to Gateway services is via the graphical user interface (GUI). We must continue to move in a direction toward developing service oriented (SOA) capabilities, as appropriate, to address the changing needs of the user community and to enable the community to leverage the ACADIS Gateway holdings.

The distributed nature of ACADIS and divergent understandings/semantics of "projects" and "datasets" are current realities to be overcome when implementing a consistent automated system. We need to work on the implementation of how the ACADIS Gateway and the other project components (e.g. ADE, *Rosetta*) exchange metadata and support user access to a broad community of researchers. ACADIS needs to get a formal contract in place for Chronopolis services.

Arctic Data Explorer

Status: From the early days of our work, we have always envisioned the role of ACADIS to include supplying access to metadata catalogs from other data systems. This reflects a core ACADIS principle: we want to expose and provide access to data to as many people as possible. To this end, the Arctic Data Explorer (ADE, http://nsidc.org/acadis/search) is a new data discovery service component of ACADIS that will be integrated with the NSIDC web presence to provide additional visibility for ACADIS collections and to broker metadata across repositories through a single search interface. Bringing together the multitude of existing data repositories containing Arctic data, metadata brokering is a necessary component of a coordinated Arctic research cyberinfrastructure. With brokering, individual repositories can maintain independent systems designed to meet the needs of local data user communities while simultaneously contributing to global initiatives.

The ADE development trajectory has been driven by input from the ACADIS Data Advisory Committee (ADAC); the group's clear priority was a useful, robust search. Access to additional metadata catalogs was deemed secondary by the ADAC, though an exception was made for datasets from the National Oceanographic Data Center (NODC). Other catalogs will be added in subsequent versions. Features of the ADE currently include:

- Access to metadata catalogs from NSIDC, the NODC, the Norwegian Meteorological Institute, the ACADIS Gateway, and the EOL Computing Data and Software (CDS) Facility;
- Spatial, temporal, and text-based search across catalogs;
- Review select metadata from datasets matching query and follow link to access data;
- Relevance ranking of results based on a usability-driven weighting scheme;
- Bookmarking of search parameters to repeat later;
- An Earth Science Information Partners (ESIP) OpenSearch access point for automated queries and metadata access;

Version 1.0 of the ADE was released on April 30, 2013 to positive reviews at the Arctic Observing Summit in Vancouver, BC. For the ADE, ACADIS has partnered with the Italian Center for National Research-funded Earth and Space Science Informatics (ESSI) Lab on development of the 'GI-Cat' metadata broker. GI-Cat is a joint development environment with ESSI-Lab and NSIDC software engineers committing code to a common repository. Stand-alone libraries and components created by NSIDC as part of ADE have been published as open source projects on GitHub (see section 4 for more details) or reuse by others. The NSIDC NASA DAAC, sharing the same core search infrastructure, has also contributed greatly to the development of the ADE.

Challenges: Search tools like the ADE are limited by inconsistency, poor form, and content gaps in metadata records. While metadata standards are a first step towards consistent content, more work is necessary. Also, the ability to discover a dataset does not imply that the data can be directly downloaded. Each harvested repository has unique system requirements, work-flows, and procedures for accessing data. Seamless discovery and download requires the adoption of common data sharing policies and potentially technical work from each of the harvested repositories. Also, performance has suffered as we have added records.

Data Citations, Content Visualization (e.g. GIS Tools)

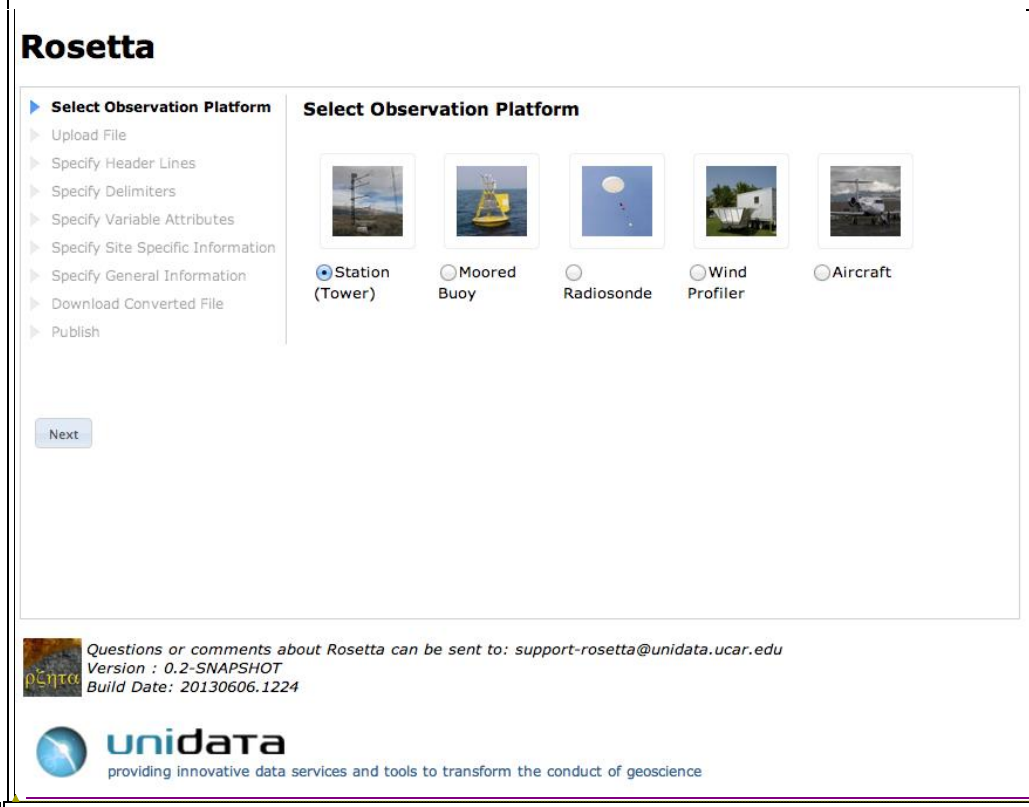
Status: EOL manages the Arctic data in its archives, and has updated its database to facilitate harvesting of the complete metadata. Alternative data access via the GIS MapServer tool is also provided. EOL initiated the creation and registration of DOIs for Arctic datasets in EOL archives, and is working closely with CISL and NSIDC in creating consistent data citations. Data from a five-year project in the Bering Sea, and a one year synthesis project encompassing data from the Chukchi and Beaufort Seas, are being readied for harvesting to ACADIS; harvesting will commence when the projects close out and password protection on the data is removed. Some very large datasets have required special attention to maximize archival, subsetting and access capabilities, and these have been managed by EOL.

Challenges: EOL data servers are separate from those used by CISL. Work needs to be done to improve interoperability, and the discovery and access of the EOL Arctic data/metadata from the ACADIS gateway. EOL- will work with CISL to be prepared to archive very large holdings of model output, including access and delivery capabilities. An automated DOI preparation capability needs to be put into place for all Arctic datasets.

CADIS, and now ACADIS, has had three simultaneous mappings of sites with data on the main ACADIS Gateway page. It has been a challenge to keep all three of these maps updated and in sync with new additions to the archive. A decision needs to be made on whether one or all of these links should remain functional. Linking metadata properly in order to maintain geo-

location information and data links and work to increase the extent of metadata and data viewable with them must be weighed against other ACADIS priorities.

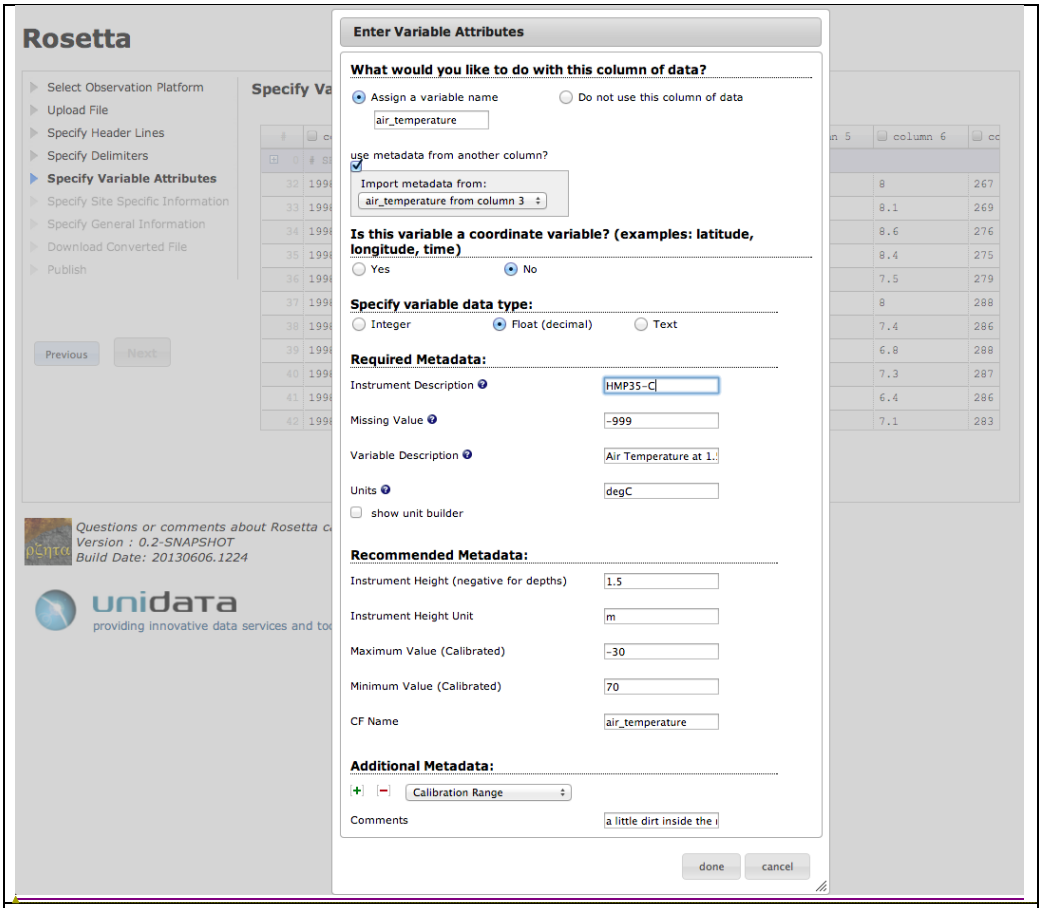
Rosetta



Rosetta can generate CF- requiring the users to intimately know the CF specifications. For example, the image above illustrates how users select the discreet sampling geometry type of their datasets.

Status: Work on *Rosetta*, the new data transformation tool of ACADIS, is progressing well. The task of generating a copy of all of the metadata and file parsing information that a user entered in the process of reformatting their dataset (called a "transaction receipt") has been completed. Once created, a transaction receipt can be uploaded to *Rosetta*, edited if necessary, and applied to a new dataset from the same site, or datasets with similar formatting (e.g. soil temperature profiles from different sites). This feature enables batch processing of data files, as well as allowing for new datasets to be added with greater ease. *Rosetta* is being tested at

NSIDC and EOL, and will be available as a community tool on the ACADIS Gateway later this year.



Collection . samples.

Challenges: A white paper describing the challenges of sharing observational data and how *Rosetta* can help can be found at: https://www.unidata.ucar.edu/software/pzhta/files/pzhta_whitepaper_201304.pdf. In addition to the general challenges outlined in the whitepaper, one of the more demanding and difficult aspects for the next year of *Rosetta* development is addressing the need to handle more diverse and complex ASCII formats. Currently, ASCII files represented as comma separated values (csv) with one observation-time per row (as most commonly produced from data

loggers) are handled by the web-based front-end to *Rosetta*. As development moves forward, the front-end will need to be improved to handle the more complex ASCII formats, such as those that include embedded comments outside of the header block, files containing multiple sites that sample different data variables, and files with blocks of height - variable tables as functions of time (such as those from vertical profiling sensors). Unidata continues to work on the web-based interface with NSIDC to make it more user friendly, and the importance of this will only grow as the complexity of the ASCII formats handled by *Rosetta* increases. The development of a standard data template for social science data remains elusive at this time.

2) Support for Datasets with Special Requirements

Status: While good progress has been made in automating the upload process through the ACADIS Gateway, there continue to be unique datasets that require special attention. EOL takes the lead for these special cases. These datasets can be 'large' in both total size and number of files or small but very unusual in format or other attributes. These may include types of data which require password protection, such as human subjects, endangered species, or DNA sequencing with GenBank accession number, as well as products such as GIS layers, images and web-based data visualizations for which relational structures may need to be maintained. Such datasets are not readily amenable to automated upload through the ACADIS Gateway and are archived at EOL. The metadata for the EOL Arctic data holdings are automatically harvested and processed for ACADIS. This harvesting process is complete for data from the more recent field projects. EOL, NSIDC, and CISL specialists work together to update the ACADIS metadata profile to better define and utilize the data from diverse disciplines.

Challenges: ACADIS accepts that some special handing may be required as a normal part of its community support services. Our biggest challenge is to make sure that these special datasets (e.g. large volume, many data files) archived at EOL can be accessed via the ACADIS gateway. Work will continue with CISL via the ACADIS Workflow subgroup to increase the capture of more special requirements data. EOL must remain agile with respect to special requirements for NSF supported process studies or individual investigator projects with new and unusual data requirements. Making sure that the new dataset metadata profiles are complete requires constant attention and interaction with some investigator teams with particularly unique datasets. Backfilling previously-received datasets with updated information and adding new metadata from older project data, while a continuing opportunity to provide the most complete Arctic database, requires priority assessment against other elements of ACADIS support.

3) Metadata Documentation, Sharing and Usability

Metadata is at the core of ACADIS activities, from capturing metadata at the point of data submission to ensuring interoperability with larger efforts, providing data citations, and supporting data discovery. ACADIS metadata efforts include: 1) Evolution of the ACADIS metadata profile to increase flexibility in search; 2) Documentation guidelines; and 3) Metadata standardization efforts. These activities are coordinated by the ACADIS Metadata Subgroup, which focuses on upgrading and assuring the continuity and quality of metadata efforts.

Metadata Subgroup

Status: The Metadata Subgroup, comprised of ACADIS partners, assessed the current state of metadata across the project, compiled and prioritized necessary metadata work, established minimum metadata sharing requirements for project partners, created a data dictionary and requirements for implementing data citations in ACADIS, and began work to determine metadata needed to support file-level usability.

Challenges: The list of needed metadata work is formidable. The biggest challenges are coordinating priorities to make progress while meeting ongoing community needs, aligning progress with technical development and dependencies, and standardizing metadata among partner data holdings. Also, new multi-disciplinary diverse datasets that are submitted to ACADIS may require further refinements to the metadata profiles. Additionally, dependencies on metadata work need to be identified by the partners and then coordinated with the ACADIS Gateway development team for implementation. With NSFs stated priority on documenting new datasets, the creation and standardization of metadata for existing and 'legacy' datasets is a challenge.

Documentation

Status: ACADIS data providers intermittently submit additional documentation beyond the metadata profile as a readme.txt files. They are given guidelines for what their documentation should cover in Appendix B of the ACADIS Data Provider's Guide (http://www.aoncadis.org/media/ProvidersGuide.pdf). The Data Provider Guide is updated by CISL as needed to highlight new gateway capabilities.

Readme files are a step towards more comprehensive documentation. For stewardship and long-term preservation, data must not only be technically secured over the long term, but also be usable by others outside the discipline that produced that data now and into the future. This means that data must have adequate documentation for human usability in addition to being technically accessible and usable. NSIDC and EOL have over 40 years of combined experience working with NASA, NOAA, and NSF to develop comprehensive documentation. In addition, NSIDC over the last year investigated the latest research and standards developments in the field to ensure future reusability of ACADIS datasets.

Challenges: Determining what documentation is necessary for new users of a dataset, obtaining that from busy PIs, and storing it in an accessible, usable way are all challenges. Also, compliance with still-emerging metadata standards, and managing standards compliance is difficult, labor intensive, and ongoing. Not all ACADIS data providers submit a readme.txt file to accompany their datasets, even though it is recommended. To address this and other issues, Curators need system notifications for events such as data upload. Manually ensuring that PIs submit standardized, compliant content and metadata isn't scalable given the increased scope of ACADIS. The experience of NSIDC and EOL in obtaining proper documentation from scientists will be brought to bear on this issue.

4) Interoperability and Initiatives

ACADIS is engaged in multiple collaborations and partnerships to enhance, develop, and improve data management practices, technology and cyberinfrastructure within the Arctic science communities. These are summarized for the ACADIS collaborators and include external partners participating in the activity

Arctic Data Explorer. Partners: EarthCube, CNR ESSI-Lab, NASA, NODC, NMI

Status: The ADE has been developed with the philosophy that inclusive collaborations produce better code. With this in mind, NSIDC is partnering with the Italian CNR ESSI-Lab and has contributed code back to the GI-Cat effort. Dedicated to the Open Source software model, we have made additional libraries accessible through GitHub (http://github.com/nsidc/gi_cat_driver). The NSIDC NASA DAAC has moved forward with a new NSIDC data search based on the same infrastructure; search modules developed by ACADIS and the NSIDC NASA DAAC are reusable by both projects and are contributed back to the ESSI-Lab infrastructure where appropriate. The ADE has also benefitted significantly from a partnership with EarthCube through an Eager award to advance metadata brokering. EarthCube will be able to put these metadata broking lessons to good use in future initiatives. Fundamental to the ADE is the understanding that those looking for data don't necessarily know (or need to know) which agency funded the collection or creation of a dataset. As a proof-of-concept, the ADE searches across NSF, NASA, NOAA, and internationally funded data sets to highlight the value of interoperability for increased data discovery.

Challenges: Long-term integration and contributing to community progress need to be balanced with delivering short-term value to the scientific community. Additionally, cutting edge efforts don't always mature into useful, stable tools. Both up-front and ongoing evaluations of initiatives are needed to maintain the biggest return on investment possible.

Arctic Social Sciences Data. Partners: Arctic Social Science Program (ASSP), Exchange for Local Observations and Knowledge of the Arctic (ELOKA), Council on Library and Information Resources (CLIR)

Status: The handling of social sciences data within ACADIS is still in its infancy. The project does not presently have sufficient expertise to know how to properly handle such data. We have engaged the ADAC in determining the best approach for handling these unique and sometimes sensitive data. One of the ADAC members is a long-time Arctic social science researcher and his guidance has been invaluable. NSIDC is in the process of hiring a Post-Doc with expertise in social sciences data. Funding has been coordinated through the Council on Library and Information Resources (CLIR) program, ELOKA, and ACADIS. ELOKA, an Arctic Local and Traditional Knowledge data management project led by NSIDC, will be submitting a request for supplemental funding through the Arctic Social Sciences Program. The goal of the post-doc will be to identify existing resources (e.g., EOL and NSIDC experience supporting projects with social components), determine if gaps exist, assist in filling gaps, and set the stage for future work grounded in understanding of, and connections with, needs of the scientific community.

Challenges: The best approach may be to handle social sciences data as an overlapping component of ACADIS. While discussion has been ongoing regarding developing this link through the ELOKA effort, a framework has yet to be solidified.

Permafrost Showcase Data Integration Pilot Study. Partners: International Permafrost Association (IPA), Global Terrestrial Network-Permafrost (GTN-P)

Status: The idea behind this pilot study (i.e. showcase) is to create a scalable model for supporting long-term data utility for analysis efforts. Essentially, the permafrost showcase seeks to create a standardized collection of ACADIS data sets that are reusable. This includes proper and complete documentation, simple functionality using common tools, and easy-to-build products. Ultimately, the goal is to generalize the showcase model for application with other ACADIS data. With standardized data we can offer value-added tools and provide further incentive for those submitting data to ACADIS. The *Rosetta* reformatting tool described previously is used to unify the data into a single format. Permafrost data were chosen in consultation with ADAC, and we engaged an NSIDC permafrost scientist to ensure connection to science needs. The showcase addresses several key components of the ACADIS vision, but it particularly hits on data stewardship and community engagement through data reuse. To have permafrost scientists excited about following data management best practices (e.g. describing their data in metadata with standard terms and using standard or well described formats), we need to build on work the permafrost community has already done. Therefore, we are engaging with the IPA and the GTN-P.

Challenges: International efforts, while extremely valuable, can proceed slowly; there is a challenge in aligning these extended efforts with short ACADIS timelines. Limited metadata presents a challenge with the showcase; the current required metadata fields are high-level and don't include information needed to integrate permafrost data.

Interoperability, Developing Archival and Access Capabilities for New Process Studies. Partners: Arctic Ocean Observing System (AOOS), GCMD, North Pacific Research Board (NPRB)

Status: ACADIS engineers and curators from EOL and NSIDC have partnered with the Arctic Ocean Observing System (AOOS) and other organizations to broaden the discovery and use of these data through metadata sharing initiatives. ACADIS submits updates to the GCMD catalog on a regular basis, and supports GCMD classifications in its metadata profile. EOL provides support for the Bering Sea Project, an ecosystem study jointly administered by the NSF and the North Pacific Research Board (NPRB). EOL provides data management support for NSF and NPRB funded process studies and synthesis projects, including the Pacific Marine Arctic Regional Synthesis (PacMARS) and utilizes AOOS tools to access oil industry data important to the project data integration.

Challenges: ACADIS offers a strong framework for archival and access to the diverse data coming from Arctic investigators. Linking to other archives (e.g. AOOS) and providing customized support for focused process studies (e.g. Bering Sea Ecosystem Study, PacMARS, and the Distributed Biological Observing). We must identify where ACADIS can best enhance and focus on those areas of support (e.g. NSF funded field projects) within the resources available.

Development of New Collaborations in the OPP Antarctic Division

Status: ACADIS will be represented by EOL and CISL PIs at an upcoming coordination meeting organized by the Antarctic Data Consortium. NSIDC also brings a long history of Antarctic research and data management to the discussion.

Challenges: ACADIS (including CADIS) has evolved over 6+ years to provide a single point of access for the NSF supported Arctic science community to upload data, metadata and documentation from many discipline in diverse formats. It remains to be seen whether this support paradigm has applicability in the Antarctic and if there are concrete next steps to unify Arctic and Antarctic data support.

Earth Cube and Collaborations Supporting Technology Integration for the ACADIS Gateway. Partners: Earth Cube (proposal in preparation), NASA/GCMD, WMO, NCAR/ESIP Workshop/RCN

Status: Successful scientific data management requires advocacy, understanding the user community, and awareness of best practices, standards and general trends of the broad scientific (and other) data management service providers. CISL has been heavily engaged in NSF's Earth Cube program, contributing to one of the community groups and the general roadmap development work. CISL demonstrated the ACADIS Gateway to multiple groups at an Earth Cube virtual meeting, noting its relevance to the general Earth Cube problem, the need for self-publishing capabilities for "long tail" science communities, and as an example of what has worked well for Arctic research needs. The ACADIS Gateway's discovery level collection metadata has been harvested by NASA/GCMD. In a similar vein, the US NWS has harvested ACADIS metadata as part of their work to serve as GISC-WASHINGTON in the global WMO Information System (WIS) network. We have engaged with the ESIP community and will broaden our interactions with ESIP in summer 2013.

Challenges: Addressing the broad needs of scientific data management is an enormous undertaking. To be effective and efficient in our collaborative work, we must identify focus areas where ACADIS can provide value. We must find a proper balance between servicing the needs of the NSF ARC community and engaging in broader community initiatives that bring value across scientific disciplines and domains.

Long-Term Data Preservation. Partnership: Chronopolis

Status: A key component of ACADIS is to assure the long-term preservation of the legacy from years of Arctic research. CISL has performed annual preservation snapshots of the ACADIS Gateway holdings including capture of the metadata database, to the Chronopolis preservation system. This area is supported by leveraging work underway for the Library of Congress initiated Chronopolis project (http://chronopolis.sdsc.edu/.)

Challenges: The current process, while sufficient, is staff intensive and has minimal automation, potentially leading to inconsistent results. Further, current procedures are applied to a sub-set of the overall ACADIS holdings and the EOL Special Requirements holdings are not easily included in the preservation archive. Lastly, ACADIS needs to get a formal contract in place for Chronopolis services.

Tool Sharing and Adaptation Initiatives

Status: *Rosetta* is gaining broad acceptance with many other groups in the geo-sciences that face similar problems. Data providers in marine sciences, hydrology, oceanography, and atmospheric sciences have expressed interest in incorporating *Rosetta* into their data infrastructure. This will continue to aid the inter-operability across domains as well as facilitate inter-operability within the Polar sciences via the ACADIS project. This inter-operability hinges

on the use of NetCDF-CF as an exchange format. This format has gained acceptance in many scientific communities and can be natively used by many scientific applications.

Challenges: Multidisciplinary adoption and use of any cyberinfrastructure will remain a continuing challenge as different communities have different practices and use different tools to conduct their research. Until conventions for borehole data are developed and broadly used by the permafrost community, we need to be adaptive and inclusive to insure inter-operability amongst disparate datasets for permafrost and other data collections. The *Rosetta* tool is adaptive and can accommodate many different platforms and data structures, but it currently is not exhaustive.

5) Science Support

Science support activities include data provider and user support through answering questions and requests, offering data management consultations, and data management plan reviews.

Provider and User Support

Status: Year 2 saw a significant increase in emails and phone calls over Year 1. Along with a greater number of community questions, the content and nature of requests has shifted. We saw increased demand for data citations and DOIs, and got our first emails regarding data use rather than data submission.

NSIDC significantly refocused and restructured data provider and user support, collectively referred to as "Community Support," in Year 2. This restructuring standardized the user experience, increased transparency, connected ACADIS Gateway development efforts more closely to the needs of the Gateway user, and better accommodated the distributed nature of the ACADIS partnership. All incoming community questions and requests are now routed through a centralized communications management system (http://Zendesk.com). Practices are reviewed and handled by the recently-formed Community Support Subgroup.

Challenges: Bridging between existing user support systems and the variety of workflows and policies from each partner is the biggest challenge to streamlining communication and user experience. Disjointed communication and services can confuse engagement with the ACADIS community. The growth of ACADIS requires scaling of community support efforts, and increasing ACADIS name-recognition means that requests are shifting in context from data submission to discovery and access. With the new centralized Zendesk system, there are new questions about what metrics to collect and what common procedures are needed.

Support to potential and active NSF funded project teams and individual investigators

Status: The technicians and post-docs on the teams of NSF funded PI laboratories often handle data from the field and organize the files for submission to the archive. EOL reaches out to these data providers to assist them in metadata and documentation procedures and best practices resulting in their submission of data into ACADIS. EOL provides guidance to potential NSF PIs developing field project plans; EOL and NSIDC review data management plans for proposal submitters. It is also important to engage other members of the Arctic and broader polar communities in working towards improved coordination and unification of practices and approaches. EOL and CISL will represent ACADIS and be involved with initial discussions to coordinate Arctic-Antarctic data management

Challenges: The focus of this support must be the near and long-term education of potential, new and ongoing ACADIS users. This begins with the NSF requirements for a Data Management Plan as part of any proposal and extends to direct support related to uploading metadata, data and documentation to the archive. There are a wide variety of metadata standards, data formats, archives and access points for Arctic data. There is a need to engage the Arctic community across agencies, institutions and international groups to work towards unified and consistent archival, access and long-term stewardship of important resources.