

Advanced Cooperative Arctic Data and Information System (ACADIS) Year 2 Consolidated Work Plan

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)

Introduction

The ACADIS Team has prepared a consolidated work plan for activities proposed for the second year of the project. The period of performance is from July 2012 through June 2013. We have divided the work into tractable and achievable milestones and they are detailed below.

Our primary objective continues to be to provide data management support to the NSF/OPP/ARC group of investigators. This includes user support to help the investigators meet their data management requirements and to provide a safe, secure, robust and long-term archive for the rich dataset coming from their work in the Arctic and surrounds.

In Year 1 of the ACADIS project we organized our work around the three major data lifecycle elements of: 1) Acquisition and Ingest; 2) Access and Integration; 3) Archiving and Preservation. However, experience from Year 1 and counsel from the ACADIS Data Advisory Committee (ADAC) has emphasized that it is often important to organize workflow around a more integrative view that cuts across these three elements. This is essential to meet the ADAC priority of consistent, discoverable and reusable data. As a result we have organized a set of major Year 2 milestones into eight areas which often span the entire data lifecycle. All of these are important, but after discussion between the PIs and feedback from the ADAC, we conclude that the first four areas are particularly critical: Dataflow and Workflow; Standardization Showcases; Metadata Enhancement; and Discovery and Access. Following an overview of project management and senior personnel, we outline these milestones, indicating for each the lead and supporting organizations.

Project Management

The ACADIS Team, comprised of about 30 individuals is distributed among NSIDC, NCAR/EOL, NCAR/CISL and UCAR/Unidata. Project management is handled by the PIs from these groups. Overall direction is the responsibility of NSIDC.

The PI team meets once per month to discuss project updates, review and address existing and new milestones, and consider new opportunities. Outcomes and decisions are then reported back to the full ACADIS team. We implemented the Basecamp Project Management support tool to consolidate milestones, calendars and project documentation in one location for all 30 team members to share and review. Complementarily NSIDC and CISL are both using an agile scrum software development model as an integral part of their work. This allows for detailed updating, tracking and reporting on a diverse set of software developments underway to support the ACADIS project.

In Year 2, EOL will contribute to ACADIS by focusing on continued user support, updating metadata for all EOL Arctic project data holdings, integration of project metadata into ACADIS, and supporting users with special datasets for inclusion in the ACADIS archive. CISL will continue to lead ACADIS Gateway development with a focus on data collection citation, data publishing work flows, data preservation, data product services, metadata interoperability and use metrics reporting. NSIDC's roles in ACADIS are to continue to coordinate the overall team, provide science leadership, lead the data stewardship activities in collaboration with UCAR, lead the ADAC-recommended data "showcase" effort (described below) and increase the visibility and usability of NSF ARC data in conjunction with Arctic data from other U.S. agencies (The Arctic Data Explorer). Unidata's contributions to ACADIS will continue to be in the format standardization, data translation, and visualization area in support of the ACADIS system and the data life cycle elements identified in the introductory section.

Senior Personnel (NSIDC)

Mark Serreze (0.05 FTE) is the PI on the NSIDC end and the overall science lead for the ACADIS project. Mark Parsons (0.3 FTE) acts as the NSIDC project manager, overall project coordinator, and community liaison to other national and international Arctic data efforts. Florence Fetterer (0.3 FTE) provides connections to the operational data community and leads the showcase effort. Lynn Yarmey (1.0 FTE) is the senior Data Curator. She oversees the overall data lifecycle and integrates all of the pieces of the ACADIS project through ongoing communication, organization and translation of "big picture" goals into day to day activities. Details on personnel and staffing for NSIDC are provided in the separate budget justification.

Senior Personnel (UCAR/NCAR)

James Moore (0.15 FTE) is the PI and Project Manager for the UCAR/NCAR ACADIS effort. Steve Williams (0.2 FTE) is responsible for the direction of the EOL CDS Data Management Group (DMG) participation in ACADIS. CO-PI Don Middleton (0.08 FTE co-sponsored) manages the CISL Visualization and Enabling Technologies Section (VETS). Eric Nienhouse (0.7 FTE) is the CISL project manager for the ACADIS portal development including management oversight for the CISL Agile development team to prioritize new features and enhancements based on user community, ACADIS PI and ADAC guidance. Mohan Ramamurthy (0.05 FTE) is Co-PI and Director of the Unidata Program. Ramamurthy's salary support for ACADIS will be co-sponsored by a separate NSF/AGS award to Unidata. Moore is responsible for overall NCAR/UCAR project coordination with NSF and any other advisory panels. All three UCAR PIs and component project managers are responsible for coordination between UCAR/NCAR and NSIDC.

1) Data Flow and Workflow

1.1) Metadata submission interface/submission process (CISL, NSIDC, EOL)

CISL will enhance data provider submission interface workflows to increase efficiency and quality of metadata record authoring for broader community use. User feedback and usability studies have generated valuable feedback regarding improvements to the data submission interface. Work areas include improving key areas of the user interface workflow based on data provider disciplines and scientific focus enabling easier and faster metadata capture. CISL will

also participate with EOL and NSIDC in defining, documenting and implementing the overall process of data provider submissions including project planning, data gathering, metadata and data publishing, archiving and access. CISL has included .75 FTE in the budget to support this activity.

1.2) Document and implement machine/human workflows (all)

The ACADIS team will first define a human curatorial and quality review process for all submitted data to ensure reasonable documentation and consistency. The process will be implemented early in Year 2 and will involve significant effort from data specialists/curators at NSIDC and EOL. EOL will include .2 FTE for the support of all aspects of this milestone.

1.3) Expand collections (EOL, CISL, NSIDC) -

EOL currently maintains project-specific archives for more than 1800 datasets. The metadata will be included in the ACADIS Gateway, enabling all these data to be discoverable and orderable through the system. In preparation for the sharing of these datasets, the metadata parameters in the EOL database will be reviewed and updated as needed. An automated EOL THREDDS server service currently in use for metadata sharing will be expanded, and EOL will ensure that the metadata are complete and consistent with the ACADIS metadata profile. EOL will also support NSIDC with the harvesting of discovery level metadata from other supporting Arctic archives, such as the National Ocean Data Center (NODC). There is .35 FTE of EOL staff time included for completing this milestone.

2) Standardization Showcases

Originally conceived as an approach to addressing impediments in data use highlighted by the ADAC, the data showcase concept has evolved into a multi-faceted package of tools and services with the goal of bridging the practices of field scientists with the needs of data users.

2.1) First Data Showcase (NSIDC, all)

Working with the ADAC, permafrost borehole temperature data was defined as a first data showcase. The Year 2 objective is to complete this showcase. We will first complete in-progress work and discussions with the International Permafrost Association (IPA) regarding field templates, metadata elements, and data logger format recommendations. Second, implementation work will begin on the remaining showcase components. Feedback from the ADAC supports creation of a pre-populated metadata template for permafrost data submission to the ACADIS Gateway; this will not only save data contributors time, but will further the standardization of metadata especially for search and discovery. A cleaned, quality controlled, and standardized Arctic permafrost data product that will be developed using data currently held by ACADIS partners. Additionally, pilot Archival Information Packages (AIPs) will be created in support of preservation and long-term reuse of the data products. All of these services and tools will be referenced in example text for use in NSF Data Management Plans for projects collecting permafrost data. Finally, the lessons learned from the permafrost showcase will be identified and generalized with an eye towards tackling a second community-identified data type as a next step in addressing the 'logger problem.' (i.e. data from automated loggers with widely differing formats and often little or no quality control)

2.2) Data Translator (standardizer) (Unidata, all)

Development of the Data Translator tool in Year 1 focused on the ability to convert unstructured data from data logger output (ASCII format) into standardized Climate and Forecast 1.6 (CF-1.6) compliant netCDF files. The netCDF format is used as an intermediate format to store data and metadata; it allows for easy translation of ACADIS data into structured ASCII and facilitates data search and discovery. Findings from the ACADIS Metadata subgroup have been used to enhance the CF-1.6 specification where necessary, such that files produced by the tool will conform to both the CF-1.6 standard as well as the ACADIS metadata schema. A web based front-end to the tool is under development and is designed to make the translation process as easy as possible for Arctic PIs, and requires no understanding of the netCDF format or the programming libraries used to create the files. Efforts in year two will focus on defining a standard structure for the ASCII output by the Data Translator (no such standard exists), as well as provide the capability to output into spreadsheet formats (.xls, .xlsx) for direct use in Excel, Calc, or other spreadsheet-based software packages. While special attention has been given to translating Arctic temperature borehole data (see data showcase milestone above), in Year 2 the Data Translator will begin to support other data types and observation platforms. Unidata proposes .5 FTE to support this expanded activity in year 2.

Unidata will provide use documentation for the Data Translator. We plan to produce videos, as well as written documentation, to reach a broader audience. The documentation will need to be tailored to the specific implementation of the Data Translator (e.g. using the web via the ACADIS portal, software downloaded to the user, etc.) Unidata will also provide documentation on how to install and maintain the Data Translator for system administrators. .2 FTE on Unidata support is needed to address this activity.

2.3) Demonstration Visualizations (Unidata)

Unidata will create visualizations of ACADIS datasets to illustrate the capabilities of the ACADIS Gateway and Data Translator. These visualizations are intended to show the value of integrating disparate Arctic data types into a single view or novel ways (i.e. 3D). Visualizations will be created in selected file formats (e.g., GIF/JPEG, Quicktime, KML/KMZ, and IDV bundles), with goal of providing such examples alongside the original and translated ACADIS data. .2 FTE is included from Unidata to continue the visualization effort.

2.4) Identify Next Data Showcase (all, ADAC)

Depending on the results of the first data showcase, we will work with ADAC to identify the next showcase to make further progress to solving the “logger problem” more generically.

3) Metadata Enhancement

Clearly defined and consistent metadata are essential for the discovery and usability of data across archives. The value of good metadata accompanying datasets increases with time, as institutional memory decreases and research methods evolve.

3.1) Improve metadata for additional functionality and interoperability (all)

Building on the Year 1 system integration progress, in Year 2 we will confront the remaining challenge of full interoperability: metadata. Field science has traditionally been conducted using local norms and ad-hoc metadata standards. Now that NSF Arctic data are closer than ever to being searchable through a single portal, we must find commonality in those ad-hoc practices to allow a truly functional discovery mechanism. Merging existing controlled vocabularies and

standardizing dataset domain categorizations are two priorities for Year 2 metadata work. However, our expectations are higher. To fulfill the requirements for application of data citations and Digital Object Identifiers (DOIs), to offer full archival services, to implement dataset versioning, to provide flexibility for inclusion of community metadata standards (such as the IPAs metadata form we are working with for the Data Showcase), and to maintain compliance with international standards organizations, we need to further augment the existing ACADIS metadata profile and balance additions with realistic expectations for data submitters. We also need to expend significant effort to bring historical metadata up to date with current requirements in both content and format. We have an updated metadata profile in draft form that needs to be vetted, split into implementation phases, and put into production. Metadata standardization work across the partner organizations will be necessary; that work has begun with the recently reconvened metadata subgroup and will be continued. This will be a significant but very valuable and necessary effort. It will require the effort of EOL data specialists, NSIDC Data Curators and student assistants.

3.2) EOL ingest to CISL, metadata sharing between partners (EOL)

EOL will work with NSIDC and CISL to put the final touches on an expanded metadata profile for ACADIS datasets. This profile will be used to facilitate the sharing of datasets among the ACADIS partner institutions, as well as to define a baseline of information on datasets being submitted to the ACADIS archive. The ACADIS metadata profile will serve as a guide when reviewing NSF Arctic project datasets archived at EOL in order to extend the metadata when necessary to match the newer ACADIS standard. This is a major effort within EOL because of the large number of existing datasets and the support level will be .8 FTE.

As ACADIS expands into disciplines not represented previously in the CADIS archive, further additions to the metadata profile may be necessary in order to better define and utilize the data. Further updates to the ACADIS profile will be undertaken based on the guidance from the ADAC, and review by scientists in the newer disciplines, while maintaining compatibility to accepted international standards.

4) Discovery and Access

4.1) Release of the Arctic Data Explorer (NSIDC, CISL, EOL)

4.1.1 Initial Release: A beta version of the Arctic Data Explorer, being developed to increase the visibility and usability of NSF ARC data in conjunction with Arctic data from other U.S. agencies, was presented during the ADAC face-to-face meeting in May 2012. The initial focus is on data from EOL, CISL, NSIDC and the Norwegian Meteorological Institute. While highly supportive of the Arctic Data Explorer, the ADAC made it clear that high priority must be placed on robust search functionality. A strong search capability will require not only solid metadata but substantial improvements on the coding side. Key steps from the development point of view are implementing a map-based search, relevance ranking of search returns, faceted search, adding a second service call for more metadata, and improving search performance (timing, consistency of returns, etc). We will release the Arctic Data Explorer when it has a minimum level of search functionality and then make additional releases over the year as functionality is added.

4.1.2 Add additional data sources. Following ADAC recommendations, we will work with the NODC to broker their Arctic data and make NODC resources discoverable through the Arctic Data Explorer.

4.1.3 Social Annotation: One recommendation by the ADAC is that the Arctic Data Explorer include the capability of listing user feedback regarding dataset quality, usability and applications. Incorporating these features is a non-trivial task and will require significant investments in developer time (see NSIDC budget justification). To address the multiple dependencies and avoid bottlenecks, development of the Arctic Data Explorer will need to progress in close coordination with metadata efforts as well as further development of the ACADIS Gateway that focuses on NSF data providers.

CISL will support the milestones discussed above by expanding existing services in support of metadata exchange and interoperability. CISL will work closely with NSIDC, where efforts will focus on improving the usability of data collected by NSF-funded projects that are held in ACADIS and providing additional services for a broad user community including data search and discovery, data access and download. The CISL contribution to accomplishing this task will be .2.FTE.

4.2) Data Download and Upload Services (CISL)

CISL will enhance data transfer services to better serve the broad user community for data download and data upload. Work areas include providing bulk data upload and download capabilities in convenient and common formats, scriptable data access services and standard data access interfaces such as THREDDS catalog representations and OPeNDAP support. We will explore data access services such as data sub-setting and visual browse capabilities based on the year two data collection needs. (0.75 FTE). This also relates closely to the dataflow and workflow milestones.

5) Education and Training (NSIDC, EOL)

Data management education for researchers and outreach to both science and scientific data curation communities has emerged as an important year 2 milestone. We will work with Arctic scientists to spread good data management practices. We plan to both deploy point-of-need materials, such as online tutorials and metadata recommendations, and proactively reach out to researchers through presentations at scientific conferences, data management and training workshops and other venues. We recognize that other groups, such as ESIP, are also investing time and resources in data management education, and we will both leverage and contribute to their efforts. We will also continue to work with the investigator teams, answering questions, offering field support especially through EOL's ongoing scientific partnerships, and address specific needs through services such as reviewing Data Management Plans before submission to NSF.

The ACADIS Team has taken a proactive role in helping as well as educating all investigators and their staff (e.g. technicians, graduate students) that bring data to ACADIS. Support begins with assistance in preparing the Data Management Plan portion of individual grant proposals to meet the new NSF requirements. Once the new grantee is known, ACADIS proposes to ask the investigator to fill out a simple questionnaire to help make their specific needs known so that we may better assist with their data and metadata submissions.

ACADIS User Support team members are also available to answer questions and assist investigators with the upload of data and metadata. As to be expected, ACADIS user support comes from all groups. This support is expanded in year 2 to better support the new disciplines and diverse types of data coming into the archive well beyond the legacy AON datasets. EOL support on this task will be .1 FTE.

6) Archiving and Preservation

6.1) Archive Information Packages (NSIDC, CISL, EOL)

Archival Information Packages (AIPs) provide a high-level archival method for managing documentation, data, and other content description through the use of descriptive and preservation metadata; the creation of AIPs promotes the goal of providing long-term archival and preservation services for NSF arctic data sets. In Year 2, NSIDC will lead an effort to create pilot AIPs based on a collection of permafrost borehole data (see Data Showcase milestone). This will involve packaging and delivering data files and their documentation together and delivering them back to the data producers and the general public. ACADIS staff will also develop specific implementation best practices for metadata standards and packaging methods so the AIPs fit the needs of the community as well as compliment the complexity of the data. These practices will in turn influence how metadata are created, and how data are organized and deposited into the current system. EOL will provide .15 FTE for this task.

6.2) Long term preservation and stewardship (CISL, NSIDC, EOL)

CISL will expand upon the pilot digital preservation work of year one and archive all appropriate ACADIS data collections to the Chronopolis preservation system. We will establish digital preservation procedures in close collaboration with NSIDC based on service levels required by project AIPs. This work area will be supported by leveraging work underway for the Library of Congress sponsored Chronopolis Project. CISL proposes 0.2 FTE to address this work.

6.3) DOI citation implementation (CISL, NSIDC, EOL)

CISL will add Digital Object Identifier (DOI) and citation support for archived data collections. This work area includes support for data collection versioning, integration with the EZID identifier service and user interface enhancements for citation generation and presentation This task will require .2 FTE of CISL staff time and .3FTE of EOL staff time..

7) Expanding Capabilities

7.1) Social Science archiving plan and implementation w/polar focus (EOL, NSIDC)

The unique challenges of handling social science data include, but are not limited to, issues related to human privacy, informed consent, and controlled and ethical access. We also recognize that we are not experts in social science data management and that we need to expand our capacity in this area. We anticipate that the special requirements of social science data will require specialized documentation and interfaces, and controls for privacy protection of sensitive information, as there are not currently part of the ACADIS system. For these reasons

the social science part of ACADIS may need to act as a somewhat uniquely identified or semi-autonomous unit.

In Year 2, we plan to develop capacity in social science data management and to work more closely with the social science community to better understand their needs. The ACADIS team has some social science-related data experience through connection to ELOKA, which has demonstrated success with handling culturally sensitive information and knowledge and building relationships with Arctic communities. As a result, NSIDC has been approached by the Inter-university Consortium for Political and Social Research, and the NSF-funded Data Conservancy led by Johns Hopkins University to explore collaboration to develop capacity to support the management of Arctic social science data (not the same as local and traditional knowledge), perhaps as part of Arctic Science, Engineering, and Education for Sustainability (ArcSEES).

At the same time we will work with social scientist Larry Hamilton of the ADAC and IASSA to better engage with the Arctic social science community to document their core data formats, standards, metadata needs, etc, compare these to ACADIS approaches, and develop a plan to evolve ACADIS to better handle these data. EOL will include .2 FTE to support this task.

7.2) Special data submissions (EOL, NSIDC, CISL)

ACADIS has developed automated procedures for metadata and data submissions, extended and refined from the original CADIS portal. It is difficult to anticipate the diverse archival needs of the larger community of Arctic scientists, their research interests and types of data ACADIS will serve. There will always be data (e.g., voluminous model grid output and “orphan” datasets) that will not fit the current archival paradigm. EOL will continue to handle these special data archival needs, and will ensure that all future ACADIS data submissions archived at EOL have metadata available in ACADIS. EOL will focus .25 FTE on this task during year 2. Examples of these new data types include:

- Underwater acoustic data from Yahtse Glacier, Icy Bay, Alaska
- Temperature, discharge and light data for Ivishak Hot Spring, Alaska
- Time-corrected high resolution time lapse images for Columbia Glacier
- Corticosteroid hormone measurements for adult and chick little auks in Hornsund Fjord, 2005-2008

8) Ongoing Activities

8.1) User Services (NSIDC, EOL)

ACADIS User Support will continue to provide services and assist NSF PIs and data users. However, the expected increase in user site traffic and data submission from scaling from AON to include all incoming active NSF ARC awards and projects will put pressure on the current ACADIS user support practices. Addition of the Arctic Data Explorer and its brokered data search will also attract a larger community. Some key steps for addressing these expectations for Year 2 include the management and tracking of users and user requests through an email management tool Zendesk, expanding the options through which users can contact project staff directly, creating detailed documentation and best practices for metadata and data submission, providing educational and arctic research related resources, creating controlled studies in order to increase usability, and contributing to policies surrounding data submission and data use. In

Year 2, NSIDC and EOL will continue to create and maintain relationships with data-producing principal investigators, other data centers with Arctic data and other users in order to advocate our services and discover current community needs. EOL will use .1 FTE for supporting this task.

8.2) Metrics (CISL)

CISL will add automated metrics reporting to the ACADIS Gateway user interface. .2 FTE of CISL staff time will be provided to develop use cases that include summary reporting for program managers, project stakeholders as well as project reports for contributing PIs, describing data access and other project related use metrics.

8.3) Ongoing ADAC activities (NSIDC, all)

The ADAC has been actively communicating and providing feedback regarding the Arctic data services we provide. The ADAC membership was largely influenced by the level of their interaction with different kinds of data in several disciplines. The committee has provided valuable feedback over the last year on impediments to data use and system and search usability testing, and has largely shaped the structure of the ACADIS Data Showcase effort focused on permafrost borehole temperature data. In Year 2, we will continue to seek ADAC's help to help align priorities with the needs of the Arctic science community. ADAC activities during Year 2 will include further usability testing on the Arctic Data Explorer at NSIDC and the CISL ACADIS Gateway, providing feedback and guidance on for the Data Showcase effort, permafrost showcase model, and providing ongoing communication and supportive feedback about the progress and future of ACADIS. The next annual meeting of the ADAC is scheduled for spring 2013. EOL has proposed .1 FTE for coordinating with the ADAC.

8.4) Sustaining EOL Arctic project websites (EOL)

EOL will continue to sustain existing specific Arctic project websites (e.g. BEST/BSIERP, SBI) and support for all Arctic data archived at EOL (<http://arctic.eol.ucar.edu/>). This stewardship approach was reinforced as a direct recommendation from the ADAC. However, the data from these projects will also be discoverable and accessible through the ACADIS Gateway. To accomplish this, EOL will review all its long-term Arctic data holdings (over the past 15 years) to update and maintain the metadata in accordance with required fields in the new ACADIS metadata profile. Much of these newer metadata parameters were not originally obtained, particularly for the oldest projects. Once all the metadata is updated, consistent, and shared with the ACADIS Gateway, EOL will also harmonize its Arctic project holdings search capabilities (e.g. Master Lists) in the event that users come directly to the EOL web pages (as many of the project PIs currently do). EOL has committed .3 FTE to this effort.