Advanced Cooperative Arctic Data and Information Service (ACADIS): Meeting our Vision

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

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 ACADIS Team has prepared a consolidated work plan for activities proposed for the third year of the project. The period of performance is from July 2013 to June 2014. We have divided the work plan into key elements as contained in the recently submitted ACADIS Status and Challenges document submitted to you. Within each element are more details about tasks and level of effort planned for year 3 from each group. The tasks in Year 3 are all focused on achieving our vision to serve and support the NSF/OPP/ARC investigators.

Project Management

The ACADIS Team is a collaboration of portions (in time) of about 30 people among four organizations including NSIDC, EOL, CISL and Unidata. Project management is Project management is handled by the PIs from these groups. Overall direction is the responsibility of NSIDC. 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. It also allows the ACADIS Team subgroups to coordinate new tools, software fixes and new user implementation procedures into the Gateway.

Activities in Year 3 will be described below in five sections that focus on;

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.

In Year 3, EOL will contribute to ACADIS by focusing on continued user support, updating metadata and Digital Object Identifiers (DOIs) or citations for all EOL Arctic project data holdings, integration of project metadata into ACADIS, and supporting users with special requirement datasets for inclusion in the ACADIS archive. CISL contributes to the ACADIS project primarily by building and operating the ACADIS Gateway in collaboration with the ACADIS project partners, community of end users and other stakeholders.   They maintain 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), with the introduction of the Arctic Data Explorer (ADE), 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 (0.08 FTE) is the NSIDC PI and the overall science lead for the ACADIS project. Florence Fetterer (0.375 FTE) provides connections to the operational data community and leads the Permafrost Data Showcase effort. Lynn Yarmey (1.0 FTE) is the senior Data Curator. She coordinates major elements of the NSIDC component of the ACADIS project with other team members through ongoing communication, organization and translation of "big picture" goals into day to day activities.

Senior Personnel (UCAR/NCAR)

James Moore (0.15 FTE) is the EOL PI and Project Manager for the UCAR/NCAR ACADIS effort. Steve Williams (0.15 FTE) is responsible for the direction of the EOL CDS Data Management Group (DMG) participation in ACADIS. Co-PI Don Middleton (0.1 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. Details on salaries are contained in the separately submitted UCAR/NCAR and NSIDC budgets

Work Plan Specific Elements

This section describes the task details for each of the 5 elements and anticipated personnel support anticipated.

There is an important caveat to note at this point related to the very effective Agile Scrum process. This applies specifically to the CISL activities in ACADIS. One of the core Agile principals is to focus on rapid delivery of the highest value possible for the project early and often. Value is defined by our stakeholders, primarily end users and data curators, based on their needs and wants. We use a value driven backlog which our Scrum team organizes into 2-week work sprints to deliverer the highest value software features in short iterations. Stated in Agile/Scrum terms, Nienhouse will continue to serve as the ACADIS Gateway Product Owner (0.2 FTE), and Wilhelmi (0.2 FTE) will continue to serve as the Chief Architect. We find that devoting an appropriate amount of time to technical management pays dividends for projects. We're requesting an additional 2.05 FTE to support the CISL Agile development team to address the broad project goals for Year 3 below which we recognize as our current priorities.

1)   Data Services

CISL: Data Provider Tools and Services

Sustainably address increasing number of data providers, by reducing workload and minimizing barriers to self-publishing.

Based on the experiences of our data provider user community and data curation staff, we have identified the following four focus areas to enable data providers to manage their data collections effectively: (1) Dataset management tools for easier data re-organization, (2) user interface support for change tracking and other provenance reporting, (3) extending data provider self service metrics reporting and (4) easy to understand service interfaces for automated data ingest. In order to increase the efficiency of data providers we'll work to automate metadata extraction where possible and integrate tools such as the Rosetta tool from Unidata to streamline data publication and still retain rich metadata capture. We'll work closely with the ACADIS Metadata sub-team to formalize metadata content related to field collection methods and use guidance.

CISL: Data Access and Reuse

Make data access faster, easier and less error prone.

CISL Year 3 development will address the community needs of data access and re-use, having focused on tools and services for the data provider as ACADIS achieved full implementation. CISL will take a lead role in addressing data access barriers, such as those related to different login mechanisms at the EOL data system and other partner organizations (identified as an annoying barrier by our ADAC). Recently, we have had requests to accept sensitive data collections requiring secure access. We plan to expose the existing data access control features of the ACADIS Gateway to provide secure access to these collections. We will also provide new capabilities to download data files in bulk and to make data available via easy to use services enabling users to automate data access or develop their own tools for data use.

CISL: Streamlining Self-Publishing Workflow

Efficiently utilize data provider time by automating processes and focus support staff on critical provider needs.

A key element to sustaining a large community science gateway is efficient use of limited data support resources. We plan to help increase this efficiency by providing automated processes, such as automated notifications of data provider updates and basic quality determinations based on provider content such as automated readme file checks. Currently, in order for data providers to begin submitting data collections to ACADIS, administrators perform a number of manual preparation steps to create work areas (projects) and perform other preparation tasks. We plan to automate project creation by integrating services such as ARLSS which exposes NSF project metadata and improving the contact database management tools. We also plan to automate portions of the current DOI and Citation generation process.

CISL, EOL: Metrics Aggregation and Reporting

In Year 3 we plan to improve metrics reporting capabilities, including the delivery of information that's useful for data publishers, program directors, and other stakeholders that have interest. Resources permitting, we wish to explore the possibility of integrating metrics from EOL and the ACADIS Gateway in order to easily provide a clearer overall usage summarye.

NSIDC: Arctic Data Explorer

We will further evolve the Arctic Data Explorer at NISDC under 2.08 FTE for software developers. Work will be shared among NSIDC's pool of 15 developers, who operate under an Agile Scrum model. This total includes 0.08 FTE support for the NSIDC software team manager. Building on deployment of Version 1 of the ADE in Year 2, key development foci, as recommended by the ADAC, will include handling duplicate metadata records, improving performance, implementing faceted search, semantic search, and taking advantage of services available for datasets. Other recommendations to be adopted include the capability of listing user feedback regarding dataset quality, usability and applications and a user dashboard for saving search returns.   ADE work will include 0.17 FTE work by T. Hoyer, the Web Designer for

NSIDC, and 0.08 FTE for Systems Administration.   ADE development will also make use the expertise of NSIDC Informatics Scientist SJS Khalsa (0.08 FTE) by leverages his Earthcube EAGER award from NSF to advance federated data search. ADE work will be coordinated closely the ACADIS Gateway work at CISL.

CISL, NSIDC, EOL:

Workflows Subgroup This ongoing activity will continue to the end of the project. Its focus on aligning partner workflows includes issues of communication, data submission and handling workflows between the ACADIS partners.   It strives to clarify the scope of how work is managed and communicated between partners, other subgroups, and stakeholders.

EOL: Fundamental technical services offered through ACADIS;

Arctic datasets will be the first collection of data at EOL to receive DOIs that are automatically registered with the DataCite publishing service. Metadata reports for DOIs of existing datasets will be generated and submitted by project, as the metadata is reviewed. An author database will be integrated with the server database to enable display of the citations with the dataset descriptions. EOL has adopted an NCAR-wide approved form for DOIs and citations, and will coordinate with CISL and NSIDC in creating consistent data citations for ACADIS. [.25 FTE]

Any new Arctic project data that will be archived at EOL (e.g. PacMARS) will be available through ACADIS. Metadata for all new Arctic project datasets will be openly available for harvesting by other institutions. [.35 FTE]

Unidata: Rosetta Development, Implementation and Integration

Development of Rosetta will continue with the various ACADIS datasets in mind. Year 3 work will focus on the addition of capabilities to handle moving sensor platforms, such as drifting buoys and CTD Gliders. In addition, Unidata will enhance Rosetta to transform data out of the standard file format used by Rosetta into standard layout CSV and spreadsheet files.

Unidata will continue to participate in the Support and Workflow subgroups to ensure smooth integration of Rosetta into the workflow for data submission to the ACADIS portal, as well as the community support systems utilized by ACADIS. Unidata, as an organization, is currently discussing its user support system, and ACADIS will be kept up-to-date as changes occur, as this will certainly change how ACADIS-related Rosetta queries will be handled.

Unidata: Data visualization

Unidata will continue to provide visualizations of the ACADIS data holdings for presentations and to provide sample imagery for use by the portal and the larger geo-science community. This will be supported in several ways including reformatting of data using Rosetta and display of products using tools such as the integrated data viewer.

2)   Special Requirements Data Collections (SRDC) support

EOL: ACADIS Handling of Unique datasets

Data that have special requirements for archiving or don't conform to the usual ACADIS workflow will be handled by EOL. In these instances the data may require special attention because of various constraints, or a very large volume of data, or concerns on effective packaging and presentation. In preparing these data for archival, EOL will work with PIs to ensure all data and metadata are submitted in acceptable form. When needed, tables will be prepared to facilitate data access. EOL will prepare and monitor controlled access to sensitive data, or data that have constraints upon use. All "special requirements" data that are archived at EOL will be available through ACADIS. [.35 FTE]

EOL will work with CISL to coordinate and develop methods to archive and access very large holdings of expected model output datasets. This would include the storage of gridded datasets to allow subsets to be easily ordered and delivered. [.10 FTE]

CISL: Expand Gateway Capabilities to automate delivery of Special Requirements Data Collections into ACADIS

 ACADIS accepts that some special handing may be required as a normal part of its community support services. Reducing the proportion of these special cases by accepting a wider variety of data provider submissions into the ACADIS Gateway offloads EOL to better target their efforts. In collaboration with the Workflows sub-group we plan to identify Gateway requirements that increase the acceptance rate into ACADIS Gateway, such as integration with storage options like HPSS tape based collections, supporting data-provider in-situ data systems and large data volume ingest tools. We will consider migrating certain Special Requirements cases from EOL's data system into the ACADIS Gateway as time allows.

Unidata: Special reformatting tasks using Rosetta

Unidata will work with the ACADIS data curators to determine how unique datasets can best be handled by Rosetta, and if necessary and possible, modify the interface to handle these unique datasets. Unidata recognizes that a declarative approach to data transformation will not work in every case, and other, complementary approaches are needed to make Rosetta a more complete tool for the community.

3)   Metadata Documentation, sharing and usability

NSIDC, EOL, CISL, Unidata: Metadata Subgroup

The metadata subgroup is an ongoing activity will continue to the duration of the project. The Subgroup assesses the current state of metadata across the ACADIS project, compiles and prioritized necessary metadata work, establishes minimum metadata sharing requirements for project partners, created a data dictionary and requirements for implementing data citations, and works to determine metadata needed to support file-level usability.

EOL: All aspects of metadata generation, updating and utilization in ACADIS

Data that are submitted through EOL for ACADIS archival are checked for adherence to the metadata and documentation standards. EOL will work with the community of arctic investigators to ensure completeness and consistency of

their submissions. In this process inconsistencies are investigated and corrected, data is reviewed, and the display and ordering of data is tested. EOL will work with NSIDC curators to share lessons learned from EOL metadata review/cleanup and quality control of datasets. [.15 FTE]

Arctic project datasets not yet in ACADIS will be prepared for harvesting. This involves confirming that metadata is consistent with ACADIS metadata sharing standards, and extending existing metadata when necessary. [.15 FTE]

The ACADIS metadata profile will be updated in order to better define and utilize data from diverse disciplines. EOL will share experience from defining biological data, and work with CISL and NSIDC specialists. The metadata profile can be extended to file level information, while adhering to recognized international standards. [.04 FTE]

Unidata: Metadata requirements for optimal Rosetta utility

Unidata will continue to participate in the Metadata subgroup to ensure Rosetta will have knowledge of required and suggested pieces of ACADIS metadata and allow for optimal interoperability.

4)   Interoperability and Initiatives

Unidata: Preparing Rosetta for Integration into the ACADIS Gateway

Preparing Rosetta for integration into the ACADIS data portal will be a year three priority for Unidata. At this point, it is not determined what mechanism will be used by CISL to allow Rosetta to publish to the data portal, but once it is defined, Unidata will move swiftly to utilize the CISL data portal submission workflow and/or API. This work will conform to standards and best practices to insure the best opportunity for interoperability.

NSIDC, Unidata: Permafrost Showcase Effort

We will complete the Permafrost Showcase effort at NSIDC focusing on permafrost borehole temperature data. The effort will be led by CO-PI F. Fetterer. K. Schaefer, a permafrost scientist at NSIDC and member of the permafrost modeling community, will contribute to the project (.08 FTE). Schaefer's involvement will guide us bridge the data gap between field scientists and modelers. Schaefer's participation the Permafrost Showcase effort offers a direct connection to the International Permafrost Association and the broader research community. NSIDC will collaborate with Unidata on use of the Rosetta tool in the showcase effort.

NSIDC: Arctic Social Sciences Data:

We will advance the handling of social sciences data through the support of NSIDC social scientist P. Pulsifer (0.08 FTE) and a postdoctoral scientist (0.5 FTE). P. Pulsifer is the PI of the ELOKA effort, an NSF project focused on Arctic

community-based monitoring and local and traditional Knowledge. His involvement will bring deeper connections between ACADIS and the NSF Arctic Social Sciences Program (ASSP). We have identified matching finding for the postdoc through the Council of Library and Information Resources Postdoctoral Fellowship Program. The postdoctoral fellow will provide recommendations and work to implement an Arctic Social Science solution for ACADIS. Also, Pulsifer will contribute to the semantic search development work planned for the ADE.

EOL: EOL Data server integration to the ACADIS Gateway

Recognizing that data in ACADIS reside on servers with different protocols for access and delivery, EOL will work on extending the interoperability of the EOL data server with the ACADIS gateway to provide a more seamless discovery and data flow experience. EOL will investigate a single sign-on to integrate with the CISL Gateway access; ensure that data order links go directly to data access or order pages, and incorporate an ACADIS look and feel when users come from the Gateway to the EOL server for data. [.15 FTE]

EOL: Project and dataset access via golocated tools (GIS)

If supported by the ADAC, EOL will update the ACADIS MapServer for all datasets, increasing the extent of metadata and data that can be directly viewed by clicking on a geolocated site. A greater use of, and further enhancements for GIS data will be explored. [.20 FTE]

EOL: Project Metrics

Metrics reporting will be expanded to incorporate more graphs and other visual representations, allowing for easier publishing to web pages and reports. Metrics from the several EOL arctic web sites and pages will be incorporated into a form easily shared with CISL and NSIDC, and retrieved in real time to be posted to a web page. [.10 FTE]

CISL: Broad Community Access to ACADIS Data Collections and Services

Increase visibility of ACADIS data collections and make them easier to find by the broad scientific community.

Much of the value of the ACADIS data holdings lies in the future re-use of these precious and difficult to reproduce digital artifacts. Broad scientific community access and discovery is imperative for re-use. We plan to continue to support and expand upon metadata harvest by the ADE, GCMD and WMO-WIS/GEOSS systems to federate these collections to a world-wide scientific audience. We'll explore the value of exposing file-level discovery metadata based on community input and needs. We will continue to harvest and provide discovery and access to EOL and NSIDC metadata records, and explore any others that the team determines will provide good value to our community.

We have been moving towards a Service Oriented Achitecture (SOA) for the ACADIS Gateway and will continue to evolve and enhance the service-oriented approach as we develop new capabilities or refine existing ones. We plan to address year three goals such as automated data ingest and certain data access paths using this approach. This has the two-fold benefit of making ACADIS more useful to other systems by providing services, and making ACADIS better by consuming external services that provide benefit to our community. Over the past year we've made substantial improvements in the SEO (Search Engine Optimization) area, such that ACADIS holdings are readily findable by general-purpose search engines. We plan to continue to improve this important area in Year 3.

CISL: Community Engagement and Metrics

ACADIS already engages with the community extensively, including assisting with the development of data management plans, assisting with self-publishing processes as needed, and handling Special Requirements Data Collections (SRDC's). In Year 3 we want to consider gathering satisfaction metrics in order to better gauge our effectiveness, and to identify the best areas to focus on for improvement (this is not just a CISL area, but project-wide). We are also enthusiastic about deepening our engagement with NSF's Earth Cube program, including potential contributions to the Earth Cube Building Block and Conceptual Design activities, an Earth Cube workshop in collaboration with EOL, and the possible development of an Arctic/Polar Research Coordination Network (RCN) for our community. We do need to consider how we support these efforts, however, as we currently do not have ACADIS resources set aside to support them.

CISL: Workflows Subgroup

ACADIS is an effort that involves a number of complex workflows, and we intend to streamline and automate these as much as possible. The idea is to maximize the effectiveness of our user community in terms of using ACADIS resources, and the efficiency of our multi-organization team. CISL will continue to lead the Workflows Subgroup in Year 3. Several areas will require particular attention, including ADE integration, Rosetta integration, analysis of Special Requirements needs and processes, issue tracking and resolution, and the general publication of data and metadata via the ACADIS Gateway.

5) Science Support

NSIDC: Data Curation and Community Support

The three data Curators, Lynn Yarmey (senior curators), Liz Schlagel (1.0 FTE) and Toni Rosati (1.0 FTE) will continue to act as a focal point for ACADIS activities. Curators define and monitor the workflow that binds the ACADIS team. They track overall project milestones and communications using the Basecamp project management system. Yarmey will continue to manage specific software

development activities in local task management systems (Pivotal, Trello, Jira) and lead or co-chair monthly meetings of the ACADIS management team. L. Schlagel will focus on Community Support emails, metrics, and educational templates through the Zendesk customer support management application. Toni Rosati will be responsible for monthly partner project status snapshots to monitor progress and allow for visibility throughout the project. The Curators will also draft promotional articles and materials for ACADIS and the scientist community. The Data Curation work at NSIDC will also require 0.08 FTE for a User Services representative (Lisa Booker), responsible for training curator staff, assistance in setting up ACADIS Community Support infrastructure, and consulting on emerging ACADIS issues (e.g., effective metrics collection, and scaling responsiveness with community growth). User Services also provides referrals to appropriate external data centers and other data and information providers as needed. They provide a conduit for user feedback on the ADE and ACADIS Gateway.

NCIDC, CISL, EOL: Community Support Subgroup

This ongoing activity includes Zendesk Agents (NSIDC Curator Liz Schlagel, NSIDC Curator Toni Rosati, Eric Neinhouse (NCAR/CISL), Sean Arms (NCAR/EOL), and Janet Scannell (NCAR/EOL)), as well Lisa Booker (NSIDC User Services Lead), and Lead NSIDC Curator Lynn Yarmey. Subgroup tasks include providing input on metrics reporting, usability testing, and review of current procedures. It addresses issues with community support workflows and recommends solutions. This includes specific services and plans for new procedures for datasets handled by EOL that require special attention.

EOL: Advice to science community on field project development and data management support

Support to Arctic Principal Investigators will be provided by EOL. Guidance to NSF PIs developing field project plans may involve planning for logistical support, data management, and defining a data policy. EOL stands ready to assist and provide support as needed. [.15 FTE]

EOL: Assistance to community in dataset preparations for submission to ACADIS

EOL will reach out to the technicians and post-docs who handle data submissions to ACADIS to assist them in the metadata and documentation procedures and archival best practices. These individuals often don't have the experience with ACADIS, and could benefit from an introduction to what is needed to create a dataset, how to prepare files for upload, etc. [.14 FTE]

EOL: Assistance to NSF on monitoring investigator data requirements

EOL will work with the Program Managers to ensure data is archived with accompanying metadata and documentation that meets the ACADIS requirements. This includes the "special requirements" data that typically has more complete or stringent documentation needs. EOL will continue to assist Program Managers with tracking data submissions and confirm archiving as needed. [.04 FTE]

EOL: Adaptation of Zendesk

Zendesk support in collaboration with NSIDC will be incorporated into the EOL workflow and EOL specialists will collaborate in ACADIS Community Support. [.04 FTE]

EOL: General community assistance

EOL will continue to support the community with the stewardship of existing Arctic holdings: serving the data, hosting the web sites, maintaining the datasets, and providing user support, to ensure the legacy of all Arctic data at EOL. [.10 FTE]

EOL and CISL: Earthcube Workshop on Polar Data

EOL and CISL will jointly plan and host a Workshop on Polar data under the sponsorship of the NSF EarthCube community domain workshop framework. The strategy here will be to bring together geoscience investigators, data providers, and archives to discuss collaboration between many diverse activities (including both Arctic and Antarctic) and gather feedback from respective scientific communities. Many of these polar activities could be better linked and integrated with ACADIS developments and also provide insight into future ACADIS developments. The goal would be to hold such a Workshop in the Fall 2013 timeframe.

CISL: ACADIS Gateway Technical Support

ACADIS strives to provide excellent, comprehensive community support. In Year 3 CISL will continue to liaise with data providers and curators in order to deliver the most valuable new features and services in the ACADIS Gateway. We will also continue to provide responsive technical support when Gateway issues arise, and we intend to streamline the integration of issue-tracking capabilities.

Unidata: Integration of Rosetta into Arctic community and broader Geosciences discipline use

Unidata will continue to advocate the use of Rosetta with the broader geoscience community by providing demos to interested groups as well as presentations at scientific meetings, such as the American Meteorological Society. Unidata has demonstrated Rosetta to groups at the Jet Propulsion Lab that archive oceanography datasets and the Integrated Earth Data Applications (IEDA), an NSF-funded facility that provides data services for global geochemistry and

marine Geoscience research at the Lamont-Doherty Earth Observatory, and the Consortium of Universities for the Advancement of Hydrologic Science (CUAHSI), all of whom have shown interest in the use of Rosetta in their respective disciplines.

NSIDC: Archive Information Packages

This work, led by Data Curator Liz Schlagel and Co-PI Florence Fetterer, will be completed in year 3.   Final steps include completion of a report on the project, addressing technical limitations (e.g., related to metadata), looking at generating information packages from the existing data and working with NSIDC informatics scientist Ruth Duerr (0.08 FTE) to identify packaging solutions and realistic community focused services for ACADIS. Duerr brings data preservation knowledge, semantic research, Data Conservancy ties, and connections to the top Library and Information Science Data Curation education programs.