

# **CADIS Report: April 2009**

**Mohan Ramamurthy and Jeff Weber**

The Cooperative Arctic Data and Information Service (CADIS) is a collaborative project, supporting the Arctic Observing Network (AON) community. AON is an NSF Office of Polar Programs funded initiative for the International Polar Year (IPY) to improve observational capabilities in the Arctic and leave a long-term legacy for the benefit of science and society. AON data will contribute to scientific research leading to (1) increased knowledge and understanding of the regional and global causes and consequences of present-day environmental Arctic Change, (2) scenarios for and prediction of the course of future Arctic Change and its regional and global consequences, and (3) the development of adaptive responses to Arctic Change.

CADIS is intended as a development project to bring community standards, data submission, archival and stewardship expertise, analysis and visualization tools, and vision to develop a comprehensive data management capability to support AON and other IPY-related projects.

The CADIS team is made up of 23 people from NCAR (EOL and CISL), Unidata, and NSIDC. Additional information about the CADIS project and data provided by CADIS can be found at: <http://www.eol.ucar.edu/projects/aon-cadis/>

Unidata is providing middleware (THREDDS), data transport (LDM) for real-time AON data transmission, and analysis and visualization capabilities (IDV). Unidata's Network Common Data Form with CF conventions is a strong contender for use as the "compatible format," so Unidata is assisting in format conversion to netCDF.

The recently submitted Annual Report to NSF provides a summary of the activities and accomplishments during the past year.

**Cooperative Arctic Data and Information System (CADIS)**  
**2008 NSF Annual Report**  
**National Center for Atmospheric Research (NCAR)**  
**National Snow and Ice Data Center (NSIDC)**  
**University Cooperation for Atmospheric Research (UCAR)**

The CADIS team is made up of two-dozen people from the supported organizations listed above. This team has been brought together to develop a coordinated data management service to begin to meet the needs of the Arctic Observing Network (AON). AON is a major contribution of the US National Science Foundation (NSF) to the International Polar Year (IPY). CADIS is intended as a development project to bring community standards, visualization tools, data archival and stewardship expertise and vision to develop a data management support capability for this important community

The NCAR Earth observing Laboratory (EOL), NCAR Computational and Information System Laboratory (CISL), UCAR Unidata Program and NSIDC Principal Investigators (PIs) provided oversight, guidance and coordination among the CADIS team members working on this project. Regular updates via email, web tools and meetings were held to develop and implement strategies for developing and implementing the key components of CADIS. EOL focused on the refinement and updating of the CADIS web pages, implementation and care of the metadata catalog for all AON CADIS datasets and coordination with the NSF Program Manager on project tasks and direction. CISL (composed of the two distinct Visualization and Enabling Technologies Section (VETS) and Data Support Section (DSS) has focused upon next-generation portal development, but also continued to develop the metadata underpinnings of the system, and collaborated on all prototype system deployments including data publication, data access (TDS) and data visualization (LAS) server configuration and data formatting, data testing and quality checking. Unidata has been involved in three aspects of CADIS to date: format conversion, remote data access, and visualization of CADIS data all contributing to the design, development and guidance of CADIS. This has resulted in a plan to guarantee the long-term engineering success of the CADIS project: understanding of data format and access requirements, definition of metadata schema, research into visualization technologies, and integration with current technologies.

Attention has been given to working with the International Arctic Systems for Observing the Atmosphere (IASOA) to develop pathways for bringing real time and post season datasets into the CADIS archive. These activities will be continued in the next year.

All groups contribute to CADIS PI and management functions, engineering design and implementation, and general interfacing with various communities and projects. EOL manages the CADIS web pages that contain sources of project background and will guide users to the CADIS AON Web Portal. The URL is:  
<http://www.eol.ucar.edu/projects/aon-cadis/>

There were several presentations and discussions held on CADIS at meetings and other fora to inform the community of our activities and seek input from the community that would help focus CADIS capabilities. These included:

American Meteorological Society	January 2008	New Orleans, LA
Arctic Observation Integration Workshops	March 2008	Palisades, NY
Sustained Arctic Observing Networks Workshop	April 2008	Helsinki, Finland
GEWEX CEOP Meeting	September 2008	Geneva, Switzerland
Sustained Arctic Observing Networks Workshop	October 2008	Edmonton, Alberta, CA
American Geophysical Union	December, 2008	San Francisco, CA

CADIS reached a major milestone in March 2008 with the release and implementation of the user interface for metadata and data upload via the EOL metadata interface and CISL Community Data Portal (CDP). **AON Investigators had archived ninety-seven data sets in CADIS by the end of 2008.** The next generation CADIS Portal was demonstrated to the community at the Fall American Geophysical Union (AGU) Fall Meeting at the end of 2008. This accomplishment now sets the stage for the long-term structure and capabilities of CADIS. Further specifics of the main development areas and highlighted accomplishments are described below:

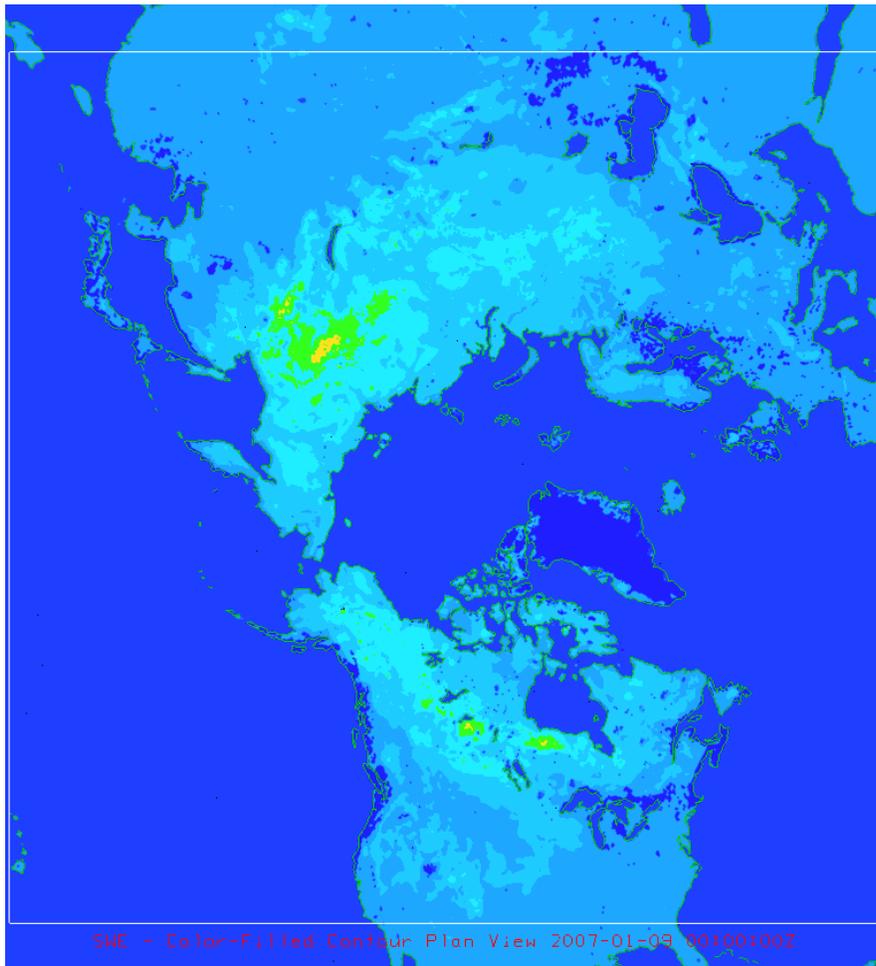
- **Metadata:** members from all CADIS partners institutions participated in a metadata working group, which was tasked with identifying the requirements for capturing all information relevant to discover, describe and access data for the Arctic domain. The result of this effort was the definition of a CADIS-specific metadata profile, which is based on the IPY guidelines, and includes additional elements either to meet the specific requirements of the CADIS project, or to facilitate export to other schemas such as ISO and DIF. The THREDDS schema, which is used as container and interoperability framework, has been extended to reflect the CADIS profile.
- **Metadata Authoring Application:** CISL and EOL developers collaborated on the development of a web-based metadata authoring tool, and its integration into the current CDP publication infrastructure. This interface allowed the CADIS PIs to capture and share the information required to describe their data holdings as well as upload data into the archive. A first prototype of this application was completed and released to the PIs for feedback in Spring 2008. As noted above, the AON investigators had documented and archived 97 datasets by the end of 2008.
- **CADIS Web Portal:** a CADIS-specific area has been established within the existing CDP web portal in order to prototype hierarchical browsing of CADIS data collections, data access, and integration with the metadata authoring tool. This work included the definition of science-based categories to organize the data holdings, as well as branding of the CDP user interface to expose a CADIS-specific look-and-feel.
- **CADIS User Assistance:** EOL and CISL engineers have been available everyday since the CADIS metadata authoring application went online to render assistance to CADIS investigators who request it. The level of assistance has ranged from answering questions, to helping lay out data sets, to engineering changes to the interface.
- **Dataset Format Use and Conversion:** Unidata developers have provided data format translation between native formats such as GRIB, EASE Grid, and NCEP Office Note 29 to NetCDF so that datasets in those formats can be accessed via the CADIS portal. NetCDF continues to be the format of choice for storage and access. Unidata tools have been applied

to perform those translations confirming the efficiency of this method to standardize CADIS data. A prototype conversion to netCDF was developed by Unidata and EOL engineers and tested for use by various data visualization tools. Combining the IDV metadata enhanced ASCII CSV output with the THREDDS netCDF JAVA implementation will allow quick and easy programmatic and/or human created CF-NetCDF files. This advancement will allow the gateway the capability to offer both netCDF files and native ASCII as provided by AON PIs.

- **Data Visualization:** The development of new effective visualization capabilities is a key CADIS focus, and we made progress on several related fronts. EOL has implemented a GIS Mapserver capability for locating all AON measurement locations on a polar project map. The Mapserver provides interactive perusal and access via the CADIS Portal. CISL engineers developed a new user interface for the NOAA/PMEL Live Access Server (LAS), exposing services for both point observation as well as gridded datasets. We also integrated the NSIDC JAZ Geographic User Interface into basic and advanced search functionality, developed data management tools for configuring Unidata's THREDDS Data Server (TDS) for data access, and began work on a mechanism for the automatic startup of IDV as part of the CADIS portal. As a precondition for data access, gridded datasets had to be converted from GRIB format to NetCDF, and in-situ data ingested from ASCII files into a relational database. Some of the CADIS datasets are being used to drive some of the development of the TDS. Unidata has been working to develop visualizations of these translated datasets using the Unidata Integrated Data Viewer (IDV). The IDV is being used as one of at the clients for data visualization (See Figure 1 below) and Unidata is documenting the process to facilitate use of the IDV for data visualization by the users and continues integration efforts.
- **Real time data ingest:** Unidata has also been working with the CADIS team and in consultation with a few selected AON PIs to determine the need and process for setting up a real-time data distribution system using the Unidata Local Data Manager (LDM) application. As the need arises, Unidata is ready to help AON PIs to install the LDM to move real-time data from Arctic observatory sites to the CADIS system in Boulder and to other users in the community.
- **Semantic Search:** A web-based application for search and discovery of Arctic data based on Semantic Web technologies (the so-called *Web 3.0*) has been prototyped and populated with some of the CADIS demo datasets. If pursued, such an approach would enable a high level of integration with other data management and access projects undergoing within NCAR and partner institutions, such as the Earth System Grid (ESG) and the Earth System Curator (ESC), along with the Community Data Portal.
- **Next Generation CADIS Web-Portal Software Base:** CISL continued to develop the Science Gateway Framework (SGF), factoring in requirements generated by the CADIS effort. We developed a new prototype gateway that included semantic search capabilities customized to Arctic metadata, adapting from the general SGF framework. This included the definition of CADIS facets, harvesting of metadata from relational database to RDF triple store, development of CADIS specific search UI, and integration of geospatial searches into query service layer. In addition, we worked on prototyping the federation of metadata records between the CADIS gateway and partner data centers, based on using the Open Archive Initiative (OAI) Protocol for Metadata Harvesting (PMH) to disseminate metadata

records in RDF and other metadata formats to/from the RDF triple store back-end. CISL culminated its contributions this year by providing live demonstrations of the next-generation system at the American Geophysical Union (AGU) Fall Meeting in San Francisco, using the eGY (Electronic Geophysical Year) booth as a shared venue.

- **Integration with Research Data Archive:** Sharing experience and building compatibility between CADIS and CISL's Research Data Archive (RDA) continues. The CADIS and RDA metadata profiles were designed to be compatible and initially three RDA collections have been identified for direct inclusion in the Arctic portal: conventional atmospheric soundings and surface observations beginning in the mid-1970s, and NCEP's 1x1 deg model output and the Final Analysis beginning in 1999. These archives are updated monthly and will be reference datasets for many Arctic studies.



**Figure 1. Example of CADIS visualization product showing digital snow water equivalent (SWE) data from NSIDC format converted to net-CDF and displayed using the IDV desktop tool. Additional data overlays can be added directly via IDV.**