

CU Proposal No. : 0311.12.0368B

A Proposal to the

**Unidata Program Center's Equipment Awards Program**

for Support of

**Improving Access to a Climate Data Repository Using Unidata Tools**

Name and Address of Institution:

The Regents of the University of Colorado  
572 UCB  
Boulder, CO 80309-0572  
Telephone: (303) 492-6221  
FAX: (303) 492-6421

Institutional Identifiers:

DUNS: 00-743-1505  
Cage Code: 4B475  
TIN: 84-6000555

Desired Starting Date:

May 1, 2011

Proposed Duration:

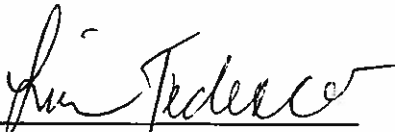
1 Year

Amount Requested:

\$5,530

Principal Investigator:

Catherine A. Smith  
Research Associate  
Cooperative Institute for Research in  
Environmental Sciences  
University of Colorado  
216 UCB  
Boulder, CO 80309-0216  
Telephone: 303 497-6263  
Fax: 303 492-1149  
Catherine.Smith@colorado.edu



Lisa Tedesco, Proposal Analyst  
Office of Contracts and Grants  
[lisa.tedesco@colorado.edu](mailto:lisa.tedesco@colorado.edu)  
303-492-0896



Catherine A. Smith  
Principal Investigator

## **Project Summary**

The current Unidata [5-year funding proposal](#) states “facilitating research and education in climate is singled out as a new priority area for Unidata.” A key component in accomplishing that goal is the use and improvement of Unidata tools and systems to provide access to and display of climate datasets.

The University of Colorado at Boulder’s Cooperative Institute for Research in Environmental Sciences (CIRES) Climate Diagnostics Center (CDC) maintains a repository of climate datasets that is used daily by researchers and educators at CIRES and around the world. These datasets are being used to answer questions about the Earth’s climate system, such as the cause and nature of extreme climate events like the 2010 Russian Heat Wave, and the predictability and impact of the El Niño/Southern Oscillation (ENSO).

Providing efficient access to these datasets is a challenge that could be addressed through the implementation of Unidata technologies such as THREDDS and RAMADDA. This proposal requests funds to purchase a new server to enhance and expand the existing THREDDS Data Server capabilities and establish a RAMADDA server at the CU/CIRES Climate Diagnostics Center in order to provide end-to-end data services that facilitate research and education in the climate sciences.

## **Project Description**

The mission of the CU-Boulder/CIRES Climate Diagnostics Center (CDC) is to advance national capabilities in interpreting the causes of observed climate variations, to apply this knowledge to improve climate models and forecasts, and to develop new climate products that better serve the needs of decision-makers and the public. The CDC develops national capabilities to analyze, interpret, and forecast important climate variations on time scales ranging from a few weeks to centuries. CIRES researchers work closely with National Oceanic and Atmospheric Administration (NOAA) scientists to advance climate science inquiry and research.

As part of that mission, the CDC maintains a publicly accessible repository of data housed at the Physical Science Division (PSD) in NOAA’s Earth System Research Laboratory (ESRL) that includes reanalysis, reforecast, sea surface temperature and radiation data. Twelve (12) terabytes (TB) of this data is currently accessible via FTP and through a THREDDS Data Server (TDS). In the past year (2010), the TDS received over 5 million hits for a transfer of approximately 1.2 TB of data. During that time, the amount of data being downloaded via TDS increased by a factor of 10, from 2-3 to 30-40 gigabytes/month. Much of this use is from educational institutions accessing reanalysis data. This load is taxing the existing system, and we plan to add more datasets in the coming year. For the CDC to continue serving these datasets through the TDS, a new system is required.

In addition to a new system, we propose to enhance the offerings on our TDS to include time aggregation of long-term datasets, data subsetting and WMS/WCS services. Currently, a proxy configuration issue on our server precludes our use of some of the integral TDS services, and we have been working on this issue with the Unidata THREDDS developers. Recently, we found a solution available in Tomcat 6, but cannot install that on the existing machine. We do anticipate a significant increase in TDS use when this issue is resolved. This issue does not affect access to the data through the Integrated Data Viewer (IDV), which is an integral part of NOAA's Climate and Weather Visualization Laboratory. The existing CDC THREDDS catalog is currently available through the main IDV data catalog so any IDV user already has access to the repository.

We also plan to set up a RAMADDA server on the new machine. Presently, the CDC is using Unidata's RAMADDA server on motherlode.ucar.edu to store climate visualizations in the form of IDV bundles. These bundles access our TDS and are used to monitor the current state of the climate. With the establishment of a local RAMADDA server, these capabilities can be moved to the CDC server and expanded. A key factor in adoption of Unidata tools like the IDV is having a rich set of examples of how it can be used. We plan to augment the existing set of climate monitoring bundles with new bundles for use by the IDV community. The access control system in RAMADDA will allow us to serve up restricted datasets via OPeNDAP to internal users instead of having to run two THREDDS servers. RAMADDA's search capabilities will make it easier for researchers and educators to find particular datasets and can be federated with other Unidata RAMADDA servers.

### **Benefit to the Climate Research and Academic Community**

The CIRES Climate Diagnostics Center has a strong record of success with its current set of web-based climate analysis tools. Examples of these include: the ability to composite data from different time periods; correlate data with index time-series such as the North Atlantic Oscillation (NAO); obtain data distributions; and subset data. The webpage <http://www.esrl.noaa.gov/psd/data/gridded/tutorial/> illustrates the many ways our products can be used. These tools are used in many academic settings as the basis of coursework, weather/climate discussions, class projects, curriculum development, and for textbook content. Institutions such as the University of Alaska, University of Albany, Penn State University, Rutgers University, and the University of Washington have indicated to us how useful the webpages and data are for teaching at the graduate and undergraduate level. The pages are used substantially by researchers at NOAA and other governmental institutions, other U.S. and international universities and by private industries.

Our downloaded datasets have been cited in numerous research papers, and products generated from our webpages are featured in talks at many research conferences. Our gridded data has been downloaded from locations as diverse as China, Iran, Israel, Brazil, Russia and Canada. Researchers from institutions such as

NOAA's IRI, PMEL, and NOAA's regional climate centers rely on the CDC's data being accessible through FTP and through TDS/OPeNDAP.

The datasets served up by the TDS are used to make products for the NOAA Climate Analysis Branch's "map room" (<http://www.esrl.noaa.gov/psd/map/>). The map room provides real-time ensemble forecast products, which are used daily by several Unidata institutions. The CDC's expertise in climate analysis enables us to serve generated products that help users both understand the current climate (e.g. the map room) as well as complex climate processes such as ENSO (<http://www.esrl.noaa.gov/psd/enso/>). Unidata sites also use the map room's ENSO and sea surface temperature pages for teaching about El Niño/La Niña.

One of the primary reasons that the CDC has been able to provide these useful tools to access our datasets is that we have standardized on netCDF as a storage format. This decision allows us to produce standardized libraries and interfaces for accessing the data files. The CDC has long been an active participant/user in Unidata's netCDF community and a proponent of netCDF convention standards. It was one of the first repositories to standardize all its datasets on netCDF using the COARDS convention prior to the Climate and Forecast (CF) convention being proposed. Most recently, the CDC has been upgrading its datasets to adhere to the CF conventions, and new datasets use this convention as they are added to the system. Additionally, the CDC is one of the first sites to use the new netCDF4 packing capabilities to store and serve data from its archive. We have been active participants in testing this through climate community tools such as the IDV, NCAR Command Language (NCL) and Climate Data Operators (CDO), netcdf4-python, PyNGL/PyNIO and GrADS.

One of the PIs (Don Murray) has been developing new capabilities in RAMADDA to provide improved access to gridded data. Due to this work, the next version of RAMADDA (1.3) will allow users to extract and plot point data from grids including long time series from climate datasets. He has also created a data ingest module for some of the CDC's customized point observation data. Any future development and bug fixes will be shared with Unidata to enhance the community capabilities.

In summary, the new equipment will allow us to continue, and substantially expand upon, providing access to data that the Unidata Program has deemed important and that many Unidata sites are already relying on. The new capabilities provided by the combination of more capable hardware and the newly installed RAMADDA server will allow us to add to the ways that users are able to interact with the data.

### **Proposed Equipment**

The CDC data repository is housed in the NOAA/ESRL/PSD facilities in Boulder. The current system (4 core, 32-bit server with 4 gigabytes (GB) memory) is used for other services (product generation, FTP, etc.) that impede the performance of the TDS, which can only be allocated 1 GB of memory. We propose purchasing a new

server, moving the TDS to the new machine and adding RAMADDA. The new machine (16 core, 64-bit server with 64 GB memory (see specifics in the budget section)) will be dedicated to serving the datasets and products through TDS and RAMADDA, offloading the burden on the other system. Instead of requesting multiple machines, we opted for a single system using virtual machines (VMs) to partition the processor and memory resources.

The new server will be physically located with the existing server at ESRL/PSD and be maintained by the ESRL/PSD system administration staff. The two systems (old and new) will constitute the CDC data serving capabilities.

### **Personnel Expertise**

PI Smith has worked at the CDC for 19 years. She has used her meteorological expertise to develop web tools that enable users to perform climate analysis on the web. She has expertise on netCDF and metadata standards and attended the Unidata netCDF-4 training workshop last Fall (2010).

Co-PI Murray worked at Unidata for 15 years before working at the CDC. He was the lead developer on the Integrated Data Viewer (IDV) and continues to contribute to its development and adoption, particularly in the climate sciences. He has experience with all Unidata products and services. He is a developer of RAMADDA and recently added capabilities in RAMADDA for displaying time series of long term climate observations.

Other CDC staff who will be involved include Don Hooper (CIRES) and Chris Kreutzer (ESRL/PSD). Don has worked at the CDC for over 25 years and has expertise in data storage and metadata. Don has attended several Unidata training workshops (IDV, TDS, netCDF) in the past. Chris will be responsible for the setup and maintenance of the system. Chris attended the netCDF and TDS training workshops last fall (2010) and has worked with the TDS developers to overcome the proxy issue that prevents effective use of TDS at the CDC now.

## **Budget**

The total requested funding for this proposal is **\$5,530**, which includes equipment and indirect costs. This will allow for the purchase of the components to build a system that will be dedicated to providing access to the CDC climate repository through TDS and RAMADDA. ESRL/PSD will provide system administration support including backup, power, networking and rack space. After discussions with Mike Schmidt at the Unidata Program Center, we are proposing a server with the following specifications:

- 1 SUPERMICRO AS-2022G-URF 2U Barebone Server
- 2 AMD Opteron 8-Core Server Processors
- 64 GB System Memory
- 2 2TB Western Digital SATA Disks
- 1 Intel Quad Port PCI Adapter
- 1 64 GB SSD drive

**Equipment Cost:     \$4,389**

Configuration notes:

- 1) The CDC datasets will be mounted via NFS from another PSD machine, so the disk space is for storing the RAMADDA database and local files (e.g. bundles, case studies).
- 2) The system will be running RedHat Enterprise Linux (version 6.0), which is the standard for PSD maintained systems. It will run separate virtual machines for the TDS and RAMADDA.
- 3) The system will run 2 RAMADDA servers, each in its own virtual machine – one publicly accessible, read-only version without login capabilities for security, and one only accessible internally for administration and product generation by CIRES and PSD staff. The 2 RAMADDA systems will share the same database and datasets.
- 4) The SSD drive will be used as the boot partition for the operating system for fast initialization.

## **Total Direct Costs**

The total direct cost (equipment) for the project is: **\$4,389**

## **Total Indirect Costs**

The indirect cost (CU overhead (26%)) is: **\$1,141**

## **Total Project Costs**

The total cost for the project is: **\$5,530**

### **Project Milestones**

We expect to purchase the equipment within one month of award. A specific Plan of Action and Milestones is:

- May 2011 – Notification of award.
- July 2011 – Acquire and install equipment.
- September 2011 – Set up and configure THREDDS and RAMADDA servers.
- November 2011 – Servers online and operational.



**UNITED STATES DEPARTMENT OF COMMERCE**  
**National Oceanic and Atmospheric Administration**  
Office of Oceanic and Atmospheric Research  
Earth System Research Laboratory  
325 Broadway – David Skaggs Research Center  
Boulder, Colorado 80305-3337

Dr. Konrad Steffen  
Director, CIRES  
University of Colorado  
216 UCB  
Boulder, Colorado 80309

Dear Dr. Steffen:

It is my understanding that Mrs. Cathy Smith will participate as a Principal Investigator on a proposal titled "Improving Access to a Climate Data Repository Using Unidata Tools". The goals of the proposal are consistent with the objectives of the ESRL Physical Sciences Division and, in the event that the proposal is funded, the Physical Sciences Division is willing and able to provide the technical support needed to complete Mrs. Smith's share of the tasks.

Sincerely,

William D. Neff  
Director, Physical Sciences Division

cc: J. Kosley

