Proposal for the 2010 Unidata Community Equipment Awards

PROPOSAL TITLE: Addition of a Community THREDDS/RAMADDA Server System at Penn State

INSTITUTION: The Pennsylvania State University
Office of Sponsored Programs
110 Technology Center Building
[redacted]

PRINCIPAL IN

[redacted]

Fax: (814) 865-3663
Email: chuckp@psu.edu

CO INVESTIGATOR(S): Dr. William Brune, Professor and Department Head
Dr. Chris E. Forest, Associate Professor
Dr. George S. Young, Professor

ADMINISTRATIVE CONTACT: Ms. Sue Lavan, Associate Director for Grants and Contracts

INSTITUTIONA

[redacted]

110 Technology Center Building
University Park, PA 16802-7000
Phone: (814) 865-1372
Fax: (814) 865-3377
Email: osp@psu.edu

EMPLOYER ID NUMBER: 24-6000376 **DUNS:** 00-340-3953
PROJECT PERIOD: 8/1/2010 – 7/31/2011
TOTAL REQUESTED AMOUNT: $19,621

The Pennsylvania State University employs individuals and accepts students and graduate research students from a multitude of national backgrounds. As an entity, the University is subject to, and works diligently to obey, federal regulations regarding the export of controlled technologies and data. Sponsor, as an independent entity, is individually responsible for ascertaining its compliance with federal export laws and procedures. If sponsor anticipates disclosure or provision of controlled technology or data to University as part of the proposed sponsored project, Sponsor should inform University, in writing, of the existence of, and information concerning the scope and extent of, such anticipated disclosures or provisions.

Project Summary:

The Pennsylvania State University (PSU), Department of Meteorology currently ingests, relays and processes most of the available Unidata Internet Data Distribution (IDD) data streams. Processing of the data is currently limited to raw and decoded gempak data sets for local use. We propose the expansion of our data collection and relay service through the addition of Unidata's Thematic Realtime Environmental Distributed Data Services (THREDDS), and the Repository for Archiving, Managing and Accessing Diverse DAta (RAMADDA) systems for the purpose of distributing IDD and locally generated products to the greater Unidata community.

Project Description:

The PSU Meteorology Department has been a user of Unidata data since program inception and has been an IDD participant since 1997. We recently upgraded our IDD relay service (hosted on idd.meteo.psu.edu) to the Tier-1 level for CONDUIT data, and currently provide IDD feeds to 14 external, downstream sites:

pollux.rcac.purdue.edu, castor.rcac.purdue.edu
aeolus.valpo.edu
wx.gmu.edu
lightning.msrc.sunysb.edu, thunder.msrc.sunysb.edu
cascade.atmos.albany.edu
omega.lsc.vsc.edu
idd.unl.edu
flightrisk.meas.ncsu.edu
shu.cup.edu
wxmcidas.csbf.nasa.gov
vortex.esc.brockport.edu
emo.unidata.ucar.edu
kepler.sca.uqam.ca
coriolis.met.tamu.edu

This relay currently ingests and relays all major IDD data streams. The next logical step forward is to consider augmenting our IDD service to the community with a THREDDS/RAMADDA data service. This service would include current data and short-term archives for a significant portion of the IDD data stream we currently ingest. We would also propose to add locally generated datasets such as BUFKIT data files, real-time, experimental forecast model output generated at Penn State, and Pennsylvania climate data. Although PSU Meteorology's investment in networking and staffing resources can support these proposed endeavors, our current hardware cannot.

We propose using Unidata Community Equipment Award funds to purchase computer hardware to advance this project with the following strategy:

1. Strengthen our current IDD relay service to meet new and future demands of the Unidata community.
2. Provide a backend storage capability for THREDDS/RAMADDA data
3. Provide computer resources to create the THREDDS/RAMADDA datasets
4. Provide computer resources to serve the data to external users

We first plan to add an additional IDD real server (Figure 1) to increase our ability to function as a reliable IDD data relay to the community. Next, a THREDDS/RAMADDA Local Data Manager (LDM) ingest system will be added to pull data from the IDD relay system (Figure 2 - top) and then decode and/or file the data onto the storage array cluster. The THREDDS/RAMADDA portal server will accept requests from the Unidata community and serve up requested data from the storage array cluster. In the initial configuration, the three storage servers will be cross-mounted via Network File System (NFS) version 4 onto the ingest system and the portal system. When parallel NFS becomes available, a metadata server will be added and the three storage array servers will be converted to parallel NFS data servers (Figure 2 - bottom). This action will unify the storage space and can be upwardly scaled to meet increasing demand by adding additional storage servers and portal servers.

The end result of the project will be a reliable THREDDS/ RAMADDA data server to compliment Penn State's Tier-1 IDD relay for the Eastern United States.

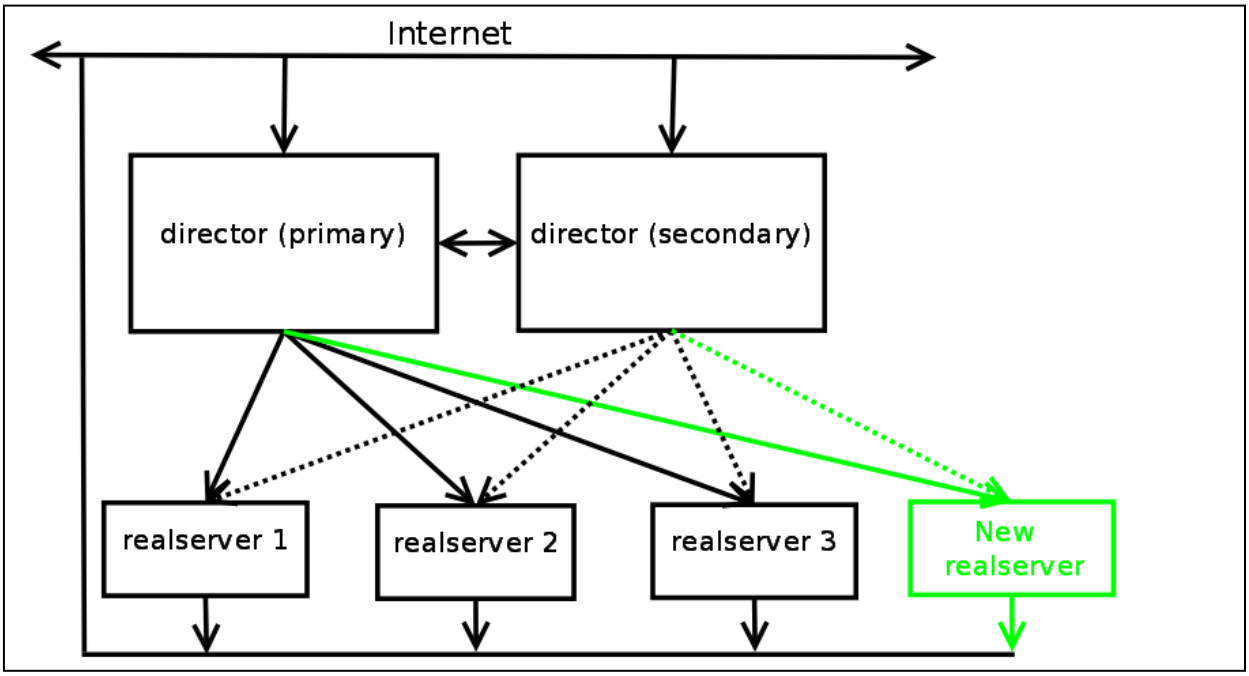


Figure 1. – New IDD data server to augment existing three bought with Unidata funds.

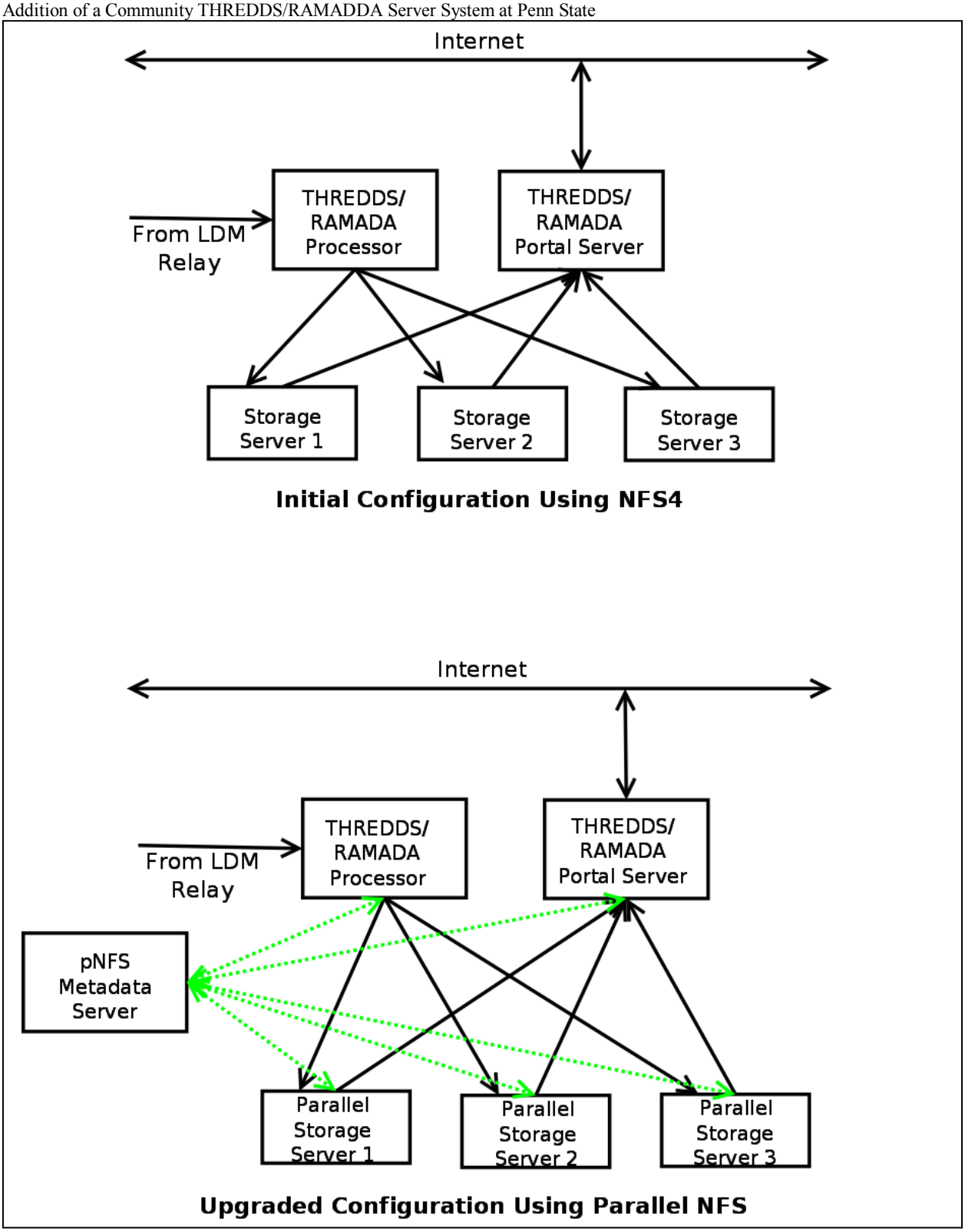


Figure 2. Initial (top) and upgraded (bottom) configuration of storage servers to support PSU Meteorology's new THREDDS/RAMADDA service to the community.

Budget:

The following hardware budget for the complete, fabricated system is proposed:

- 1 each: Dual Xeon E5520 2.26 Ghz quad-core processor, 16 GB memory system as an additional IDD relay "real" server. *Cost: $2,472*

- 3 each: Dual Xeon E5520 2.26 Ghz quad-core processor, 16 GB memory, 6 TB disk servers. This provides 9 TB of RAID 10, high-performance storage (expandable to 18 TB of RAID 10) across 3 servers. *Cost $11,925*

- 1 each: Dual Xeon E5520 2.26 Ghz quad-core processor, 16 GB memory server as an LDM data ingest/processor system for the creation of THREDDS/RAMADA datasets. *Cost $2,792*

- 1 each: Dual Xeon E5520 2.26 Ghz quad-core processor, 16 GB memory system as a THREDDS/RAMADA data portal server that interfaces to the high-performance 9 TB storage cluster and serves data to the Unidata community. *Cost: $2,432*

Total estimated hardware cost for the entire system: $19,621

Configuration notes:

1) Memory on these systems is deliberately kept on the low end to allow for future expansion at lower memory prices
2) Storage cluster systems are equipped with only 6 of 12 possible disks per chassis. This provides adequate storage to begin operations while providing opportunity for expansion. The RAID 10 arrangement of disks along with 3 independent systems provides for high performance. Should more storage be required after the chassis's are full, expansion racks with additional disks can be added if the systems are capable of handling the additional bandwidth load.
3) The storage servers will run the Solaris operating system with the ZFS file system to promote reliability and prevent filesystem instability issues. The THREDDS/RAMADA ingest and portal servers and the IDD relay server will run Linux.

Project Milestones:

By end July 2010	Award in house and ready for implementation
By August 2010	Acquire hardware, assemble and integrate into existing LDM/IDD data array
October 2010	Test THREDDS / RAMADDA server internally to ensure reliability and functionality with existing IDD data
December 2010	Integrate BUFKIT data stream into THREDDS/RAMADDA Begin allowing external access to THREDDS/RAMADDA
Spring 2011	Enable framework for the integration into THREDDS/IDD/ RAMADDA of realtime MM5 and WRF output and climate data
May 2011	Project Goals Completed

College of Earth and Mineral Sciences/The Pennsylvania State University
Addition of a Community THREDDS/RAMADDA Server System at Penn State
Unidata (NSF prime)
Project Dates: 8/1/2010 - 7/31/2011

	8/1/2010 - 7/31/2011	TOTAL
DIRECT COSTS		
Salaries (Category I)		
Chuck Pavloski, Principal Investigator	0	0
Per guidelines, no PI salary is permitted		
William Brune, Co-I	0	0
Per guidelines, no Co-I salary is permitted		
Chris Forest, Co-I	0	0
Per guidelines, no Co-I salary is permitted		
George Young, Co-I	0	0
Per guidelines, no Co-I salary is permitted		
Subtotal	0	0
Total Salaries and Wages	0	0
Modified Total Direct Costs	0	0
Fabricated Equip. - Community Thredds/RAMADDA Serv	2,472	2,472
Dual Xeon E5520 2.26 Ghz quad-core processor, 16 GB memory system as an additional relay "real" server		
Fabricated Equip. - Community Thredds/RAMADDA Serv	11,925	11,925
3 Dual Xeon E5520 2.26 Ghz quad-core processors, 16 GB memory, 6 TB disk servers. This provides 9 TB of RAID 10, high performance storage $3975 x 3 = $11,925		
Fabricated Equip. - Community Thredds/RAMADDA Serv	2,792	2,792

Proposal 04-024--UP-00269-2009-hmg108 Budget approved by hmg108 on 3/30/2010

	8/1/2010 - 7/31/2011	TOTAL
Dual Xeon E5520 2.26 Ghz quad core-processor, 16 GB memory server as an LDM data ingest/processor system for the creation of THREDDS/RAMADA datasets		
Fabricated Equip. - Community Thredds/RAMADDA Serv	2,432	2,432
Dual Xeon E5520 2.26 Ghz quad-core processor, 16 GB memory system as a THREDDS/RAMADA data portal server that interfaces to the high performance 9 TB storage cluster and serves data to the Unidata community		
Total Direct Costs	19,621	19,621
INDIRECT COSTS	0	0
TOTAL REQUESTED FROM SPONSOR	19,621	19,621

Recovery of F&A - Rates are negotiated and approved annually by the Office of Naval Research, the cognizant federal agency for this institution. Approved Rates for the period between July 1, 2009 and June 30, 2010 are at 48% of MTDC. The last approved rate quoted above has also been used for any project period occurring after July 1, 2010 and forward in lieu of negotiated rates for the forward period.