UNIDATA COMMUNITY EQUIPMENT REQUEST

TITLE: ESTABLISHMENT OF THREDDS SERVER AND RAMADDA FOR COMMUNITY ACCESS TO 78-YEAR WEATHER DATA ARCHIVE

by

William A. Gallus, Jr.
Title: Professor of Meteorology
Phone: (515) 294-2270
Fax: (515) 294-2196
Department of Geological and Atmospheric Sciences
3025 Agronomy Building
Iowa State University
Ames, IA 50011
Email: wgallus@iastate.edu

[redacted]

Department Chair: Carl Jacobson
Phone: (515) 294-4477
Fax: (515) 294-6049
Email: [redacted]

1 Project Summary

An extensive archive of meteorological data has been collected at Iowa State University and is currently accessible at the web site http://mtarchive.geol.iastate.edu. The data collection includes textual information (severe weather statements and other National Weather Service products), much numerical model output in gempak format, and gif images of weather maps created daily since 2006, along with gempak-format surface and upper air data back to 1933, much of which was provided to us by NOAA's National Severe Storms Laboratory. We have been told by users across the country and by Unidata staff that this archive is frequently used.

This proposal is designed to simplify the ability of the community to access all of the information by establishing a powerful server that will be both THREDDS and RAMADDA-equipped. Specifically, this proposal seeks funding to purchase a high-end machine with extensive memory and storage space.

2 Project Description

For the last 5 years, the meteorology program at Iowa State University has been archiving many of the products it receives through the Unidata datastream. On the web site mentioned above, users can find three types of data. Raw gempak-format model output is available for RUC211 grids every 3 hours, and AVN211, AVN213, AVN-THIN, ETA211, and ETA212, along with MOS, profiler, surface and upperair data. Text data is present and includes climate summaries, public information statements, severe weather statements and warnings, MOS data, record statements, hourly METARs, and raw upperair information. In addition, during this 5 year period, we have archived all of the images that we create in real-time and display on a weather forecasting page. These gif images include model forecast charts of many traditional parameters at many levels for the GFS, ECMWF, NAM, RUC, and UKMET models, MOS forecasts, profiler charts, radar images, lightning plots, surface parameters, skew-Ts, upperair charts, and satellite images. In addition, thanks to a gift from NOAA's National Severe Storms Laboratory, our archive was extended back to 1933 for surface data in gempak format, and back to 1946 for upperair data in gempak format. In the past year, we have added the new NMQ estimates of precipitation to the archive. Interested users can download these files on their own systems and then use gempak to create analyses.

In recent years, Unidata has created both RAMADDA and THREDDS to allow more convenient access to Unidata and other products. THREDDS, the Thematic Realtime Environmental Distributed Data Services project, is a middleware product that bridges the gap between data providers and data users. It relies on the concept of catalogs which describe on-line datasets. RAMADDA, the Repository for Archiving, Managing and Accessing Diverse Data, is a comprehensive content repository and publishing platorm. It allows for data file ingest, organization, meta-data creation, and access control.

An advantage of funding this proposal is that it will make available to the entire community what is perhaps the largest university-based weather data archive in existence. In addition, the meteorology program normally funds a team of undergraduate weather lab helpers during the academic year. These students are available to answer questions from students taking a large-lecture introductory meteorology course, but frequently have spare time on their hands. We will use these students to develop a catalog of interesting weather events that can help users find the cases they want to work with in this archive. We will also create a contributions folder that will allow the community to add images and other information for the appropriate days in the archive. We will set the permissions within RAMADDA so that the data itself cannot be tampered with, but outside users will be permitted to add to the contributions folder. We will also use some of the student team to add products to the archive.

2.1 Motivation

Although the existing archive already has been helpful to the community, the establishment of a THREDDS server on the new machine and installation of RAMADDA to access the archive will

facilitate its use and allow newer Unidata software such as IDV to be used to view the data. These recent developments at Unidata should allow broader community use of the archive.

The **educational and research benefits** of this proposal are large. The archive has been helpful at Iowa State in the design of weather lab exercises used in synpotic and mesoscale courses, and improved access should assist instructors around the world. In addition, easy access to the archived data will facilitate its use in research projects and in consulting work.

2.2 Serving the Unidata community

Unidata has made a concentrated effort to expand its access to a broader community through the development of THREDDS servers, RAMADDA, and IDV display software. Our installation of a THREDDS server and RAMADDA will allow more users to access the weather data, will help others access the gempak-formatted data via IDV, and will permit the archive to become richer as users can contribute to it themselves.

2.3 Importance of this project

This proposal directly addresses Unidata's current mission and its desire to broaden its community. The proposed activity fits well within the two core areas Unidata is emphasizing in this year's call for proposals. Not only does it make use of some of Unidata's most current technology but it will potentially broaden Unidata's user base as the newly archive NMQ data likely will be most desired by members of the hydrology community.

3 Computing Facility Overview

The machines requested would fit nicely into our dedicated computing facility. The computing room provides dedicated air handling and gigabit ethernet access with a direct link to Internet2 providing high speed access to the IDD. Iowa State University's commitment to Red Hat Enterprise Linux provides a stable platform for these systems to run on for many years to come. We believe the $7,500 we are requesting allows us to have enough memory and data storage to allow smooth access to the data and continued expansion of the archive for many years to come. We have some experience already with THREDDS servers and RAMADDA, currently running a THREDDS server on shared hardware with limited capacity, and using RAMADDA to transfer small amounts of WRF model generated QPF to collaborators at the University of Iowa who use it in a hydrological model. Our prior experiences should help us to quickly get the new system operational.

4 Budget

Iowa State University will not provide direct funds for the purchase of the equipment proposed. ISU will provide the support infrastructure for the successful development of these systems. Staff time will be provided by ISU for the installation and administration of these systems. In addition, we will be able to leverage some university funding that pays for our student weather lab helpers so that these students will also contribute to the RAMADDA archive.

4.1 Hardware Justification

Because of the vast size of the data archive, we believe that we need a machine for the THREDDS server and RAMADDA installation that will have large amounts of fast disk space, processing power, and memory. We estimate that our proposed system will work well for these needs.

	Item Description	Price
(1)	Dell Intel Xeon 5520 32GB of Memory 7 TB of SATA storage	
	Total	$7,500

5 Project Milestones

Assuming that this project was funded on 1 June 2010:

Date	Task
1 Jul 2010	Purchase and delivery of the Dell server.
1 Aug 2010	Testing and deployment of the server completed.
1 Aug 2010	Installation of THREDDS and RAMADDA completed
1 Sep 2010	All data accessible through THREDDS/RAMADDA with contributions being added
1 Jan 2011	Initial cataloging of many significant events completed