UNIDATA COMMUNITY EQUIPMENT REQUEST

Upgrade of Existing Computer Hardware to Facilitate Processing and Distribution of Large Oceanographic and Environmental Datasets from the Rutgers University Coastal Ocean Observation Laboratory

March 18, 2011

Principal Investigator: John Kerfoot
Title: Research Computing Specialist
Phone: (732) 932-6555 x527
Fax: (732) 932-1821
Email: kerfoot@marine.rutgers.edu
Institute of Marine and Coastal Sciences, Rutgers University
71 Dudley Road
New Brunswick, NJ 08901

[redacted]

Office of Research & Sponsored Programs: Sarah Dumais
Research Contract /Grant Specialist
Phone: (732) 932-0150 x2107
Fax: (732) 932-0162
Email: duamis@grants.rutgers.edu
3 Rutgers Plaza
New Brunswick, NJ 08901

[redacted]

Project Summary

Over 1.2 billion people (23% of the world's population) live within 100 meters of sea level and 150 kilometers from the coast (Small and Nichols 2003, Small and Cohen 2004) and these coastal communities are disproportionately important to their national economies (NOAA, 1998). Given the size and importance of the coastal economies and ports in the Mid-Atlantic Bight, understanding the physical and biological ocean processes impacting this region will lead to improvements in maritime safety and ecosystem based management strategies. The **R**utgers **U**niversity **C**oastal **O**cean **O**bservation **L**aboratory (**RU-COOL**) has assembled an extensive collection of real-time and historical oceanographic datasets dating back to 1997. A large portion of these datasets has been made available to the broader research community via Unidata's **T**hematic **R**eal-time **E**nvironmental **D**istributed **D**ata **S**ervices (**THREDDS**). This data sharing has resulted in significant improvements to existing numerical modeling and forecasting efforts for the Mid-Atlantic Bight as well as increased cooperation between federal, state and local agencies and institutions.

In order to meet the current and future increases in data processing and access demands on these datasets, RU-COOL seeks to purchase 2 new servers to upgrade the hardware currently performing these duties. The acquisition of these 2 servers will significantly improve our ability to host the entire archive as well as bring new datasets online as they are collected.

Project Description and Justification

Since 1997, RU-COOL (http://rucool.marine.rutgers.edu) has operated a regional coastal ocean observing network in an effort to provide the public as well as a wide range of federal, state, local research and educational institutions with real-time information products for improving safety at sea, search and rescue, ecosystem based management and understanding of oceanic processes in the Mid-Atlantic Bight and around the world. Over the last 15 years, the observatory has invested significant financial and human resources in 4 core hardware technologies:

- 2 satellite tracking antennas (L-band and X-band) for real-time acquisition of a variety of data products from the international satellite constellation.
- A high-frequency radar network (13 - 20 sites) for real-time measurements of surface current fields in harbors, estuaries and the coastal oceans.
- A fleet of more than 20 Slocum autonomous underwater gliders to measure sub-surface features in the coastal and open oceans around the world.
- Computer hardware and infrastructure for real-time acquisition, processing and dataset serving.

These investments have provided researchers from around the world with high-resolution spatial and temporal datasets used to study and understand the physical, biological and optical processes of the coastal and open oceans. RU-COOL has been a leader in the deployment and recovery of these platforms and real-time acquisition of the datasets.

As a member of the **IOOS** **M**id-**A**tlantic **R**egional **A**ssociation of **O**cean **O**bserving **S**ystems (**MARACOOS**, http://maracoos.org), RU-COOL is focusing on the primary mission of the NOAA **I**ntegrated **O**cean **O**bserving **S**ystem (**IOOS**, http://www.ioos.gov) program: "to lead the integration of ocean, coastal, and Great Lakes observing capabilities, in collaboration with Federal and non-Federal partners, to maximize access to data and generation of information products, inform decision making, and promote economic, environmental, and social benefits to our nation and the world." To begin accomplishing this goal, **RU-COOL** installed and maintains a THREDDS data server (http://tashtego.marine.rutgers.edu:8080/thredds/catalog.html) which serves the following datasets in real or near real-time:

- Hourly surface current fields from high frequency (**HF**) radar (real-time with a growing 2 year archive)
- Sea surface temperature from the **A**dvanced **V**ery **H**igh-**R**esolution **R**adiometer (**AVHRR**) (real-time with a growing 12 year archive)
- Sea surface temperature from the **Mo**derate **R**esolution **I**maging **S**pectroradiometer (**MODIS**) (real-time with a growing 2 year archive)
- Sub-surface temperature, salinity and optical observations from the glider fleet (real-time with a growing 4 year archive)

These real-time datasets are regularly accessed and used as inputs to a range of modeling efforts including, but not limited to, the United States Coast Guard's **S**hort **T**erm **P**redictive **S**ystem (**STPS**) and **S**earch **a**nd **R**escue **O**ptimal **P**lanning **S**ystem (**SAROPS**), **R**egional **O**cean **M**odeling **S**ystem (**ROMS**) and the NCEP **N**orth **A**merican **M**esoscale **M**odel (**NAM**). Results of the **ROMS** and **NAM** runs are also available via the THREDDS server.

In addition to providing real-time datasets to the broader research community and local, state and federal agencies, RU-COOL has embraced the role that undergraduate and graduate students play in making important contributions to environmental science and research. We believe that early entrainment and involvement of students in real-time ocean and atmospheric research efforts is critical. To help ensure this participation, it is our desire to make these real-time and archived datasets accessible to them early in their careers. By providing this access, we hope to accelerate the timeline over which the students can develop and polish the

software tools required. Their focus will then shift to answering questions rather than dataset acquisition, manipulation and organization.

Finally, there has been significant work and time invested by the broader Unidata community on the development of standards for representing spatially gridded datasets such as the HF Radar and Sea-Surface Temperature. However, the use and recent acceptance of autonomous underwater vehicles as legitimate ocean sampling platforms requires a similar level of effort and participation to develop corresponding standards. In 1999, RU-COOL began investing significant time, resources and effort in the development of a global ocean glider fleet. To date the fleet, which currently consists of over 20 gliders, has been deployed over 250 times around the world. The fleet has flown over 93,000km (2.3 times the circumference of the Earth) over 3,987 in-water days. We believe this to be the largest and most extensive fleet in the world with deployments in, among other regions, the Southern Ocean, the Mediterranean Sea, North America and Australia. In December 2009, RU-COOL became the first institution to successfully fly a glider across an ocean basin (http://rucool.marine.rutgers.edu/atlantic). This achievement was only possible through the participation and coordination of an international community of researchers and students. While we currently distribute deployment datasets from the Mid-Atlantic Bight (~60 deployments total) via THREDDS, we have yet to make all of the deployments available to the broader user-community for 2 primary reasons: 1) a lack of developed and accepted standards for representing this point/temporal/spatial hybrid dataset and 2) worries about overhead associated with increased access by the research community. We believe that ingestion of the unique sub-surface datasets provided by the gliders will result in increased model predictive accuracy and further understanding of local, regional and global ocean processes. We intend to make all deployments available in the hopes of accelerating the development of accepted **C**limate and **F**orecasting (**CF**) metadata conventions for these new types of datasets.

Due to a significant increase in both the size of and access demands to the datasets over the last year, the current hosting server (`tashtego.marine.rutgers.edu`) is beginning to show signs of age. Given the size of the datasets will continue to grow and the number of organizations accessing them will increase significantly over the coming years, we are requesting funding to purchase a replacement server which will scale well to future use. The current server went into service in 2007 and has 2 dual-core Opteron processors with 6 GB of RAM. It runs a THREDDS, ncWMS and ERDDAP server via Apache Tomcat. It also holds myroms.org and ocean-modeling.org tools, which includes 6 MySQL databases, 7 Subversion repositories with matching TRAC portals, secure and unsecure web services and several cronjobs to backup these sites and send daily emails to the ROMS user list. We envision using the new server to host the THREDDS, ncWMS and ERDDAP services while hosting the ROMS

services on the current server. This should reduce the load considerably on `tashtego` and minimize bottlenecks in data access.

We are also requesting funding to purchase an additional compute server which will allow us to process all of the current and future THREDDS datasets. We currently use a single server (`ironman.marine.rutgers.edu`), purchased in 2005, to perform this processing and to fulfill the data processing requirements of the individual members of RU-COOL (approximately 30 users, including staff, undergraduate and graduate students). The current compute server went into operation in 2006 and has 2 dual-core AMD Opteron (1.81 GHz) processors with 8 GB of RAM. As the number of users requiring access to this sever for data processing has increased greatly, this server now experiences long periods of high CPU utilization and load averages, resulting in a significant slowdown of overall data processing times. These increased times have significantly hampered our ability to process and provide access to the datasets in real-time.

The **I**nstitute of **M**arine and **C**oastal **S**ciences (**IMCS**) at Rutgers University currently utilizes an EMC Celerra NS-480 fileserver with ample available file space (> 1Tb) and expansion capabilities. For this reason, we have chosen hardware configurations that are low on hard drive space, instead opting to put the bulk of the cost into performance (processors and RAM). The servers will be housed in the existing IMCS computing facility, where they will be monitored and maintained by the system administration staff, guaranteeing high uptime at a minimum of additional resources.

Budget

We intend to purchase 2 servers. The first will host the THREDDS, ncWMS and ERDDAP servers. The selected configuration for this machine is as follows:

- Dell PowerEdge R610 with 2 Intel Xeon X5667 3.06 GHz processors
- 48GB, 1333MHz RDIMM RAM
- PERC H700 Integrated RAID Controller, 512Mb cache
- RAID1/RAID5 for H700 Controller
- 2 160GB SATA Hot Plug Hard Drives
- 4 250GB SATA Hot Plug Hard Drives

The total cost of this server is **$7,377.80**.

The second will be used to process the datasets which will be served via THREDDS. We have selected the following configuration for this machine:

- Dell PowerEdge R610 with 2 Intel Xeon X5680 3.33Ghz processors

- 96Gb 1333Mhz RDIMM RAM
- PERC H200 Integrated RAID Controller
- (2) 160GB SATA Hot Plug Hard Drives

The total coast of this server is **$9,059.82**.

The total cost of both servers as configured is **$16,437.62**.

Project Milestones

Assuming a funding date of June 1, 2011, we propose the following milestone timeline:

- June 2011: Purchase and delivery of the 2 new servers
- July 2011: Configuration (CentOS, THREDDS, ncWMS, ERDDAP) and of the 2 new servers
- September 2011: Deployment and testing completed
- October 2011: Current product catalog will be made available. Previous server will remain operational until this step is completed, preventing data dropouts.
- January 2012: Cataloging of glider datasets completed and accessible by the broader research community.

References

Small, C., and J.E. Cohen. 2004. Continental physiography, climate, and the global distribution of human populations. *Current Anthropology* 45(2):269–288.

Small, C., and R.J. Nicholls. 2003. A global analysis of human settlement in coastal zones. *Journal of Coastal Research* 19(3):584–599.

National Oceanic and Atmospheric Administration (NOAA). 1998. *Year of the Ocean. Discussion Papers*. National Oceanic and Atmospheric Administration, Silver Spring, MD.