Unidata Community Equipment Awards Cover Sheet
Title: University of Miami data server 2017

A proposal submitted to Unidata Program Center (UPC) Equipment Awards

Project Dates:
May 1, 2017 – April 30, 2018

Budget:
University of Miami: $18,649

Principal Investigator:
Dr. Brian Mapes, Associate Professor
Department of Atmospheric Sciences
University of Miami - Rosenstiel School of Marine and Atmospheric Science
4600 Rickenbacker Causeway, Miami, FL 33149-1031
phone: (305) 421-4275
fax: (305) 421-4696
email: mapes@miami.edu

Signature: _____ **Date:** Mar 29, 2017

Institutional Representative:

David Nolan
Chair, Department of Atmospheric Sciences
University of Miami - Rosenstiel School of Marine and Atmospheric Science
4600 Rickenbacker Causeway, SA 111
Miami, FL 33149-1031
phone: (305) 421-4930
fax: (305) 421-4696
email: dnolan@miami.edu

Signature: _____ **Date:** 3/29/2017

B. Project Summary:

We propose to obtain a new Linux machine, to shadow and then replace (upon failure) our current aging machine (*weather.rsmas.miami.edu*), a significant geographical node in Unidata's IDD network. *Weather* is also a large public data repository, a showcase for Unidata technologies both to the Miami research community and to the online community at large. The continuity of *weather.rsmas* is important: For instance, it is advertized in my YouTube videos enthusing about Unidata's IDV and *The Mapes IDV Collection*. That Collection now includes visualizations for our new Lab Manual supplement to a popular synoptic meteorology textbook (Lackmann). The manual goes on sale soon from AMS press, and many users of the Lackmann textbook. A *weather.rsmas* failure as fall classes start hitting the server would be embarrassing and inconvenient, so we want to obtain a hot spare this summer. We have had one scare, and there is a concerning upswing in opaque RAID controller warning messages.

We will put the aging machine (which will be refurbished at Miami expense upon first failure) to good use. For intellectual merit, it will serve as a backup of our substantial (~50 TB) community data repository, including large dataset offerings that showcase Unidata technologies to climate, oceanography, field-campaign, and other potential user bases at the margin of Unidata's traditional realtime weather community. Community service functions such as Internet Data Distribution will be sustained and enhanced, as new capacity may allow longer rolling archives and larger new LDM feeds (GFS 0.25 deg, HRRR, GOES-R). Redundant capacity will also be used as a test platform for the PI's (collaborative with Unidata) Earthcube research project to integrate Unidata's powerful IDV with other technologies (notebooks and repositories), to make Unidata more relevant to statistical and climatic research. High-capacity internal testing (with goodwill) of any software stack is crucial to avoid spoiling the broader community's patience via a disappointing rollout. The spare capacity will facilitate such testing, using classrooms as synchronous loads. These students, at the bleeding edge of Jupyter Notebook-IDV integration, will also be entrained into the production of promotional materials such as the PI's YouTube videos.

C. Project Description

We propose to obtain a hot spare and ready replacement for our aging Linux server (*weather.rsmas.miami.edu*). The new machine will be set up as a quasi-clone of the existing machine, including a daily synchronized copy of many TB of high value data, which was itself set up as a motherlode.ucar.edu clone, for rapid failover and minimal maintenance burden. Unidata staff who help manage the resource for community as well as local good will have capable Miami sysadmin staff assistance, where a good working relationship exists already. After 4 years of service, the weather.rsmas machine is giving off more and more RAID warnings. It could still have years of service life, but redundancy is needed to prevent a major service interruption upon failure. In fact we have had one scare already (two drives failed in the RAID, requiring a nervous week or two to rebuild). The virtues of such redundancy for reliability, including cross mounted disks with a daily synchronized copy of many TB of the highest value data, will not be belabored here, but they are significant.

In the meantime, before failure of the old system, the overlapping computer and storage capacity will be highly valued and put to good use toward intellectual merit, community data services, and promotion of Unidata technologies to both researchers in adjacent but under-engaged communities (climate, oceanography) and the world at large (through YouTube videos, and the hoped-for reliability of the PI's new textbook-related IDV exercises, in the upcoming Lackmann, Mapes, Tyle lab manual, see https://tinyurl.com/LMTmanual.

Special Considerations for 2017

This proposal addresses Special Consideration items detailed in the RFP http://www.unidata.ucar.edu/community/equipaward/RFP2017.html, especially these:

- Production of online training materials (code notebooks, video tutorials, online documentation, or similar resources) that can be shared with the Unidata community

- Installation of equipment that provides student access to modern visualization tools such as AWIPS, IDV, or Jupyter notebooks

In particular, PI Mapes has a substantial presence in online IDV training materials. Besides the LMT Lab Manual described above, he has produced many IDV-based resources, under the name *Mapes IDV Collection* (http://www.rsmas.miami.edu/users/bmapes/MapesIDVcollection.html). These resources include well-received introductory YouTube videos, but also a substantial deeper body of tools. A library of curated bundles is the most visible. My self-updating IDV plugin includes not only a Favorites folder pointing to those bundles, but also advanced formulas for statistics, filtering, interactive scatterplot, publication-quality vertical sections on a pressure coordinate, and many other useful additions to the IDV's basic capabilities, aimed at keeping it relevant for modern research and not just weather depiction. The PI is eager to keep up the effort, but worried about the frailty of its hardware underpinnings. Cloud solutions do not seem ready yet to replace "bare metal" of Linux racks hosted locally, but the frailty of my current machine (from a 2013 Unidata equipment grant) is wearing on me.

Another truly novel contribution by the PI's group is the *IDV notebook* (see examples and code at https://github.com/suvarchal/JyIDV). This allows an IDV session to be operated and annotated from a Jupyter Notebook environment (http://jupyter.org). This coupling transforms an IDV session from a personal experience into a documented and repeatable exploration. The PI has collaborated with current and former Unidata staff, including some consulting paid at my own expense as well as an Earthcube grant with co-PI Yuan Ho at Unidata (the main IDV developer), to help firm up this coupling, and bring IDV strongly into the world of Jupyter. These technologies will need explaining, or really *showing*, and continuity of the platform for their development and testing is essential.

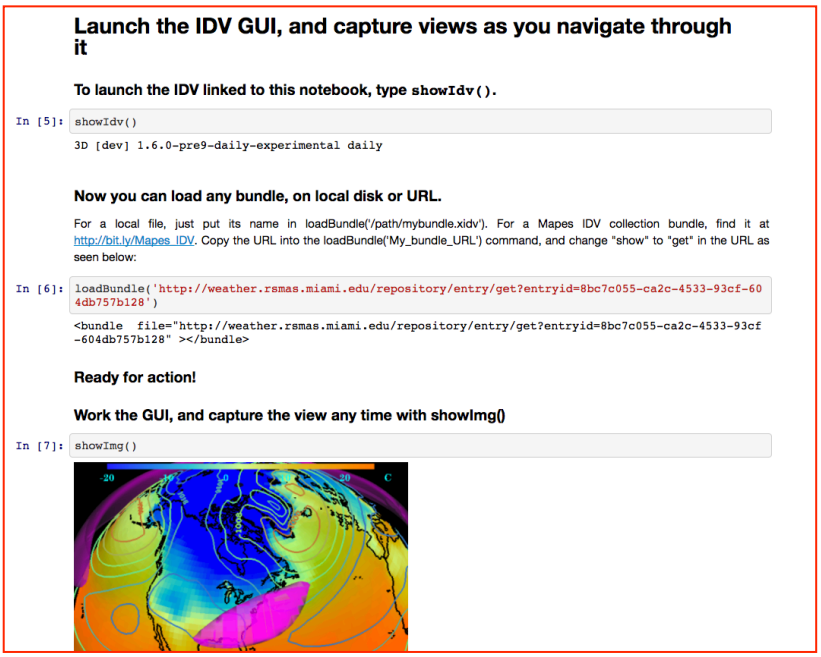


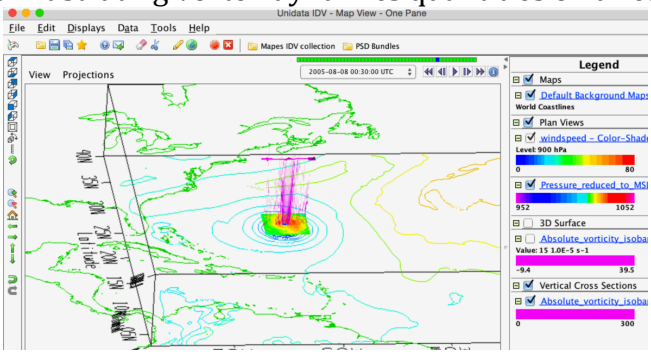
Figure 1: IDV Notebook, an innovation from the PI's group (https://github.com/suvarchal/JyIDV). It couples the IDV to Jupyter notebooks to create a saved, replicable, literate (human-reader-oriented) wrapper around the otherwise ephemeral GUI experience.

Based partly on this prototype, the PI received (collaborative with Unidata) an NSF-Earthcube research grant to further integrate Unidata's IDV with notebooks and repositories, to build a software stack that is much more useful than its parts to analysis challenges in the Big Data age. Specifically, our goal is to fulfill the vision of *Visualization for Algorithm Development*, the original basis for the IDV's underpinnings (VisAD is in fact the corresponding Java library name). Algorithms are essential to extract meaning from big datasets, but edge cases of their performance need to be examined (Visualized) carefully, in order to further their Development. In this sense, we are fulfilling very elemental goals at the base of Unidata's history. A stable computing platform with some available capacity will allow both research and classroom testing of these "Unidata-plus" software stacks. Classroom testing is especially powerful: Besides being educational for the students, it quickly exposes jargon and workarounds that experts have ceased to notice they are using.

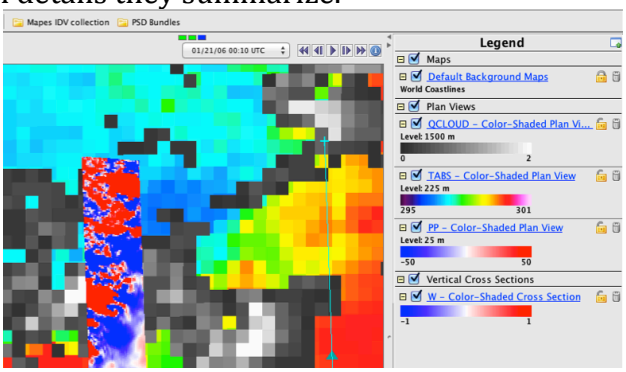
Research uses of the weather.rsmas.miami.edu RAMADDA respository will be sustained by the safe backup and enhanced with any surplus capacity. In addition to the realtime

weather data streams, we host a number of datasets that utilize RAMADDA's services to aid community goals. Readers are invited to browse at http://tinyurl.com/Miami-RAMADDA-datasets, but here is a sample:

- Large nested hurricane "nature run" simulation datasets, with associated IDV bundles – valuable for showcasing multiscale hurricane dynamics, and for computing and illustrating vortex dynamics quantities on a realistic 3D dataset.



- Huge "Giga-LES" (2048 x2048 x128) convection simulation outputs, and coarse-grainings (like 256x256), all with associated IDV bundles displaying both sets -- valuable for turbulence studies, allowing statistics to be superposed on the gridpoint-level details they summarize.



- Datasets from the DYNAMO-AMIE field campaign, all gathered together, with IDV bundles that overlay and juxtapose them. While these data can be downloaded one by one from "catalogs" elsewhere on the web, data integration is so much easier when they are hosted on a repository with services.
- Energy and Water Cycle datasets archive, synthesizing multiple estimates of difficult to observe quantities in one place for easy comparison (in collaboration with the NASA NEWS program and its research community).

- Oceanographic datasets (coming soon)

A common theme is that research communities at the margins of Unidata's traditional meteorology roots are being educated about the power of these tools, and some

of the limitations of the old "download, folderize, and write code" model of science for data *integration*, an increasingly important part of research.

Personnel, expertise and institutional setting

Professor Mapes is an educator and researcher with experience and skills in computing and scientific programming. He serves on the IDV steering committee, and until recently on Unidata's Strategic Advisory Committee. He is also PI on an EarthCube project, with UM oceanographer Mohammed Iskandarani and Unidata co-PI Yuan Ho, designed to keep the strengths of the IDV relevant in the age of Python and Jupyter notebooks, as sketched briefly above.

The Department of Atmospheric Science consists of 12 faculty and more than 50 students, undergraduate and graduate. There is no central computing for this group. The Rosenstiel School of Marine and Atmospheric Sciences (RSMAS) has a faculty of about 80, and total research computer users in the hundreds. The RSMAS Computing Facility (RCF) has University-leading skills due to the data-intensive nature of our field. In addition, the University maintains a Center for Computational Science (CCS). Together these facilities and expertise and services (from power to bandwidth to maintenance) are more than adequate to successfully deploy and maintain the proposed purchase.

D. Budget

The total requested funding for this proposal is $18,649, a capital equipment line exempt from indirect costs. This quote was based on the existing *weather.rsmas.miami.edu* machine, and uses the same vendor. Again, our goal is low-maintenance continuation of existing services, enabling a safe platform for research and outreach novelties, while continuing to foster the core Unidata-related community goods.

Server-room rack space, electricity and cooling will be provided by the School as an expense of its research and instructional mission. Optionally, the University has offered to host the machine at its University's Center for Computational Science, in a crosstown facility with major regional-infrastructure grade internet lines and electrical and physical security. This decision will be made upon recommendation for funding, in consultation

with experts at Unidata, the School, and the University. The PI contributes time and travel costs, including frequent visits to Unidata, and is active (to the point of a personal mission) in transferring knowledge about Unidata technologies and services to faculty, students and other interested members of the UM-RSMAS community and beyond. This time is considered an investment in my teaching and research.

E. Project Milestones

We expect a timeline like the following:

- May 2017 – Notification of award and purchase of quoted machine.
- June 2017 – Install equipment and software, perhaps literally cloning an image of existing *weather.rsmas* as a starting point. Consider broadening and deepening IDD subscriptions (LDM feeds) to utilize increased capacities.
- July 2017: Prepare and test failover schemes for *weather.rsmas.miami.edu* IP.
- Summer 2017: LMT textbook lab manual released... In preparation for autumn semester synoptic meteorology courses, retest all materials for readiness and prepare backup plans, with Albany (via collaborator K. Tyle) as a further failsafe. Prepare and post more YouTube videos and other training materials.
- Autumn 2017: Problem-solve any LMT textbook manual issues as they arise.
- Ongoing: Utilize server for hosting of additional openly shared data sources such as high-resolution ocean grids, testing and sharing improvements to the *Mapes IDV Collection*, new capabilities (such as *IDV Notebook*), etc.

- April 2018 – Submit a final report, in the form of a short article, to the Unidata Program Center that describes: 1) Lessons learned about computer setup for failureproofing, 2) Activities and services to the Miami research and broader community, made possible by the availability of overlapping capacity, 3) Assessments on approaches that worked/failed in order to improve the program, and 4) Recommendations and benefits of participation in the program.