Unidata Community Equipment Awards Cover Sheet

Proposal Title

A Jupyter server for the Oregon State University Climate Science Program

March 22, 2016

Principal Investigator Name: Karen M. Shell
Title: Associate Professor, Climate Science Program Head
Institution: Oregon State University
Telephone number: 541-737-0980
FAX number: 541-737-2064
Street Address: College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, 104 CEOAS Admin Bldg., Corvallis, OR 97331
Email address: kshell@coas.oregonstate.edu

[redacted]

Co-Principal Investigator: Simon de Szoeke, Associate Professor, Atmospheric Sciences, sdeszoek@coas.oregonstate.edu

Co-Investigator: Robert Kennedy, Assistant Professor, Geography, rkennedy@coas.oregonstate.edu

Co-Investigator: Jamon Van Den Hoek, Assistant Professor, Geography, vandenhj@oregonstate.edu

Institutional Authorizing Official: Patricia A Hawk
Title: Asst. Vice President for Research
Phone: (541)737-4933
Fax: (541)737-3093
sponsored.programs@oregonstate.edu

Acting for:

[redacted]

A Jupyter server for the Oregon State University Climate Science Program

B. Project Summary

The College of Earth, Ocean, and Atmospheric Sciences (CEOAS) at Oregon State University (OSU) is requesting funds to purchase a **Jupyter Notebook server** for use across the Climate and related Earth Sciences curriculum and research enterprise. The server will provide an invaluable learning tool for students in data-intensive courses such as Climate Data Analysis, Climate Modeling, and Geospatial Programming. It will facilitate experiential learning opportunities outside of class and provide a platform for students to develop their own visualization and data-serving products. As the first Jupyter/Python server in the College, it will serve as a pilot project for CEOAS educators and researchers to explore the effectiveness of cloud-based storage, analysis, and visualization for their needs. The server will also enable the development of interfaces for distributing data to the broader geoscience community.

C. Project Description

This project proposes a **Jupyter Notebook server running Python** that will provide a standard interface for data-proximate analysis and visualization across the Climate and other Earth Sciences curriculum. Students will become increasingly proficient in the use of Jupyter/Python as they progress through their programs of study, with some eventually using the server for their senior research projects. This project will also develop the capacity for using Jupyter servers in educational and research settings for CEOAS faculty and staff. This project will serve as the first instance of a group within the College running a dedicated Jupyter Notebook server and will thus be a prototype and resource for wider implementation.

The server will be intended for three main audiences: Climate/Earth Science students, CEOAS researchers, and external geoscientists. The primary goal is **educational**: to enable meaningful climate/geospatial data experiences for undergraduate Earth Science students. The secondary goal is to streamline **research** by providing a usable data analysis and visualization platform for CEOAS graduate students, postdocs, and researchers who wish to migrate their research workflow to a Jupyter server. Finally, over time, students will develop **data-serving** interfaces for data hosted by the server to the broader Unidata and geoscience community. Figure 1 shows how the data-proximate analysis capabilities of the Jupyter notebook server meaningfully connect student-researchers and geoscience data sets.

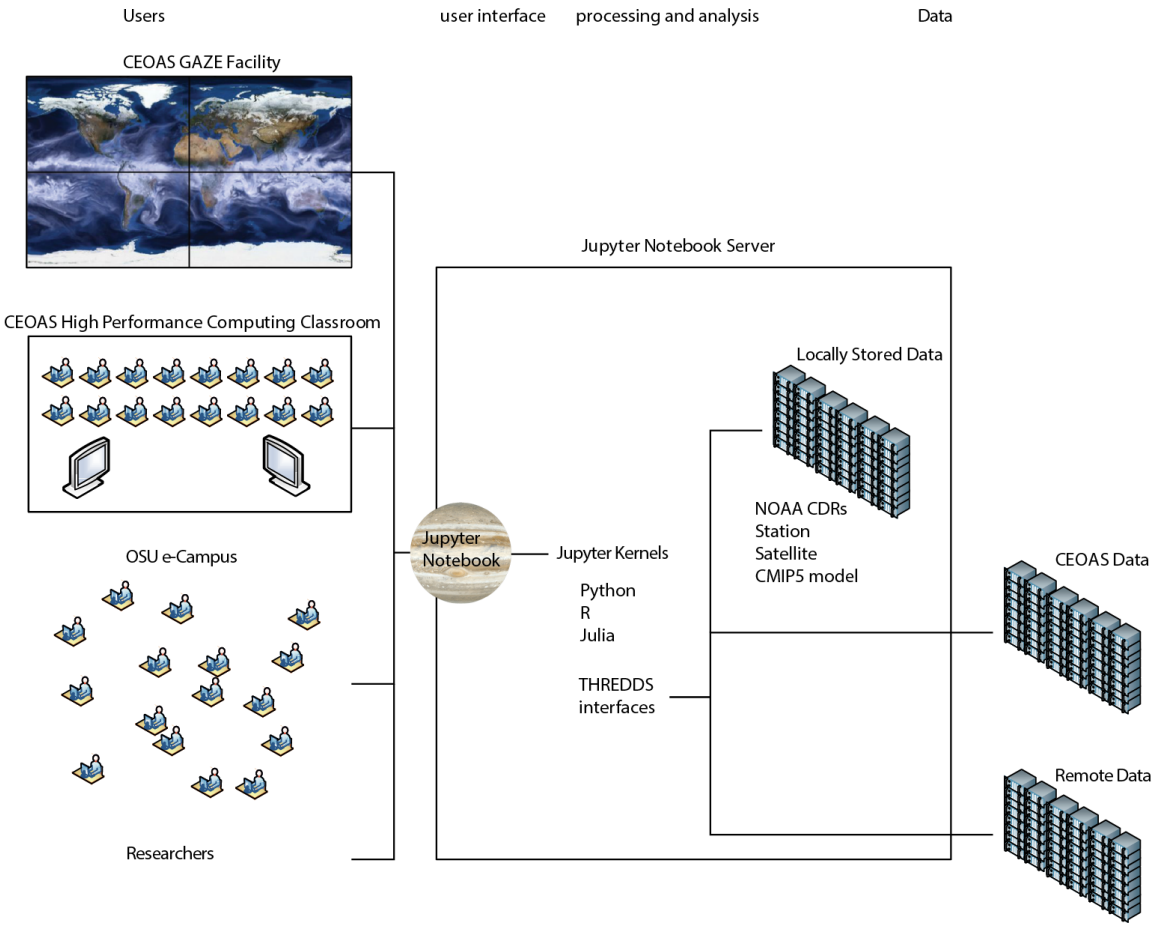


Figure 1: Relationships of diverse users to data sets provided by the the Jupyter Notebook Server.

This project will help CEOAS faculty develop experience using Jupyter Notebook as a tool for teaching data analysis and recording methodologies (e.g., digital lab notebooks). With instructor feedback, students can learn to add descriptions and discussions inline with code and plots in their notebooks. Additionally, having all student projects on a single server, with a single coding environment, will facilitate feedback and assistance (e.g., finding bugs) from peers and instructors. Students can easily share scripts and notebooks with each other to see various methodologies, or to compare the results of using different data sets (e.g., different climate models) in an analysis. Students produce "portfolios" of their analyses that they can share with potential employers.

CEOAS's new Climate Science undergraduate program includes new and revised classes highlighting experiential learning as the students work with real data. The server will introduce sophomore and junior Climate Sciences students to Python and Jupyter in Climate Data Analysis (ATS 301), to be first taught by the PI in Fall 2016. Students will build on these skills in later classes, like the senior-level Climate Modeling (ATS 421). Field data collected during classes such as Observing Climate (ATS 295) and Meteorology (ATS 301) will be stored on the server, so that students can compare observations to those from previous years. By their senior year, climate students will

have developed proficiency in using the platform for a range of purposes, allowing smooth transitions to independent research and projects with faculty members.

Two courses in CEOAS's Geography program, a senior/graduate level Python-focused geospatial programming class (GEOG 578) and a follow-on class in experimental satellite image analysis and advanced geospatial programming (GEOG 599), will use the Jupyter server initially. Co-Investigators Kennedy and Van Den Hoek will bring core satellite time series data for monitoring terrestrial carbon flux and land use/land cover changes to the server, and they will also work with students to develop data-serving map interfaces to allow students to follow the analysis arc from data exploration and analysis through visualization and distribution.

While most of these classes are standard in-person classes, the College offers a number of classes online through OSU's large, nationally ranked e-campus program (top 10, U.S. News & World Report). The Jupyter Notebook server will be particularly helpful here, allowing remote students from around the country (or world) to perform cloud-based computing through the Web browsers of their personal computers or devices. For example, an online version of GEOG 578 is part of the Geographic Information Science Graduate Certificate. Centralization of course activities to the Jupyter Notebook server will streamline and strengthen the class.

This project will **contribute to the capacity of the Unidata community** to transform undergraduate geoscience education, towards expansion of a data-literate student population and workforce. In addition to learning lessons about configuring and running a Jupyter server for educational purposes, this project will build a library of Python notebooks for analyzing climate and geospatial data, which we can make available to the geoscience education community. Furthermore, given the interdisciplinary nature of climate science, this project will **broaden the Unidata community scope and capabilities**, incorporating data from outside atmospheric science (especially the satellite data of the Co-Investigators). The Jupyter Notebook format allows for these data analysis and visualization procedures to be shared, and the server will support the development of interfaces for distributing land surface datasets.

Equipment description

The requested Jupyter Notebook and Computational Server will have
24 processors, 512 GB system memory, 24 TB of local disk space, and a 72 TB disk array.

Item 1: Dell Power Edge R730 Server

- Total Processors: 24
- 2 x12C/24T Intel Xeon E5-2690 v3 2.6GHz,
- 30M Cache
- 512 GB System Memory
- 6 x 4 TB Drives

- PERC H830 RAID Adapter for External MD Storage Array
- 3 Year Service agreement

Item 2: Dell MD 1400 Storage Array
- Total Storage 72 TB
- 12 x 6 TB Drives
- 3 Year Service agreement

CEOAS will provide all necessary infrastructure, including racks and networking integration, and CEOAS staff will provide planning, construction, setup, classroom integration, and testing of the facility, with assistance from Unidata staff. The PI and co-Is will configure the server as necessary for their respective classes. Priority will be given to educational uses. Because this proposal does not fund staff time, faculty will be developing experiential learning activities for classes they have already taught or developed as part of their normal teaching roles. With their own external support, faculty may use the server to explore cloud-based Jupyter Notebook computing in their research or develop pilot projects for data serving.

How the equipment will benefit Earth sciences education

Educational programs within the College of Earth, Ocean, and Atmospheric Sciences have strong data analysis, visualization, and distribution components. Climate science is data-rich and inherently interdisciplinary, involving atmospheric science, oceanography, geology, and geography. Climate scientists collect make detailed measurements, study satellite and ice core data, and simulate the climate using advanced computer models. All students in the Climate Science Option (http://ceoas.oregonstate.edu/earthsci/climate/) of the Earth Sciences Bachelor of Science program participate in hands-on experiential learning through field classes, project-focused classes such as Climate Data Analysis and Climate Modeling, and senior research or internships. In addition to the climate sciences, Geography faculty will use the Jupyter server as an integral component of their undergraduate and graduate programs.

The core capability of the server is to provide educational access to extensive Earth sciences data sets through Jupyter notebooks running Python kernels. With an array of user-contributed libraries for analysis and visualization, Python is open-source and widely used in education, government, and industrial sectors within and beyond the Earth sciences. The server will store climate, atmospheric, and surface data sets chosen to meet the learning outcomes of the relevant courses. These data sets will include NOAA Climate Data Records, paleoclimate proxy time series, CMIP5 model output, and other weather and climate data sets hosted by Unidata or the NCAR CISL Research Data Archive. The Jupyter notebook server allows users to download, subset, format, analyze, and visualize data quickly within a seamless flexible development environment. The web notebook interface permits users distributed across and beyond the OSU campus to use the server to efficiently perform scientific computations at the location these data sets are stored, reducing network traffic.

The Jupyter notebook interface promotes learning by facilitating the student data analysis experience. Its common user interface enables seasoned Earth scientists to mentor fledgling scientists. The simple and powerful notebook interface also helps Earth scientists migrate from other development environments to adopt effective new open-source analysis tools.

Python is likely to remain a valuable tool for the future climate and weather services workforce, and we expect much of the data analysis to use Python. Additionally, we intend to extend the server's appeal to a broader cross section of scientific users by providing R and Julia kernels. Both languages are used by CEOAS researchers and educators.

The data sets that can be accessed from the server's Jupyter notebook environment can be extended from the locally stored data to networked data sets using the THREDDS protocol, thus opening access to many more data sets served by Unidata, NCAR, and other participating hosts. This may be as simple as importing the Siphon package into Python. We intend to consult with Unidata experts to enable the Jupyter notebook and its kernels to most efficiently access data provided by THREDDS Data Servers and their NetCDF Subset Services.

Furnished with Unidata and other open source software, the proposed hardware provides new enhanced data sharing and analysis capabilities to a broad group of new scientific educational and research users.

Existing computing facilities

CEOAS has 90 teaching and research faculty, 26 research staff, 750 Earth Sciences and Environmental Sciences undergraduate students (including 320 online students), and 180 graduate students. The CEOAS Environmental Computing Center (ECC) is the nerve center for the College's computing. The center contains modern, high-throughput computing-class infrastructure in order to facilitate, deliver, and meet the challenges of next-generation instructional and scientific workflows.

Integrated into to the ECC is a 16 seat high-performance instructional classroom (Figure 2). The classroom contains 18 Apple Macintosh Pro Workstations as well as an Instructor podium connected to two 60-in 2K LCD presentation monitors. Each Workstation is equipped with a Quad-Core Intel Xeon E5 processor running at 3.7 GHZ, 32 GB of system memory, and Dual AMD Firepro Graphics processors each with 1280 stream processors.

The classroom is connected end-to-end, to a low-latency 20GbE network crossbar switching fabric allowing for high-bandwidth intra-Data Center and classroom connectivity for moving large data sets between resources. From the CEOAS local network domain, the college's infrastructure is connected to the Oregon State campus network, which in turn is connected to the commodity Internet as well as Internet-2.

The Jupyter Server and associated storage will be housed in the College's ECC, a 4500 sq ft data center that has been engineered and built to deliver reliable and scalable power, cooling and network infrastructure. Power is distributed to individually managed server racks arranged in a hot/cold aisle layout. Current top-of-rack TOR implementation includes both 10 and 40GB Ethernet optimized to deliver high-bandwidth I/O connectivity. In addition, the entire data center is backed up by a battery-based uninterruptible power supply (UPS) with an emergency generator capable of running 72 hours before needing refueling. At the heart of the Environmental Data Center are supercomputer-class machines ranging from large shared memory systems to tightly coupled clusters with and without GPU-enabled acceleration providing a wide range of services, including compute, storage, application, print, project and scratch services.



Figure 2: CEOAS high-performance instructional classroom

The College and Co-I Van Den Hoek were recently awarded a competitive $98K OSU Learning Innovation Grant to construct the GAZE (Geospatial Analysis and visualiZation for Education) facility. Located in a fully networked 1,750 sq ft classroom, GAZE will be a shared space that promotes spatial reasoning and understanding of human-environment interactions and Earth processes through data-driven visualizations and hands-on learning. The centerpiece of the GAZE facility is the Hyperwall, an immersive (12 ft wide, 6 ft tall) visualization system composed of six 60" monitors (4k definition) for displaying dynamic Earth processes and imagery. The proposed Jupyter/Python server will seamlessly integrate with the Hyperwall's controller workstation (a Zeon OctoCore CPU, 256 GB RAM, an Nvidia Quadro M6000 processor, and 10 Gigabit dual network cards) to deliver geoscience datasets to be visualized for in-class education as well as public outreach. In the GAZE Facility, the Jupyter server will support collaborative problem solving and data exploration, through on-demand visualization, and aid students in collaborative interactive programming, e.g., for comparing climate change scenarios or land cover change processes. This competitive grant demonstrates OSU's commitment to cutting-edge geosciences cyberinfrastructure. The visualization facility will be operational in time for Fall 2016 classes.

D. Budget

Funds are requested for a Dell Power Edge R730 Server ($13,025.98) and a Dell MD 1400 Storage Array ($5,967.60), for a total of $18,994, based on attached quotes (below) dated 3/15/2016.

E. Project Milestones

We have already received vendor quotes. While these will expire before notification of an award (i.e., after 30 days), we will re-request quotes immediately, in consultation with Unidata personnel, so that equipment can be received and installed in the ECC before the start of the Fall quarter (end of September). The server will first be used in the Fall 2016 Climate Data Analysis class (taught by the PI) as a test case and then be made available to other classes as capacity allows. The ECC already contains rack space and power/cooling capacity/infrastructure. Thus, there are no dependencies that would alter the project goals and deadlines.

GROUP: 1 QUANTITY: 1 SYSTEM PRICE: $12,995.99 GROUP TOTAL: $12,995.99

Description	Quantity
PowerEdge R730 Server (210-ACXU)	1
PowerEdge R730/R730xd Motherboard (591-BBCH)	1
Dell Hardware Limited Warranty Plus On Site Service (976-8706)	1
Basic Hardware Services: Business Hours (5X10) Next Business Day On Site Hardware Warranty Repair 3 Year (976-8778)	1
Dell ProSupport Service Offering Declined (991-2878)	1
Declined recommended ProSupport service - Call your Dell Sales Rep if Upgrade Needed (996-8029)	1
On-Site Installation Declined (900-9997)	1
US Order (332-1286)	1
On-Site Installation Declined (900-9997)	1
PowerEdge R730 Shipping (340-AKKB)	1
R730/xd PCIe Riser 2, Center (330-BBCO)	1
R730 PCIe Riser 3, Left (330-BBCQ)	1
R730/xd PCIe Riser 1, Right (330-BBCR)	1
Broadcom 5720 QP 1Gb Network Daughter Card (540-BBBW)	1
PERC H830 RAID Adapter for External MD14XX only, 2GB NV Cache, Low Profile (405-AAEZ)	1
iDRAC8 Express, integrated Dell Remote Access Controller, Express (385-BBHN)	1
Chassis with up to 8, 3.5" Hard Drives (350-BBEO)	1
No Bezel (350-BBBW)	1
Performance BIOS Settings (384-BBBL)	1
UEFI BIOS (800-BBDM)	1
RAID 6 for H730/H730P (4-16 HDDs or SSDs) (780-BBJW)	1
PERC H730 Integrated RAID Controller, 1GB Cache (405-AAEG)	1
Intel Xeon E5-2690 v3 2.6GHz,30M Cache,9.60GT/s QPI,Turbo,HT,12C/24T (135W) Max Mem 2133MHz (338-BFFL)	1
Upgrade to Two Intel Xeon E5-2690 v3 2.6GHz,30M Cache,9.60GT/s QPI,Turbo,HT,12C/24T (135W) (374-BBGS)	1
32GB RDIMM, 2133 MT/s, Dual Rank, x4 Data Width (370-ABVW)	16
2133MT/s RDIMMs (370-ABUF)	1
Performance Optimized (370-AAIP)	1
4TB 7.2K RPM SATA 6Gbps 3.5in Hot-plug Hard Drive,13G (400-AEGJ)	6
Electronic System Documentation and OpenManage DVD Kit, PowerEdge R730/xd (631-AAJG)	1
DVD ROM, SATA, INTERNAL (429-AAPU)	1
ReadyRails Sliding Rails Without Cable Management Arm (770-BBBQ)	1
Dual, Hot-plug, Redundant Power Supply (1+1), 750W (450-ADWS)	1
C13 to C14, PDU Style, 12 AMP, 2 Feet (.6m) Power Cord, North America (492-BBDH)	2
No Operating System (619-ABVR)	1
No Media Required (421-5736)	1
DIMM Blanks for System with 2 Processors (370-ABWE)	1
Standard Heatsink for PowerEdge R730/R730xd (374-BBHM)	1
Standard Heatsink for PowerEdge R730/R730xd (374-BBHM)	1

***Total Purchase Price:**	**$13,025.98**
Product Subtotal:	$12,995.99
Tax:	$0.00
Shipping & Handling:	$29.99
State Environmental Fee:	$0.00
Shipping Method:	LTL 5 DAY OR LESS
	(Amount denoted in $)*

GROUP: 1 QUANTITY: 1 SYSTEM PRICE: $5,947.61 GROUP TOTAL: $5,947.61

Description	Quanti
Dell Storage MD1400, 12 HDs 3.5", Rackmount, 2Us (210-ACZB)	1
Dell Hardware Limited Warranty Plus On Site Service Initial Year (976-9985)	1
Dell Hardware Limited Warranty Plus On Site Service Extended Year (976-9986)	1
ProSupport: Next Business Day Onsite Service After Problem Diagnosis, Initial Year (976-9993)	1
ProSupport: Next Business Day Onsite Service After Problem Diagnosis, 2 Year Extended (976-9994)	1
ProSupport: 7x24 HW / SW Tech Support and Assistance, 3 Year (977-7802)	1
Thank you choosing Dell ProSupport. For tech support, visit http://www.dell.com/support or call 1-800- 945-3355 (989-3439)	1
US Order (332-1286)	1
On-Site Installation Declined (900-9997)	1
Declined Remote Consulting Service (973-2426)	1
Dell Storage MD1400 Shipping - DAO (340-AKOE)	1
No Raid, no Controller (405-AAFC)	1
Bezel MD1400, 2U, 12 Drives (325-BBHR)	1
Enclosure Management Module, dual (403-BBGM)	1
6TB 7.2K RPM NLSAS 12Gbps 512e 3.5in Hot-plug Hard Drive (400-AIUC)	12
Rack rail, 2Us, Static (770-BBJE)	1
MD1400/SC400 Enclosure Regulatory Label, Mexico (340-AKPJ)	1
Power supply, AC 600W, Redundant (450-AEBJ)	1
C13 to C14, PDU Style, 12 AMP, 6.5 Feet (2m) Power Cord, North America (492-BBDI)	1
C13 to C14, PDU Style, 12 AMP, 6.5 Feet (2m) Power Cord, North America (492-BBDI)	1
12Gb HD-Mini to HD-Mini SAS cable, 2m (470-ABDO)	2
12Gb HD-Mini to HD-Mini SAS cable, 0.5m (470-ABDN)	2

***Total Purchase Price:**	**$5,967.60**
Product Subtotal:	$5,947.61
Tax:	$0.00
Shipping & Handling:	$19.99
State Environmental Fee:	$0.00
Shipping Method:	LTL 5 DAY OR LESS
	(Amount denoted in $)*