

**State University System
Education and General
2015-2016 Legislative Operating Budget Issue
Form I**

University:	SSERCA Collaboration - UF, FSU, USF, UCF, FIU, FAU in SUS with UM (non-SUS)
Issue Title:	SSERCA Collaborative-Research Big-Data Infrastructure
Priority Number	
Recurring Funds Requested:	\$906,000
Non-Recurring Funds Requested:	\$6,000,000
Total Funds Requested:	\$6,906,000

I. **Description** *(Describe the service or program to be provided if this initiative is funded. Include whether this is a new or expanded service/program. If expanded, what has been accomplished with the current service/program?)*

a. Executive summary

This issue is a new service to build a statewide infrastructure to support collaborative research in the age of big data in which only the most competitive efforts can obtain funding for new projects or for continuing existing projects. This infrastructure will make researchers in the SUS more competitive in their research activities and in obtaining external funding for their research projects. This issue addresses all three of the goals articulated in the Scholarship, Research, Innovation section of the 2012-2025 Board of Governors Strategic Plan.

The service will be provided by the Sunshine State Education and Research Computing Alliance (SSERCA) through the member institutions (UF, FSU, USF, UCF, FIU) and consists of the following components:

1. The one-time acquisition of data storage systems for \$1,000,000 at each of the six institutions acting as part of the SSERCA organization (**non-recurring**), with \$60,000 per year annual maintenance contracts at each of the six institutions (**recurring**).
2. Provide one expert person at each of the six institutions to support the use of the service for \$91,000 per year (**recurring**).

There is an aligned budget request for consideration by the Legislature through the Department of Education from partner institution UM, also a member institution of SSERCA. That request is for UM to provide the seventh component of the state-wide integrated storage system for \$1,000,000 with \$60,000 per year annual maintenance contract and a full time expert support staff at \$91,000 per year.

b. Definitions

Research computing is defined as all computing, networking, and data management activities in support of research activities in all areas or science, engineering, and scholarship. Research computing includes, but is distinct from and much broader than, computing research, which is a part of computer science and engineering and focuses on doing research in computing activities.

Big data is a general term that describes the fact that modern research very often involves very large and complex sets of data that need to be searched and explored for knowledge. Often the research is done by collaborative groups with members spread over large geographic regions or all over the world. This leads to the challenge of sharing large data sets over large distances. These challenges are beyond the capability of research groups to handle and require properly architected and professionally managed infrastructure to overcome.

c. New Service resulting from prior BOG investment

This is a new service that is a much enhanced and expanded evolution of a successful \$450,000 project called “Sunshine Grid” funded by the BOG in 2010 within the New Florida Clustering Award Program. In that project FSU (award #15 \$150,000), UF (award #26 \$200,000) and USF (award #37 \$100,000) supported three selected collaborative research projects (in life sciences, weather and ocean modeling, and high energy physics) by providing storage infrastructure that could be shared by researchers located at different institutions to collaborate more effectively. The matching funds from the institutions created positions for permanent support personnel who are now an integral part of the support for education and research computing at their campuses and within SSERCA.

d. About SSERCA

SSERCA was created as a collaborative organization in 2010 by several SUS institutions (UF, FSU, USF, UCF) and one private institution (UM) who have a campus-wide effort to support research computing. The organization holds three to four summits per year and has developed a governance structure documented in the bylaws. These can be found at the website <http://www.sserca.org>. In 2012, FIU joined as a member (it also has a campus-wide effort to support research computing), and FAMU and UNF became affiliates. In 2014, FAU joined as a member and UWF became an affiliate.

SSERCA is the organization that provides advanced services in support of education and research computing on top of the statewide network infrastructure operated by the Florida LambdaRail (FLR, <http://www.flrnet.org>). The SUS institutions are founding members of FLR.

In the four years since SSERCA was founded, it has already provided a number of benefits to the faculty and researchers at the member and affiliate institutions. To name a few:

- Joint booth exhibiting the research of its faculty at the international Super Computing conference in 2011 in Seattle, in 2012 in Salt Lake City, and 2013 in Denver. The booth for 2014 in New Orleans is being planned now.
- SSERCA sponsored workshop on programming accelerators at FSU and XSEDE and SSERCA sponsored workshop at FIU
- Joint research by research computing staff of the member institutions on shared storage infrastructure, authentication methods, and sharing of expertise and experience on resource scheduling, system provisioning, and billing research grants for services and infrastructure costs.

The long-term vision for SSERCA includes supporting education and research computing at all institutions of higher learning, including state colleges, in the State of Florida. As an organization, SSERCA also plans to provide access to advanced resources and training for high-school projects. Another path for future expansion is to provide advanced high-performance computing (HPC) and big data resources and consulting to startup and mature companies with emerging needs in the HPC and big data. But these two long term goals are not achievable with currently available human resources.

e. Motivation and need for the infrastructure

Modern research is increasingly complex and increasingly associated with large amounts of data. These two requirements lead to two realities:

1. Researchers have to and do collaborate to be competitive.
2. Managing large amounts of data requires sophisticated, complex and expensive storage, computer, and network equipment.

The result is that research teams that consist of one or a few faculty members and their graduate students and research associates are spending an increasing fraction of their time managing the infrastructure for doing research, instead of doing the research. Institutions that provide state-of-the-art infrastructure with professional staff to design, build, maintain the infrastructure and provide advanced training and consulting for using it efficiently will provide their researchers with a significant competitive advantage. This advantage shows in two ways:

1. The time it takes to get results in the research is shortened because of increased researcher efficiency.
2. The likelihood to obtain external funding for the research is increased because reviewers and funding agencies recognize that projects with advanced infrastructure have a lower risk of failure and increased effectiveness of the researchers.

The infrastructure funded with this issue, to be described in detail next, will provide advanced support for collaborative research in all disciplines that involve generating, collecting, analyzing, and sharing large and complex data sets.

Furthermore, there is an increasing need for the ability to work with data that is restricted in some way, for example research on data with patient health information, work that involves export controlled data and software, and data and software with intellectual property restrictions. An improved and well-managed infrastructure to allow researchers to work with such data will increase the productivity of the research teams as well as reduce the risk for the institutions. The cost of loss or compromise of restricted data comes not only in the form of fines and the expenses of damage control, but also in long-term loss of revenue from the damage to researchers and institutional reputation, which reduces the likelihood of securing further grant and contract funding.

f. Infrastructure and service implementation details

Equipment The service consists of a coherent network of storage servers deployed at each of the six SSERCA member institutions UF, FSU, USF, UCF, FIU, and FAU in their data centers. A seventh identical unit will be deployed at UM with funding to be obtained separately outside of this issue. Each storage unit will cost \$1,000,000 and provide 3 PB (1 Petabyte = 1,000 Terabyte = 1,000,000 Gigabyte = 100,000 high-definition movies) of storage for a total of 21 PB across the State. The annual maintenance cost for the equipment is \$60,000 per year per site. This includes replacement of failed parts as well as expert support from the storage system vendor.

This storage has been architected to provide a number of important features that will enable easy-to-use, effective, and efficient sharing of data by researchers at any of the SUS institutions with their collaborators at other SUS institutions and elsewhere in the world.

1. One of the features is an easy interface to upload and download data that is familiar to researchers and anyone using the Internet with the same functionality as provided by cloud providers such as Dropbox, Google Drive, and others.
2. Another feature is automatic replication of data across multiple sites. This will provide extra data security in our State where a hurricane may jeopardize the data integrity at one site, but not at all six service sites simultaneously.

During 2014 SSERCA issues an ITN (Invitation to Negotiate) lead by FSU to engage the storage vendors to build a system with the above requirements. In November the vendor DataDirect Networks was selected as a partner. Three institutions (UF, FSU, and USF) have already made an investment of over \$500K to acquire and deploy a proof-of-concept system with total capacity around 1 PB. In partnership with the vendor DDN, this system is now being tested and validated.

Positions The second component of the service is equally important and consists of expert staff to be hired at each of the six sites, with the position at UM to be

funded separately. The team of 7 staff members will coordinate the operation and maintenance of the research data storage infrastructure within the existing organization of SSERCA. The expertise of the staff is such that a competitive salary for people in these positions is \$70,000 plus fringe benefits, or \$91,000 per site.

These individuals are also the primary people to provide the human interface to the faculty members, their students, and research associates for training and expert consulting services. Together with the other, existing staff members of the research computing support centers at each of the institutions, they will organize outreach and training sessions and workshops to make sure that the faculty at each institution is aware of the service and knows how to make the most effective use of it for their research.

Fund allocation for this issue should be directly to the participating SUS institutions: The equipment will be owned by each institution that operates and maintains it and the positions are part of the institutions as well. SSERCA is a lightweight organization that will coordinate the activities and will ensure that the new service functions in a coherent and efficient way. This is an ability it has already demonstrated in several projects. SSERCA does not own equipment or positions.

Leverage data centers By installing the storage systems in the data centers at SUS institutions that already operate and maintain research computing systems for complex scientific and engineering research, that existing infrastructure can be leveraged for the data processing, analysis, and visualization of the collaborative research data. Only the storage systems need to be acquired as part of this issue to provide a collaborative research infrastructure and service for big data research activities.

Leverage network The Florida LambdaRail (FLR) already connects the institutions with a high speed network at 20 Gigabits per second and this network will provide the underlying infrastructure to support the sharing of data. All institutions are members of FLR and have unlimited access to the FLR transport at no extra cost as part of existing and ongoing agreements. FLR has committed to upgrade its backbone network capacity to 100 GB, which will make the proposed storage system work even better.

Examples of the types of collaborations that will be enabled abound. SSERCA has already participated in a number of projects who needed to share data and has deployed specific solutions for a few such projects. The collaborations supported by the Sunshine Grid project listed above were the first projects.

A recent example on 2012 involved a collaboration between FSU, UF, and UM. The climate modelers at FSU and UM generated the climate evolution data over a period of several centuries. Then the researcher at UF needed to use the climate data as background for generating a crop model to determine which type of crop

would produce the most yield under the given climate conditions temperature, humidity, rainfall across the state of Florida.

A third example active in 2013 and 2014 involves gene sequencing data generated at FSU that needed to be analyzed by a team at UF.

A shared infrastructure like the one proposed in this issue will simplify the process of supporting these and similar researchers to the point where they can make a simple request that is provisioned in hours instead of days or weeks.

g. Future evolution and sustainability of the service

As will be discussed in section II below, the service proposed in this issue addresses a fundamental need expressed in the BOG Strategic Plan. As such we envision a multi-stage development to provide both accountability and mitigate risk to ensure a successful deployment.

The seven SSERCA members providing the service,-- UF, FSU, USF, UCF, FIU. FAU in the SUS plus UM with separate funding,-- have been chosen because they already provide research computing support for the faculty on their campus. They have the staff and expertise to deploy the proposed infrastructure and support its use by the collaborative research teams of which the faculty on their respective campuses are part.

Once the infrastructure has been deployed, SSERCA will extend its use to the affiliates. In addition, SSERCA will continue its efforts to engage SUS institutions, like FIT and FGCU, to join SSERCA. The infrastructure proposed in this issue will certainly provide an attractive value for these institutions to join so that their faculty members can benefit as well. The conditions to become an affiliate of SSERCA include appointing a contact person on the campus to act as the liaison between the researchers on each campus and the SSERCA service providers. This support person is essential in order to provide a good user experience for the faculty and their students at each campus when using SSERCA resources. A time table for the project is shown in the table below. Because the pilot project between UF, FSU, and USF, is already underway during 2014, the production system will be deployed as soon as funding becomes available.

Jul 2015	Oct 2015	Jan 2016	Apr 2016
Acquisition	Testing	Service for members	Service for members
Installation	Friendly user mode		
Hiring			

		Explore use by affiliates	Service for affiliates
--	--	---------------------------	------------------------

Expansion beyond SUS After extending the proposed research data storage service to all SUS institutions, it will be possible to consider extending the service further to state colleges. But that will require evaluation of resources, mainly personnel, to ensure that proper quality of service can be delivered.

The business model of the service will be to provide a basic level of storage with some limit to be determined for any project with as sponsor any faculty researcher at any SUS institution for free. However, to get storage limit larger, there will be a fee. The value of the fee will be determined to cover the cost of the hardware. The cost of the personnel will be treated as a subsidy from the State as an investment in advanced infrastructure for its researchers. Thus the storage system can grow as the demand grows with funding from the sources that fund the demand. This will also cover the replacement of the system after its expected useful life time of five years.

This mixed model of funding has been shown to work at other places and provides a sustainable balance between low cost for the researchers and sustainable investment by the State and the SUS to provide its faculty and researchers with advanced infrastructure.

II. Return on Investment *(Describe the outcome(s) anticipated, dashboard indicator(s) to be improved, or return on investment. Be specific. For example, if this issue focuses on enhancing marine bioscience research, indicate the anticipated outcome associated with the research, the accountability metric(s) that could be improved upon, and any expected return on investment to the State.)*

a. Alignment with BOG Strategic Plan

The “State University of Florida Board of Governors Strategic Plan 2012-2025” lists on pages 16 and 17 a number of goals that this issue is addressing in a direct way.

The opening paragraphs state that “...the Board of Governors will work to increase federal and private funding for collaborative research that targets STEM initiatives...” The proposed infrastructure will allow researchers at SUS systems to write competitive proposals to do exactly that.

The Plan continues with “...the Board of Governors will more sharply focus the research agenda of the State University System ... by strengthening research collaboration among the universities.” SSERCA is an organization that has the mission to foster and support collaboration on the computing infrastructure for

education and research and the infrastructure proposed in this issue is a concrete, enabling, and cost effective step towards that objective.

The first Goal in the Plan is on Excellence and seeks to “...strengthen the quality and reputation of scholarship, research, and innovation.” The proposed state-of-the-art infrastructure for collaborative research on Big Data Science and Engineering is an investment that will enable the highly talented researchers in the SUS to reach this goal.

The second Goal in the Plan is to increase productivity. By providing shared, professionally managed infrastructure, the researchers will be able to spend more time on their research and commercialization efforts than on dealing with infrastructure needed to do their work.

The third Goal in the Plan seeks to “increase collaboration and external support for research activity.” Providing the faculty, the students, and research associates in the SUS institutions with the proper infrastructure will allow them to make this goal a reality.

b. Metrics for the value to the State

The specific metrics to be used to establish the return on investment from building and maintaining the infrastructure proposed in this issue will be the ones developed recently by SSERCA to measure its own effectiveness.

1. Number of collaborations between faculty at different SUS institutions supported by data storage, high-performance computation, and high-speed data transmission infrastructure, training, and consulting.
2. Number of shared resources deployed under the umbrella of SSERCA.
3. Number of researchers, faculty and students reached by SSERCA sponsored training sessions and workshops, including online participation.
4. Number of grants funded at SUS institutions using the provided infrastructure and services.

That the investment by institutions in advanced infrastructure pays off to obtain very competitive grant funding has been shown many times over the past decades at numerous institutions across the Nation. An example was the \$10M award to the University of Florida by NNSA of the PSAAP II Center for Compressible Multiphase Turbulence <http://www.eng.ufl.edu/ccmt>, where the recent investments by UF in research computing infrastructure played an important role. Another example is the Southeast Center for Integrated Metabolomics <http://secim.ufl.edu> funded by a \$9.2M NIH award.

III. Facilities (*If this issue requires an expansion or construction of a facility and is on the Capital Improvement List complete the following table.*):

No new facilities need to be constructed as part of this issue.

	Facility Project Title	Fiscal Year	Amount Requested	Priority Number
1.				
2.				