

**State University System  
Education and General  
2016-2017 Legislative Budget Request  
Form I**

<b>University(s):</b>	<b>UF, FSU, USF, UCF, FIU, FAU</b>
<b>Issue Title:</b>	<b>Florida Collaborative Data Initiative</b>
<b>Priority Number</b>	
<b>Recurring Funds Requested:</b>	<b>\$2,161,800</b>
<b>Non-Recurring Funds Requested:</b>	<b>\$0</b>
<b>Total Funds Requested:</b>	<b>\$2,161,800</b>
<b>Please check the issue type below:</b>	
<b>Shared Services/System-Wide Issue</b>	<input checked="" type="checkbox"/>
<b>2015-2016 Non-Recurring Issue</b>	<input type="checkbox"/>
<b>New Issue for 2016-2017</b>	<input checked="" type="checkbox"/>

**I. Description**

**a. Executive summary**

Six Florida universities (UF, FSU, USF, UCF, FIU and FAU) seek to build a shared research data storage and sharing system that allows researchers to easily collaborate, store enormous quantities of data securely and be more competitive for federal grants. By working together as a system, tremendous cost savings and efficiencies can be realized. A single system can be built for considerably less money than six separate systems. The shared infrastructure provides researchers from around the state opportunities they would not have otherwise, simplifies computer systems support, and speeds up data transfer to days rather than weeks. It is a shared system, yet each university will build a component part on its own campus, hire its own support staff and pay for operations. This project scales and is easy to use and maintain. Additional universities can be added to the system in the future. UNF, UWF, and FAMU are already participating affiliate members. Universities across Florida have unique faculty, data and research expertise. A shared research data storage system encourages more cross-university collaboration, leveraging the best researchers in the state and making Florida researchers more competitive for federal grants. All Florida universities are elevated with the increased research capabilities and opportunities for collaboration this shared system will provide. Harvard College, Massachusetts Institute of Technology, and Boston College have a similar system to allow their researchers to collaborate.

Modern research is becoming increasingly complex and increasingly associated with vast amounts of data. The universities in the SUS have a great potential to collaborate on these issues, but lack the proper tools. This project is a new service to build a statewide data infrastructure to support collaborative research within and across the State of Florida. This infrastructure will make researchers in the SUS more competitive in their research activities and in obtaining external funding for their research projects in which only the most contending efforts can obtain funding. This is more important than ever as the amount of federal funds for grants continues to decline, thus driving competition up. The service proposed in this issue addresses all three of the goals articulated in the Scholarship, Research, Innovation section of the 2012-2025 Board of Governors Strategic Plan and will fill fundamental needs expressed in that plan, as will be discussed in section II below.

This is a new service that is a much enhanced and expanded evolution of a successful project called “Sunshine Grid” funded by the BOG in 2010 within the New Florida Clustering Award Program. In that project, FSU (award #15 for \$150,000), UF (award #26 for \$200,000) and USF (award #37 for \$100,000) supported three selected collaborative STEM research projects by providing storage infrastructure that could be shared by researchers located at different institutions to collaborate more effectively.

The service will be provided by the Sunshine State Education and Research Computing Alliance (SSERCA) through six of its member institutions (UF, FSU, USF, UCF, FIU, FAU), and will consist of recurring funds in the amount of \$360,300 at each of the six SUS institutions acting in collaboration as part of the SSERCA organization.

#### **b. Business model**

The business model will provide easy access to the service, but also become sustainable in the long term. It will provide a free, limited level of storage for any faculty researcher at any SUS institution, thereby immediately supporting the SUS as a whole. If needed, additional storage can be acquired at an affordable cost. This rate structure will be based on hardware cost-recovery and determined by the service providing institutions. The cost of the personnel will be treated as a subsidy from the State as an investment in advanced infrastructure for its researchers. Future growth will be funded by new external funding supported by the Florida Collaborative Data Initiative.

This mixed model of funding has been shown to work at other campuses 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.

### **c. Motivation**

Technological advances have enabled researchers to study more complex and challenging problems than ever before. However, this has led to a large increase in amounts of data that need to be processed, analyzed, and stored on computer systems that have become more complex and costly. Research grants often pay for these expenses, but the added expense associated with the collaborative nature is typically too cost prohibitive to be competitive for the grants. A State investment is necessary for the researchers to succeed.

The result is that research teams spend an increasing fraction of their budget and time managing the infrastructure for doing research, instead of conducting the research. Having institutions provide the required state-of-the-art infrastructure with support from professional staff to provide advanced training, consulting and security will provide researchers with significant competitive advantages:

1. A portion of storage costs are eliminated from grants and will be noted as State contribution.
2. The time it takes to get research results is shortened because of increased researcher efficiency.
3. The likelihood of obtaining external funding for 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.

This initiative will address these major issues in the following ways:

1. Promote research on combined data sets between SSERCA members and/or its affiliate members, and with other Florida educational institutions and researchers world-wide.
2. Provide a cost-effective disaster recovery solution.
3. Provide access to resources for workforce development.

### **d. Infrastructure and service implementation details**

**Positions:** A major component of the service consists of three expert staff to be hired at each of the six research universities. The team of 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 an average competitive salary for people in these positions is \$77,000 plus fringe benefits, or \$100,100 per person.

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 ensure that the faculty at each institution is aware of the service and knows how to make the most effective use of it in their research.

**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. Each initial storage unit will cost \$650,000 and provide 2 Petabytes (1 Petabyte is equivalent to 100,000 high-definition movies) of storage for a total of 12 PB across the State. It will be purchased using first year salary savings, because a phased approach for staff hiring will be implemented. When demand dictates, additional and replacement storage will be purchased using grants and institutional sources of funds.

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 cost-effective, easy-to-use, 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. Its attributes include the following:

1. An important feature is automatic replication of data across multiple sites. This facilitates easy sharing of research data by researchers and educators.
2. Data replication will also provide extra data security, for example during an extreme weather event, such as a hurricane, which can jeopardize the data integrity at one site, but not at all six service sites simultaneously.
3. An easy interface will be implemented to upload and download data with the same functionality as provided by cloud vendors such as Dropbox, Google Drive, and others. By leveraging economies of scale, SSERCA can offer this service at competitive rates.

During 2014 SSERCA issued an ITN (Invitation to Negotiate), led by FSU, to engage storage vendors to build a system with the above requirements. In November DataDirect Networks was selected as a partner. Four institutions (UF, FSU, USF, and FIU) 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 DDN, this system has been tested and validated. At the SuperComputing 2014 Conference, this system was demonstrated to a global audience. This initiative will build upon the experience gained from this proof-of-concept system.

**Fund allocation:** Funds for this initiative should be allocated equally to each of the six participating SUS institutions. The equipment will be acquired and maintained by each institution and the positions will become part of the institutions' IT staff as well. SSERCA 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.

**Leveraging 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.

**Leveraging the network:** The Florida LambdaRail (FLR) already connects the institutions with a high speed network at 100 Gigabits per second and this network will provide the underlying infrastructure to support real time collaboration on joint data sets. All six institutions are members of FLR and have unlimited access to the FLR transport at no extra cost as part of existing and ongoing agreements.

#### **e. Future evolution and sustainability of the service**

We envision a multi-stage development both to provide accountability and to mitigate risk in order to ensure a successful deployment. The six SSERCA members providing the service --UF, FSU, USF, UCF, FIU, and FAU-- have been selected because they already provide research computing support for the faculty on their campus. The new staff will complement existing 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 its affiliates at no cost. The infrastructure proposed in this issue will 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 among the researchers on each campus and the SSERCA service providers. These support personnel are essential in order to provide a positive 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 among UF, FSU, and USF, is already underway, the production system can be deployed as soon as funding becomes available.

Jul 2016	Oct 2016	Jan 2017	Apr 2017	July 2017
Acquisition Installation Hiring	Testing Pilot Members	Service for members	Service for members	Begin Recruitment of Permanent Staff
		Explore use by affiliates	Service for affiliates	

Initially, the proposed research data storage service will be set up for all SUS institutions exclusively. However, the long term goal is to extend the service to state colleges.

**f. About SSERCA**

SSERCA is the organization that provides advanced services in support of education and research computing in the State of Florida on top of the statewide network infrastructure operated by the Florida LambdaRail (FLR, <http://www.flrnet.org>) of which the SUS institutions are founding members. SSERCA was created as a collaborative organization in 2010 by several SUS institutions (UF, FSU, USF, UCF) and one private institution (the University of Miami [UM]), each of which has a campus-wide effort to support research computing. In recent years, FIU and FAU joined as a members and FAMU, UNF, and UWF became affiliate members.

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

- SSERCA sponsored workshop on programing 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. A recent example in 2012 involved a collaboration among 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 used the climate data as background for generating a crop model to determine which crops would produce the greatest yield under the given climate conditions temperature, humidity, rainfall across the state of Florida. In another joined project, gene sequencing data generated at FSU needed to be analyzed by a team at UF.

- Joint booth exhibiting the research of its faculty at the international Super Computing conference starting in 2011 at various locations. The booth for 2015 in Austin is being planned now.

The long-term vision for SSERCA includes supporting education and research computing at all Florida institutions of higher learning, including state colleges. As an organization, SSERCA also plans to provide access to advanced resources and training for high-school projects.

## **II. Return on Investment**

### **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 addresses.

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 aligns directly with the plan.

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.

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.

### **Metrics for the value to the State**

The 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 grants funded at SUS institutions using the provided infrastructure and services.
2. 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.
3. Number of shared resources deployed under the umbrella of SSERCA.
4. Number of researchers, faculty and students reached by SSERCA sponsored training sessions and workshops, including online participation.

The investment by institutions in advanced infrastructure has paid off to obtain very competitive grant funding 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

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