NSF Campus Cyberinfrastructure Program (CC) and OSPool/OSDF*

HTC'24

Kevin Thompson

Program Director, kthompso@nsf.gov

Office of Advanced Cyberinfrastructure

Computer & Information Science & Engineering

July 9, 2024

The Campus Cyberinfrastructure (CC*) Program

- Networking as a fundamental layer and underpinning of Cyberinfrastructure, driven by scientific R&E needs

- Most awards go to 10/100+ Gbps Campus networking upgrades, external connectivity to the national R&E fabric, and campus border redesign prioritizing science traffic.

- >500 awards totaling >$100M have been made across 50 states and jurisdictions in CC* since 2012

- CC* emphasizes strong campus level partnerships between researchers/teachers and campus IT leadership

Simple Science DMZ Diagram

A simple Science DMZ has several essential components. These include dedicated access to high-performance wide area networks and advanced services infrastructures, high-performance network equipment, and dedicated science resources such as Data Transfer Nodes. A notional diagram of a simple Science DMZ showing these components, along with data paths, is shown below:
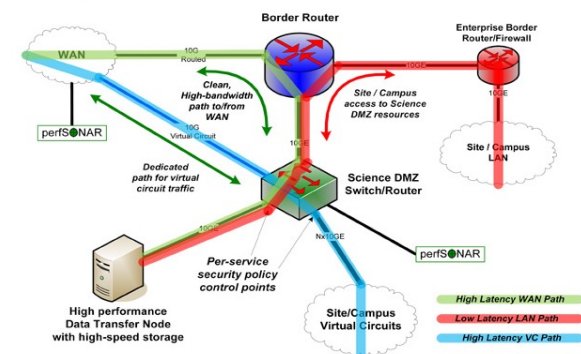
The essential components and a simple architecture for a Science DMZ are shown in the Figure above. The Data Transfer Node (DTN) is connected directly to a high-performance Science DMZ switch or router, which is connected directly to the border router. The DTN's job is to efficiently and effectively move science data to and from remote sites and facilities, and everything in the Science DMZ is aimed at this goal. The security policy enforcement for the DTN is done using access control lists on the Science DMZ switch or router, not on a separate firewall.

Campus Cyberinfrastructure (CC*)

Must be SCIENCE DRIVEN
Must have a campus Cyberinfrastructure Plan (except Strategy awards)
Seek to create partnerships – researchers, educators, IT organization

Area 1: Data Driven Networking Infrastructure	Area 2: Computing and the Computing Continuum	Area 3: Network Integration and Applied Innovation	Area 4: Data Storage and Digital Archives	Area 5: Strategy
Campus up to $700K **Region up to $1.4M**	**Campus up to $700K** **Region up to $1.4M**	**Small up to $500K** **Large up to $1M**	**Campus up to $700K** **Region up to $1.4M**	**Campus up to $100K** **Region up to $200K**
Technical solution; network management plan and diagram; leverage community	Multiple science drivers and needs; architecture; 20% is shared, typically through PATh	Networking R&D applied to the campus network with graduate student involvement	Multiple science drivers and needs; architecture; 20% is shared, typically through OSDF	A grant to help teams plan for a full proposal! No CI plan; Funds community building activities; No hardware

- **Proposals for the Campus** must address campus-wide needs, not a single research project.
- **Proposals for the Region** must address needs of multiple small or under resourced institutions. These can be led by a regional research lead organization.
- $15-$20M in FY24 in expected funding

See NSF 24-530 for details. Next deadline: October 15, 2024.
Program Officers: Amy Apon, awapon@nsf.gov and Kevin Thompson, kthompso@nsf.gov

CC* Area#2 - Computing and the Computing Continuum for the Campus or Region

- **Campus awards up to $700,000** total for up to 2 years
- **NSF encourages campus-level proposals in this program area from under-resourced institutions and strong preference will be given to proposals demonstrating a compelling need for access to campus and distributed computing resources.**
- Campus proposals may request funding for the acquisition of a shared, high-performance network-connected compute resource available to scientific computing users on campus and outside of campus.
- **The program area promotes a coordinated approach incentivizing multi-campus and national resource sharing as enabled by the OSG Consortium, an NSF-supported fabric of distributed scientific computing services that federate computing capacity across more than 150 institutions that delivered 2.6 billion CPU hours of scientific computing in calendar year 2023.**
- Current and previous awardees in Campus area within the last 5 years are not eligible to apply to this area.
- See solicitation for further guidance.

CC* Area#2 - Computing and the Computing Continuum for the Campus or Region

- **Region awards up to $1,400,000** total for up to 2 years
- Proposals for the Region in this area include all of the guidance and requirements of campus-level proposals and expand the impact and usage of the compute resource on multiple small and under-resourced campuses identified in the proposal. Proposals for the Region:
  - promote coordinated approaches in scientific computing resources serving scientific computing needs spanning a state's or region's small and under resourced institutions.
  - are led by regional and state research and education leadership organizations.
  - may request funding for the acquisition of a shared, high-performance network-connected compute resource available to scientific computing users across a defined set of campuses.
  - are required to address computing needs spanning multiple under-resourced institutions. A proposal focusing on a single campus but submitted as a proposal for the Region will be returned without review.
  - fund computing resources primarily used by the targeted small and under resourced institutions.

More on Area#2 – 20% Sharing

- All Area (2) proposals should commit to a minimum of 20% shared time and describe their approach to making the computing resource available as a shared resource external to the state/region and the institution(s) being primarily served. **Proposals are strongly encouraged to address this requirement by joining the Partnerships to Advance Throughput Computing (PATh)** campus federation (https://path-cc.io) and adopting an appropriate subset of PATh services to make the resource available to researchers on a national scale. **Proposals are encouraged to include a letter of collaboration from the selected platform and describe how they will track and report on meeting the 20% extramural usage goal each year.** Institutions in need of technical direction/expertise during their proposal development are encouraged to engage the NSF-funded PATh project at: https://path-cc.io.

Path-cc.io/services/research-computing

PATh's Support for Research Computing and the CC* Program

We are here to help with your CC* Proposal (NSF 24-530)!

Campuses with awards from the NSF Campus Cyberinfrastructure (CC*) Program play an important role in supporting Open Science. To date, 37 CC* campuses contribute to the processing and storage capacity of the Open Science Pool (OSPool) that is harnessed weekly by more than 3M jobs.

[Contact Us] [How We Can Help]

Enhancing the capacity of Research Computing of US campuses through local deployment and cross campus sharing is fully aligned with the vision of our NSF funded project - Partnership to Advance Throughput Computing (PATh). Our project is committed to support CC* projects from proposal, through deployment, to operation.

Proposal

Proposals in response to the 2024 CC* program solicitation (NSF 24-530) are due on 15 October 2024. Please contact us at cc-star-proposals@osg-htc.org (the earlier the better!) with any questions or requests you may have regarding the involvement of PATh in your proposed project. Our technology and services are readily available to support a spectrum of CC* projects.

The NSF 24-530 solicitation explicitly mentions the OSG services we provide as a means to meet requirements for the following areas:

(2) Computing and the Computing Continuum for the Campus or Region

NSF notes that "All Area (2) proposals should commit to a minimum of 20% shared time and describe their approach to making the computing resource available as a shared resource external to the state/region and the institution(s) being primarily served. Proposals are strongly encouraged to address this requirement by joining the Partnerships to Advance Throughput Computing (PATh) campus federation and adopting an appropriate subset of PATh services to make the resource available to researchers on a national scale. Proposals are encouraged to include a letter of collaboration from the selected platform and describe how they will track and report on meeting the 20% extramural usage goal each year. Institutions in need of technical direction/expertise during their proposal development are encouraged to engage the NSF-funded PATh project at: https://path-cc.io."

(4) Data Storage and Digital Archives for the Campus or Region

NSF states that "All Area (4) proposals are required to have interoperability with a national and federated data sharing fabric such as PATh/OSDF(see: http://www.osg-htc.org/about/osdf). At least

Contents

Sharing capacity with OSG services

The fabric of distributed computing services provided by the **OSG** enables campuses to share their computing capacity with the national S&E community via the Open Science Pool (**OSPool**)

- Any researcher affiliated with a US research institution can harness the **OSPool** capacity via Access Points provided by the **PATh** NSF Project

- Any US institution can contribute capacity to the **OSPool**

- **OSG** services enable science collaborations and campuses to build and operate private distributed computing environments across >130 sites that deliver over **2B** core hours annually

OSPool on July 5
(osg-htc.org/services/open_science_pool.html)

Who Contributes?
(https://osg-htc.org/services/open_science_pool/institutions)
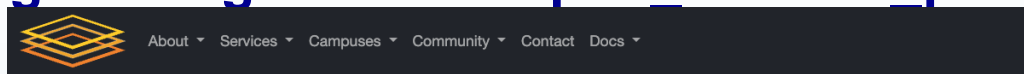
OSPool Contributing Institutions

The OSPool is powered by contributions from the Open Science community, specifically the institutions listed below. The scale of research being conducted has reached new heights through the capacity provided by these institutions and the Consortium's technology suite.

Institutions provide administrative control and oversight over the services they provide. At universities, the name of the registered institution is typically the name of the university rather than the name of the department that operates the service.

All statistics listed below are a summary of the last year of contributions.

Click on a row to view institution details.

Name	Jobs Ran	Impacted Fields of Science	Impacted Research Projects
University of California San Diego	17,839,448	43	122
University of Wisconsin	13,956,152	51	157
University of Chicago	11,131,565	49	152
Syracuse University	8,086,559	40	109
Great Plains Network	5,400,556	47	130
Indiana University	5,180,792	39	100
Fermi National Accelerator Laboratory	4,543,817	39	92
Purdue University	3,571,375	44	114
Clemson University	3,192,981	49	140
Montana State University	3,164,113	40	102
University of Connecticut	3,095,381	49	142
University of Michigan	2,513,006	38	103
Kansas State University	1,446,782	38	97
American Museum of Natural History	1,405,632	47	135
Lehigh University	1,331,038	34	82
University of Washington	1,279,622	26	61
University of Nebraska	1,248,051	43	113

Summary Statistics:	67	106,976,830	56	199

Institution drill down example

Franklin and Marshall College   ✕

Science Impact

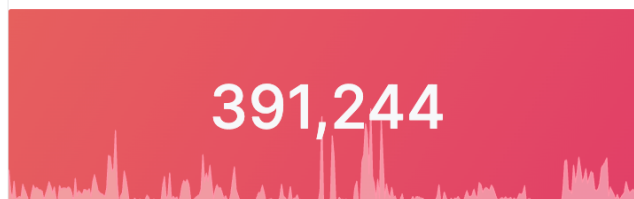**Projects supported by Franklin and Marshall College Resources** ⓘ

OSG Project Name	Provided Core Hours
WSU_3DHydro	97,997
NCSU_Hall	72,863
SBU_Jia	69,623
TelescopeArray	50,134
REDTOP	47,731
KOTO	21,454
CSUN_Katz	15,770
xenon	14,799
TG-CHE200122	14,368

Fields of Science Supported by Franklin and Marshall College Resources ⓘ

NSF Field of Science	Provided Core Hours
Nuclear Physics	98,501
Physics	85,918
Chemical Engineering	72,863
High Energy Physics	69,262
Astrophysics	66,593
Mathematics	16,500
Chemical Sciences	14,368
Biological Sciences	14,261
Chemistry	9,292

Organizations Supported by Franklin and Marshall College Resources ⓘ

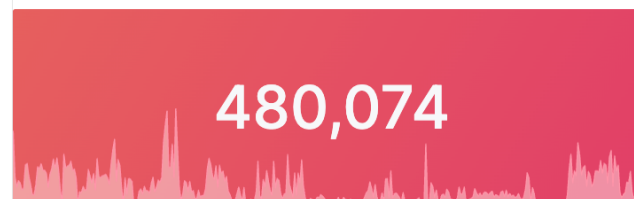| Organization	Provided Core Hours
Wayne State University	97,997
North Carolina State University	76,862
State University of New York at St...	71,238
University of Utah	50,134
Fermilab	47,731
Carnegie-Mellon University	23,379
The University of Chicago	21,454
University of Chicago	16,431
California State University, Northri...	15,770

Resources Provided

Total Jobs Ran on Franklin and Marshall College Resources ⓘ

391,244

CPU Hours Provided by Franklin and Marshall College ⓘ

480,074

OSPool Active Projects
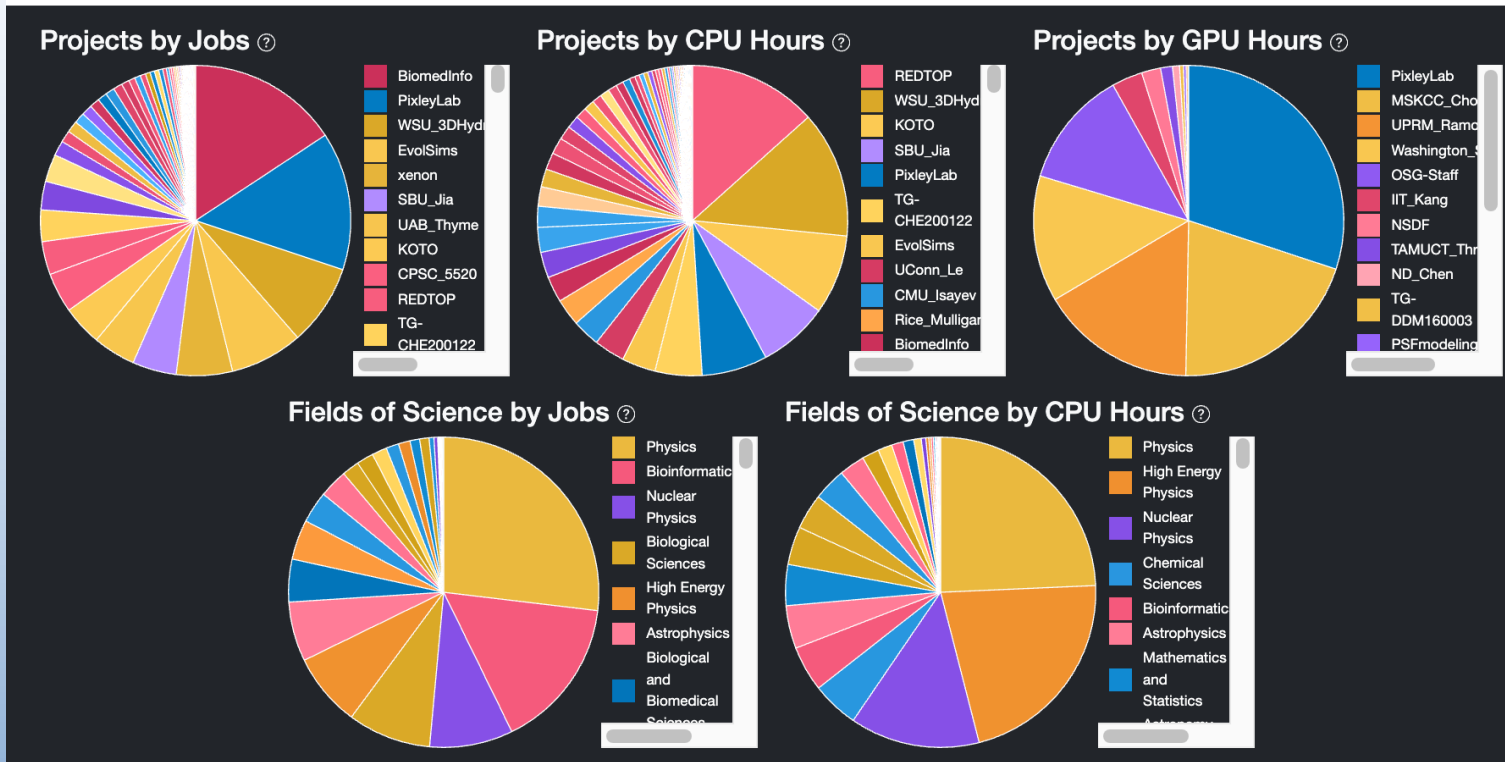
Most Recent Set of Awards in CC* Campus and Regional Computing from 23-526
(Total active awards: ~20 campus, ~10 regional)

| PI name	Institution	Title
Page, Richard	Miami U	CC* Campus Compute: Interdisciplinary GPU-Enabled Compute
Barraza-Lopez, Salvador	U of Arkansas	CC* Campus Compute: A High-Performance Computing System for Research and Education in Arkansas
Johnson, Ryan	U of South Dakota	CC* Campus Compute: Lawrence 2.0: Advancing Multi-Disciplinary Research and Education in South Dakota
Williams, Greg	U of Colorado Colo Springs	CC* Campus Compute: Building a Computational Cluster for Scientific Discovery
Moore, Shirley	U Texas El Paso	CC* Campus Compute: UTEP Cyberinfrastructure for Scientific and Machine Learning Applications
Brenn, James	Claflin U	Equipment: CC* Campus Compute: Advanced Computing for Claflin University Research and Teaching Ecosystem
Davray, Niranjan	Colgate	CC* Campus Compute: Colgate Compute Acceleration for Research and Education
Knutson, Ryan	Montana State University	Research Infrastructure: CC* Campus Compute: Next-Generation Shared HPC Cluster to Coordinate Campus Cyberinfrastructure and Accelerate Interdisciplinary Research
Petersen, Andrew	North Carolina State University	Equipment: CC Campus Compute: Expansion of GPU Compute Capacity for NC State University HPC to Support Research and Education

PI name	Institution	Title
Sheehan, Jerry	San Diego St U	CC* Regional Computing: The California State University System Technology Infrastructure for Data Exploration
Chadalapaka, Sarvani	UC Merced	CC* Regional Computing: CENVAL-ARC: Central Valley Accessible Research and Computational Hub
Pazner, Will	Portland State University	CC* Regional Computing: ORCA: Oregon Regional Computing Accelerator
Woodford, Jeffrey	Missouri Western State University	CC* Regional Computing: Development of Undergraduate High-Performance Computing Capacity with the CIMUSE Consortium
Scott, Grant J	University of Missouri-Columbia	CC* Regional Computing: Great Plains Extended Network of GPUs for Interactive Experimenters (GP-ENGINE)
Knuth, Shelley L	University of Colorado at Boulder	CC* Regional Computing: Enhancing Computing At Regional Schools in the Rocky Mountain Advanced Computing Consortium (RMACC)

CC* Campus Compute: UTEP Cyberinfrastructure for Scientific and Machine Learning Applications

Shirley Moore (PI), Deepak Tosh (Co-PI), Paras Mandal (Co-PI)
University of Texas at El Paso   NSF Award #: 234717
Start date: April 1, 2024   End date (estimated): March 31, 2026

UTEP HPC+NVIDIA Cluster Design
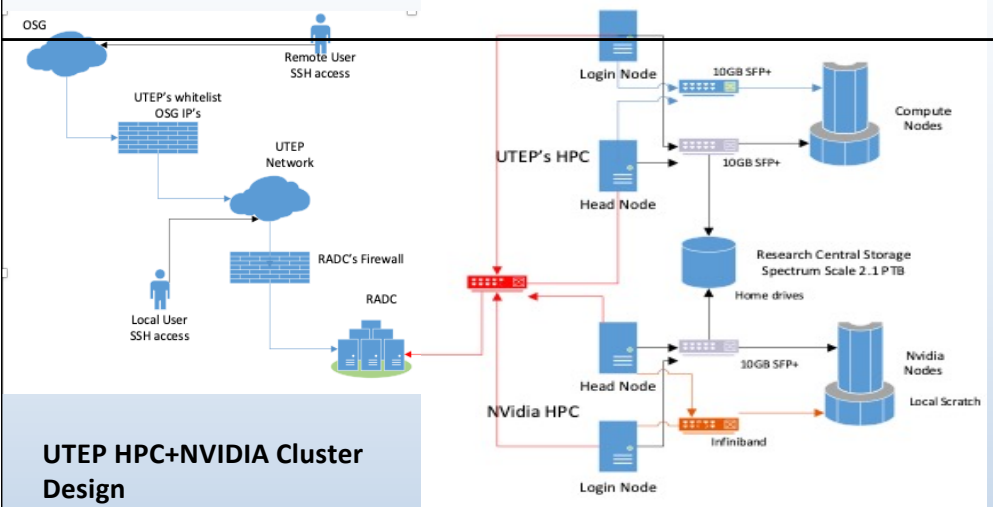
Technical Approach

- Integrate new NVIDIA H100 DGX and HGX systems with existing campus HPC cluster
- **Share GPU resources with the broader academic research community through participation in the Open Science Grid**
- Implement sharing using SLURM and NVIDIA multi-instance GPU (MIG)
- Collaborate with UTEP Information Resources and CSTECH to provide application onboarding, tutorials, account management, and user support and to assess progress towards project goals
- Advisory Committee consisting of representatives from science drivers, colleges of science and engineering, and students

Project Goals

- Provide researchers in computational science and engineering fields with access to state-of-the-art GPU-accelerated hardware and software resources for scientific simulation and machine learning
- Meet the requirements of science drivers that address important societal and national problems in the areas of renewable energy, advanced manufacturing, advanced materials, electric power systems, cybersecurity, and quantum computing
- Increase use of HPC and skills in using state-of-the-art HPC technologies across scientific disciplines
- Equip students with skills and experience needed for future careers in HPC and large-scale machine learning

Project Tasks
Year 1

- Order/install/test equipment
- Documentation and tutorials
- Launch project/onboard science drivers/education users
- Install/test perfSONAR servers
- Connect UTEP cluster with OSG
- Evaluate project launch and progress towards project goals
- Analyze network monitoring/perfSONAR data

Project Tasks
Year 2

- Update software stack and network configuration based on data analysis
- Revise documentation and tutorials
- Participate in OSG All-hands meeting
- Conduct quarterly surveys of NVIDIA cluster users
- Collect data/write final report

CC* Area#4 – Data Storage and Digital Archives for the Campus or Region

- **Campus Award size up to $700,000** total for up to 2 years.

- This program area promotes coordinated approaches in **scientific storage, data management, and digital archives and incentivizes multi-campus and national resource sharing.** Awards in this area reflect NSF principles and guidance in the community's stewardship of data from NSF funded research, and particularly aim at supporting the data lifecycle. As described below for the Data Management Plan, **proposals should adhere to the findable, accessible, interoperable, and reusable (FAIR) principles** and guidance from NSF described in the Dear Colleague Letter (DCL) on Effective Practices for Data: https://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf19069.

- Storage proposals for the Campus **address campus-wide storage needs** in the proposal. A proposal focusing on a single science domain or project use will not be considered for funding.

- **New this year**: Storage proposals for the **Region (up to $1.4M)** adopt all the requirements and guidance of a campus storage proposal and target the research and education data needs spanning multiple small and under-resourced institutions in a region.

Sharing Storage Requirement in Area#4

- **All Area (4) proposals are required to have interoperability with a national and federated data sharing fabric** such as PATh/OSDF (see: http://www.opensciencegrid.org/about/osdf). At least 20% of the disk/storage space on the proposed storage system should be made available as part of the chosen federated data sharing fabric. Proposals should describe the interaction with data sharing fabric beyond simply naming it.

Recent CC* Data Storage Awards from 23-526 (~18 total active campus storage awards)

PI	Institution	Title
Reimer, Torsten	University of Chicago	CC* Data Storage: A Secure Hub for Access, Reliability, and Exchange of Data (SHARED)
Maloney, John	U of Ill Urbana-Champaign	CC* Data Storage: High-Capacity Active Archive to Enable Economical Data Access and Distribution for Illinois Researchers and the Na
Doyle, Damian	U of MD Baltimore County	CC* Data Storage: Broadening UMBCs Data Storage footprint to Advance Scientific Research and Discovery
Hanna, Chad	PA St U University Park	CC* Data Storage: Cost-effective Attached Storage for High throughput computing using Homo- geneous IT (CASH HIT) supporting P
Kennedy, Michael	U of Cal Riverside	CC* Data Storage: Foundational Campus Research Storage for Digital Transformation
Bobadilla, Leonardo	Florida International University	CC* Storage: EnviStor: A Repository for Supporting Collaborative Interdisciplinary Research on South Florida's Built and Natural Environments
Perez, Lisa M	Texas A&M University	CC* Data Storage: FASTER Data Infrastructure to Accelerate Computing
Huffman, John N	University of Delaware	CC* Data Storage: Shareable, Equitable, and Extensible Data Storage for Collaborative Data-intensive Research
Calyam, Prasad	University of Missouri-Columbia	CC* Data Storage: Remote Instrumentation Science Environment for Intelligent Image Analytics
Meneveau, Charles V	Johns Hopkins University	Research Infrastructure: CC* Data Storage: 20 Petabyte Campus Research Storage Facility at Johns Hopkins University
Yan, Le	Louisiana State University	CC* Data Storage: Tiger Den: sustainable campus-wide data system to support CI-enabled research and education
Altiparmak, Nihat	Foundation Inc	CC* Data Storage: Cardinal Academic Research Data Storage (CARDS)
Engquist, Erik	William Marsh Rice University	Research Infrastructure: CC* Data Storage: Rice Collaborative Object Store
Morrell, Thomas E	California Institute of Technology	CC* Data Storage: Closing Caltech's data storage gap: from ad-hoc to well-managed stewardship of large-scale datasets
Rossmiller, Zachary	University of Montana	Equipment: CC* Data Storage: Improving Research Ability with Data Storage at the University of Montana

CC* Storage: EnviStor: A Repository for Supporting Collaborative Interdisciplinary Research on South Florida's Built and Natural Environments (award # 2322308)

Leonardo Bobadilla, Mike Kirgan, Julio Ibarra, Jason Liu, Jayantha Obeysekera

Goals

The goal is to build "EnviStor: A Repository for Supporting Collaborative Interdisciplinary Research on South Florida's Built and Natural Environments." which will provide a robust database management system (DBMS) with efficient integration essential for interdisciplinary modeling.

The system will receive data streams from various research activities including data from sensors, satellites, simulations, and high-resolution scientific instruments

Intellectual Merit

EnviStor will be transformative in facilitating multidisciplinary research by significantly reducing constraints when generating copious amounts from research built and natural environments. It will facilitate inter and intra-campus capabilities augmenting the current FIU storage capabilities from the Terabyte to the Petabyte scale.
 The design and implementation will follow the FAIR principles to maximize reusability and interoperability. Our system will provide an unprecedented benefit to the research community in the region as a clearing house for environmental and engineering data products, with interoperability features that will help fundamental research and multidisciplinary research.

Broader Impacts

High-impact applications: Scientific advances in this project will enable cost-effective data storage and sharing for many important applications, including the assessment of reef ecosystem health, rising sea levels and flooding monitoring and modeling, harmful algae blooms, water contamination, and wildlife habit loss.

Research Involvement from Underrepresented Groups: EnviStor will provide opportunity for students at FIU (75% of them coming from underrepresented groups in STEM) to enhance their knowledge of database management and focus on interoperability

Encouraging Inter and Trans-disciplinary research: Envistor will encourage researchers to use a centralized and secure location to store and share their data across. This will promote interoperability, convergence and accelerate innovation disciplines.

Approach
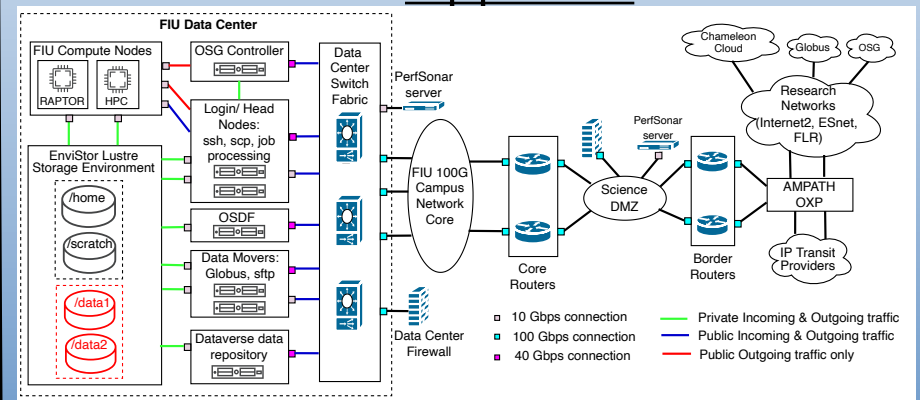
EnviStor architecture will add a new high-capacity Petabyte-scale Lustre storage system to FIU's storage environment. The Lustre file system provides low-latency, high-performance, and cost-effective architecture for computational resources that can scale to PBs of data.

CC* Futures

- NSF CISE/OAC is likely to issue another CC* solicitation in FY25
    - TBD on 1 or 2 deadlines
    - Computing and Storage areas are popular…
    - CC* remains committed to campus level resource sharing
- NSF always welcomes community feedback on its programs and solicitations

Thanks!