NSF and Campus Cyberinfrastructure

HTC'23

Madison, Wisconsin

July , 2023

Kevin Thompson
National Science Foundation
Office of Advanced Cyberinfrastructure

Programmatic directorates and offices supporting the NSF Mission: promote the progress of science; advance the national health, prosperity, and welfare; and to secure the national defense

NSF's CISE/OAC and Scientific Cyberinfrastructure

- **Office of Advanced Cyberinfrastructure**: Supports and coordinates the development, acquisition and provisioning of state-of-the-art cyberinfrastructure resources, tools and services essential to the advancement and transformation of science and engineering.

- **Cyberinfrastructure (CI)**: Compute, data, software, networking and people to facilitate scientific discovery and innovation.

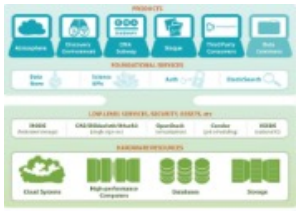
NSF Office of Advanced Cyberinfrastructure (OAC)

Foster a cyberinfrastructure (CI) ecosystem to transform science and engineering research… through Research CI and CI research

Observation

People, organizations, and communities

Coordination & User support



Gateways, Hubs, and Services

Data Infrastructure

Software and Workflow Systems

CI-Enabled Instrumentation

Pilots, Testbeds

Discovery

Computing Resources

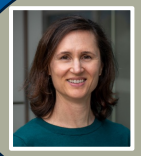
R&E Networks, Security Layers

Cloud Resources & Services

NSF Office of Advanced Cyberinfrastructure (OAC)

July 2023

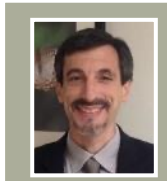
Directorate for Computer and Information Science and Engineering



Katie Antypas *
Office Director

Amy Walton
Deputy Office Director

Bob Chadduck



Bill Miller

Sheikh Ghafoor *

Andrey Kanaev

Sharon Geva

Marlon Pierce *

Ed Walker

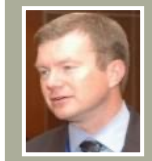
Rob Beverly

Amy Apon *

Alejandro Suarez

Seung-Jong (Jay) Park *

Juan (Jen) Li *



Kevin Thompson

Varun Chandola *

Ashok Srinivasan *

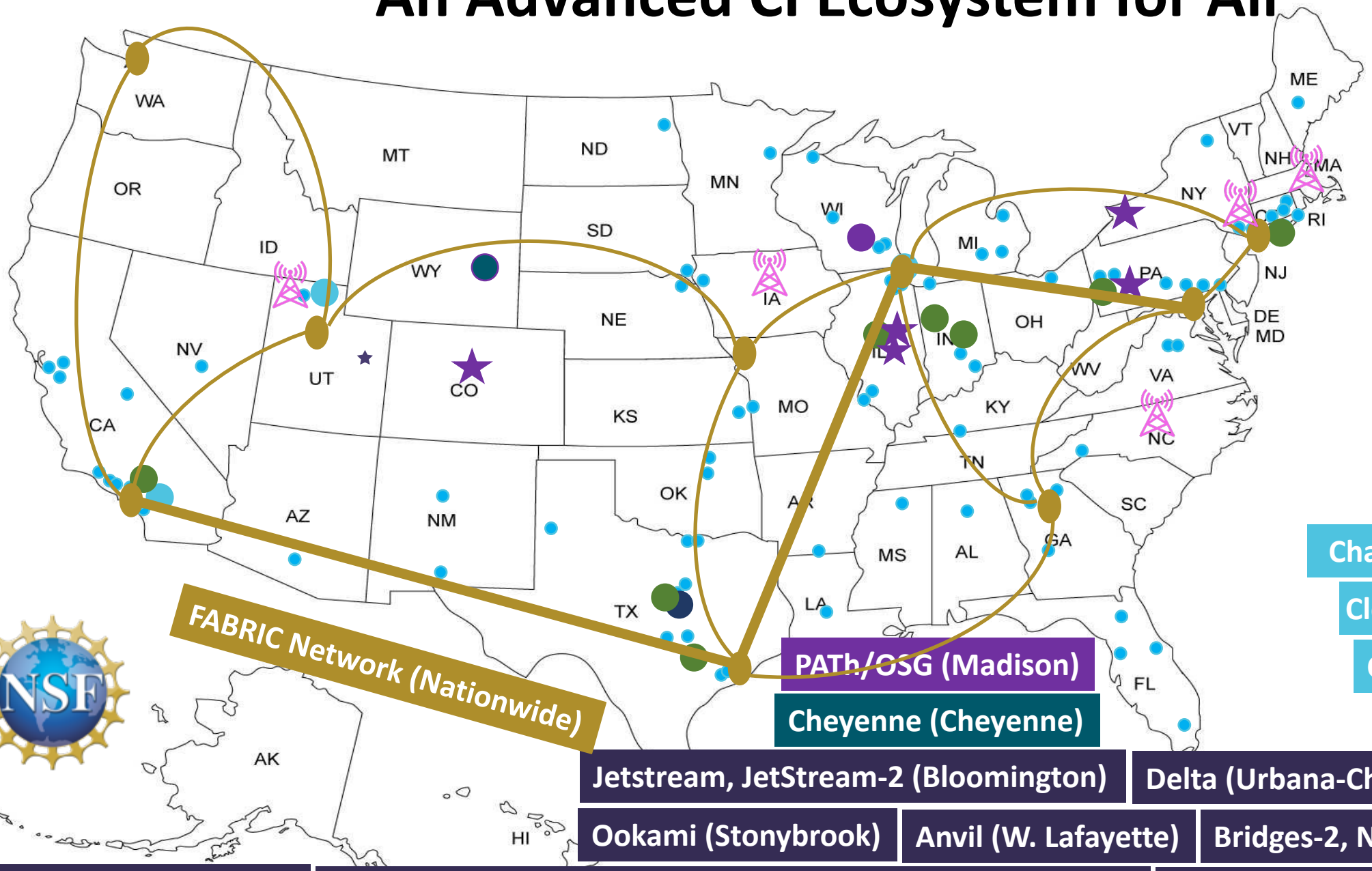
Tom Gulbransen *

Sharmistha Bagchi-Sen

Plato Smith *

* IPA Appointment

An Advanced CI Ecosystem for All



ACCESS PIs

Leadership-class

Innovative systems

HTC Services

NCAR

Cloud Resources

Shared Campus Resources

PAWR Testbeds

FABRIC Network (Nationwide)

PATh/OSG (Madison)

Cheyenne (Cheyenne)

Jetstream, JetStream-2 (Bloomington)

Delta (Urbana-Champaign)

Ookami (Stonybrook)

Anvil (W. Lafayette)

Bridges-2, Neocortex (Pitt)

ACES (College Station)

Expanse, Voyager, National Research Platform (San Diego)

Stampede 2, Wrangler (Austin)

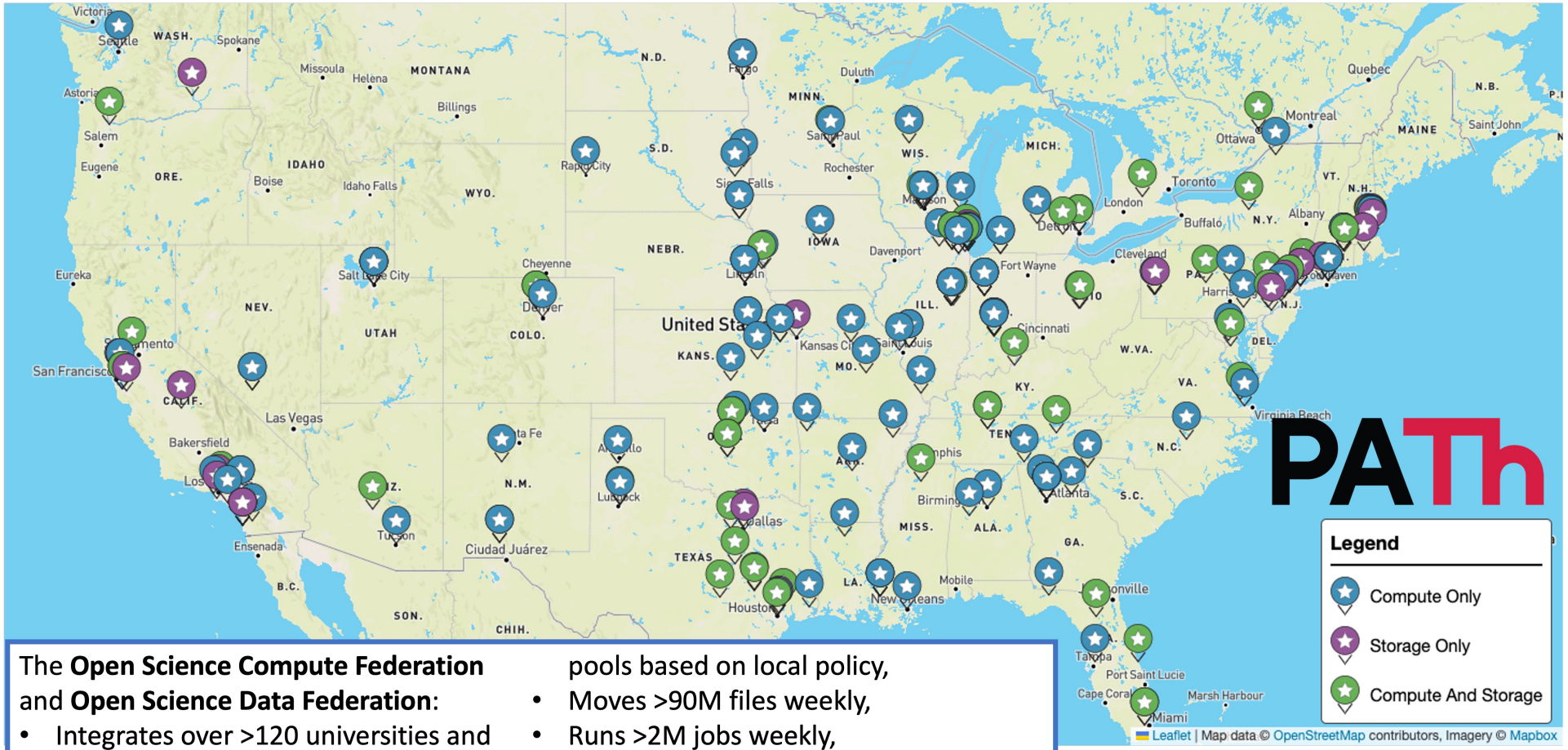
Chameleon Lab (Chicago)

CloudLab (Salt Lake City)

CloudBank (San Diego)

Frontera (Austin)

Federating Data and Compute (PATh/OSG, OSDF)

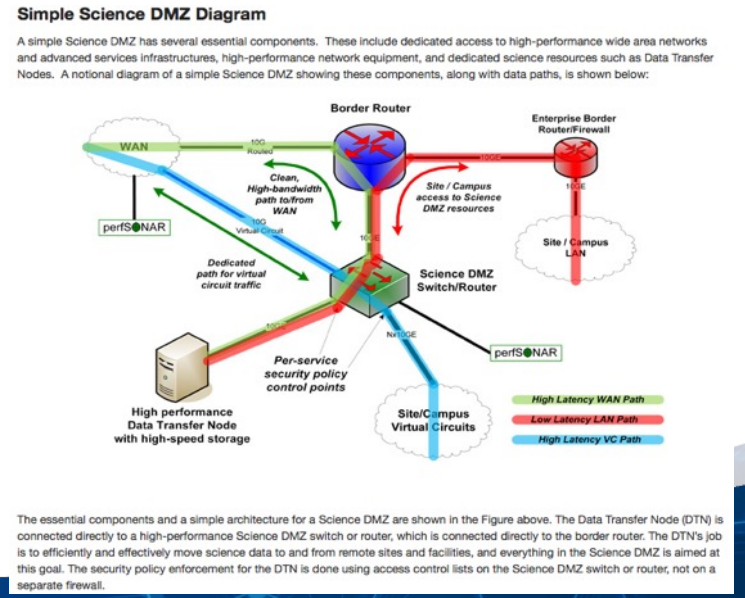


The **Open Science Compute Federation** and **Open Science Data Federation**:

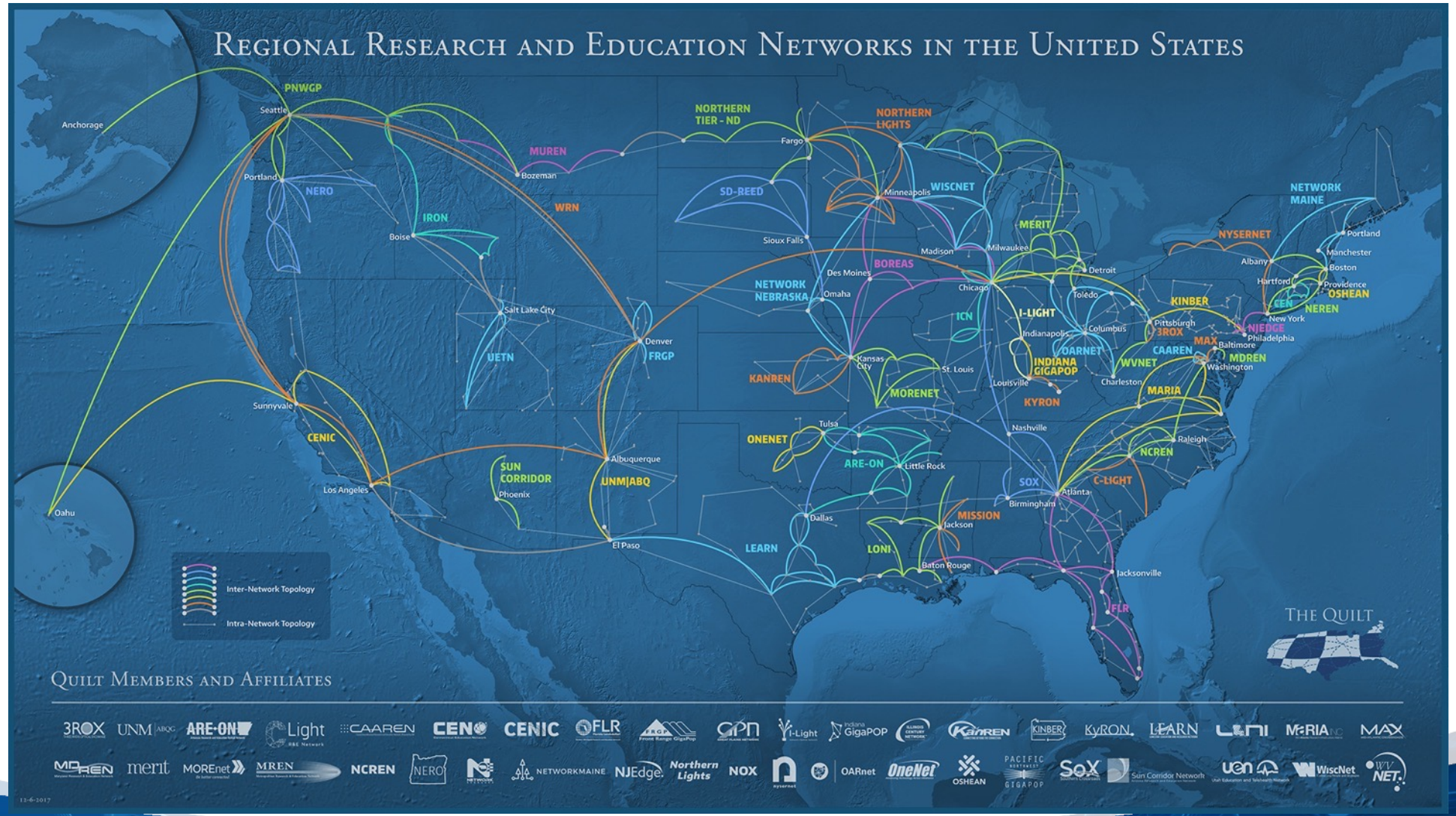
- Integrates over >120 universities and laboratories across the planet, about half outside the US.
- Organizes the capacity to resource pools based on local policy,
- Moves >90M files weekly,
- Runs >2M jobs weekly,
- Manages jobs from more than a dozen Access Points.

The Campus Cyberinfrastructure (CC*) Program

- Networking as a fundamental layer and underpinning of Cyberinfrastructure, driven by scientific R&E needs
- Most awards go to High Speed Campus networking upgrades, external connectivity to the national R&E fabric, and campus border redesign prioritizing science traffic.
- Program is in its 12th year with different program areas expanding into other components of CI including computing and data
- CC* emphasizes strong campus level partnerships between researchers/teachers and campus IT leadership



State and Regional R&E Fabric



CC* 23-526 - Campus Cyberinfrastructure

- https://www.nsf.gov/pubs/2023/nsf23526/nsf23526.htm
- $15M-$20M in expected award funding
- 30-50+ expected awards
- **Proposals due ~~March 1 and~~ September 11, 2023**
- Area #1 – **Campus Network upgrades**
- Area #2 – **Regional Connectivity for Small Institutions**
- Area #3 – **Networking Integration and Applied Innovation**
- *Area #4 – **Campus Computing**
- *Area #5 – **Regional Computing**
- *Area #6 – **Data Storage**
- Area #7 – **Planning Grants and CI-Research Alignment**

All 3 mention PATh/OSG/OSDF

Area 4: Campus Computing and Computing Continuum

This program area promotes coordinated approaches in scientific computing at the campus level and invests in the seeding of new and shared computing resources through investments in capacity computing in campus clusters. The program area promotes a coordinated approach.

It is expected that the campus-wide computing needs are addressed in the proposal.

All proposals submitted to this area must address:

- Scientific and engineering projects and their research and education computing needs, describing project-specific scenarios for scientific computing tied to the proposed computing resources;
- Features, capabilities, and software platforms representing the proposed computing resources; and
- Scientific computing codes expected to run on the resources.

NSF encourages 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/cloud resources

Excerpt from CC* Area#4 on Campus Computing

The Cluster as a Shared Resource Intra-campus and Inter-campus: Proposals should describe (1) their approach to sharing the proposed computing resource across the science drivers and researchers at their institution; (2) how the resource will be accessed by external research groups; and (3) how the resource is coordinated with external resources allowing the institution's researchers to seamlessly access computing resources at other campuses, regional and national computing resources, and/or production cloud resources, if appropriate.

Proposals are required to commit to a minimum of 20% shared time on the cluster and describe their approach to making the cluster available as a shared resource external to the campus, with access and authorization according to local administrative policy. Conversely, the proposal should describe the approach to providing on-demand access to additional external computing resources for its targeted on-campus users and projects. **One possible approach to implementing such a federated distributed computing solution is joining a multi-campus or national federated system such as the Open Science Grid**. Whatever opportunistic, federated, scalable, distributed computing platform is chosen, the proposal is expected to justify the choice by including a discussion on the shared platform's track record in the community, its current scientific computing production capability, and its scaling properties. 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.

Area 5: Regional Computing

This program area promotes coordinated approaches in scientific computing at the regional level through investments in computing clusters serving scientific computing needs spanning a state's or region's small and under resourced institutions.

This area solicits proposals led by regional and state research and education leadership entities and universities. This area represents an opportunity as well for potential collaborations between EPSCoR and non-EPSCoR jurisdictions.

It is expected that multi campus-wide computing needs are addressed in the proposal. It is expected that the resources funded by this track will be primarily used by the targeted small and under resourced institutions.

All proposals submitted to this area must address:

- Scientific and engineering projects and their research and education computing needs, describing project-specific scenarios for scientific computing tied to the proposed computing resources;
- Features, capabilities, and software platforms representing the proposed computing resources;
- Scientific computing codes expected to run on the resources; and
- Approaches/mechanisms that will be used to ensure fair and equitable use of the resources by the targeted **small and under resourced institutions**.

CC* Area#6 Data Storage

- Proposals must include in the Project Description:
  - A summary table of the science drivers and their data storage environments - these requirements should be specified in clear terms reflecting a specific understanding of the required storage resources and environment, for example, storage type, data movement characteristics and data curation approach as part of a scientific workflow profile;
  - The platform architecture and open-source software/platform, with under-resourced institutions allowed to include commercial software license fees for only the storage management platform;
  - An open source-based approach to storage system monitoring, measurement, management, and instrumentation;
  - A sustainability plan addressing the institution's commitment to providing an ongoing level of sustained access to storage resources;
  - High-Performance Network Connectivity and Specification - see below for more details; and
  - A description of the storage system as a Shared Resource Intra-campus and Inter-campus - see below for more details

CC* Area#6 Data Storage

- **The Storage system as a Shared Resource Intra-campus and Inter-campus**: Interoperability is required with a national and federated data sharing fabric such as PATh/OSDF (see: <www.opensciencegrid.org/about/osdf>). At least 20% of the disk/storage space on the proposed storage system must 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.
- Staffing required to configure, operate, and support data management and access of the storage resource is an acceptable component of the budget at up to 25% of the overall budget request. Staff associated with training and facilitating adoption, integration, and use of the resource on campus may be included. Staff and user training costs may extend to the open source software platforms and federated data fabric sharing platform.
- Tangible metrics addressing measures of success should be included.
- Proposals should discuss the storage system in the larger context of campus CI and a vision inclusive of supporting shared resources and computing at the campus level.
- Proposals should also address the long-term plan for archival storage and sustainability. Proposals may address these topics in the Project Description, Data Management Plan, or the Campus CI Plan.

2019 - 1st year of Campus Compute Awards (12)

Award#	PI	Institution	Title
1925558	Hawkins, Ronald B	University of California-San Diego	CC* Compute: Triton Stratus
1925590	Montes, Juan	American Museum Natural History	CC* Compute: High Performance Campus Computing for Institutional Research at the American Museum of Natural History
1925541	Belgin, Mehmet	Georgia Tech Research Corporation	CC* Compute: Integrating Georgia Tech into the Open Science Grid for Multi-Messenger Astrophysics
1925596	Fiske, Joshua A	Clarkson University	CC* Compute: Accelerating Computational Research for Engineering and Science (ACRES) at Clarkson University, A Campus Cluster Proposal
1925645	Lannon, Kevin	University of Notre Dame	CC* Compute: CAML - Accelerating Machine Learning via Campus and Grid
1925467	Cinabro, David A	Wayne State University	CC*: Campus Computing and the Computing Continuum: Campus Cluster Resource: Expanded High Performance Computing at Wayne State
1925603	Skjellum, Anthony	University of Tennessee Chattanooga	CC* Compute: A Cost-Effective, 2,048 Core InfiniBand Cluster at UTC for Campus Research and Education
1925192	Rampp, Carrie	Franklin and Marshall College	CC* Compute: Building a state-of-the-art campus compute resource at Franklin & Marshall College
1925267	Rossmiller, Zachary	University of Montana	CC* Compute: Improved Computing for Advanced Research and Education (ICARE)
1925716	Jones, Richard T	University of Connecticut	CC* Compute: Shared Computing Infrastructure for Large-scale Science Problems
1925766	Hauser, Thomas	UC Boulder	CC* Compute: A Hybrid Cloud Environment for the Rocky Mountain Advanced Computing Consortium
1925717	Delaney, Kris	UC Santa Barbara	CC* Compute: A high-performance GPU cluster for accelerated research

2019 award example

NSF Campus Cyberinfrastructure PI and
Cybersecurity Innovation for Cyberinfrastructure PI Workshop

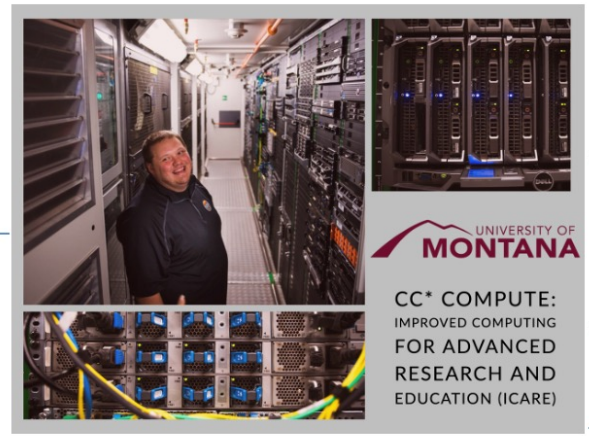
Quad Chart for: *CC* Compute: Improved Computing for Advanced Research and Education (ICARE)*

Challenges:

- Supporting the academic and research community's increasing computing demands during a time of declining budgets.
- Bringing together faculty researchers and IT professionals to identify needs and develop solutions on a mixed centralized/decentralized IT campus.

Solutions:

- Build an ICARE Team made up of Central & Distributed IT, Faculty, and Researchers (collaboration ensures everyone is invested in the campus cyberinfrastructure and IT focuses effort & investment where it is needed).
- Develop a University of Montana Shared Computing Cluster (UMSCC) devoted to advancing research and education.
- Partner with the Open Science Grid (OSG) for bidirectional resource sharing beyond UM.



Scientific Impact:

- immediate progress on "science drivers" (Biomedical & Pharmaceutical Sciences, Computer Science, Biological Sciences, Public and Community Health Sciences, & Geosciences). Including 19 publications.
- **End of grant period: 124 researchers accessing the cluster running 690,484,179 compute minutes (approximately 500,000 compute days of activity).**
- Participation in OSG impacts access to additional resources and builds a broader research community.

Metadata tags:

- *Shared Compute Cluster*
- *Graphics Processing Unit*
- *Open Science Grid*
- *Open Source Solutions*
- *Need Collaborators*
- *More Funding Opportunities*

2020 – Year2 of Campus Compute area (19)

Award#	PI	Institution	Title
2018846	Carver	U of Alabama Tuscaloosa	CC* Compute: Accelerating Advances in Science and Engineering at The University of Alabama Through HPC Infrastructure
2018149	Sedore	Tufts University	CC* Compute: GPU Infrastructure to Explore New Algorithmic & AI Methods in Data-Driven Science and Engineering at Tufts University
2018766	Andresen	Kansas State University	CC* Compute: GP-ARGO: The Great Plains Augmented Regional Gateway to the Open Science Grid
2020446	Yan	LSU	CC* Compute: Deep Bayou: Accelerating Scientific Discoveries with A GPU Cluster
2018758	Larkins	Rhodes College	CC* Compute: A high-performance computing cluster to accelerate research, education, and training at Rhodes College
2018551	Chace	Southern Ill U Edwardsvill	CC* Compute: SIUE Campus Cluster
2018851	Segee	University of Maine	CC* Compute: High-Memory Compute Resources for Maine
2019089	Mandel	U of Colorado Denver	CC* Compute: Accelerating Science and Education by Campus and Grid Computing
2018926	Smith	Purdue University	CC* Compute: Private Campus Cloud for Data Analytics and Machine Learning
2019007	Jelinkova	William Marsh Rice Univ	CC* Compute: Interactive Data Analysis Platform
2019000	Dugas	New Mexico St University	CC*Compute: From classroom to the lab: NMSU responds to the changing HPC landscape in New Mexico
2019216	Liu	Portland State University	CC* Compute: GPU-based Computation and Data Enabled Research and Education (G-CoDERE) at PSU
2019035	Webb III	Lehigh University	CC* Compute: Acquisition of a Lehigh University HPC cluster to enhance collaboration, research productivity and educational impact
2018936	Taylor	LA St Univ Health Center	CC* Compute: Compute Cluster for Computational Sciences at Louisiana State University Health Sciences Center – New Orleans (LSUHSC-NO)
2018933	Wemhoff	Villanova University	CC* Compute: High-Performance Computing Backbone for Accelerating Campus-Wide and Regional Research
2019220	Kesselman	USC	CC* Compute: A Customizable, Reproducible, and Secure Cloud Infrastructure as a Service for Scientific Research in Southern California
2018822	Scozzafava	Syracuse University	CC* Compute: A High Performance GPU Cluster at Syracuse University
2018841	Pal	West Texas A&M University	CC* Compute: GROWTH - Gateway for Increased Research Output at a West Texas Higher-education Campus
2019194	Castillo	San Diego State Univ Fdn	CC* Compute: Central Computing with Advanced Implementation at San Diego State University

2020 Award example

NSF VIRTUAL CAMPUS CYBERINFRASTRUCTURE PI WORKSHOP

SEPTEMBER 19 – 21, 2022

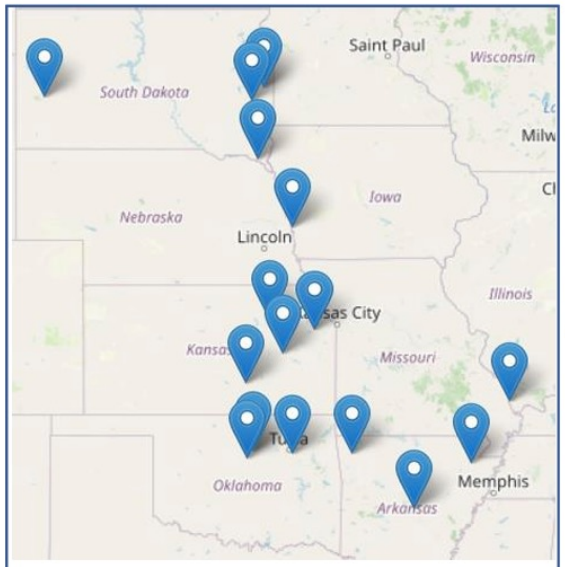
Quad Chart for: ***CC* Compute: GP-ARGO: The Great Plains Augmented Regional Gateway to the Open Science Grid***

Challenge Project Seeks to Address:

- General lack of HTC resources at regional scale
- Lack of local HTC resources – leads to lack of interest and buy-in by students and staff
- Lack of qualified researcher-facing staff in region

Solution(s) :

- Train and develop local/regional researcher-facing staff
- Deployed 18 HTC GPU-equipped compute nodes across region
- Developing gateways to local HPC resources
- Augment CyberTeam CC* to train users, engage administration



GP-ARGO node locations

Broader Impact:

- Increased regional engagement by students & staff in HTC computing
- Increased resources for bioinformatics and pandemic science
- Model for other distributed regional HTC onboarding efforts

Metadata tag:

- ***Overview:*** *https://gp-argo.greatplains.net/*
- ***Metrics:***
- *https://ksu.gp-argo.greatplains.net/grafana/d/hYceI5X7k/gp-argo-home?orgId=1&var-Node=All*
- *Collaborators welcome!*
- *Student engagement ideas needed*

2021/early FY2022 – 2 deadlines…17 awards

Award#	PI	Institution	Title
2127188	Rogers, Michael D	Tennessee Technological University	CC* Compute: A HPC Cluster for Science Research and Education at Tennessee Tech University
2126303	Jennewein, Douglas M	Arizona State University	CC* Compute: The Arizona Federated Open Research Computing Enclave (AFORCE), an Advanced Computing Platform for Science, Engineering, and Health
2125646	Salathe, Eric	University of Washington	CC* Compute: A campus-wide computing resource for research and teaching at the University of Washington Bothell
2126229	Chang, Philip	University of Wisconsin-Milwaukee	CC* Compute: A Balanced Cluster For Science and Engineering in the Great Lakes Region
2126253	Liu, Jason X	Florida International University	CC* Compute: RAPTOR - Reconfigurable Advanced Platform for Transdisciplinary Open Research

Award#	PI	Institution	Title
2201561	Neeman, Henry J.	University of Oklahoma Norman Campus	CC* Compute: OneOklahoma Cyberinfrastructure Initiative Research Accelerator for Machine Learning (OneOCII-RAML)
2201599	Tavakkoli, Alireza	Board of Regents, NSHE, obo University of Nevada, Reno	CC* Compute: Nevada Bridge to AI-enabled Scientific & Engineering Computing (NvBAISEC)
2201497	Skjellum, Anthony	University of Tennessee Chattanooga	CC* Compute: Augmenting a 2,560-core EPYC2 Computational Cluster with GPUs for AI, Machine Learning, and other GPU-Accelerated HPC Applications
2201428	Cleveland, Sean B.	University of Hawaii	CC* Compute: Koa - A High Performance and Flexible Research Computing Resource
2201435	Wheat, Stephen R.	Oral Roberts University	CC* Compute: GPU HPC Cluster Partition for Research, Education, and Student Success
2201479	Fondjo Fotou, Franklin	Langston University	CC* Compute: Collaboration in Computing Infrastructure for Research and Education (CO-InResE)
2201538	Pasricha, Sudeep	Colorado State University	CC* Compute: HPC Services for the Colorado State University System
2201106	Zink, Michael H.	University of Massachusetts Amherst	CC* Compute: COllaborative Next-generation Technology In the Northeast: the UMassUnity Machine (CONTINUUM)
2201558	Khan, Javed I.	Kent State University	CC* Compute: Accelerating Compute Driven Science Through a Sharable High Performance Computing Cluster in Kent State Multi-Campus System
2200792	O'Shea, Brian W.	Michigan State University	CC* Compute: The MSU Data Machine - a high-memory, GPU-enabled compute cluster for data-intensive and AI applications
2201445	Hanna, Chad	Pennsylvania State Univ University Park	CC* Compute: An Open Science Grid shared computing platform at Penn State
2201105	Kneifel, Charles	Duke University	CC* Compute: NCShare Compute as a Service

2021/2022 Award Example

NSF VIRTUAL CAMPUS CYBERINFRASTRUCTURE PI WORKSHOP

SEPTEMBER 19 – 21, 2022

Quad Chart for:

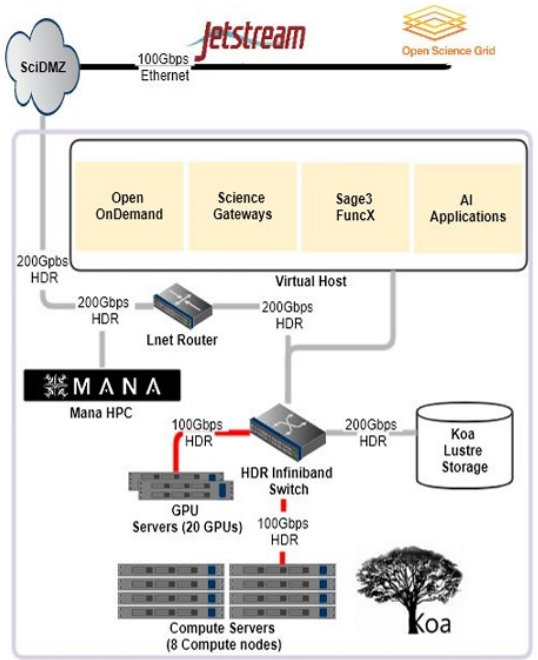
#2201428 CC* Compute: Koa - A High Performance and Flexible Research Computing Resource

Challenge Project Seeks to Address:

- Lack of computational resources to support data science, modeling and DL/AI workflows end-to-end.
- Research workflows/applications requiring sustained resources and interfaces.
- Lack of computational scratch storage to to support scaling large workflows.
- Large research dataset movement requirements from/to sites across the globe

Solution(s):

- Increase computational resource access with focus toward ML,AI and large scale simulation by deploying new advanced computational storage, HDR networking , CPU/GPU resources, and VM hosts
- Increase I/O throughput and compute capacity by deploying high performance computational storage resources.
- Provide fast data transfers in and out of the cluster to national, commercial cloud and academic resources over 100Gb science DMZ external and connected with HDR to internal resources



Broader Impact:

- Lower barriers of adoption by providing on-demand interactive & notebook environments with direct access to large scale storage & compute,
- Serve non-traditional users, i.e students from community colleges & non-technical domains.
- Advance data intensive research, education, and broaden participation for the students and citizens of the State of Hawai'i

Metadata tag:

- *Just Awarded 6/1/2022*
- *Equipment Procurement Finalization Stage*

Quad Chart for:

CC* Data Storage: Multi-Petabyte Open Storage (MPOS) at the American Museum of Natural History

Challenges Project Seeks to Address:

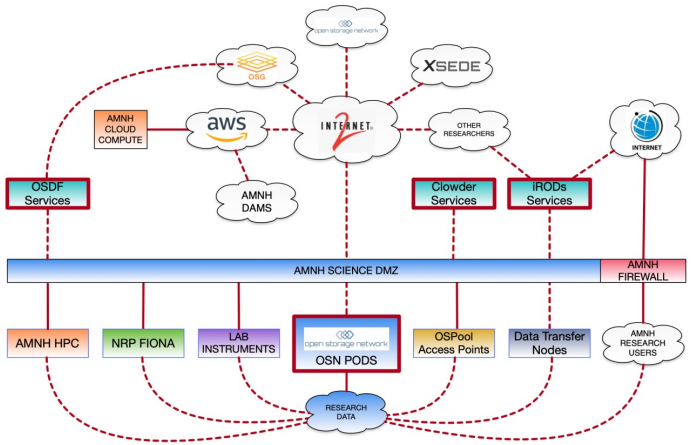
- Reduce or remove limitations on researchers arising from a lack of sufficient data storage.
- Limit duplication of datasets across various systems by providing a centralized storage system.
- Provide access to research data regardless of location (onsite, cloud, etc.)
- Improve data management capabilities and adherence to FAIR principles.

Deliverables:

- Installation of Open Storage Network (OSN) Pods providing 2.4+ petabytes of storage for active research data.
- Provisioning of 20% of the total OSN pod storage to the national data storage cloud.
- Provide interconnects between local and remote computational services and the OSN pods.
- Deployment iRODS Data Management Platform and the Clowder Framework at AMNH.

150 YEARS

AMERICAN MUSEUM OF NATURAL HISTORY



Scientific Impact:

- Provides researchers crucial storage resources accessible from a variety of onsite and remote computational and analytic systems.
- Allows for local storage of large-scale data sets (source and derivative).
- Enables the management and automation of research data ingest, tagging, indexing, and sharing..
- Facilitates the curation of data across the Museum and drives compliance with FAIR principles.

Metadata Tag:

- Museums
- Data Storage
- Data Management
- Open Storage Network
- iRODS
- Clowder
- Science DMZ
- Science Engagement
- GLOBUS
- Contact: nsf-ci@amnh.org

Late FY2022 New Areas: Data Storage (9) and Regional Compute (5) Awards

Award#	PI	Institution	Title
2232810	Kneifel, Charles	Duke University	CC* Data Storage: Flexible Affordable Scalable Technology for Research Storage (FAST-Research Storage)
2232816	Aygun, Ramazan S	Kennesaw State University Research and Service Foundation	CC* Data Storage: High Volume Data Storage Infrastructure for Scientific Research and Education at Kennesaw State University Shared as Open Science Data Federation Data Origin
2232803	Nabrzyski, Jaroslaw	University of Notre Dame	CC* Data Storage: Institutional Storage for the University of Notre Dame (NDStore)
2232817	Sun, Qi	New York University	CC* Data Storage: Distributed, Fast, Scalable Infrastructure for Emerging Media Research Data
2232872	Gough	Purdue University	CC* Data Storage: Software Defined Storage for Composable and HPC Workflows
2232601	Li	University of Washington	CC* Data Storage: Supporting Big-Data Edge Computing using Hybrid and Cloud-Native Storage Infrastructure
2232862	Cleveland	University of Hawaii	CC* Data Storage: KoaStore: A High Performance and Flexible Research Storage Resource
2232857	Montes	American Museum Natural History	Research Infrastructure: CC* Data Storage: Multi-Petabyte Open Storage (MPOS) at the American Museum of Natural History
2232851	Weitzel	University of Nebraska-Lincoln	CC* Data Storage: NRDStor: Nebraska Research Data Storage
2232895	Chakravorty, Dhruva	Texas A&M University	CC* Regional Computing: Launch: Advancing Cyberinfrastructure at Minority Serving Institutions
2232860	Kramer, William T	University of Illinois at Urbana-Champaign	CC* Regional Computing: Taylor Geospatial Institute Regional AI Learning System
2232917	Kesselman, Carl F	University of Southern California	CC* Regional Computing: Building Cyberinfrastructure to Forge a Regional Research Computing Alliance in Southern California
2232873	Florinski, Vladimir	University of Alabama in Huntsville	Research Infrastructure: CC* Regional Computing: A Regional Computing Hub for Alabama Universities
2232628	McKee, Shawn	University of Michigan	CC* Regional Computing: Helping Our Researchers Upgrade their Science (HORUS)

2023+

- CC* is making awards this summer off 23-526's 1st deadline
- 49 campus compute awards and counting
- 7 regional compute awards
- 12 data storage awards and counting
- There will be another CC* solicitation in 2024
- Looking for ways to broaden participation (MS-CC)
- Looking for ways to increase scientific usage in OSPool from CC* awardees

Thank You

Kevin Thompson
Office of Advanced Cyberinfrastructure
kthompso@nsf.gov

To subscribe to the OAC Announce Mailing List
Send an email to: OAC-ANNOUNCE-subscribe-request@listserv.nsf.gov

25