The Ecosystem for Research Networking (ERN): Democratizing Access to Research Instruments and Data

All Hands Meeting
March 23 – 24, 2023
Welcome & Introductions
Ecosystem for Research Networking (ERN)

Vision:
Simplify, support, catalyze, and foster multi-campus collaborations and partnerships between academic institutions of all types and sizes across the U.S. that advance the frontiers of research, pedagogy, and innovation.

Mission:
To achieve the vision through a consortium of academic institutions, research facilities, core service providers, network providers, and industry partners, both public and private, organized around a shared interest in supporting and enabling collaborative data and computation-enabled science by providing standards, blueprints, policies, and training associated with the design and implementation of an infrastructure to access data and research instruments, a distributed federated environment designed to simplify, support, and encourage collaborative science, scholarship, and education.

To realize the mission and vision, ERN will enable collaborations for democratization of access to research instruments, technical expertise, infrastructure, services, and resources to lower barriers to participation for scientists engaged in collaborative research across institutional and disciplinary boundaries.
Purpose of this AHM
Purpose of this AHM

To connect with the ERN community to:

– Discuss current and future ERN projects, collaborations and partnerships
– Give updates on the ERN Working Groups’ efforts and activities
– Discuss next steps on democratizing access to research instruments
– Share workflow and remote access initiatives
– Discuss scope of the ERN
– Discuss future funding opportunities, future workshops, and other topics of interest
The ERN All Hands Meeting

Topics of Interest:

• Security
• Network Infrastructure
• Organization - Governance, Bylaws, Committees
• Research Instrument and Data Abstractions for Accelerating Discovery
• Machine-actionable Data Management and Curation
• Leveraging the Current CI Ecosystem
• Translational Computer Science
• Education, Training, and Workforce Development
• Diversity, Equity, and Inclusion
Day 1 Agenda

1:00 - 1:15 Welcome & Introductions

1:15 - 1:30 Purpose of this AHM

1:30 - 2:15 Keynote - Democratization of Research
   Speaker Alan Blatecky - The Missing Millions

2:15 - 3:30 State of the ERN
   - Architecture and Federation
   - Broadening the Reach
   - Structural Biology, Materials Discovery, CS, Policy and Next Steps

3:30 - 4:00 Break

4:00 - 4:30 End of Day 1 Wrap-up

5:30 - 7:00 Reception
Day 2 Agenda

8:00 - 9:00 Meet and Greet Breakfast

9:00 - 9:15 Welcome and set tone for the rest of the morning

9:15 - 10:15 Breakout Groups - Discussion Topics:

10:15 - 10:45 Break

10:45 - 11:30 Open Discussion - Summary of Breakout Groups

11:30 - Noon Action items and Final Thoughts
State of the ERN
Eastern Regional Network 2019
Vision & Mission: Original Version

- **Vision**: To simplify multi-campus collaborations and partnerships that advance the frontiers of research and innovation.

- **Mission**: Through a partnership of educational institutions, research facilities, regional network providers, and Internet2, the ERN is committed to providing layered and transparent access to shared data and computing facilities for research projects located at and between partner sites to address the growing need for a regional research platform designed to support a diverse set of science drivers and education needs. The resulting layered approach offers the research community across the region access to a broad range of services and resources that are not available on any one campus alone.
Vision & Mission: Current Version

Vision:
Simplify, support, catalyze, and foster multi-campus collaborations and partnerships between academic institutions of all types and sizes across the U.S. that advance the frontiers of research, pedagogy, and innovation.

Mission:
To achieve the vision through a consortium of academic institutions, research facilities, core service providers, network providers, and industry partners, both public and private, organized around a shared interest in supporting and enabling collaborative data and computation-enabled science by providing standards, blueprints, policies, and training associated with the design and implementation of an infrastructure to access data and research instruments, a distributed federated environment designed to simplify, support, and encourage collaborative science, scholarship, and education.

To realize the mission and vision, ERN will enable collaborations for democratization of access to research instruments, technical expertise, infrastructure, services, and resources to lower barriers to participation for scientists engaged in collaborative research across institutional and disciplinary boundaries.
ERN Origins and Evolution

- **Germ of an idea:** at 2017 National Research Platform meetings
  - Can the regional research platform idea work in the Northeast?
  - Rutgers, OSHEAN, KINBER

- **Coalition of the Willing:** January 2018 gathering at Rutgers
  - Can we find something of common interest to work on?
  - KINBER, Rutgers, OSHEAN, MGHPCC, Internet2, NYSERnet

- **Resource Federation Proof of Concept:**
  - Can we work together to prototype a potentially beneficial resource federation idea?
  - Rutgers, MGHPCC, Syracuse, NJ Edge, U Maine, Google, Internet2
  - “Early if not elegant” approach; emphasis on getting people to work together
Origins and Evolution

● **Northeast perfSONAR Mesh:**
  ○ Can we work together toward a regional science DMZ?
  ○ Federation sites plus
    ■ BU, Brown, Colby, CEN, KINBER, Network Maine, OSHEAN, UNH
  ○ Less-specialized technology allows expanded participation

● **Growing interest:**
  ○ Maybe we really can do this?
  ○ [in progress] expanded steering committee; per-project working groups
  ○ All of the above plus
    ■ U Delaware, NJIT, U Buffalo, Bucknell, UMass Amherst, Yale, Princeton, BTAA, U Kentucky, Case Western Reserve
Resource Federation Proof of Concept

- Six sites, five states
- Ability to launch containerized (Singularity) jobs as well as traditional HPC jobs, from any site to any other site including commercial cloud, via SLURM scheduler.
- Working with Google, Cisco, Internet2, SchedMD (SLURM), and OSG on enhancements
- InCommon authentication and authorization
- Policy based federation of computing resources
- Reserving slices of the network
- Cloud bursting
- OSG federation with containerized SLURM (slurmd)
- Testing different data sharing approaches as well as filesystems

Recipe for joining:
https://github.com/rutgers-oarc/ern-poc
Federation with Commercial Cloud

- Partnership with Google Cloud and SchedMD (SLURM) to develop federated hybrid environment
  - Allows federation across resources (including cloud services)
  - Allows us to offer one stop shop for users
  - Allows on demand elastic computing
- Initial Proof of Concept
  - Set up federated environment between Rutgers and Google Cloud
  - Google Cloud instance is set up similar to local resource, includes SLURM
  - Connection through Internet2 between both environments
  - Able to submit jobs directly from Rutgers HPC to Google Cloud using SLURM submit script
- Elasticity/bursting capability now being tested
- Next steps: Extend to Jetstream, OSG
ERN perfSONAR Mesh

- 23 nodes, 14 sites, six states
- New sites welcome
- Still in teething phase
- Improves visibility on connectivity/bandwidth/delay
- Starting point for connectivity improvements (where they matter)
- Prerequisite for a regional data DMZ

Site:
http://mesh.hpc.rutgers.edu/maddash-webui/

Recipe for joining:
https://github.com/rutgers-oarc/ern-poc
ERN’s Evolution - How we got to where we are now
NSF CC* CRIA OAC-2018927

• Received notice of funding during AHM last year (June)
• Project Team: Goodhue, Honavar, Pitt, Segee, and von Oehsen
• Funding period July 1, 2020, to June 30, 2023
• Working Groups
  – Materials Discovery (Shashank Priya, Chair)
  – Structural Biology (Stephen Burley, Chair)
  – Architecture/Federation/Computer Science (Maureen Dougherty, Michael Zink, Co-Chairs)
  – Policies (Ron Hutchins, Chair)
  – Broadening the Reach (Forough Ghahramani, John Hicks, Co-Chairs)
• Collect information to be used for future initiatives and funding opportunities
Policy Working Group

- Melissa Cragin
- Maureen Dougherty
- Anthony Elam
- Jenni Evans
- Clark Gaylord
- John Goodhue
- Jim Griffioen
- Helen Hill
- John Huffman

- Vasant Honavar
- Ron Hutchins*
- Barr von Oehsen
- Mike Zink
Broadening the Reach Working Group

- Galen Collier
- Scott Frees
- John Goodhue
- Forough Ghahramani*
- Ventsi Gotov
- John Hicks*
- John Huffman
- Jim Kyriannis
- Derek Leydig
- Julie Ma
- Dave Marble
- John McNutt
- Linh Ngo
- Todd Price
- Barr von Oehsen
- Jennifer Oxenford
- Andy Sherman
- Scott Valcourt

[Bar chart showing topics of interest for the workshop]

[Bar chart showing barriers to adequate research computing resources]
Architecture & Federation Working Group

- Tom Barton
- Steve Cousins
- Eva Deelman
- Heidi Dempsey
- Bala Designhu
- Maureen Dougherty*
- Adam Focht
- Forough Ghahramani
- Jim Griffioen
- Ventsi Gotov
- John Goodhue
- Vasant Honavar
- Dave Hudak
- Ron Hutchins
- Mahmut Kandemir
- Orran Krieger
- Eric Lyons
- John McNutt
- Barr von Oehsen
- Chuck Pavloski
- Bruce Segee
- Marc Sugarman
- Robert Settlage
- Scott Valcourt
- Ann West
- Boyd Wilson
- Scott Yockel
- Matt Zekaukas
- Mike Zink*
Structural Biology Working Group

- Stephen Burley*
- Chris Hill
- Arek Kulczyk
- Chen Xu
- John Goodhue
- Barr von Oehsen
- Ed O'Brien
- Mike Zink
- Bala Desinghu
- Maureen Dougherty

Materials Discovery Working Group

- Liping Yu
- Shashank Priya*
- Bala Desinghu
- Vasant Honavar
- Alisa Kang
- Emanuela Del Gado
- Barr von Oehsen
- John Goodhue
- Maureen Dougherty
Planning Grant Workshops and Results

• Three information gathering workshops
  – ERN Architecture and Federation Virtual Workshop, December 2-4, 2020
  – ERN Broadening the Reach Virtual Workshop, December 10-11, 2020
  – Structural Biology The Voice of the Customer Online Workshop, February 11, 2021

• Year One Significant Results
  – NSF Mid-Scale RI-1 solicitation, pre-proposal
  – ERN Federated OpenCI Labs Design
  – ERN Policy Planning
    • Bylaws and charter
  – Four white papers (two publications)
  – Data CI Pilot
OpenCI Labs

- Distributed Federated Cloud Architecture
  - Building Block design
- Gateway to instruments on the edge and Core Facilities
  - CryoEM, X-ray diffraction, Scanning electron microscope, Focused ion beam, Transmission electron microscopy, and gene sequencers
- Data access & Distributed Workflows
  - Materials-centric structured databases
  - Data-sharing policies and standardizations
  - Machine learning/Deep Learning for materials discovery across all scales
  - Distributed research computing (virtualization/containers, bare metal)
- Minimal configuration:
  - Open OnDemand, OSN Gateway (or node), and federation
Step 2 Design

Edge Resource

Research Scientific Instrument

Instrument System
- Static VNC
- Port Forwarding

Access/Bridge System

Camera Server
- Raw Data

Data Hop Server
- Raw Data Access

Instrument Cloudlet

Instrument Portal pod

- Open OnDemand Container Frontend Proxy

noVNC Client

VNC Server Container
- Flask App
- TEM Access Option

Per User Nginx OOD Container

CryoSPARC Container

Instrument Portal

Edge Computing

- Raw Data Access
- /Projects/Data Access

GPU

Amarel Cluster

Performance Storage

Compute Cluster

Remote Researcher Browser

Edge Router

100G 10G 1G Direct Mount Mixed
Publications


Publications

Participation

**Academic Institutions:**
Alabama State University, Arcadia University, Boston University, Brown University, Case Western Reserve, Clemson University, Columbia University, Cornell University, Delaware State University, Franklin and Marshall College, George Washington University, Georgetown University, Harvard University, Kentucky State University, Lafayette College, MIT, Montclair State University, New Jersey Institute of Technology, Northeastern University, Pace University, Penn State University, Princeton University, Ramapo University, Rutgers University, Southern Connecticut State University, Stevens Institute of Technology, Syracuse University, The College of New Jersey, Trinity College, Tufts University, USMA, University of Arkansas, University of Buffalo, UC Santa Barbara, University of Chicago, University of Delaware, University of Illinois, University of Kentucky, University of Maine-Orono, University of Massachusetts-Amherst, University of Michigan, University of New Hampshire, University of New Haven, University of Utah, University of Virginia, West Chester University of Pennsylvania, and Yale University.

**Organizations/Industry/RENS:**
Massachusetts Green High Performance Computing Center (MGHPCC), Northeast Big Data Innovation Hub, Ohio Supercomputer Center, SDSC, Internet2, CAAREN, CEN, KINBER, NetworkMaine, NJEdge, NYSERNET, Great Plains Network, Omnibond, OSHEAN, CILogon, CaRCC, Redhat, SHI, IBM
What’s Next?
Special Thanks

• Maureen Dougherty
• Joylynn Almeida
• All the ERN working groups
• The ERN Steering Committee
• The National Science Foundation
• University of Delaware
Thank You!

Interested in learning more or participating, please contact info@ernrp.org

Website: https://ernrp.org