INTERNET2
2023
COMMUNITY
exchange

May 8-11, 2023   Atlanta, GA

Remote Instrument Collaboration

Maureen Dougherty: Ecosystem for Research Networking
Amy Cannon: Omnibond Systems

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.

ECOSYSTEM for RESEARCH NETWORKING

Remote Instrument Collaboration

Research community feedback received by the ERN Structural Biology Working Group through community outreach activities, voiced the need for multi-institutional collaboration at the interface of computing and electron microscopy leveraging real-time result monitoring.

Benefits
- Adjust experiment parameters live
- Identify target achieved/fruitless runs
- Broaden collaborative efforts, science discovery
- Optimize instrument utilization

Barriers
- Security
- Access limitations
- Significant latency issues
- Insufficient data transfer rates
- HPC queue wait times (public and private)
- Training and expertise

ERN CryoEM Remote Instrument Pilot Project Objectives

Facilitate and simplify multi-institutional collaborative research by removing many of the barriers encountered when attempting to access remote scientific instruments.

- Secure, Easy to use, web-based resource portal
- Simplified, federated authentication, authorization and access
- Real-time workflow adjustments
- Edge computing
- Access to additional analysis resources – private and/or public
- Portable, easily duplicated, managed and maintained
- Ease expertise and training limitations
- Secure data management system
- Do not reinvent the wheel
- Institution/owner maintains ownership and oversight of resources
- Security should augment existing institution/lab security policies and procedure
- Share efforts with the research community

Challenges

- Identity Management

- Data Management

- Security

- Infrastructure

- Reproducible, Reliable, Reusable

- Policies

ECOSYSTEM for
RESEARCH
NETWORKING

Pilot Project Design
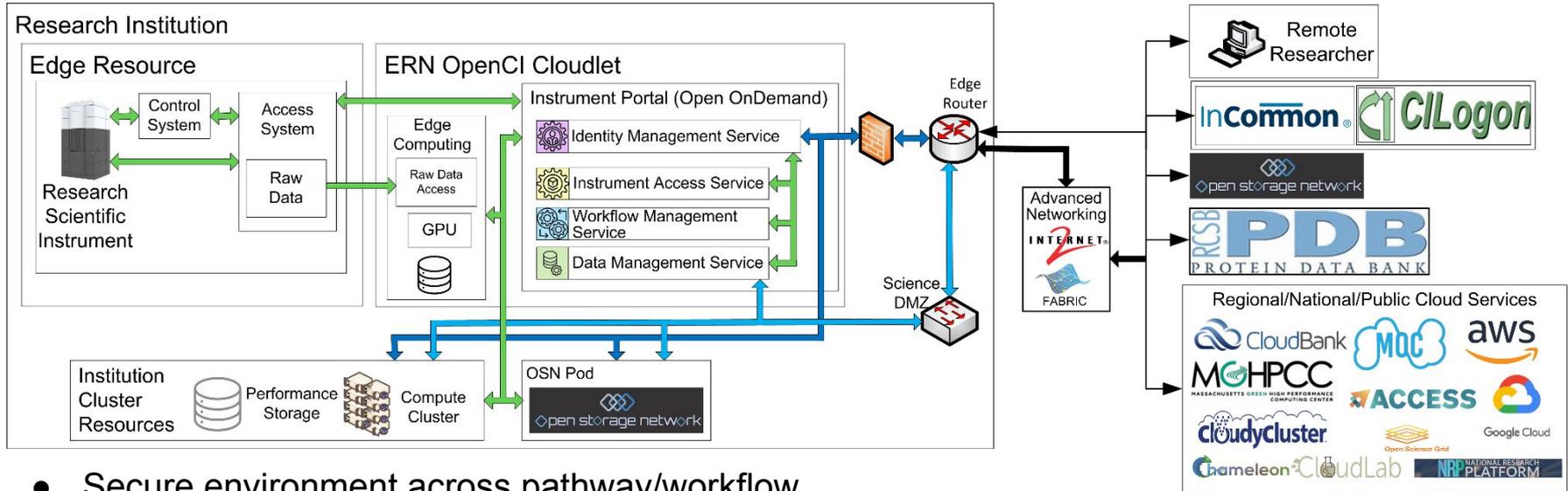
- Secure environment across pathway/workflow
- Common framework for federated authorization, authentication, and access
- Reproducible, reliable, portable, simplified support
- Edge computing
- Open Source project

Phase 1 – Instrument Accessibility

Basic framework for Instrument Cloudlet scientific instrument access

Collaborators
- Rutgers CryoEM & Nanoimaging Facility (RCNF)
- ERN Structural Biology and Architecture and Federation Working Groups members

Step 1: Basic Remote Accessibility

- VNC/noVNC communications between remote researcher and scientific instrument
- Real-time processing workflows leveraging Cloudlet edge computing

Step 2: Instrument Portal Implementation

- Deploy Instrument Portal interface
- Incorporate federated authentication
- Leverage open source resources
- Validate with Step 1 workflow

Step 1 - Basic Remote Accessibility
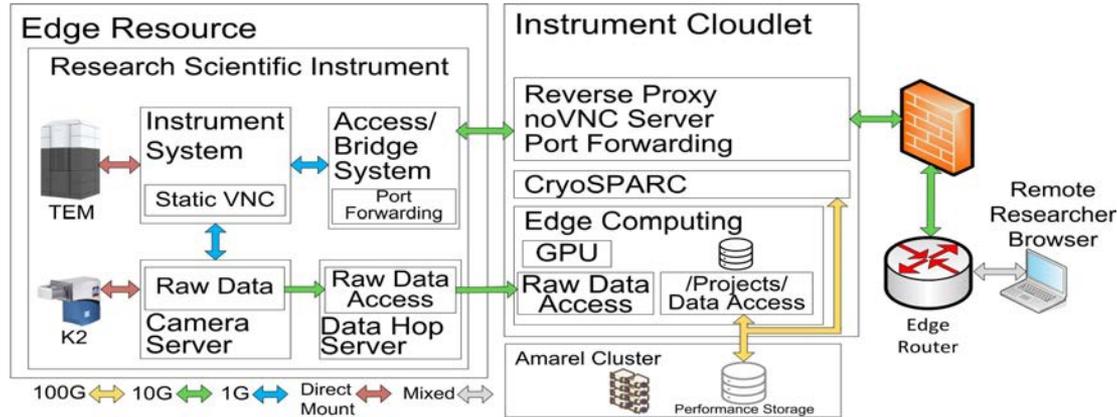
Workflow

- Off-institution remote system establish Rutgers VPN session
- noVNC access through remote web browser to Instrument System's static VNC
- Workflow launched
- CryoSPARC application
  - Real-time adjustment decisions made
  - Raw datasets pre-process image with edge computing GPUs
  - Amarel cluster job submission 2D alignment and 3D structure refinement of pre-processed image files

ECOSYSTEM for RESEARCH NETWORKING

Step 1 - Basic Remote Accessibility
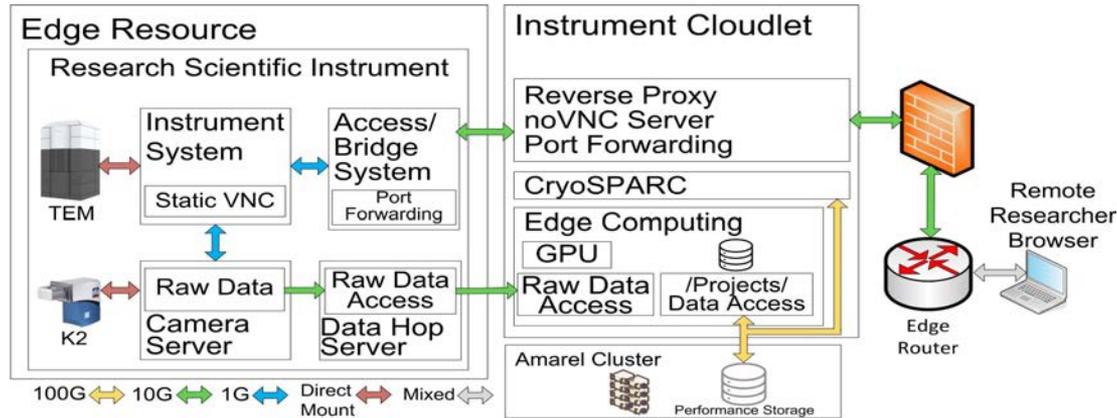
Results

- 320 images/hour novel complex of the transmembrane protein TolC
- Processed 2.5TB over 2 days
- Computational output < 3 minutes behind actual data acquisition
- Real-time experiment adjustments made based on quality of incoming TEM data

- Bandwidth measurements confirmed data transfer rate from instrument > 1GB
- Data transfer rate from cloudlet to Amarel cluster < 1GB
- Network I/O reduced by 1-2 orders of magnitude due to Cloudlet edge computing

Step 2 - Instrument Portal Implementation
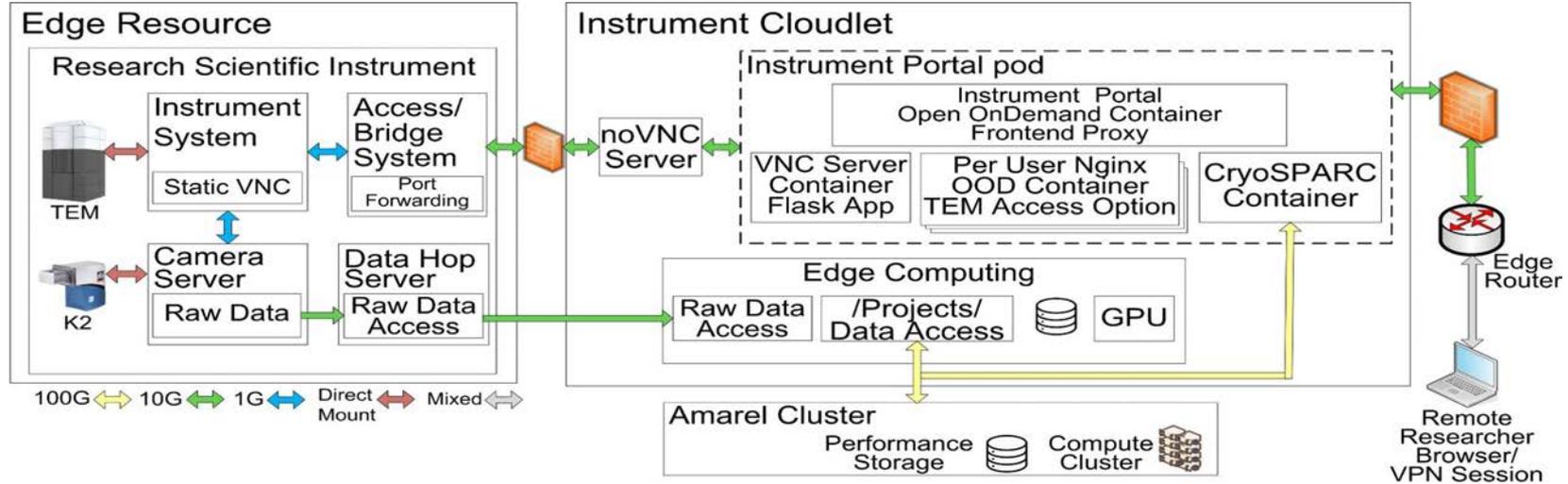
Simplify Step1 VNC access through the Instrument Portal

- Integration of Open OnDemand(OOD)
  - CILogon and local LDAP
  - Configuration files Navbar customization
  - GitHub repository for solutions and documentation
- Validate successful migration of containerized OOD with workflow utilizing edge computing

Step 2 – Testbed Implementation
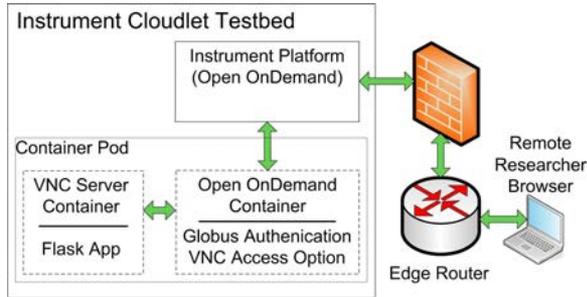
- Development in testbed environment without access to a TEM
- Podman-containerized Open OnDemand (OOD)
- Globus for authentication, mapping against local user file
- Local Apache SSL certificates leveraged
- Parameterized configuration files,

CryoEM Remote Instrument Access Demonstration

https://vimeo.com/776242831?embedded=true&source=video_title&owner=188132628

Phase 1 Conclusion

Remote access to edge scientific instruments for real time analytical workflows using edge computing is both feasible and beneficial

Benefits

- **Remote access to scientific instrument in secure environment**
- **Real-time decision making and adjustment**
- **Edge computing**
- **Decreased network I/O** for pre-processed image data
- **Reliability, reproducibility, reusability, portability, ease of use/management/support**
- **Github repository** for community participation and contributions
- **Foster team science and democratization of scientific instruments with emphasis on under-represented and under-resourced colleges and institutions**

Lessons Learned

- security: traffic isolation, rootless container, per-user permissions
- expertise: subject matter experts, researcher and technical expertise important

Next Steps

- CryoSPARC interaction through portal

  - Open OnDemand engagement for web based application

- Pegasus Workflow and Data Management System integration

- FABRIC integration

- Enable access to external researchers

  - Engage with additional interested institutions

The ERN Federated CryoEM Instrument Pilot Project Site Map

Participating Member Sites (alphabetical order)
- Massachusetts Green High Performance Computing Center
- Omnibond
- Pennsylvania State
- Rutgers University
- University of Massachusetts, Amherst
- University of Minnesota
- University of Southern California

Future Partnering Sites (alphabetical order)
- American Indian Higher Education Consortium
- Arizona State University
- Harvard University
- Kennesaw State University
- Rowan University
- University of California, Santa Cruz
- University of Florida, Gainesville
- University of Utah

INTERNET2 2023 COMMUNITY EXCHANGE

ECOSYSTEM for RESEARCH NETWORKING

Acknowledgements

- James Barr von Oehsen, Pittsburgh Supercomputing Center
- Ken Dalenberg, Bala Desinghu, Jason Kaelber, and Jeremy Schafer, Rutgers University
- Wolf Hey, Penn State University
- John Goodhue, MGHPCC
- Morgan Ludwig, TechSquare
- Amy Cannon, Boyd Wilson and Cole McKnight, Omnibond
- Michael Zink, University of Massachusetts, Amherst
- Ewa Deelman, Mats Rynge, University of Southern California
- Maureen Dougherty, Ecosystem for Research Networking
- ERN Steering Committee
- The Open OnDemand team
- The FABRIC team
- NSF OSN-2018927

Interested in learning more or participating, please contact <u>info@ernrp.org</u>

GITHUB: <u>https://github.com/mghpcc/ERN-Remote-Scientific-Instrument</u>

Website: https://ernrp.org

ECOSYSTEM for RESEARCH NETWORKING