

Community Engagement at Scale: NSF Centers of Expertise

PEARC 2019 Panel

Daniel Crawford (MoISSI), Ewa Deelman (CI CoE Pilot),
Von Welch (ResearchSOC, Trusted CI),
Frank Wuerthwein (OSG), Mike Zentner (SGCI),
Jason Zurawski (EPOC),
Moderator: Ruth Marinshaw (Stanford University)

Agenda

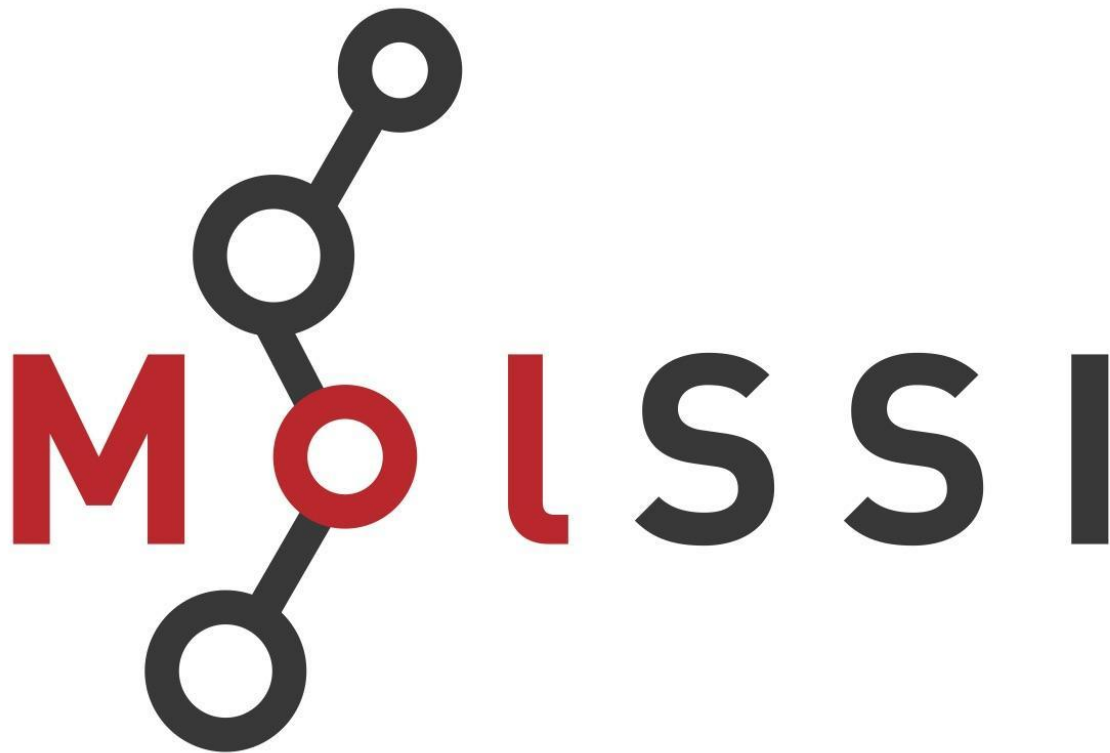
Five minutes each:

- Daniel Crawford (MolSSI)
- Ewa Deelman (CI CoE Pilot)
- Jason Zurawski (EPOC)
- Von Welch (ResearchSOC, Trusted CI)
- Frank Wuerthwein (OSG)
- Mike Zentner (SGCI)

Moderated discussion

Moderator: Ruth Marinshaw (Stanford University)

Daniel Crawford (MoISSI)

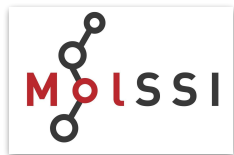


The Molecular Sciences Software Institute

... a nexus for science, education, and cooperation
for the global computational molecular sciences community.

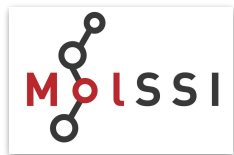
What is the MolSSI?

- Launched August 1st, 2016, funded by the National Science Foundation.
- Collaborative effort by Virginia Tech (TDC), Rice U. (C. Clementi), Stony Brook U. (R. Harrison), U.C. Berkeley (T. Head-Gordon), Stanford U. (V. Pande), Rutgers U. (S. Jha), U. Southern California (A. Krylov), and Iowa State U (T. Windus).
- Part of the NSF's commitment to the White House's National Strategic Computing Initiative (NSCI).
- Total budget of \$19.42M for five years, potentially renewable to ten years.
- Joint support from numerous NSF divisions: Advanced Cyberinfrastructure (ACI), Chemistry (CHE), and Division of Materials Research (DMR)
- Designed to **serve** and **enhance** the software development efforts of the broad field of computational molecular science (CMS) – a broad domain that includes quantum chemistry, computational materials science, and biomolecular simulation.



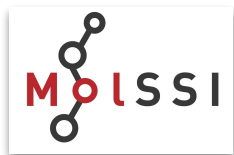
Who is the MolSSI?

- Software Scientists: A team software engineering experts, drawn both from newly minted Ph.D.s and established researchers in molecular sciences, computer science, and applied mathematics.
- Software Fellows: A cohort of ~24 graduate students and postdocs supported simultaneously and selected from research groups across the U.S. by the MolSSI's Science and Software Advisory Board.
- Board of Directors: Seven co-PIs who oversee the MolSSI's activities and provide guidance and expertise.
- Science and Software Advisory Board: Representatives from academia, industry, national laboratories, and international facilities who advise the MolSSI on the most important software priorities for the community.
- Community-Code Partners: Approximately 40 computational molecular science software packages whose developers work with the MolSSI on standards, training, and infrastructure.



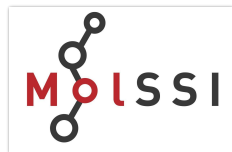
The MolSSI's Goals

- To provide the molecular sciences community with software expertise and infrastructure;
- To provide education and training in modern software engineering best practices and tools;
- To provide community engagement and leadership to help the molecular sciences domain establish standards of data and code interoperability.



MolSSI Highlights So Far

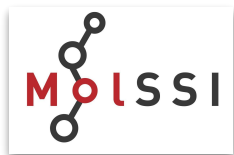
- Hired **twelve Software Scientists** – the full contingent as originally planned, but we are considering hiring an additional team member.
- **17 software workshops with more than 500 participants** so far; at least another eight to be held in 2019.
- New software components currently under development including an **open QM database (QCArchive)**, a **general QM/MM driver**, a **new basis set exchange**, a **reference integral** implementation, and more.
- **Community-driven working groups** established in forcefield interoperability, quantum chemistry data exchange, and tensor algebra interfaces.
- **Nearly 400 students** directly engaged so far in two Software Summer Schools, five “Best Practices” workshops, three Software Fellowship bootcamps, and three undergraduate programming schools; many more educational workshops and schools coming in 2019-20.
- **24 Software Fellows currently supported**, plus **11 new Fellows** who started July 1st, for a **total of 50 Fellows** funded overall.



Watch molssi.org for the latest information!

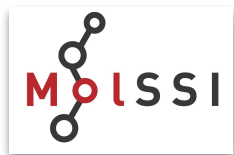
Community Engagement Mechanisms

- Software Fellowship program
- The MoISSI Educational Initiative
- Software Scientist outreach and collaborations
- Software community-led workshops
- Community code partners
- Discovery process for new software infrastructure projects
- The MoISSI Associates program
- Industrial Affiliates Programs
- Social media and other marketing efforts
- Direct outreach by the Board of Directors and the Science and Software Advisory Board



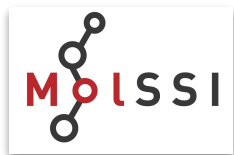
MolSSI Software Scientists (MSSs)

- A team of ~12 software engineering experts, drawn both from newly minted Ph.D.s and established researchers in molecular sciences, computer science, and applied mathematics.
- Dedicated to multiple responsibilities:
 - Developing software infrastructure and frameworks;
 - Interacting with CMS research groups and community code developers;
 - Providing forums for standards development and resource curation;
 - Serving as mentors to MolSSI Software Fellows;
 - Working with industrial, national laboratory, and international partners;



MolSSI Software Fellows (MSFs)

- A cohort of ~24 Fellows supported simultaneously – graduate students and postdocs selected by the Science and Software Advisory Board from research groups across the U.S.
- Fellows work directly with the Software Scientists and the MolSSI Directors, thus providing a conduit between the Institute and the CMS community itself.
- Fellows work on their own projects, as well as contribute to the MolSSI development efforts where appropriate, and they engage in outreach and education activities under the Institute guidance.
- Funding for MolSSI Software Fellows follows a flexible, two-component structure, providing up to two years of support.



MolSSI Software Fellowship Program



MolSSI Coding Workshops at the 2017-2019 MERCURY Conferences



Focus: Undergraduate Training in Python, Git, GitHub, Jupyter Notebooks, in QM and MD Simulations
Furman University

Organizers: Paul Nerenberg, Aurelia Ball, Lee-Ping Wang, Olaseni Sode, and Theresa Windus

Participants: 60 Undergraduates (40 Female, 18 URM) & 8 Faculty from 14 PUIs
All materials available in the MolSSI GitHub repo



2019 MolSSI Software Workshops

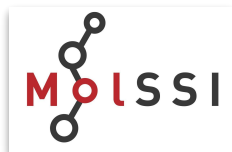
- Stochastic Approaches to Electronic Structure Calculations, U. Pittsburgh, July 29-August 2, 2019 (**Organizer**: K. Jordan)
- Software for Advanced Potential Energy Surfaces, 1-2 August 2019 (**Organizers**: T. Head-Gordon and S. B. Sinnott)
- Parallel Computing in Molecular Sciences, Stony Brook U., 5-8 August 2019 (**Organizers**: Bert de Jong, Edward Valeev, Carlos Simmerling, Robert Harrison)
- Open Molecular Science Cloud, Perugia and Rome, Italy, September 3-6, 2019 (**Organizer**: A. Lagana)
- Molecular dynamics software Interoperability, New York, NY, 3-5 November 2019 (**Organizers**: M. Chavent, J. Chodera, C. Clementi, E. Lindahl, J.-P. Piquemal)
- Rovibrational Molecular Spectroscopy: Setting standards for software packages and toolkits, MolSSI HQ, November 14-15, 2019 (**Organizers**: B. Poirier, A. Császár, J. Tennyson)
- Machine learning and chemistry: challenges on the way forward, College Park, MD, November 16-18, 2019 (**Organizers**: P. Tiwary, O. Isayev, A. Roitberg)
- Continuing the Development of Flexible and Robust Software in Computational Atomic and Molecular Physics, National Institute of Standards and Technology, December 11-13, 2019 (**Organizer**: B. Schneider)



MolSSI Marketing



- The MolSSI signed a contract with a new and hungry marketing firm, Blue Mobius in April 2018:
 - 113% increase in overall website traffic and an 83% increase in “organic” website traffic (e.g., traffic arising from Google searches);
 - Social media engagements have increased by a factor of 10 (4k/month vs. 40k/month)
 - Twitter followers have increased significantly from 274 in April 2018 to 957 as of July 28, 2019.
 - Designed a project logo (QCArchive) and the new website (coming soon!)
 - They have produced a set of short films in the next few months highlighting efforts by members of our Software Scientist team (to be released with the new website)
 - They have produced a series of blog posts highlighting both the MolSSI’s goals/activities and our Software Scientist team.
- We hired Invision Communications (UK) to produce a film on the MolSSI for the European Chemical Society meeting in Liverpool in August 2018.



MolSSI Board of Directors



Cecilia Clementi, Rice University, Co-Director for International Engagement

T. Daniel Crawford, Virginia Tech, Director



Robert J. Harrison, Stony Brook University, Co-Director for Parallel Computing and Emerging Technologies

Teresa Head-Gordon, U.C. Berkeley, Co-Director for Laboratory, Industrial, and Academic Outreach and Education



Shantenu Jha, Rutgers University, Co-Director for Software Engineering Process, Middleware, and Infrastructure

Anna Krylov, U. Southern California, Co-Director for Quantum Chemistry and Materials



Theresa Windus, Iowa State University, Deputy Director and Co-Director for Code and Data Interoperability



Cyberinfrastructure Center of Excellence Pilot

Ewa Deelman, USC (PI)

Co-PIs:

Anirban Mandal, RENCi

Jarek Nabrzyski, Notre Dame University

Valerio Pascucci and **Rob Ricci**, University of Utah



Searching for
gravitational waves

Understanding ocean
and coastal
ecosystems

Looking for exoplanets

Studying
climate

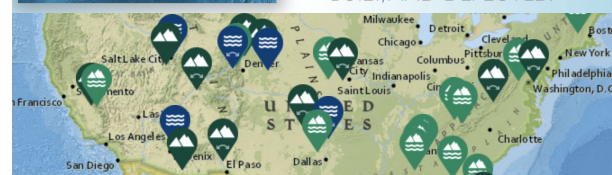


 **National Radio
Astronomy
Observatory**



THE INFRASTRUCTURE

89 PLATFORMS
CARRYING OVER
830 INSTRUMENTS
PROVIDING OVER
100,000 DATA PRODUCTS
HAVE BEEN DESIGNED,
BUILT, AND DEPLOYED.



The National Ecological Observatory Network: Open data
to understand how our aquatic and terrestrial ecosystems are
changing.


National Ecological Observatory Network

There are limited interactions
and limited knowledge sharing
among large facilities and large
CI projects.

Manish Parashar (PI and Chair), Rutgers University and OOI
Stuart Anderson, LIGO
Ewa Deelman, USC
Valerio Pascucci, University of Utah
Donald Petravick, LSST
Ellen M. Rathje, NHERI

NSF Large Facilities Cyberinfrastructure Workshop



IceCube

- **Establish a center of excellence** (following a model similar to the NSF-funded Cybersecurity Center Trusted CI as a resource providing expertise in CI technologies and effective practices related to large-scale facilities as they conceptualize, start up, and operate.
- Foster the creation of a facilities' CI community and establish mechanisms and resources to enable **the community to interact, collaborate, and share.**

September 2017 Workshop report at <http://facilitiesci.org/>

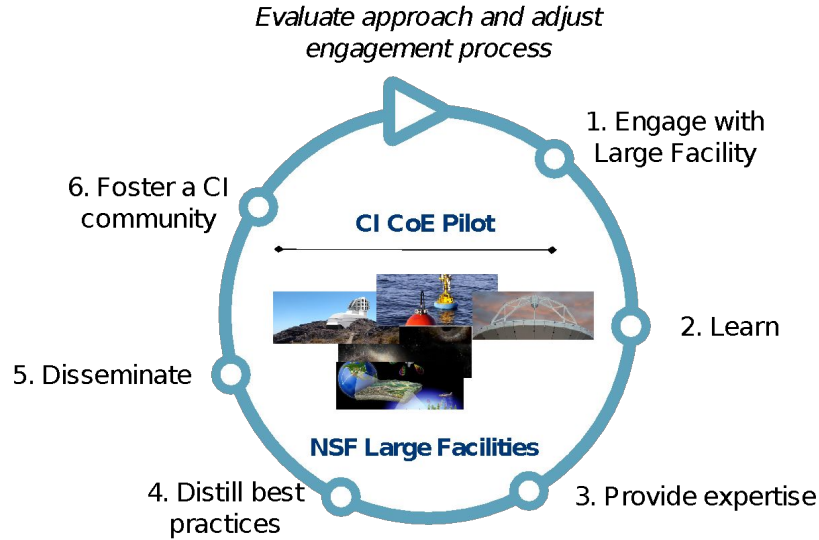
Develop a model and a plan for a Cyberinfrastructure Center of Excellence

- Platform for knowledge sharing and community building
- Key partner for the establishment and improvement of Large Facilities with advanced CI architecture designs
- Grounded in re-use of dependable CI tools and solutions
- Forum for discussions about CI sustainability and workforce development and training
- Pilot a study for a CI CoE through close engagement with NEON and further engagement with other LFs and large CI projects.

10/2018-9/2020

1. Recognize the expertise, experience, and mission-focus of Large Facilities
2. Engage with and learn from current LFs CI
3. Build on existing knowledge, tools, community efforts
 - Avoid duplication, seek providing added value,
4. Prototype solutions that can enhance particular LF's CI
 - Keep a separation between our efforts and the LF's CI developments
5. Build expertise, not software
6. Work with the LFs and the CI community on a blueprint for the CI CoE

Developing and improving Engagement Processes

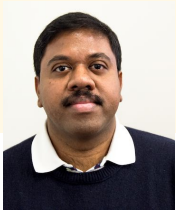
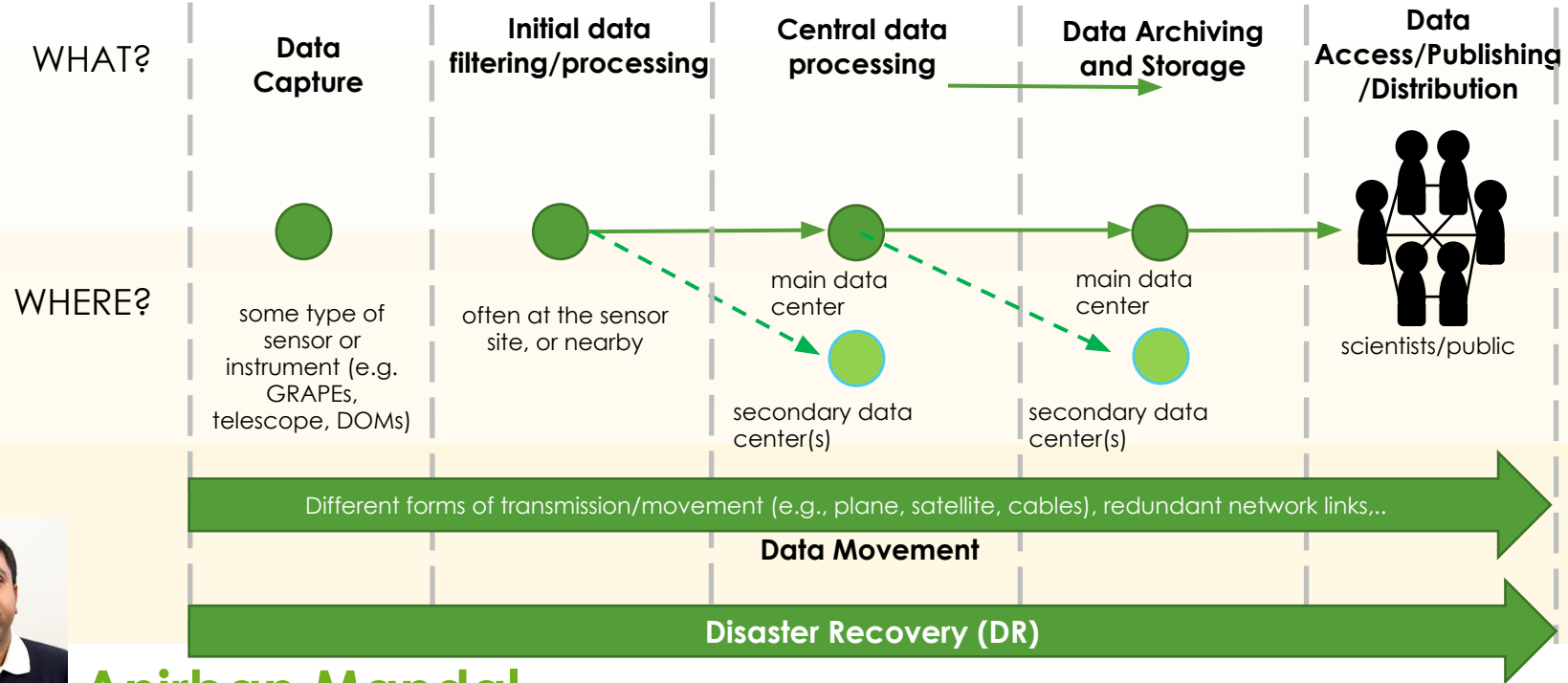


Engagement with a Facility

- Engage at the management level, potentially seek introductions from NSF PO, participate in community meetings (LF Workshop)
- Initial virtual technical group discussions to define possible avenues of engagement
- In person meeting with a number of technical personnel
- Identity topics for engagement
- Set up working groups
- Follow up email and conference call discussions focused on particular topics/working groups
- Bigger group discussions/checkpointing
- Reports of engagement, gather feedback from the project engaged

- **Deep engagement:**
 - Identify a topic that is important and not-yet fully solved by the LF,
 - Conduct focused discussions, mix of virtual and in-person presence, hands-on work
 - Includes an engagement template that defines scope, sets expectations, identifies products
 - Work products: documents/papers, prototypes, schema implementations, demos
- **Topical discussions:**
 - Identify a topic that is important to a number of LFs
 - Facilitate virtual discussions, sessions at conferences, collect and share experiences, distill best practices
 - Discover opportunities for shared infrastructure
- **Community building: bringing in new members to the CI CoE Pilot effort**
 - Identify related efforts
 - Collect information and disseminate information about the broad community activities
 - Maintain a living resource for community information
- **Each engagement has a working group with a leader and a set of work products.**

- Data capture
- Data processing
- Data storage/curation/preservation
- Data access/visualization/dissemination
- Disaster recovery
- Identity management
- Engagement with Large Facilities



Anirban Mandal

1. Work with NEON to identify concrete collaboration points and actionable information we can provide
2. Based on NEON interactions develop a blueprint that can scale out the approach to other Large Facilities and beyond
3. Engage individual LFs: workshops, meetings, tele-conferences, attendance at project meetings and community events
4. Develop an online presence
5. Build a community around the planned CI CoE

Learn from and collaborate closely with Trusted CI



The Engagement and Performance Operations Center

Dr. Jennifer M. Schopf (PI)
Indiana University International Networks

Jason Zurawski (Co-PI)
ESnet / Lawrence Berkeley National Laboratory

Dave Jent (Co-PI)
Indiana University GRNOC



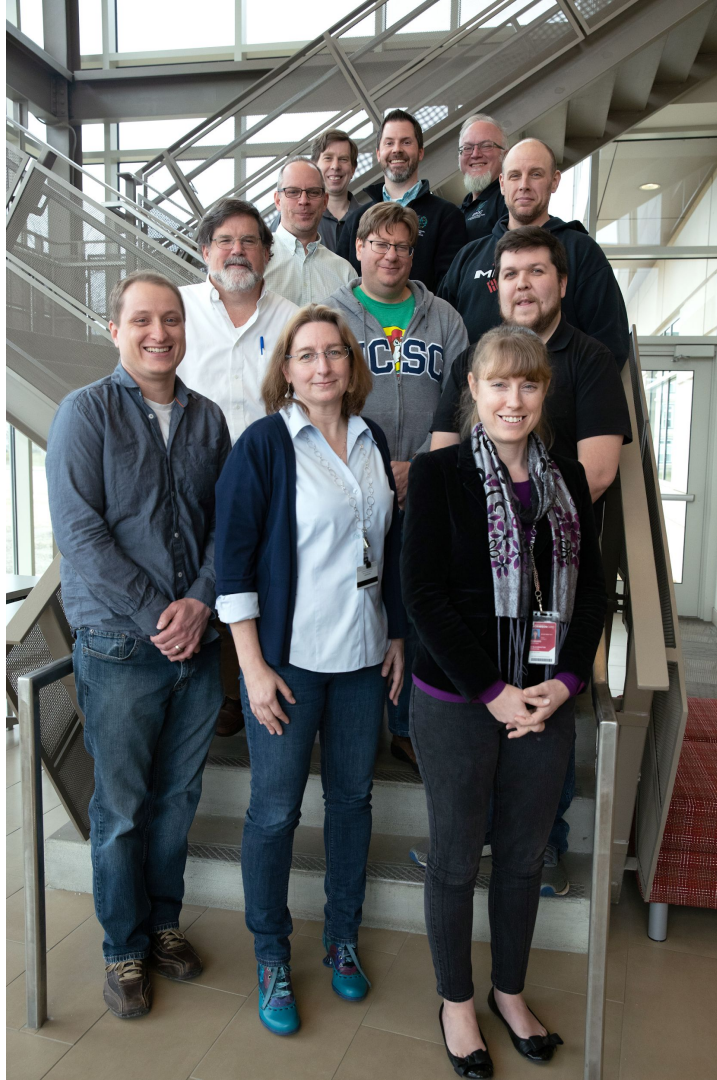
ESnet
ENERGY SCIENCES NETWORK

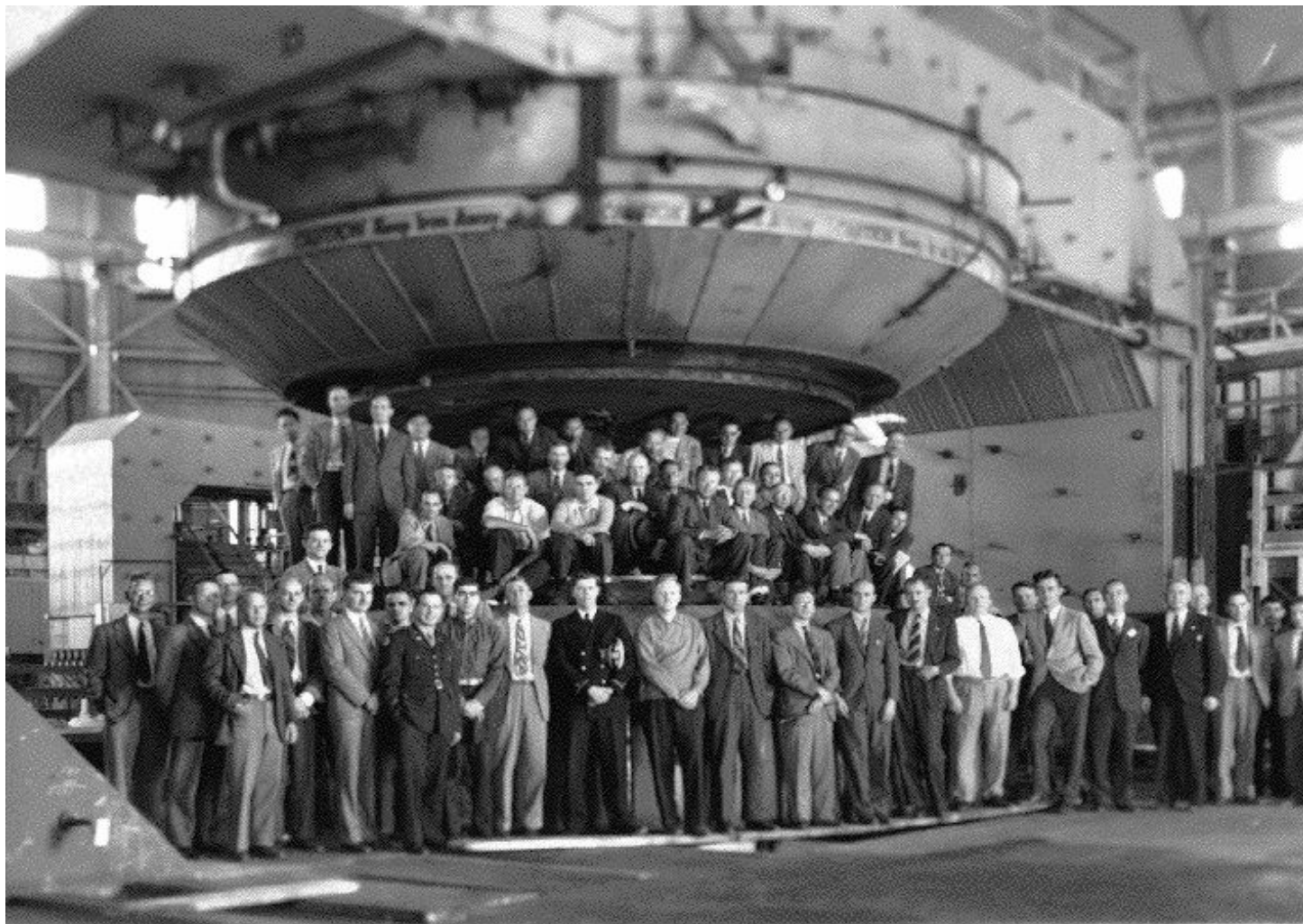


INDIANA UNIVERSITY

Engagement and Performance Operations Center

- Joint project between Indiana University and ESnet
- Part of CC* program for domestic science support
- Award #1826994, \$3.5M over 3 years
- Partnerships with regional, infrastructure, and science communities that span the NSF and DOE continuum of funding





Why an Engagement Operations Center?

- Today's science is collaborative science
- Collaborative science
 - Multiple partners
 - Multiple data sets
 - Many points of connection
 - Cross agency cooperation
- With better access to data we ask harder questions
- Interactive data sources change the science we do

Network as Infrastructure *Instrument*



Connectivity is the first step – **usability** must follow



EPOC Five Main Focus Areas

1. Roadside Assistance for Performance Problems
2. Application Deep Dives
3. Network Analysis (NetSage)
4. Services “in a box” (DMZ, testpoint in a box, etc)
5. Training

Move Slow, and Build Relationships

- "Move fast and break things. Unless you are breaking stuff, you are not moving fast enough."
- Mark Zuckerberg



YOU CAN PRESENT THE
MATERIAL, BUT YOU CAN'T
MAKE ME CARE.



The Golden Spike



- We don't want scientists to have to build their own networks, hardware, or software
- Engineers don't have to know the nuances of scientific instruments
- Meeting in the middle is the process of science engagement:
 - Engineering staff learning enough about the process of science to be helpful in how to adopt technology to needs
 - Science staff having an open mind to better use what is out there



Von Welch

NSF Awards 1547272 and 1840034



Trusted CI: The NSF Cybersecurity Center of Excellence

Our mission: to lead in the development of an NSF Cybersecurity Ecosystem with the workforce, knowledge, processes, and cyberinfrastructure that enables trustworthy science and NSF's vision of a nation that is a global leader in research and innovation.



<https://trustedci.org/>

Trusted CI: Impacts

Trusted CI has positively impacted over 260 NSF projects since inception in 2012.

Members of more than 180 NSF projects have attended our NSF Cybersecurity Summit.

Members of more than 80 NSF projects have attended our monthly webinars.

We have provided more than 300 hours of training to the community.

We've had engagements with 41 projects, including nine NSF Large Facilities.



The Trusted CI Broader Impacts Project Report

June 28, 2018
For Public Distribution

Jeannette Dopheide¹, John Zage², Jim Basney³

<https://hdl.handle.net/2022/22148>



ResearchSoc



UC San Diego

Operational cybersecurity services for research.

Building on existing services (OmniSOC, STINGAR) and expertise to bolster the NSF cybersecurity community's incident response capabilities.

Ramping up in 2019, initial clients in 2020, sustaining in 2021.

<https://researchsoc.iu.edu/>

Timeline and Initial Clients

2019

Project start

Development of tech and contracts; outreach to InfoSec and Researchers

2020

Beta Testing

2021

Sustainability and for-free services



NATIONAL RESOURCE FOR TRANSLATIONAL
AND DEVELOPMENTAL PROTEOMICS

Engagement Challenges

- ❖ Funding cycle has made community skeptical - projects come and go.
- ❖ Community loath to accept risk of relying on external partner.
- ❖ Different cultures and “languages” among research areas.
- ❖ Outreach venues are rare.

Engagement Challenges and Lessons

- ★ Built trust: Listen, talk their talk, and don't oversell,
- ★ Show early wins - get word of mouth.
- ★ Rule of seven and a half: Communicate, communicate, communicate.
- ★ Find and leverage community champions.

Advancing Open Science through distributed High Throughput Computing



Frank Wuerthwein
OSG Executive Director
Professor of Physics
UCSD/SDSC

Advancing All of Open Science

- All of Open Science **irrespective of discipline.**
- Advance the **maximum possible dynamic range** of science, groups, and organizations:
 - From **individual undergraduates** to international collaborations with thousands of members.
 - From **small colleges, museums, zoos** to national scale centers of open science.
- Advancing this entire spectrum requires us to have a **diversified portfolio of services and strategies.**

OSG thinks of its “customers” as 4 Distinct Groups

- **Individual researchers** and small groups across all fields of science.
 - OSG supports everything from login platform, to storage, to training on how to use the OSG compute and data federations.
- **Campus Research Support Organizations**
 - Teach IT organizations & support services so that campuses can integrate into the OSG compute and Data Federations
 - “Train the trainers” to support their researchers
- **Multi-institutional science teams**
 - XENION, GlueX, SPT, and many many more
 - Science collaborations that span multiple campuses, often internationally.
- **The big 4 Science Projects**
 - US-ATLAS, US-CMS, LIGO, IceCube

Following the “eat your own dog food” model

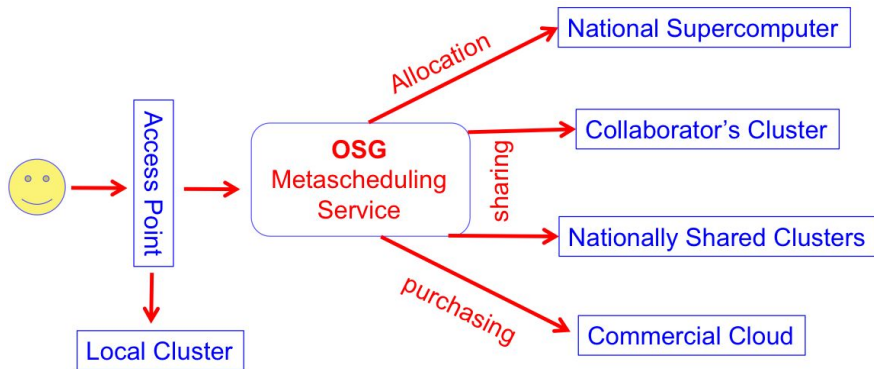
- To successfully engage the full dynamic range of organizations, we find ourselves mediating a creative tension between “doing as we preach” and employing “scalable engagement models”
- **Doing as we preach:**
 - We operate instances of the services that we offer as downloadable software
 - We follow the training and facilitation strategy that we promote
 - Focus on being solutions oriented ... even if it occasionally compromises our principles.
- **Scalable engagement models:**
 - Our goal is to “train the trainers”, i.e. engage organizations and help them support their communities.
 - Constantly trying to engage with new partners on projects of common interest.
 - Strong interest in helping projects that support science communities

OSG Compute and Data Federations

Federation means policy of use is local. => **“Owners Rule”**

Compute Federation utilization of 200-300,000 CPUs and ~2,000 GPUs.

Any campus can join and provide as little or as much as they want.



Data Federation is a **network** of “data origins” and “caches” to facilitate uniform data access across global compute federation.

Towards a Content Delivery Network for Open Science

6 Data Origins
12 Data caches across US, EU, and Asia



Open Governance

- The OSG Consortium is **governed by an unfunded Council** that represents the major stakeholders.
 - Council controls its own membership and elects the Executive Director
 - The **day-to-day operations is lead by an Executive Director**
 - Complex funding streams across half a dozen awards from multiple directorates and agencies.
 - Annual planning exercise that provides the bulk of the direction for the day-to-day work, but leaves open changes at the margins throughout the year.
- **Value responsiveness to change !!!**



Connecting people and resources
to accelerate discovery by empowering
the science gateway community



The Science Gateways Community Institute

Michael Zentner

Director, Science Gateways Community Institute

Director for Sustainable Scientific Software, San Diego Supercomputer Center

Director, HUBzero Platform

sciencegateways.org

Award Number
ACI-1547611



(expensive) Resources



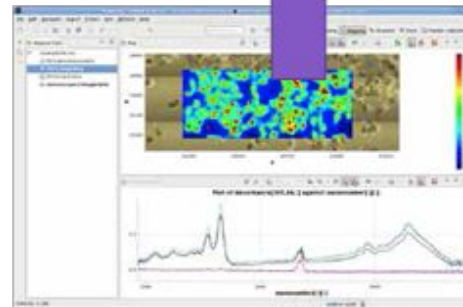
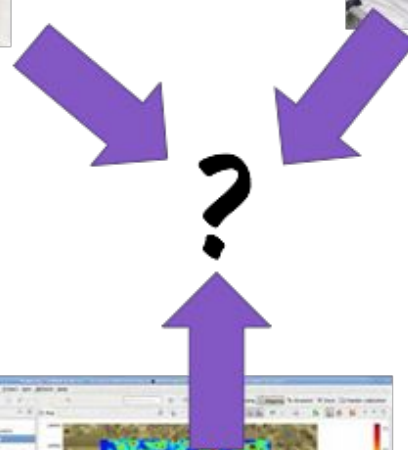
Audience(s)

Pain Points

Connecting expensive
resources & data to

scientist authors that
write codes to utilize
them to

audiences that use
those codes, data,
instruments,
collaborate, educate,
etc. en masse



Collaborating Scientist
Software / Data
Authors

Our Products:

Technology Consulting

- ✓ Technology selection, planning
- ✓ Development process implementation
- ✓ Cybersecurity scanning, design & remediation
- ✓ Usability assessment / design
- ✓ Co-developer placement



Our Products:

Operations Consulting

- ✓ Marketing approach development
- ✓ Communication assistance
- ✓ Graphic design
- ✓ Licensing issues
- ✓ Activity based budget & resource planning
- ✓ Understanding your market
- ✓ Funding / sustainability planning

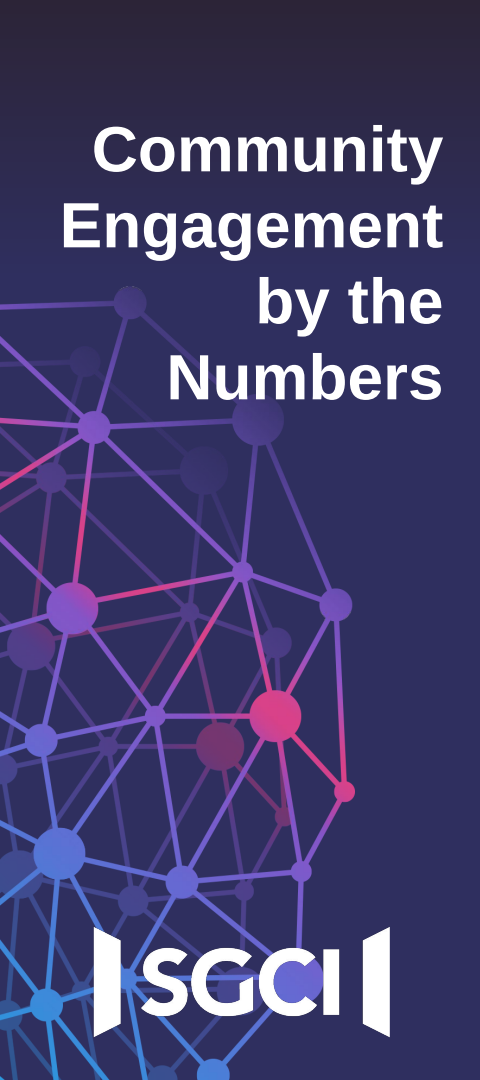


Our Products:

Education and Engagement

- ✓ Gateway “Focus Week” (Sept 9-13)
- ✓ Annual Gateways conference (Sept 23-15)
- ✓ Gateway catalog
- ✓ Hackathons (Wednesday 3:30 PM)
- ✓ Internships
- ✓ Gateway ambassadors
- ✓ Science ambassadors
- ✓ Technical forums





Community Engagement by the Numbers



- 1509 event participants
 - 771 webinar attendees
 - 210 Focus Week attendees
- 26 partners & affiliates
- 32 success stories
- 67 consultations
- 106 letters of collaboration
- 586 gateway catalog entries



Results of Community Engagement: Approaching Sustainability

Recognized Value of Services

\$614,000

Additional Funding from External Projects

\$25,000

Additional Funding from Focus Week Events

Demonstrated Value Propositions

41%

Collaborating Project Hit Rate

8.5x

Demonstrated Return on Investment
More Money for Science

Nearly All Services Have A Waiting List

Lessons learned...

...so far

- ✓ Seek opportunities that provide leverage
- ✓ Do things that get people talking about what you do
- ✓ People remember interactive experiences
- ✓ Create opportunities for follow-up “check-ins”
- ✓ Good work without telling the story does not sell itself
- ✓ “Sales is never not your problem!”
- ✓ *Our challenge: identify more points of leverage but be able to scale in different ways to meet the anticipated increase in demand*



Discussion

How have you effectively worked with universities, colleges, research labs, etc.?

How have you (or could you) work with the boots-on-the ground CI teams who support NSF-funded researchers at institutions?

as you listened to each other, what
“lessons” or approaches are
generalizable across all?

Related ... how do you all (your Centers)
stay in touch with each other and share
practices, information and strategies?

What workforce challenges do you see
specific to community engagement?
What would you propose to “solve” or
help address those?

How do you know you have succeeded?
How to you evaluate, measure, assess,
gauge, think about the value of your
program?

Wait - the PEARC19 theme is “Rise of
the Machines”. Can’t AI do this???

To Contact Presenters

- Daniel Crawford - crawdada@vt.edu
- Ewa Deelman - deelman@isi.edu, CI CoE Pilot, <http://cicoe-pilot.org>
- Ruth Marinshaw - ruthm@stanford.edu
- Von Welch - vwelch@iu.edu - trustedci.org - researchsoc.iu.edu
- Frank Wuerthwein - fkf@ucsd.edu
- Michael Zentner - mzentner@ucsd.edu
- Jason Zurawski - zurawski@es.net

Acknowledgements

- NSF Award #1842042 - Pilot Study for a Cyberinfrastructure Center of Excellence
- NSF Award #1826994 - CC* NPEO: The Research and Science Engagement Center: A Production Platform for Operations, Applied Training, Monitoring, and R&E Support (a.k.a EPOC)
- ResearchSOC: Award NSF award 1840034
- Trusted CI: Award 1547272

The views expressed do not necessarily reflect the views of the National Science Foundation or any other organization.