

Cyberinfrastructure Center of Excellence Pilot: Connecting Large Facilities Cyberinfrastructure

Ewa Deelman,

University of Southern California, PI

Co-Pls: Anirban Mandal, RENCI Jarek Nabrzyski, Notre Dame University Valerio Pascucci and Rob Ricci, University of Utah



Funded by the National Science Foundation Grant #1842042







CENTER FOR APPLIED CYBERSECURITY RESEARCH

TI







Cyberinfrastructure "consists of computing systems, data storage systems, advanced instruments and data repositories, visualization environments, and people, all linked together by software and high performance networks to improve research productivity and enable breakthroughs not otherwise possible."¹

 ¹ M. Parashar, S. Anderson, E. Deelman, V. Pascucci, D. Petravick, and E. M. Rathje,
 "2017 NSF Large Facilities Cyberinfrastructure Workshop," 2017. [Online]. Available: http://facilitiesci.org/assets/reports/facilitiesci-workshop-report-11-17.pdf











Clis a critical component of Science: Large Facilities (LFs)





Searching for gravitational waves

Understanding ocean and coastal ecosystems

Looking for exoplanets

Studying climate

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

HE INFRASTRUCTURE



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









CENTER FOR APPLIED CYBERSECURITY RESEARCH







October 16th 2017:

"LIGO and Virgo make first detection of gravitational waves produced by colliding neutron stars"



"aftermath of the BNS merger....... 16 days after the merger NRAO's Jansky Very Large Array (VLA) captured the radio image" from LIGO.org Images credit: LIGO Scientific Collaboration



ISC Viterbi School of Engineering Information Sciences Institute











Manish Parashar (Pl 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

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

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

USC Viterbi School of Engineering Information Sciences Institute





CENTER FOR APPLIED CYBERSECURITY RESEARCH renci

Ū



Project Goals

Ш

CENTER FOR APPLIED CYBERSECURITY RESEARCH



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

- Dedicated to the enhancement of CI for science
- 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









Project Team



USC Ewa Deelman Mats Rynge Karan Vahi Loïc Pottier Rafael Ferreira da Silva Ryan Mitchell



Automation, Resource Management, Workflows

RENCI

Anirban Mandal Ilya Baldin Laura Christopherson Paul Ruth Erik Scott



Resource Management, Networking, Clouds, Social Science















University of Notre Dame Jarek Nabrzyski Jane Wyngaard **Charles Vardeman**

University of Utah Valerio Pascucci, Rob Ricci, Timo Bremer, Attila Gyulassy, Steve Petruzza





Workforce development, Sensors, Semantic technologies

Data management, visualization, clouds, large-scale CI deployment

Indiana University Susan Sons (co-funded by Trusted CI) Von Welch (unfunded collaborator)



Cybersecurity









Ш





Overall Strategy



- 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

Build partnerships:

- Trusted CI (identity management): share personnel
- Open Science Grid (data and workload management): share expertise
- Campus Research Computing Consortium (CaRCC): workforce development















Developing and improving Engagement Model Evaluate approach and adjust engagement process 1. Engage with Large Facility 6. Foster a CI community **CI CoE Pilot** 2. Learn 5. Disseminate **NSF Large Facilities** 4. Distill best 3. Provide expertise practices

Process for Engagement with a Facility

- Engage at the management level, potentially seek introductions from NSF PO, participate in meeting (LF Workshop, LF CI 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









National Ecological Observatory Network Mission

NEON provides a coordinated national system for monitoring critical ecological and environmental properties at multiple spatial and temporal scales. ...transformative science ...workforce development



NSF

20 ecoclimatic domains

distinct landforms, vegetation, climate, and ecosystem dynamics.

Terrestrial sites:

terrestrial plants, animals, soil, and the atmosphere,

Aquatic sites: aquatic organisms, sediment and water chemistry, morphology, and hydrology.

Data collection over 30 years

27 Relocatable terrestrialsites13 Relocatable aquatic sites



Slide courtesy of Tom Gulbransen, NEON





Anirban Mandal, lead

Data Lifecycle for LFs



What services correspond to the data lifecycle stages?



Working Groups and Products co-lead by the Pilot and NEON



Working group	Goals	Products
Data Capture	Develop demonstrators and comparisons of the multiple architectures for data capture at the sensor to data deposition in a repository	 Prototype: architecture demo on github: <u>https://github.com/cicoe/SensorThingsGost-Balena</u>
Data Life Cycle & Disaster Recovery	Develop a general set of DR requirements and policies that can inform the LFs about best practices for DR and how those can be adapted for specific facilities.	 Document: Disaster recovery template Document: Filled out template example (IceCube) Webinar: Best Practices for NSF Large Facilities: Data Life Cycle and Disaster Recovery Planning
Data Processing	Provide support and distill best practices for workflows and services related to the processing of data.	• Paper: "Exploration of Workflow Management Systems Emerging Features from Users Perspectives" (in submission
Data Storage, Curation, & Preservation	Compare and be able to consult on different data storage, curation and preservation technologies.	 Document: Competency questions based on scenarios that domain experts may use Google dataset search for NEON dataset discovery Presentation: at ESIP on schema.org Small containerized prototype of publishing neon vocabularies as linked data and linked data connection









CENTER FOR APPLIED CYBERSECURITY RESEARCH





Working Groups and Products



Working group	Goals	Products
Data Visualization & Dissemination	Understand the access, visualization and user interaction workflows in large facilities. Distill best practices and provide solutions to improve the access and usability of the available data.	 Document describing AOP data visualization cyberinfrastructure Online demo and video: Visualizing AOP Data <u>https://cert-data.neonscience.org/data-</u> <u>products/DP3.30010.001</u>
Identity Management	Understand current practice in authentication and authorization and help mature practice across the NSF Large Facilities.	 Production deployment: Connection to CI Logon NEON data download (using existing university / organization credentials) <u>https://cert-data.neonscience.org/home</u> Paper: <u>NEON IdM Experiences</u> (in submission to NSF Cybersecurity Summit)
Engagement with Large Facilities	Engage with Large Facilities and other large cyberinfrastructure projects to foster knowledge and effective practice sharing; 2) define avenues of engagement, modes of engagement, and plan community activities.	 Document: LF engagement template Presentations: SCIMMA project meeting, 2019 LF meeting, PEARC'19 Paper: Invited e-Science 2019 paper













- 1. Importance of f2f discussions, building relationships and trust
- 2. Benefits of formalizing the engagement: expectation, timelines, resources to use
- 3. Importance of LF priorities and challenges, importance of good timing
- 4. Organizing work around working groups and work products
- 5. Be open to learn about what works, don't fix it (workflow management)
- 6. Co-existence of old and new systems, making for a heterogeneous CI landscape

















CoE Pilot Benefits to NEON Thus Far

- Short ramp-up due to receptivity/readiness to change
- Broadened network of expert CI colleagues
- Major upgrade to Data Portal's remote sensing visualization
- Accelerated Data Portal completion plan
- Affirmed strategies for workflow, messaging, & DR
- Raised critical mass of attention on semantics & schema.org
- Excited software developers
- Escalated accountability of CI
- More coming

Slide courtesy of Tom Gulbransen, NEON









Tom Gulbransen









• 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:

- Connect CI professionals
- Collect information and disseminate information about the broad community activities
- Maintain a living resource for community information
- Develop new partnerships

• Each engagement has a working group with 1-2 leaders and a set of work products.













To find out more

Things to come: engagement with other large facilities and the broader CI community

- http://cicoe-pilot.org
- <u>ci-coe-pilot@isi.edu</u>
- Ewa Deelman <u>deelman@isi.edu</u>
- Connecting LF CI workshop, 2019: http://facilitiesci.org









CENTER FOR APPLIED CYBERSECURITY RESEARCH

