



# 2019 LARGE FACILITIES WORKSHOP

APRIL 2-4 | TEXAS ADVANCED COMPUTING CENTER

## Information Section

<b>Topic:</b>	Facility Cyberinfrastructure
<b>Speaker(s) Name, Title:</b>	Bill Miller, Ewa Dellman, Manish Parashar
<b>Session Description:</b>	The session will highlight important recent initiatives in the area of cyberinfrastructure for facilities
<b>Session Time Slot:</b>	Thursday, April 4, 2:20pm
<b>Purpose and Desired Outcome:</b>	Present on NSF OAC funded CI efforts in support of Large Facilities; provide an overview of NSF's thinking, programs and recent grants

## Notetaking Section

<b>Scribe Name</b>	
--------------------	--

**Disclaimer:** *These are raw notes that were captured by the assigned scribe during this session at the 2019 Large Facilities Workshop. This is one individual's interpretation of what took place during the session, and its content does not necessarily represent the viewpoint of the National Science Foundation.*

### Notes and Key Points:

- Ewa: Goal to develop a model and plan for CI center of excellence; platform for knowledge sharing; partner with existing and new facilities; grounded in re-use of dependable CI tools; forum for discussion; working with NEON as case
- Engage with LF CI; find out goals, challenges; build on existing tools; prototype solutions when needed; work with community of blueprint for CI Center of Excellence
- Developing engagement processes: engage; learn; provide expertise; distill best practices; disseminate; foster community
- Set up working groups; engage bigger discussions/groups; report out
- First case: NEON. Had meetings; shared design docs (i.e: data management); started developing data pipelines; data presentation options; understanding data lifecycle and disaster recovery
- Spun up groups to work closely with NEON: data capture; data processing; vis/dissemination; disaster recovery and identity management – working more broadly with other facilities.
- Ex: data pipelines for sensors: found useful tools being used: Airflow, Pachyderm;
- Ex: data portal/visualization: developing prototypes of interactive portal to view data in different resolution that's more intuitive for a user; being considered
- Ex: Data Lifecycle: mapping needs, tools, and disaster recovery solutions
- Digital recovery is cross-cutting across LFs. Working on DR planning template to share.
- Lessons: face-to-face discussions are invaluable; formalizing engagement; efforts help organize future CI CoE activities
- Tom from NEON: readiness was high; fast ramp-up
- Broadened expertise of staff
- Will put remote sensing vis into action in coming year
- Accelerated data portal plan by months
- Improved accountability of NEON staff
- Would be useful to have methods for CI self-assessment

- Ewa: engagement beyond NEON: helping with identity management, important to lots of LFs; lots of solutions; working to understand pros/cons; developing doc to outline solutions
  - Starting to work with IceCube, SCIMMA (multi-messenger astronomy software institute), LSST
  - Collaborating with TrustedCI, researchSOC (community building, training)
  - Hoping to work with more LFs; publish results; explore workforce development
  - Goal: at end of 3 year effort have blueprint for CI CoE
- 
- Manish Parashar: Director of OAC, CISE, NSF
  - Workshop on LF CI: saw gaps; recognized need for org to collect/disseminate best practices et al
  - CI CoE could fill gap; running pilot to determine value
  - LFs have huge potential impact; often CI not done right can prevent impact from being realized; done right can amplify impact
  - OAC overview: \$224M FY2018 budget; 305 awards
  - Foster CI ecosystem to transform S&E research (through CI and CI research)
  - Through all components of CI: support research and production services; and people/orgs
  - “Transforming Science through Cyberinfrastructure”: draft released 2/27/19: included interaction with LFs
  - Plan to build Leadership Class Facility: phase 1: Frontera; phase 2: anticipated to be part of MRFEC process
  - NSF Big ideas: 6 research; 4 process. \$30 million/idea in budget; also “convergence accelerator” – bring academia and industry together to solve societally relevant problems;
  - CI is a key enabler for NSF Big Ideas
  - CI is central to NSF’s large facilities; workshop in 2017 led to report.
  - Found many LFs were facing same issues; volume of data: how to keep up? Meet needs of evolving community?
  - Lack of community; lots of reinventing the wheel
  - Opportunities: i.e.: on demand data processing; end-to-end science workflows
  - Holistic view of CI for facilities: Instruments -> internal CI: collect data; process; present to user [enterprise practices exist; challenges of cybersecurity; workforce retentions] -> data products out to users: workflow; analysis
  - Observations:
  - Facility science complex:
    - solution: develop CI at beginning in conjunction with LF planning; refresh frequently;
  - Science workflows evolving:
    - Develop missing middle
  - CI designs are challenging;
    - Don’t do it alone; re-use best practices
  - Results/programs to address: Community planning for SCIMMA (Scalable CI for Multi-Messenger Astronomy)
  - IRIS-HEP: Institute for Research and Innovation in software for High-Energy Physics (for LHC upgrade and beyond)
  - Developing the Missing Middle (between facility and user’s responsibilities): i.e. data streaming; Intelligent delivery:
  - Virtual data Set Services Enabling New Science at NSF Facilities; et al
  - Everyone building custom CI; can we find commonalities, best practices? Pilot for CI CoE
  - Meeting tomorrow to discuss CI at length; recommendations for...
  - 2019 NSF CI for Facilities Workshop

**Best Practices:**

- To improve CI for LFs, face-to-face discussions are invaluable; identify/understand needs; find areas where improvements can be made; create working groups to attack problems.
- In the case of NEON this led to development of new remote sensor visualizations, speed up of data portal development.
- Process of collaboration helps CI CoE Pilot improve/formalize engagement; helps organize future activities
- From OAC perspective, CI is critical for Large Facilities; can amplify impact or hinder development
- Looking to develop mechanisms to identify/re-use/disseminate best practices for CI: [led to creation of CI CoE Pilot and other grants]
- Also working to fund develop tools for Missing Middle – products that sit between LFs and user workflows -- to make science more impactful

**Actionable Recommendations (Action Owner Name & Organization):**

- CI CoE should engage with NIST FFRDC in Cybersecurity (NITRD-funded);
- Look to NASA for ideas for developing re-usable CI tools, centers; they have good models
- Utilize CI CoE as sounding board for thinking through big CI problems that LFs face/will face in the future

**Decisions:**

- 
-

## Session Summary

Manish Parashar, director of NSF's OAC, presented an overview of NSF's thinking, programs and recent grants for funded CI efforts specifically in support of Large Facilities; Ewa Dellman, a recent recipient of an NSF award to develop a Cyberinfrastructure Center of Excellent Pilot program, described recent efforts to collaborate with NEON and other projects to improve large facility cyberinfrastructure. Both presenters highlighted the goal of identifying best practices and gaps/common needs and developing re-usable solutions that can serve a number of large facilities, as well as better mechanisms for improving the processes of developing CI including workforce development and community-building across facilities.