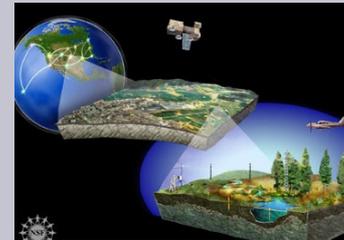# Large Facilities & Cyberinfrastructure
Follow-up from December 2015 workshop

Bill Miller

Science Advisor
Division of Advanced Cyberinfrastructure (ACI)
CISE Directorate
National Science Foundation

WLMiller@nsf.gov

Cyberinfrastructure (CI)

- Science-related computational and data capabilities, resources and services that serve the research and education end-users (generally, outward facing).

  - Computing, software, data infrastructure, workflows, portals, networking, and related workforce…

  - Distinguish from Information Technology (IT): not science related, part of business operations to benefit facility's own personnel.

  - Some elements like cybersecurity may span CI and IT.

- Facility CI (in-house capabilities)…
  - vs. external shared research CI
  - vs. commercial resources (e.g. cloud).

Facilities and Shared Research CI

Facility CI
Capabilities

Products &
Services
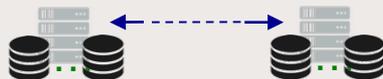
Portals, Data, Software, Analysis tools

External

"Middleware"
CI Services

- Data Services – Access, Discovery, Analytics, Semantics
- Science APIs, Portals, Gateways
- Workflow systems and other CI

Computing
Resources

**NSF-supported
CI ecosystem**

campus, regional
resources

commercial
cloud

international

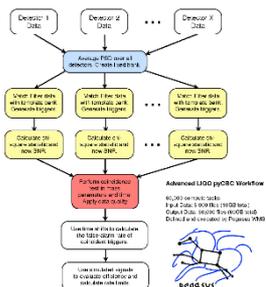Networks

National/International Research and Education Network

LIGO science enabled by shared research cyberinfrastructure (CI)

Courtesy SXS.

Open Science Grid

Computing

- **Open Science Grid (OSG):** High throughput parallelized event searches of 100,000 models for neutron binary star mergers, black holes,...
- **Comet** (SDSC) and **Stampede** (TACC) **HPC** computations, via **XSEDE** allocations system**.**

Workflows

- **Pegasus** and **HTCondor**: create, distribute, monitor OSG jobs, manage data transfers, analyze LIGO data.

Networking

- NSF-funded upgrade from 10Gbps to 100Gbps WAN enabled huge gains in throughput.

NSF programs: Data Building Blocks (DIBBs), Software Infrastructure (SI2), Campus Cyberinfrastructure Network Infrastructure and Engineering (CC*NIE, DNI), and others. OSG and Pegasus also supported by the DOE.

NSF's goals

❖ **Gain a deeper understanding of cyber needs <u>within</u> and <u>external to</u> large facilities to support large scale science.**

- Identify needs, gaps and trends that can guide future NSF investments.

❖ **Foster dialog and collaboration <u>among large facilities</u> and <u>with the CI community</u>.**

- Exchange practices, success cases, and technology awareness, trends.
- Maximize use of existing CI resources, minimizing duplication of effort.
- Develop new partnerships to address challenges, and inspire new R&D.

❖ **Ultimately: Create a dynamic *national CI ecosystem* that** supports the changing needs of the research community.

Cyberinfrastructure for NSF Large Facilities Workshop

December 1st, 10 a.m. – 6:00 p.m. & December 2nd, 8:30 a.m. to 2:30 p.m., 2015

Westin Arlington Gateway, Arlington, VA

http://idies.jhu.edu/symposium/cyberinfrastructure-ci-for-nsf-large-facilities-workshop/

Organizers

Workshop Chair:

Alex Szalay, Johns Hopkins University

Program Committee:

- ☞ Christine Borgman, University of California, Los Angeles
- ☞ Peter Couvares, Syracuse University
- ☞ Brian Glendenning, National Radio Astronomy Observatory
- ☞ Kerstin Lehnert, Columbia University
- ☞ Chuck Meertens, UNAVCO
- ☞ Manish Parashar, Rutgers University

Goal was to "create a forum for direct interaction between the NSF large facilities and CI developer community"

75 participants from over 45 institutions, large facilities, and CI resources.

Some take-homes from the workshop

Issues raised

- CI workforce – recruitment, retention, finding expertise and partners
- Dealing with data – discoverability, sharing, archiving/curation…
- Software – computational codes, analysis, visualization, integrated workflows
- Accessing best practices, guides, consultative processes (like for cybersecurity)
- Deciding when to build in-house vs. outsourcing or leveraging existing resources
- Interoperability, re-use, accessing/using external resources
- Disconnect between facility lifetime (long) and external CI lifecycles (shorter)
- User needs for analysis tools, computational resources and other services.
- New trends – e.g. streaming data to advance computing

Next steps?

- Additional workshops – exploring CI needs/challenges in specific areas
- Incentivizing community building, partnerships, collaboration and exchange, among facilities and with CI projects
- Identifying new ways of leveraging existing shared CI resources and capabilities

CTSC

CENTER FOR TRUSTWORTHY SCIENTIFIC CYBERINFRASTRUCTURE

The NSF Cybersecurity Center of Excellence

Home

About CTSC   +

Getting Help From CTSC

2016 NSF Cybersecurity Summit
for Large Facilities and Cyberinfrastructure

Theme: Strengthening Trustworthy Science

- Complete **Online Registration**
- View the **Call for Participation**
- View the **Student Program**
- View Information from the **2013**, **2014** and **2015** Summits

When: August 16 through August 18, 2016

Where: **The Westin Arlington Gateway** near NSF headquarters. A group rate is available for lodging until July 29, 2016. Hotel reservations may be **made online.**

http://trustedci.org/2016summit/