Topic:	Cyberinfrastructure for Facilities
Speaker(s) Name and Title:	Manish Parashar, CISE Office Director Ivan Rodero, Rutgers, Rutgers Discovery Informatics Institute (RDI^2)
Scribe Name:	
Session Description:	Session covered (1) updates on the NSF Office of Advanced Cyberinfrastructure, (2) findings and reports from the 2017 Workshop on Cyberinfrastructure (CI) for Large Facilities, and (3) defining expectations/best practices for CI in large facility projects.
Session Time Slot:	Tuesday May 1, 2018 at 3 pm
Purpose and Desired Outcome:	Participants develop an understanding of future challenges and strategies for addressing these challenges.

Notes & Key Points	*Capture best practices, actionable recommendations, and decisions.*

Disclaimer: *These are raw notes that were captured by the assigned scribe during this session at the 2018 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:

Manish Parashar - CISE Office Director

- FY18 Budget is 2nd largest increase in NSF research budget in 15 years. FY19 Request Budget is flat with respect to FY17.
- NSF Big Ideas have cross-cutting big cyberinfrastructure challenges.
- NSCI, National Strategic Computing Initiative – multi-agency effort to maximize the benefits of High Performance Computing
- Request for Information (RFI) – Building on Community Input:
    - Growing need for on-demand computing for new ways to process data
    - Big Data – discoverability, accessibility, and reproducibility
    - Processing and integrating data
    - Secure access, dynamic and high bandwidth workflows
    - Reliable & Sustainable software
    - Workforce development
- Many of these are cross-cutting the issues raised during last September's cyber-infrastructure workshop. Discussed below.

Data from large science facilities are essential part of science. There are many different data systems, like islands. The science is requiring more integrated science data. Data is far away from where you want to process the data.

Wholistic cyberinfrastructure ecosystem going beyond hosting the science. Need the integration to be more homogenous.

Large Facilities Challenges:
- Incorporation of technology in the "waterfall" model for facility development
- Data lifecycle management
- Interoperability across data/compute islands
- On-demand data processing
- Enhanced data delivery – open access but data is too massive for the typical user to be able to download and process.
- Data provenance, reproducibility
- CyberSecurity – The NSF Cybersecurity Center of Excellence (trustedci.org)

  - Large Facilities are advancing science but collectively they can transform science. Challenges, combined with the MREFC process, with technology advances and sustainability of cyberinfrastructure.
  - Cyberinfrastructure (CI) is central to facilities and their success. It is essential to build the right CI and build the CI right.

<u>Ivan Rodero – Rutgers Discovery Informatics Institute (RDI^2)</u>
Motivations of NSF Large Facilities CI Workshop:
- CI critical component of NSF facilities and is growing in scale and complexity
- Need to integrate data across multiple facilities
- Need to provide and sustain CI to meet current and future needs

Workshop Goals – detailed on Slide 4. Primarily to identify commonalities in CI practices and generate recommendations for current and future NSF-CI related programs.

Workshop had three phases.
- In preparation, there were white papers and questionnaires. Slides 6 through 16.

- During the Workshop, as part of the panels and breakouts, key findings and recommended actions were identified. Slides 17-22.
  - Panel 1: Facilities typically address CI challenges independently of each other. Facilities can benefit from a trusted forum.to host institution
  - Panel 2: Workforce Development. Poor mission alignment to host institution HR policies. Increased intra-facilities communications.
  - Panel 3: Sharing "best practices" exist across the facilities. Common location of information.
  - Panel 4 – Sustaining Facilities CI: long-term commitment, need support to build the community.
  - Recommended Actions from the Workshop are on Slide 22.

- Post Workshop included a survey, report, and study of white papers. Survey resulted on Slides 24-26.

Challenges – Looking Forward, Slide 37

"Take Home" Messages:
- Need to establish facilities CI community
- Need for re-using existing ACI investment
- Need for "CI best practices and trusted entity

Best Practices & Lessons Learned from Pre-Workshop Survey (Slides 9 & 10)**:**
- Use Systems engineering to manage CI lifecycle and interfaces
- Bake in redundancy to provide high availability
- Lessons learned – ability to trace CI features to requirements and business needed.