



Morning Session Cyberinfrastructure at Major Facilities

11:00am – 12:30pm

LARGE FACILITIES WORKSHOP



2021 WEBINAR SERIES

OCTOBER 14, 2021



National
Science
Foundation



Morning Session Cyber Infrastructure at Major Facilities

11:00am – 12:30pm

Panel Discussion on Cyber Infrastructure challenges and opportunities for facilities in the 21st century and approaches to cyber infrastructure and data management.

- **Panel includes**

- Moderator: **Manish Parashar**, Director, Office of Advanced Cyberinfrastructure, NSF
- **David Halstead**, Chief Information Officer at National Radio Astronomy Observatory, Federally Funded Research and Development Center (FFRDC)
- **Inder Monga**, Director for Scientific Networking Division, LBNL and Executive Director of ESN
- **Ewa Deelman**, Director of CI-Compass, NSF CI Center of Excellence
- **Ilya Baldin**, PI and Project Director for FABRIC, NSF Mid-Scale Research Infrastructure project
- **Nicola Ferrier**, Deputy Director of Sage, NSF Mid-Scale Research Infrastructure project
- **Chad Trabant**, Deputy Director of Data Services, Incorporated Research Institutions for Seismology



Cyberinfrastructure Panel background

Moderator: Manish Parashar, Director, Office of Advanced Cyberinfrastructure, NSF



- The Office of Advanced Cyberinfrastructure (OAC) supports and coordinates the development, acquisition, and provision of state-of-the-art cyberinfrastructure resources, tools and services essential to the advancement and transformation of science and engineering.
- OAC also supports forward-looking research and education to expand the future capabilities of cyberinfrastructure specific to science and engineering. By fostering a vibrant ecosystem of technologies and a skilled workforce of developers, researchers, staff and users, OAC serves the growing community of scientists and engineers, across all disciplines, whose work relies on the power of an advanced research cyberinfrastructure.



Morning Session Cyberinfrastructure at Major Facilities

11:00am – 12:30pm

Panel Discussion on Cyberinfrastructure challenges and opportunities for facilities in the 21st century and approaches to cyberinfrastructure and data management.

- Cyberinfrastructure (CI), comprising computing, data, software, and networking and related expertise are critical components of most NSF major facilities and essential to their operations, equitable use, and overall impact. Architecting, deploying and operating a robust, secure, performant and scalable CI can be challenging, especially considering new security challenges/threats, new and changing science needs, and rapidly evolving (and often disruptive) technology landscapes. Furthermore, emerging technologies can enable novel approaches for operating and using facilities.
- This session will include a panel and audience discussion on these CI challenges and opportunities for NSF facilities in the 21st century.
 - The discussions will explore existing best practices, approaches for developing shared (and sustainable) CI services and for leveraging national CI services, as well as opportunities for transforming facilities' science through CI innovations.



- David Halstead is the CIO for the National Radio Astronomy Observatory, a facility of the NSF operated under cooperative agreement by AUI, where his responsibilities are divided between Data Management for the Observatory's HPC infrastructure in support of the national radio telescopes, and the general IT support for 500+ employees.
- He has served on multiple SuperComputing Conference committees and is a founding member of the ACM's SIGHPC Education Chapter. Prior to joining NRAO, he worked in the DOE Scalable Computing Laboratory in Ames Lab, and in the private sector with Celera Genomics.



Very Large Array

ALMA



The Atacama Large Millimeter/submillimeter Array. Teams from North America, East Asia, and Europe merged projects to develop this breakthrough scientific instrument in northern Chile.





- Indermohan (Inder) S. Monga serves as the Division Director for Scientific Networking Division, Lawrence Berkeley National Lab and Executive Director of Energy Sciences Network, a high-performance network user facility optimized for large-scale science, interconnecting the National Laboratory System in the United States.
- His efforts are directed towards advancing the science of networking for collaborative and distributed research applications, as well as contributing to ongoing research projects tackling programmability, analytics and quality of experience driving convergence between application layer and the network.



- ESnet is the Department of Energy's dedicated science network, helping researchers meet their goals from experiment to discovery.
- ESnet's Mission is to enable and accelerate scientific discovery by delivering unparalleled network infrastructure, capabilities, and tools.



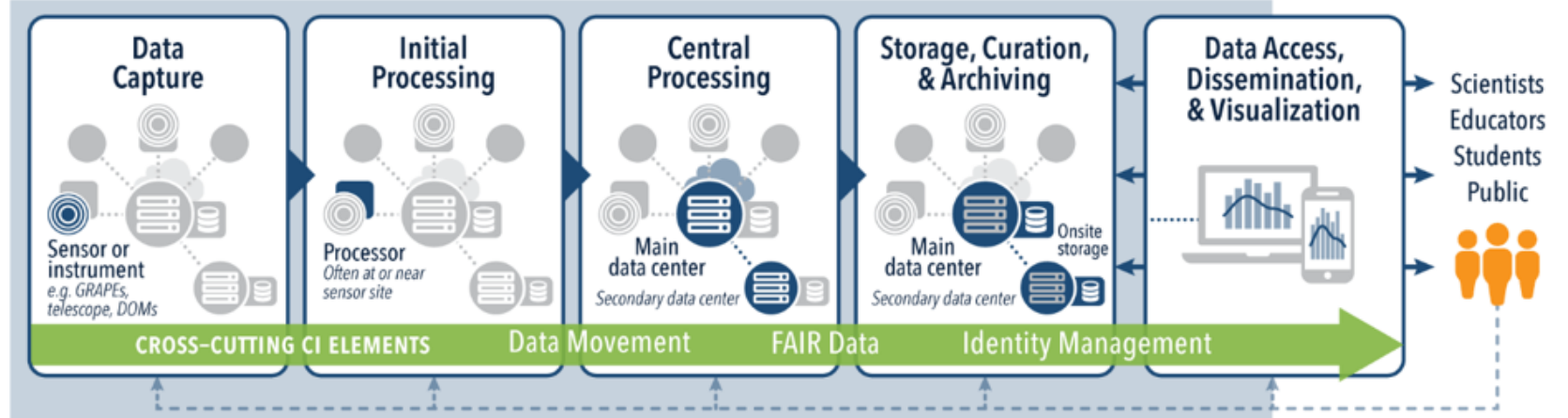
Ewa Deelman, Director of CI-Compass, NSF CI Center of Excellence



- Ewa Deelman is Principal Investigator for CI Compass, which provides leadership, expertise, and active support to CI practitioners at NSF Major Facilities and throughout the research ecosystem in order to enable ongoing evolution of our technologies, our practices, and our field, ensuring the integrity and effectiveness of the CI upon which research and discovery depend.
- CI Compass helps MFs navigate their data lifecycle (DLC). The DLC identifies specific stages of the data flow and captures the transformation of raw data into more interoperable and integration-ready data products that can be visualized, disseminated, and transformed into insights and knowledge.



ci-compass.org/





- Ilya Baldin is the PI and Project Director for FABRIC, a unique national research infrastructure to enable cutting-edge and exploratory research at-scale in networking, cybersecurity, distributed computing and storage systems, machine learning, and science applications.
- FABRIC is an everywhere programmable nationwide instrument comprised of novel extensible network elements equipped with large amounts of compute and storage, interconnected by high speed, dedicated optical links.



- FABRIC will connect a number of specialized testbeds (5G/IoT PAWR, NSF Clouds) and high-performance computing facilities to create a rich fabric for a wide variety of experimental activities.
- FABRIC Across Borders (FAB) extends the network to nodes in Asia and Europe for expanded scientific impact.



Nicola Ferrier, Deputy Director of Sage, NSF Mid-Scale Research Infrastructure project

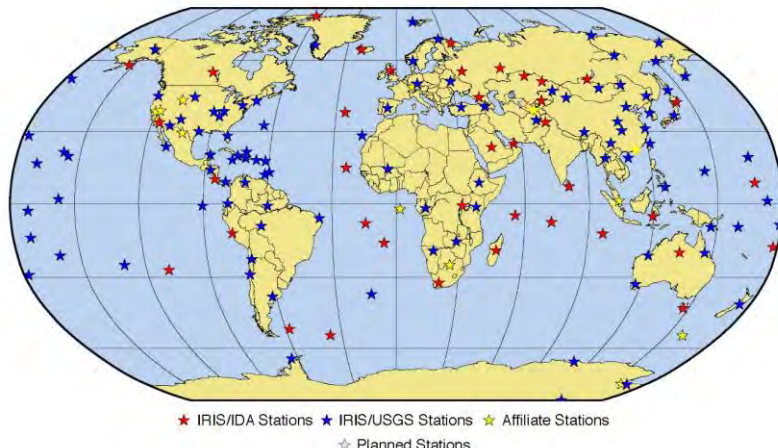


- Nicola Ferrier is Deputy Director of Sage. The goal of Sage is building a continent-spanning network of smart sensors.
- With Sage, advanced machine learning algorithms will be moved to “the edge.” Edge computing is a way to streamline data flowing from Internet of Things (IoT) devices by providing data analysis very near the site where the data is gathered.
- By linking small, powerful, computers directly to high-resolution cameras, air quality and weather sensors, and experimental Light Detection and Ranging (LIDAR) systems weather sensors, this new distributed infrastructure will enable researchers to analyze and respond to data almost instantly.
- From early detection of wildfire smoke plumes in California to identifying ultrasonic calls of bats or the patterns of pedestrians in a busy crosswalk, Sage’s artificial intelligence-enabled sensors will give scientists a new tool to understand our planet.
- The same cyberinfrastructure can be used at large facilities



- Chad Trabant is Deputy Director & Chief Architect of Data Services, Incorporated Research Institutions for Seismology (IRIS).
- IRIS is a consortium of over 100 US universities dedicated to the operation of science facilities for seismological data, contributing to scholarly research, education, earthquake hazard mitigation, and verification of the Comprehensive Nuclear-Test-Ban Treaty.
- IRIS operates Seismological Facilities for the Advancement of Geoscience (SAGE) facility on behalf of the NSF.

Global Seismographic Network



- The IRIS Data Management Center (DMC) is the largest facility in the world for the archiving, curation and distribution of seismological data for use by the research community.
- Currently, the DMC archive holds over 750+ terabytes of data. It receives data in real-time from ~3800 stations across the U.S. and around the world and manages data from over 80,000 seismic stations. Each year nearly 1 petabyte of data is distributed to users in more than 150 different countries.



Morning Session Cyberinfrastructure at Major Facilities

11:00am – 12:30pm

Panel Discussion on Cyberinfrastructure challenges and opportunities for facilities in the 21st century and approaches to cyberinfrastructure and data management.

Questions bank for the panelists

- 1. What are the key cyberinfrastructure challenges for facilities in the 21st century?**
- 2. What opportunities does the broader CI ecosystem of resources, services and expertise present to facilities and how can facilities best leverage them?**
- 3. What are the emerging use cases and user requirements and how do they impact cyberinfrastructure?**
- 4. How is the changing technology landscapes, such as the increasing use of AI/ML, impacting facilities? What new opportunities and challenges does it present?**



Morning Session Cyberinfrastructure at Major Facilities

11:00am – 12:30pm

Panel Discussion on Cyberinfrastructure challenges and opportunities for facilities in the 21st century and approaches to cyberinfrastructure and data management.

- 5. Ensuring equitable access to the services and data provided by facilities is extremely important to NSF. How are facilities ensuring that the entire science community has equitable access? What are associated challenges?**
- 6. Access to a skilled cyberinfrastructure workforce remains a crosscutting challenge. How can facilities attract, train, nurture and retain its workforce of cyberinfrastructure professionals?**
- 7. The dramatic increases in the availability of data across all science is presenting new challenges for managing, processing and storing this data. What are the data-related challenges faced by the facilities? What are possible solutions? Would it be necessary to discard data and how would this decision be made?**



Morning Session Cyberinfrastructure at Major Facilities

11:00am – 12:30pm

Panel Discussion on Cyberinfrastructure challenges and opportunities for facilities in the 21st century and approaches to cyberinfrastructure and data management.

- 8. What are the barriers to combining services and data provided by multiple facilities as part of science workflows? What new opportunities does such and integration provide for facilities science?**



Morning Session Cyberinfrastructure at Major Facilities

11:00am – 12:30pm

Panel Discussion on Cyberinfrastructure challenges and opportunities for facilities in the 21st century and approaches to cyberinfrastructure and data management.

**Many thanks to all of today's
participants**

If you have a question that cannot be answered during this panel discussion, please email it to LFWorkshop@nsf.gov.

Reminder that we will have a webinar session on Cybersecurity at Major Facilities this afternoon starting at 2pm