

# APS Scientific Computation Seminar Series

- Speaker:** Valerio Pascucci  
John R. Parks Endowed Chair, University of Utah  
Professor, School of Computing  
Faculty, Scientific Computing and Imaging Institute  
Director, Center for Extreme Data Management Analysis and Visualization (CEDMAV)
- Title:** The National Science Data Fabric: Democratizing Data Access for Science and Society
- Date:** Monday, November 29, 2021
- Time:** 1:00 p.m. (Central Time)
- Location:** Microsoft Teams meeting  
**Join on your computer or mobile app**  
[Click here to join the meeting](#)  
**Or call in (audio only)**  
[+1 630-556-7958,707344060#](#) United States, Big Rock  
Phone Conference ID: 707 344 060#  
[Find a local number](#) | [Reset PIN](#)  
[Learn More](#) | [Meeting options](#)
- Hosts:** Mathew Cherukara and Nicholas Schwarz
- Abstract:** Effective use of data management techniques for the analysis and visualization of massive scientific data is a crucial ingredient for the success of any experimental facility, supercomputing center, or cyberinfrastructure that supports data-intensive scientific investigations. Data movements have become a central component that can enable or stifle innovation in the progress towards high-resolution experimental data acquisition (e.g., APS, SLAC, NSLS II). However, universal data delivery remains elusive, limiting the scientific impacts of these facilities. This is particularly true for high-volume/high-velocity datasets and resource-constrained institutions. This talk will present the National Science Data Fabric (NSDF) testbed, which introduces a novel trans-disciplinary data fabric integrating access to and use of shared storage, networking, computing, and educational resources. The NSDF technology addresses the key data management challenges involved in constructing complex streaming workflows that take advantage of any data processing opportunities that arise while the data is in motion. This technology finds practical use in many research and industrial applications, including materials science, precision agriculture, ecology, and telemedicine. This NSDF overview will include several techniques that allow building a scalable data movement infrastructure for fast I/O while organizing the data in a way that makes it immediately accessible for analytics and visualization. For example, I will present a use case for the real-time data acquisition from an APS beamline to allow remote users to monitor the progress of an experiment. We accomplish this with an ephemeral NSDF installation that can be instantiated via Docker or Singularity at the beginning of the experiment and removed right after. In general, the advanced use of containerized applications with automated deployment and scaling makes the practical use of clients, servers, and data repositories straightforward in practice, even for non-expert users. Full integration with Python scripting facilitates the use of external libraries for data processing. The scan of a 3D metallic foam can be easily distributed with the following Jupyter notebook <https://bit.ly/NSDF-example01>. Overall, this leads to a flexible data streaming workflow that allows working with massive imaging models without compromising the interactive nature of the exploratory process, the most effective characteristic of discovery activities in science and engineering. The presentation will be combined with a few live demonstrations of the technology.