



Contribution ID: 155

Type: **Keynote Talk**

## **KEYNOTE 6: Scientific Data Democratization: Enabling Efficient Access and Analysis of Large-Scale Scientific Data**

*Thursday, 7 December 2023 09:45 (45 minutes)*

The vast amounts of data generated by scientific research hold immense potential for advancing knowledge and discovery. However, the complexity and sheer volume of this data often pose significant challenges in terms of accessibility, analysis, and interpretation. Scientific data democratization aims to address these challenges by enabling researchers to easily access, analyze, and share scientific data, regardless of their technical expertise or the location of the data. My group along with our close collaborators have worked with hundreds of scientific communities in the past 20 years, in many disciplines ranging from astronomy, fusion, combustion, seismology, weather, climate, accelerator science, material science to clinical pathology. In our partnerships we have created sustainable software components which help address the following needs 1) Creating a self-describing I/O framework which allows data to be read/written at terabytes/sec, 2) Having the ability to query PBs of data efficiently even for derived quantities which are NOT contained in the data, 3) Having the ability to subscribe to data (in memory) without modifying the codes such that I/O is abstracting from data-at-rest to data-in-motion, 4) Creating new mathematical formulations which allows data to be reduced in both the size and in the degrees of freedom to allow for faster access, and 5) having the ability to work with federated data, as if it was local. In this presentation, we will explore the challenges and opportunities associated with scientific data democratization, along with some of the work we have done in these fields.

### **Student or Postdoc?**

No. Not a student nor Postdoc.

**Primary author:** Dr KLASKY, Scott (University of Tennessee at Knoxville, Georgia Institute of Technology, Oak Ridge National Laboratory)

**Presenter:** Dr KLASKY, Scott (University of Tennessee at Knoxville, Georgia Institute of Technology, Oak Ridge National Laboratory)

**Session Classification:** Keynote

**Track Classification:** DIRISA