Title: Accessing Multi-TB-sized Datasets: RDA

The NCAR Research Data Archive (RDA, http://rda.ucar.edu) maintains a variety of data discovery and access capabilities as part of it’s curation of 600+ datasets. In-house developed and standards-based community tools offer services to more than 10,000 users annually. By number of users the largest group is the worldwide research community that access the RDA through web based protocols; the NCAR HPC users are another important group. They are fewer in number, but typically access more data volume directly from a high performance central file disk system. This talk will detail the data discovery and access services maintained by the RDA to support both user groups, show metrics that illustrate how the community is using the services, and highlight how HPC computing at NCAR is addressing access to TB-sized data collections.

Bio:

Doug Schuster is a meteorologist, database engineer, and technical lead with the Data Support Section at NCAR which manages the CISL Research Data Archive (RDA). He started his education in Civil Engineering at the University of Minnesota, and pivoted into the Atmospheric Sciences with graduate study and research at Colorado State University (graduating in 2001) focusing on operational mesoscale modeling. In 2003 Doug joined the team of software engineers that support the RDA. This marked a transition into large scale data management to support weather and climate research. His goals cover all phases of the data lifecycle including complete and well documented data ingest into the RDA, implementing advanced and multiple methods to support data discovery, providing numerous data access options suitable for a diverse user community, offering high-performance data interoperability directly into data analysis tools, maintaining consulting and two-way communications with users, and continuously seeking system improvements to meet users evolving expectations.