The Big Data Repository at the IRIS Data Management Center: Developing Standards for Services and Federated Access for Globally Distributed Data Centers

*Timothy Keith Ahern[1], Chad Trabant[1], Mick Van Fossen[1], Rob Casey[1]

1. Incorporated Research Institutions for Seismology

The IRIS DMC is the largest repository of seismological data from permanent observatories in the world. The archive at IRIS is nearly 1/2 petabyte in size and distributes nearly 1 petabyte per year to the scientific and monitoring communities. IRIS works closely with the International Federation of Digital Seismograph Networks (FDSN) in coordinating standards for data exchange formats, metadata descriptions, standardized web services, standardization of request parameters and federated concepts for data centers.

IRIS currently operates a large analytics engine that automatically assesses quality of data from seismic observatories. This "Big Data" problem has resulted in a powerful system that will enable researchers to receive data that have been filtered by their quality in addition to the specification of space-time constraints. As the volume of data increases in the expanding federated system of seismological data centers, IRIS is also developing a system called Research Ready Data Sets (RRDS) that will allow a users data request to be filtered by their quality as measured by the quality assurance system at the DMC that is now fully operational.

IRIS is currently testing the concept of operating its data center infrastructure in both High Performance Computing (HPC) as well as Cloud Computing environments with the ultimate goal of this project to identify an appropriate environment in which to run a data center.

This presentation will describe several aspects of current IRIS DMC systems, how web services have totally changed our ability to service huge numbers of user requests, how have adopted new access mechanisms, and how quality assurance systems are now being used to improve seismic network performance as well new tools being developed that will make scientific researchers more efficient in conducting their research.

Keywords: Seismology, Web Services, Federation, Quality Assurance