Andrew Michaelis
NCTS# 25415-17 AGU Fall Meeting 2016

**Using Cloud-based Storage Technologies for Earth Science Data**

Cloud based infrastructure may offer several key benefits of scalability, built in redundancy and reduced total cost of ownership as compared with a traditional data center approach. However, most of the tools and software systems developed for NASA data repositories were not developed with a cloud based infrastructure in mind and do not fully take advantage of commonly available cloud-based technologies. Object storage services are provided through all the leading public (Amazon Web Service, Microsoft Azure, Google Cloud, etc.) and private (Open Stack) clouds, and may provide a more cost-effective means of storing large data collections online. We describe a system that utilizes object storage rather than traditional file system based storage to vend earth science data. The system described is not only cost effective, but shows superior performance for running many different analytics tasks in the cloud. To enable compatibility with existing tools and applications, we outline client libraries that are API compatible with existing libraries for HDF5 and NetCDF4. Performance of the system is demonstrated using clouds services running on Amazon Web Services.

**Authors**

Andrew Michaelis
andrew.r.michaelis@nasa.gov
- California State University Monterey Bay

John Readey
jreadey@hdfgroup.org
- HDF Group

Petr Votava
petr.votava-1@nasa.gov
- California State University Monterey Bay