



Creating Value through Open Data, a Perspective from AWS

Conor Delaney and Jed Sundwall

Public Data Sets Program of Amazon Web Services

Providing open access to data encourages innovation by making data available to everyone, including stakeholders that may not have traditionally had access to data, such as students, startups, SMEs, and enterprises. Recently, some national meteorological services have been providing open access to datasets that traditionally required payment to access. However, the infrastructure usually used to share data is not keeping up with growth in data volumes. Data is still being provided with the assumption that users will download and store their own copies of data. Requiring users to download and store copies of large meteorological datasets requires time and expense that can limit data access and impede research.

The AWS Public Datasets program, through collaborations with NOAA, the UK Met Office, and the Finnish Meteorological Institute, has gained an overview of the challenge of improving access to large meteorological data sets. To encourage fresh approaches from the community AWS has provided credits for cloud computing resources to researchers who are investigating ways to expand access to scientific data in the cloud. These community led investigations have produced a number of different approaches.

In this presentation, we discuss some of the challenges that have been encountered by Earth science communities in moving data to the cloud. We then describes some techniques that AWS customers are using to improve access to NetCDF and GRIB files in a cloud computing environment. Finally, we discuss the idea of undifferentiated heavy lifting from the perspective of serving large scientific data sets to new and larger communities.