



Toward a Vision for Enabling Science as a Service to Facilitate Open Science and Reproducible Research

Mohan Ramamurthy and Julien Chastang

University Corporation for Atmospheric Research, Unidata, Boulder, United States (mohan@ucar.edu)

The geoscience disciplines are either gathering or generating data in ever-increasing volumes. To ensure that the science community and society reap the greatest benefits in research and societal applications from such rich and diverse data resources, there is a growing interest in broad-scale, open data sharing to foster myriad scientific endeavors. However, open access to data is not sufficient, research outputs must be reusable and reproducible to accelerate scientific discovery and catalyze innovation.

As part of its mission, Unidata, a cyberinfrastructure facility, has in recent years been developing and deploying data infrastructure and data-proximate scientific workflows and analysis tools using cloud computing technologies for accessing, analyzing, and visualizing geoscience data.

Specifically, Unidata has developed techniques that combine robust access to well-documented datasets with easy-to-use tools, using workflow technologies. In addition to fostering the adoption of technologies like pre-configured virtual machines through Docker containers and Jupyter notebooks, other computational and analytic methods are enabled via “Software as a Service” and “Data as a Service” techniques with the deployment of the Cloud IDV, AWIPS Servers, and the THREDDS Data Server in the cloud. The collective impact of these services and tools is to enable scientists to use the Unidata Science Gateway capabilities to not only conduct their research but also share and collaborate with other researchers and advance the intertwined goals of Reproducibility of Science and Open Science, and in the process, truly enabling “Science as a Service”.

Unidata has implemented the aforementioned services on the Unidata Science Gateway (<http://science-gateway.unidata.ucar.edu>), which is hosted on the Jetstream cloud, a cloud-computing facility that is funded by the U. S. National Science Foundation. The aim is to give geoscientists an ecosystem that includes data, tools, models, workflows, and workspaces for collaboration and sharing of resources.

In this presentation, we will present our work to date in developing the Unidata Science Gateway and the hosted services therein, as well as our future directions to advance data and software transparency, open science, and reproducible research. We will share our experiences in how the geoscience and information science communities are using the data, tools and services provided through the Unidata Science Gateway to advance research and education.