Unidata Science Gateway for Enabling Science as a Service to Facilitate Open Science and Reproducible Research

M Ramamurthy
Unidata, University Corporation for Atmospheric Research, Boudier, CO, USA
E-mail: mohan@ucar.edu

Abstract: Unidata, an NSF-funded geoscience cyberinfrastructure facility, has deployed data infrastructure and data-proximate scientific workflows and analysis tools, using cloud computing technologies, for analyzing and visualizing well-documented datasets that combine robust access to well-documented datasets. Docker containers and Jupyter notebooks, and other analytic methods are enabled via “Software as a Service” and “Data as a Service”. The collective impact of these services is to enable scientists to not only conduct their research but also share their work with other researchers.

1. Introduction
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.

The Unidata Program Center (http://www.unidata.ucar.edu/), a cyberinfrastructure facility, is sponsored primarily by the National Science Foundation. Several hundred institutions worldwide participate in the Unidata data sharing network and many more institutions use Unidata tools and technologies in education, research, and operations.

As the enabler of a broad community, Unidata

- Acquires and distributes data to facilitate Earth System Science and education
- Develops software for accessing, managing, analyzing, visualizing, and effectively using those data
- Provides comprehensive support to users
- Conducts annual training workshops on Unidata software packages
- Facilitates advancement of standards, conventions, and interoperability
- Provides leadership in cyberinfrastructure and fosters technological change
- Assesses and responds to community needs
- Advocates on behalf of the community on data issues and negotiates data agreements
- Fosters community interaction to promote sharing of data, tools, and ideas
- Grants equipment awards to universities to enable and enhance participation

2. Science as a Service
In recent years, Unidata has 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”.

3. Unidata Science Gateway

To achieve the above goals, Unidata is developing data-driven scientific workflows using cloud computing technologies [1] for accessing, analyzing, and visualizing geoscience data. The Unidata Science Gateway [2] (http://science-gateway.unidata.ucar.edu) has implemented Unidata-related technologies hosted on Jetstream (https://jetstream-cloud.org), a cloud facility 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.

Through the Science Gateway, researchers can make use of well-integrated resources either directly in their browser or using one of the client applications. They include

- Access to large volumes of meteorological data, via OPeNDAP, ADDE, WCS, WMS, netCDF APIs and protocols, including access to subsets of data
- Data transformation and format conversion
- Extensive data analysis capabilities
- Visualization of data provided by the gateway
- Access to a collection of Jupyter Notebooks for data analysis
- Publishing of products and results
- Access to Advanced Weather Information Processing System

Real-time data from over 30 meteorological data streams, in excess of 1 TB/day, flow into the Gateway data servers via the Local Data Manager software. Those data include radar, satellite, surface, upper-air, ship, aircraft and other observations as well as forecast model output from operational weather prediction centers. Our hope is that by combining the resources available in this gateway – and coupling them with XSEDE (https://www.xsede.org) HPC resources – community members can create end-to-end workflows in the Jetstream Cloud.

The Unidata Science Gateway and the hosted services therein, as well as our future directions to advance data and software transparency, are enabling open science and reproducible research. Our Science Gateway is also advancing our educational goals by supporting classrooms to teach data science courses at universities. Numerous software training workshops also took advantage of the resources on the Gateway for their instruction.

References

[1] Lazarus, S. M., and Coauthors, 2013: 2012 Unidata Users Workshop Navigating Earth System Science Data. Bull. Amer. Meteor. Soc., 94, ES136–ES143.