

NOAA Open Data Dissemination Overview

CPASW | May 11, 2023 | Asheville, NC

Patrick Keown
Program Manager

NODD Team
Adrienne Simonson | Kate Szura

NODD CISESS/NCSU Team
Otis Brown | Jonathan Brannock | Jenny Dissen | Denis Willett

BDP → NODD Evolution



April 2015

NOAA signs 3-year CRADA with AWS, Google, IBM, Microsoft and OCC

May 2018

CRADA Extended

FY 2022

- FY22 Omnibus provides approval of NODD
- Included in FY22-FY26 NOAA Strategic Plan
- Big Data Analytics Chapter

FY 2024

- LOs pushing data / NODD integrated :
- Metrics
 - Cloud training
 - User Engagement

Oct 2015

NEXRAD as first NOAA dataset moved under the CRADA

October 2019

BDP Initial Operational Capability

FY 2023

- FY23 Omnibus supports NODD
- Metrics portal with email log in
- Customer Software Coordination
- NODD Service Delivery & Internal Engagement

NODD Builds on Key CRADA Lessons



- ❑ Access on cloud platforms has resulted in increased usage of NOAA's data
- ❑ Integration of NOAA data into industry tools is most effective
- ❑ Key NOAA asset is **expertise** to understand and support the data
- ❑ The role of an intermediate **"Data Broker"** has emerged as a valuable function & possible Enterprise Service that could support NOAA in provision of data to the commercial cloud
- ❑ A **defined commitment and level of service** are needed by NOAA, the Collaborators, and Users



Open &
Free

NODD Disseminates NOAA Line Office Data

NOAA data is growing exponentially....



TECHNOLOGY MODERNIZATION

Reduces stress on NOAA's on-premise dissemination systems

Improves services for Users

FULL & OPEN PUBLIC ACCESS

Supports Federal Data Strategy & Evidence Act
Open Data Requirements

No egress costs

ENABLES & ENGAGES USERS

Catalyzes innovation in environmental services

Enables interoperability

- Open data with value to the public
- No use restrictions or user registration
- Appropriate metadata provided



Via Public-Private Partnerships

- ❑ **NOAA Decision** – data goes into the NOAA Allocation on the CSP, minimum of 5 PBs per CSP
- ❑ **CSP or User Request** - data is not counted against the NOAA allocations
- ❑ **Based on an LO contract for storage or processing with one of the three CSPs** - data does not count against the NOAA allocation
- ❑ **Two year base contract** with four options of two years each. *Oct 1, 2022 start of year four of ten.*
- ❑ Very low cost contract with **no outyear costs**.
- ❑ Three Cloud Service Providers (CSPs) host NOAA data in the cloud - with **no cost for public egress or data download**.

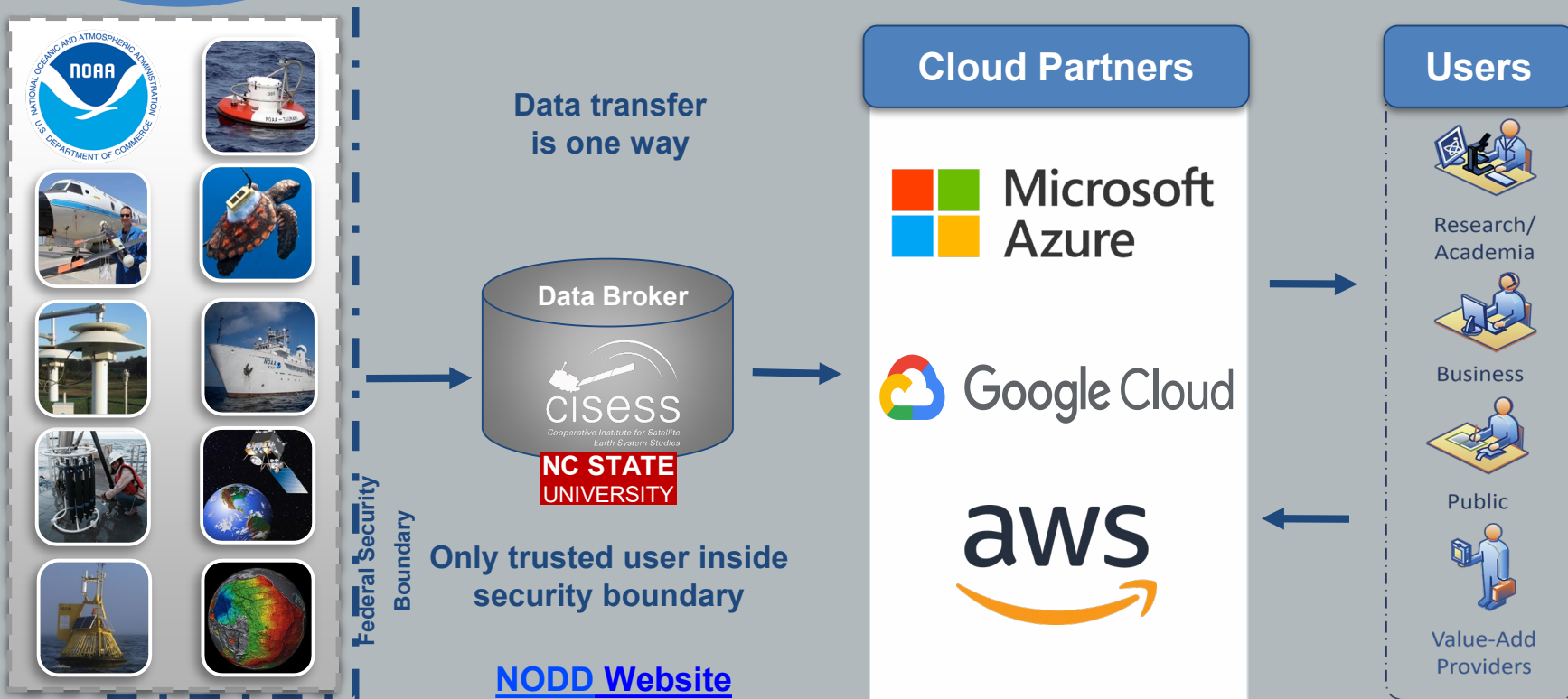
“Let’s leverage the CSPs expertise, platforms and tools.”

In order to avoid agency costs and risks associated with public access



Architecture

NODD Accelerates Access to NOAA Open Data



The NODD Data Broker Role and Activities

Data Transfer

- Event driven data flows
- Cloud Ingest, Apache NiFi
- Key Products: GOES, JPSS, GFS, GEFS, HRRR, etc.
- Leverage Cloud Native Technologies

Infrastructure Provisioning

- Storage Buckets
- Transfer Clusters
- Credentials
- On-Prem and Multi-Cloud

Monitoring

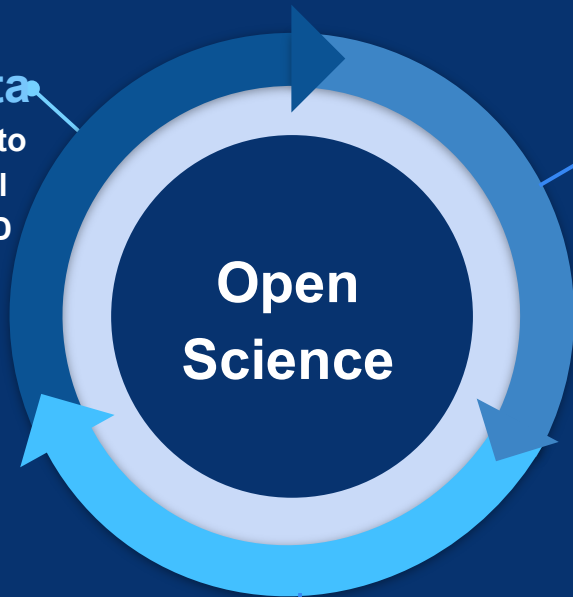
- Key Metrics
- Real Time Monitoring and Alerting of Data Flows
- Single-Pane of Glass
- Reporting

User Engagement

NODD Accelerates Open Science & Interoperability

Open Data

Open access to environmental data via NODD



Open Science

Open Reproducible Analysis

Open scientific process and reusable analytics pipelines

Open Products

Open access to research products, results, and new data for reproducibility



Public Datasets Programs



- The three cloud partners host NOAA data made available via NODD on each of their public datasets programs
 - Microsoft - [Planetary Computer](#)
 - AWS - [Registry of Open Data](#)
 - Google Cloud - [Marketplace](#)

Registry of Open Data on AWS



NOAA Coastal Lidar Data

[climate](#) [disaster response](#) [elevation](#) [geospatial](#) [lidar](#)

Description

Lidar (light detection and ranging) is a technology that can measure the 3-dimensional location of objects, including the solid earth surface. The data consists of a point cloud of the positions of solid objects that reflected a laser pulse, typically from an airborne platform. In addition to the position, each point may also be attributed by the type of object it reflected from, the intensity of the reflection, and other system dependent metadata. The NOAA Coastal Lidar Data is a collection of lidar projects from many different sources and agencies, geographically focused on the coastal areas of the United States of America. The data is provided in Entwine Point Tiles (<https://entwine.io>) format, which is a lossless streamable octree of the point cloud. Datasets are maintained in their original projects and care should be taken when merging projects. The coordinate reference system for the data is The NAD83(2011) UTM zone appropriate for the center of each data set and the orthometric datum appropriate for that area (for example, NAVD88 in the mainland United States, PRVD02 in Puerto Rico, or GUVDO3 in Guam). The geoid model used is reflected in the data set resource name.

Update Frequency

Resources on AWS

Description

NOAA Coastal Lidar Dataset

Resource type

S3 Bucket

Amazon Resource Name (ARN)

`arn:aws:s3:::noaa-nos-coastal-lidar-pds`

AWS Region

`us-east-1`

AWS CLI Access (No AWS account required)

```
aws s3 ls --no-sign-request s3://noaa-nos-coastal-lidar-pds/
```

Description

NOAA Coastal Lidar Dataset New Dataset Notification

Data Catalog

The Planetary Computer Data Catalog includes petabytes of environmental monitoring data, in consistent, analysis-ready formats. All of the datasets below can be accessed via Azure Blob Storage, and can be used by developers whether you're working within or outside of our Planetary Computer Hub.

Datasets matching "NOAA"



NOAA GFS Warm Start Conditions

Six-hourly initial conditions for the NOAA Global Forecast System

[Global](#) [NOAA](#) [Weather](#) [Temperature](#) [Precipitation](#)

Select a project

Marketplace

NOAA

Marketplace > "NOAA" > Datasets

26 results

Marketplace home

Your products

Your orders



noaa-global-forecast-system

BigQuery Public Data

The Global Forecast System (GFS) is a weather forecast model produced by the National Centers for Environmental Prediction (NCEP). The GFS dataset consists of selected model outputs (described below) as gridded forecast variables. The 384-hour forecasts, with 3-hour forecast interval, are made at 6-hour temporal resolution (i.e. updated four times daily). Use the 'creation_time' and 'forecast_time' properties to select...

Filter Type to filter

Category

Maps (1)

Big data (1)

Analytics (1)



NOAA Monthly U.S. Climate Gridded Dataset (NClimGrid)

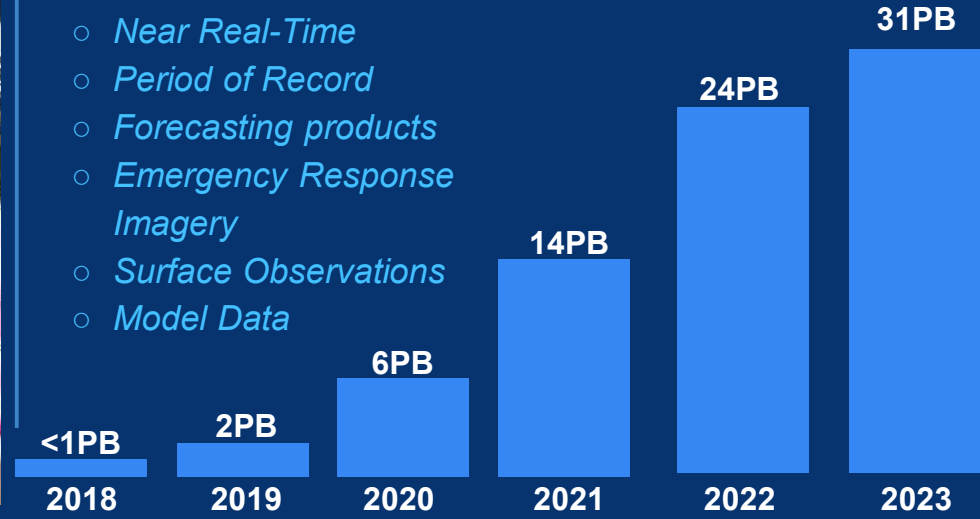
NOAA

The NOAA Monthly U.S. Climate Gridded Dataset (NClimGrid) consists of four climate variables derived from the GHCN-D dataset: maximum temperature, minimum temperature, average temperature and precipitation. Each file provides monthly values in a 5x5 lat/lon grid for the Continental United States. Data is available from 1895 to the present. On an annual basis, approximately one year of "final" NClimGrid will be...

NOAA Datasets Accessible via NODD



- Atmospheric
- Oceanic
- Fisheries
- Weather
- Climate
- Near Real-Time
- Period of Record
- Forecasting products
- Emergency Response
- Imagery
- Surface Observations
- Model Data



[NOAA.GOV/NODD](https://www.noaa.gov/nodd)



NWS Datasets on NODD



NWS Data

NWS Datasets:

- Climate Forecast System (CFS)
- Global Forecast System (GFS)
- Global Ensemble Forecast System (GEFS)
- National Blend of Models (NBM)
- National Digital Forecast Database (NDFD) [Historical]
- National Digital Forecast Database (NDFD)
- National Water Model (NWM)
- Next Generation Weather Radar (NEXRAD)
- Space Weather Prediction Center (SWPC) Forecasts
- Yesterday's Storm Reports
- Multi-Radar/Multi-Sensor (MRMS)
- Global Real-Time Ocean Forecast System (Global RTOFS)
- Rapid Refresh Forecast System (RRFS) Ensemble (Prototype)

Registry of Open Data on AWS

The Registry of Open Data on AWS is now available on AWS Data Exchange

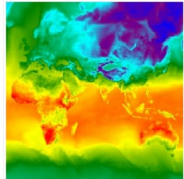
NOAA Global Forecast System (GFS)

Description
The Global Forecast System (GFS) is a weather forecast model produced by the National Centers for Environmental Prediction (NCEP). Dozens of atmospheric and land-soil variables are available through this dataset, from temperatures, winds, and precipitation to soil moisture and atmospheric ozone concentration. The entire globe is covered by the GFS at a base horizontal resolution of 16 miles (28 kilometers) between grid points, which is used by the operational forecasters who predict weather out to 16 days in the future. Horizontal resolution drops to 44 miles (70 kilometers) between grid point for forecasts between one week and two weeks.

Resources on AWS
Description: GFS data
Resource type: S3 Bucket
Amazon Resource Name (ARN): `arn:aws:s3:::noaa-gfs-bdp-pds`
AWS Region: us-east-1
AWS CLI Access (No AWS account required): `aws s3 ls --no-sign-request s3://noaa-gfs-bdp-pds/`

Earth Engine Data Catalog

GFS: Global Forecast System 384-Hour Predicted Atmosphere Data



Dataset Availability
2015-07-01T00:00:00Z–2022-09-28T06:00:00

Dataset Provider

NOAA Global Forecast System (GFS)

Notebooks and documentation for AI-for-Earth-managed datasets on Azure

[View on GitHub](#)

NOAA Global Forecast System (GFS)

Overview

The NOAA Global Forecast System (GFS) is a global numerical weather prediction system containing a



National Weather Service

RRFS Research to Operations Delivery from WCOSS to the Public Cloud

AT A GLANCE

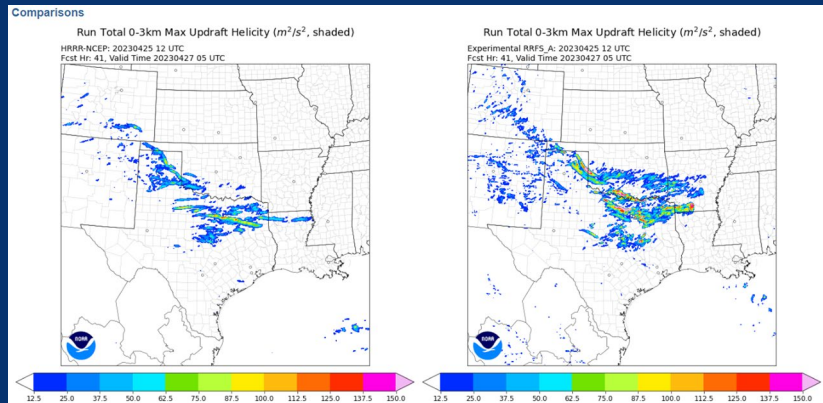
NOAA's National Weather Service (NWS) is currently developing a new data assimilation, and forecasting system to provide enhanced weather information to its customers.

HOW NODD IS UTILIZED

Rapid Refresh Forecast System data is provided to researchers, academia, and the public via NODD on AWS. Data is pushed from NOAA supercomputers directly out to the Cloud.


IMPACT PROVIDED

It is critical to the development of this model to provide prototype data to researchers and academia. Cloud access to RRFS will allow for enhanced testing and development of the model.



NODD Metrics Portal Now Available

nodd.ncics.org






NOAA Open Data Dissemination
Cloud Data Analytics Summary
Analytics Team: CISESS-NC
Version 1.0.0 | Generated: May 11, 2023

31.8 PB
Total Storage

NODD-Wide Weekly Metrics

Combined metrics from Amazon Web Services (AWS), Microsoft Azure (Azure), and Google Cloud Platform (GCP).

Week Month

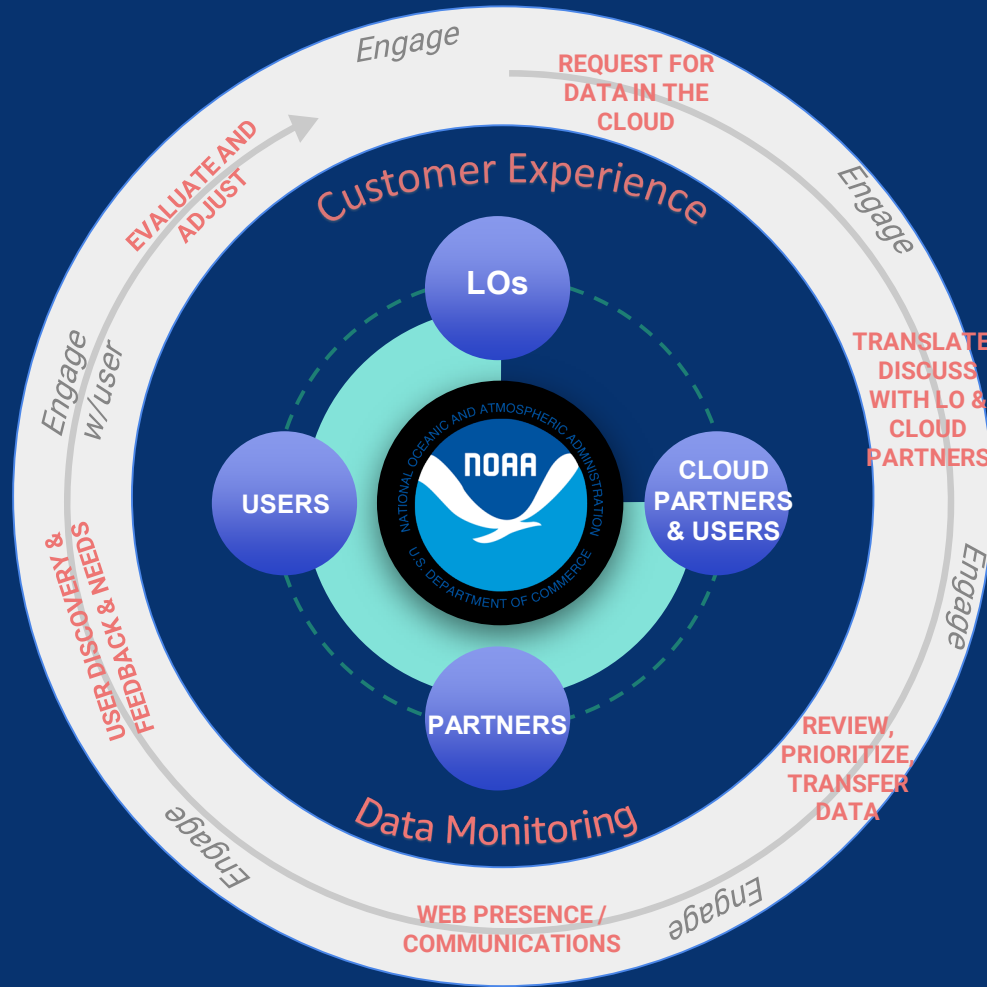
Accessions 	Interactions 	Uploads 
2.79 PB	1.25 B	432.44 TB

NODD Metrics Portal



NWS Report
National Weather Service Cloud Holdings Metrics updated daily →

NOAA SERVICE DELIVERY & CX Functionality



- Users look to NOAA for a range of data, information, tools and services...via the CLOUD
- Need support to apply NOAA's data, information, and tools
- Others want answers, guidance, training, or a helping hand
- Access and discoverability continues to be a common theme
- Continuous user engagement via variety of modes provides insights on how data products and tools are, or are not, serving specific localities or sectors

“Data have value when they are used in decision making. If not, then the economic value of such data is effectively zero.”

NWS NODD AWS Office Hours on HRRR Update



- NODD co-hosted HRRR Office Hours with the NWS Office of Organizational Excellence and AWS on March 15th
- Roughly 40 attendees from public and private sectors, and academia
- Attendees answered a couple of initial engagement polls (provided right) to gauge attendee interests

- HRRR SME Geoff Manikin and the NODD team answered user questions
- Meeting summary can be referenced [here](#)

AWS Open Data + NOAA Open Data Dissemination (NODD) Pilot Environmental Data Office Hours

MARCH 15th, 2023 | 12-1PM EDT | [REGISTER HERE](#)

- ❖ Hear about NOAA Open Data Dissemination (NODD)
- ❖ Connect with NOAA data subject matter experts
- ❖ Share your experience, case studies and lessons learned

Speakers:

- Chris Stoner, AWS Open Environmental Data Lead
- Adrienne Simonson, Patrick Keown, Jenny Disson (NODD)
- Cindy Eisenheimer, Partnership Engagement Lead, NOAA NWS Office of Organizational Excellence
- Geoff Manikin, Environmental Modeling Center NWS

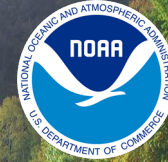
We invite your input or questions about environmental data via the cloud — in advance when you register and during the discussion!

Poll 1		
Question	Answer	Count
How do you access HRRR data today?	On-prem via NOAA	2
	Cloud	11
	Both/Either	7
	3rd party / web-based viewer	4
	Other	3

Poll 2		
Question	Answer	Count
My primary goal for attending today is:	Understand technical use and access of HRRR data	5
	To learn about the HRRR-to-RRFS transition	9
	To learn about cloud access to data (e.g. NODD Program)	7
	Meet and engage with NOAA staff scientists	3
	Learn about AWS access and tools	1

Cloud Data Analytics Faculty Fellows

- Expose faculty and students to NOAA data on NODD. Provide expertise, including connection to SMEs.
- Provide experiential cloud training to faculty and students via Microsoft access and tools.
- Allows faculty to incorporate cloud analytics as a strategy and NOAA data's broad commercial applicability as a resource in solving societal or economic problems in the classroom
- Serves as a pilot for other universities.
- 4 faculty selected with interdisciplinary backgrounds





**VISIT US
NOAA.GOV/NODD**

**Let's
Connect.
EMAIL: nodd@noaa.gov**

*Adrienne Simonson, NOAA OCIO, Director
Patrick Keown, NOAA NOS / OCIO, Program Manager
Katelyn Szura, NOAA/OCIO, Communications Lead
Otis Brown, Director, CISESS / NC State University
Jonathan Brannock, Cloud Software Engineer, CISESS / NC State University
Jenny Dissen, Engagement & Partnerships, CISESS / NC State University
Denis Willett, Cloud Software Engineer, CISESS / NC State University*