Data Expedition Platform

Integrating Multi-agency Data Products in a Cloud-based Platform for Streamlined Discovery, Visualization, and Use

AGU 2021 · IN32B-03 · December 2021

Motivation

- Earth data users often incorporate geospatial Earth observations provided by both NASA and NOAA
- Typically, these users must complete search, discovery, and access/download of data from each federal agency via largely disparate workflows
- Growing need to facilitate search, discovery, and scientific use of multi-agency geospatial observations ***via a single platform***

Motivation

NOAA & NASA have formed a working group to address cloud architectures, data access and discovery challenges, and best practices via joint projects.

*This group has collaborated on a **cloud-based data expedition pilot, demonstrating how a single platform with inputs from both NASA and NOAA can answer a single question.***

Objectives:

- Identify and explore opportunities for collaboration through sharing of information, lessons learned, and future initiatives, focusing on developing data governance (provenance) guidelines and common practices through the data life cycle in the cloud.
- Establish proof-of-concept for integrating data products from both NASA and NOAA agencies within a single, cloud-oriented platform

Goals

Project Use Case:

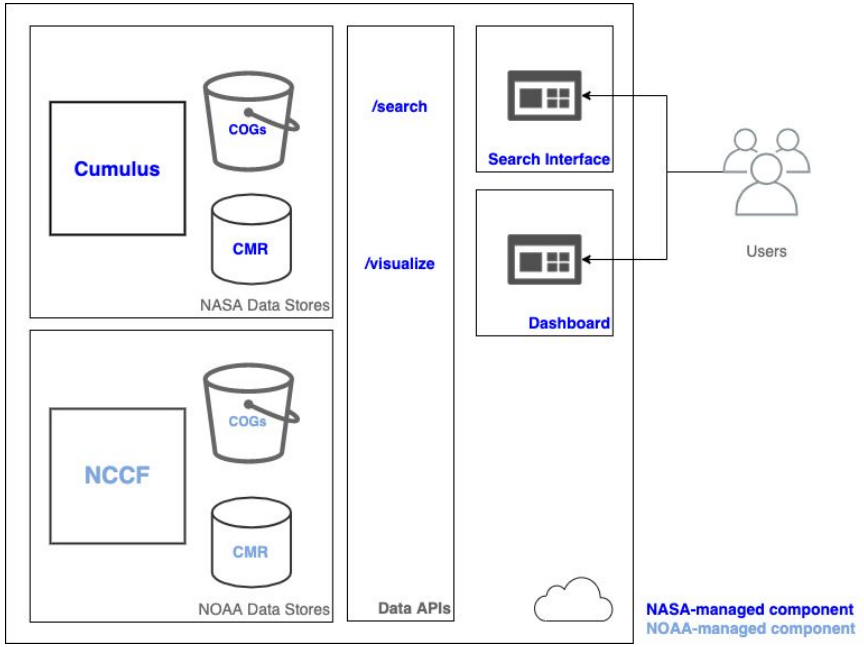
Data products from both agencies are used to assess fire weather and burn scars from a major fire.

Key Components:

- Cross-agency search API:
  Single search API, using the SpatioTemporal Asset Catalog (STAC) metadata standard, allows data discovery from 2 distinct instances of NASA's Common Metadata Repository (CMR), individual cloud storage frameworks (Cumulus and NESDIS Common Cloud Framework (NCCF))
- Metadata curation:
  Designed and generated for products of interest for fire weather
- Cloud-optimized data formats for dynamic visualization:
  Data products transformed to cloud-optimized formats (eg: Cloud-Optimized GeoTiffs (COGs)), facilitating backend for dynamic map visualizations
- Interfaces for search and discovery

Architecture - Current

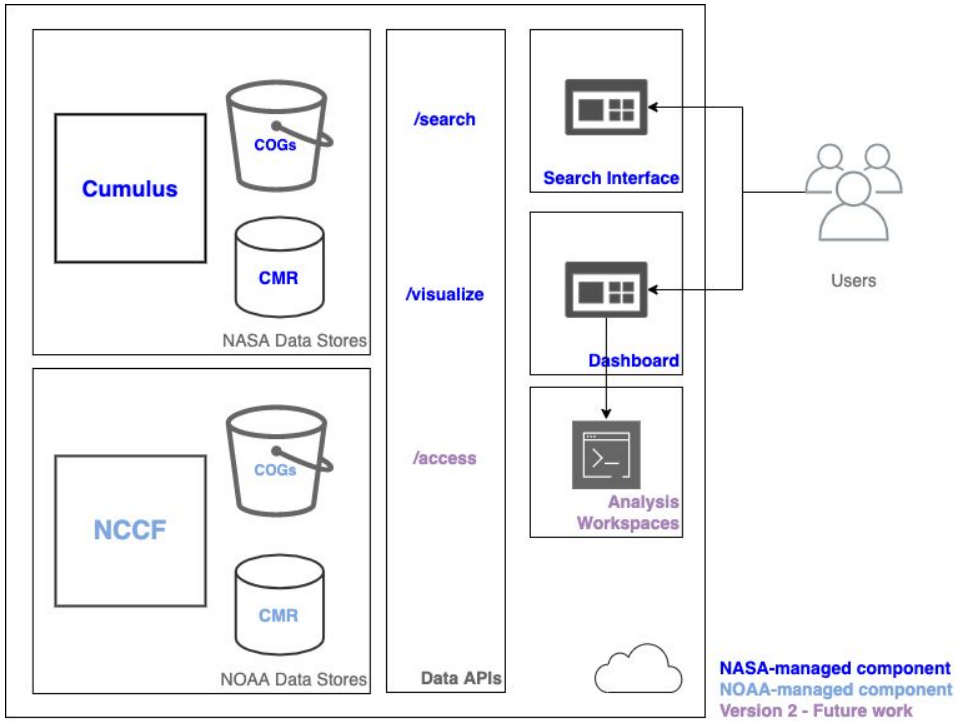
- Separate Cumulus and NCCF framework
  - S3 buckets, Cloud Optimized GeoTIFFs (COGs)
- 2 CMR instances
  - Separate per agency
- Joint capabilities via API:
  - Visualization
  - Search



Architecture - Future

Future work:

- Users may require more robust search and discovery, such as faceted search of granules agencies
- Users want to migrate from search and dashboard into analysis workspaces.
- Users need cloud-optimized access APIs via cloud-optimized formats and APIs for efficient subsetting and analysis.



Target Data Products

Format Conversion:

- Data transformations to cloud optimized formats: Zarr, Cloud Optimized GeoTIFF (COG)
- These transformations support performant analysis and dynamic visualization

Focus Variables:

- GOES ABI:
  - Fire Temperature, Fire Area, Fire Radiative Power
- MODIS VI:
  - NDVI, EVI, Reflectances
- MODIS TA:
  - Fire Mask, Fire Radiative Power

Metadata Curation:

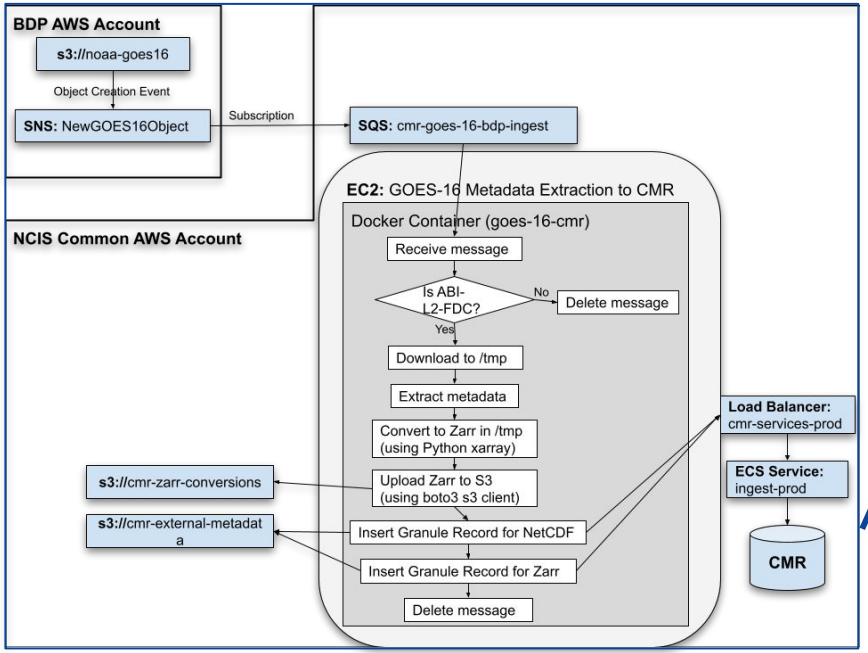
- New elements conforming with emerging best practice recommendations for cloud optimized data product metadata

Target Data Products

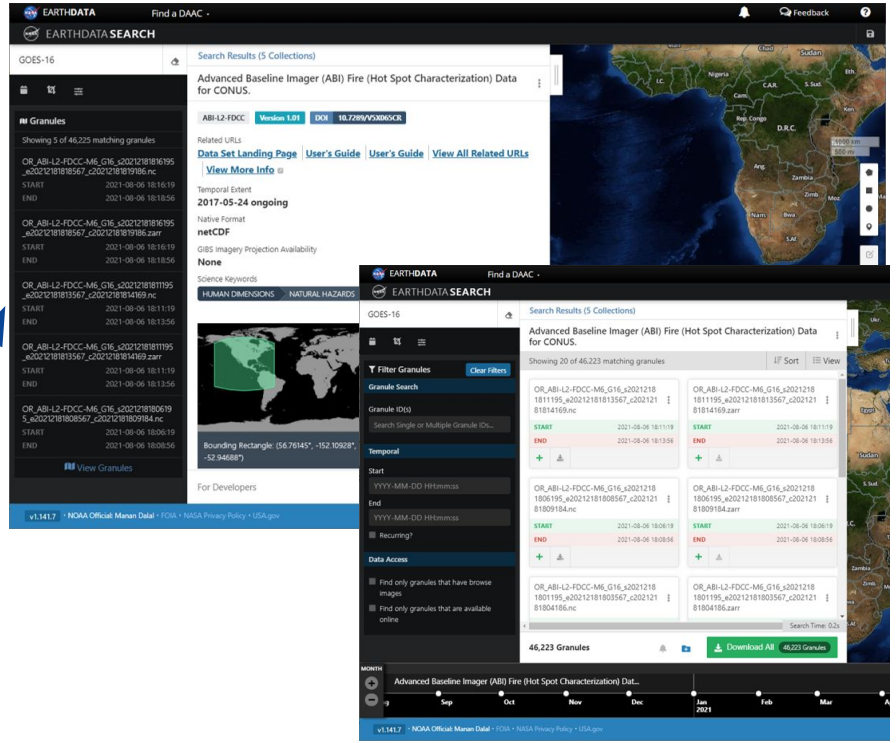
NOAA Products	Ingested (Native)	Zarr	COG
GOES-16 Advanced Baseline Imager Level 2 Fire (Hot Spot Characterization) CONUS (ABI-L2-FDCC)	Y	Y	Y*
GOES-17 Advanced Baseline Imager Level 2 Fire (Hot Spot Characterization) CONUS (ABI-L2-FDCC)	Y	Y	N
GOES-16 Advanced Baseline Imager Level 2 Fire (Hot Spot Characterization) Full Disk (ABI-L2-FDCF)	Y	Y	Y*
GEOS-17 Advanced Baseline Imager Level 2 Fire (Hot Spot Characterization) Full Disk (ABI-L2-FDCF)	Y	Y	N
Lightning for fire initiation (GLM)	Y	N	N
Himawari-8	Y	N	N
JPSS I-band Active fire product	N	N	N
NASA Products	**Ingested (Native)**	**Zarr**	**COG**
MODIS Terra Vegetation Indices 16-Day L3 Global 250 m (MOD13Q1)	Y	N	Y*
MODIS Aqua Vegetation Indices 16-Day L3 Global 250 m (MYD13Q1)	Y	N	Y*
MODIS Terra Thermal Anomalies/Fire Daily L3 Global 1km (MOD14A1)	Y	N	Y*
MODIS Aqua Thermal Anomalies/Fire Daily L3 Global 1km (MYD14A1)	Y	N	Y*

* Limited subset of data converted to COG for this project.

Architecture for NOAA Zarr Stores

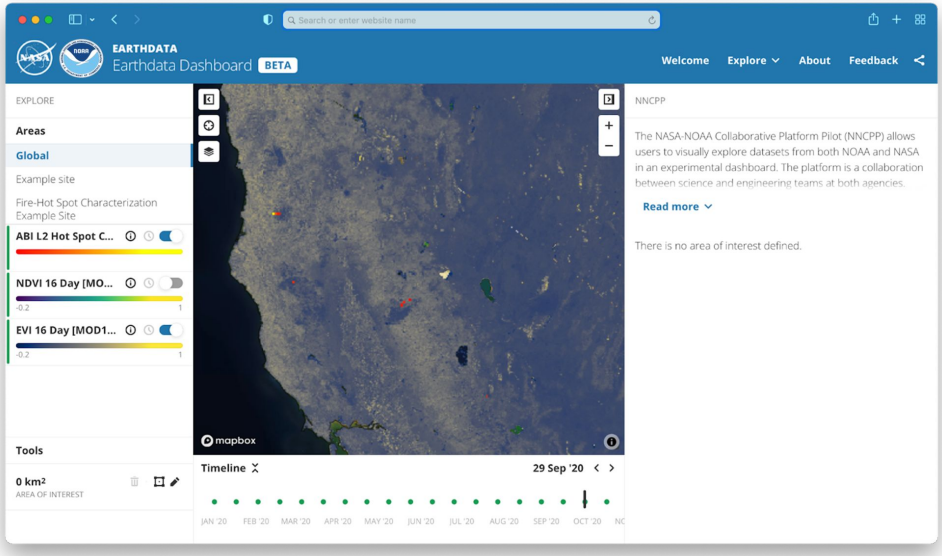


Earthdata Search UI



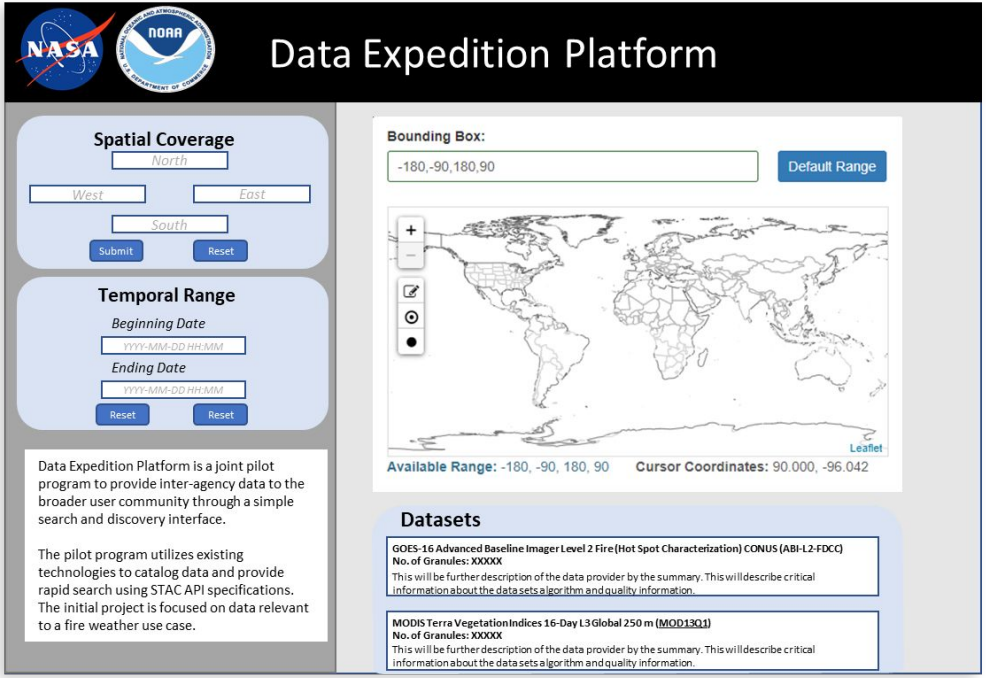
Visualization

- Leveraged existing technology by reusing the NASA's COVID-19 Earthdata Dashboard architecture to deploy a new instance of an interactive visual interface.
- Ingested target data products from both agencies for the period 1 Aug - 31 Oct 2020 (California's August Complex Fire)
- A core challenge addressed was data products' difference in temporal resolution.



User Interface - *Future Effort*

- Future phase of the project's effort will be to create a user-driven search for data through an inter-agency web-based interface.
  - Leveraging the STAC API backend.
  - Basic search functionalities such as temporal range, spatial coverage, and collection.
  - Search results providing access to data sets.



*This is a mockup of a potential user interface

Summary

The Data Expedition Platform team as part of the NASA-NOAA collaboration benefited from **continuous technology and information sharing**, improving the relationship between the agencies and advancing existing search and discovery technologies for Earth observation data.

Challenges Faced:

- *How to store the data in the existing agencies' cloud frameworks and implementing a common metadata repository.*
- *Best approaches to convert the native data to Cloud Optimized formats (Zarr, Cloud Optimized GeoTIFF).*
- *Tackling different data temporal resolution for dashboard integration*
- *Determining search functionality with the capability to search across multiple repositories from a single UI.*

Pilot project ***successfully demonstrates foundational capabilities of making NASA and NOAA data searchable through a unified mechanism using common metadata terminology***, and if further pursued could improve the openness of data, and facilitate more open science.

BACK UP

Goal

- ***Phase I: Configure NASA CMR within NOAA and NASA environment*** - prototype NASA's Common Metadata Repository (CMR) in the NESDIS Cloud Infrastructure Sandbox Service (NCIS) to: i.) evaluate if its capabilities can address NESDIS inventory and catalog needs; and ii.) Enable a cross agency data catalog with both NESDIS and NASA that can be used for data discovery and access.
  - Status
    - CMR has been successfully built in NCIS in a rapid timeframe
      - Ran sample data through the publishing to ensure functionality
      - Worked with NASA in determining services and enabling common population for metadata terminology
  - In progress
      - Use case is fire weather related data – determining pre-processing software conversion to zarr
      - Implementation of STAC API
      - Collecting information about candidate data sets recommended by NCAP leadership
- ***Phase II: Perform visualization or develop a service to provide ability to visualize the data***
      - Exploring Cloud Optimized Geotiff format conversion for a visualization dashboard
- ***Phase III: Enable common data access to end users***
      - Common interface for search and discoverability of inter-agency data

Goal

Key Activities/Outcomes

- Refining fire burn impacts use case: defining target event & vegetation properties/indices for visualization; selecting NASA and NOAA data products aligned with goals & spatio/temporal criteria
- Deployed Cumulus ingest and archive infrastructure on the NASA pilot platform
- Common Metadata Repository (CMR) deployment on the NASA pilot platform and NESDIS Cloud Infrastructure Service (NCIS)
- Conducted multiple information sharing sessions on data format conversions
- Stood up the Earthdata Search UI in NCIS with NASA support
- Successfully ingested first several candidate data sets into NCIS CMR, in native and zarr format in NCIS.

Cloud Architectures Data Access & Discovery

a.) NESDIS Cloud Archive Project

Key Activities/Outcomes:

- Held multiple technical information sharing sessions between NOAA and NASA on Common Metadata Repository (CMR) and Cumulus
- Began crosswalking metadata to ensure proper translation and common metadata terminology

Next Steps:

- Stand up Cumulus in the NCIS framework
- Communicate back with NASA metadata findings
- Implement access service capabilities through integrated UI

b.) Joint Data Expedition Pilot

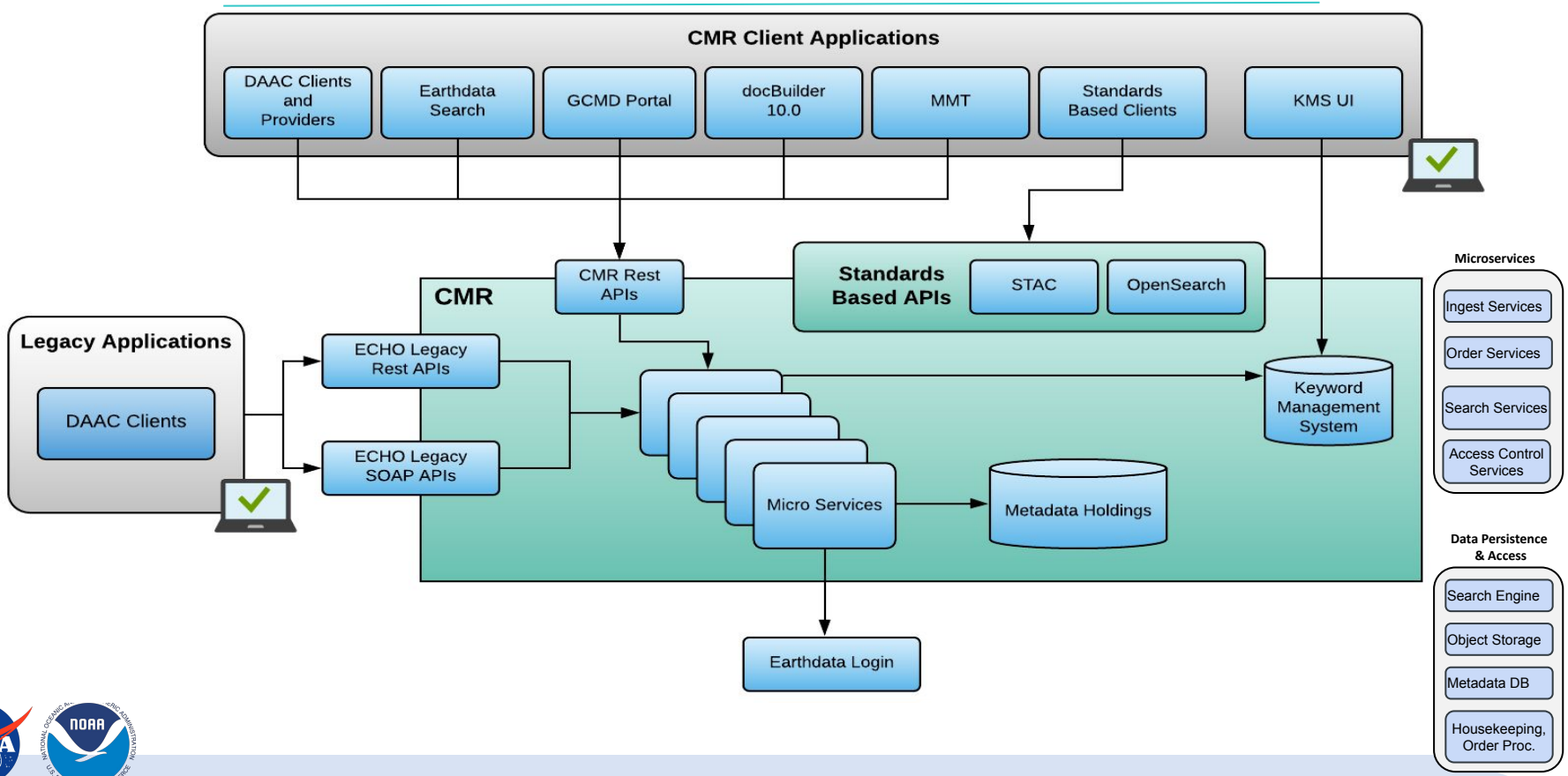
Key Activities/Outcomes

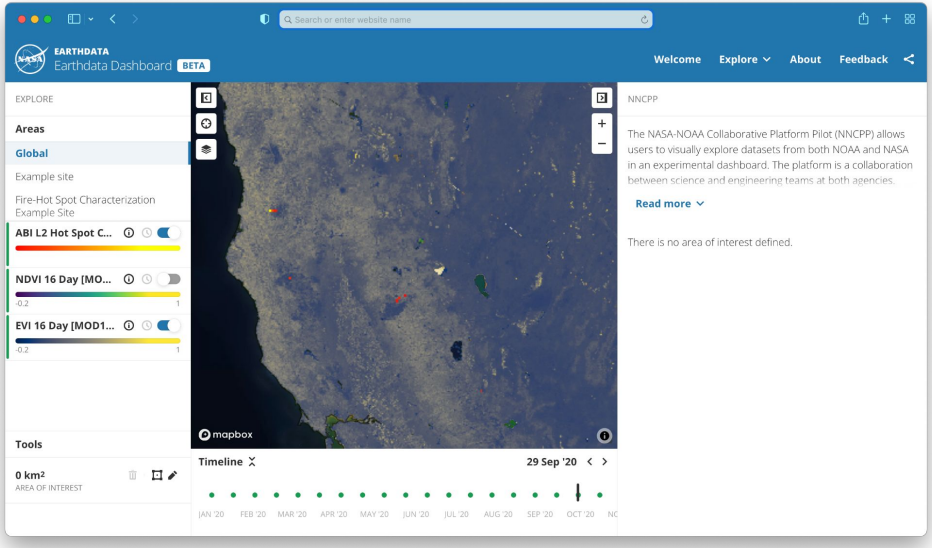
- Refining fire burn impacts use case: defining target event & vegetation properties/indices for visualization; selecting NASA data products aligned with goals & spatio/temporal criteria
- Deployed Cumulus ingest and archive infrastructure on the NASA pilot platform
- Common Metadata Repository (CMR) deployment on the NASA pilot platform and NESDIS Cloud Infrastructure Service (NCIS)
- Conducted multiple information sharing sessions on data format conversions
- Stood up the Earthdata Search UI in NCIS with NASA support
- Successfully ingested first candidate data sets into NCIS CMR, in native and zarr format in NCIS.

Next Steps

- Discuss fire use case details and data product selection criteria and reasoning with subject matter expert teams supporting the project

Overview of CMR Framework




Search or enter website name
EARTHDATA
Earthdata Dashboard
BETA
Welcome
Explore
About
Feedback
EXPLORE
Areas
Global
Example site
Fire-Hot Spot Characterization Example Site
ABI L2 Hot Spot C...
NDVI 16 Day [MO...
-0.2
1
EVI 16 Day [MOD1...
-0.2
1
Tools
0 km²
AREA OF INTEREST
mapbox
Timeline
29 Sep '20
JAN '20
FEB '20
MAR '20
APR '20
MAY '20
JUN '20
JUL '20
AUG '20
SEP '20
OCT '20
NNCPP
The NASA-NOAA Collaborative Platform Pilot (NNCPP) allows users to visually explore datasets from both NOAA and NASA in an experimental dashboard. The platform is a collaboration between science and engineering teams at both agencies.
Read more
There is no area of interest defined.