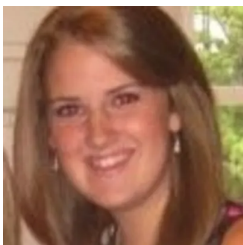


Advancing Open Science through Public-Private Partnerships



Elizabeth Fancher, Rahul Ramachandran

Barrios Technology, NASA Marshall Space Flight Center

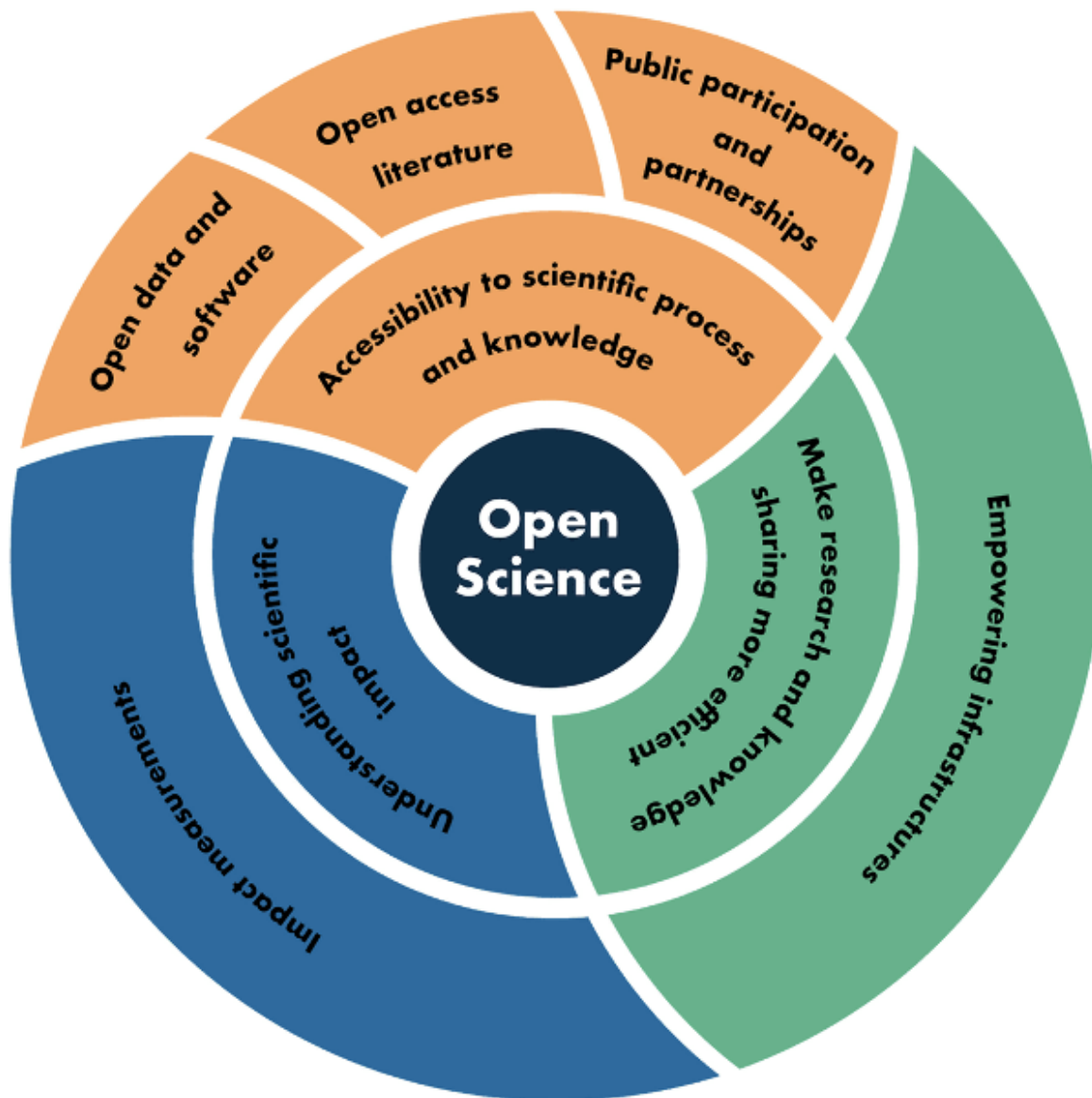


PRESENTED AT:



OPEN SCIENCE DRIVER

Collaborations are a crucial component of Open Science (Ramachandran et al., 2021). While governments and public and private organizations are making research advances in technologies such as cloud computing, artificial intelligence/machine learning, etc., the lack of formal collaboration mechanisms often hinders open knowledge transfer.



Our goal within NASA's Earth Science Data Systems Program is to overcome these barriers by establishing formal collaboration with external partners in the private sector. The goals of these partnerships are to:

- Have a formal mechanism to work on projects of mutual interest
- Share information and knowledge to coordinate future directions and leverage partners expertise
- Broaden the network of technical experts working on similar problems

Ramachandran, R., Bugbee, K., & Murphy, K. J. (2021). Moving from Open Data to Open Science. *Earth and Space Science*, Wiley Publication (<https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2020EA001562>)

CLOUD-BASED DISCOVERY

Under the SAA with **Amazon Web Services (AWS)** and **Microsoft**, NASA is investigating the transfer, storage, and value of making large volumes of NASA science datasets available on AWS's Open Data Sponsorship Program and Microsoft's Azure platform.

AWS's Open Data Sponsorship Program is providing public access to high value NASA datasets in a cloud optimized format. This open forum is facilitating open science within the community and providing valuable information back to NASA on how the data is being used. A registry of NASA's open data on the AWS cloud platform can be found at <https://registry.opendata.aws/collab/nasa/>. (<https://registry.opendata.aws/collab/nasa/>.)

Registry of Open Data on AWS



NASA Space Act Agreement



Amazon Web Services and the National Aeronautics and Space Administration (NASA) have entered into a Space Act Agreement to explore best practices around discovery, access, and use of high-value NASA science datasets. Making analytics-optimized data stores available to the science community will minimize the need for data wrangling and preprocessing within the community, leading to a faster time to insight and quicker innovation.

Search datasets (currently 4 matching datasets)

Add to this registry

If you want to add a dataset or example of how to use a dataset to this registry, please follow the instructions on the [Registry of Open Data on AWS GitHub repository](#).

Unless specifically stated in the applicable dataset documentation, datasets available through the Registry of Open Data on AWS are not provided and maintained by AWS. Datasets are provided and maintained by a variety of third parties under a variety of licenses. Please check dataset licenses and related documentation to determine if a dataset may be used for your application.

Multi-Scale Ultra High Resolution (MUR) Sea Surface Temperature (SST)

[climate](#)
[earth observation](#)
[environmental](#)
[natural resource](#)
[oceans](#)
[satellite imagery](#)
[sustainability](#)
[water](#)
[weather](#)

A global, gap-free, gridded, daily 1 km Sea Surface Temperature (SST) dataset created by merging multiple Level-2 satellite SST datasets. Those input datasets include the NASA Advanced Microwave Scanning Radiometer-EOS (AMSR-E), the JAXA Advanced Microwave Scanning Radiometer 2 (AMSR-2) on GCOM-W1, the Moderate Resolution Imaging Spectroradiometers (MODIS) on the NASA Aqua and Terra platforms, the US Navy microwave WindSat radiometer, the Advanced Very High Resolution Radiometer (AVHRR) on several NOAA satellites, and in situ SST observations from the NOAA iQuam project. Data are available fro...

[Details →](#)

Usage examples

- [HTTPS server by PO.DAAC](#)
- [Python Reader Software by PO.DAAC](#)
- [THREDDS server by PO.DAAC](#)
- [AWS Improving our knowledge about the oceans by providing cloud-based access to large datasets by Chelle Gentemann](#)
- [OPeNDAP server by Physical Oceanography Distributed Active Archive Center \(PO.DAAC\)](#)

[See 10 usage examples →](#)

NASA is also exploring ways to collaborate with the science community by hosting internal workshops and hackathons which aim at transforming data access and the usage of cloud-computing resources.

"Cloud Hackathon: Transitioning Earthdata Workflows to the Cloud" (<https://nasa-openscapes.github.io/2021-Cloud-Hackathon/>) was a virtual event co-hosted by Openscapes and NASA to explore and promote science on the cloud using NASA Earth Observations data from within the NASA Earthdata Cloud, hosted on AWS's platform.



Image Credit: Openscapes

Microsoft's Planetary Computer Data Catalog includes vast amounts of environmental data in analysis-ready formats which can be searched on Azure. NASA datasets such as Landsat 8 Collection 2 Level 2 can be queried on the platform by using the tag "NASA" in the search bar. Explore more at <https://planetarycomputer.microsoft.com/catalog?tags=NASA> (<https://planetarycomputer.microsoft.com/catalog?tags=NASA>)

[Planetary Computer](#)
[Explore](#)
[Data Catalog](#)
[Hub](#)
[Applications](#)
[Documentation](#)
[Request access](#)

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 available through the Planetary Computer API

Our largest data sets can be queried and accessed through our Planetary Computer API. We are continuing to expand the data available through the API, and continuing to bring new data sets to Azure. If you are interested in seeing additional data on-boarded or published through our API – or if you have data you'd like to contribute – [contact us](#).

OVERVIEW OF PUBLIC-PRIVATE PARTNERSHIPS

Strategic public-private partnerships have been established as non-reimbursable Space Act Agreements (SAA) between **NASA, Amazon Web Services, Microsoft, Google and the Frontier Development Lab.**



Non-reimbursable SAAs are collaborative partnerships in which NASA and another party agree to execute a non-procurement relationship that does not require acquisition of goods or services (no exchange of funds).

Each SAA centers around the collaboration theme of discovery, access, and use of high-value NASA science datasets and explores the following objectives:

- Enable new types of analytics with NASA data collocated to promote open science in the cloud
- Demonstrate the value of bringing algorithms to NASA data in the cloud
- Improve discovery of NASA data and other important resources (services, tools, etc.) through commercial search engines
- Explore new applications of artificial intelligence/machine learning (AI/ML) to address data search, discovery and use of NASA data

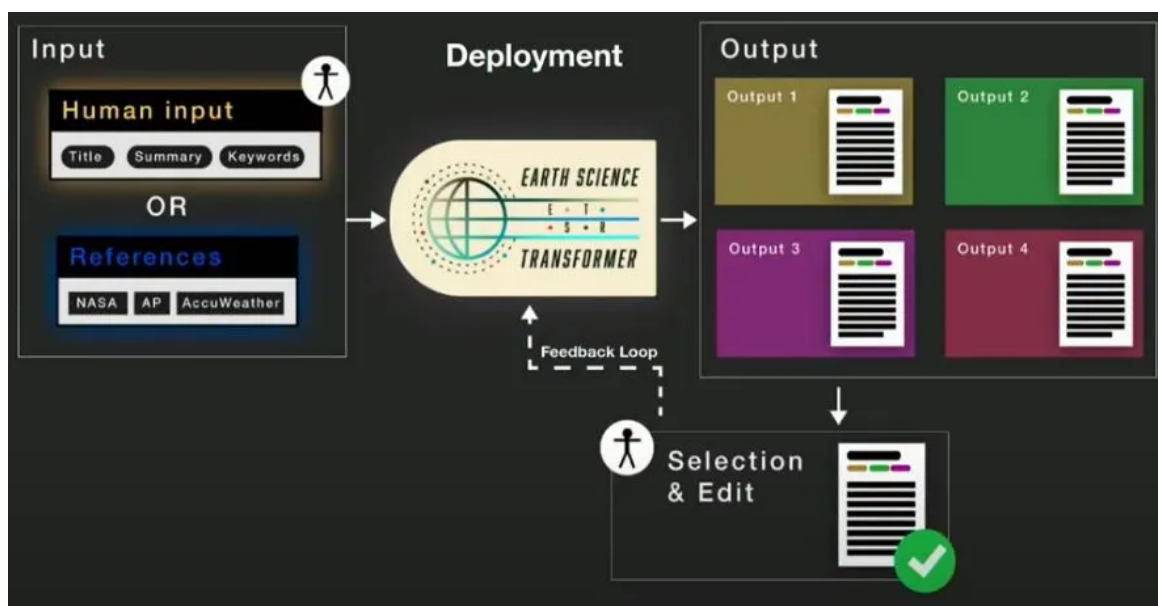
Image Credit: Amazon Web Services, Inc., Frontier Development Lab, and Google, LLC.

APPLICATIONS OF AI/ML

Under the SAA with **Google**, NASA has partnered with the **Frontier Development Lab (FDL)** to demonstrate and evolve our artificial intelligence (AI) and machine learning (ML) capacities. FDL 21 is an applied AI research partnership with Google focusing on developing future AI applications in data-intensive challenges related to space science, planetary science, and Earth science. Two success stories are outlined below.

Automated Reporting Framework

NASA sponsored an Earth science challenge to automate the reporting of natural events using AI and NASA's satellite imagery and metadata. The challenge participants utilized a natural language processing technique and transformer language model to generate descriptions of natural hazards and phenomena. These narratives in turn could be used for distribution on public websites or on popular social media platforms without the need of a human writer. The full video demonstration is here (<https://www.hashcut.com/v/xkF5RFa>) and framework is shown below.



NASA GIBS Worldview Similarity Search is a one of seven SpaceML projects (spin off from FDL 2020) applying AI and ML to NASA data in an effort to develop a new search engine pilot to enable content based searches. The prototype allows science users to sift through NASA's vast archive of Earth science data to identify patterns or events given a single image. Access to open repository is here (<https://spaceml.org/repo/project/605b7b751644770011e850c3/false/>) and shown below.



WorldView Search

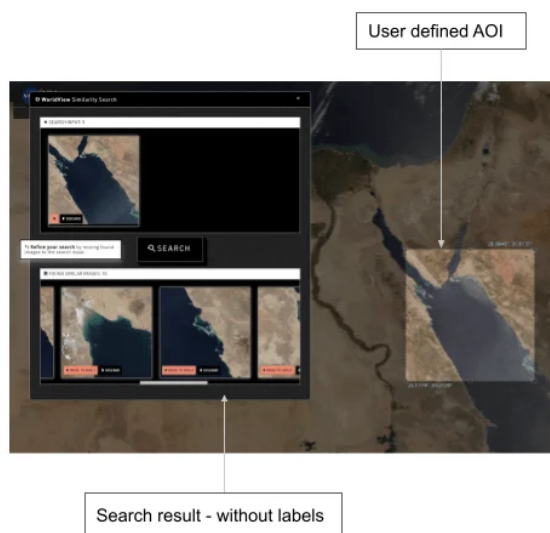
Searching unlabeled petabyte-scale imagery with self-supervised Machine Learning.



FRONTIER
DEVELOPMENT
LAB



ARTIFICIAL INTELLIGENCE
RESEARCH FOR SPACE SCIENCE,
EXPLORATION & ALL HUMANKIND



www.spaceML.org

Both open source science prototypes could be used to support societal applications such as disaster response and monitoring environmental changes as well as addressing gaps in existing search and discovery tools hosted by NASA.

NEXT STEPS

-

Continue to collaborate with each partner on the **discovery, access and use of high-value NASA science datasets** on commercial cloud platforms

-

Broaden and increase access to NASA's data and other data resources by executing **new Space Act Agreements** with other interested partners

-

Investigate and foster **new AI applications using NASA** data which will enable and accelerate open science

-

Build capacity in the science research and applications community by supporting training workshops which promote utilization of NASA data on the cloud

-

Investigate innovative ways to combine **interactive storytelling platforms** with current NASA data systems

-

Explore new strategies for hosting and showcasing NASA data on the AWS Open Data Sponsorship Program (demonstration sandbox/playground) and other commercial cloud platforms

Notice!

Your iPoster has now been unpublished and will not be displayed on the iPoster Gallery.

You need to publish it again if you want to be displayed.

ABSTRACT

Rapid technology developments are changing the way data-driven research is performed within the science community. With the emergence of cloud computing, this has quickly become a viable approach for enabling “science at scale”. Researchers are no longer hindered by obstacles of data management and data wrangling, allowing them to quickly discover, access and perform analysis on extremely large datasets. Infrastructures that move data out of institutional silos and into a computational platform, will ensure that data and tools are accessible to all users.

NASA’s Interagency Implementation and Advanced Concepts Team (IMPACT) seeks to address these rapid technology developments by establishing Space Act Agreements with selected partners from the public-private sector working in the area of cloud computing. These agreements aim to explore new opportunities with commercial cloud providers to accelerate open science and enable discovery, access and use of data sets on the cloud. In addition, they will also help establish training workshops for the science community to help researchers utilize the cloud for science. In this talk, we will present an overview of current and new partnerships we are developing to support open science and open data initiatives.