Getting Ready for NISAR—and for Managing Big Data using the Commercial Cloud

The upcoming NISAR satellite mission is expected to add as much as 85 TB of data each day to the EOSDIS archive. The commercial cloud is being explored as a way to archive and disseminate this extremely high volume of data efficiently.

The launch of the upcoming joint NASA/Indian Space Research Organisation (ISRO) Synthetic Aperture Radar (NISAR) mission, currently scheduled for late 2021, will be a landmark undertaking. NISAR is not only the first joint mission between NASA and ISRO scheduled to launch (the two organizations also have an agreement to work on joint Mars exploration missions), but also the launch of the first dual-frequency synthetic aperture radar (SAR). The data collected by the L-band (produced by NASA) and S-band (produced by ISRO) SAR systems aboard the NISAR satellite and processed into cloud-free, ultra-sharp imagery will facilitate cutting-edge research into some of the planet’s most complex processes, including ecosystem disturbances, ice-sheet dynamics, earthquakes, tsunamis, volcanoes, and landslides.
“NISAR will produce copious amounts of data and these data will be in high demand, not only for the NISAR products themselves, but as ingredients in the generation of multiple higher-level informational products,” says NISAR Program Scientist Craig Dobson. “The global scope of NISAR science combined with NASA's open data policy will stimulate and facilitate vast interest in these data.”

As Dobson notes, NISAR is expected to generate a tremendous volume of data over its scheduled three-year mission—as much as 140 petabytes (PB). In comparison, the total volume of data in NASA's Earth Observing System Data and Information System (EOSDIS) archive at the beginning of 2017 was about 22 PB, according to metrics from NASA's Earth Science Data Systems (ESDS) Program. In fact, NISAR is expected to generate a data volume close to 85 terabytes (TB) each day. This is much greater than any currently operating NASA Earth observing mission. “This places considerable demands on the logistics of shipping data and on computational speed and efficiency,” Dobson says.

![NISAR](image)

“NISAR has a two-part problem,” observes Chris Stoner, the Project Office Manager at NASA's Alaska Satellite Facility Distributed Active Archive Center (ASF DAAC), which is one of 12 discipline-specific EOSDIS DAACS and the future home for NASA's NISAR data. “The file sizes are large and the overall volume of data will be huge. This means we have to do something different to ensure that NISAR data users have a good user experience and are able to do their research.”

The “different” approach ASF DAAC is exploring to efficiently store and distribute the tremendous amount of data expected from NISAR is to use the commercial cloud. While the ASF DAAC will archive and distribute NISAR data, these data will be processed at NASA's Jet Propulsion Laboratory (JPL), located at the California Institute of Technology in Pasadena, CA. ASF DAAC is working collaboratively with JPL to test and prototype ways of archiving and distributing NISAR data using the commercial cloud. This three-year project began in 2016 and is called Getting Ready for NISAR, or GRFN (pronounced Griffin).

Now in its second year, GRFN is part of the ongoing EOSDIS efforts to evolve NASA Earth observing data and EOSDIS services to the commercial cloud. The primary GRFN goals are to:

1. Obtain a better understanding of the costs and technical challenges associated with cloud-based and hybrid architectures for processing and storing NISAR data, and
2. Provide the opportunity for the science community to become comfortable interacting with and working with large SAR datasets in the cloud.

Since NISAR data will not be available until after launch, SAR data from the European Space Agency’s (ESA) Sentinel-1 mission (which are archived and distributed by the ASF DAAC) are being used as a surrogate to prototype a cloud-based system using Amazon Web Services (AWS). AWS is currently the only commercial cloud provider approved by NASA for NASA data.

Stoner notes that the huge size of NISAR files is a primary reason for using the commercial cloud for processing, archiving, and distributing mission data. Modern SAR sensors produce extremely high-resolution images, day or night, without the need for outside illumination (such as from the sun). By comparing SAR imagery from different days, subtle changes easily can be observed and measured, such as uplift from earthquakes or subsidence from excessive groundwater pumping.

The trade-off for this detailed imagery, though, is extremely large data files necessitating high data transfer rates. According to the ASF DAAC, Sentinel-1 data products average 5 gigabytes (GB) per frame. NISAR data products will be much larger than Sentinel-1 products, and average 25 GB per frame. According to Stoner, users of Sentinel-1 data typically download hundreds or sometimes thousands of scenes for their research. The time researchers need to download data is time away from conducting research using the data.
“The difference between a 5 gigabyte Sentinel file and a 25 gigabyte NISAR file is significant,” says Stoner. “Researchers on Sentinel-1 now are asking us if there are alternatives to downloading a bunch of files because it takes a while. Well, ‘it takes a while’ is going to evolve for NISAR to ‘I cannot work this way.’ It will simply take too long to process NISAR data using conventional processing technology.”

The commercial cloud provides an elegant alternative to downloading huge data files. In a cloud-based system, researchers have the ability to work with data directly in the cloud and only have to download the finished, derived product. “Data users can bring their algorithms to the cloud and process [data] next to data storage; no downloads,” Stoner says, noting that while the derived product still could be very large in size, a researcher will not have to wait to download all the individual files necessary to produce the finished product.

During the first year of GRFN, JPL built a prototype science data system (SDS) environment in AWS to generate Sentinel-1 interferograms and deliver these to the ASF DAAC. Meanwhile, the ASF DAAC built a prototype system to manage all aspects of the data lifecycle for these products in AWS, including data ingest, storage, discovery, distribution, and on-demand product generation. This not only allows data users to become familiar with using SAR products in the cloud, but, by analyzing how these products are being used, also provides a better understanding of the associated costs of cloud-based storage and dissemination of these data.

GRFN is simulating several processing scenarios applicable to SAR data and the NISAR mission:

- Forward stream processing, or “keeping up” with normal data flow from the satellite
- Bulk reprocessing
- On-demand processing
- Urgent response (such as after a tsunami, an earthquake, or other natural event)

The GRFN team already has demonstrated sustained processing speeds of 10 gigabits per second (Gbps), which is the speed required for NISAR forward stream processing. The next objective will be in achieving 40 gigabits per second, which is the NISAR bulk reprocessing rate. The goal, says Stoner, is 50 gigabits per second, which is the rate required for receiving both forward and bulk processing. Processing rates this fast will occur when the satellite is sending data to the SDS for processing while the SDS is coincidentally reprocessing a year’s worth of data and sending these data to the ASF DAAC as rapidly as possible.

“Right now, we’ve achieved very fast processing, and it’s going to get faster,” Stoner says. “Our challenge is to keep up with the data processing needs while keeping the whole process cost-effective. Overall, the cloud shouldn’t be excessively expensive [compared with more traditional data processing and storage methods]. The cloud should be cost effective, especially at scale.”

Although only in its second year, GRFN already is successfully demonstrating the viability of using the commercial cloud for SAR data missions, with tools and techniques that easily can be applied to similar high-volume data missions. “There is certainly a time constraint to research, and we’re trying to react to that. Cloud computing really answers a lot of the problems [in dealing with big data] and it offers probably a better and a broader opportunity,” says Stoner.

By laying the groundwork for processing, archiving, and disseminating mission data using the commercial cloud now, well in advance of the scheduled NISAR launch, JPL, the ASF DAAC, NASA’s EOSDIS, and data users will be better prepared to maximize the tremendous amount of data expected from this groundbreaking mission.
Earth Science Data in the Cloud: The EOSDIS Cumulus Project

As part of the ongoing evolution of EOSDIS data and services, testing and prototyping are underway to see how DAAC data collections can be archived and disseminated using the commercial cloud.

Clouds in the sky constantly grow and shrink as they adjust to evolving atmospheric conditions. A cloud computing environment, like an atmospheric cloud, also easily can adjust to evolving conditions, expanding or contracting as needed based on data storage requirements and the needs of data users. This flexibility helps make the commercial cloud a viable option for archiving and disseminating large volumes of data or for managing data holdings that are expected to change rapidly over a short amount of time.

NASA’s Earth Observing System Data and Information System (EOSDIS) is responsible for a data collection that is both large in volume and projected to grow rapidly over the next several years. From its current size of almost 22 petabytes (PB), the volume of data in the EOSDIS archive is expected to increase to almost 247 PB by 2025, according to estimates by NASA’s Earth Science Data Systems (ESDS) Program.

To prepare for this tremendous growth and efficiently provide access to these data, the EOSDIS is investigating the evolution of its data and services to run in the commercial cloud. As part of these efforts, staff at NASA’s Earth Science Data and Information System (ESDIS) Project (which manages EOSDIS data) are prototyping and testing how EOSDIS data collections can be archived collectively and disseminated in the cloud. As befitting the cloud environment, this prototype is called Cumulus.

A primary feature of Cumulus is a cloud-based framework for data ingest, archive, distribution, and management, which are the primary activities of the discipline-specific EOSDIS Distributed Active Archive Centers (DAACs). The overall Cumulus goal is to provide the following functionality in the commercial cloud:

- Data acquisition from data providers (such as NASA science teams),
- Data ingest (including validation and processing),
- The harvest, creation, and publication of dataset metadata to the EOSDIS Common Metadata Repository (CMR),
- The storage and distribution of data, including disaster recovery, and
- Publication of metrics to the ESDIS Metrics System (EMS), which collects and organizes various metrics from the DAACs and other data providers.

The DAACs would still serve as gateways to EOSDIS Earth science data and continue to provide a wide range of support services for data users. EOSDIS data users likely would not even notice any difference in interactions with their discipline-specific DAACs and Earthdata Search when searching for and downloading data that happens to be stored in the cloud.

Selected EOSDIS data and services already are operating in Amazon Web Services (AWS), which currently is the only NASA-approved commercial cloud provider. Earthdata Search and the CMR evolved to the cloud in September 2016 and April 2017, respectively. Global Imagery Browse Services (GIBS), which provides access to over 400 satellite imagery products that can be viewed using client applications such as EOSDIS Worldview, is expected to evolve to the cloud as a prototype starting in 2018. The next step in this evolution is to prototype and test EOSDIS data collections in the cloud. This is an important undertaking given the expected significant growth of the EOSDIS archive.
Between 2015 and 2017, the volume of data in the EOSDIS archive more than doubled, from roughly 10 PB to almost 22 PB (see graphic). This dramatic growth in the EOSDIS data archive volume is expected to continue over the next several years. The ESDIS Project believes that having these data archived and disseminated in the commercial cloud provides the EOSDIS with a cost-effective, flexible, and scalable data system that can keep pace with mission advancements and capabilities. The cloud also gives EOSDIS data users the ability to efficiently access and process significantly larger data volumes from multiple DAACs and do more with these data without having to download large data collections (including the ability to work with data directly in the cloud). Other benefits of having EOSDIS data in the cloud are described in detail in “Evolving NASA Earth Science Data and Services to the Cloud,” which is available through the EOSDIS Cloud Evolution page on the Earthdata website. Along with testing data storage and dissemination in the cloud, ESDIS Project staff are working with the DAACs to develop compatible tools and libraries that can be used across DAACs and with multiple data collections (such as widely applicable geographic information system [GIS] components or sub-setters) as well as enabling discipline-specific data customizations (such as the ability to work with specific subsets of data). If prototypes developed as part of Cumulus are successful, an entire DAAC could be running in the cloud by 2019 or 2020.

Of course, not every DAAC or DAAC data collection may be appropriate for evolution to the cloud, and DAAC functions and data products are being reviewed to determine which collections might work best in the cloud. This includes an evaluation of current DAAC data volume size and growth along with DAAC data distribution characteristics.

Along with EOSDIS data products, ESDIS Project Cumulus team members will continue testing dataset-specific tools and applications in the cloud. These tools and applications will be migrated to the cloud on a case-by-case basis. The overall objective is to containerize specific tools and applications to run in the commercial cloud, especially those that can be used across multiple DAACs and with multiple datasets.

While Cumulus is still in the testing and prototyping phase with individual DAACs and selected data, these efforts will create a firm foundation for managing the expected significant growth of the EOSDIS archive and provide options for efficiently disseminating these data. More importantly, EOSDIS cloud efforts like Cumulus ensure that data users will continue to be able to maximize both their research using NASA Earth science data and the amount of data with which they can work.
AppEEARS 2.0: New Features and Data to Further Improve User Workflow Efficiency

Lindsey Harriman, LP DAAC Science Integration Lead; SGT, Inc.

Aaron Friesz, LP DAAC Data Scientist; Innovate, Inc.

As the amount of remote sensing data increase in number and volume, users are continually looking for ways to simplify and streamline their data access and processing methods. The release of the Application for Extracting and Exploring Analysis Ready Samples (AppEEARS) in March 2016 by NASA’s Land Processes Distributed Active Archive Center (LP DAAC) significantly reduced the amount of data users needed to download, enabling similar datasets to be co-discovered more easily and providing a way to more efficiently explore results. The recent release of AppEEARS 2.0 further increases the usefulness of this powerful application with the addition of an Area Sampler along with access to more datasets.

The initial release of AppEEARS included a Point Sampler that immediately lifted the data access and processing burden from users conducting studies at discrete geographic point locations using data distributed by the LP DAAC. The LP DAAC is a partnership between the U.S. Geological Survey (USGS) and NASA, and is one of 12 discipline-specific Distributed Active Archive Centers (DAACs) in NASA’s Earth Observing System Data and Information System (EOSDIS).

The AppEEARS Point Sampler allows users to input tens, and even hundreds, of geographic coordinates, select specific data variables of interest from NASA Earth science remote sensing data, and set constraints to receive time series information. Quality data associated with raw data values also are returned with AppEEARS requests, which is an element of remote sensing data that often is overlooked. Sample request outputs are downloaded as comma separated value (csv) files. The use of the csv format significantly reduces the amount of data users have to download to their local machine. Users also can explore data values, trends, and relationships within the AppEEARS interface using a variety of interactive charts and graphs.

The release of AppEEARS Version 2.0, announced on September 29, 2017, adds greater functionality through the addition of an Area Sampler. The Area Sampler enables users to subset and transform geographic areas (polygons) in a variety of ways. Users can set parameters to subset by time and by layer, as well as mosaic, reformat, and reproject multiple datasets in one simple step. In addition to returning area-specific raw data values, AppEEARS also returns definitions for quality data. A recent follow-on release provides an additional download option for retrieving summary statistics relating to data visualizations offered within the Area Sampler in the csv file format. With users now able to subset specific geographic areas along with setting variable and temporal constraints, they no longer need to download gigabytes or terabytes worth of data. In fact, in some cases AppEEARS can reduce the amount of data users are required to download by up to 10,000 percent—a significant savings in time and computational power.

The growth of the AppEEARS user base has led to the demand for not only more features and functions, but...
also for more datasets. AppEEARS now offers access to more than 100 datasets from several EOSDIS DAACs. Data products in AppEEARS are provided through a collaborative effort by the LP DAAC, the National Snow and Ice Data Center DAAC (NSIDC DAAC), and the Socioeconomic Data and Applications Center (SEDAC). The LP DAAC also is working with the Oak Ridge National Laboratory DAAC (ORNL DAAC) to add Daymet data and with the NSIDC DAAC to add a selection of datasets from NASA’s Soil Moisture Active Passive (SMAP) mission. In addition, AppEEARS provides access to a wide range of environmental variables derived from data collected by the Moderate Resolution Imaging Spectroradiometer (MODIS) aboard NASA’s Terra and Aqua Earth observing satellites, as well as socioeconomic variables from SEDAC’s Gridded Population of the World (GPW) Version 4 collection. Finally, the LP DAAC is working with the USGS’ Landsat Program at the Earth Resources Observation and Science (EROS) Center to prototype access to Landsat Analysis Ready Data (ARD).

AppEEARS offers users a simple and efficient way to access and prepare data for analysis by enabling users to download only the data they need. The release of AppEEARS 2.0 provides greater functionality for this application as it continues to grow across the EOSDIS enterprise, link archived data across federal agencies, and extend its reach in the Earth science user community.

Additional Resources

• Access the AppEEARS Point and Area Samplers, find help documentation, and browse a list of available data products at https://lpdaacsvc.cr.usgs.gov/appears/

• Read more about the collaboration with SEDAC and the importance of linking human dimensions with land dynamics in this World Data System blog post

• Learn more about how data from the LP DAAC and NSIDC DAAC are being used for data analysis in this Data in Action story ■

The Global Change Master Directory: Data, Services, and Tools Serving the International Science Community

Over more than a quarter-century, the GCMD has evolved into the primary international resource for discovering Earth science and environmental data collections.

Over the past quarter-century, NASA’s Global Change Master Directory (GCMD) has become an integral system facilitating Earth science and global change studies. The metadata and keyword structures of the GCMD are pivotal components of NASA’s Earth science data collection. The GCMD also is a cornerstone of NASA’s international collaboration, and one of NASA’s contributions to the international Committee on Earth Observation Satellites (CEOS), where it is known as the CEOS International Directory Network (IDN).

Since its inception, the GCMD and NASA’s Earth Observing System Data and Information System (EOSDIS) have remained separate systems. Now, through the development of the EOSDIS Common Metadata Repository (CMR), these systems are in the process of becoming unified through the use of the CMR as the metadata source for both systems. For the GCMD’s broad base of international data users, this means a more robust system and the ability to drill-down even more deeply in their searches for Earth science and environmental collection-level data. “GCMD’s original purpose, and its continuing purpose, is to support the discovery of Earth science and environmental data collections,” says Dr. Stephen Wharton, the former GCMD Project Manager and Chief of NASA’s Global Change Data Center (GCDC).
To put this significant recent evolution of the GCMD into perspective, it is worth reviewing the development of the GCMD and the many innovations adopted and created by the GCMD team. A look at the future direction of the GCMD shows how this directory will remain a premiere collaborative international resource linking scientists, researchers, policy makers, and the general public with Earth science and environmental data.

The GCMD was established at a fortuitous time, and filled a need for discovering Earth science and environmental data. The 1980s saw not only the development of computers with the required power and cost-effectiveness to support such a directory, but also a literal turning point in Earth’s environmental systems. Earth observing data from numerous sources indicate that during this decade “abrupt, substantial, and persistent changes in the state of natural systems” occurred, according to recent research. This, in turn, led to a growing need for researchers, scientists, and managers to discover Earth science data related to these changes. However, this was easier said than done. “I think it’s fair to say that these [data] collections were not necessarily searchable online [in the late-1980s],” says Dr. Wharton.

In 1987, NASA released the NASA Master Directory (NMD) as a source for Earth and space data described at the collection level. While collections of data were being exposed as available to users, users would have obtained the file-level data by ordering media offline in the form of tapes or the then-new technology of compact disks; there was no easy way to find file-level data. By the early 1990s, NASA Earth science data were separated into their own directory—the GCMD. In 1994, the GCMD became part of NASA’s Global Change Data Center at NASA’s Goddard Space Flight Center (GSFC) in Greenbelt, MD. Also around this same time, the EOSDIS was conceived as NASA’s premiere system for archiving and disseminating Earth science data at the file level. It was natural that the EOSDIS and the GCMD would be managed under the same program, yet remain separate entities.

It is important to note the distinction between collection-level and file-level (or what the EOSDIS refers to as “granular”) data. A data collection is a description of data where people can understand what the data are about. A data granule, on the other hand, is an individual data value that is part of a larger collection. For example, you might have a data collection comprising 10 years of data, but you might want one day of data from one month in this 10-year collection; this is the data granule. As established, the GCMD and the EOSDIS served different needs (collection-level data searches vs. file-level/granular data searches). This, in turn, required that the GCMD and the EOSDIS have separate systems describing their data and enabling these searches. Data used to describe data are called “metadata,” and are what make data discoverable and searchable. As a result, the GCMD and the EOSDIS remained separate systems.

The recent development of the EOSDIS Common Metadata Repository (CMR) created the opportunity...
to finally unify the separate metadata systems used by the GCMD and the EOSDIS into a single system. The CMR was developed by NASA’s Earth Science Data and Information System (ESDIS) Project to be the authoritative management system for all EOSDIS metadata and facilitate rapid searches through the EOSDIS archive. The CMR serves as the metadata source for the EOSDIS’ Earthdata Search and now also serves as the metadata source for the GCMD.

Having the CMR as the metadata source for the GCMD is considered to be a win-win for data users by the GCMD staff in that the CMR not only speeds up GCMD searches, but also enables GCMD users to drill-down even more deeply into Earth science data collections. “Prior to this, the GCMD had its own backend system for serving data and information on the GCMD website,” says Alicia Aleman, the GCMD Senior Science Coordinator. “Once the CMR was in place, we migrated all of our content from our own servers and databases to the CMR. Now we’re part of this much stronger, more robust infrastructure.”

The use of the CMR as the source for GCMD metadata is only the latest evolution of the directory, and builds on many innovations developed by the GCMD team. These include the establishment of Science Keywords and data portals for easy data collection discovery, the adoption of the Directory Interchange Format (DIF) standard for exchanging information about scientific datasets and the development of docBUILDER for ensuring complete dataset metadata, and the implementation of automated quality assurance (QA) rules to ensure the highest quality metadata.

**GCMD Science Keywords** remain the heart of the GCMD system and an international resource used throughout the Earth science community. The keywords describe Earth science data and services consistently and comprehensively in a hierarchical format, and follow a codified governance process. The power of the keywords is in their ability to enable scientists to tag their data using a taxonomy of controlled scientific categories. This, in turn, allows those searching for data to discover datasets easily through the use of an established hierarchy. “The GCMD science keywords are an authoritative source that can be integrated into search interfaces, used for metadata authoring tools, and serve as the foundation for building ontologies,” says Aleman. “The main strengths of the keywords are the breadth of their content and how they allow for precision in searching for and retrieving data.” Since their introduction more than 20 years ago, GCMD keywords continue to be refined and expanded in response to user needs. As part of a more formalized keyword governance process, a **Keywords Community Forum** recently was established. The forum gives GCMD users an even greater say in keyword development and evolution. “The forum is intended to be a place where users can come to us with questions about keywords, we can respond to them, and users can comment on our responses,” explains Dr. Wharton.

While the science keywords are a guide to finding large data collections, scientists studying specific areas looked to the GCMD for a way to make their smaller, more specific data collections more accessible. “They said, ‘We have all these datasets, but we don’t want to build a system to make all these data accessible. Can you build something for us so that all these datasets are shown and we can have something out there that has our name and our institution and allows us to highlight the datasets we have?’” says Dr. Wharton.

This request evolved into the **GCMD portals**. Portals facilitate focused views for organizations to maintain and document their data within the GCMD without having to create a separate online directory for these data. When new datasets are defined and submitted to the GCMD, they are automatically recognized as being part of a specific portal without having to tag each individual dataset. GCMD portals are an easy way to put datasets on the map, and benefit from the high level of quality given them through the GCMD science keywords describing the information.
An important element in GCMD evolution was the adoption of the DIF standard as a means for exchanging information about scientific datasets. DIF was developed in the late-1980s to provide a specific set of attributes for describing Earth science data. “How do we structure the information in the [GCMD] consistently? This led to development of the DIF,” says Dr. Wharton. “The DIF was a consistent format for representing all this information.”

The DIF standard is the basis for constructing directory entries that describe a group of data, that is, metadata. Having high-quality metadata provides two important benefits for GCMD users—it increases the likelihood that researchers will find their datasets of interest and it decreases the likelihood that datasets will become undiscoverable.

The DIF is the “container” for the metadata elements that are maintained in the GCMD database. To ensure that all required dataset metadata are entered in a DIF record and the record is complete, the GCMD team developed the innovative docBUILDER tool that allows metadata authors to easily create or modify dataset descriptions. The most recent iteration of docBUILDER, docBUILDER-10, ensures that DIFs comply with CMR requirements, and allows metadata authors to validate and submit DIFs directly to the CMR.

The GCMD team further ensures the quality of GCMD metadata using quality assurance (QA) rules. QA recently was enhanced through automation that enables GCMD Science Coordinators to automatically make fixes and update metadata. “Quality review used to be done manually, which was a time-consuming process and things could be missed,” says Dr. Wharton. “Now we have an automated QA system with a formal set of QA rules. We can [conduct] QA [of] all of our records very quickly.”

QA is especially important when dealing with keyword releases. “When we change a science keyword, we have to update the associated metadata,” Aleman says. “We now have tools in place that can facilitate these changes very rapidly, assuring that the metadata remain in compliance with QA rules.”

All these enhancements, tools, and services come together in the redesigned GCMD website. “When you go to the website, you have an option to search by GCMD keywords, either by facets or traditional ‘drill-down,’ or you can search by free text,” says Aleman. “This offers different options for different users. Because our user community is so broad, we’ve found this to be a really successful implementation of our search interface.”

Through its continuous development over more than a quarter-century, the GCMD has established itself as a key resource for its global user community. The overall objective of the EOSDIS is the continuous improvement of all systems, with better access to Earth science resources for the entire science community. Unifying the GCMD and EOSDIS metadata systems through the CMR is the most recent evolution in this continuous improvement.

Future evolution includes the development of a single metadata management tool that will incorporate all of the features of both the GCMD and the EOSDIS, meaning users will only have to learn one tool. Also in the works is a combined approach to QA for the metadata supplied by the CMR to both the GCMD and EOSDIS systems, further improving metadata quality. “In terms of capabilities, NASA has a commitment to the GCMD services,” says Dr. Wharton. “Putting GCMD tools into one system that has more resources and capabilities than the GCMD improves the long-term viability of [the GCMD].”

While GCMD users might not notice what’s going on in the background, the services, tools, and enhancements developed by the GCMD team along with the evolution to using the CMR as the source for GCMD metadata ensure that this directory remains a powerful resource. Through the ongoing support of NASA and the EOSDIS, this evolution will continue.

**Additional Resources**

- GCMD Keywords Community Forum: [https://wiki.earthdata.nasa.gov/display/gcmdkey/GC...](https://wiki.earthdata.nasa.gov/display/gcmdkey/GC...)
- GCMD Website: [https://gcmd.nasa.gov](https://gcmd.nasa.gov)
Earthdata Login is the Key to Unlocking NASA Earth Science Data

An Earthdata Login is all that’s needed to download data from the huge EOSDIS archive.

NASA Earth science data are fully and openly available to the public at no charge. This policy applies to not only data, but also metadata, algorithms used to produce the data, application program interfaces (APIs)—everything. All NASA asks in return is that data users indicate their interest and need for these data through a simple registration process before they download data.

For those interested in downloading and using the more than 22 petabytes of Earth science data archived by NASA’s Earth Observing System Data and Information System (EOSDIS) and managed by the Earth Science Data and Information System (ESDIS) Project, this registration is called Earthdata Login. Earthdata Login provides numerous benefits for data users and helps ESDIS Project managers improve EOSDIS data and services. A key benefit of registration for users is that it allows the user to be contacted with information about the data products and services they use. This information includes data updates, new data releases, and alerts about potential data quality issues, along with notices about new services or service outages.

Earthdata Login also helps improve how the ESDIS Project manages NASA Earth science data. During the registration process, data users are requested to note their fields of study, affiliation, and country in which they reside and/or conduct research. This information, combined with metrics on the specific data products being downloaded and how often they are downloaded, helps the ESDIS Project prioritize the development of new products and services and improve existing products and services. This ensures that NASA Earth science data and services constantly evolve to meet changing user needs.

An important point is that user registration via Earthdata Login is only required for downloading or retrieving EOSDIS data, such as from a Distributed Active Archive Center (DAAC). Registration is not required for non-data related activities, such as reading articles on the Earthdata website or on DAAC websites, or for using EOSDIS Worldview to explore data imagery.

Establishing an Earthdata Login account is easy. Simply go to the Earthdata Login website (urs.earthdata.nasa.gov) and click the “Register” button, which is next to the green “Log In” button under the Password entry box. Fill in the required boxes (indicated with a red asterisk), then click on the “Register for Earthdata Login” green button at the bottom of the page. An email with instructions for activating the registration completes the process.

Along with the current benefits noted above, future enhancements to Earthdata Login will allow EOSDIS data users to more efficiently access and download data from multiple DAACs by maintaining a user’s spatial and temporal data search parameters between platforms. For example, if a data user goes to the National Snow and Ice Center DAAC (NSIDC DAAC) to download
Soil Moisture Active Passive (SMAP) products covering North Africa between 1 June and 30 September 2017, these search parameters will be remembered if they want to download Moderate Resolution Imaging Spectroradiometer (MODIS) data from the Land Processes DAAC (LP DAAC), greatly reducing the time needed to find relevant data. Similar planned enhancements will further streamline the process for finding data products.

EOSDIS and the ESDIS Project are committed to enabling the most efficient use of NASA Earth science data. User feedback helps make this possible, and comments, suggestions, and recommendations for improving Earthdata Login are welcome and encouraged. NASA data belong to everyone; making these data easy to access and use is not only a vital part of NASA’s mission, it also is a key EOSDIS objective.

Collaborators and Colleagues: Summer Interns are an Integral Part of the EOSDIS Team

More than 30 EOSDIS interns are supporting tasks at the ESDIS Project Office, NASA’s Science Data Systems Branch, and at DAACs across the country this summer.

The word “collaborate” comes up often when talking with the 2017 Earth Observing System Data and Information System (EOSDIS) summer interns and their mentors. Of course, this is expected when 32 high school, college, and graduate school students spend 10 weeks working together accomplishing vital tasks and projects that help further the EOSDIS mission. After 10 weeks, though, another word also becomes quite common for both interns and EOSDIS staff—“colleague.”

This summer’s EOSDIS interns are supporting critical tasks at five EOSDIS Distributed Active Archive Centers (DAACs), the Earth Science Data and Information System (ESDIS) Project Office, and the NASA Science Data Systems Branch. For many of the interns, this is their first opportunity to apply their knowledge to real-world tasks in a professional environment. For some, this also may be the start of a NASA career.

A NASA career “is definitely a possibility” for ESDIS Project Office intern Jonathan Rosenberg, a senior mechanical engineering and computer science double-major at the University of Maryland-College Park. Jonathan is using the Grafana analytics platform to create graphs and prototype dashboard set-ups to enable visual interaction with ESDIS Project metrics. “At first I thought it would be distracting sharing a conference room [with other ESDIS Project interns],” he says. “But it’s really not. It’s a collaborative environment that helps us focus on our work. We’ll often bounce ideas back and forth, or take a break to help each other with our code.”

ESDIS Project Office interns Ben Banavige, a rising senior statistics major at Harvard University, and Karthik Garimella, an entering computer engineering Master’s student at Washington University in St. Louis, are working together on a project converting Godard Earth Sciences Data and Information Services Center (GES DISC) data recipes into Jupyter notebooks. From there, they hope to put these data recipes in the cloud. “I want to help Earth scientists get more comfortable with cloud computing; this is a big part of our project,” Karthik says.
Two of this summer’s ESDIS Project Office interns are returning interns. Zac Lamb, a computer science Ph.D. student at the University of Cincinnati, was an intern last summer and is working on a text mining effort to provide better metrics for overall ESDIS Project data usage. “We have thousands of scientific articles that have used NASA data and many of these articles do not provide a full or correct citation of the datasets,” he observes. “From my work here, I hope I can apply the new tools and techniques I learn during this summer to my own thesis research.”

Meanwhile, this is the third summer internship at the ESDIS Project Office for Dionne Wright, a rising junior at Penn State University majoring in information science technology. Dionne started as an ESDIS Project intern when she was in high school and continues her work converting paper documents in configuration change request (CCR) folders into an electronic database. “I can take what I’m learning here at NASA and connect it with my classwork,” Dionne says. “It really helps connect the dots between school and the real world.”

Working with Dionne on this configuration management project is Adetumara (Tumie) Olateru, a rising high school senior at Chesapeake Math and Information Technology Academy in Laurel, MD. “It’s a great experience,” Tumie says. “Just being around here and meeting new people; getting to learn new processes and explore career paths. You feel really good about yourself when you get into NASA.”

These six interns supporting the ESDIS Project Office are among the almost 450 interns working at NASA’s Goddard Space Flight Center in Greenbelt, MD, this summer. They are part of the hundreds of NASA interns supporting tasks and missions at NASA centers, laboratories, and facilities across the country.

NASA internships are available throughout the year and provide unique NASA-related research and operational experiences for students as well as educators. Each fall, NASA puts out a call for projects to be added to the NASA Interns, Fellows & Scholars One Stop Shopping Initiative (OSSI) website. Through OSSI, students can search and apply for all NASA internship, fellowship, and scholarship opportunities in one location.

NASA internships are available for high school, undergraduate, and graduate students. High school interns must be at least 16 years old and a current sophomore, junior, or senior; undergraduate and graduate students must be accepted or enrolled full-time in an accredited U.S. college or university in a degree-granting course of study “appropriate to NASA’s long-term professional workforce needs.” Along with a GPA of at least 3.0 on a 4.0 scale, perspective interns must be U.S. citizens. (NASA internships for international students from select countries are available through the NASA International Internship Program.)

Every intern works under the supervision of a mentor, and sometimes more than one mentor. Mentors play an important role in not just helping interns with their assigned projects, but also passing along the experience and knowledge they have acquired through their years at NASA.

Dr. Justin Rice, an ESDIS Project systems engineer, is the primary mentor for Ben and Karthik, and assists ESDIS System Architect Dr. Chris Lynnes, who is the primary mentor for Zac and Jonathan. To Dr. Rice, learning, working hard, and having fun are the highest priorities of a successful internship. He also believes it’s important to give the interns insights into his experiences. “Determining what’s next after school can often prove challenging,” he says. “During my time as an intern, I sought out the advice of seasoned professionals to help me along, and they were gracious enough to oblige my requests. Now, in many ways, the roles are reversed. I do my best to pass on things
that I have learned as a NASA intern, graduate student, researcher, engineer, and software developer in an effort to assist our students in their careers.”

Dr. Rice sees his interns as collaborators, and notes that he is learning from them just as they are learning from him. “Technology rapidly evolves and changes, and it’s good to be open to the fresh ideas and perspectives of our interns. Sometimes I advise them; other times, they advise me,” he observes. “Two-way mentoring is the best way to cultivate an environment conducive to learning; everyone benefits.”

Along with the six EOSDIS interns supporting the ESDIS Project Office, interns also are supporting projects at NASA’s Science Data Systems Branch at Goddard and at the EOSDIS’ discipline-specific DAACs. Nine interns are at the Atmospheric Science Data Center (ASDC) in Hampton, VA; five are supporting projects at the Alaska Satellite Facility (ASF) DAAC in Fairbanks, AK; five are working at the National Snow and Ice Data Center (NSIDC) in Boulder, CO; and two are supporting projects at the Socioeconomic Data and Applications Center (SEDAC) in Palisades, NY.

Three interns also are at the GES DISC at Goddard: Meghan Andrews, a rising senior geography major at George Mason University in Fairfax, VA; Kendy
Edmonds, a rising senior studying unmanned aircraft systems (UAS) at Kansas State Polytechnic in Salina, KS; and Zachary Bruick, an entering atmospheric science Master's student at Colorado State University in Fort Collins, CO. Zachary also is one of 12 recipients this summer of a John Mather Nobel Scholarship, which was established by Goddard Senior Astrophysicist, Goddard Fellow, and Nobel Laureate Dr. John Mather and is awarded to selected Goddard interns. The scholarship provides a travel allowance towards the cost of presenting research papers at professional conferences.

GES DISC principal support scientist Dr. Jennifer Wei is the mentor for Meghan and Zachary, who are working on geographic information system (GIS)-based projects. “Because we’re a NASA data center, I hope they gain an understanding of how we do data—how the data are structured, how we handle data, how data are processed, and how data can be visualized in ways people can understand them,” she says.

Dr. Wei notes that along with their individual projects, GES DISC interns attend a Python programming language boot camp, seminars, and weekly meetings, and are encouraged to create an internship experience that best suits their needs and interests. “If they want to do research, we let them do research. If they want to be more engaged with programming, we let them do this,” she says. “The interns come here with a dream and I hope they leave here with their dream fulfilled.”

Of course, the intern experience involves more than simply working on tasks. Interns participate in many enrichment, networking, and social events. Intern activities at Goddard include a film festival, tours through laboratories and facilities, and the opportunity to participate in the same lectures and programs open to all Goddard employees.

As the interns prepare to return to their schools, they leave with not only a sense of accomplishment, but also a greater understanding of NASA and the EOSDIS—and their potential future roles in both.

For ESDIS Project intern Tumie, who would like to attend the United States Military Academy at West Point after high school, seeing NASA engineering in action is one highlight of his summer. “It’s great to get an overview of the whole ESDIS Project,” he says.

Karthik also is enjoying his summer. “It’s just awesome to be at a place like this; every week we have tours and meetings we can go to,” he says. “Yesterday we went on a tour of the robotics laboratory and saw behind the scenes about how they’re going to work with satellites in space. This is just really cool for me as a computer engineer.”

Karthik’s collaborator Ben agrees. “This seems like a normal office, but right down the hall is where they’re controlling orbiting Earth satellites,” he observes. “You can see where the [James Webb Space Telescope] was built. It’s brought a much greater appreciation for NASA.”

This appreciation is reciprocated by the mentors, who note that the interns are not only accomplishing tasks that help further their work, but sharing new knowledge being taught to the next generation of EOSDIS staff. “They are part of our family; we treat them as a colleague,” says Dr. Wei.

Dr. Rice concurs, and notes the special nature of working as a NASA intern. “I’m pretty sure the relationships they form here will last far beyond this summer,” he says. “You meet someone at an internship and it’s definitely not the end, but just the beginning.”
USER PROFILES:

NASA Earth Science Data User Profiles highlight our diverse end-user community worldwide and show you not only how these data are being used for research and applications, but also where these data are being used – from the plains of West Texas to the Sea of Oman and everywhere in between. You’ll also learn where you can download the data sets in each feature. [https://earthdata.nasa.gov/user-resources/who-uses-nasa-earth-science-data-user-profiles](https://earthdata.nasa.gov/user-resources/who-uses-nasa-earth-science-data-user-profiles)

**Dr. Sergio Signorini**

Who uses NASA Earth science data? Dr. Sergio Signorini, to study the impact of climate variability on phytoplankton and Earth’s oceans.

Dr. Sergio Signorini, Principal Research Scientist, Science Applications International Corporation (SAIC)/NASA Goddard Space Flight Center

Research interests: The ocean carbon cycle and the impact of climate variability on the ocean’s physical and biological properties.


**Dr. Rachel Albrecht**

Who uses NASA Earth science data? Dr. Rachel Albrecht, to pinpoint global lightning hotspots.

Dr. Rachel Albrecht, Professor, Department of Atmospheric Sciences, University of São Paulo, Brazil

Research interests: Cloud electrification, lightning, nowcasting of severe weather, cloud-aerosol-precipitation interactions (CAPI), and local circulation effects in deep convection.


**Dave Johnson**


David M. Johnson, Senior Geographer, United States Department of Agriculture (USDA)/National Agricultural Statistics Service (NASS)

Research interests: Using remotely-sensed data to provide quantitative estimates of crop production for major U.S. commodities, such as corn, soybeans, cotton, wheat, and rice.

ANNOUNCEMENTS

MOPITT Data now Available in LANCE

Global carbon monoxide data from the MOPITT instrument aboard NASA’s Terra satellite are the newest near real-time products available through the Land, Atmosphere Near real-time Capability for EOS (LANCE) system.

Carbon monoxide (CO) is a colorless, odorless, and tasteless gas that is a by-product of incomplete combustion of carbon-based fuels such as coal, wood, and natural gas. While it is best known for the deadly effects it can have in homes with faulty ventilation or appliances, it also impacts Earth’s atmosphere. Global CO concentrations vary through activities such as seasonal agricultural burning as well as from natural events such as wildfires and volcanic eruptions. Higher atmospheric CO concentrations can affect oxygen transport in the blood and can increase levels of ground-based ozone, which also can lead to health problems.

Knowledge about global CO concentrations increased tremendously with the launch of the Measurements of Pollution in the Troposphere (MOPITT) sensor aboard NASA’s Terra Earth observing satellite in 1999. Since becoming operational in 2000, MOPITT has provided an invaluable record of daily, monthly, and annual global CO concentrations and movement. The MOPITT Retrieved CO (Thermal Infrared Radiances) product is now available in near real-time (NRT) through the Land, Atmosphere Near real-time Capability for EOS, or LANCE system.

The MOPITT CO product is generally available through LANCE within 180 minutes of the sensor overflight. While MOPITT NRT data do not have the extensive processing necessary for use in scientific research, their rapid availability make them a vital resource for forecasting air quality and atmospheric chemistry.

The MOPITT instrument, which was funded by the Canadian Space Agency and built by a consortium of Canadian companies, is one of five instruments aboard NASA’s Terra Earth observing satellite. MOPITT measures concentrations of CO in the troposphere, which is the lowest layer of Earth’s atmosphere and extends from ground level to 9-16 km (5.5-10 miles) above the surface. CO remains in the troposphere for about a month, according to NASA, which means it persists long enough to be transported long distances by high altitude winds, but not long enough to mix evenly throughout the atmosphere. Likewise, CO does not accumulate in the atmosphere like carbon dioxide (CO₂). MOPITT has a sensor swath of 640 km (almost 400 miles) and can measure CO concentrations in 5 km (about 3 mile) layers down a vertical atmospheric column. Each MOPITT pixel is about 22 km (about 13.7 miles) square. Complete global coverage is provided every three days.

These two images created from MOPITT data show the yearly averages of tropospheric CO concentrations at 3,700 meters (slightly more than 12,000 feet) for 2000 (left image) and 2014 (right image). Yellow areas indicate low CO concentrations; orange/red areas indicate higher CO concentrations (gray areas north of India and in Antarctica are areas where no data are available). Note the significant decrease in tropospheric CO between 2000 and 2014. As noted by the MOPITT team, this is likely due to more intense agricultural burning and wildfires in 2000. NASA Earth Observatory images by Jesse Allen and Joshua Stevens, using data from the National Center for Atmospheric Research (NCAR) and the University of Toronto MOPITT Teams.
LANCE is part of NASA’s Earth Observing System Data and Information System (EOSDIS), and distributes NRT data products from almost a dozen satellite-borne instruments. EOSDIS provides end-to-end capabilities for managing NASA’s Earth science data. These data represent some of the most complex and diverse Earth science data sets on the planet from satellites, aircraft, field measurements, and numerous other EOSDIS programs. The primary services provided by EOSDIS are data archive, management, and distribution; information management; product generation; and user support services. These services are managed by NASA’s Earth Science Data and Information System (ESDIS) Project.

Terra is NASA’s flagship Earth observing satellite. Data from Terra’s five instruments are incorporated into more than 80 standard data products for use in scientific research. These standard products are archived and disseminated through several EOSDIS discipline-specific Distributed Active Archive Centers (DAACs). Originally designed to last six years, Terra is still going strong in 2017.

To download MOPITT NRT data or for more information about LANCE, visit the EOSDIS LANCE page.  

https://youtu.be/l205NLJxDGY

7/19/17
Global Navigation Satellite System (GNSS) Radio Occultation Science and Applications

https://youtu.be/Y_f0L62TleE

8/30/17

https://youtu.be/8VHzCiupTVo

10/19/17
Using NASA’s AppEEARS to Slice and Dice Big Earth Data
Getting Started with VIIRS Surface Reflectance Data

Part 1: All about Accessing the Data
https://youtu.be/x3qxxL-zFVM

Part 2: Using the Data
https://youtu.be/eORFLFdnu4

Choosing a Data Access Tool: NASA AppEEARS Area Sampler
https://youtu.be/NUM4znhkzI
Latest NASA Earthdata Images

Larsen-C Ice Shelf Calving
https://earthdata.nasa.gov/larsen-c-ice-shelf-calving
(Published on 7/31/17)

Beaufort Sea, Alaska
https://earthdata.nasa.gov/mastheads

Fires and Smoke in Canada
https://earthdata.nasa.gov/fire-and-smoke-in-canada
(Published on 8/21/17)

Sensing Vegetation at Mono Lake, California
http://earthdata.nasa.gov/mastheads

Hurricanes Katia, Irma and Jose
https://earthdata.nasa.gov/hurricanes-katia-irma-and-jose

Larsen-C Ice Shelf Calving
Beaufort Sea, Alaska
Fires and Smoke in Canada
Sensing Vegetation at Mono Lake, California
Hurricanes Katia, Irma and Jose

Sensing Lightning from Space
http://earthdata.nasa.gov/mastheads
(Published on 9/5/17)

Hurricane Irma Approaching the Leeward Island
https://earthdata.nasa.gov/hurricane-irma-111-approaching-the-leeward-island
(Published on 9/11/17)

Sensing Soil Moisture after Hurricane Harvey
https://earthdata.nasa.gov/mastheads

Fires in Northern California
https://earthdata.nasa.gov/fires-in-northern-california-usa

Tree Mortality from Fires and Bark Beetles
http://earthdata.nasa.gov/mastheads
(Published on 10/10/17)

About Us
Discover NASA Earth Observing System Data and Information System (EOSDIS) data, information, services, and tools. Tap into our resources! To learn more, visit our website at https://earthdata.nasa.gov

Connect with Us
Follow NASA Earthdata for Earth science data, services and tools, news, and updates on social media:

Twitter: https://twitter.com/NASAEarthdata
Facebook: https://www.facebook.com/NASAEarthdata
Google+: http://google.com/+NASAEarthdata

Subscribe to EOSDIS News
https://lists.nasa.gov/mailman/listinfo/eosdis-news

Feedback?
Email: support@earthdata.nasa.gov

Webinars, Tutorials, and Recipes
Data discovery and data access webinars, tutorials, and short data “how-to” recipes on YouTube at http://www.youtube.com/c/NASAEarthdata.
View schedule and sign-up to receive webinar announcements https://earthdata.nasa.gov/user-resources/webinars-and-tutorials