

# **CDDIS FAIR Commitment and NASA's Year of Open Science**



Justine Woo<sup>1</sup>, Nathan Pollack<sup>1</sup>, Taylor Yates<sup>1</sup>, Benjamin Patrick Michael<sup>2</sup>

<sup>1</sup>Science Systems and Applications INC./ NASA Goddard Space Flight Center, Code 61A, Greenbelt, MD, USA (<u>Justine.y.woo@nasa.gov</u>) <sup>2</sup>NASA Goddard Space Flight Center, Code 61A, Greenbelt, MD, USA

# ABSTRACT

The Crustal Dynamics Data Information System (CDDIS) is one of twelve NASA Earth Observing System Data and Information System (EOSDIS) Distributed Active Archive Centers (DAAC) and supports the space geodesy and geodynamics community through the International Association of Geodesy (IAG) services. As an EOSDIS DAAC, the CDDIS is required to meet best archival practices including the Findability, Accessibility, Interoperability, and Reuse (FAIR) Guiding Principles for scientific data management and stewardship which ultimately serves data and product providers and users. This poster will introduce some common tools to ensure clarity for users on how these best practices support their efforts and how to utilize and find the tools built to support them. These include ensuring data is findable via the use of Digital Object Identifiers (DOIs) and facilitating discovery using the Earthdata Search API and NASA's Earthdata Search. As the CDDIS continues to grow during NASA's year of open science, the CDDIS will start migrating to the cloud further ensuring clarity and accessibility of data.

# THE CDDIS AND GEODESY DATA/PRODUCTS

### **Techniques**:

- Global Navigation Satellite Systems (GNSS)
- Laser ranging
  - Artificial satellites (SLR)
  - Lunar (LLR)
- Very Long Baseline Interferometry (VLBI)
- Doppler Orbitography and Radio-positioning Integrated by Satellite (DORIS)

### Data

- Observation data
- Broadcast orbit information
- Meteorological parameters

### Products

- Precise orbits and orbit predictions
- Clock products
- Reference frames
- Troposphere and ionosphere products
- Displacement time series and Plate boundary aseismic transients
- Water storage time series



Figure 1: Number of Distinct Users of the CDDIS Archive By Country



SLR	٩	14 Matching Collections			
* * *		Showing 14 of 14 matching collections	Export	J≞ Sort	i≣ View
⊞ 4 <del>‡</del>		Ground-Based Satellite Laser Ranging (SLR) Obse	ervation Data		0
▼ Filter Collections		(normal points, daily, 24 hour files) from NASA C	DDIS	No	o image
Categories		109,455 Granules 2016-01-01 ongoing		a)	allable
		This dataset consists of ground-based Satellite Laser Ranging observation			
Features	^	data (normal points, daily 24 hour files) from the NASA Crustal Dynamics			
Available from AWS Cloud		GEOSS • CDDIS SLR data daily npt v1 - NASA/GSFC/SED/ESD/GGL/CDDIS			
Customizable					
Map Imagery		Ground-Based Satellite Laser Ranging (SLR) Obse	rvation Data		0
Keywords		(normal points, hourly files) from NASA CDDIS		No	o image
		24,408 Granules 2010-01-01 ongoing		av	/ailable
Platforms	$\sim$	This dataset consists of ground-based Satellite Laser Rar	nging observati	on	
		data (normal points, hourly files) from the NASA Crustal	Dynamics Data	э	
Instruments	~				
Organizations 1 Selected	~	GEOSS • CDDIS SLR data hourly npt v1 - NASA/GSFC/SED/ESD/	GGL/CDDIS		

Figure 2: Earthdata Search – CDDIS SLR Collections



New Users: Earthdata Search

Earthdata Search provides a visual tool for scientists and engineers to search for data available from all the DAACs. The CDDIS currently has more than 270 collections which users can find by narrowing down the "Organization" to the CDDIS. For more information on how to narrow down your search, please see the Interoperable section.

#### **CDDIS Archive Search**

create queries that can be integrated into custom software allowing a wide range of searches to be built. Below you will find examples ranging from very broad searches to very specific; these examples will be shown both usin standard tool such as cURL as well as custom python software ese examples mix and match various search parameters and return types. All of the examples should be nterchangeable with some minor modifications made to the location of the CDDIS data you need. Collections = data and derived product sets Granules = files Python Code Examples cURL Examples Get all CDDIS Collections Get all CDDIS Collections Get CDDIS - GNSS Get all VLBI Collection Collections Get CDDIS - DORIS roducts Collectio Get CDDIS Granules for a Date Range Get CDDIS Granules for Granules a Date Range

Figure 3: CDDIS Website Archive Search

#### General Use: The CDDIS Website and Earthdata Search API



The CDDIS website contains information on the space geodesy data it has available, broken down by technique. Collection information is under the "Data and Products" tab. This website is useful for those who know that they need geodesy data but need more information.

The Earthdata Search API also allows users to get a listing of the CDDIS collections and granules. To simplify the process for users, the CDDIS website has samples queries and guidance.

#### Stream Status Stream Late



Figure 4: CDDIS Website GNSS Real-time Data Streaming Latency Status Specific Use: Real-time (NTRIP) System Status

The CDDIS has been supporting GNSS real-time data streaming since 2018 via the CDDIS-Caster NTRIP broadcaster. This year, to provide additional usability and clarity to users, the CDDIS created a real-time map showing which streams are active and their latencies.

CDDIS-Caster NTRIP
Broadcaster



#### INTEROPERABLE REUSABLE ACCESSIBLE CDDIS The CDDIS ensures the Earth Science Keywords Structure On October 2020, the To ensure NASA's Archive of Space Geodesy Data NASA's Archive of Space Geodesy Data Home About CDDIS Programs Publications Citing our Data CDDIS Text Search reusability of its data and CDDIS shifted away interoperability with The Earth Science Keywords have a six-level keyword structure with the option fo GNSS 🕨 About CDDIS Data and Products Techniques Programs eference frame product a seventh uncontrolled field. Category and Topic levels define how the keywords products by ensuring they from FTP downloads other DAACs, the The CDDIS archive of laser ranging products consists of <u>solutions</u> derived by analysis centers in support of the ILRS. Products derived from these SLR observations include precise satellite ephemerides, station positions an SLR 🔻

Parent Directory	
🦲 campaign	
🦲 daily	
ighrate	
iourly	
I2ctest	
I5test	
rinex3test	
🧮 rt	
satellite	

Figure 5: CDDIS GNSS Archive via Web Browser



For more information on archive access and download via command line:

ormation on ss and download d line:

line.

and transitioned to

supports human and

computer readability.

which can be created

for free. The archive

browser and command

can be accessed via

Access requires an

Earthdata account

HTTPS which

are organized and the associated Earth science discipline within the hierarchy. The Term and Variables levels define the subject area, measured variables/parameters, and the hierarchical-type relationship for the subject area.

Keyword Level	Example	
Category	Earth Science	
Торіс	Atmosphere	
Term	Weather Events	
Variable Level 1	Subtropical Cyclones	
Variable Level 2	Subtropical Depression	
Variable Level 3	Subtropical Depression Track	
Detailed Variable	(Uncontrolled Keyword)	

Figure 6: Screenshot from the GCMD Website



CDDIS follows the Global Change Master Directory (GCMD) structure and keywords to provide insight into its data and products. This allows for clearer

connections between different DAACs and their data/products and is primarily used for Earthdata Search. velocities of sites in the ILRS network, and EOPs (polar motion and rates, length-of-day). ILRS analysis centers deliver the station position/EOP product to the ILRS data centers, including the CDDIS. The ILRS combination centers retrieve these solutions and produce the combined ILRS A and B solutions (generated by the primary a backup ILRS combination centers, respectively) which are in turn archived in the CDDIS. Prior to 2012, ILRS analysis centers generated the products on a weekly basis; today, these products are generated daily. Station positions and EOP date from 1976 through the present. SLR products are available from the CDDIS in subdirectories by year and date.

unar laser ranging data	subdirectories by year and date. Reference Frames:				
luct holdings					
eference frame	<ul> <li>ITRF2020, DOI: <u>https://doi.org/10.5067/SLR/slr_itrf20200_repro2020_001</u></li> </ul>				
ITRF2020	Product formats:				
recise orbits	<ul> <li>Software INdependent EXchange (SINEX) format</li> </ul>				
redicted orbits	The starting directory for these files is: <u>https://cddis.nasa.gov/archive/slr/products/pos+eop/</u>				
orts					
ted links	Append the following directory and file names to the starting directory:				
•	YYYY/YYMMDD/ACNAME.pos+eop.YYMMDD.v###.TYP.gz				
NS 🕨	as described in the table below.				
er products					
	Code	Meaning			
ive search information	YYYY	4-digit year			
ive access information	YY	2-digit year			
	ММ	2-digit month			

Figure 7: CDDIS Website DOI Example

Normal point data

Full-rate data

can be cited via Digital Object Identifiers (DOIs). DOIs can be added to references in papers and publications to ensure traceability. The DOIs can be found on the CDDIS website by selecting the "Data and Products" tab, followed by the specific data or product you hope the cite.

The link will either lead directly to a landing page with the DOI or the DOI link will be available which leads to the landing page. If you have any issues finding an applicable DOI please contact the CDDIS.

## DATA DESCRIPTION FILES (DDF)

# CDDIS file definition Version 1.0 of 2019.10.22

Figure 8: Sample DDF

# 2019.07.16 Leonid P	etrov created		
# 2019.08.29 Dirk Beh	rend updated		
# 2020.08.07 Dirk Beh	rend changed Format_file_name location		
#			
Short_description:	24-hour master file		
Long_description:	Master file for 24-hour sessions		
Format_file_name:	/vlbi/ivsformats/master-format.txt		
Reference:	n/a		
DOI:	n/a		
Filenaming_scheme:	master@{date{%y}}.txt		
File_location:	/vlbi/ivscontrol		
Product_ID:	VLBI_AUX_MASTER_24		
Data_type:	VLBI		
Data_content_type:	Misc		
Data_format:	ascii		
Validate_proc:	master.py		
Magic:	## Master file format version 1.0	2001.08.21	CCT&NRV
Compression_type:	none		

In 2020, the CDDIS collaborated with the International VLBI Service (IVS) to create a new processing structure for VLBI data. The software architecture emphasizes a modular design maintained using DDFs with quality checking software provided by the IVS. A sample DDF for processing VLBI master files is given in Figure 8.

The file contains all the information needed to identify and process the file, including some of the metadata needed for findability.

## **OPEN SCIENCE - EARTHDATA CLOUD TRANSITION**

#### Current Interaction



Figure 10: Current User Interaction with the CDDIS Server

Once the CDDIS has shifted to the Earthdata Cloud, users will have access to the following benefits:

• Access to full features on Earthdata Search including searchability and download directly from Earthdata

The CDDIS is transitioning to the Earthdata Cloud in order to fully realize NASA's goals for open science. The primary focus of these efforts is to ensure the accessibility of NASA data by reducing the barriers to entry within the science community while increasing traceability and reproducibility of research.

**Future Options** 



## PROCESSING V4 - PRE-CLOUD



Given the simplicity and effectiveness of DDFs, the CDDIS has been working to transition its processing software to this format for all the techniques it supports.

The new architecture also considers the CDDIS's future transition to the Earthdata Cloud by structuring its software to fit AWS services.

The new architecture have the following improvements that providers and users will experience:

- Automatic processing after upload; processing will no longer be run at discrete time intervals
- Faster availability on Earthdata Search
- Simplified process for adding or removing data or products from the processing stream
- Ability to work with the CDDIS data and products in the cloud – downloads are no longer required, reducing the need to invest in onpremises resources
- Easier identification of collocated data from across the NASA DAACs
- Access to new tools for subscribing and analyzing data
- \* Please note that, even as the CDDIS transitions to the cloud, users will still be able to download data to their servers and the original archive structure will still be maintained. EOSDIS is also investing in helping scientists transition to the cloud through training initiatives

Figure 10: Future Options for User Interaction with the CDDIS Server

The CDDIS is also working to ensuring the following as data and products are migrated to the cloud:

- Ensuring the data and products have a DOI and landing page, especially for older data sets
- Reorganize the data/products into clearer collections with the input from representatives of the techniques supported
- More through QC is created for each of the file
- types
- Data can be downloaded based on temporal specifications