CDDIS Services to the ILRS

Justine Woo (1), Sandra Belvins (1), Benjamin Patrick Michael (2), Taylor Yates (1), Rebecca Limbacher (1), Nathan Pollack (1)

(1) Science Systems and Applications INC./ NASA Goddard Space Flight Center, Code 61A, Greenbelt, MD, USA (Justine.y.woo@nasa.gov)
(2) 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, which includes the International Laser Ranging Service (ILRS). 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. These best practices and artifacts built to support them are not always visible to users although utilizing them ultimately benefits the community, for example citing with Digital Object Identifiers (DOIs) to ensure traceability and that proper credit is given to the data and product creators. To address this gap, this paper steps through the CDDIS's ingest, quality control, and archive systems that ensure data are searchable, citable, and that credit is given to providers. The CDDIS will then review ESDIS tools for exploration of its archive. Beginning next year, the CDDIS will start migrating to the cloud to support open science.

1. Introduction

The EOSDIS serves as a large, centralized data system for NASA's earth observing systems that seeks to reconcile the needs of developers and users. The specific goals of EOSDIS evolve as best practices on data systems become clearer with recent goals focusing on increasing interoperability among DAACs, providing consistency in the user experience, and an increased focus on security. For several years, EOSDIS has focused its efforts on ensuring its DAACs meet the FAIR data principles which ensure data is Findable, Accessible, Interoperable, and Reusable.

FAIR was first established to clarify what constitutes "good data management" and to provide routes toward supporting the reuse of scholarly data and to improve transparency, reproducibility, and reusability of research among other goals[1]. For the CDDIS community, one of the primary benefits FAIR may bring is ensuring proper credit is given to data and product creators.

The EOSDIS set up the following outlines to meet FAIR requirements for DAACs:

Principle	**EOSDIS Actions**[2,3]
Findable	- Adoption and on-going assignment of Digital Object Identifiers (DOIs) to all datasets

	- Facilitating discovery using the Common Metadata Repository (CMR) with rich metadata
Accessible	- Use of https and provision of both human- and machine-accessibility - Data and services consistent with NASA's free and open policy - User login for tracking usage metrics and informing users about changes to data and/or services - Landing pages for superseded versions of collections and providing pointers to later versions
Interoperable	- Controlled keyword vocabulary of Global Change Master Directory - Structured metadata in a Unified Metadata Model (UMM) - Use of internationally accepted standard formats for data and metadata
Reusable	- Metadata fields and documentation references provided in the datasets and their landing pages to inform users about the scope and limitations of the data - Metadata are associated with detailed provenance - Documentation of known issues - Quality flags and indicators to enable users to filter data before use

The CDDIS is working to meet each of these goals with progress being made in the creation of DOIs, increasing findability by using CMR which powers Earthdata Search, and documenting existing issues with older data.

2. Findable and Reusable: CDDIS Website and DOI Landing Pages

The CDDIS website contains information about the SLR data and products archive including landing pages describing each of the filetypes, version information, links to its location in the archive, and digital object identifiers (DOIs).

DOIs are globally unique, persistent identifiers that chronicle the provenance of digital objects (in our case data and product sets) and allows them to be cited. This ensures proper credit is given to data and product creators and archive and data centers. DOIs are included in the data/product metadata, helping increase the use of citations in papers and journals.

In total, the CDDIS has published **120 DOIs** representing Global Navigation Satellite System (GNSS), Doppler Orbitography and Radio-positioning Integrated by Satellite (DORIS), and Satellite Laser Ranging (SLR) data and derived products.

A list of SLR related DOIs are below with links to the landing page which contains the data citation information:

File Type	**DOI**
SLR hourly normal point data	doi: 10.5067/SLR/SLR_DATA_HOURLY_NPT_001

SLR daily normal point data	doi: 10.5067/SLR/SLR_DATA_DAILY_NPT_001
SLR monthly normal point data	doi: 10.5067/SLR/SLR_DATA_MONTHLY_NPT_001
SLR monthly normal point data summary	doi: 10.5067/SLR/SLR_DATA_MONTH-LYSUM_NPT_001
SLR hourly full-rate data	doi: 10.5067/SLR/SLR_DATA_DAILY_FR_001
SLR daily full-rate data	doi: 10.5067/SLR/SLR_DATA_DAILY_FR_001
SLR monthly full-rate data	doi: 10.5067/SLR/SLR_DATA_MONTHLY_FR_001
SLR monthly full-rate data summary	doi: 10.5067/SLR/SLR_DATA_MONTHLYSUM_FR_001
SLR Satellite Orbit Prediction Product (multi-day, daily, and sub-daily files)	doi: 10.5067/SLR/SLR_ILRSORBPRED_001
SLR Analysis Center (AC) Position and ERP Product (daily files, generated daily)	doi: 10.5067/SLR/SLR_ILRSACSNX_001
SLR Final Station Position and ERP Product (daily files, generated daily)	doi: 10.5067/SLR/SLR_ILRSSNX_001
SLR products ITRF2020 Station Positions and Earth Orientation Parameters Time Series (REPRO2020)	doi: 10.5067/SLR/slr_itrf20200_repro2020_001
SLR Analysis Center (AC) Orbit Product (daily files, generated weekly)	doi: 10.5067/SLR/SLR_ILRSACORB_001
SLR Final Orbit Product (daily files, generated weekly)	doi: 10.5067/SLR/SLR_ILRSORB_001

Although the existing DOIs only cover a fraction of the data and products available at the CDDIS, the CDDIS is actively working towards creating additional DOIs, ensuring their availability via landing pages, and that data and products creators are added to the landing pages. The CDDIS is also working to ensure clarity and access to the existing landing pages. As new data and products become available, the CDDIS will work closely with the data providers to assign DOIs and create landing pages.

For more information on how to cite our data, please see:
https://cddis.nasa.gov/About/Data_citation_and_acknowledgment.html

3. Findable: Earthdata Search

NASA's Earthdata Search is a web-based interface that allows users to search and access data from all NASA's DAACs. The tool was created to help researchers find the right data sets and to create consistency in the tools and services available across the DAACs. Currently, the CDDIS data and derived product collections can be found through Earthdata Search but collections need to be downloaded via https from the CDDIS archive or via ftp-ssl from EOSDIS – see section 4.

The CDDIS archive can also be searched via the NASA Earthdata Search API with instructions on the following website: https://cddis.nasa.gov/Data_and_Derived_Products/CDDIS_Earthdata_Search.html

The CDDIS will be working to fully utilize the capabilities available using Earthdata Search. This effort has started with backlogging all data and products in the archive, so they can be found, and ensuring the metadata is available on NASA's Common Metadata Repository (CMR). CMR is a metadata system that catalogs all data and service metadata records for EOSDIS and serves as a primary component of Earthdata Search. CMR utilized a Unified Metadata Model (UMM) for services and variables – discussed in section 5 – which allows data to be cataloged, searched, and for relevant data to be extracted from archives.[5]

4. Access: CDDIS Archive and Download via HTTPs

On October 31, 2020, the CDDIS terminated the use of unencrypted ftp based on a NASA requirement set March 2018. This was done because unencrypted ftp did not meet security requirements set by NASA. To ensure users could continue to easily access its archive, https and ftp-ssl protocols are allowed. Both require users to have an Earthdata account before being able to download data.

Examples on how to download data are given here:
https://cddis.nasa.gov/Data_and_Derived_Products/CDDIS_Archive_Access.html.

5. Interoperable – UMM, GCMD Keywords

Various NASA DAACs supply metadata differently which makes searchability more difficult. To correct this, the Unified Metadata Model (UMM) translates between the different metadata standards available in CMR. UMM ensures collection and file (granule) metadata is collected and plays a core role in reusability by ensuring the origin details such as authorship, instruments, and versions are included.

To simplify and allow for interoperability, Global Change Master Directory (GCMD) keywords provide a controlled vocabulary. This ensures Earth science data, services, and variables are described in a consistent and comprehensive manner.[6]

6. Reusable – Documentation of Known Issues

Reusability via DOIs and landing pages is addressed under section 2. In addition to ensuring that data and collections can be cited, the CDDIS is working to document known issues with SLR CRD V1 data that were addressed over time but continue to exist in earlier data. In 2021, the CDDIS reprocessed all existing CRD V1 data to ensure its availability on Earthdata Search. After reprocessing, several issues were identified and presented in the following poster: https://agu2021fallmeeting-agu.ipostersessions.com/Default.aspx?s=A2-51-EF-5A-1B-B2-2B-F4-A7-97-6F-91-28-B0-B9-CA. A working group has been established to discuss how to document these known issues for future users.

7. Conclusion

Starting in 2023, the CDDIS will begin transitioning from a NASA on-site archive to a cloud-based system, Amazon Web Services (AWS), to meet EOSDIS's new infrastructure objectives. The migration is expected to take several years.

Currently, users need to download data from the CDDIS onto their computers or servers which requires, at times, substantial resources and support from additional infrastructure needs. By migrating to the cloud, users will still have the option to download data to their servers but will also be able to create instances in the cloud (similar to having a virtual machine) to download and perform their work without needing additional hardware on-site. This will enable users to have access to large volumes of data without investing in additional hardware and removes barriers to cross-DAAC data access and tools. In addition, users will be able to have a subscription download which retrieves files based on when a collection was last updated

References

Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

Ramapriyan, H. and Behnke, J., "NASA's Earth Observing System Data and Information System (EOSDIS) and FAIR - A Self-Assessment", vol. 2020, 2020.

Ramapriyan, Hampapuram; Behnke, Jeanne (2022): NASA's Earth Observation System Data and Information System (EOSDIS) and FAIR – A Self-Assessment. figshare. Presentation. https://doi.org/10.6084/m9.figshare.20315130.v1

Blevins, S., Tyahla, L., Michael, B., and Noll, C., "DOIs for geodetic data and derived product collections at the NASA GSFC CDDIS", AGU 2020, Online, Dec. 7-11.

What is CMR? https://www.earthdata.nasa.gov/eosdis/science-system-description/eosdis-components/cmr (2021) Accessed 30 December, 2022

Global Change Master Directory (GCMD) Keywords https://www.earthdata.nasa.gov/learn/find-data/idn/gcmd-keywords (2022) Accessed 30 December, 2022