DORIS Data Ingest Improvements at NASA CDDIS

Abstract: NASA's Crustal Dynamics Data Information System (CDDIS) continues to maintain and expand archive services for the International DORIS Service (IDS) and the Space Geodesy community. Data from Haiyang 2C (HY2C) were introduced in 2020, and from Haiyang 2D (HY2D) and Sentinel 6A (S6A) in 2021. Historical data have been reprocessed according to modern Quality Analysis (QA) standards, and metadata from these files have been collected. Having these metadata allows DORIS data and products to be discovered through Earthdata Search via the Common Metadata Repository (CMR). Planned future improvements include moving to a new QA architecture that utilizes common standards across all techniques (Data Definition file Formats, or DDFs), and transitioning the datasets from on-premises servers to a cloud deployment with Amazon Web Service (AWS).

[1]Taylor Yates (taylor.a.yates@nasa.gov)
[1]Justine Woo (justine.y.woo@nasa.gov)
[1]Sandra Blevins (sandra.blevins@nasa.gov)
[2]Patrick Michael (benjamin.p.michael@nasa.gov)

S6A / HY2C / HY2D Additions

Data and products from S6A, HY2C, and HY2D have been added to the DORIS archive at CDDIS. In this same time period, the Haiyang 2A (HY2A) satellite mission came to an end, and the International DORIS Service (IDS) now has 8 satellites in operation.

Data and products from these missions are already available at the CDDIS archive via the QR code in Figure 1.

Figure 1. DORIS archive at CDDIS (https://cddis.nasa.gov/archive/doris)

Reprocessed Data

DORIS data and products were reprocessed according to modern QA standards in September 2021. This included over 300,000 data and products. Reprocessing files from before 2019 was necessary to add their metadata to CMR and the CDDIS internal database. CMR is the source for Earthdata Search, and therefore adding these data and products to CMR is necessary to improve data visibility in the cloud environment.

This is useful not only for users working in a cloud environment themselves, but also useful for users who wish to browse the CDDIS data holdings in Earthdata Search.

Ingest Processing with DDFs

DDFs are a standardized format developed jointly by CDDIS and the International VLBI Service (IVS). After the successful implementation of DDFs for VLBI datasets, CDDIS is now in the process of converting our ingest software into DDF format for all techniques.

DDFs are text files that specify metadata about each dataset. These include CMR descriptions, file naming schemes, DOIs, file locations, data formats, compression types, etc. As shown in Figure 2, each technique has a class that inherits the general class, allowing universal QA to be performed on all files while technique specific QA is performed as required. The new architecture also allows external QA. Further specificity is allowed by the DDF, that performs specific QA for each dataset. External QA is defined as any QA script that is owned by a non CDDIS entity. As an example, the International VLBI Service (IVS) provides additional scripts that they request to be performed on incoming files. Uploads are added to the archive in the case of no fatal errors.

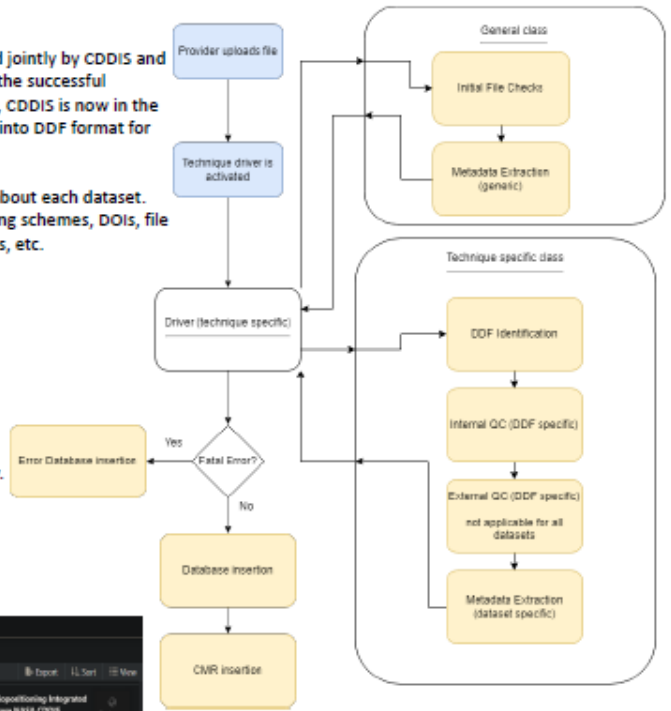

Figure 2. CDDIS Ingest Architecture using DDFs

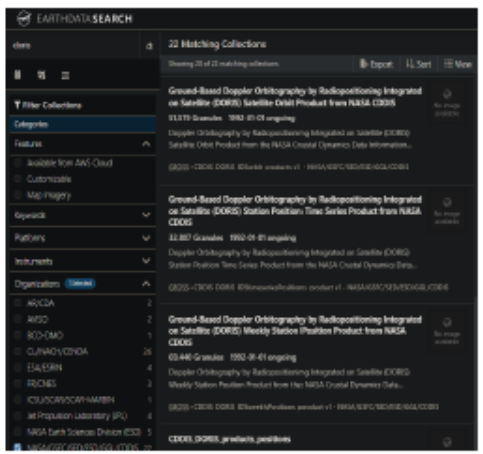

Figure 3. Earthdata Search results for CDDIS DORIS collections

Using this new ingest architecture allows CDDIS to process files more quickly as they arrive to the archive. CMR insertion is currently done in a batch process. With the new ingest architecture, CMR insertion will be moved to real-time.

This is particularly important in conjunction with the planned cloud deployment, because Earthdata Search will be used as a data discovery tool in the cloud environment. Earthdata Search has an intuitive and useful interface for data discovery, as shown in Figure 3.

DORIS datasets are planned to be moved to the new ingest architecture first, and other technique datasets will follow.

Planned Cloud Deployment

In partnership with other Distributed Active Archive Centers (DAACs) in NASA's Earth Science Data and Information System (ESDIS) Project, CDDIS is planned to transition from the current on-premises archive to a cloud-based system, NASA's Earthdata Cloud (EDC).

With data located in Amazon Web Service (AWS) S3 buckets, users can choose to either download the data via HTTPS or use it in place. This option bypasses the current requirement to download data from the CDDIS archive to a local machine.

Unnecessary downloads represent a waste of bandwidth and storage space. Removing the requirement to download data before use will greatly increase the future availability and usability of all Geodetic datasets going forward. While this is an improvement for users with an AWS account, others will still be able to download CDDIS data as before.

Work is planned to begin in late FY23, with this transition being planned to be complete for DORIS datasets by the end of FY24.

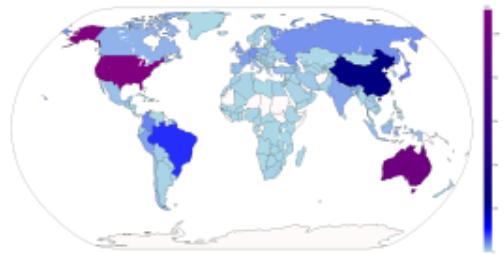

Figure 4. CDDIS 2021 Unique Users

For Feedback/Questions, contact Taylor Yates (Taylor.a.Yates@nasa.gov)

References and Information

How To Cite CDDIS Data:

C. Noll, The Crustal Dynamics Data Information System: A resource to support scientific analysis using space geodesy, Advances in Space Research, Volume 45, Issue 12, 15 June 2010, Pages 1421-1440, ISSN 0273-1177, DOI:10.1016/j.asr.2010.01.018.

(1) Science Systems and Applications, Lanham, MD, USA
(2) NASA Goddard Space Flight Center, Greenbelt, MD, USA