CDDIS Data Center Summary for the IVS 2017–2018 Biennial Report

Carey Noll, Patrick Michael

Abstract This report summarizes activities during the years 2017 through 2018 and the future plans of the Crustal Dynamics Data Information System (CDDIS) with respect to the International VLBI Service for Geodesy and Astrometry (IVS). Included in this report are background information about the CDDIS, the computer architecture, archive contents, and future plans for the CDDIS within the IVS.

1 General Information

The Crustal Dynamics Data Information System (CDDIS) has supported the archiving and distribution of Very Long Baseline Interferometry (VLBI) data since its inception in 1982. The CDDIS is a central facility that provides users access to data and derived products to facilitate scientific investigation. The full CDDIS archive of GNSS (GPS, GLONASS, Galileo, etc.), laser ranging, VLBI, and DORIS data is available online for remote access. Information about the system is available via the web at the URL https://cddis.nasa.gov. In addition to the IVS, the CDDIS actively supports other IAG services including the International GNSS Service (IGS), the International Laser Ranging Service (ILRS), and the International DORIS Service (IDS), as well as the International Earth Rotation and Reference Systems Service (IERS), and the IAG's observing system, the Global Geodetic Observing System (GGOS). The current and future plans for the CDDIS support of the IVS are discussed below.

The CDDIS is one of NASA's Earth Observing System Data and Information System (EOSDIS) Distributed Active Archive Centers (DAACs) (see https://earthdata.nasa.gov); EOSDIS data centers serve a diverse user community and are tasked to provide facilities to search and access science data and products. The CDDIS is also a regular member of the International Council for Science (ICSU) World Data System (WDS, https://www.icsu-wds.org) and the Earth Science Information Partners (ESIP, https://www.esipfed.org).

2 System Description

The CDDIS archive of VLBI data and products is accessible to the public through anonymous ftp (ftp://cddis.nasa.gov) and https (https://cddis.nasa.gov/archive).

2.1 Computer Architecture

The CDDIS is operational on a dedicated server, cddis.nasa.gov, located at NASA GSFC. The system is accessible to the international user community 24 hours per day, seven days per week. By the end of 2018, over 325 Gbytes of storage was devoted to VLBI-related data, derived products, and information.

The CDDIS computer facility is fully redundant with primary and secondary/failover systems utilizing a virtual machine (VM) based system, configured with

NASA Goddard Space Flight Center

CDDIS Data Center

IVS 2017+2018 Biennial Report

100 Tbytes of unified storage operating within the EOSDIS computer facility and network infrastructure. This system configuration provides reliable environment (power, HVAC, 24-hour on-site emergency personnel, etc.) and network connectivity; a disaster recovery system is installed in a different location on the GSFC campus for rapid failover if required. Multiple, redundant network switches are available to take full advantage of a high-performance network infrastructure by utilizing fully redundant network paths for all outgoing and incoming files along with dedicated network connections between its primary operations and its backup operations. The use of the virtual machine technology provides multiple instance services for a load balancing configuration and allows for VM instances to be increased or decreased due to demand. Furthermore, the VM technology allows for system maintenance (patching, upgrades, etc.) to proceed without any downtime or interruption to user access. The large, unified storage system will easily accommodate future growth of the archive and facilitate near real-time replication between its production and disaster recovery sites. The entire archive is also mirrored to traditional storage arrays for additional complete copies of the archive. This system architecture has allowed the CDDIS to achieve an uptime figure of over 99.9 in recent years; a few brief interruptions occurred in 2018 which were outside CDDIS control, due to issues with EOSDIS and NASA infrastructure.

As shown in Figure 1, the providers of files for the CDDIS archive push their files (data, derived products, etc.) to the CDDIS ingest server, utilizing the Earthdata Login system for validating access. Incoming files are then handled by the processing system which performs file/content validation, quality control, and metrics extraction. Metadata and metrics (ingest/archive and distribution) information is pushed to the EOSDIS Common Metadata Repository (CMR) system. Content metadata, describing collections and granules, are available for access by a broad user community through the CMR. Valid files are then moved to the CDDIS archive for public access through the CDDIS ftp and web servers.

2.2 File Submissions

The CDDIS utilizes an https-based protocol method for delivery of files from suppliers of data and products. The validation is performed through the EOSDIS Earthdata Login system, the same system used for access to the CDDIS real-time caster. The file uploads can be performed through a webpage interface or a command line application that can perform an http "post" operation, which is more commonly used for scripting. This process allows data suppliers to authenticate through the Earthdata Login system and provide their files through https to CDDIS for ingest into the archive. More information on the CDDIS file upload system, including an FAQ, is available at URL: https://cddis.nasa.gov/About/CDDIS_File_Upload_Documentation.html

2.3 File Ingest Processing

New file ingest processing software was implemented at CDDIS in 2016 for incoming GNSS files; over the next year, the software was updated to process incoming SLR and DORIS files. This new software suite incorporated numerous disparate programs developed over the years into a single, easily maintained software base which incorporates all the CDDIS requirements for data ingest while also allowing additional flexibility in meeting future metadata requirements. This file ingest processing system allows staff to check for errors in a more consistent fashion, regardless of data type or file provider; the automated system allows the staff to identify several error types, such as problems with file naming, compression, and content. The software then moves validated incoming files to the appropriate directory based on the filename.

Starting in 2018, the CDDIS has worked with the GSFC VLBI staff to transition code for processing incoming VLBI-related files into this new, common file ingest software. This effort was more complicated since the process for archiving VLBI files had always been unique and actually developed by the GSFC VLBI staff. The transition to the common ingest processing is not yet complete but work continues.

3 Archive Contents

The CDDIS has supported GSFC VLBI and IVS archiving requirements since 1979 and 1999, respectively.

The IVS Data Center content and structure is shown in Table 1 (a figure illustrating the flow of information, data, and products between the various IVS components was presented in the CDDIS submission to the IVS 2000 Annual Report). As described above, the CDDIS has established a file upload system for providing IVS data, product, and information files to the archive. Using specified filenames, Operation and Analysis Centers upload files to this system. Automated archiving routines peruse the directories and move any new data to the appropriate public disk area. These routines migrate the data based on the filename to the appropriate directory as described in Table 1. "Mirroring" software on the CDDIS host computer, as well as all other IVS Data Centers, facilitates equalization of data and product holdings among these Data Centers. At this time, mirroring is performed between the IVS Data Centers located at the CDDIS, the Bundesamt fur Kartographie und Geodasie in Leipzig, and the Observatoire de Paris.

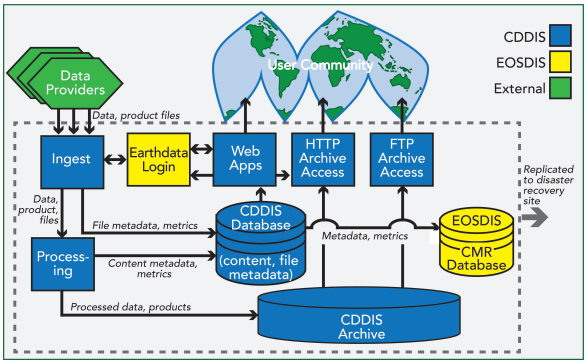

Fig. 1 System architecture overview diagram for the CDDIS facility installation within the EOSDIS infrastructure.

4 Accessing the CDDIS Archive

The CDDIS has a large international user community; over 243K unique hosts accessed the system in 2018. Today, users access the CDDIS archive through anonymous ftp and https. The ftp protocol allows users to easily automate file downloads but has problems from a system/security standpoint. As per U.S. Government and NASA directives, the CDDIS has begun to move users away from reliance on anonymous ftp. Despite this requirement, the CDDIS staff is committed to ensuring continued, easy, open access to its archive.

The CDDIS has configured servers to utilize protocols that allow two new methods for system access: https (browser and command line) and ftp-ssl (command line). The https protocol is as efficient as ftp transfer without firewall/router issues of ftp. The access to the CDDIS archive through both methods continues to present the same structure as that provided through anonymous ftp. For the near-term, access to data in the CDDIS archive will continue through ftp but users are strongly encouraged to explore the https and ftp-ssl (address: gdc.cddis.eosdis.nasa.gov) capabilities as soon as possible. The major reason for changing the archive access methods at CDDIS is system security and data integrity.

Archive access through the https protocol utilizes the same NASA single sign-on system, the EOSDIS Earthdata Login utility, as is used for the file upload and real-time caster user authentication. Before using the https protocol to access the CDDIS archive, new users must initially access the webpage, https://cddis.nasa.gov/archive, to establish an account and authorize access; this page will then redirect the user to the Earthdata Login page. Earthdata Login allows users to easily search and access the full breadth of all twelve EOSDIS DAAC archives. Earthdata Login also allows CDDIS staff to know our users better, which will then allow us to improve CDDIS capabilities. Once an account is established, the user has all permissions required to access the CDDIS archive using the https protocol, via a web browser or via a command line interface (e.g., through cURL or Wget) to script and automate file retrieval.

In addition, ftp-ssl access, an extension of ftp using TLS (transport layer security), can be used for scripting downloads from the CDDIS archive. The ftp-ssl is the option most similar to standard anonymous ftp. As with https, ftp-ssl will satisfy U.S. Government/NASA requirements for encryption.

Examples on using these protocols, including help with the cURL and Wget commands, are available on the CDDIS website; users are encour-

Table 1 IVS data and product directory structure.

Directory	Description
Data Directories	
vlbi/ivsdata/db/yyyy	VLBI database files for year yyyy
vlbi/ivsdata/ngs/yyyy	VLBI data files in NGS card image format for year yyyy
vlbi/ivsdata/vgosdb/yyyy	VLBI data files in vgosDB format for year yyyy
vlbi/ivsdata/aux/yyyy/ssssss	Auxiliary files for year yyyy and session ssssss
	these files include: log files, wx files, cable files, schedule files, correlator notes
Product Directories	
vlbi/ivsproducts/crf	CRF solutions
vlbi/ivsproducts/eopi	EOP-I solutions
vlbi/ivsproducts/eops	EOP-S solutions
vlbi/ivsproducts/daily sinex	Daily SINEX solutions
vlbi/ivsproducts/int sinex	Intensive SINEX solutions
vlbi/ivsproducts/trf	TRF solutions
vlbi/ivsproducts/trop	Troposphere solutions
Project Directories	
vlbi/ITRF2013	IVS contributions to the ITRF 2013 efforts
vlbi/ITRF2014	IVS contributions to the ITRF 2014 solution
vlbi/ivs-pilotbl	IVS Analysis Center pilot project (baseline)
Other Directories	
vlbi/ivscontrol	IVS control files (Master Schedule, etc.)
vlbi/ivsdocuments	IVS document files (solution descriptions, etc.)
vlbi/raw	Raw VLBI data
vlbi/dserver	dserver software and incoming files

aged to consult the available documentation at: https://cddis.nasa.gov/About/CDDIS_File_Download_Documentation.html as well as various presentations on these updates to the CDDIS archive access (see Section 7 below and https://cddis.nasa.gov/Publications/Presentations.html)

5 System Usage

During the 2017-2018 time period, nearly 2500 distinct hosts accessed the CDDIS to retrieve VLBI related files. These users, which include other IVS Data Centers, downloaded over 1.8 Tbytes (13.6 M files) of VLBI related files from the CDDIS in this two year period.

Future Plans

The CDDIS staff will continue to work closely with the IVS Coordinating Center staff to ensure that our system is an active and successful participant in the IVS archiving effort. A major area of focus will be the completion of the modifications to the CDDIS ingest processing software to accommodate all incoming VLBI-related files.

As discussed above, in the near future, the CDDIS will no longer support non-encrypted anonymous ftp access to its archive; access to the archive through https and ftp-ssl have already been implemented. The staff is also testing providing a WebDAV (Web Distributed Authoring and Versioning) interface to provide another method for accessing CDDIS archive. If feasible for CDDIS, this interface method would allow users to securely connect to the CDDIS archive as if it were a local drive on their computer.

The CDDIS has established Digital Object Identifiers (DOIs) for several of its collections of GNSS, SLR, and DORIS data and products; website landing pages have been established, linking to these published DOIs. DOIs for additional items, including VLBI data and products, are under development and review prior to registering and implementation.

The CDDIS has received funding to procure a refresh of its system servers, storage, and network hardware. Staff members have begun the engineering design for this next system; plans are to have the up-

graded system installed by the end 2019. The server and network hardware will remain within the same physical infrastructure as today's system, thus providing a reliable hosting environment with fully redundant networking paths and backup sites.