

IN23B-0087 NASA CDDIS: NEXT GENERATION SYSTEM

ABSTRACT: The Crustal Dynamics Data Information System (CDDIS) supports data archiving and distribution activities for the space geodesy and geodynamics community. The main objectives of the system are to make space geodesy and geodynamics related data and derived products available in a central archive, to maintain information about the archival of these data, to disseminate these data and information in a timely manner to a global scientific research community, and to provide user based tools for the exploration and use of the archive. As the techniques and data volume have increased, the CDDIS has evolved to offer a broad range of data ingest services, from data upload, quality control, documentation, metadata extraction, and ancillary information. As a major step taken to improve services, the CDDIS has transitioned to a new hardware system and implemented incremental upgrades to a new software system to meet these goals while increasing automation. This new system increases the ability of the CDDIS to consistently track errors and issues associated with data and derived product files uploaded to the system and to perform post-ingest checks on all files received for the archive. In addition, software to process new data sets and changes to existing data sets have been implemented to handle new formats and any issues identified during the ingest process. In this poster, we will discuss the CDDIS archive in general as well as review and contrast the system structures and quality control measures employed before and after the system upgrade. We will also present information about new data sets and changes to existing data and derived products archived at the CDDIS. Several applications have recently been developed at the CDDIS to aid users in data discovery, both within the international space geodesy community and beyond. This poster will include background information about the system and its user communities, archive contents and updates, enhancements for data discovery, new system architecture, and future plans.



CDDIS
Crustal Dynamics Data Information System

Carey Noll/CDDIS Manager (Carey.Noll@nasa.gov)
 Patrick Michael/Deputy (Patrick.Michael@nasa.gov)
 Justine Woo (justine.y.woo@nasa.gov)
 Rebecca Limbacher (rebecca.l.limbacher@nasa.gov)
 NASA Goddard Space Flight Center
 Code 61A, Greenbelt, MD 20771, USA

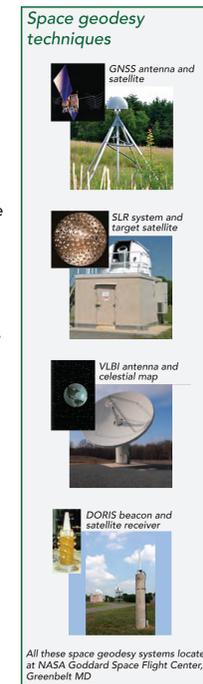


Presented at: 2017 Fall AGU
New Orleans LA, December 11-15, 2017

CRUSTAL DYNAMICS DATA INFORMATION SYSTEM

Background:

- The Crustal Dynamics Data Information System (CDDIS) is NASA's active archive of space geodesy data, products, and information (Global Navigation Satellite System/GNSS, Satellite Laser Ranging/SLR, Very Long Baseline Interferometry/VLBI, and Doppler Orbitography and Radio-positioning Integrated by Satellite/DORIS).
- CDDIS is one of 12 Distributed Active Archive Centers (DAACs) within NASA's Earth Observation System Data and Information System (EOSDIS)
- CDDIS became a regular member of the World Data System (WDS) in March 2013 and a member of Earth Science Information Partners (ESIP) in August 2017
- The largest CDDIS user community comes from the services within the International Association of Geodesy (IAG).
- The contents of the CDDIS archive are utilized for geodetic studies, e.g., plate tectonics, earthquake displacements, Earth orientation, Earth's surface deformation, Earth's gravity field, etc.
- The CDDIS archive also plays an interdisciplinary role in supporting the derivation of a Terrestrial Reference Frame (the foundation for virtually all airborne, space-based and ground-based Earth observations), precise orbit determination (POD) for NASA/international missions, atmospheric studies, etc.
- CDDIS has extensive partnerships through the International Association of Geodesy (IAG) serving as one of the primary data centers for the geometric services and its observing system, GGOS (Global Geodetic Observing System):
 - International GNSS Service (IGS)
 - International Laser Ranging Service (ILRS)
 - International VLBI Service for Geodesy and Astrometry (IVS)
 - International DORIS Service (IDS)



Archive contents:

- Point data from permanent stations in the global GNSS, SLR/LLR, VLBI, and DORIS networks
 - Data from over 1500 observing sites located at about 1000 locations around the world, going back in time as far as 1975
 - Stations in the GNSS, SLR/LLR, VLBI, and DORIS networks generate point data on a multi-day, daily, hourly, and/or sub-hourly basis
- Products derived from these data (some since 1976) including:
 - Precise network station positions (for ITRF)
 - Satellite orbits (for POD)
 - Station and satellite clocks (for timing)
 - Earth rotation parameters
 - Positions of celestial objects (for CRF)
 - Atmospheric parameters (ionosphere TEC, troposphere ZPD)

Current space geodesy site locations



Archive usage:

- The CDDIS contains data and derived products from over 1500 observing sites located at about 1000 locations around the world, going back in time as far as 1975.
- The archive is updated with new data/product files on varying time scales, dependent on the data type, from a sub-daily basis to weekly basis.
- Users require continuous access to data for generation of products on pre-determined schedules.
- The average user of the CDDIS accesses the contents of the archive through anonymous ftp by means of automated scripts executed on predefined schedules (typically sub-daily).
- Analysts can use this method for data transfer because they are familiar with the structure of the online archive and thus know what files they require, their availability schedule, and where to find them within the online structure.
- In 2016, the CDDIS distributed nearly 1.5B files totaling 170TB in volume from over 320K distinct hosts; we have seen at least a 20% increase in these figures thus far in 2017.

CDDIS INFRASTRUCTURE IMPROVEMENTS

Background:

- In December 2016, CDDIS transitioned to new computer system architecture
- Transition included implementation of an upgraded ingest processing system
- Goals of the new ingest system
 - Consolidation of key software components across data types
 - Improved quality control measures
 - Improved timeliness of data availability
 - Automation

Why update the archive processing system?
Issues with CDDIS incoming file processing prior to the computer system upgrade and implementation of new ingest system

- Multiple programs depending on data type (GNSS, SLR, DORIS, VLBI, derived products, etc.) and source
- Processing was inconsistent across data/file types and file providers, which resulted in inconsistent file processing and quality checking
- Code maintenance was a problem
- Processing at irregular intervals affected timeliness of archival
- Not all error checking was automated

Post-upgrade improvements addressed these issues and include:

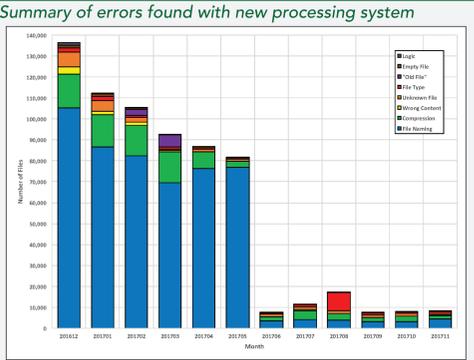
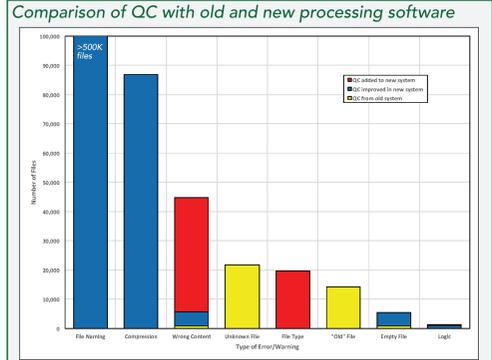
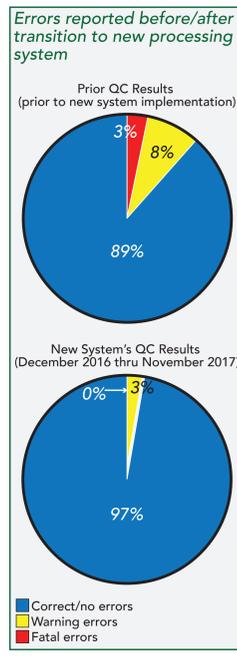
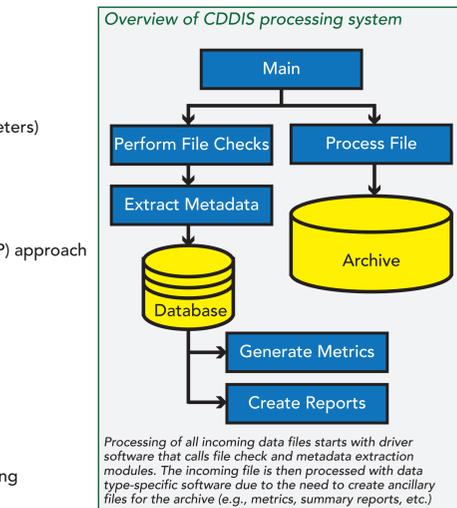
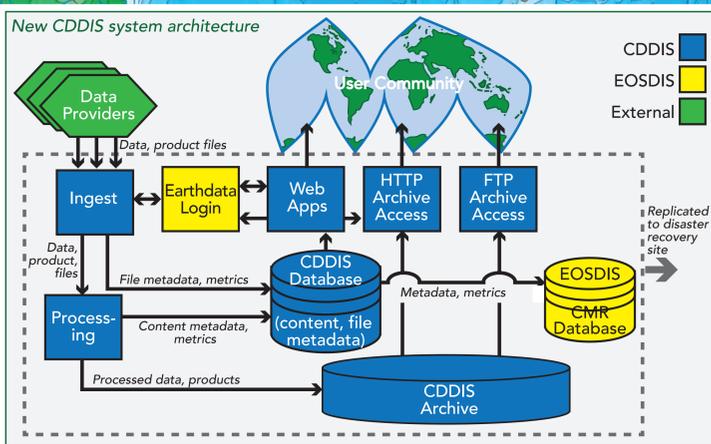
- Improved QC on all incoming files
 - Known file type
 - Non-empty file
 - Virus free
 - Valid compression
 - Filename consistent with file contents (e.g., temporal and spatial parameters)
 - Filename follows format naming conventions
 - "Logic" checks (data for future timeframe or very old data)
 - Some internal format validation
 - Checksums (MD5SUM and SHA512SUM)
- Checks for duplicate delivery of previously delivered identical files
- Consolidated and simplified code using object-oriented programming (OOP) approach
 - Reduced number of modules and programs
 - Generalized processing path for incoming files
 - Easier to implement and track software modifications
 - Code more versatile and reusable
- Same code base used for all data types
- Improved automation
 - Software creation of new data type and temporal directories
 - All content checking
 - Metrics creation of metrics files
- Error processing
 - All errors are recorded to the database allowing for efficient error tracking
 - Consistent checks of incoming files regardless of the file type
 - Fatal errors result in quarantine

New file ingest processing system more integrated with CDDIS database management

- Use of database simplifies archive QC (e.g., consistency checking)
- Improved reporting on archive contents and error processing
 - File-level metadata extracted on ingest
 - Content metadata extracted on ingest
- All processing errors are tracked/ pushed to database for improved monitoring
- Improved monitoring of system performance (e.g., processing times, file availability)

Future plans for updates to archive processing system

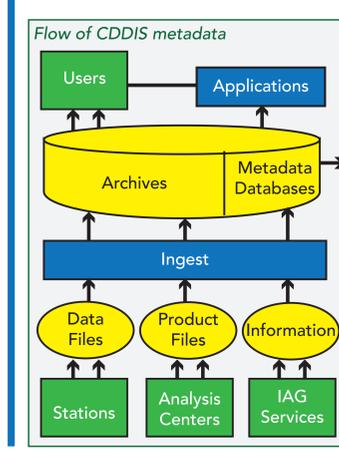
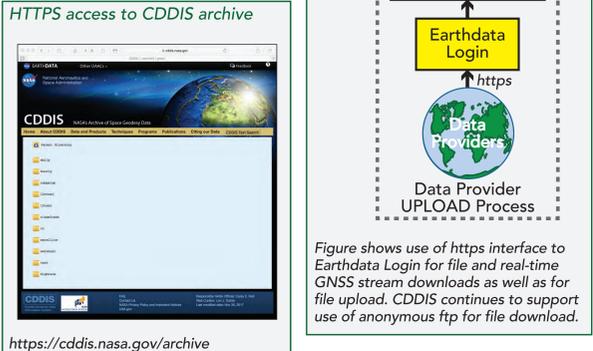
- CDDIS currently developing updated version of the next-generation processing software
- Features of new version:
 - Processing of incoming VLBI data files
 - Processing of derived product files
 - Re-organization of internal error handling
- Software enhancements will allow CDDIS to re-process data and product files archived prior to 2017 to ensure quality and allow valuable metadata to be collected



WHAT'S UP NEXT?

Updates to archive access methodologies

- Users access the CDDIS archive through anonymous ftp
- This protocol allows users to easily automate file downloads but has problems from a system and user standpoint
- HTTPS access to CDDIS archive
 - Archives and users continue to move away from using FTP
 - Therefore, CDDIS will implement access to it's full archive through HTTPS
 - HTTPS access will continue to use same structure as provided through FTP
 - HTTPS access is as efficient as FTP transfer without the firewall/router issues of FTP
 - FTP is a two-port protocol; users can have connectivity problems (e.g., with firewall, DNS, etc.)
 - HTTP is a one-port protocol, fewer issues with downloads
 - Earthdata Login system will be used for access through HTTPS
 - FTP access to CDDIS archive will continue but users are encouraged to explore HTTPS capabilities
 - Early results from user testing are promising
- Advantages of using HTTPS to access CDDIS archive:
 - Earthdata Login allows users to easily search and access the full breadth of all twelve EOSDIS DAAC archives
 - Earthdata Login will allow CDDIS to know our users better which will then allow us to improve CDDIS capabilities
 - HTTPS protocol avoids issues inherent with FTP protocol



Improvements to CDDIS metadata

- CDDIS currently updating collection and granule level metadata to satisfy EOSDIS requirements
- CDDIS is an EOSDIS DAAC
- EOSDIS implemented a "Common Metadata Repository" (CMR)
- Single source of unified, high-quality, high-performance, and reliable Earth Science metadata
- Metadata can be discovered and accessed through programmatic interfaces leveraging standard protocols and APIs
- CDDIS modifying existing metadata to develop relationships with EOSDIS-required metadata
- Initial focus on completion of collection-level metadata records describing CDDIS holdings of GNSS, SLR, DORIS, and VLBI data and derived products
- Once complete, CDDIS will review and update granule-level metadata as needed

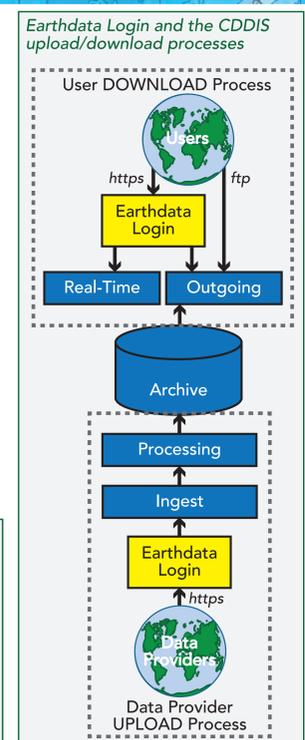


Figure shows use of https interface to Earthdata Login for file and real-time GNSS stream downloads as well as for file upload. CDDIS continues to support use of anonymous ftp for file download.

FOR MORE INFORMATION

- Data and products are acquired as part of NASA's Earth Science Data Systems and archived and distributed by the Crustal Dynamics Data Information System (CDDIS): 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.
- The staff welcomes feedback on the CDDIS and in particular the ideas expressed in this poster; contact Carey Noll (Carey.Noll@nasa.gov)
- For more information, visit the CDDIS website: <https://cddis.nasa.gov>

