CDDIS Global Data Center Technical Report 2014

C. Noll

NASA Goddard Space Flight Center, Code 690.1
Greenbelt, MD 20771 USA
Carey.Noll@nasa.gov

1 Introduction

The Crustal Dynamics Data Information System (CDDIS) is NASA's data archive and information service supporting the international space geodesy community. For over 30 years, the CDDIS has provided continuous, long term, public access to the data (mainly GNSS-Global Navigation Satellite System, SLR-Satellite Laser Ranging, VLBI-Very Long Baseline Interferometry, and DORIS-Doppler Orbitography and Radiopositioning Integrated by Satellite) and products derived from these data required for a variety of science observations, including the determination of a global terrestrial reference frame and geodetic studies in plate tectonics, earthquake displacements, volcano monitoring, Earth orientation, and atmospheric angular momentum, among others. The specialized nature of the CDDIS lends itself well to enhancement to accommodate diverse data sets and user requirements. The CDDIS is one of NASA's Earth Observing System Data and Information System (EOSDIS) Distributed Active Archive Centers (DAACs); 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).

The CDDIS serves as one of the primary data centers and core components for the geometric services established under the International Association of Geodesy (IAG), an organization that promotes scientific cooperation and research in geodesy on a global scale. The system has supported the International GNSS Service (IGS) as a global data center since 1992. The CDDIS activities within the IGS during 2014 are summarized below; this report also includes any recent changes or enhancements made to the CDDIS.

2 System Description

The CDDIS archive of IGS data and products are accessible worldwide through anonymous ftp (*ftp://cddis.gsfc.nasa.gov*). The CDDIS has recently implemented web-based access to the archive (*http://cddis.gsfc.nasa.gov/archive*). The CDDIS is located at NASA's Goddard Space Flight Center (GSFC) and is available to users 24 hours per day, seven days per week.

The CDDIS computer system is fully redundant with the primary and secondary/failover system. Each system utilizes a distributed functionality (incoming, outgoing, processing, database, and map servers) and is configured with a local backup system as well as a full backup system located in a third building at GSFC. The archive is equipped with a multi-Tbyte RAID storage system and is scaled to accommodate future growth. All ftp and web access is performed on the outgoing server. Data centers, stations, and analysis centers push files to the CDDIS incoming server. Processing of incoming files for the on-line archive is performed in a separate environment that also includes a database server for managing metadata extracted from incoming data.

3 Archive Content

As a global data center for the IGS, the CDDIS is responsible for archiving and providing access to GNSS data from the global IGS network as well as the products derived from the analyses of these data in support of both operational and working group/pilot project activities. The CDDIS archive is

approximately 11.7 Tbytes in size of which 11 Tbytes (95%) is devoted to GNSS data, products (710 Gbytes), and ancillary information. All data and products are accessible through subdirectories of *ftp://cddis.gsfc.nasa.gov/gnss* (a symbolic link to *ftp://cddis.gsfc.nasa.gov/gps*).

3.1 GNSS Tracking Data

3.1.1 Operational Data Archive

The user community has access to GNSS data available through the on-line global data center archives of the IGS. Over 50 operational and regional IGS data centers and station operators make data (observation, navigation, and meteorological) available in RINEX format to the CDDIS from selected receivers on a daily, hourly, and sub-hourly basis. The CDDIS also accesses the archives of the other three IGS global data centers, Scripps Institution of Oceanography (SIO) in California, the Institut Géographique National (IGN) in France, and the Korea Astronomy and Space Science Institute (KASI) to retrieve (or receive) data holdings not routinely transmitted to the CDDIS by an operational or regional data center. Table 1 below summarizes the types of IGS operational GNSS data sets archived at the CDDIS.

Table 1a: GNSS Data Type Summary.

Data Type	Sample Rate	Data Format	Available On-line
Daily GNSS	30 sec.	RINEX and compact RINEX	Since 1992
Hourly GNSS	30 sec.	Compact RINEX	10+ years
High-rate GNSS	1 sec.	Compact RINEX	Since May 2001
Satellite GPS	10 sec.	Compact RINEX	Since 2002

Table 1b: GNSS Data Archive Summary for 2014.

Data Type	Avg. No. Sites/Day	No. Unique Sites	Avg. Volume/Day	Total Volume/Year	No. Files	Directory Location
Daily GNSS	475	534	1,100 Mb	400 Gb	735K	*/gnss/data/daily*
Hourly GNSS	310	341	383 Mb	140 Gb	6,705K	*/gnss/data/hourly*
High-rate GNSS	166	200	2,096 Mb	765 Gb	9,750K	*/gnss/data/highrate*

Data, in RINEX V2.10 or V2.11 format, from GPS and GPS+GLONASS receivers are archived within the main GNSS directory structure */gnss/data*.

The CDDIS archives four major types/formats of GNSS data, all in RINEX format, as described in Table 1a. Daily RINEX data are quality-checked, summarized, and archived to public disk areas in subdirectories by year, day, and file type; the summary and inventory information are also loaded into an on-line database. Nearly 175K daily station days from 534 distinct GNSS receivers were archived at the CDDIS during 2014. A complete list of daily, hourly, and high-rate sites archived in the CDDIS can be found in the yearly summary reports at URL *ftp://cddis.gsfc.nasa.gov/reports/gnss/*.

Within minutes of receipt, the hourly GNSS files are archived to subdirectories by year, day, and hour. Although these data are retained on-line, the daily files delivered at the end of the UTC day contain all data from these hourly files and thus can be used in lieu of the individual hourly files. A total of 341 hourly sites (over 6.7 million files) were archived during 2014.

High-rate (typically 1-second sampling) GNSS data are archived in files containing fifteen minutes of data and in subdirectories by year, day, file type, and hour. Many of these data files are created from real-time streams. Data from 200 high-rate sites (nearly 10 million files) were also archived in the CDDIS in 2014.

The CDDIS generates global broadcast ephemeris files (for both GPS and GLONASS) on an hourly basis. These files are derived from the site-specific ephemeris data files for each day/hour. These files are appended to a single file that contains the orbit information for all GPS and GLONASS satellites for the day up through that hour. The merged ephemeris data files are then copied to the day's subdirectory within the hourly data file system. Within 1-2 hours after the end of the UTC day, after sufficient station-specific navigation files have been submitted, this concatenation procedure is repeated to create the daily

broadcast ephemeris files (both GPS and GLONASS), using daily site-specific navigation files as input. The daily files are copied to the corresponding subdirectory under the daily file system. Users can thus download this single, daily (or hourly) file to obtain the unique navigation messages rather than downloading multiple broadcast ephemeris files from the individual stations.

The CDDIS generates and updates status files, (*/gnss/data/daily/YYYY/DDD/YYDD.status*) that summarize the holdings of daily GNSS data. These files include a list of stations. The archive status files of CDDIS GNSS data holdings reflect timeliness of the data delivered as well as statistics on number of data points, cycle slips, and multipath. The user community can receive a snapshot of data availability and quality by viewing the contents of such a summary file.

3.1.2 MGEX Archive

During 2014 the CDDIS continued the archiving of data from participating multi-GNSS receivers as well as products derived from the analysis of these data. The data include newly available signals (e.g., Galileo, QZS, SBAS, and BeiDou). The summary of the MGEX data holdings at the CDDIS is shown in Table 2 below. Daily status files are also provided that summarize the MGEX data holdings; however, data quality information, generated for operational GNSS data holdings, is not available through the software created by CDDIS to summarize data in RINEX V3 format. Products derived in support of MGEX by three to six ACs are also available through the CDDIS (*/gnss/products/mgex/WWWW*).

Table 2: GNSS MGEX Data Archive Summary for 2014.

Data Type	Avg. No. Sites/Day	No. Unique Sites	No. Files	Avg. Volume/Day	Directory Location
Daily GNSS	100	115	35.1K	500 Mb	*/gnss/data/campaign/mgex/daily*
Hourly GNSS	50	54	17.1K	100 Mb	*/gnss/campaign/mgex /data/hourly*
High-rate GNSS	40	47	13.6K	1,350 Mb	*/gnss/campaign/mgex /data/highrate*

The CDDIS also added a merged, multi-GNSS broadcast ephemeris file containing GPS, GLONASS, Galileo, BeiDou, QZSS, and SBAS ephemerides from MGEX stations. This file, generate by colleagues at the Technical University in Munich (TUM) and Deutsches Zentrum f¨ur Luft- und Raumfahrt (DLR), is similar to the daily and hourly concatenated broadcast message files provided by the CDDIS for the operational data sets; it contains all the unique broadcast navigation messages for the day. The file is denoted brdmDDD0.YYp.Z and found in daily subdirectories within the MGEX archive at CDDIS (*/gnss/data/campaign/mgex//daily/rinex3/YYYY/DDD/YYp*).

In order to promote usage of RINEX V3 and allow users (and data centers) to become familiar with the format and file naming conventions, ESA began delivery of data from MGEX stations using both RINEX V2 and V3 filename formats. The CDDIS established a daily subdirectory for these files within the daily MGEX directory structure (*/gnss/data/campaign/mgex//daily/rinex3/YYYY/DDD/crx*).

Colleagues at TUM and DLR are also providing GPS and QZSS CNAV (civilian navigation) data on an operational basis within MGEX. These messages are collected from a sub-network (ten stations) of MGEX stations and are provided in a merged daily file in a format similar to RINEX. These files are named *brdxDDD0.YYx.Z* and stored in a daily subdirectory within the MGEX archive at CDDIS (*/gnss/data/campaign/mgex//daily/rinex3/YYYY/cnav*).

Colleagues at DLR provided a new product, differential code biases (DCBs) for the MGEX campaign. This product was derived from GPS, GLONASS, Galileo, and BeiDou ionosphere-corrected pseudorange differences and is available in the bias SINEX format. Two files are available, daily satellite and daily satellite and station biases, for 2013-2014 in CDDIS directory */gnss/products/mgex/dcb*. Additional details on the DCB product are available in IGSMail message 6868 sent in February 2014.

3.2 IGS Products

The CDDIS routinely archives IGS operational products (daily, rapid, and ultra-rapid orbits and clocks, ERP, and station positions) as well as products generated by IGS working groups and pilot projects

(ionosphere, troposphere, real-time clocks). Table 3 below summarizes the GNSS products available through the CDDIS. The CDDIS currently provides on-line access through anonymous ftp to all IGS products generated since the start of the IGS Test Campaign in June 1992 in the file system */gnss/products*; products from GPS+GLONASS products are available through this filesystem. Products derived from GLONASS data only continued to be archived at the CDDIS in a directory structure within the file system */glonass/products*.

Table 3: GNSS Product Summary.

Product Type	Number of ACs/AACs	Volume	Directory
Orbits, clocks, ERP, positions	14+Combinations	1.2 Gb/week	*/gnss/products/WWWW (GPS, GPS+GLONASS)* */glonass/products/WWWW (GLONASS only)*
Troposphere	Combination	2.6 Mb/day, 940 Mb/year	*/gnss/products/troposphere/YYYY*
Ionosphere	4+Combination	4 Mb/day, 1.5 Gb/year	*/gnss/products/ionex/YYYY*
Real-time clocks	Combination	6.0 Mb/week	*/gnss/products/rtpp/YYYY*
Repro2 products	9	500 Mb/week	*/gnss/products/WWWW/repro2*

Note: *WWWW*=4-digit GPS week number; *YYYY*=4-digit year

The CDDIS also continues to archive combined troposphere estimates in directories by GPS week. Global ionosphere maps of total electron content (TEC) from the IONEX AACs are archived in subdirectories by year and day of year. Real-time clock comparison products have been archived at the CDDIS in support of the IGS Real-Time Pilot Project, and current IGS Real-Time Service, since 2009.

In 2014, the IGS analysis centers provided products for the second IGS reprocessing campaign (repro2). The CDDIS will provide support through upload of files from the ACs and online archive of these products (*/gnss/products/WWWW/repro2*).

3.3 Real-Time Activities

In 2013, the CDDIS staff configured a server and began testing a real-time caster to provide a real-time streaming capability at GSFC and support the IGS Real-Time Service (IGS RTS). The CDDIS successfully tested obtaining product streams from the BKG and IGS casters and providing access to these streams to authorized users. Work continued in 2014 to make the system operational and in spring 2014, the CDDIS caster was fully installed for broadcasting product streams in real-time. The caster runs the NTRIP (Network Transport of RTCM via internet Protocol) format.

The majority of the work in 2014 involved development of a registration process that satisfied NASA security requirements and collected information required by the IGS RTS. As stated previously, the CDDIS is one of NASA's EOSDIS DAACs and through EOSDIS, has access to a world-class user registration process, the EOSDIS User Registration System (URS), with over 100K users in its system. Since the NTRIP-native registration/access software was not compatible with NASA policies, the CDDIS developed software to interface the caster and the URS within a generic Lightweight Directory Access Protocol (LDAP) framework. The module was specifically developed to easily interface with multiple user verification systems and was given back to the NTRIP community for possible inclusion in future releases. The user registration form is available on the CDDIS website; once completed, the data are passed to the URS, which generates an email to the user with a validation link. The user accesses the link and the URS validates the form's data; this process is accomplished within a minute or less. The user's validated access request is submitted to CDDIS staff for access authorization to the CDDIS caster. This second step is not yet automated and can take several hours to configure depending on the time of day. Furthermore, users registering in this system have access to the entire suite of EOSDIS products across all 12 EOSDIS DAACs.

Initially, the CDDIS caster is providing access to product streams from both the BKG and IGS casters. Data streams have also been tested, provided through JPL for receivers in NASA's Global GPS Network. This network of roughly seventy globally distributed, geodetic quality, dual frequency receivers, will add 1 Hz data streams to those current available from the IGS RTS.

Once the CDDIS caster is operational, the system will serve as a third primary caster for the IGS RTS, thus providing a more robust topology with redundancy and increased reliability for the service. User registration, however, for all three casters is unique; therefore current users of the casters located at the IGS and BKG will be required to register through the CDDIS registration process in order to use the CDDIS caster.

The CDDIS has also developed software to capture real-time data streams into fifteen-minute high-rate files. This capability requires further testing as the CDDIS caster becomes operational and data streams from real-time stations are added.

3.4 Supporting Information

Daily status files of GNSS data holdings, reflecting timeliness of the data delivered as well as statistics on number of data points, cycle slips, and multipath, continue to be generated by the CDDIS. By accessing these files, the user community can receive a quick look at a day's data availability and quality by viewing a single file. The daily status files are available through the web at URL *ftp://cddis.gsfc.nasa.gov/reports/gnss/status*. The daily status files are also archived in the daily GNSS data directories.

In preparation for the analysis center's second reprocessing campaign, the CDDIS developed site-specific reports detailing missing data. Station operators and operational data centers can consult these lists (*ftp://cddis.gsfc.nasa.gov/gnss/data/daily/reports/missing*) and if available, supply missing files to the CDDIS for inclusion in the global data center archives.

Ancillary information to aid in the use of GNSS data and products are also accessible through the CDDIS. Weekly and yearly summaries of IGS tracking data (daily, hourly, and high-rate) archived at the CDDIS are generated on a routine basis. These summaries are accessible through the web at URL *ftp://cddis.gsfc.nasa.gov/reports/gnss.* The CDDIS also maintains an archive of and indices to IGS Mail, Report, Station, and other IGS-related messages.

4 System Usage

Figure 1 summarizes the usage of the CDDIS for the retrieval of GNSS data and products in 2014. This figure illustrates the number and volume of GNSS files retrieved by the user community during 2014, categorized by type (daily, hourly, high-rate, MGEX data, products). Nearly 370 million files (over 50 Tbytes), excluding robot downloads, were transferred in 2014, with an average of nearly 30 million files per month. Figure 2 illustrates the profile of users accessing the CDDIS IGS archive during 2014. The majority of CDDIS users were once again from hosts in Europe, Asia, and North America.

5 Recent Developments

5.1 CDDIS Website

Work on an update of the CDDIS website was completed in early 2014. In addition to a refresh of the appearance of the website, the content was reviewed and updated. An application for the enhanced display and comparison of the contents of IGS, ILRS, and IDS site logs was completed in 2014. The Site Log Viewer is an application for the enhanced display and comparison of the contents IAG service site logs. Through the Site Log Viewer application, users can display a complete site log, section by section, display contents of one section for all site logs, and search the contents of one section of a site log for a specified parameter value. Thus, users can survey the entire collection of site logs for systems having particular equipment or characteristics.

A second application, the CDDIS Archive Explorer, is currently under development to aid in discovering data available through the CDDIS. The application will allow users, particularly those new to the CDDIS, the ability to specify search criteria based on temporal, spatial, target, site designation, and/or observation parameter in order to identify data and products of interest for download. Results of these queries will

include a listing of sites (or other metadata) or data holdings satisfying the user input specifications. Such a user interface will also aid CDDIS staff in managing the contents of the archive.

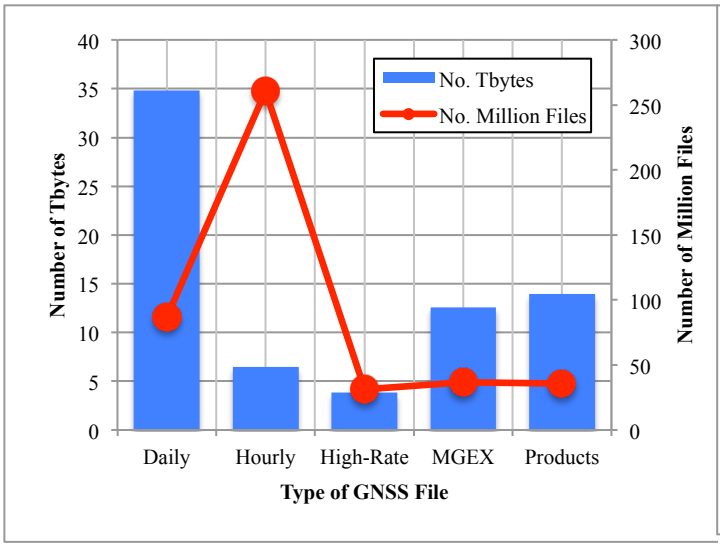

Figure 1: Number and volume of GNSS files transferred from the CDDIS in 2014.

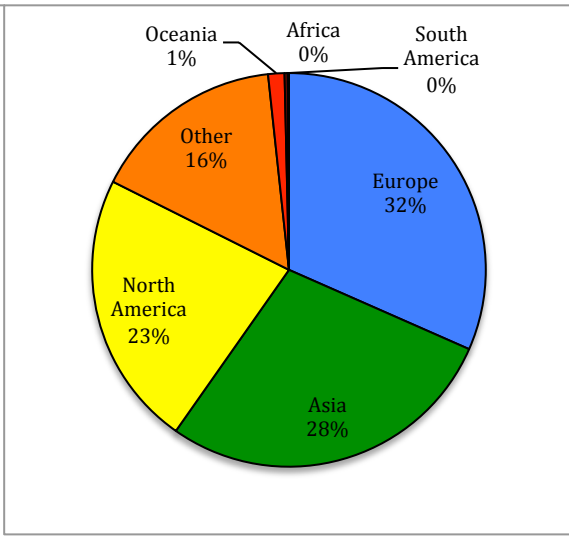

Figure 2: Geographic distribution of IGS users of the CDDIS in 2014.

5.2 Next Generation Hardware

Funding was identified in 2013 to procure a computer system refresh for the CDDIS. The CDDIS system engineer reviewed current and near-term requirements and developed a hardware procurement strategy. Hardware was procured in mid-2014 with installation beginning in late 2014 and scheduled for completion and testing in early 2015. The system will be installed within the EOSDIS computer facility and network infrastructure providing a more reliable/redundant environment (power, HVAC, 24-hour on-site emergency personnel, etc.) and network connectivity; a disaster recovery system will be installed in a different location on the GSFC campus. The new system location will address the number one operational issue CDDIS has experienced over the past several years, namely, the lack of consistent and redundant power and cooling in its existing computer facility. Multiple redundant 40G network switches will also be utilized to take full advantage of a high-performance network infrastructure by utilizing fully redundant network paths for all outgoing and incoming streams along with dedicated 10G network connections between its primary operations and its backup operations. The CDDIS will also transition approximately 85% of its operation services over to virtual machine (VM) technology for both multiple instance services in a load balancing configuration which will allow additional instances to be increased or decreased due to demand and will allow maintenance (patching, upgrades, etc.) to proceed without interruption to the user or any downtime. CDDIS will be utilizing a large (XX Tbyte) storage system to easily accommodate future growth of the archive.

5.3 Metadata Developments

The CDDIS has recently made modifications to the metadata extracted from incoming data and product files pushed to its archive. These enhancements have facilitated cross discipline data discovery by providing information about CDDIS archive holdings to other data portals such as Earth Observing System (EOS) Clearinghouse (ECHO) and future integration into the Global Geodetic Observing System (GGOS) portal. The staff has begun a metadata evolution effort, re-designing the metadata extracted from incoming data and adding information that will better support EOSDIS applications such as ECHO and the metrics collection effort.

The CDDIS has implemented Digital Object Identifiers (DOIs) to select IGS data sets (daily GNSS data). DOIs can provide easier access to CDDIS data holdings and allow researchers to cite these data holdings in publications. Landing pages are available for each of the DOIs created for CDDIS data products and linked to description pages on the CDDIS website; an example of a typical DOI description (or landing) page, for daily Hatanaka-compressed GNSS data files, can be viewed at: *http://cddis.gsfc.nasa.gov/Data_and_Derived_Products/GNSS/daily_gnss_d.html*. DOIs will be assigned to additional GNSS data and product sets in the near future.

6 Publications

The CDDIS staff attended several conferences during 2014 and presented papers on or conducted demos of their activities within the IGS, including:

C. Noll, P. Michael, N. Pollack, L. Tyahla, "Supporting GGOS through the Crustal Dynamics Data Information System", Abstract EGU2014-7174 presented at 2014 EGU General Assembly, Vienna Austria, 28 Apr.-02 May.

C. Noll, F. Boler, H. Habrich. "Data Centers: Status and Progress", presented at IGS 20th Anniversary Workshop, Pasadena CA, 23-27 Jun.

C. Noll, P. Michael, "Recent Developments at the CDDIS", presented at IGS 20th Anniversary Workshop, Pasadena CA, 23-27 Jun.

P. Michael, C. Noll, J. Roark. "CDDIS Real-Time Developments", presented at IGS 20th Anniversary Workshop, Pasadena CA, 23-27 Jun.

C. Noll, P. Michael, N. Pollack. "Recent Developments in Space Geodesy Data Discovery at the CDDIS", Abstract IN11C-3623 presented at 2014 Fall Meeting, AGU, San Francisco, Calif., 15-19 Dec.

P. Michael, C. Noll, J. Roark. "CDDIS Near Real-Time Data for Geodesy Based Applications", Abstract IN43C-3709 presented at 2014 Fall Meeting, AGU, San Francisco, Calif., 15-19 Dec.

Electronic versions of these and other publications can be accessed through the CDDIS on-line documentation page on the web at URL *http://cddis.gsfc.nasa.gov/Publications/Presentations.html*.

7 Future Plans

The CDDIS will continue to support the IGS MGEX. The experiment is an excellent opportunity to prepare the data centers for archive of data in RINEX V3. The CDDIS will coordinate with the Infrastructure Committee, the Data Center Working Group, and other IGS data centers to develop a transition plan and introduce RINEX V3 data into the "operational" GNSS data directory structure, making it easier for users to access these data.

The CDDIS plans to make its real-time caster operational in the coming year as part of the IGS Real-Time Service. CDDIS staff will work with the IGS to identify stations for streaming to its caster. Future activities in the real-time area include capturing the streams for generation of 15-minute high-rate files for archive. This capability requires further testing as the CDDIS caster becomes operational and data streams from real-time stations are added. The CDDIS staff will need to develop a revised interface software to the EOSDIS' next generation URS (version 4). The staff will also automate the process of adding users to the CDDIS caster configuration files.

CDDIS has traditionally used ftp for delivery of data for the archive from both data centers and analysis centers. While this has worked well over the years, transition to the new system provides an opportune time time to look at updating this method to a web-based approach that can utilize the EOSDIS URS infrastructure. CDDIS will investigate the best methods to incorporate a web-based approach that will

continue to allow suppliers to use existing scripts without significant modification but also tie authentication into the URS.

8 Contact Information

To obtain more information about the CDDIS IGS archive of data and products, contact:

Ms. Carey E. Noll	Phone:	(301) 614-6542
Manager, CDDIS	Fax:	(301) 614-6015
Code 690.1	E-mail:	*Carey.Noll@nasa.gov*
NASA GSFC	WWW:	*http://cddis.gsfc.nasa.gov*
Greenbelt, MD 20771		*http://cddis.nasa.gov*

9 Acknowledgments

The author would like to thank the CDDIS contractor staff, Patrick Michael, Maurice Dube, Nathan Pollack, and Rebecca Limbacher (Science Systems and Applications, Inc./SSAI), Lori Tyahla (Stinger Ghaffarian Technologies/SGT), and James Roark (ADNET Systems). The recognition and success of the CDDIS in many international programs can be directly attributed to the continued dedicated, consistent, professional, and timely support of its staff.

References

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.

C. Noll, Y. Bock, H. Habrich and A. Moore, “Development of data infrastructure to support scientific analysis for the International GNSS Service”, Journal of Geodesy, Feb 2009, pages 309-325, DOI 10.1007/s00190-008-0245-6.