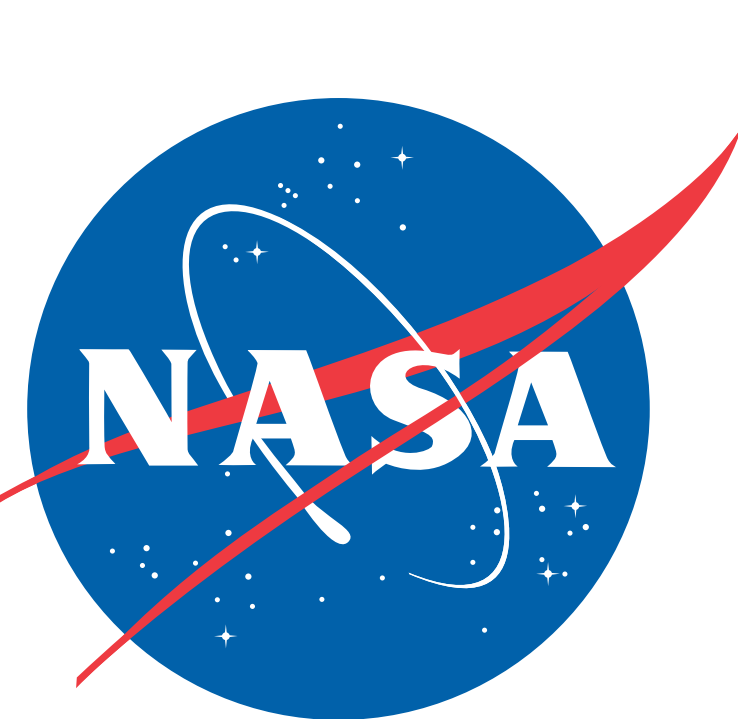# Improvements to the Crustal Dynamics Data Information System

Carey Noll (Carey.Noll@nasa.gov)
Maurice Dube/SSAI (Maurice.P.Dube@nasa.gov)
Patrick Michael/Catholic University of America (Patrick.Michael@nasa.gov)
Nathan Pollack/SSAI (nathan.pollack@ssaihq.com)
Lori Tyahla/SGT, Inc. (ltyahla@sgt-inc.com)

NASA Goddard Space Flight Center
Code 690, Greenbelt, MD 20771
USA

Abstract: NASA's space geodesy archive, the Crustal Dynamics Data Information System (CDDIS), provides data archiving and distribution support to a global research community. The archive consists of GNSS, laser ranging, VLBI, and DORIS data sets and products derived from these data. The CDDIS is one of NASA's Earth Observing System Data and Information System (EOSDIS) distributed data centers; EOSDIS data centers serve a diverse user community and are tasked to provide facilities to search and access science data and products. The CDDIS data system and its archive are vital components to many national and international science communities, in particular several of the operational services within the International Association of Geodesy (IAG) and its project the Global Geodetic Observing System (GGOS), including the International DORIS Service (IDS), the International GNSS Service (IGS), the International Laser Ranging Service (ILRS), the International VLBI Service for Geodesy and Astrometry (IVS), and the International Earth Rotation and Reference Systems Service (IERS). Several activities are nearing completion at the CDDIS to aid users in data discovery, both within the international space geodesy community and beyond. The CDDIS has completed enhancements to its implementation of the Geodetic Seamless Archive Centers (GSAC). The service provides web services to facilitate data discovery within and across participating archives. The CDDIS is currently implementing modifications to the metadata extracted from incoming data and product files pushed to its archive to permit information about CDDIS archive holdings to be made available through other data portals such as Earth Observing System (EOS) Clearinghouse (ECHO). The CDDIS hopes to soon install an Ntrip broadcast relay to support the activities of the IGS Real-Time Pilot Project (RTPP) and the future Real-Time IGS Service. 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.
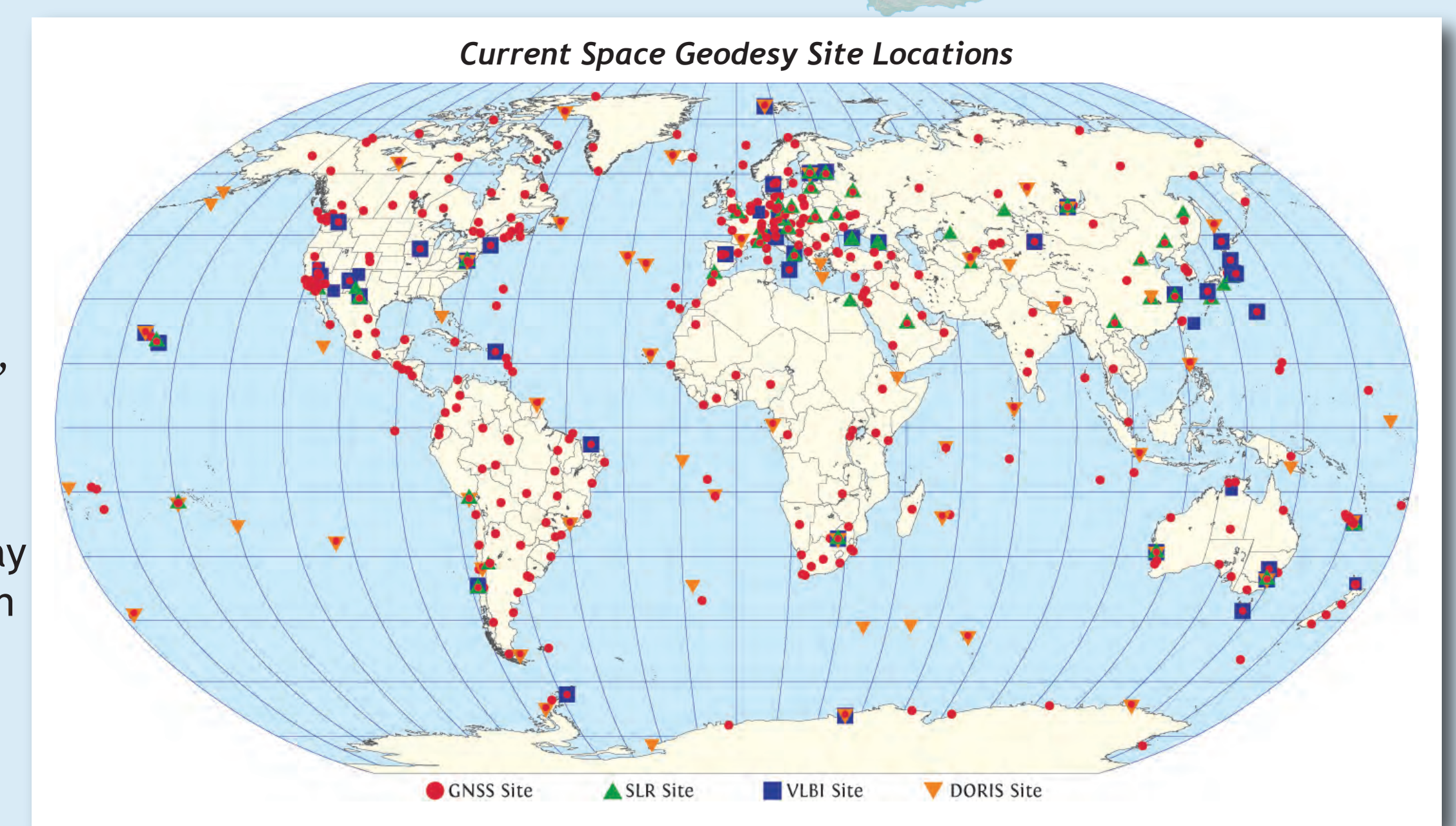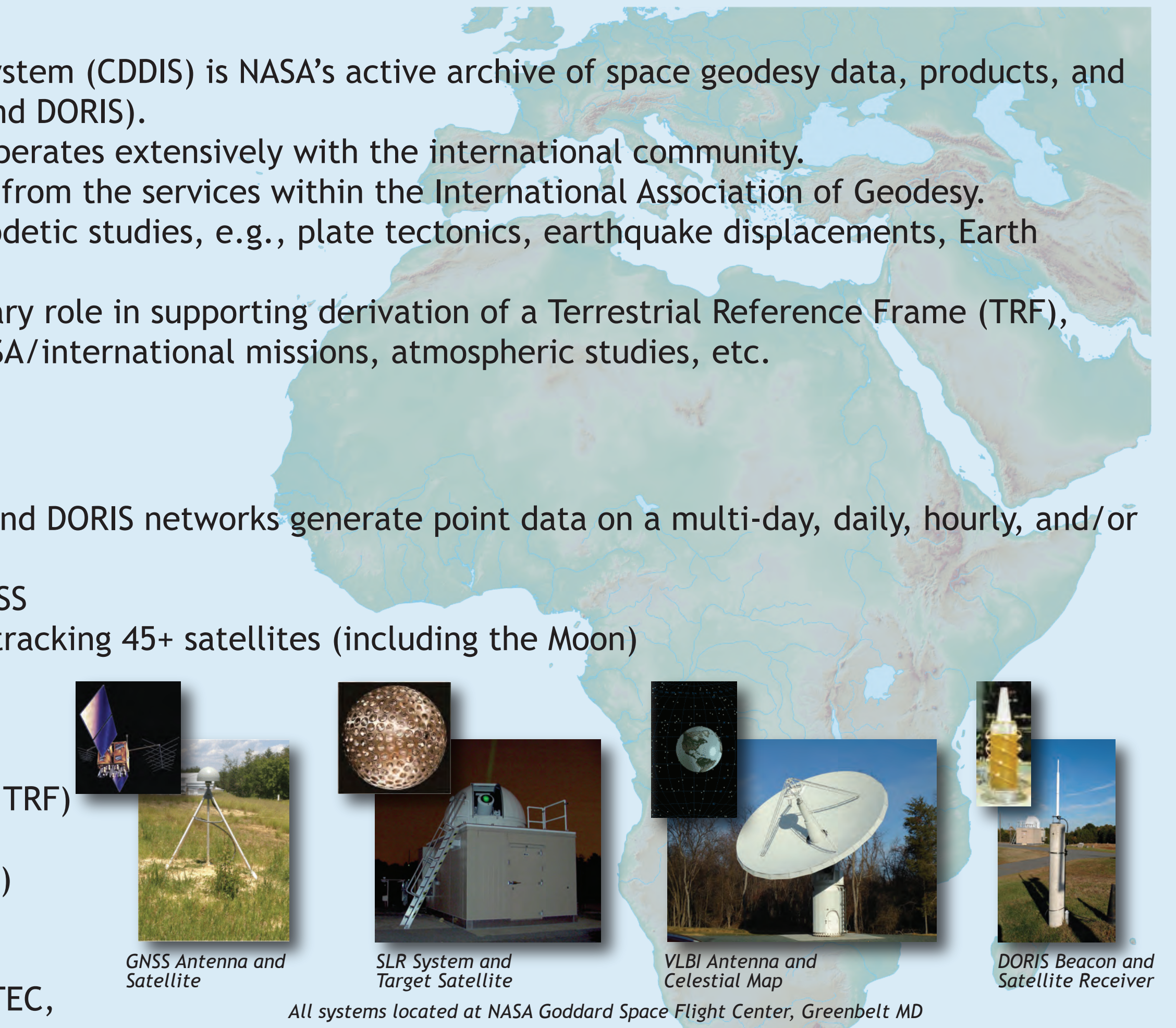
Crustal Dynamics Data Information System

Background:
- The Crustal Dynamics Data Information System (CDDIS) is NASA's active archive of space geodesy data, products, and information (GNSS, laser ranging, VLBI, and DORIS).
- The CDDIS funded by NASA/ESDIS but cooperates extensively with the international community.
- The largest CDDIS user community comes from the services within the International Association of Geodesy.
- Contents of CDDIS archive utilized for geodetic studies, e.g., plate tectonics, earthquake displacements, Earth orientation, etc.
- CDDIS archive also plays an interdisciplinary role in supporting derivation of a Terrestrial Reference Frame (TRF), precise orbit determination (POD) for NASA/international missions, atmospheric studies, etc.

Archive Contents:
- Data:
  - Stations in the GNSS, SLR/LLR, VLBI, and DORIS networks generate point data on a multi-day, daily, hourly, and/or sub-hourly basis
  - GNSS: 450+ sites tracking GPS, GLONASS
  - Laser Ranging (SLR and LLR): 42 sites tracking 45+ satellites (including the Moon)
  - VLBI: 45 sites
  - DORIS: 58 sites tracking 6 satellites
- Products:
  - 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) …
- Metadata information:
  - Non-standard metadata, data type specific
  - Extracted from incoming files
  - Internal access to metadata database

GNSS Antenna and Satellite *SLR System and Target Satellite* *VLBI Antenna and Celestial Map* *DORIS Beacon and Satellite Receiver*

All systems located at NASA Goddard Space Flight Center, Greenbelt MD

Current Space Geodesy Site Locations

● GNSS Site ▲ SLR Site ■ VLBI Site ▼ DORIS Site

Archive Statistics:
- File size is typically <2Mb/data "granule", <10Mb /derived product "granule"
- Archive size: ~8Tb
- Ingest rate: ~8.25Gb (75K files)/day
- Distribution rate: ~150Gb (~1.4M files)/day
- Data (L1, L1B), products (L2) derived from these data, and information about data and products
- Multi-day, daily, hourly, sub-hourly
- Varying latencies (minutes, hours, days)

Current Access Methods:
- 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.
- The majority of CDDIS user community are analysts supporting the services within the International Association of Geodesy.
- These groups produce derived products (e.g., positions of observing stations, Earth orientation parameters, precise satellite orbits, etc.) for use by a broader scientific community.
- 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.

User Interface Enhancements:
- New users of the CDDIS, both those familiar with space geodesy techniques as well as new research communities, would prefer a browsing interface to the archive contents.
- Furthermore, users also need to browse the archive for new or historic data sets.
- Therefore, the CDDIS has designed a web interface based search tool that queries the CDDIS metadata.
- Users have the ability to specify search criteria based on temporal, spatial, target, site designation, and/or observation parameters 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.

For More Information:
- 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.

## CDDIS Data Discovery Application Development

Background:
- Several years ago, the Scripps Orbit and Permanent Array Center (SOPAC) and UNAVCO collaborated to create the GPS Seamless Archive Centers (GSAC)
- GSAC enabled research by facilitating data discovery and distribution of GPS data from distributed GPS data centers
- CDDIS participated in the GSAC effort by its sharing GPS metadata to allow discovery of its GPS data holdings
- Today, SOPAC, UNAVCO, and CDDIS data centers are participating in an effort to modernize the GSAC (now GSAC-WS, the Geodesy Seamless Archive Center - Web Services) through a NASA ACCESS (Advancing Collaborative Connections for Earth System Science) proposal "Discovery and Delivery of Space Geodetic Data Products from Distributed Archives"
- These data centers have different data types and holdings as well as different metadata schemas

Development:
- The GSAC-WS ACCESS project involves three partner geodetic data centers (UNAVCO, SOPAC, and CDDIS) to facility data exchange and query web services; the University of Nevada, Reno is a user partner who will test the web services in their daily GNSS data processing activities.
- The ACCESS team has developed the GSAC Service Layer (GSL), a Java-based application that is used at each repository (CDDIS, UNAVCO, SOPAC) to interface to the data center-developed GSAC Repository Implementation. The GSL is a middleware framework that provides different types of output (HTML, XML, wget scripts, etc.) from the data repository metadata databases. The GSL handles the incoming web service requests and routes it to the repository. A federated Repository Implementation has been developed that allows for an aggregate search across all repositories.
- The CDDIS has developed a custom interface, tailored to CDDIS user requirements, through its own GSAC Repository client based on an open source application framework. This application currently interacts with the GSL for both Site and File searches. This custom interface allows CDDIS to leverage the capabilities of the GSL while providing an interface tailored to CDDIS users. Parts of this interface could eventually be made more generic and rolled into the GSL.

Data Discovery Enhancements:
- Develop a search/metadata interface tool for CDDIS to:
  - Aid users in discovery of CDDIS data, products, and information
  - Aid staff in archive management
  - Promote CDDIS data holdings to a larger community (e.g., through metadata standards)
- Specify (any/all):
  - Temporal: Year, date/time, range
  - Spatial: Region, lat/lon, range
  - Target: Satellite (SLR, DORIS)
  - Designation: Station name/number/code
  - Parameter: Receiver type (GNSS), event timer (SLR), antenna type (GNSS, VLBI), …
- Results:
  - List of sites satisfying specifications
  - List of data holdings satisfying specifications
 - Metadata relevant to selection
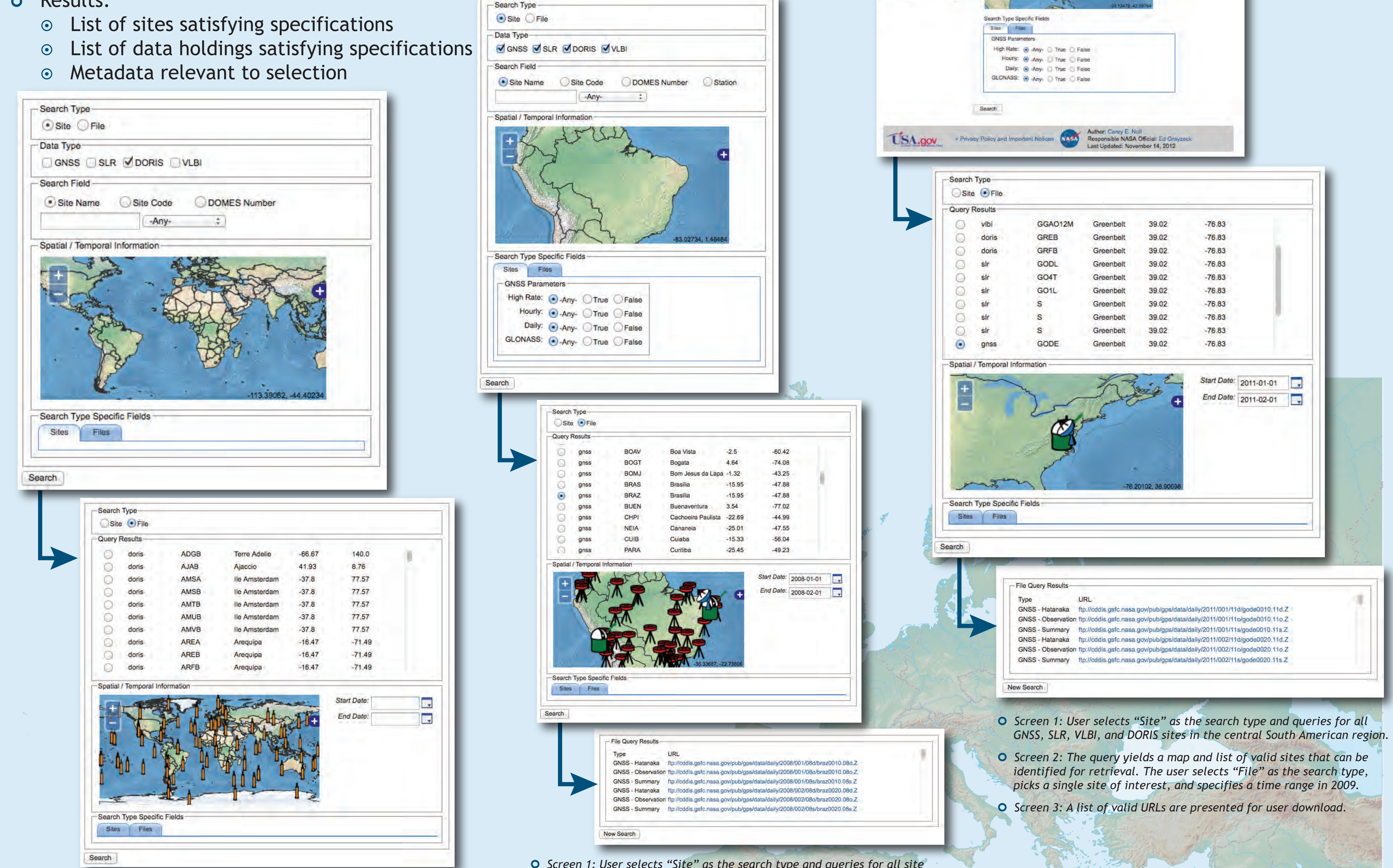
○ Screen 1: User selects "Site" as the search type and queries for all DORIS sites.
○ Screen 2: The query yields a global map and list of valid sites that can be identified for retrieval. The user can then select "File" as the search type, pick a site of interest and a time range.

○ Screen 1: User selects "Site" as the search type and queries for all GNSS, SLR, VLBI, and DORIS sites in the central South American region.
○ Screen 2: The query yields a map and list of valid sites that can be identified for retrieval. The user selects "File" as the search type, picks a single site of interest, and specifies a time range in 2011.
○ Screen 3: A list of valid URLs are presented for user download.

○ Screen 1: User selects "Site" as the search type and queries for all GNSS, SLR, VLBI, and DORIS sites by name (Greenbelt).
○ Screen 2: The query yields a map and list of valid sites that can be identified for retrieval. The user selects "File" as the search type, picks a single site of interest; and specifies a time range in 2009.
○ Screen 3: A list of valid URLs are presented for user download.
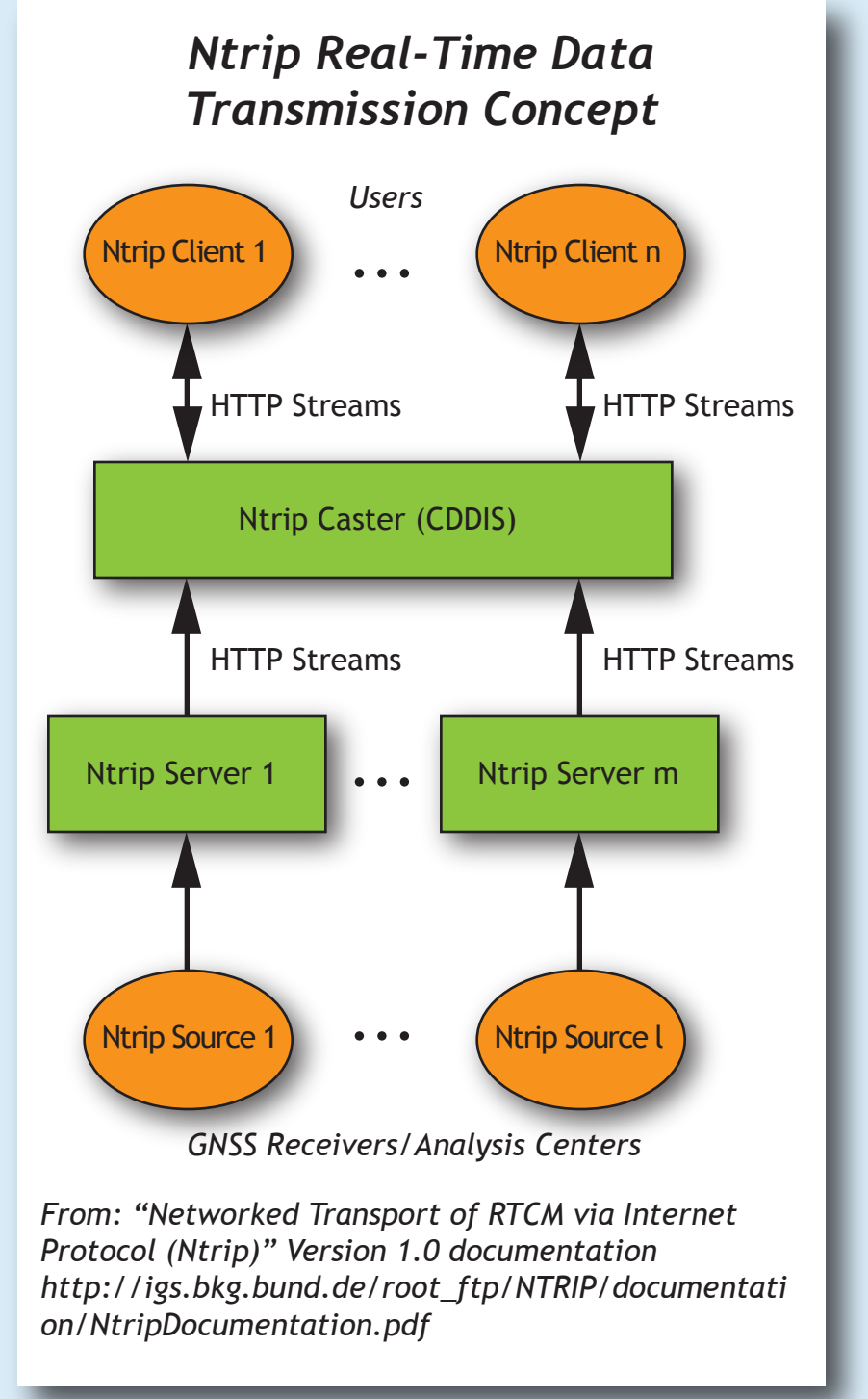
Other Developments and Future Plans

CDDIS Support of IGS Multi-GNSS Experiment (MGEX):
- The CDDIS is one of three IGS data centers supporting the IGS Multi-GNSS Experiment (MGEX).
- CDDIS expanded its data archive and distribution service to include data from participating multi-GNSS receivers, products derived from the analysis of these data, and any required metadata for the experiment.
- Archive now contains daily and hourly 30-second and sub-hourly 1-second data from an additional 40+ stations in RINEX V3 format.
- CDDIS developed software to extract metadata from the RINEX headers for use in reporting and archive management.
- Software also used to generate daily status files summarizing data holdings by site; however, data quality information available for operational GNSS data holdings is not available through this software.
- MGEX data are available within the campaign subdirectory structure in the CDDIS archive.

IGS MGEX Site Locations

◇ DAILY SITE ◇ HOURLY SITE ◆ HIGH-RATE SITE
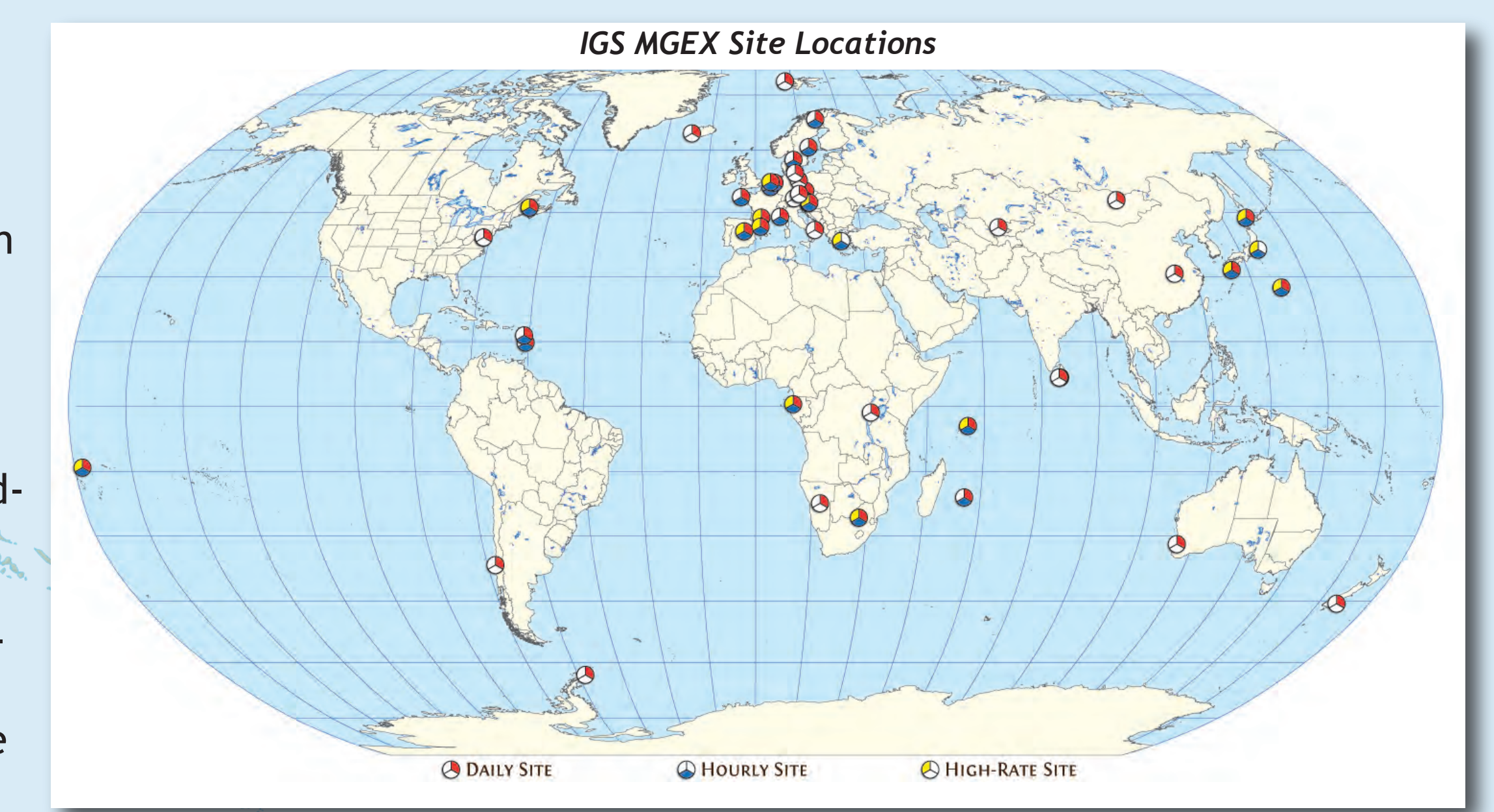
CDDIS Support of Real-Time Data and Products:
- In support of the IGS Real-Time Pilot Project (RTPP) and Real-Time IGS Service, the CDDIS has investigated the feasibility of establishing a real-time streaming capability at GSFC.
- System disseminates differential correction data or other kinds of GNSS streaming data to stationary or mobile users over the Internet.
- Mobile users obtain corrections/data from reference stations in real-time to improve positioning.
- CDDIS installed open source Ntrip (Networked Transport of RTCM via Internet Protocol) software on an available, dedicated server (called a caster) for receiving and serving (broadcasting) real-time GNSS data streams.
- Real-time operations:
  - NtripSource: Generates data streams at a specific location
  - NtripServer: Transfers the data streams from a source to the NTRIPCaster
  - NtripCaster: Major system component
  - NtripClient: Accesses data streams of desired NtripSources on the NtripCaster
- Real-time products include:
  - Data streams from a global network of high-quality GNSS receivers
  - Orbits
  - Satellite clock solutions
  - Ionosphere information
- Products enable real-time precise point positioning at global scales used in scientific research and hazard detection applications.
- CDDIS will also investigate process for capturing incoming streams and generating/comparing high-rate data files for archive.
- For operation at CDDIS, the NtripCaster software did not provide the NASA-required user authentication security.
- A module to plug into Ntrip will interface with an established User Registration System (URS) at NASA GSFC.
- Testing continues on the CDDIS caster installation and the Ntrip user registration module.

Ntrip Real-Time Data Transmission Concept

From: "Networked Transport of RTCM via Internet Protocol (Ntrip)" Version 1.0 documentation http://igs.bkg.bund.de/root_ftp/NTRIP/documentation/NtripDocumentation.pdf

CDDIS System Improvements:
- In May 2012 CDDIS operations were transferred to new server environment.
- This new configuration allows for efficient and timely processing of incoming files as well as enhanced system reliability by separating user/archive functions. Distinct servers handle incoming data and product files, outgoing ftp and http requests, real-time caster, and archive operations to the RAID storage.
- The new system is fully redundant with backup servers/clusters which can take over operations should a failure occur with the prime server. A secondary system installed in a second location at GSFC can accommodate CDDIS operations should the prime server be offline for an extended period of time. Both primary and secondary systems have backup systems.
- The archive is equipped with a multi-Tbyte RAID storage system and is scaled to accommodate future growth. The CDDIS archive increases in size by approximately 12 million files/1 Tbyte per year; the existing storage will accommodate the archive requirements for the near future.
- Additional archive backups are made to both local and off-site disk arrays.
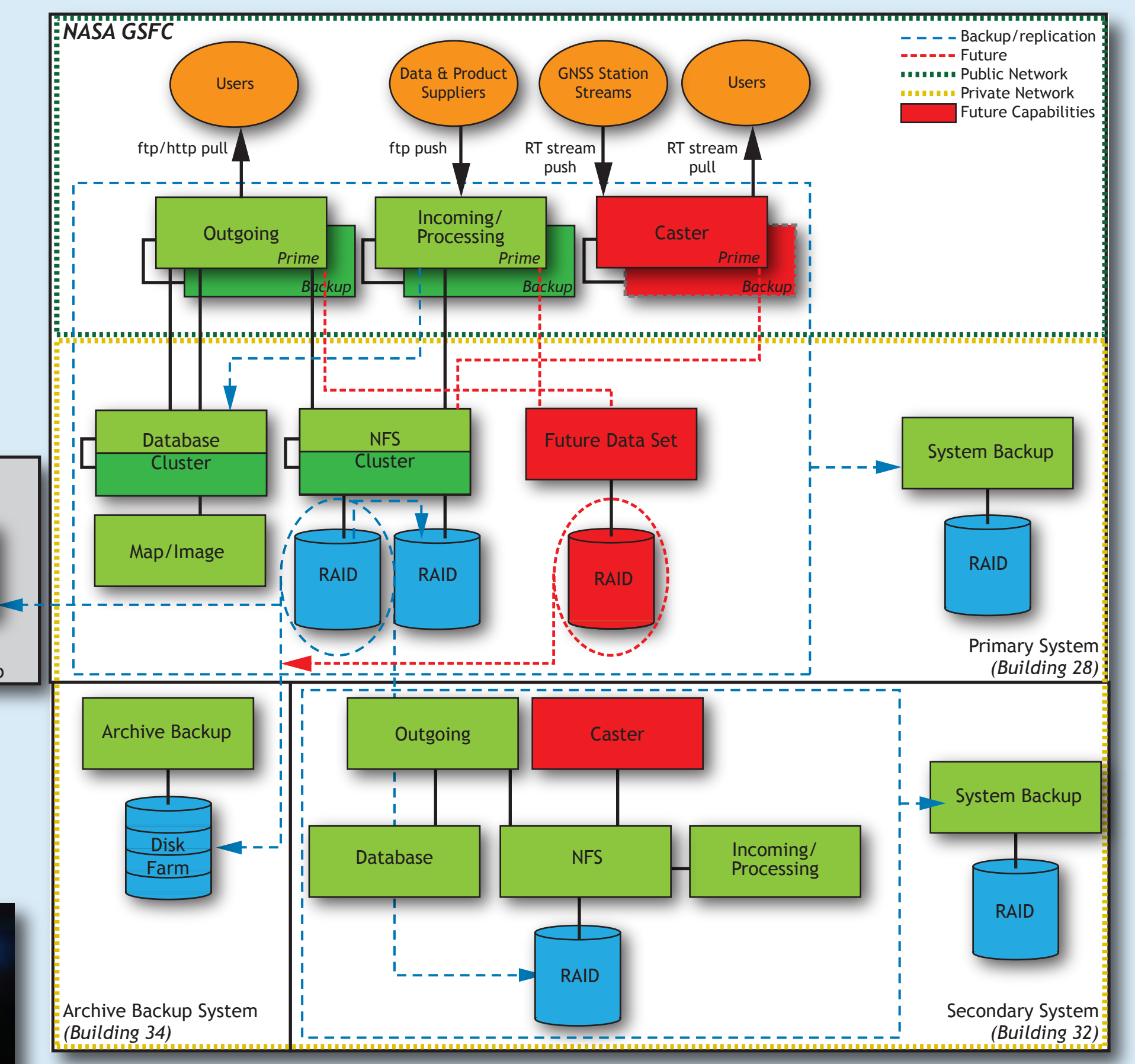
CDDIS Website Improvements:
- A design for a CDDIS website upgrade is currently underway.
- The new website will feature access to a map selection tool as well as the CDDIS data discovery application.
- An improved navigation design will also be incorporated.
- The new website will also provide links to the EOSDIS EarthData web applications.