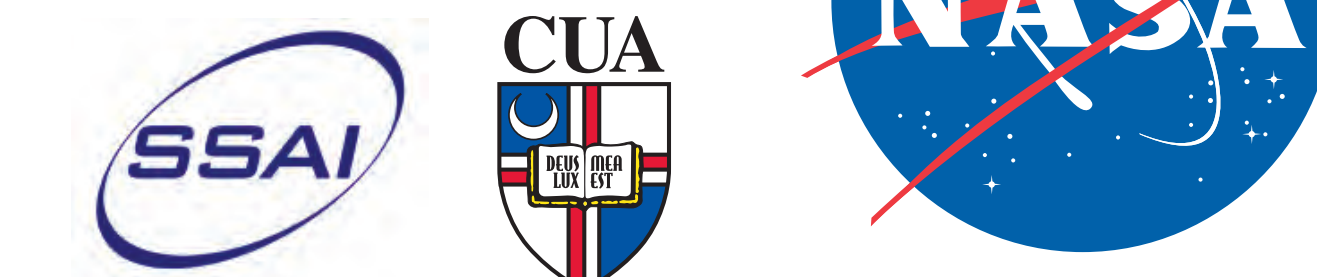


# IN41B-1410 Improvements in Space Geodesy Data Discovery at the CDDIS

**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 store space geodesy and geodynamics related data products in a central data bank, to maintain information about the archival of these data, and to disseminate these data and information in a timely manner to a global scientific 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. Several activities are currently under development at the CDDIS to aid users in data discovery, both within the current community and beyond. The CDDIS is cooperating in the development of Geodetic Seamless Archive Centers (GSAC) with colleagues at UNAVCO and SIO. The activity will provide web services to facilitate data discovery within and across participating archives. In addition, the CDDIS is currently implementing modifications to the metadata extracted from incoming data and product files pushed to its archive. These enhancements will permit information about CDDIS archive holdings to be made available through other data portals such as Earth Observing System (EOS) Clearinghouse (ECHO) and integration into the Global Geodetic Observing System (GGOS) portal. This poster will present the prototype implementation of these GSAC web services at the CDDIS as well as plans for the metadata enhancements to facilitate cross discipline data discovery.

Carey Noll (Carey.Noll@nasa.gov)  
Nathan Pollack/SSAI (nathan.pollack@ssaihq.com)  
Patrick Michael/Catholic University of America (Patrick.Michael@nasa.gov)

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

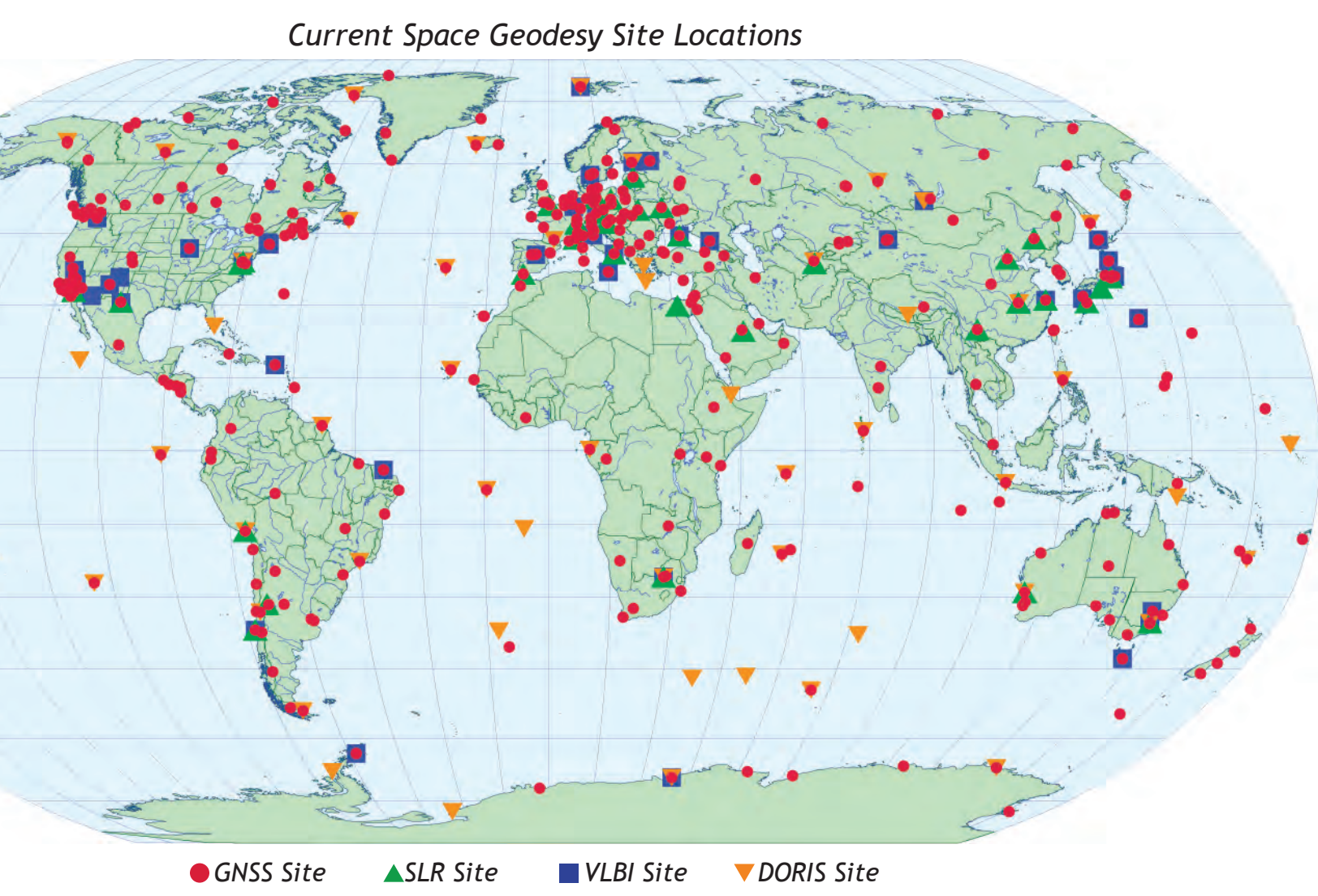
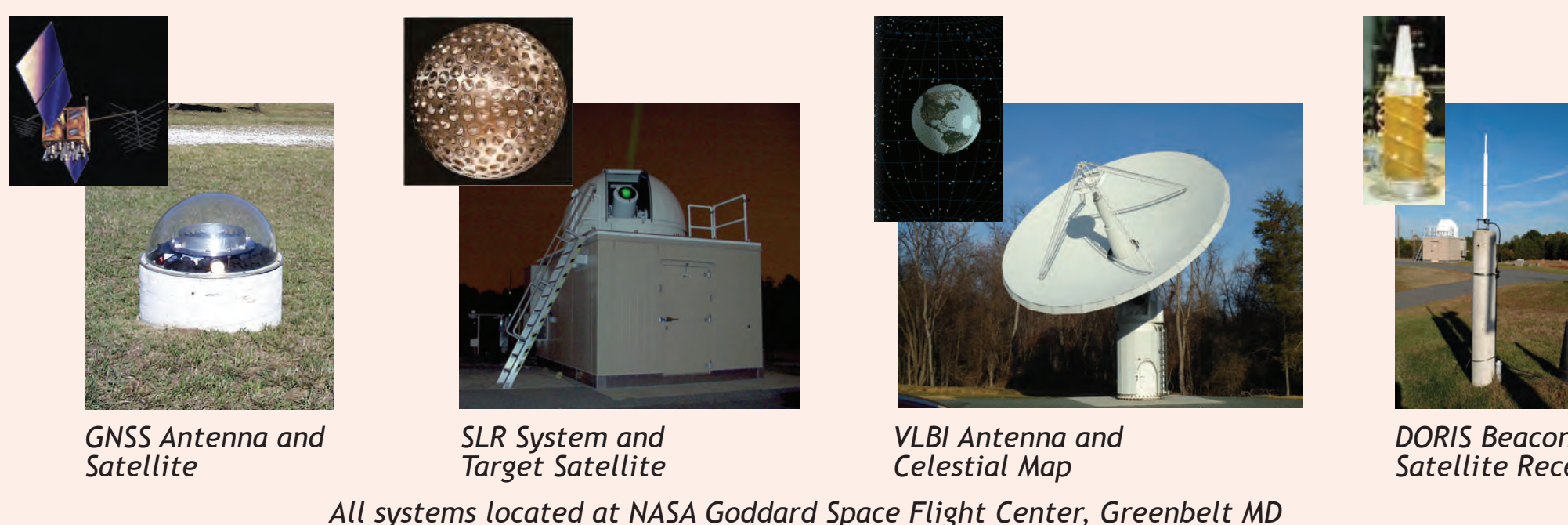


## Introduction to the CDDIS

- 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/EOSDIS 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 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



### Archive statistics:

- File size is typically <2Mb/data "granule", <10Mb/derived product "granule"
- Archive size: ~6Tb
- Ingest rate: ~3Gb/day
- Distribution rate: ~125+Gb/day, ~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)

### 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.

## Accessing the Contents of the CDDIS Archive

### 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.

## Data Discovery User Interface for the CDDIS

### 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

*The user specifies a query for all laser ranging (SLR) data to the LAGEOS-1 and LAGEOS-2 satellites from all sites in September 2011.*

File	Type	Site	Date	File size
lago1_1109	SLR	LAGEOS-1	2011-09-01 - 2011-09-30	N/A
lago2_1109	SLR	LAGEOS-2	2011-09-01 - 2011-09-30	N/A

*Query returns list of files containing LAGEOS-1 and -2 data for September 2011.*

Temporal search for SLR data from specified satellites

*The user queries for all space geodesy instruments found at the Greenbelt MD site.*

Site Code	Name	Type	Location (lat,lon,m)
CODE	Greenbelt	GNSS	39.05, 76.83, 0.0
GOSL	Greenbelt	SLR	39.05, 76.83, 0.0
GOSB	Greenbelt	DORIS	39.05, 76.83, 0.0

*The query returns a map with the Greenbelt location and list of space geodesy occupations in Greenbelt.*

Search for geodesy instruments at specified location

*Spatial and temporal search for GNSS sites and data in specified region*

*User queries for all GNSS sites providing daily 30-second GNSS data in Africa during October, 2011.*

*The query yields a map and list of valid sites that can be identified for retrieval.*

*The user asks for sites with all techniques (GNSS, SLR, VLBI, and DORIS) located in the region of Australia.*

File	Type	Site	Date	File size
slr00740114.2	GNSS - Hebridean	ABPO	2011-10-01 - 2011-10-01	N/A
slr00740114.2	GNSS - Observing	ABPO	2011-10-01 - 2011-10-01	N/A
slr00740114.2	GNSS - Summary	ABPO	2011-10-01 - 2011-10-01	N/A

*A list of valid filenames are presented for user download.*

*The query returns a map and list of sites satisfying the search criteria.*

Search for geodesy sites in specified region

## Future Plans

- Phase 1 of the data discovery development at the CDDIS is currently underway. This effort will continue through early 2012, concentrating on presenting a form interface to retrieve information on sites and files for download satisfying various search criteria (temporal, spatial, and other parameters).
- Phase 2, beginning in mid-2012, will provide enhancements to the forms developed in Phase 1 based on user feedback. Additional development will address allowing users to download identified data sets of interest through other methods, such as wget and export site information in various formats such as KML. Development on a client script interface will also be pursued to allow users to create scripts for automated downloads.
- Phase 3 development, beyond 2012, will further refine the data discovery interface and expand the application to include queries for derived products.

## Acknowledgements

The authors would like to acknowledge the development performed under the NASA ACCESS Proposal "Discovery and Delivery of Space Geodetic Data Products from Distributed Archives" (proposal number 09-ACCESS09-0034), PI Fran Boler, and Co-Is/collaborators/team members Jeff McWhirter, Louis Ratzesberger, Melinda Squibb, Yehuda Bock, Geoff Blewitt, Corné Kreemer, and Carey Noll.

For more information on the development of the GSL please visit poster IN23B-1457, "GSL: An Open Source Framework for the Rapid Development of Data Archive Access Services" Jeff McWhirter, Frances Boler, Yehuda Bock, Melinda Squibb, and Louis Ratzesberger in Software Reuse and Open Source Software in Earth Science II Posters session.