Discovery and Delivery of Space Geodetic Data Products from Distributed Archives

Final Technical Report

Submitted by Frances Boler, PI
UNAVCO, Inc.
6350 Nautilus Dr.
Boulder, CO 80305

Performance Period: April 1, 2010 through December 31, 2012

NASA Grant Number NNX10AF07A

Submitted to

Stephen Berrick, NASA Headquarters, Technical Officer
Martha Maiden, NASA Headquarters, Technical Officer
NASA/NSSC, Grants Officer

Project Accomplishments

The GSAC-WS project effort centers on modernizing the mechanisms for exchange of metadata and data for space geodetic data products. The technology development and implementation to enable exchange of metadata for space geodetic data and products has been completed and the three partner archives (UNAVCO, SOPAC, and CDDIS) all have working implementations of the GSAC-WS software, which was developed through this project. In addition, the science partner (University of Nevada Reno, UNR), has completed a reanalysis of 15 million GPS files and computation of the QA/QC parameters proposed for this project. The science partners use the GSAC-WS services to gather data and metadata, and continue daily analysis of over 10,000 stations. The GSAC-WS system that provides catalog and file information via web services improves discoverability and access to available products including the new QA/QC products.

Geodesy archive repository implementations are running at the three project partner archives at UNAVCO, SOPAC, and CDDIS. A federated repository that allows querying of all three archives is operational at UNAVCO.

In addition to the three partner archive implementations of the GSAC-WS system, UNAVCO has implemented two additional GSAC repositories for the purpose of facilitating metadata exchange. One of these repositories provides metadata for the International GNSS Service (IGS) stations system as part of UNAVCO's NASA funded support to the IGS Central Bureau. The second GSAC repository provides metadata for the Plate Boundary Observatory's (PBO) analysis of a set of several hundred stations not in the UNAVCO, SOPAC, or CDDIS repositories. The metadata are gathered via text site log files from the source repository, ingested into a metadata database running on a system with the GSAC-WS repository software at UNAVCO. Metadata is then served to the PBO Analysis Centers through GSAC web services.

In order to fully realize the potential for this technology to enable data and product discovery and access, it would be highly advantageous if there was broad adoption of this technology among the numerous geodesy archives both in the US and elsewhere. Recently, UNAVCO has promoted the GSAC-WS project within the US and internationally as a mechanism to facilitate federation of data and product search and delivery, and has made progress in its effort towards expanding adoption of this technology. On the US side, UNAVCO has had discussions with the University of California, Berkeley, and NOAA's National Geodetic Survey about potentially installing GSAC repository software at their archives, and having these repositories participate in the federated system. In an international context, and in synergy with an NSF and European cooperative effort in cyberinfrastructure development, UNAVCO is promoting GSAC technology with several European geodesy data repositories that are collaborating through EPOS, the European Plate Observing System. EPOS participants who are potentially interested in GSAC-WS technology include Istituto Nazionale di Geofisica e Vulcanologica (INGV), the European Reference Frame (EUREF), and the University of Beira Interior, Portugal.

Summary of No-Cost Extension Period Activities

The project has completed a nine-month no-cost extension phase in addition to the original two-year award. The accomplishments achieved during the no cost extension are the following:

- Implementation of weekly runs of the UNR QA/QC product generation completed by UNR.
- Documentation of the QA product was developed by UNR and is available from UNAVCO.
- Archiving and dissemination of the QA products in a production capacity and using GSAC-WS for dissemination completed by UNAVCO.
- Migration of SOPAC's GSAC-WS implementation to a production server was completed.
- SOPAC has formulated plans for incorporating links to their GSAC-WS API and GUI from their website when they perform a planned web site upgrade.
- Completion of the CDDIS GSAC-WS implementation and migration to a production, publicly query-able system was completed.
- Expanded documentation aimed at repository managers and software engineers for implementation of a GSAC WS repository was developed by UNAVCO.
- Documentation for users of GSAC web services was expanded by UNAVCO.
- Documentation of standard and optional GSAC-WS metadata was developed by UNAVCO and will be further refined.
- A project wrap-up meeting was held in November with representatives of all partners present (UNAVCO, SOPAC, UNR) or attending via WebEx (CDDIS).

Although the technology development is complete, ongoing refinements and functionality enhancements will be supported by UNAVCO as GSAC-WS expands into new repositories. As new users and repositories exercise the system and software, documentation will be refined and enhanced. All three partner archives are committed to ongoing operation of their repositories through existing funding, and UNR will continue generation of the QA products.

Project Activities Summary

During the first year of the project, the initial release of the basic technology component, the GSAC Service Layer (GSL), was developed and implementation began at each of the partner archives. Because the details of database structure are different for each archive, implementation of the GSL includes customization of queries within the Repository component.

The major components of the GSAC-WS system are shown in Figure 1. The ability to utilize the GSL as a mechanism to federate queries is achieved by implementing the Federated Repository component, shown in Figure 2. The federated repository is a separate instance of the GSL software that is configured to send queries in parallel to the selected, affiliated repositories and to aggregate the results for presentation to the user.

The GSAC Repository implementation includes a web Graphical User Interface (GUI) that is an optional front end to the web services. Figure 3 shows this interface to UNAVCO's GSAC. In addition, a Java client that can be downloaded to the users system provides a second interface to the web services. Of course, web services users can directly access the web services using RESTful URLs. Online documentation of the web services is available through the Repository GUI; UNAVCO's Repository web documentation is shown in Figure 4.

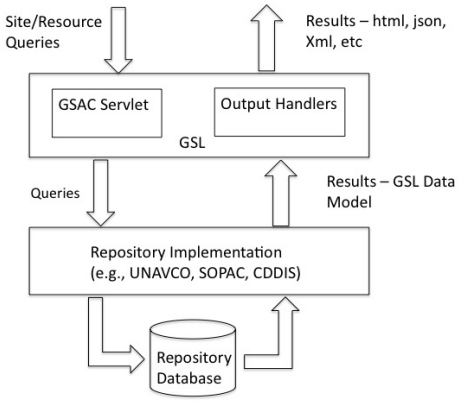

Figure 1: GSAC Service Layer Architecture

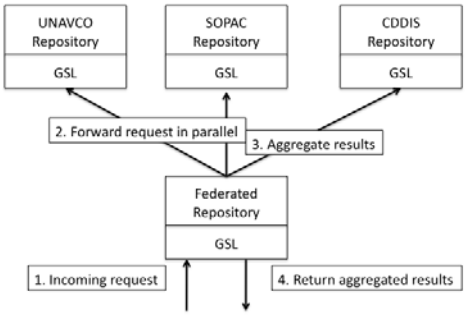

Figure 2. Federated Repository

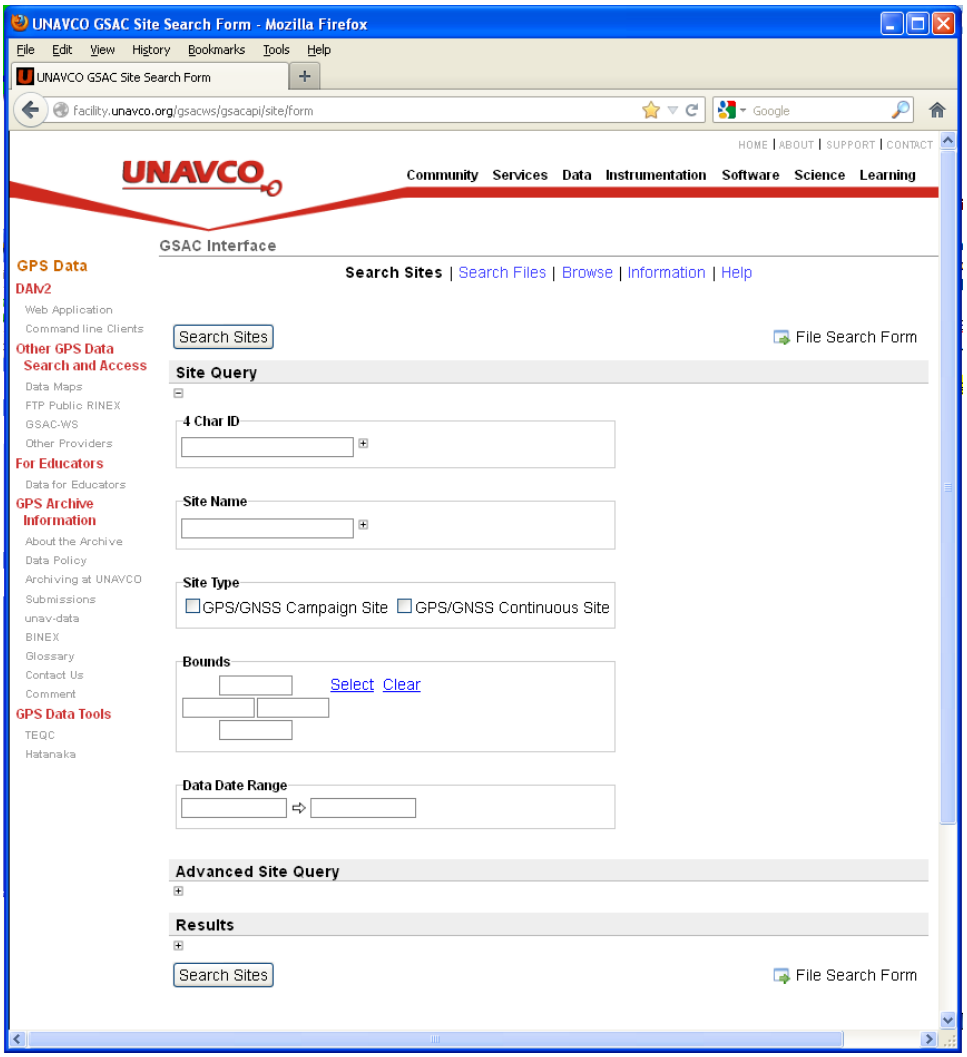

Figure 3: UNAVCO's GSAC Web GUI. Each GSAC Repository can utilize the Web GUI as part of their GSAC implementation if desired. Elements may differ based on the Repository's capabilities.

The GSL technology was refined during the second year of the project as each of the archive partners worked to achieve implementations that provided robust access to their metadata, data and products; for CDDIS this involved defining and implementing extensive customization to their GSAC Repository in order to facilitate access to their full suite of geodesy data types. While the component software went through several releases during the year, the bulk of the work at the partner archives was in implementation, customization, and readying the systems for production and metrics reporting. During the second year the GSAC -WS system metadata for search and access to real time streams for GPS/GNSS data was finalized and put in production in UNAVCO's GSAC Repository. SOPAC extended its Site Log XML format to incorporate real time stream metadata. Metrics reporting systems were defined and implementing, with logging and parsing scripts provided by UNAVCO.

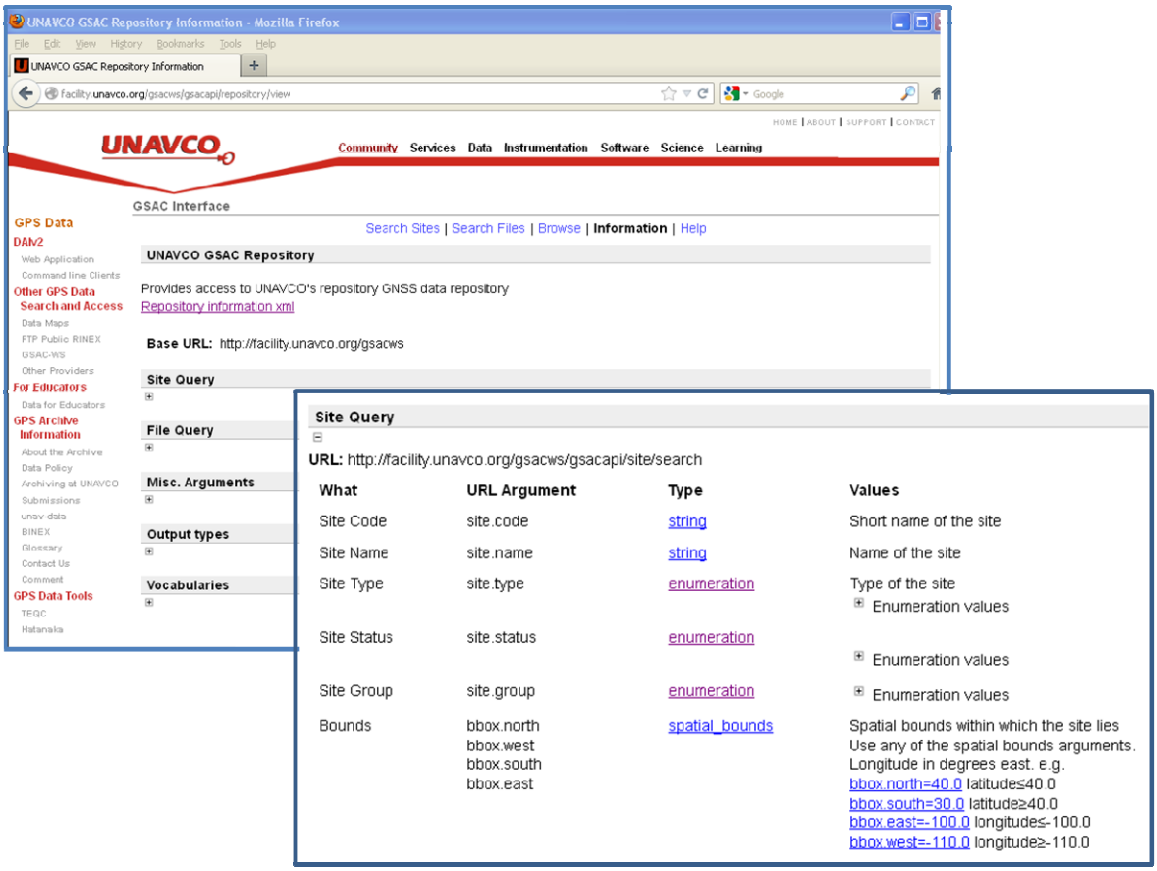

Figure 4: UNAVCO's web-based documentation for using GSAC web services. Each repository has unique documentation that is auto-generated by the GSL software.

During the final nine months of the project (no-cost extension period), UNR's QA computation system was brought into final production, with archiving of the products and metadata distribution through GSAC we b services handled by UNAVCO. UNAVCO efforts focused on refining documentation available to users of the GSAC system, and adding output metadata formats that were requested by the IGS and PBO including the SINEX format and the GAMIT station.info format. For geodesy repositories, additional documentation was created to assist the repositories in system installation. SOPAC moved its GSAC system onto a production server and finalized metrics computation. CDDIS also moved its GSAC Repository to a new production system, with access to the public.

The project efforts by each partner are summarized below.

Partner Collaboration Activities

The teams from UNAVCO, CDDIS, SOPAC, and the University of Nevada, Reno met in a face-to-face meeting at UNAVCO in Boulder, twice and at SOPAC twice over the course of the project. Team members also met informally at the UNAVCO Science Workshop in 2012 and at AGU. Monthly teleconferences were held with partner PIs and technical personnel in attendance.

A collaborative team project wiki and mail forum system was created on Google Sites/Groups to facilitate project management and interactions. The mail list and Google sites website was used throughout the project for planning, to post documents, and track the project schedule.

During the kickoff meeting in 2010, the partners had an initial agreement on metadata for incorporation in the GSAC and the architecture of the system. A document was created and the initial GSAC-WS software was based on this definition the GSAC metadata.

UNAVCO

During the first year of the project, UNAVCO developed the base repository software for the GSAC Service Layer, with input from the other partners. The GSAC-WS Repository provides the GSAC-WS Web Service API and a web query interface that utilizes the API. It provides the ability to do a variety of GNSS site queries (e.g., spatial, temporal, equipment, etc) as well as queries on data holdings.

The GSL is based around a Model-View-Control paradigm. The GSAC Servlet is the controller, the underlying Repository is responsible for constructing the model and the Output Handlers create the views of the content. This decoupling of model from view is critical in that it enables the rapid development of new encodings and result formats. Various representations of the results are user selectable including Site Lox XML for metadata, Google Earth KML, Javascript Object Notation (JSON) and CSV.

The GSAC Repository template package and code generation capabilities allow a repository developer to rapidly generate the Java code that provides a basic implementation of a repository, vastly reducing the overall development effort. This package generation process also includes the generation of the build and release scripts. We have made use of this template mechanism in all three implementations of the GSAC-WS repositories (UNAVCO, SOPAC, CDDIS, the partner federated repository, an IGS metadata repository, and a PBO expanded analysis metadata repository).

In addition to adding to the overall system capabilities and working with the other partners to improve their implementations, UNAVCO developed the real-time query capabilities of the GSAC-WS software system. UNAVCO experimented with a separate repository for the Plate Boundary Observatory real-time holdings because that system employs an independent database within UNAVCO. Ultimately, the PBO real-time metadata delivery was converted to an in-house web service built on the PBO real-time database. This system provides connection information and statistics about each real-time stations data completeness and latency.

UNAVCO also developed code for logging GSAC-WS system queries and URLs returned (URLs returned are a measure of products delivered via the GSAC-WS system). These logging capabilities were included in a system release for dissemination to project partners. Separate scripts for parsing the GSAC-WS logs were developed and released to the archive partners. In addition, a system for reporting partner metrics to UNAVCO utilizing Google Docs spreadsheet was implemented.

SOPAC's Site Log XML with extensions for real-time station metadata was incorporated as an output encoding for GSAC-WS metadata.

The UNAVCO Repository web GUI can be accessed at:

http://facility.unavco.org/gsacws/gsacpi/site/form (site search)

http://facility.unavco.org/gsacws/gsacapi/file/form (file search)

http://facility.unavco.org/gsacws/gsacapi/repository/view (repository documentation)

The UNAVCO Federated Repository web GUI can be accessed at:

http://facility.unavco.org/gsacfederated/gsacpi/site/form (site search)

http://facility.unavco.org/gsacfederated/gsacapi/file/form (file search)

http://facility.unavco.org/gsacfederated/gsacapi/repository/view (repository documentation)

SOPAC

SOPAC staff contributed to the definition and refinement of the GSL and Repository software developed principally at UNAVCO. SOPAC added to the XML encodings of real-time metadata for augmenting their Site Log XML. SOPAC contributed documentation to the project which provides a comprehensive description of the GSAC-WS system in comparison to the legacy GSAC flat file exchange system.

The SOPAC Repository web GUI can be accessed at:

http://geogsac.ucsd.edu:8080/gsacws/gsacapi/site/form (site search)

http://geogsac.ucsd.edu:8080/gsacws/gsacapi/file/form (file search)

http://geogsac.ucsd.edu:8080/gsacws/gsacapi/repository/view (repository documentation)

SOPAC has GSL and Repository implementation is running on a production virtual server.

SOPAC extended their Site Log XML format to incorporate real time metadata.

Over the course of the project, SOPAC developed documentation for GSAC users comparing the original GSAC system to the new GSAC-WS system. The document allows users of the old GSDAC system to migrate to the new GSAC-WS.

CDDIS

The CDDIS staff completed development of a custom interface to the GSAC Service Layer (GSL) for use on the CDDIS website. This work leverages the capabilities of the GSL while maintaining a consistent user experience for those exploring the CDDIS website for data holdings. The interface allows for more customization and will allow users to search for sites and files by specifying spatial, temporal and other parameters in a way that is unique to CDDIS data holdings. The more generic parts of this work can be rolled into the GSL for use by other partners.

The new interface was developed using an open source application framework and interacts with the GSL for searches. The custom user interface allows collection of fields related to both site and file searches. The new client requests raw data from the GSL and displays the results in JavaServer Pages, combining the results with a customized OpenLayers map.

The CDDIS Repository web GUI can be accessed at:

http://cddis.gsfc.nasa.gov/gsacws/gsacapi/site/form (site search)

http://cddis.gsfc.nasa.gov/gsacws/gsacapi/file/form (file search)

http://cddis.gsfc.nasa.gov/gsacws/gsacapi/repository/view (repository documentation)

UNR

The primary task of the UNR team was to provide quality assessment statistics on the GPS data, which UNAVCO could then distribute to users. The statistics are based on post-analysis rather than pre-screening, as it more closely meets user needs. To meet this objective the UNR team downloads data from all continuous GPS stations that are in the UNAVCO/SOPAC/CDDIS archives. They process the data as part of their global analysis. During this project UNR changed to using JPL's ITRF2010/IGS08 products and GIPSY 6.1 software with a new single-station ambiguity resolution method known as WLPB (Wide-Lane and Phase Bias), recently developed by JPL. The re-analysis of all data (which include more than those in the aforementioned archives) took ~3 months. From this analysis, UNR now routinely extracts various parameters concerning the quality of each RINEX file. These parameters include number of phase biases, residual scatter of the observations, satellites tracked, number of outliers, formal errors, and level of multipath. These products have been provided to UNAVCO, thus closing the loop. The positive benefit of using the WLPB method of single-station ambiguity resolution (instead of the now-deprecated Ambizap network algorithm) is that data that arrives late can be included and quality-assessed without affecting all other stations. The routine update of the solutions and QA parameters is uploaded to UNAVCO weekly.

The second major task of the project was for the UNR partners to use the gsac client in their data ingestion software. Because that software is operating well, it was decided not to replace it with the gsac client and to use the gsac client only as a new feature in UNR's software that it currently lacks: to easily pick up data that has arrived late (i.e., after UNR's latest fetch) in the UNAVCO/SOPAC/CDDIS archives.

In parallel activity at UNR that closely relates to this project, UNR in September 2011 started a prototype public "next-day" service that provides time series and post-analysis statistics on the scatter of 5-minute precise point positions with ambiguity resolution. These QA statistics are currently provided for over 1,000 stations in the western US, and provide rapid feedback on low-latency data, such as the growing number of stations with real-time capability. As they are public, these data are also available for UNAVCO pickup, and could also be made part of the existing QA stream. Results are very encouraging, as the RMS scatter of 5-minute positions are strongly site-dependent, and they perhaps provide the most relevant statistic to assess the damaging effect of multipath on GPS positions.

ESDSWG Participation

The UNAVCO software engineer participated in the Technology Infusion Working Group (TIWG), and attended the Software Reuse Working Group telecons and the Service Interoperability and Orchestration (SIO) group telecons. Posters were presented at the 2010 and 2011 ESDSWG meetings.

Project Presentations

Blewitt, G. and C. Kreemer, A System to Produce Precise Global GPS Network Solutions for all Geodetic GPS Stations in the World, 2010, Eos Trans. AGU, Fall Meeting Suppl., Abstract G22A-02.

Boler, F., C. Noll, Y. Bock, C Kreemer, and G. Blewitt, Discovery and Delivery of Space Geodetic Data from Distributed Archives, Poster presented at the IGS Infrastructure Meeting, 2010.

Boler, F., G. Blewitt, C. Kreemer, Y. Bock, C. Noll, J. McWhirter, P. Jamason, and M. Squibb, Geodetic Seamless Archive Centers Modernization – Information Technology for Exploiting the Data Explosion, 2010, Eos Trans. AGU, Fall Meeting Suppl., Abstract G23B-827.

McWhirter, J., F. Boler, Y. Bock, P. Jamason, M. Squibb, C. Noll, G. Blewitt, and C. Kreemer, The Geodetic Seamless Archive Centers Service Layer: A System Architecture for Federating Geodesy Data Repositories, 2010, Eos Trans. AGU, Fall Meeting Suppl., Abstract IN11B-1085.

McWhirter, J., F. Boler, Y. Bock, M. Squibb, and L. Ratzesberger, An Open Source Framework for the Rapid Development of Data Archive Access Services, 2011, Eos Trans. AGU, Fall Meeting Suppl., Abstract IN23B-1457.

Noll, C., N. Pollack, P. Michael, Improvements in Space Geodesy Data Discovery at the CDDIS, 2011, Eos Trans. AGU, Fall Meeting Suppl., Abstract IN41B-1410.