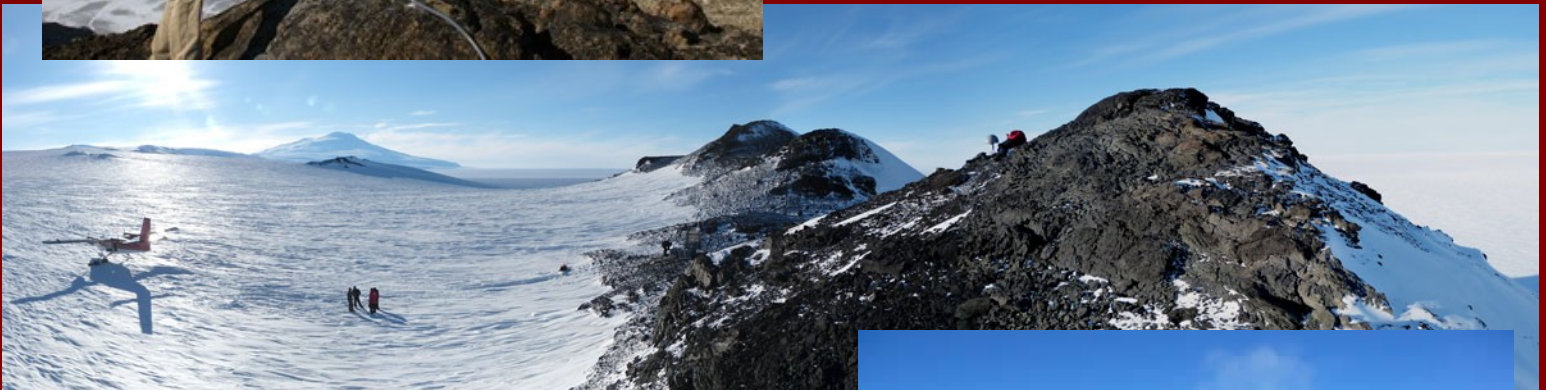# UNAVCO GSAC WS:

Web Services for Geodesy Data Repositories

UNAVCO GSAC WS: Web Services for Geodesy Data Repositories

Stuart Wier, Fran Boler, Jeff McWhirter (UNAVCO), and SOPAC Staff

GSAC is online at http://facility.unavco.org/data/gsacws/

November 2012; April 2013; September 2013; November 27, 2013.

Cover photographs from UNAVCO:

Hjornefjedet, Greenland. GNET 2007
Uturuncu volcano, Bolivia. PLUTONS Project Part II
Toney Mountains, Antarctica. POLENET
Green Mountain Reservoir, Colorado. Monument RG16, Rio Grande Rift Network
Mt. Erebus, Antarctica. LIDaR station G-081, 2009

UNAVCO: http://www.unavco.org/

CDDIS: http://cddis.nasa.gov/

SOPAC: http://sopac.ucsd.edu/

UNAVCO
6350 Nautilus Drive
Boulder, Colorado U.S.A.

GSAC Web Services

The Geodesy Seamless Archive Web Services, called GSAC Web Services, GSAC-WS, or simply GSAC, is a software package from UNAVCO and partners SOPAC and CDDIS, providing comprehensive user access to a broad range of geodetic data, including GPS/GNSS, DORIS, SLR, and VLBI. Any organization with geodesy data files, and information about that data, in a database can readily implement a GSAC repository utilizing GSAC code. GSAC is based on the basic concept of *sites* at precise locations with *instruments* making *data files*. Metadata in the database includes details about sites, instruments, times, and data files.

GSAC is a suite of free, open source, code, with related documentation for installation and use. GSAC is middleware connecting your database and file system on one hand and remote users needing to query metadata or to access data files on the other. From the user perspective, GSAC provides web services for data search and access from a data repository, allowing queries about sites and instruments, and downloads of instrument data files. A separate document on the UNAVCO GSAC web site, the *UNAVCO User Guide for GSAC-WS Data Repository Access*, explains how to use GSAC services.
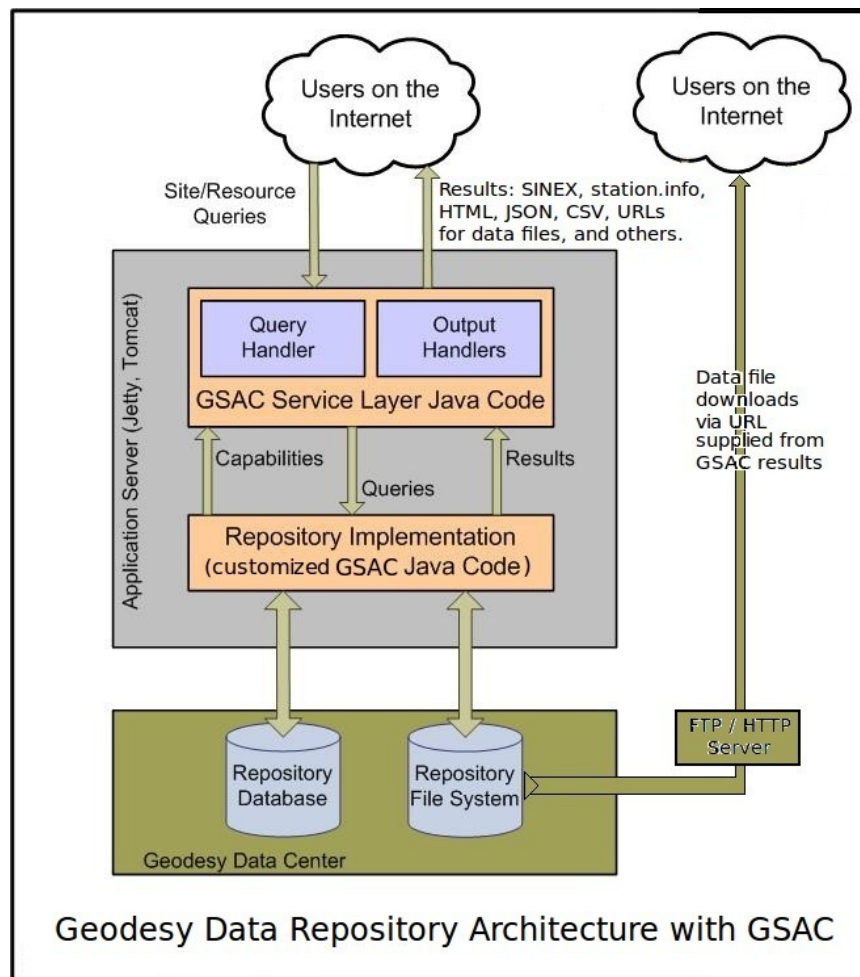
GSAC has a data model common to many geodesy data centers and to some other Earth and environmental sciences data centers. For example, the sites-plus-instruments paradigm is utilized within GSAC, and metadata about sites and the instrumentation installed at sites are typically elements of the database schema for such a

repository. For files containing the observations, metadata about the data files is stored in the database and typically includes file directory location paths, instrument types, start and end times of data, file sizes, etc. Searches and downloads of data from geodesy data centers often use the same or similar functions such as queries for site codes, IDs and names, and data types, geographic locations, and times. The GSAC implementation provides code for these common needs, plus an ability to tailor details specific to a particular repository.

GSAC code provides an extensive working framework with rich functionality, and with the flexibility to implement queries on parameters about instruments, sites, and instrument data files. Like other web services, GSAC accepts incoming URL-based requests, handles the request, and returns a result. The format of GSAC query results is user-selectable, including HTML (web pages), SINEX, GAMIT station.info, CSV ASCII files, SOPAC XML station logs, and plain text for easy human verification. For data access, GSAC can return a list of URLs for FTP, and WGET scripts for downloading data. There is also a GSAC command line client, a Java-based client for accessing a GSAC repository, which allows users to do programmatic searches of a GSAC repository and to download files without use of web page forms or menu clicks. Each GSAC web site has a home page about the agency, and online help about GSAC use for remote users.

A GSAC service publishes on its web pages its repository information in XML, a capabilities document for the API search. This information is read by a federated GSAC implementation to work with other GSAC repositories. Any GSAC federated repository may represent two or more GSAC data centers. The GSAC package can support independent stand-alone repositories, and any number of independent federations of repositories.

GSAC enables powerful tools for discovery, sharing, and access to data. GSAC has a well-defined purpose, to provide complete and consistent web services, for existing geodesy data repositories, for example. GSAC provides a complete suite of web services for remote users to discover and download geodesy information and data files from a data repository, by hand or with automated queries.

GSAC is designed for the data sharing principles (http://www.earthobservations.org/geoss_dsp.shtml) of the Group on Earth Observations (GEO, http://www.earthobservations.org/):

> *All shared data, metadata and products being free of charge or no more than cost of reproduction will be encouraged for research and education*.

In general GSAC offers free access to data files and site metadata without restriction. GSAC provides a way to hide some selected data files and/or some site metadata in your data center from public searches. If you wish to provide user authentication, you need to add that as part of your data file download server. This applies to user data file downloads, not to GSAC access. GSAC does not have user accounts. GSAC does not track or authenticate users, or manage accounts and passwords.

New GSAC web servers are being installed at INGV/RING (Grottominarda IT), SEGAL (University of Beira Interior, Covilha PT), NOANET (National Observatory of Athens), Geoazur (Valbonne, Nice, France), the European Permanent GNSS Network Central Bureau (Royal Observatory of Belgium), and for the IS-CGPS network, Icelandic Meteorological Office (IMO), Rejkyavik.

As part of the NASA ROSES ACCESS Project "Discovery and Delivery of Space-based Geodesy Data Products," several GSAC implementations were installed, including the UNAVCO GSAC Repository, the SOPAC GSAC Repository at Scripps Orbit and Permanent Array Center, the CDDIS GSAC Repository at NASA's Crustal Dynamics Data Information System, and a federated GSAC Repository at UNAVCO. The federated GSAC Repository at UNAVCO provides searches of the UNAVCO, SOPAC, and CDDIS repositories in parallel, with aggregation of the results. Together these repositories represent some 10 000 instrument sites and 90 million data files.

Single and Federated GSAC Data Repositories

GSAC provides up-to-date and complete web services at geodesy data repositories, for search, discovery, and retrieval of data. GSAC software supports single repositories, and also becomes particularly powerful when used in the federated mode for search and download from several different GSAC-enabled data repositories. A federated GSAC system dynamically queries participating GSAC repositories directly through each one's GSAC interface, and aggregates results from two or more GSAC data centers. A federated GSAC repository does not maintain a database about other data holdings, nor does it copy or mirror collections of data files.

A single GSAC repository has an associated database and (usually) related data files. It also has a web site and API access capabilities for data discovery and download by remote users, from that repository. In the diagram above, the GSAC code is inside the gray box. The GSAC Service Layer has the core GSAC code to provide the functionality; it is complete and included in the GSAC source package. The local GSAC Repository Implementation is software tailored to read your database and with information about data files to download.

GSAC at a data repository provides data search, discovery and download mechanisms to remote users, for that one repository. A *federated* GSAC server runs without a local database or data file system of its own; it is a GSAC web service for joint searches and data retrieval across multiple GSAC-enabled data centers.
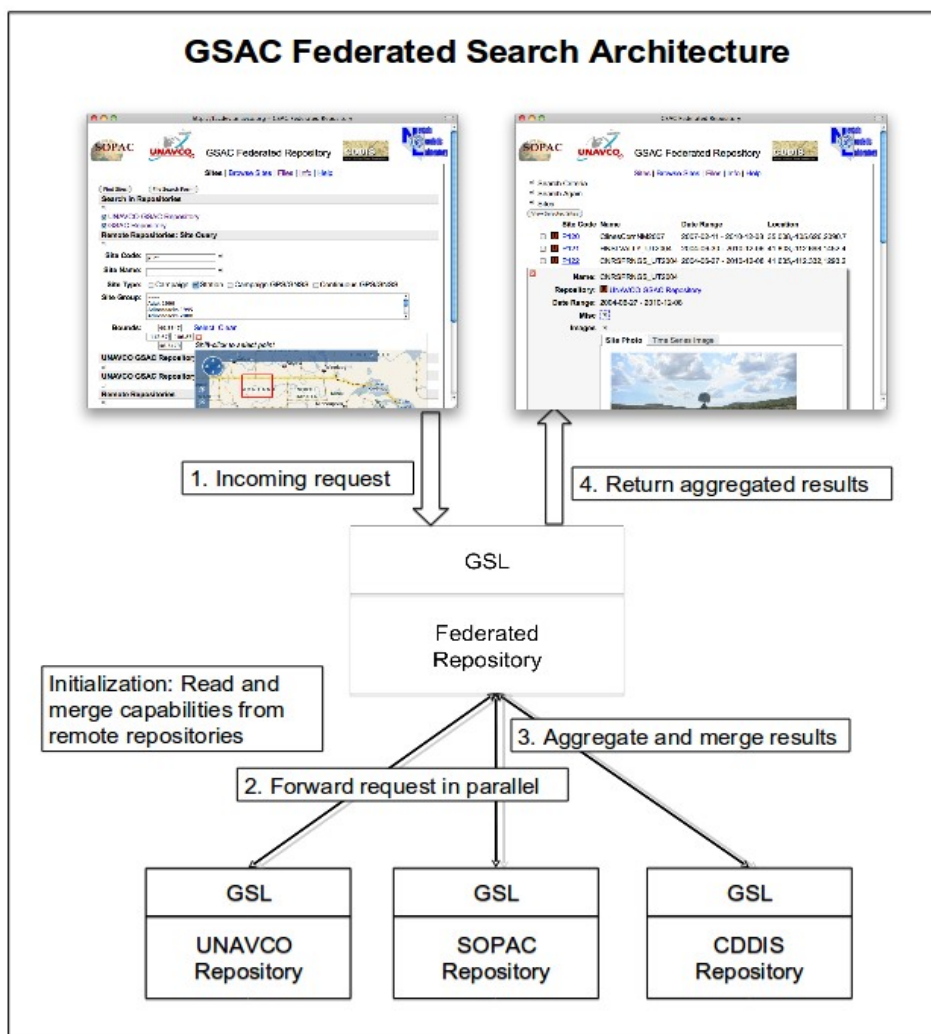

A federated GSAC offers the power of merging many data archives into one search and retrieval tool. A federated

GSAC server operates a single GSAC at any location, which uses the services of other GSAC web service repositories.

A federated GSAC web service does not copy any data or files from remote servers, but rather knows how to query remote servers for information they hold. Federated searches are made possible by a GSAC web service being able to read and use the API capabilities documents (XML information) of GSAC repositories (confirmed as part of the federation). Federated GSAC servers query in parallel two or more associated GSAC repositories.

Shown below is the search page of the UNAVCO Federated Repository showing the ability to simultaneously search for sites on one to three GSAC repositories (SOPAC, UNAVCO, and CDDIS), and the ability to search for one or more of the four data types (GPS, DORIS, SLR, and VLBI) offered by those repositories. Note that not all of the repositories have all the data types.
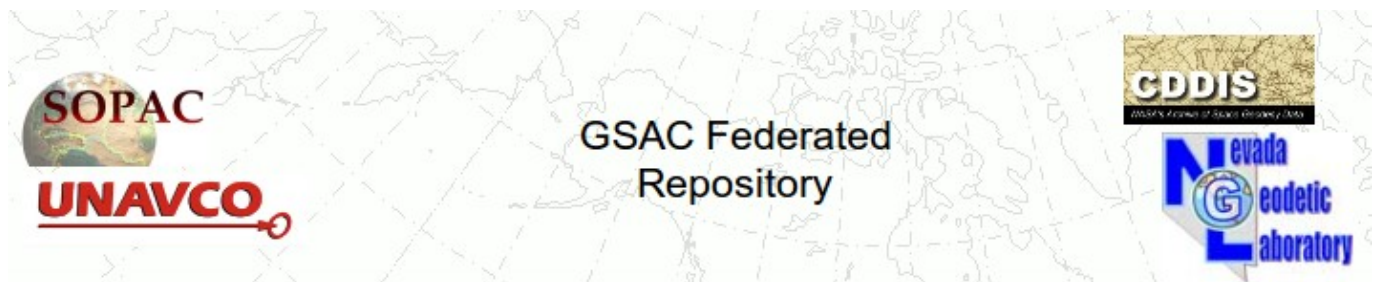
What is needed for a GSAC Repository at Your Agency?

You can install GSAC web services for a data repository when you have:

- a database with complete metadata about your data center's holdings, including metadata about sites (stations or monuments) and their instruments, and about their data files (if any). GSAC can read Oracle, Postgres, or MySQL. You can use an existing database you have, with no changes to it, but new Java code for GSAC code will be needed to read the database. Or you can make a new database using the Prototype GSAC database schema, populating it with your data values, which allows you to use GSAC code with very few changes. See more in the GSAC Database section, below.
- a web server for GSAC web pages. Running GSAC with the Tomcat application server is common, but Tomcat is optional. GSAC can use its included application server, Jetty.
- usually GNSS data files from instruments, for remote users to download. GSAC can provide discovery and download of geodesy data files like RINEX files. This is optional.
- Also, if you want to offer data file downloads, an FTP / HTTP download server the data files. GSAC does not download files. GSAC tells users how and where to download files via a URL to the file. GSAC itself can run on a web server different from the server(s) for downloads, a benefit.
- A Linux system for the initial installation and tests of your GSAC web site, with related software tools bash, Java JDK 1.6 or 1.7, subversion ("svn"), and ant, and with access to your GSAC database.
- the GSAC software package provided by UNAVCO. GSAC is free and open source code in the GSAC package on SourceForge (http://sourceforge.net/projects/gsac).

A typical Linux system for the initial installation and test of GSAC could be Ubuntu, version 11, 12 or 13. Use the bash shell. Ubuntu with bash has very simple installation of the software tools you need. Debian Linux should work well; Mac may also use these tools.

The GSAC Database

An essential part of GSAC installation and operations is a database which you create and maintain, about sites (stations), their instruments, and their data files (if any). The GSAC database contains the information about sites which you wish to offer online, and about sites' data files which you wish to offer for download. The database must contain up-to-date, complete, and correct information if GSAC services are to represent your data repository correctly. (Your database read by GSAC may also be one way to manage contents of your data repository, aside from GSAC operations, if it contains all essential information.. To manage a data center, a well-designed database is much less error prone, and much more suited to computer operations, than text files such as IGS site logs or SINEX files.)

There are two ways to supply a database for GSAC. First, you can use an existing Oracle, Postgre, or MySQL database about your data holdings and sites, if it has the type of fields required by GSAC. See the Prototype GSAC Database schema notes in the UNAVCO GSAC Installation web page for the database tables and fields needed by GSAC to provide its complete suite of web services for searching for geodesy sites information and for site instrument data files, and for data file downloads by GSAC. The Prototype GSAC database schema notes are provided so that persons considering GSAC may see if they have adequate types of metadata in an exisiting database to support GSAC web services. This schema is for MySQL but the parallel to Postgres and Oracle should be clear. You will modify the PrototypeSiteManager.java and PrototypeFileManager.java files during GSAC installation to use the table names and field names in your database. The files PrototypeSiteManager.java and PrototypeFileManager.java are a good guide to writing new Java code using GSAC Java classes. You can download the PrototypeSiteManager.java and PrototypeFileManager.java files from the UNAVCO GSAC Installation web page. The more you are willing to modify PrototypeSiteManager.java and PrototypeFileManager.javas files, the more your database schema can differ from the Prototype GSAC database.

Alternately you can create a new a database for GSAC for your data center, using the Prototype GSAC Database schema. You can download a MySQL Prototype GSAC Database schema .sql file for the prototype GSAC database, a MySQL "dump file," from the UNAVCO GSAC Installation web page. This schema supports geodesy data search and discovery, and all the GSAC formats of search results including web pages, and SINEX and GAMIT station.info files. This prototype schema may be revised in future, but it is quite complete now.

The tables and fields (columns) in the prototype GSAC database schema show the essential parameters needed for GSAC, and their data types. The prototype design corresponds to common geodesy needs, such as encapsulating site receiver-antenna sessions into "equipment sessions," and helps avoid error-prone archive practices, such as storing latitude values or date values in a plain character strings. The prototype design includes features to promote reduced problems in an archive database. For example, only actual dates and numbers can be stored in date and number fields. Databases which store dates or numbers as character strings are error-prone and difficult to manage. For example, GSAC and the prototype database holds all times and dates in ISO 8601 format in the UTC time zone.

Installing GSAC

To implement GSAC web services for your data repository, you download GSAC code from SourceForge, create a database about your data holdings, write some new Java code to specify repository-specific details, build the Java executable, and run the new GSAC server.

Follow the GSAC installation instructions (http://facility.unavco.org/data/gsacws/gsac-install.html) in the GSAC web site which will lead to detailed installation instructions in two README files in the GSAC package itself.

The style, decoration, and layout of your GSAC web pages can look like pages on your web site. See the GSAC installation README Part 2 file for help about GSAC web page HTML and image files.

Installation of GSAC, not including creation of a database for your data center, can take from one day to two weeks depending on the situation. GSAC has been installed in two centers in less than a day each with UNAVCO staff assistance, with the database being ready. Installing GSAC is much faster than creating a new web site and complete modern web services all anew.

UNAVCO can aid and advise about GSAC installations and operations. UNAVCO has a responsibility to ensure that GSAC itself installs properly. UNAVCO can help install GSAC, in some cases including providing a provisional working database with some of your data center's information, if some conditions are met including your use of the provisional database schema. UNAVCO will populate your database with an initial but incomplete set of data, just enough to ensure that GSAC is working.

Completely populating your database, and maintaining it, is necessary for continuing GSAC operations, but doing so is entirely outside of GSAC -- see the diagram on page one -- and is your responsibility. Also, GSAC does not check data file quality or otherwise manage a data archive. GSAC is web services, not a content management system, and it does not "know" (much) about geodesy.

Conforming GSAC Geodesy Repositories and the Prototype GSAC Database

All GSAC geodesy repositories should operate in consistent ways, to provide standard GSAC operations so that scientists have a standard interface at many data centers, and to allow any GSAC to be part of federations of GSAC data repositories, supporting the same kinds of search queries. GSAC has the concept of a *conforming GSAC Geodesy Repository*. What makes a GSAC geodesy implementation conforming is that it supports the standard GSAC geodesy parameters, the fields in the Prototype GSAC Database schema. The database which a conforming GSAC reads for a repository needs to have fields corresponding to the standard parameters.

The prototype GSAC geodesy database schema for GSAC at geodesy centers includes all standard GSAC

parameters. Going beyond its role in GSAC, the Prototype GSAC Database schema is also one suitable schema for maintaining a geodesy data archive, based on UNAVCO operational experience running geodesy archives. See the section "Operating a Basic Geodesy Data Center," below.

If you already have a database you use to support manage and run your repository, we recommend that you install GSAC to use that database. Two databases for one archive is error-prone and very difficult to manage. You will must write some new Java code for your GSAC installation, which should be a roughly comparable effort to creating a new database, and be sure your database has the essential fields used by GSAC. Installing GSAC will quickly tell you if some fields are missing.

Formats of Results from GSAC Site and File Queries

GSAC provides these choices for types of results from queries about sites:

- SINEX
- GAMIT station.info
- SOPAC XML station log
- a JSON format, for remote computer services and processing; more compact than XML.
- an HTML format, used for a GSAC site search result pages on a GSAC repository web site
- a KMZ or KML file for a map of GSAC-selected sites in Google Earth.
- GSAC Plain Text, a format for quick and easy visual inspection of site and instrument metadata values
- GSAC full CSV , a comma separated value format file with complete site metadata values
- GSAC short CSV, with only site name and location values (a legacy format of early GSAC)

Any of these results can be retrieved from a GSAC web site, from the GSAC command line client, or with API calls. See samples of the GSAC results at http://facility.unavco.org/data/gsacws/documents.html. You can add new fields to the results in any of these output formats with some Java coding.

When you do searches for instrument data files, GSAC can show you individual links to all the files found with the GSAC query on one web page, or provide a file of wget commands to run to get all the files found with the GSAC query, or can give you a link to one zip file which contains all the files found with the GSAC query. With GSAC you can easily download many instrument data files with a few commands. See examples in the next section.

Data File Downloads and Site Queries from a GSAC Repository

One of the key services of a GSAC-enabled data repository is providing downloads of instrumental observation data files. A common use is to search a GSAC data repository for files using the web page "Search Files." This will create a table on a web page, with links to download files by clicking on one link for each file.

Or, on the web site, with the same file search page, choosing output ("Results") format of "File CSV" will return a list with URLS to files and the MD5 check sum value of each file, which you can use to confirm uncorrupted file transfers. You can write a script to read a GSAC "File CSV" file and download all the files with one command.

Or, using the API or the GSAC client program and specifying result output format of "file.url" you will get a list of one or more URLs like `ftp://data-out.unavco.org/pub/rinex/obs/2012/158/palm1580.12d.Z.` This is of course a URL which can be used to download a file with FTP from the given server.

Alternately, with the API approach, if you choose GSAC output of "file.wget" the result is a file called "gsacwget.sh," a script file, with one line for each available data file, such as

```
wget ftp://data-out.unavco.org/pub/rinex/obs/2012/158/palm1580.12d.Z
```

which is the wget command to download the file using the LINUX wget tool from the given server.

(GSAC relies on the agency running GSAC to provide the IT infrastructure (FTP or HTML) to serve file downloads, and to provide GSAC with the information about how to download files, typically a URL for each FTP file download. Normally the complete FTP/HTML URL for each data file is available from the database GSAC reads, in a "gnss_data_file" table, with an associated data file type name, MD5 check sum, site name, and times the data was recorded. See for example the GSAC prototype database, table gnss_data_file.)

Here are some examples of site metadata retrieval from a GSAC repository directly from the user's command line. This is a simple way to get metadata, and also quite powerful. On a Linux machine which has the "curl" utility (usually at /usr/bin/curl), copy this next line and enter it as a one-line command on a command line:

```
curl http://facility.unavco.org/gsacws/gsacapi/site
/search/sitesoutput=site.snx\&site.code=p123
```

This should list on your terminal a SINEX file created by the main UNAVCO GSAC repository, for the GNSS site whose code (ID) is P123. The 'arguments' in the command follow the "?" and the "\&".

The command (one line):

```
curl http://facility.unavco.org/gsacws/gsacapi/site/search/sites?
output=site.snx\&site.code=P30*
```

This should list on your terminal a SINEX file for all GNSS sites whose code (ID) begins with "P30" (ten sites in this case).

For several distinct sites by each code(ID) :

```
curl "http://facility.unavco.org/gsacws/gsacapi/site/search?
site.code=P120&site.code=MTWO&site.code=AB04&output=site.snx"
```

See more curl examples on the GSAC use web page, http://facility.unavco.org/data/gsacws/gsac-use.html.

Operating a Basic Geodesy Data Center

You can create and operate a new geodesy data center for public access to your geodesy holdings, and using GSAC for web services. In addition to GSAC, you need a collection of data files from instruments, and an FTP or HTTP server to provide them for download. See for example this typical URL for an FTP geodesy data file server, ftp://data-out.unavco.org/pub/rinex/obs/2007/055/ , at UNAVCO. You also need a database with complete information about the data files, and information about the sites (stations, monuments) and about the instruments where the data was collected. The GSAC prototype database schema, available from the GSAC web site on the GSAC installation page is a suitable database. You need to populate the database yourself. You will need to maintain the data files, the FTP server, and the database as new data arrives and as instruments and sites change. Using a database is a much better approach to archive management than, for example, keeping site information in free-form (ASCII) files, such as SINEX files or IGS site log files which are extremely error-prone.

UNAVCO may, in future, provide a software package to help manage a geodesy archive, but this will be new software, and separate from GSAC. That new code may use GSAC for the public interface to the data repository, but the archive management software will not be GSAC. GSAC is web services.

GSAC-WS Software Notes

GSAC is built on the concept of three basic units, *sites* with *instruments* making *data files*. GSAC is not just for GPS receivers and RINEX files. Two tables in the prototype database let a GSAC installation handle different kinds of sites, instruments, and data files from instruments. The database information about each site has one of the "station_style" values, such as "GPS/GNSS Continuous" or "Strainmeter." Each data file's information in the database has one of the "file_type" values, such as "RINEX GLONASS navigation file," "BSM (borehole strainmeter) Processed," or "Tiltmeter Plot." GSAC servers in operation already handle GNSS, VLBI, SLR and DORIS instruments and data files. No code changes are needed to GSAC to use new instruments or data file types. As a practical matter, to allow consistent use of all GSAC's, GSAC supplies standard unchanging names for types of instruments and files in the prototype database schema. You can see the current prototype database schema in the installation page on this web site. If you add a new type of instrument or data file to your GSAC, please inform UNAVCO so we can include it in new GSACs.

The GSAC Service Layer (GSL) in the diagram on page one is the core GSAC code, common to all GSAC installations, which handles user queries about sites, data files, and resources with the GSAC application, and which provides output handlers to return results of queries. During GSAC installation, you tell the GSL what are your repository's capabilities using new Java code automatically created in a new Java package, and (possibly) writing new code to call your database. GSL code knows how to handle incoming requests, and interprets the queries for your database, and has the code to pass the results of the queries to the output handlers.

GSAC uses an application container (such as Apache Tomcat), the GSAC Service Layer (GSL) interface, and files for the local implementation configuration. A GSAC-enabled archive has a full API to query and download information and data files through URLs. GSAC provides a separate client program to programmatically query any GSAC repository (without use of a web site) and send results directly to your terminal or to files. The client program can be used for automated data retrieval of many files without use of the web interface. The client program should run on any system with recent Java support.

Queries to the GSAC API specify both desired parameters or filters and the format of the returned data using the guidelines called Representation State (REST) web services. The URI parameters in a GSAC API request select GSL resources and can return results in any of several formats.

You can publish your GSAC repository server with Tomcat, or you can run a package release as a stand-alone GSAC Repository server using Jetty. Jetty is included in the GSAC package. You test GSAC during installation and development with Jetty on a desktop using a normal browser, which provides rapid checks of progress in GSAC installation, without changing a public web server until GSAC is ready for public operations.

UNAVCO GSAC Resources, Contacts, and Acknowledgements

GSAC Support

UNAVCO supplies support for GSAC installation and use. Email gsac-ws@unavco.org.

GSAC and COOPEUS

The U.S. NSF (http://www.nsf.gov/) has funded UNAVCO to work with COOPEUS (http://www.coopeus.eu), including provision of GSAC support, from June 2012 through December 2013.

GSAC Information online

UNAVCO GSAC home page: http://facility.unavco.org/data/gsacws

GSAC project at SourceForge: http://sourceforge.net/projects/gsac

Selected GSAC Repositories

UNAVCO GSAC Repository: http://facility.unavco.org/gsacws

CDDIS GSAC Repository: http://cddis.gsfc.nasa.gov/gsacws

SOPAC GSAC Repository: http://geogsac.ucsd.edu:8080/gsacws

Federated GSAC Repository at UNAVCO: http://facility.unavco.org/gsacfederated/

UNAVCO GSAC personnel

Fran Boler, PhD, Geophysics.   GSAC Principle Investigator   boler@unavco.org

Stuart Wier, PhD, Geophysics.   Science Software Engineer   wier@unavco.org

Acknowledgements

Initial development of GSAC-WS was a NASA ROSES ACCESS Program funded project during 2010-2012, Cooperative Agreement NNX10AF07A, developed at UNAVCO, NASA CDDIS, and SOPAC UCSD. The initial GSAC-WS software was developed at UNAVCO (Boulder, Colorado) by Jeff McWhirter, with assistance from staff at CDDIS (NASA's Crustal Dynamics Data Information System) and SOPAC (the Scripps Orbit and Permanent Array Center, at the University of California at San Diego).

GSAC development and user support continued at UNAVCO (June 2012 through December 2013) under NSF support for UNAVCO in COOPEUS.

See also

COOPEUS   http://www.coopeus.eu

Supersites   http://supersites.earthobservations.org/gnss.php

GEO   http://www.earthobservations.org/index.shtml; http://www.earthobservations.org/gsnl.php