GSL: An Open Source Framework for the Rapid Development of Data Archive Access Services

Jeff McWhirter, Fran Boler - UNAVCO Boulder Facility

Yehuda Bock, Melinda Squibb, Louis Ratzeberger - Scripps Orbit and Permanent Array Center, University of California, San Diego

Overview

The Geodetic Seamless Archive Center (GSAC) project is a NASA ROSES ACCESS funded effort with the goal to define and implement a web service API for uniform querying across multiple geodetic data repositories. The project information technology partners include two International GNSS Service (IGS) Data Centers, CDDIS and SOPAC, and a third global NSF and NASA-funded Data Center, UNAVCO. A science partner, the University of Nevada, Reno Geodetic Laboratory is utilizing the services and contributing new quality assessment products.

Each of the participating data centers has its own internal database schema and various web-based access mechanisms to suit the needs of their individual user communities. The software engineering challenge that we faced was to unify these access mechanisms into a common web service API.

To facilitate the development of the GSAC interfaces within these disparate repositories we have developed a Java-based open source middleware framework called the GSAC Service Layer (GSL). It was recognized early in the project that much of the required functionality within each repository could be provided by a common software layer, thus achieving a substantial reduction in the development effort required by each partner institution.

The GSL provides a range of common services for repository configuration, initialization, database management, web interface generation, and data model and metadata creation. Each repository implementation then simply focuses on the core task of responding to query requests and delivering to the GSL layer the internal results. The GSL layer is then responsible for encoding the results in the desired format.

A fundamental challenge in developing a framework such as the GSL is balancing the trade-off between the agnosticism needed within the GSL so that it is applicable to a wide range of problem domains and the specific needs required for a particular repository implementations. The GSL addresses this challenge through an extensible data model and object-oriented framework that allows for the injection of repository specific functionality as needed.

GSL Features

- Repository package code generation facilities
- Stand-alone server or runs under Tomcat
- Rich query capabilities model
- Automatic search form generation
- Browse facilities
- Extensible data model
- Site, files, resources, etc
- Extensible metadata model
- Multiple content views
- HTML, mapping, CSV, JSON, XML, ATOM, RSS
- Integrated file downloader
- Self-documenting web service API
- Command line client
- Federated search across multiple GSAC repositories

Results and Current Status

Currently there are five separate repository implementations that use the GSL:
- UNAVCO, SOPAC, NASA CDDIS, GSAC/RAMODA and GSAC Federated.

The use of the GSL in each of these repository implementations has dramatically reduced the overall development time for each repository and has increased the level of functionality each repository provides. For example, the table below shows a rough estimate of the lines of code in each of the packages. While the absolute numbers don't necessarily reflect the complexity of a particular package it serves to illustrate the effort required for the implementation of a repository.

<table>
<thead>
<tr>
<th>Package</th>
<th>Lines of Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSL</td>
<td>22000</td>
</tr>
<tr>
<td>UNAVCO</td>
<td>2100</td>
</tr>
<tr>
<td>CDDIS</td>
<td>2800</td>
</tr>
<tr>
<td>SOPAC</td>
<td>1200</td>
</tr>
<tr>
<td>Federated</td>
<td>2800</td>
</tr>
</tbody>
</table>

While our focus in the GSAC project has primarily been on geodetic data repositories the GSL can provide the same level of services for other domains as well. We plan to expand its scope to other facilities and continue to grow its open source developer community.

Project website
http://facility.unavco.org/data/gsacws

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

Acknowledgements

The authors gratefully acknowledge support for this work from NASA through the ROSES ACCESS program Cooperative Agreement NNX10AF07A.

GSAC Repository Architecture

GSAC Capabilities – An Alternative to OpenSearch

A type based mechanism to describe what can be searched for in a repository:

```xml
<complexType name="UNAVCO GSAC Repository">
  <complexContent>
    <extension base="Site">
      <sequence>
        <element name="Site Search" type="Site Search"/>
        <element name="Site Code" type="string"/>
        <element name="Station" type="string"/>
        <element name="Station Code" type="string"/>
        <element name="Site Type" type="Site Type"/>
      </sequence>
    </extension>
  </complexContent>
</complexType>
```

Types include:
- String
- Enumeration
- Date range
- Geospatial bounds
- Numeric range
- Boolean

Results – GSL

Web service API documented automatically:

HTML form and browse interfaces generated automatically: