

C. Trabant, T. Ahern, M. Stults,
IRIS Data Management Center, Seattle Washington

Abstract

At the IRIS Data Management Center (DMC) we have been developing web service data access interfaces for our, primarily seismological, repositories for five years. These interfaces have become the primary access mechanisms for all data extraction from the DMC.

For the last two years the DMC has been a principal participant in the GeoWS project, which aims to develop common web service interfaces for data access across hydrology, geodesy, seismology, marine geophysics, atmospheric and other geoscience disciplines. By extending our approach we have converged, along with other project members, on a web service interface and presentation design appropriate for geoscience and other data.

The key principles of the approach include using a simple subset of RESTful concepts, common calling conventions whenever possible, a common tabular text data set convention, human-readable documentation and tools to help scientific end users learn how to use the interfaces. The common tabular text format, called GeoCSV, has been incorporated into the DMC's seismic station and event (earthquake) services.

In addition to modifying our existing services, we have developed prototype GeoCSV web services for data sets managed by external (unfunded) collaborators. These prototype services include interfaces for data sets at NGDC/NCEI (water level tides and meteorological satellite measurements), INTERMAGNET repository and UTEP gravity and magnetic measurements. In progress are interfaces for WOVOdat (volcano observatory measurements), NEON (ecological observatory measurements) and more.

An important goal of our work is to build interfaces usable by non-technologist end users. We find direct usability by researchers to be a major factor in cross-discipline data use, which itself is a key to solving complex research questions. In addition to data discovery and collection by end users, these interfaces provide a foundation upon which federated data access and brokering systems are already being built.

Background

At the IRIS Data Management Center (DMC), seismic data delivery volume has improved noticeably in the past few years in part do to using a framework composed of simple web services.

Based on lessons learned and concepts developed to deliver this data, IRIS and its partners proposed a project to NSF to generalize and extend these concepts to more types of services and geoscience data.

As a result, in 2014, IRIS and its partners started the GeoWS project to determine common practices and develop guidelines for establishing consistent implementations of these practices.

Project Members

IRIS, Caltech Seismology, GeodeSystems, LDEO, SDSC - CUAHSI, UNAVCO, Unidata

Key Objectives and Our Approach

Ease the tasks of data discovery and access

- Coordinate with Project Members to
 - develop a consistent look and feel for respective services
 - identify and document commonly used terms - like minlatitude, maxlatitude, etc.
 - identify tabular text data as the first data exchange format and create a set of guidelines called GeoCSV

Ease the tasks of data usability

- use RESTful style to design service API
- design query URLs where all the information is specified in a single URL
- support use of URL builders to document and interactively access respective service

Coordinate with other NSF EarthCube projects

- Develop Accessors for plug-in connection to B-Cube Broker
- Enter service URLs into CINERGI registry

Provide both human and machine readable documentation

- Test use of Swagger Specification 2.0, JSON file format for service description
- Evaluate use of Swagger-ui to dynamically document services

Unfunded Collaborators

In addition to project members, a variety of other institutions agreed to work with us to make respective data collections available. These institutions are not funded by this project, so integration activities have been performed by personnel at the IRIS Data Management Center. The work performed involves retrieving data from existing web applications, reorganizing the data, and presenting the results in a GeoWS service. The data is present in GeoCSV format or in already established text formats.

Unfunded Collaborators	Status
Global Geodynamics Project (GGP) superconducting gravimeters	Data being consolidated in new repository in Germany
Gravity and Magnetics - data collected by Randy Keller - UTEP	data in GeoCVSVformat
INTERMAGNET - consortium of magnetic observatories	summary data in GeoCSV format and data in IAGA2002 format
Structural Geo. & Tect. - U. of Kansas	in discussions
National Ecological Observing Network (NEON)	In discussions
NCEI - National Centers for Environmental Information - Waterlevel Tides & meteorological satellite image (DMSP) data	selected overview data in GeoCSV format
OOI - Ocean Observatory Initiative - Seismic data	data available via IRIS FDSN services
WOVOdat - A database of volcanic unrest data from World Organization of Volcano Observatories	In discussions

Links

GeoWS project: <http://earthcube.org/group/geows-geoscience-web-services>
GeoWS technical home: <http://geows.ds.iris.edu>

GeoCSV

The GeoCSV is a set of guidelines and annotations that describe a format for exchange of tabular text data. The tabular data is formatted as comma separated values (CSV) as defined by W3C - CSVW recommendations for best practices. Additionally, to facilitate exchange of scientific data, optional, straight-forward extensions are defined to further further describe the CSV data.

Web Service Shell

Web Service Shell is a key technology developed over the past few years that enables access to backend data applications and delivers the returned data to HTTP clients.

By configuration, Web Service Shell -

- links URLs to respective backend command line data applications
- manages HTTP media types as needed for respective applications
- provides parameters validation and offloads HTTP responsibilities from data applications

By programming, WebServiceShell -

- can be extended using standard Java web service technology
- can be used to meet more complex data delivery needs

