The Impact of Web Services at the IRIS DMC

Tim Ahern, Robert Weekly, Chad Trabant, Mick Van Fossen, Yazan Suleiman, Bruce Weertman, Mike Stults
IRIS Data Management Center, Seattle Washington

Abstract

The IRIS Data Management Center (DMC) has served the seismological community for nearly 25 years. In that time we have offered data and information from our archive using a variety of mechanisms ranging from email-based to desktop applications to web applications and web services. Of these, web services have quickly become the primary method for data extraction at the DMC. In 2011, the first full year of operation, web services accounted for 25% of the data shipped from the DMC. In 2015, over ~400 TB of data was delivered directly to users through web services, representing about 50% of all shipments from the DMC that year. In addition to handling requests directly from users, the DMC switched all data extraction methods to use web services in 2014. On average the DMC now handles between 10 and 20 million requests per day submitted to web service interfaces.

The rapid adoption of web services is attributed to the many advantages they bring. For users, they provide on-demand data using an interface technology, HTTP, that is widely supported in nearly every computing environment and language. These characteristics, combined with human-readable documentation and existing tools make integration of data access into existing workflows relatively easy. For the DMC, the web services provide an abstraction layer to internal repositories allowing for concentrated optimization of extraction workflow and easier evolution of those repositories. Lending further support to DMC's push in this direction, the core web services for station metadata, time series data and event parameters were adopted as standards by the International Federation of Digital Seismograph Networks (FDSN).

We expect to continue enhancing existing services and building new capabilities for this platform. For example, the DMC has created a federation system and tools allowing researchers to discover and collect seismic data from data centers running the FDSN-standardized services. A capability in development, called Research-Ready Datasets, will leverage the DMC's quality assurance system to select data based on quality measurements. Within five years, the DMC's web services have proven to be a robust and flexible platform that enables continued growth for the DMC. We expect continued enhancements and adoption of web services.

Key Advantages of Web Services

The DMC strives for efficiency in implementing a set of RESTful web services that provide broad access to the data center's holdings. These services perform a wide range of data extraction and processing techniques that include retrieving seismological time series and related metadata, formatting earthquake hypocentral parameters, evaluating instrument responses, and (soon) calculating synthetic seismograms using an arbitrary Earth model.

Our web services focus on 5 key concepts for simplicity and efficiency:

-Implement methods native to HTTP. This allows our web services to be adaptable to more complex network topologies.

-Provide easily readable, standardized documentation for end users. Everyone appreciates clear directions.

-Develop tools and clients. Clients that retrieve and process data are sourced from the DMC, research community and widely available HTTP clients.

-Provide data in useful formats. Formatting data easily consumed by human or machine means more time for processing and analysis.

-Standardization across global data centers.

Core Web Services from the DMC

The following web services represent the most persistently used services offered by the DMC, and the returned data format.

fdsnws-dataselect	Raw time series data in standardized miniSEED format
fdsnws-station	Station metadata in standardized StationXML and simple text format
fdsnws-event	Event information in standard QuakeML and simple text format
irisws-timeseries	Processed or converted time series data in miniSEED, SAC, text, or graphical format
irisws-resp	Instrument sensor response in common SEED RESP format
irisws-sacpz	Instrument sensor response in common SAC Poles and Zeros format

Web Services and Data Extraction

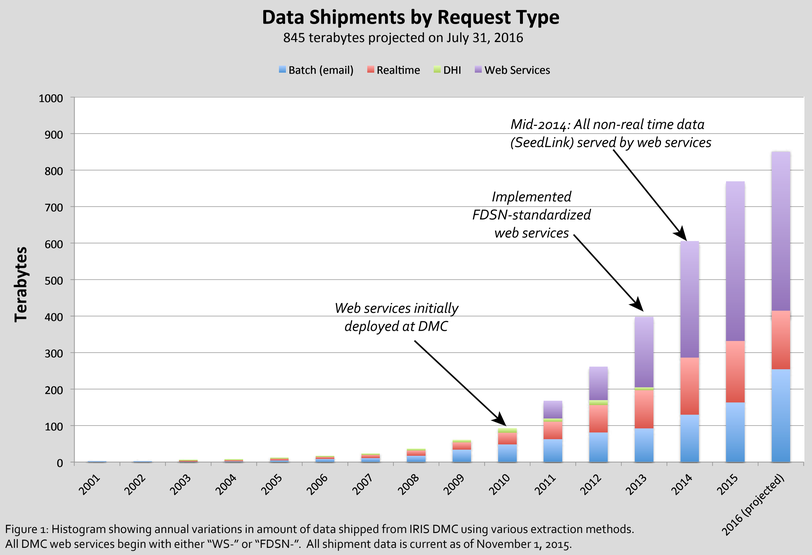

Figure 1: Histogram showing annual variations in amount of data shipped from IRIS DMC using various extraction methods. All DMC web services begin with either "WS-" or "FDSN-". All shipment data is current as of November 1, 2015.

The adaptability of web services have greatly impacted the DMC's ability to serve data to a broad user base across a variety of platforms. In fact, since the summer of 2014, the role of web services have been expanded to include all data extraction using the popular BREQ_FAST format in addition to requests coming from software clients or web browsers.

Since becoming the primary method for all non-real time data requests (blue and red bars, Figure 1), web services account for shipping nearly 1,175 terabytes of data (~75% of all shipped data) from the DMC. Much of the shipped data is sent using *fdsnws-dataselect*, the DMC's most popular web service.

Standardization of Web Services

The International Federation of Digital Seismographs (FDSN) recently approved specifications for 3 web services that define access protocols for:

- *fdsnws-dataselect* -- raw time series data
- *fdsnws-station* -- related station metadata
- *fdsnws-event* -- earthquake (event) parameters

These services have been implemented at the DMC since 2013 (Figure 1) and have been implemented by 14 other data centers located throughout North America, Europe, and South America. A complete list of data centers that currently host FDSN-compatible web services can be found at http://www.fdsn.org/webservices/datacenters/

The FDSN web services provide a foundation for systems that benefit from aggregating data from multiple data centers. Examples of these types of systems include data request brokers that distribute requests to multiple data centers, or data cataloguers that organize information for multidisciplinary discovery.

Re-use of data access software (clients) is another important benefit to standardization.

A Variety of Data Request Methods

In addition to common HTTP clients such as **wget**, **curl**, and web browsers, a number of programs and toolkits have been developed to allow access to data from a variety of platforms

command line, **Fetch scripts**	**http://service.iris.edu/clients**
Java, **IRIS-WS library**	**http://service.iris.edu/clients**
MATLAB, **irisFetch**	**http://service.iris.edu/clients**
web, **Wilber** (event oriented)	**http://ds.iris.edu/wilber3/**
Python, **ObsPy**	**http://docs.obspy.org/**
email, **BREQ_FAST**	**http://ds.iris.edu/ds/nodes/dmc/manuals/breq_fast/**

The IRIS Federator

The standardization and adoption of FDSN web services provides a common interface for seismological data access. Leveraging these common services, the DMC has built the IRIS Federator to help users discover and access data across data centers. The IRIS Federator allows simple, client-side federation and involves two steps: first the client sends the request to the *irisws-fedcatalog* service, and then the client parses the catalog response and sends the request(s) to the identified data centers.

Data are requested by and returned directly to the client application from each federated data center. Data are not routed through a central location.

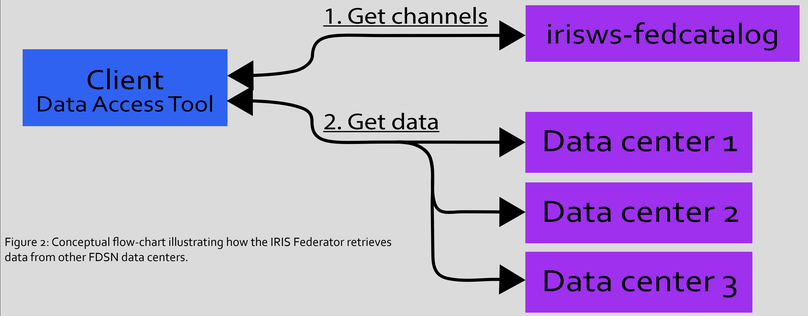

Figure 2: Conceptual flow-chart illustrating how the IRIS Federator retrieves data from other FDSN data centers.

The main interface for the IRIS Federator is the *irisws-fedcatalog* web service: **http://service.iris.edu/irisws/fedcatalog/1/.**

Support for using the Federator to collect data from multiple centers is included in the DMC's command-line FetchData and MATLAB irisFetch.m clients. An example:

```
$ FetchData -F -L -- -C LHZ --radius 3.5:96:70 -s
2004-12-26T00:30:00 -e 2004-12-26T02:00:00 -o fetch.mseed
-m fetch.meta
Fetching data from GEOFON (http://geofon.gfz-potsdam.de)
Processed metadata for 13 channel epochs in 0.4 seconds
Received 200.0 KB of time series data in 1.7 seconds (119.4 KB/s) - written to GEOFON-fetch.mseed
Fetching data from IPGP (http://centrededonnees.ipgp.fr)
Processed metadata for 4 channel epochs in 0.3 seconds
Received 80.0 KB of time series data in 1.4 seconds (58.6 KB/s) - written to IPGP-fetch.mseed
Fetching data from IRISDMC (http://ds.iris.edu)
Processed metadata for 22 channel epochs in 0.0 seconds
Received 424.0 KB of time series data in 6.2 seconds (68.7 KB/s) - written to IRISDMC-fetch.mseed
```

This work was supported through NSF awards: EAR-1261681, ICER-1343709 & ICER-1321600