Award # G10AC00533
Title: Development of Innovative Data Access Techniques and Processing Methods to Support Seismological Research

Author and Affiliations(s) with address and zip code
Dr. Timothy Keith Ahern
Director of Data Services
IRIS
1200 New York Avenue
Suite 400
Washington DC 20005
Author's telephone numbers, fax numbers and email address
206 547-0393 phone, 206 547-1093 Fax, tim@iris.washington.edu

Term covered by the award and the report (start and end dates)
August 1, 2010 thru July 31, 2011

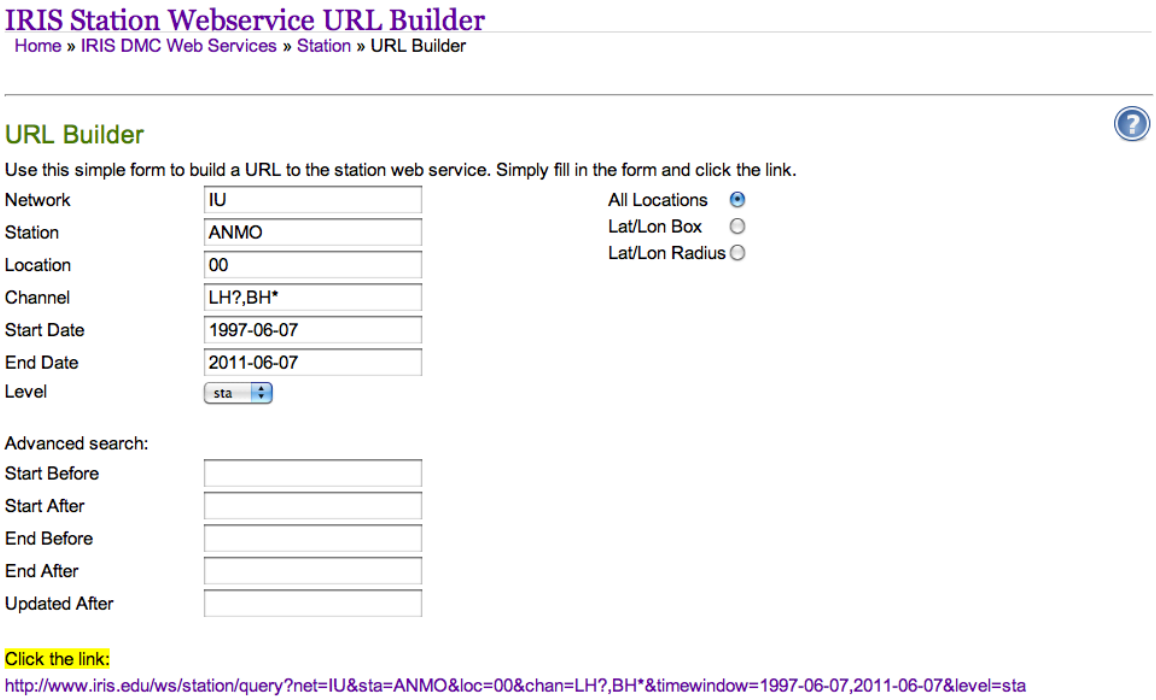
Abstract
All work originally proposed to the USGS was completed during the term of this award. This included the development of two web services to provide access to seismic station metadata (ws_station) and event catalogs (ws_event). IRIS met with USGS personnel at the NEIC to transfer the technology and summarize the developments with the USGS. We demonstrated that a scientific workflow engine could connect to all IRIS web services allowing them to be inserted as steps in a workflow. IRIS demonstrated that a sequence of operations that was implemented as a hardwired workflow could be replicated by executing individual web services in the workflow. In addition to these software development activities IRIS also acquired 5 Dell Servers that have been incorporated into the virtualized environment operating at the IRIS DMC. The web services are now in operation at the IRIS DMC and are being used by the research community.

Report
The development of web services at the IRIS DMC has been a coordinated effort funded by the USGS as part of this proposal and from the Instrumentation and Facilities program of the GEO Directorate of the National Science Foundation. By the end of this award period IRIS developed a total of 14 web services. All of them are listed at the URL http://www.iris.edu/ws. Two of these 14 services were developed using salary support that was received by the USGS through this proposal. The two web services were ws_station and ws_event.

Ws_station Ws_station is a web service that delivers results as an XML document that adheres to the station XML schema jointly developed by Caltech, Berkeley, the USGS/NEIC and IRIS. While station XML is the payload, staff at IRIS largely developed the various query parameters and service implementation. Documentation for ws_station can be found online at http://www.iris.edu/ws/station and it is replicated in Appendix A in this report.

The station web service can be invoked in many ways including through a URL in a browser, using GNU wget, using scripts written in PERL or other scripting languages, or through programming languages such as Java.

To assist users in understanding how to form the URL, IRIS has developed URL builders as GUIs that run in a browser. To access the URL builder for the station web service go to URL http://www.iris.edu/ws/builder/station/. The builder GUI is shown below:

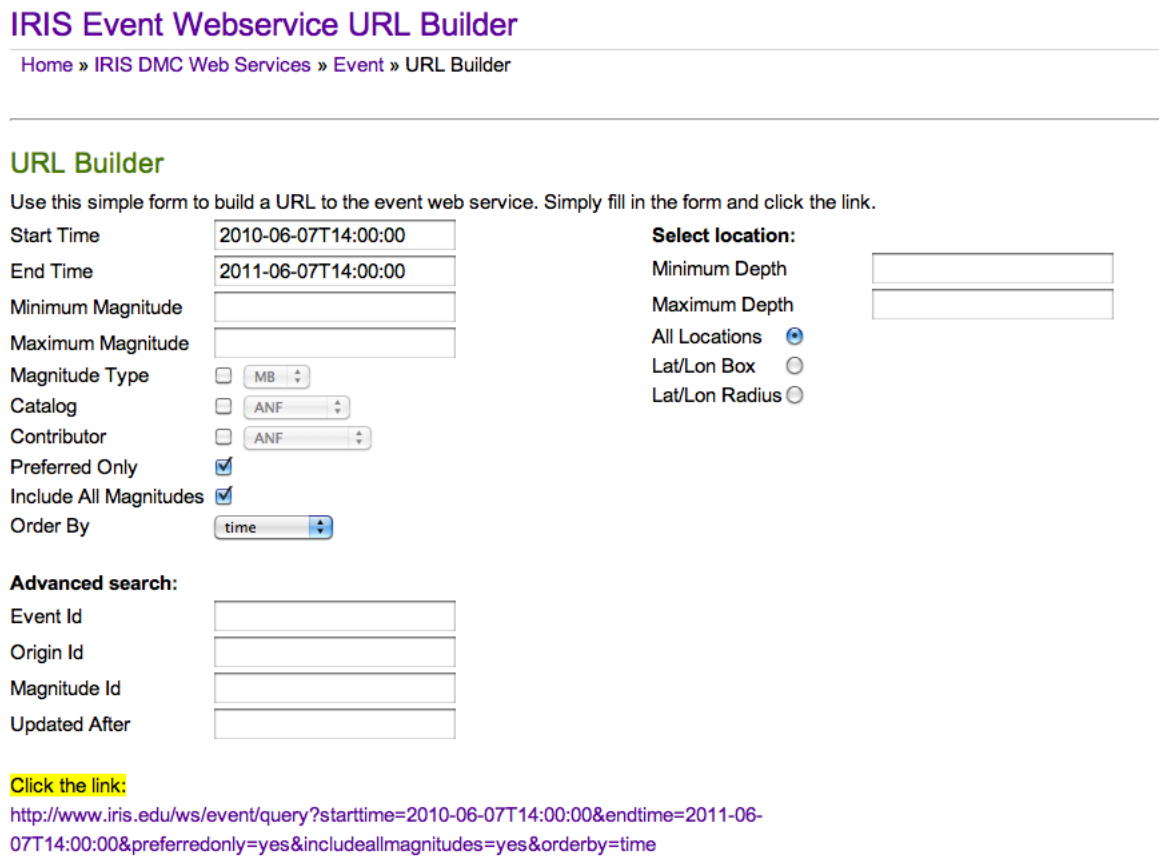


As one enters values in the various fields, the URL near the bottom of the figure changes accordingly. This results in a user easily understanding how to form a URL for the query they wish to submit to the IRIS DMC. The builder can also be used to receive information simply when the number of results of interest is small but it is not intended to be used when the number of results to be returned is large.

Ws_event The second web service developed as part of this proposal is the event service. In this case the payload, QuakeML, has been defined by a group led by ETH in Zurich and has become a standard for exchange of event related information. The USGS (through the NEIC) and IRIS were involved in the definition of QuakeML. Documentation for ws_event can be found online at http://www.iris.edu/ws/event and it is replicated in Appendix B in this report.

The event web service can be invoked in many ways including through a URL in a browser, using wget from GNU, from scripts written in PERL or other scripting languages, or through programming languages such as Java.

To assist users in understanding how to form the URL, IRIS has developed URL builders as GUIs that run in a browser. To access the URL builder for the event web service go to URL http://www.iris.edu/ws/builder/event/. The builder GUI is shown below:



As with the station URL builder, when one enters values in the various fields, the URL near the bottom of the figure changes accordingly. This results in a user easily understanding how to form a URL for the query they wish to submit to the IRIS DMC. Again the builder is a method to understand the structure of the required URL. It is not intended to retrieve large amounts of event information.

Other Web Services not using USGS funds in their development

IRIS, using NSF funding, has developed twelve other web services. These are all documented at http://www.iris.edu/ws/. These web services include:

1. Ws_resp, a web service that returns instrument response information in the RESP format
2. Ws_sacpz, a web service that returns instrument response information in the format used by SAC, SAC_PZ
3. Ws_bulkdataselect, a web service that returns time series for one or more seismograms that have been specified in a list
4. Ws_dataselect, a web service that returns a single seismogram

5. Ws_timeseries, a web service that can recover a seismogram, process the seismogram through a variety of digital signal processing algorithms, and convert the data into one of a variety of output formats
6. Ws_availability, a web service that returns what data exist within a given space-time domain
7. Ws_evalresp, a web service that returns the response of a seismogram in amplitude and phase values by frequency. The metadata is extracted from the IRIS DMC's database of response information
8. Ws_distaz, a web service that returns the distance and azimuth between two points on the earth's surface.
9. Ws_ms2ascii, a web service that will convert a miniSEED file into one of two simple ASCII formats
10. Ws_plotter, a web service that will produce simple graphics representations of a miniSEED file.
11. Ws_tracedsp, a web service that will provide several digital signal-processing algorithms to a time series.
12. Ws_flinnengdahl, a web service that will convert a latitude, longitude pair into either a Flinn-Engdahl region code or name.

All 14 of the web services work seamlessly together and have been shown to be callable by an external workflow engine; in our case we used the Trident Scientific workflow engine developed by Microsoft Research.

Example Client Applications. Using NSF funding, IRIS developed a series of client applications written in PERL. These can be found at http://www.iris.edu/ws/wsclients/. One of these clients is FetchEvent and it can be used to easily return events meeting specific criteria. The documentation for FetchEvent is provided within the PERL script. Fetch_resp is a PERL script that accesses the ws_station service and returns response information in the RESP format. The ws_sacpz also accesses ws_station and returns the response information in the SACPZ format.

The clients are well documented and clearly written. They can be used as starting points for users that desire to modify the scripts for their own purposes.

Hardcoded work flows.

During the period of this award IRIS produced one hardcoded workflow. This workflow is ws_timeseries. It can be accessed at http://www.iris.edu/ws/timeseries/. In essence ws_timeseries is a combination of 3+ underlying workflows including ws_dataselect ws_tracedsp, and a series of web services providing output format options such as SAC format, SAC ASCII format, audio files, 2 forms of simple ASCII output, a miniSEED file, or a png file for a picture of a seismogram. Ws_dataselect is used to retrieve a waveform segment from the IRIS DMC, ws_tracedsp is used to filter, gain correct, deconvolve the instrument response, convert units of ground motion, and a variety of related functions. The functions available through ws_timeseries have also been duplicated using the Trident workflow engine by combining ws_dataselect, ws_tracedsp, and output services that for instance convert a miniSEED file to a graphics display of the waveform.

Complete documentation for ws_timeseries can be found at http://www.iris.edu/ws/timeseries/ and will not be reproduced here.

Servers: USGS funds were used to acquire 5 Dell servers that are being used to augment other Dell servers to provide the computational engine upon which these web services will be implemented. The servers are running RedHat Enterprise Virtualization (RHEV) and RedHat Enterprise Linux (RHEL). At the present time the web services are running on a number of SUN servers. Delays in getting the RHEL and RHEV system software functioning delayed the deployment of the web services to the virtual server environment but this should take place before the end of 2011. These details are transparent to the users of the IRIS web services.

International Coordination

The results of our work in developing web services have not been formally published but they have been disseminated to the research and monitoring communities in a variety of venues. These include

- e-Science for Seismology meeting in Edinburgh, Scotland in March 2011
- Meetings at ORFEUS in deBilt, Netherlands in March 2011
- Bilateral meetings in China in April 2011
- CTBTO meetings in Vienna in June 2011
- IUGG/FDSN meetings in Australia in June-July 2011
- European Union and NSF coordination meetings in Brussels in September 2011
- FDSN

Appendix A – Definition of Query parameters for ws_station.

Station Web Service

Home » IRIS DMC Web Services » Station

The ws-station web service returns station metadata in StationXML format (schema). Results are available at multiple levels of granularity: network, station, channel and response.

Metadata may be selected based on network, station, location or channel names in addition to time windows, geographic regions or metadata update times.

- For more overview information visit http://www.iris.edu/ws/doc/station_help.htm
- For a list of changes for each release visit http://www.iris.edu/ws/doc/station_revisions.htm

Below is a full list of service parameters and their usage.

Query: /ws/station/query?[query-parameters]	
where query-parameters:	
Empty/No parameters	will return all stations - level defaults to sta
net="network"	Specify SEED network code, example net=IU
sta="station"	Specify SEED station code, example sta=ANMO
loc="location"	Specify SEED location code mask, example: loc=00. substitute dash characters for white space, example: loc=--
chan="channel"	Specify SEED channel code mask, example: chan=BH1
Wildcard wildcard search is available for net|sta|chan|loc paremeters. A wildcard can appear anywhere in the search criteria. Example: ▪ net=IU&sta=ANM* ▪ net=IU&sta=AN?O ▪ net=IU&sta=A*O	
List users can specify multiple networks, stations, channels or locations in their search by providing comma separated list. Example: ▪ sta=ANM*,ANTO ▪ chan=BH1,BHZ	
minlat="degrees"	Specify minimum latitude, northern boundary [-90 <= X <= 90]
maxlat="degrees"	Specify maximum latitude, southern boundary [-90 <= X <= 90]
minlon="degrees"	Specify minimum longitude, western boundary [-180 <= X <= 180]
maxlon="degrees"	Specify maximum longitude, eastern boundary [-180 <= X <= 180]
[lat,lon,maxradius [minradius]]="degrees"	Optional, geographic circular bounding area. All lat, lon, maxradius parameters must be given together. minradius parameter is optional and defaults to zero. Cannot be combined with minlat, maxlat, minlon, maxlon. Example: ▪ lat=40.0&lon=-100.0&minradius=0.0&maxradius=2
timewindow=YYYY-MM-DD,YYYY-MM-DD	Specify a time range that stations/channels must be operating within
updatedafter=YYYY-MM-DD	Query stations updated after a specific date
level=net|sta|chan|resp	Specify whether to include channel/response metadata or not
startbefore=YYYY-MM-DD	Specify stations starting before this date
startafter=YYYY-MM-DD	Specify stations starting after this date
endbefore=YYYY-MM-DD	Specify stations ending before this date
endafter=YYYY-MM-DD	Specify stations ending after this date

Appendix B – Definition of Query parameters for ws_event.

IRIS Event Webservice

URL Builder

Home » IRIS DMC Web Services » Event

Description

The ws-event web service returns event information in the QuakeML format (schema 1.2 RC3). Events may be selected based on location, time, catalog, contributor and internal identifiers.

Metadata may be selected based on network, station, location or channel names in addition to time windows, geographic regions or metadata update times.

- To launch a simple form to help you construct the query URL, visit the URL builder
- For more overview information visit http://www.iris.edu/ws/doc/event_help.htm
- For a list of changes for each release visit http://www.iris.edu/ws/doc/event_revisions.htm

Below is a full list of service parameters and their usage.

Query Usage:

```
/query?[geographic-constraints] [depth-constraints] [temporal-constraints] [magnitude-constraints]
   [organization-constraints][misc-parameters]
or
/query?[id-options] [includeallmagnitudes] [preferredonly]
```

where

```
geographic-constraints ::[boundaries-rect] OR [boundaries-radial]
boundaries-rect        ::[minlat=<latitude>] [maxlat=<latitude>] [minlon=<longitude>] [maxlon=
                         <longitude>]
boundaries-radial      ::[lat=<latitude>&lon=<longitude>&maxradius=<degrees>] [minradius=<degrees>]
depth-constraints      ::[mindepth=<km>] [maxdepth=<km>]
temporal-constraints   ::[starttime=<time>] [endtime=<time>]
magnitude-constraints  ::[minmag=<magnitude>] [maxmag=<magnitude>] [magtype=<magnitude-type>]
organization-          ::[catalog=<catalog>] [contributor=<contributor>]
constraints
misc-parameters        ::[limit=<number-of-events>] [orderby=<time|magnitude>] [updatedafter=<time>]
                         [includeallmagnitude=<yes|no>] [preferredonly=<yes|no>]
id-options             ::[eventid=<id_value>] OR [originid=<id_value>] OR [magnitudeid=<id_value>]
```

Sample Queries:

- http://www.iris.edu/ws/event/query?mindepth=34.9&maxdepth=35.1&catalog=NEIC PDE&contributor=NEIC PDE-Q&magtype=MB&lat=-56.1&lon=-26.7&maxradius=.1
- http://www.iris.edu/ws/event/query?eventid=3316989
- http://www.iris.edu/ws/event/query?minmag=8.5
- http://www.iris.edu/ws/event/query?starttime=2011-01-07T14:00:00&endtime=2011-02-07&minlat=15&maxlat=40&minlon=-170&maxlon=170&preferredonly=yes&includeallmagnitudes=yes&orderby=magnitude

Detailed Descriptions of each Query Parameter

parameter	examples	discussion
Geographic constraints - bounding rectangle The following four parameters work together to specify a boundary rectangle. All four parameters are optional, but they may not be mixed with the parameters used for searching within a defined radius. - ex. .../ws/event/query?minlat=35&maxlat=40&minlon=-150&maxlon=-148		
minlat	15.5	Specify the southern boundary. The minimum latitude must be between -90 and 90 degrees inclusive (and less than or equal to maxlat). If not specified, then this value defaults to -90
maxlat	25.0	Specify the northern boundary. The maximum latitude must be between -90 and 90 degrees inclusive and greater than or equal to minlat. If not specified, then this value defaults to 90
minlon	170.0	Specify the western boundary. The minimum longitude must be between -180 and 180 degrees inclusive. If not specified, then this value defaults to -180. If minlon > maxlon, then the boundary will cross the -180/180 meridian
maxlon	-170.0	Specify the eastern boundary. The minimum longitude must be between -180 and 180 degrees inclusive. If not specified, then this value defaults to +180. If maxlon < minlon, then the boundary will cross the -180/180 meridian

Geographic constraints - bounding radius The following four parameters work together to specify a boundary using a radius around a coordinate. `lat`, `lon`, and `maxradius` are all required, and must be used together. These parameters are incompatible with the boundary-box parameters described above. ▪ ex. .../ws/event/query?lat=40.0&lon=-100.0&minradius=0.0&maxradius=5.0		
lat	`35`	Specify the central latitude point, in degrees. This value must be between -90 and 90 degrees. This MUST be used in conjunction with the lon and maxradius parameters.
lon	`170`	Specify the central longitude point, in degrees. This MUST be used in conjunction with the lat and maxradius parameters.
maxradius	`60`	Specify the maximum radius, in degrees. Only earthquakes within maxradius degrees of the lat/lon point will be retrieved. This MUST be used in conjunction with the lat and lon parameters.
minradius	`25`	This optional parameter allows for the exclusion of events that are closer than minradius degrees from the specified lat/lon point. This MUST be used in conjunction with the lat, lon, and maxradius parameters and is subject to the same restrictions. If this parameter isn't specified, then it defaults to 0.0 degrees.
Depth constraints		
mindepth	`-1`	Specify minimum depth (kilometers), values increase positively with depth
maxdepth	`20`	Specify maximum depth (kilometers), values increase positively with depth

Temporal constraints The following two parameters impose time constrants on the query. For additional information about allowed time formats, see: Time Formats		
starttime	`2011-05-21T03:12:00.0`	Limit results to the events occurring after the specified start time.
endtime	`2001-05-21`	Limit results to the events occurring before the specified end time.
Magnitude constraints		
minmag	`-1.0`	Specify a minimum magnitude.
maxmag	`8.3`	Specify a maximum magnitude.
magtype	`Ml`	Specify magnitude type. Some common types (there are many) include: ▪ `Ml` - local (Richter) magnitude ▪ `Ms` - surface magnitude ▪ `mb` - body wave magnitude ▪ `Mw` - moment magnitude

Subset the data by organization Both of these parameters may contain spaces. While actually entering spaces into the query is acceptable, it is preferrable to represent them with `%20` as shown in the example		
catalog	`NEIC PDE`	Specify a catalog [available catalogs]. Results will include any origins which contain the specified catalog text, i.e. "PDE" will match "NEIC PDE"
contributor	`NEIC PDE-Q`	Specify a contributor [available contributors]. When selecting a contributor, the result includes a preferred origin as specified by the contributor. Results will include any origins which contain the specified contributor text, i.e. "NEIC" will match "NEIC PDE-Q".

Miscelleneous parameters These parameters affect how the search is conducted, and how the events are returned		
limit	`25`	Limit the results to the specified number of events. This value must be 10 or greater. By default, the results are not limited.
orderby	`time`	Sort the resulting events in order of descending `time` or `magnitude`. By default, results are sorted in descending time order.
updatedafter	`2000-05-23`	Select origins updated after a certain date, of the format `YYYY-MM-DD`. This is most useful for synchronization purposes.
The default download is restricted to QuakeML event, preferred origin and magnitude elements. optional flags are available to include additional information:		
includeallmagnitudes	`yes`	Can be either yes or no. Defaults to yes. This will include all the magnitudes in search and print criteria. If magnitudes do not exist for a certain origin, the search algorithm will consider it a miss and therefore will not include the event.
preferredonly	`yes`	Can be either yes or no. Defaults to yes. Include preferred estimates only. When catalog is selected, the result returned will include the preferred origin as specified by the catalog.

Specifying an event using an IRIS ID number Individual events can be retrieved using ID numbers assigned by IRIS. When these parameters are used, then only the `includeallmagnitudes` and `preferredonly` parameters are also allowed.		
eventid	`1234`	Retrieve an event based on the unique IRIS event id.
originid	`2345`	Retrieve an event based on the unique IRIS origin id.
magnitudeid	`13720660`	Retrieve an event based on the unique IRIS magnitude id.

Time Formats:

Year, Month, Day in Month - Time:

`YYYY-MM-DDThh:mm:SS[.sss]` ex. `1997-01-31T12:04:32, 1997-01-31T12:04:32.123`

Year, Month, Day in Month (where time is assumed to be 00:00:00)

`YYYY-MM-DD` ex. `1997-01-31`

Where:

```
YYYY = four-digit year
MM = two-digit month (01=January, etc.)
DD = two-digit day of month (01 through 31)
T = date-time separator
hh = two digits of hour (00 through 23) (am/pm NOT allowed)
mm = two digits of minute (00 through 59)
SS = two digits of second (00 through 59)
sss = one to three digits representing a decimal fraction of a second
```