NERA

Network of European Research Infrastructures for Earthquake Risk Assessment and Mitigation

Report

Webservice protocols used to transfer data between networks (webservice protocols)

Activity:	*Networking school seismology activities*
Activity number:	*NA8: task 8.2*
Deliverable:	*Webservice protocols used to transfer data between networks*
Deliverable number:	*D8.2*
Responsible activity leader:	*Paul Denton*
Responsible participant:	*NERC/BGS*
Author:	*Paul Denton/ Iestyn Evans (NERC/BGS)*

Please use the first page of this template for the cover of your deliverable report.

Summary

School seismology projects in the UK, France and Switzerland currently archive data recorded by their schools in different ways.   One of the aims of this project is to develop robust systems to enable school seismology networks to share their results online.   This will be done using webservices.

Introduction

At the start of this project each national network collected data from their own schools and presented the data back to schools through a website interface.   However in order for a school in one country to vie w data from a different country they would have to visit a separate website (maybe in a different language) and navigate a different web system.   It would be simpler is a mechanism was in place to allow national networks to share their data behind eth scenes and present a unified data stream to schools.   In 2009 the UK, Ireland and the USA developed a webservices protocol that allowed such behind the scenes data sharing.   This protocol has been developed and modified in the light of our experiences in using it and to take account of the differing data streams recorded by European countries.

 Workshop meeting : October 5-7 2011

A three day technical workshop was held at NERC/BGS offices in Keyworth UK with technical experts from the school seismology projects in France the UK, Switzerland and Italy, together with invited participation of representatives from Ireland and the USA. During this meeting the discussion focussed on what data should be shared and how could such data be accessed.   After the meeting NERC/BGS prepared a draft document outlining the proposed data sharing protocols using REST webservice principles. This document is kept as a Live reference document at

http://www.bgs.ac.uk/discoveringGeology/hazards/earthquakes/schoolSeismology/webServices/home.html

Web Services

These pages describe how we see Seismology in Schools International (SISI) being able to share data.

We have designed two possible solutions:

The [first solution](#) is an extension of the existing IRIS web service, with the minimum of changes required to support the new features. The proposed changes should not affect existing clients of the IRIS web service, and other providers will only need to implement a subset of the current system. The issue with this implementation is that it may not be easily extensible to future alterations.

The [second *recommended* solution](#) is based around a model that more accurately describes the core concepts required by a Seismology in Schools system. Consequently a number of logical extension points exist within the model to cater for future additions, such as new or multiple sensor types (accelerometer, etc).

Required features

- To distribute data from multiple service providers
- To allow stations to provide multiple channels of data
- To deal with differing event estimates from multiple service providers, such as USGS, EMSC, etc.

Optional features

- To support automated trace data for events, such as professional stations, via IRIS TimeSeries web service
- To be able to receive updated data in bulk from a service provider

Web Services: First Solution

Contents

API Changes

API changes are formatted as:

- **Additional element**
- ~~Deprecated or unused element~~

URLs are described using URL Templates. The `{serviceRoot}` variable is the location of the web service, e.g. for IRIS this is http://www.iris.edu/hq/ssn/api/.

Dates should be supplied in ISO 8601 format.

The API described here defines a subset of the full API provided by IRIS. Any additional or crossed-out URLs could be supported by an implementing web service, but the URLs and responses described here are sufficient to synchronize a database of network, station, and event information.

Networks

The network response remains almost unchanged.

The ID of a network must be globally unique across all service providers. It is recommended that the ID is registered with FDSN.

URL Template

```
{serviceRoot}/get_all_networks
```

Response Format

- `<networks>`
- o `<network id="IRI">`
- ~~`<iris_network_url>IRIS</iris_network_url>`~~
- `<name>IRIS</name>`
- `<description>IRIS's Seismographs in Schools Program serves teachers across the country and around the world using seismic instruments or real-time seismic data in K-16 classrooms. Our hope is to bridge the gap between science classrooms to create an international educational seismic network.</description>`
- `<contact>Tammy Bravo</contact>`
- `<website>http://www.iris.edu/hq/sis</website>`

- ▪ `<logo_url>http://www.iris.edu/hq/ssn/img/misc/iris.png</logo_url>`
- ▪ `<color>660099</color>`
  `</network>`

- o `<network id="UK">`
- ▪ `...`
  `</network>`

- o `...`
  `</networks>`

Stations

The stations response includes moving the network code to be an attribute of the station tag (essentially to make identifying information clearly visible, and for consistency with other elements), and the addition of multiple channels.

The station ID must be unique within a network, consisting of 3 to 5 alphanumeric characters.

Instrument type, channel orientation, and channel band codes are expected to use the conventions described in SEED Manual - Appendix A: Channel Naming.

For channels, the channel ID is unique for a given station, but no significance is assigned to it's value - the service is free to choose any identifier. A channel label is provided for websites to use when displaying channel information, or related SAC files.

URL Template

```
{serviceRoot}/get_all_stations

{serviceRoot}/get_all_stations/{networkCode}

{serviceRoot}/get_all_stations/{networkCode}/{dayCount}
```

Template Variables

`networkCode` - A network code found from a networks response, or 'ALL' for all networks.

`dayCount` - A station will only be included in the response if it has uploaded data in the previous number of days.

Response Format

- • `<stations>`
- o `<station network="UK" code="BHS">`
- ▪ `<iris_station_url>http://www.iris.edu/hq/ssn/schools/view/BHS</iris_station_url>`
- ▪ `<station_name>Belvoir High School</station_name>`
- ▪ `<network>UK</network>`
- ▪ `<install_date>0000-00-00</install_date>`
- ▪ `<instrument_type code="L">SEP seismometer</instrument_type>`
- ▪ `<latitude>52.93712</latitude>`

- `<longitude>-0.807728</longitude>`
- `<elevation/>`
- `<address_1/>`
- `<address_2/>`
- `<city>Barkestone Lane, Bottesford, Nottinghamshire</city>`
- `<state/>`
- `<zip>NG13 0AX</zip>`
- `<country>UK</country>`
- `<age/>`
- `<website>http://www.belvoirhigh.leics.sch.uk/</website>`
- `<seismogram_photo/>`
- `<realtime_updated>0000-00-00 00:00:00</realtime_updated>`
- `<realtime_image/>`
- `<date_closed>0000-00-00</date_closed>`
- `<created>2010-04-29 06:15:03</created>`
- **`<channel code="SZ">`**
- **`<label>Z</label>`**
- **`<orientation code="Z"/>`**
- **`<band code="S"/>`**
  `</channel>`

- **`<channel code="1234">`**
- **`...`**
  `</channel>`

- `...`
  `</station>`

o `<station `**`network="UK"`**` code="ABC">`
- `...`
  `</station>`

o `...`
  `</stations>`

Events

Station elements and USGS URNs have been deprecated. Instead a seismology agency code has been added to declare who provided the event estimate, and SAC files are provided as part of the event information (to simplify the synchronization process).

The expected agency codes have not yet been defined, but are likely to include values like: ISC, PDE, PDEW, PDEQ, JMA, EMSC, etc...

The event id is unique for a given service provider, it is **not** globally unique. To combine event information from multiple services, an algorithm for resolving event estimates should be applied.

URL Template

```
{serviceRoot}/get_events

{serviceRoot}/get_events/{startDate}

{serviceRoot}/get_events/{startDate}/{endDate}
```

Template Variables

`startDate` - Only return events that have taken place on or after this date. *Only calendar days (not timestamps) are supported.*

`endDate` - Only return events that have taken place on or before this date. *Only calendar days (not timestamps) are supported.*

Response Format

- `<events>`
- `<event id="23124"` ~~`usgs_id="urn:earthquake-usgs-gov:us:c0006390">`~~
  - ~~`<iris_event_url>http://www.iris.edu/hq/ssn/events/view/23124</iris_event_url>`~~
  - **`<agency code="PDEQ" />`**
  - `<date>2011-10-01 12:48:59</date>`
  - `<region>Andaman Islands, India region</region>`
  - `<magnitude>5.1</magnitude>`
  - `<depth>29.4</depth>`
  - `<lat>13.0114</lat>`
  - `<long>95.8531</long>`
  - ~~`<stations>`~~
  - ~~`<station>TATH</station>`~~
  - ~~`...`~~
    ~~`</stations>`~~

  - **`<sacfile network="UK" station="BHS" channel="SZ">`**
  - **`<sacfile_url>http://www.iris.edu/hq/ssn/events/download/7151</sacfile_url>`**
  - **`<sacfile_image>http://www.iris.edu/hq/ssn/img/seis/1109300723eptx.png</sacfile_image>`**
    `</sacfile>`

  - **`<sacfile network="UK" station="BHS" channel="1234">`**
  - **`<sacfile_url>http://www.iris.edu/hq/ssn/events/download/7151</sacfile_url>`**
  - **`<sacfile_image>http://www.iris.edu/hq/ssn/img/seis/1109300723eptx.png</sacfile_image>`**
    `</sacfile>`

  - `...`
    `</event>`

- `<event id="23116"` ~~`usgs_id="urn:earthquake-usgs-gov:nn:00349692">`~~
  - `...`
- `...`
  `</events>`

Synchronization Process

To fully populate an empty database, a subscriber should call:

1. `{serviceRoot}/get_all_networks`
2. `{serviceRoot}/get_all_stations`
3. `{serviceRoot}/get_events`

To update an existing database, a subscriber could call:

- Once a day (or if parent information appears to be missing after an events call):
    1. `{serviceRoot}/get_all_networks`
    2. `{serviceRoot}/get_all_stations`
- Once every ten minutes (or however frequently near-realtime updates are required):
    1. `{serviceRoot}/get_events/{startDate}` - where `startDate` is in the range of 30 days previously

Resolving Event Estimates

Since event information is an estimate at any given time, subscribers should be prepared for updates to the time, location, magnitude etc... Additionally, multiple web service providers will likely provide different estimates for the same event.

Events discrepancies should be resolved by comparing the time and location information using a suitable tolerance. Estimates that fall within the tolerance levels are likely to be the same event.

When choosing between event estimates provided by each service, a priority ranking could be applied via the declared agency. A possible priority list could be:

1. ISC
2. PDE
3. PDEW
4. PDEQ
5. JMA
6. EMSC
7. Other regional/country networks (BGS, etc)
8. Any unknown networks

Second (recommended) solution

This content is a collaborative effort and should be seen as a work in progress. If you have any improvements, alterations, or suggestions please let use know.

The Seismology In Schools Event Language (**SISEL**) is designed to publish information relating to seismic events in a form suitable for **Seismology In Schools** projects.

It expands on [IRIS'](#) work to create their [Seismographs in Schools Web Service](#), while providing support for additional types of information needed to support other Seismology In Schools websites. The system defines a way for sites to publish event information in such a way that other sites can integrate it into their own websites.

Contents

1. [Concepts](#)

   A definition of the concepts and terminology used in the system.

   This section focuses on seismology (and not IT) terms.

2. [Format](#)

   A discussion of how the concepts are encoded into the system.

   This section builds on the previous section, but focuses on the lower-level IT details of the language format.

3. [Publishing](#)

   A description of how the information can be published and synchronized across distributed locations.

Concepts

This content is a collaborative effort and should be seen as a work in progress. If you have any improvements, alterations, or suggestions please let use know.

Seismographs in Schools Web Service

The existing IRIS Seismographs in Schools Web Service defines three core concepts: Network, Station, and Event. In this IRIS service, a station implicitly is made up of one sensor which provided one channel of data. Event provides a definition of where, when, and how strong an earthquake was. For a given event, a station may provide one recording of trace data from it's channel.

In order to accommodate stations which provide multiple channels of data, and to allow multiple web sites with different event data to collaborate, this model needs to be expanded.

Conceptual Model

A **network** can have one or more stations belonging to it. It needs a unique ID (that we recommend is registered with FDSN), a **name**, a **description**, a technical **contact**, a **website**, a **logo** and a **color** to distinguish it from other networks.

A **station** is registered with a network. A station installs and manages one or more **sensors**, which emit timeseries data. These sensors have a **location**, **model type**, **bandwidth** and contain multiple channels. A **channel** has an **orientation**, a **band** and ultimately is the source for a stream of timeseries data.

An **event** nominally occurs at a given location at a particular time. However, our understanding of the time and location as well magnitude may change and at best we have a series of estimates. An **estimate** has a **region**, **date**, **magnitude** and **location**. It's changes over time will be recorded by an **agency**'s catalogues, PDE, etc. We will have a **preferred** estimate at a given point in time.

A station may provide a recorded **trace** of data for a particular event. The trace may contain timeseries data captured on one of the stations channels or be a static data file that has been manually uploaded. There may be one or more **images** associated with each of the traces.

Format

This content is a collaborative effort and should be seen as a work in progress. If you have any improvements, alterations, or suggestions please let use know.

- General principles
    - Be clear and unambiguous where possible
    - Publish as much information as possible, consumers can always filter content.

  XML format principles
    - Borrow from existing standards where possible - e.g. geo, vCard, html...
    - If a property can't be added, leave it out, don't add a blank entry (e.g. don't add something like `<bandwidth />`).

  Approach is a little different to the IRIS web service. This is primarily concerned with describing information rather than defining an API.

  View a working example of an XML document.

Document hierarchy
- sisel
- network *
- link *
- station *
- address
- link *
- sensor *
- location
- channel *
- event *
- estimate *

- location
- resources *
- link *
- trace *
- link *

NB * **= multiple allowed**

Definitions for the core types of element

Attributes / children for each are provided **and explain which are optional and which are required**.

Sisel (root element)

The root element is used in the same way <html> is used in a web page. It is used to group top-level concepts. Sisel stands for Seismology In Schools Event Language.

Name	XML Type	Data Type	Multiplicity	Comments
xmlns	Attribute	URN	1	Namespace identifier for the language version
network	Element	Network	0..*	
station	Element	Station	0..*	
event	Element	Event	0..*	

- `<sisel xmlns="http://purl.org/sisi/sisel/1.0">`

Network

Name	XML Type	Data Type	Multiplicity	Comments
code	Attribute	String	1	Globally unique identifier
name	Element	String	1	
description	Element	String	1	
contact	Element	String	1	
link	Element	Link	0..*	rel types: *related* for website *icon* for network logo
color	Element	Hex string	1	

- `<network code="`**`FR`**`">`
- o `<name>Édusismo</name>`
- o `<description>The Édusismo sismos a lEcole project etc...</description>`
- o `<contact>Joe Bloggs</contact>`
- o `<link rel="related" href="http://www.edusismo.org/" />`
- o `<link rel="icon" href="http://www.edusismo.org/images/bando.jpg" type="image/jpeg" />`

- o `<color>990000</color>`
- `</network>`

Station

Name	XML Type	Data Type	Multiplicity	Comments
network	Attribute	String	1	Network code from FDSN the station belongs to
code	Attribute	String	1	Unique identifier within a network
name	Element	String	1	
registered	Element	Date	1	
link	Element	Link	0..*	rel types: *related* for website
sensor	Element	Sensor	1..*	
address	Element	Address	0..1	

- `<station network="`**`FR`**`" code="`**`VLIB`**`">`
- o `<name>Collège International de Valbonne (06)</name>`
- o `<registered>2011-05-06T12:45:16Z</registered>`
- o `<link rel="related" href="http://www.civfrance.com/CIV/index.php?option=com_content&task=view&id=29&Itemid=105" />`
- `</station>`

Sensor

Name	XML Type	Data Type	Multiplicity	Comments
type	Attribute	String	1	SEED instrument type code
installed	Element	Date	1	
model	Element	String	1	
bandwidth	Element	String	1	
location	Element	Location	1	Expressed as an elevation
channel	Element	Channel	1..*	

- `<sensor type="`**`L`**`">`
- o `<installed>2011-04-06</installed>`
- o `<model>Guralp CMG-3T</model>`
- o `<bandwidth>120s - 50Hz</bandwidth>`
- `</sensor>`

Channel

Name	XML Type	Data Type	Multiplicity	Comments
code	Attribute	String	1	Unique identifier within a station
label	Element	String	1	Human-readable title

Name	XML Type	Data Type	Multiplicity	Comments
orientation	Element	String	1	code attribute should come from SEED orientation values
band	Element	String	1	code attribute should come from SEED band values

- `<channel code="1">`
  - o `<label>Z</label>`
  - o `<orientation code="Z" />`
  - o `<band code="B" />`
- `</channel>`

Event

Name	XML Type	Data Type	Multiplicity	Comments
code	Attribute	String	1	Identifier unique only within the source service provider
estimate	Element	Estimate	1..*	At most 1 estimate can be *preferred*
resources	Element	Resources	0..*	

- `<event code="1">`
- `</event>`

Estimate

Name	XML Type	Data Type	Multiplicity	Comments
preferred	Attribute	String	0..1	Value is *preferred* or attribute is not present
agency	Element	String	1	*Code* attribute indicates the source of the estimate. A complete list of values is not yet defined. Potential values include: ISC PDE PDEW PDEQ JMA EMSC
date	Element	Date	1	
location	Element	Location	1	
magnitude	Element	String	1	scale attribute may be *mb* (body) *ML* (local/Richter) *Mw* (moment)

Name	XML Type	Data Type	Multiplicity	Comments
				Ms (surface wave)
region	Element	String	0-1	*code* attribute and value derived from Flinn-Engdahl Regions
locality	Element	String	0-1	Localised description of where event occurred (such as *Market Rasen, UK*).

- `<estimate preferred="`**`preferred`**`">`
  - `<agency code="`**`CSE`**`" />`
  - `<date>2011-09-15T19:31:01Z</date>`
  - `<magnitude scale="`**`mb`**`">7.1</magnitude>`
  - `<region code="`**`181`**`">Fiji Islands region</region>`
- `</estimate>`

Resources

Name	XML Type	Data Type	Multiplicity	Comments
network	Attribute	Network	1	
station	Attribute	Station	1	
trace	Element	Trace	0..*	
link	Element	Link	0..*	rel types: *related* Or other useful link, such as teaching aid

- `<resources network="`**`FR`**`" station="`**`VLIB`**`">`
  - `<link rel="related" href="http://www.example.org/event/1234" type="application/pdf" />`
- `</resources>`

Trace

Name	XML Type	Data Type	Multiplicity	Comments
channel	Attribute	Channel	1	
link	Element	Link	1..*	rel types: trace preview

- `<trace channel="`**`3`**`">`
  - `<link rel="`**`trace`**`" href="`**`http://www.edusismo.org/ftpsismo/base/20110915/2011.09.15-19.31.01.CIVF.BH-E.SAC`**`" />`
  - `<link rel="`**`preview`**`" href="`**`http://www.edusismo.org/ftpsismo/base/20110915/2011.09.15-19.31.01.CIVF.BH-N.png`**`" type="`**`image/png`**`" />`
- `</trace>`

Location

Name	XML Type	Data Type	Multiplicity	Comments
lat	Element	Decimal	1	WGS:84
long	Element	Decimal	1	WGS:84
depth	Element	String	0..1	A location should include at most 1 depth or 1 elevation. *units* attribute should be *km* or *m*
elevation	Element	String	0..1	A location should include at most 1 depth or 1 elevation. *units* attribute should be *km* or *m*

- `<location>`
- o `<lat>-21.44</lat>`
- o `<long>-179.29</long>`
- o `<depth units="km">605</depth>`
- `</location>`

Address

Name	XML Type	Data Type	Multiplicity	Comments
street	Element	String	0..1	
locality	Element	String	0..1	City, etc
region	Element	String	0..1	County, State, etc
code	Element	String	0..1	Postcode, zip, etc
country	Element	String	0..1	

- `<address>`
- o `<locality>Valbonne</locality>`
- o `<region>Provence-Alpes-Côte d'Azur</region>`
- o `<country>France</country>`
- `</address>`

Link

Name	XML Type	Data Type	Multiplicity	Comments
rel	Attribute	String	1	[Possible existing rel types](#)
href	Attribute	URL	1	
type	Attribute	Media type	0..1	

- `<link rel="related" href="http://www.bgs.ac.uk/schoolSeismology" type="text/html" />`

Publishing

This content is a collaborative effort and should be seen as a work in progress. If you have any improvements, alterations, or suggestions please let use know.

Service requirements:

- Support publication of a number of key concepts (networks, stations, and events)
- Allow subscribers to maintain a local copy of the data

The Atom Publishing Protocol is an existing well-supported standard that meets these requirements.

A dataset can be thought of as the result of a list of changes. For example:

1. Add UK network
2. Add RND-UK station
3. Add Event 1234
4. Add IRI network
5. Update Event 1234 (with a trace from RND-UK station)
6. Add Event 5678
7. Update UK network (correcting a typo somewhere)

The current state of the whole dataset can be copied by running the complete list of changes.

If some more changes become available after 7 in the above list, (e.g. 8, 9, 10) the dataset can be brought up to date by just applying the new changes to the copy.

If a copy exists in state 7, and a number of changes (6, 7, 8, 9, 10) are provided. The changes 6 and 7 can be ignored because they have already been applied to the copy.

An Atom **feed** describes a list of entries.

An **entry** provides information about the state of a 'thing', including an identifier, an update time/version, and links relating to the 'thing'. The entry may also include a copy of the 'thing' itself, in any suitable format.

If more than one entry exists with the same identifier, the one with the most recent update date is used - therefore providing support for publishing modifications to data.

It is therefore possible to publish all the additions and updates to the data in a single feed with an entry for each key concept. An interested website or user can monitor the feed for changes, and can synchronize a local copy of the data. Google use this technique to publish most of their data.

One problem is that a giant list of changes is not very fast or simple when a copy is only missing a few items from the end of the feed. This is of course the most common request: "give me the most recent changes please".

This problem is solved by the related Feed Paging and Archiving standard.

The giant logical feed is split into a sequence of smaller ones, one 'recent' feed and a list of 'archive' feeds. The service is free to choose when to archive content - it could be hourly, daily, monthly, etc... Once a feed is marked as archived, a subscriber never needs to fetch it again.

The process of synchronizing a feed is therefore:

1. Fetch the 'current' feed
2. Process any new entries in the feed
3. Look for a special link to the previous 'archive' feed
4. If there is a link and it has not been visited before, fetch it and repeat from 2

The definition of 'recent' is entirely left up to the service provider, they are free to pick whatever definition best meets their publishing schedule. The subscribers don't need any special knowledge about dates, times, or the periods when data is likely to change. They just poll the current feed, and follow any archive links if they appear. Also, new concepts can be added into the feed if required, existing clients should simply skip any entries that are not understood.

The process is probably best understood by looking at some examples. Because most web browsers natively understand Atom, they will likely provide a special feed view of the data. To see what is going on select 'view source'.

The following example Atom feeds (with embedded SISEL) are available:

- Daily
- Monthly

Also provided is a Java application that was used to generate the above example feeds. You can use it to pick ideas or borrow code.

- Application jar
- Source code