Topic	Seismic waveform data retrieval
Author	Joachim Saul, GFZ German Research Centre for Geosciences, Telegrafenberg, 14473 Potsdam, Germany E-mail: saul@gfz-potsdam.de
Version	December 2014; **DOI: 10.2312/GFZ. NMSOP-2_IS_8.3**

Abstract

This information sheet gives an overview of some of the currently most popular and efficient tools for the retrieval of seismic waveform data from data centers. We do not attempt to provide a complete list of tools or exhaustive documentation. Instead it is our aim to provide some guidance and ideally whet the reader's appetite.

1 Overview

Digital recording methods have been used in seismology since about 1970. Digital methods have greatly increased data quality but they have also created new challenges. In order to allow efficient data exchange, the data had to be homogenized across different computer software and architectures. At the same time the exchange techniques had to keep pace with evolving new technologies. In particular, the advent of the Internet a little over 20 years ago and the continuously decreasing costs for network bandwidth have had great impacts on the way seismic data are acquired, archived, exchanged and processed today.

This chapter gives an overview of the techniques that are now (2014) "state of the art" for retrieving seismic waveform data from data archives. Data retrieval methods can be basically categorized into three groups:

- interactive, web-based solutions
- non-interactive techniques suitable for automated batch-type requests
- continuous, real-time data transfer

We will present some of the more popular solutions for each of these groups and discuss their individual merits and disadvantages.

1.1 Definition of the subject

We want to request seismic waveform data in order to display, analyze or otherwise use them. In general, the request of waveform data takes place at the beginning of a possibly long work flow, which is, of course, very application specific. Data requests nonetheless do very often follow a certain sequence of steps:

A. Discovery of

- seismic events based on hypocenter location, magnitude, type, etc.
- seismic stations based on sensor location (often within a specific epicentral distance range), instrument characteristics, etc.

- a combination of both in order to highlight a certain region, e.g., for tomographic studies

B. Selection of stations and channels based on what we found in step A. This selection process often combines event and station criteria.

C. Selection of time windows, often based on a combination of event and station locations

D. Formulation and submission of the request

E. Wait

F. Pick-up of the data

G. Data pre-processing (extraction, conversion, scaling, etc.)

H. Data processing - the actual data analysis, display, etc.

The subject of this information sheet is to provide some guidance to the novice user through steps A-F. Our aim is not to provide a comprehensive recipe, but to discuss concepts, weaknesses and strengths of various data request techniques. For more in-depth and technical documentation, the (potential) user is referred to external documents.

2 Relevant data formats used for data exchange

Chapter 10.3 gives a comprehensive overview about data formats commonly used in seismology, both for parametric and waveform data. In the following, we briefly summarize only those data formats, which are currently the most relevant for the data retrieval from seismic data archives.

2.1 SEED format

The SEED data format (SEED - Standard for the Exchange of Earthquake Data) was designed for use by the earthquake research community, primarily for the exchange between institutions of unprocessed raw waveform data. It's specification is part of the SEED manual (FDSN, 2012). Today, SEED is also the format that is most commonly provided to the end user by all major seismic data centers, including IRIS DMC, GEOFON, ORFEUS DC, GEOSCOPE and other data centers of the FDSN.

SEED is a binary format and while it is in fact quite complicated, there is de-facto standard software available for decoding and thus helping the end user to convert it to formats used by the most common data processing software packages. This conversion software (rdseed), available from IRIS and add the URL, allows the seismologist to conveniently decode SEED and encode the data into application specific formats like the SAC or AH formats.

A SEED volume consists of an arbitrary number of so-called “blockettes”. The format of these blockettes is described in the SEED manual. There are many different blockette types for raw waveform data as well as for metadata describing the raw data (e.g., station location, channel gain to convert digital counts into units of ground motion, etc.). Raw waveform data and metadata may also be shipped separately, then normally referred to as miniSEED and dataless SEED, respectively.

2.2 MiniSEED format

The miniSEED format is the raw data part of the SEED format. It only consists of those blockette types that are used for storing information about the actual waveforms. That includes the encoded waveform data themselves, but also basic metadata like station code, channel code, start time of a data record and -optionally- timing quality. However, neither event nor instrument response information (not even a gain factor) is encoded in MiniSEED, because these metadata generally change very rarely if at all. They are therefore normally transmitted separately from the waveforms (often in the form of dataless SEED). This makes miniSEED a very efficient, light-weight format for archival as well as real-time exchange. MiniSEED consists of fixed-size blockettes as SEED. A miniSEED data blockette is also often referred to as a “record”. The physical size of a record, which is the smallest block containing contiguous data, is normally a multiple of 512 bytes. The most commonly used record sizes are 512 or 4096 bytes, but 512 bytes is presently the most common record size and is also used by data exchange protocols such as SeedLink described below. Generally the number of data samples that can be stored within one miniSEED record depends on the sampling frequency and efficiency of data compression, besides the physical record size. A 512-byte record can store approximately 400 samples of data if data compression is used, corresponding to data snippets of 20 seconds length at a sampling rate of 20 Hz. For 100-Hz data, the data snippets are accordingly shorter (approximately 4 seconds). This causes a latency of typically 20 seconds in acquiring 20-Hz data, which might be of concern for certain early-warning applications. Such applications, however, commonly record data at a higher sampling rate like 100 Hz, which already dramatically reduces the latency to less than 5 seconds. The processing can be further reduced if the data recorder sends out the data records prior to filling them with the maximum possible number of samples. Currently miniSEED is the basis for or is at least supported by all relevant real-time data processing and archival systems.

2.3 Dataless SEED format

While miniSEED contains only the raw waveforms (usually in digital counts) without any metadata, dataless SEED contains only metadata such as station location and instrument response without any waveform data. This separation allows storage, exchange and configuration of station inventories with the smallest possible overhead.

The main disadvantage of SEED is that it is not easily extensible. For instance the station identifier consists of up to five alphanumeric characters. This limitation is simply due to a pre-defined fixed space limit. There are many other limitations in SEED and the FDSN has therefore decided to introduce a new, XML (eXtensible Markup Language)-based format for station metadata. Dataless SEED will nevertheless continue to be of importance for the seismological research community for many years to come.

2.4 FDSN stationXML

Being an XML (eXtensible Markup Language) based format, FDSN Station XML was designed to overcome the limitations of dataless SEED, particularly the limited extensibility. The format has been approved by the FDSN in 2013 and is meant to become – in the long term -- a replacement of the dataless SEED format.

Some data centers and services make their station inventories available as FDSN StationXML already. Software exists for both the conversion to/from dataless SEED as well as for direct use as inventory data format.

2.5 Further information

For further information about the FDSN standard formats the reader is referred to

- the SEED format version 2.4 specification (FDSN 2012)
- the FDSN StationXML specification (FDSN 2013)

3 Access to archive data

20 years ago, the access to seismic data could be a lengthy and time-consuming procedure, especially for large data volumes. Requests had to be submitted to the data centers and data were returned, e.g., on magnetic cartridges or CD's. Due to the availability and greatly reduced costs of high-speed Internet connections, the data exchange through physical media is now obsolete (2014). The Internet can be and is nowadays used even for synchronization of large data archives. E.g., seismic noise studies utilize huge amounts of data that can now be transferred over the Internet. The cost of hard disk space has decreased greatly, with the benefit for data centers of being able to now store their entire archives online on hard disk, which has brought additional dramatic speed advantages.

This offers many new possibilities for data centers to offer their data in ways more convenient to the users than ever before. While it may have taken weeks to process data requests at the data center and to ship the data tapes to the user, it is now a matter of minutes or at worst hours between the arrival of the user request and the data shipment. Some request mechanisms are even fast enough to provide data "on the fly" within seconds. New approaches for data archival, like distributed data archives, have become possible.

Below we will describe interactive and non-interactive data access techniques.

3.1 Interactive methods

In this section we will explore several web pages that provide interactive access to seismic waveform data. At each of the data portals we will perform an example data request for broadband waveforms of the 2011 Tohoku earthquake. We are interested in teleseismic P waves and want to request all public broadband data from stations within 90 degrees epicentral distance. For the P waves we can restrict our request to the vertical component only. Due to the magnitude of the earthquake and long rupture duration, we request time windows of 5 minutes before to 10 minutes after the predicted P-wave arrival time. Such a data set may be used, e.g., for magnitude computation or body-wave moment tensor inversion.

3.1.1 IRIS Wilber3

IRIS Wilber is a web application for requesting event-oriented data from the IRIS DMC. It allows users to easily find earthquakes from DMC's database and access waveform data based on a particular event. Waveform data are prepared using earthquake travel times and may be downloaded in a variety of data formats including SEED, SAC, and miniSEED. It is worth mentioning that WILBER3 interacts with information at the IRIS DMC totally through FDSN web services deployed at the IRIS DMC.

- URL: http://www.iris.edu/wilber3 (see Figure 1)
- At *Load Event Data* select *Since 1990, M5.5+*
- At the right panel, *Date* field, enter **`2011-03-11`** as both start and end date
- You are now offered a list of events that occurred on the day 2011-03-11. Scroll to the bottom of the list and select the Tohoku main shock.
- Go to the *Select Stations* dialog (see Figure 2). A world map with available stations is now displayed, with the seismic travel times as isolines.
- Seismic Network). To select a different (virtual) network, click on the text field and select the network from the pop-up menu.
- On the bottom panel the selection can be fine-tuned on a per-station level if needed.
- Finally, hit *Request Data* to get a pop-up window in which you have to specify time windows and data format (see Figure 3). Most users will request either SEED or MiniSEED format, but options like SAC can be selected, too. In our example we select output as SEED volume. In this case, the request will be processed by the IRIS BREQ_FAST, which means that once the request is ready (usually within a few minutes) the user is informed by email which SEED file to pick up at the IRIS FTP server.

On the panel to the right of the map, you can further limit the station selections according to distance and azimuth with respect to the event, but also channels (default is BH? - all broadband channels) and (virtual) networks (default is GSN - IRIS Global

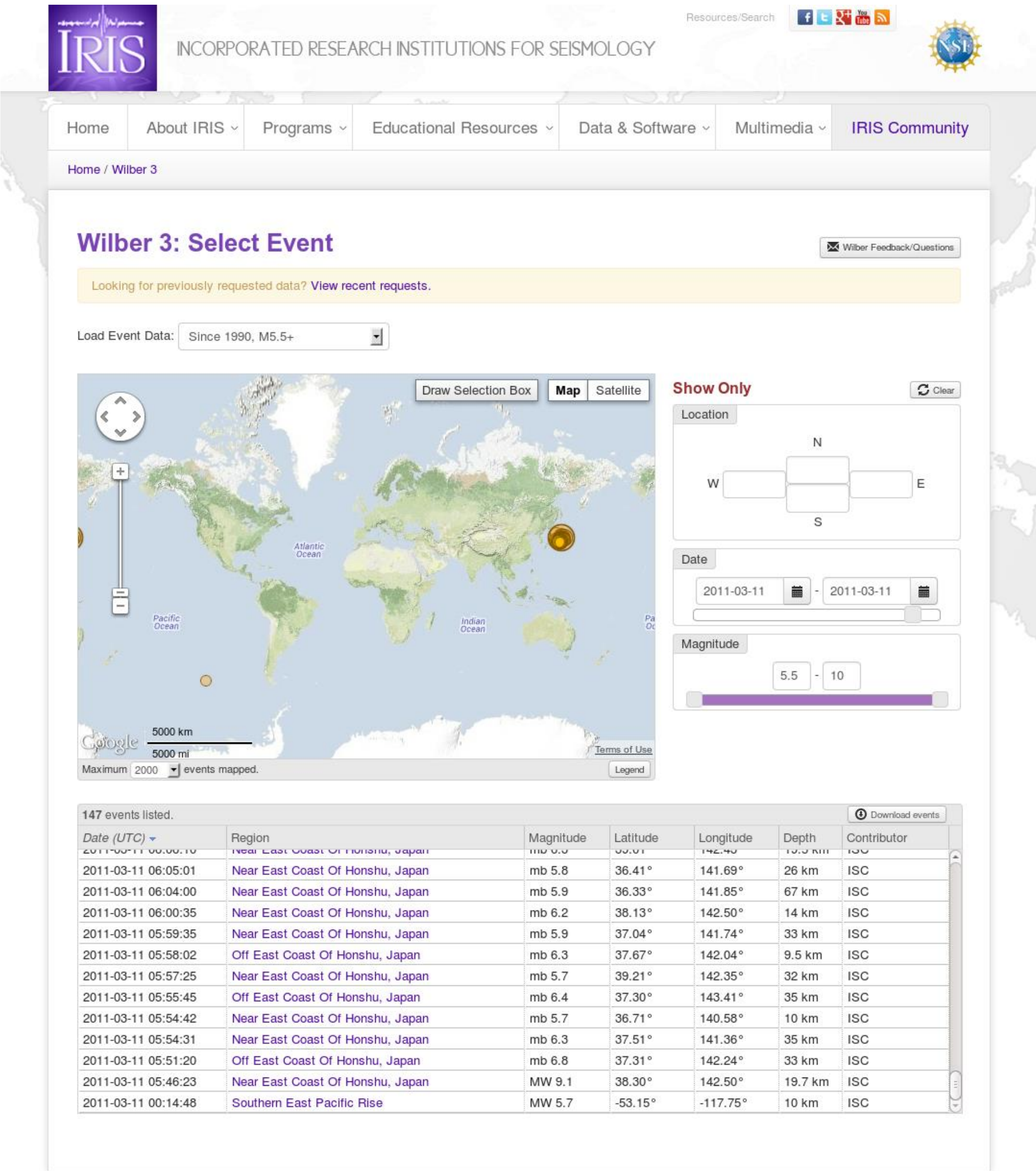

Figure 1 Selection of seismic events using the IRIS Wilber 3 interface

3.1.2 EIDA WebDC

The WebDC web interface is a portal to the seismic waveform data archives at GEOFON, including the European Integrated Data Archive (EIDA) data holdings. It allows to request

- seismological waveform data and
- station inventories (Dataless SEED, Inventory XML).

from different data centers across Europe participating in the EIDA initiative. In order to retrieve data for the Tohoku earthquake follow these steps:

- URL: http://webdc.eu
- We start our request session at the *Explore events* tab (see Figure 4)
- Select a catalog: GFZ is default, USGS and EMSC are other options. Alternatively, a user-supplied event list may be used and uploaded to the WebDC server. This is particularly useful to request data for events not covered by the above earthquake catalogs.
- As in the Wilber3 example above, we again select **`2011-03-11`** as start and end date. By limiting the magnitude to 7 and above, we restrict our search to the Tohoku main shock and two major aftershocks.
- Once the main shock is selected, we move on to the *Explore stations* tab (see Figure 5) and select the GEOFON network (GE) and streams by code *BH*, which means broadband streams. By hitting "Search" the station selection is displayed on the map as well as in a list at the bottom of the page, where stations can be (de)selected individually if desired.
- Finally, to submit the request, we move on to the *Submit request* tab, where we can choose between relative and absolute time windows (see Figure 6). Relative time windows are specified relative to a selectable wave type (P by default). The data formats that can be requested include SEED, MiniSEED as well as station metadata (dataless SEED). After specifying a valid email address the “submit” button is hit to submit the data request to WebDC.
- After submitting the request, the user will be notified about it being processed by a blinking “Download data” tab at the top of the page. By clicking on that tab, a list of the most recent data request of the user (identified by the email address) will be displayed, along with the processing status and -once the request is completed- one or several download links (see Figure 7). Note, however, that the data may have been split into several volumes as EIDA is an integrated data center with several nodes in Europe, each of which is responsible for its own data. This also means that the user downloads the data directly from the respective node.

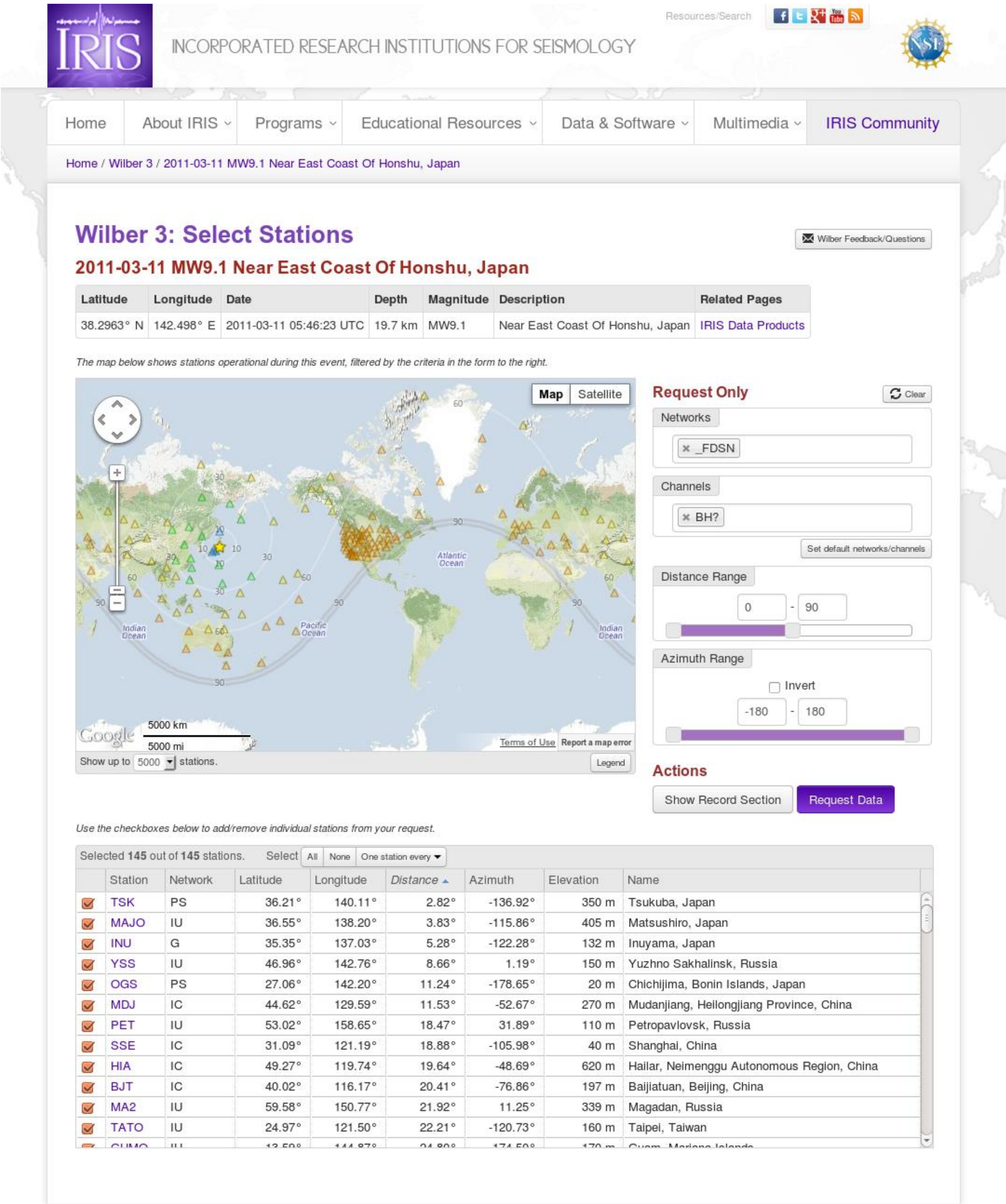

Figure 2 Station selection using the IRIS Wilber 3 interface

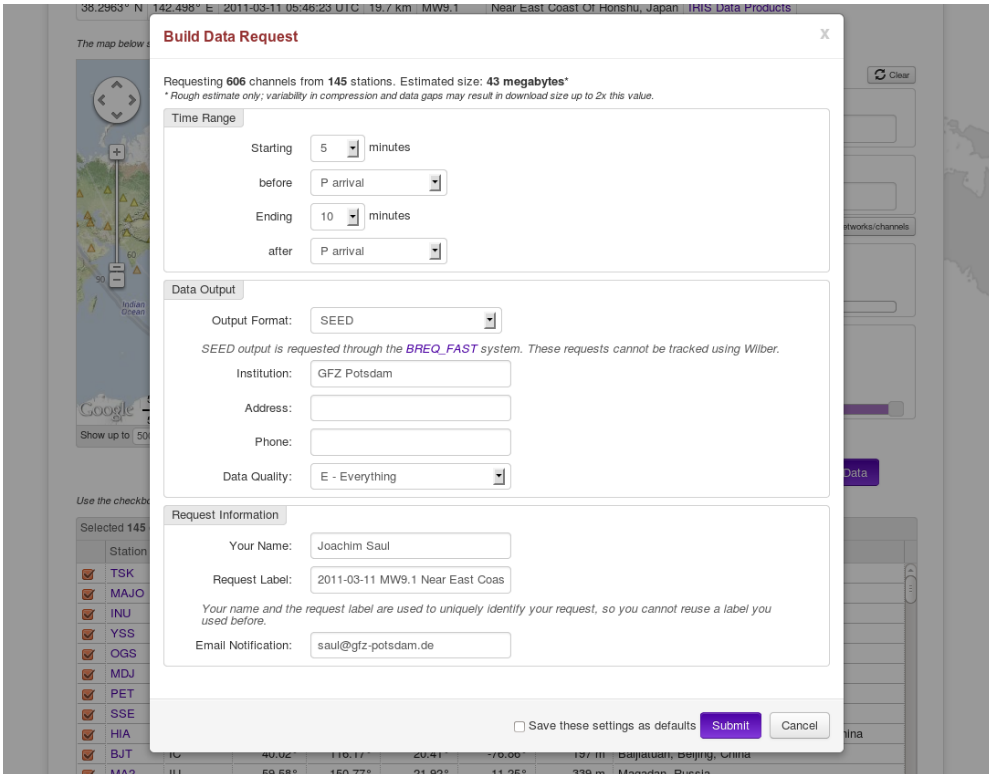

Figure 3 Formulation of the actual data request using the IRIS Wilber 3 interface

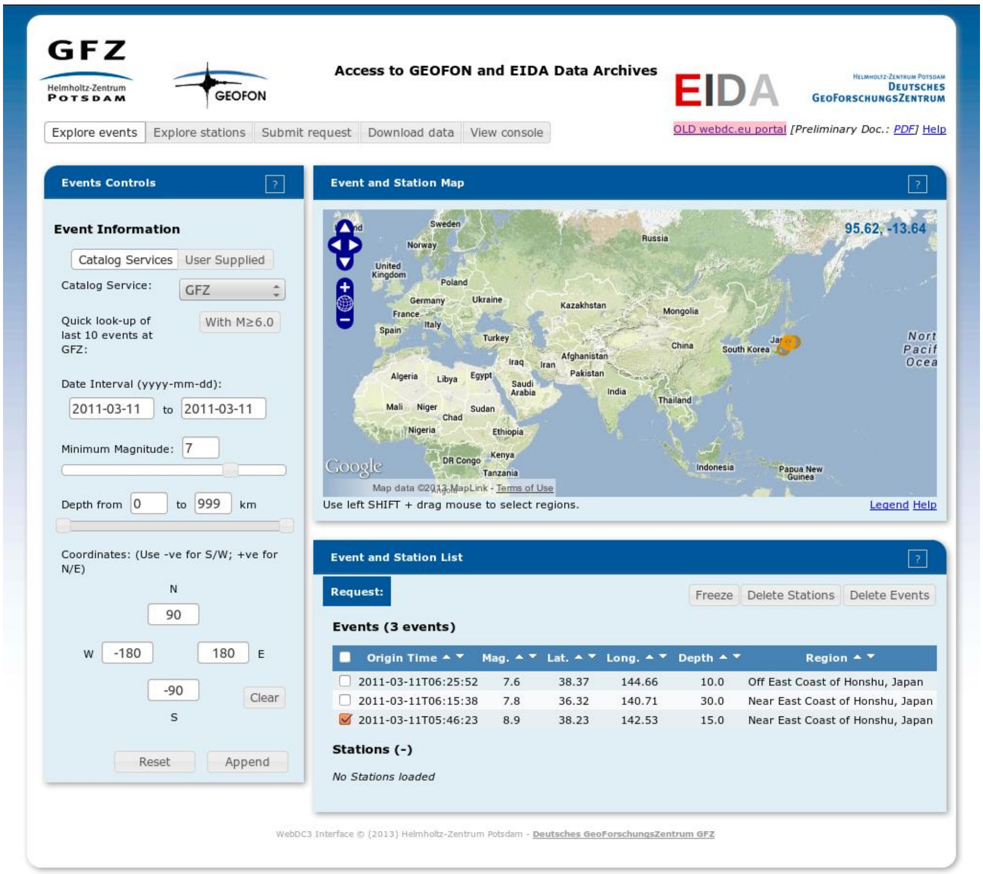

Figure 4 Explore seismic event catalogs using the WebDC interface

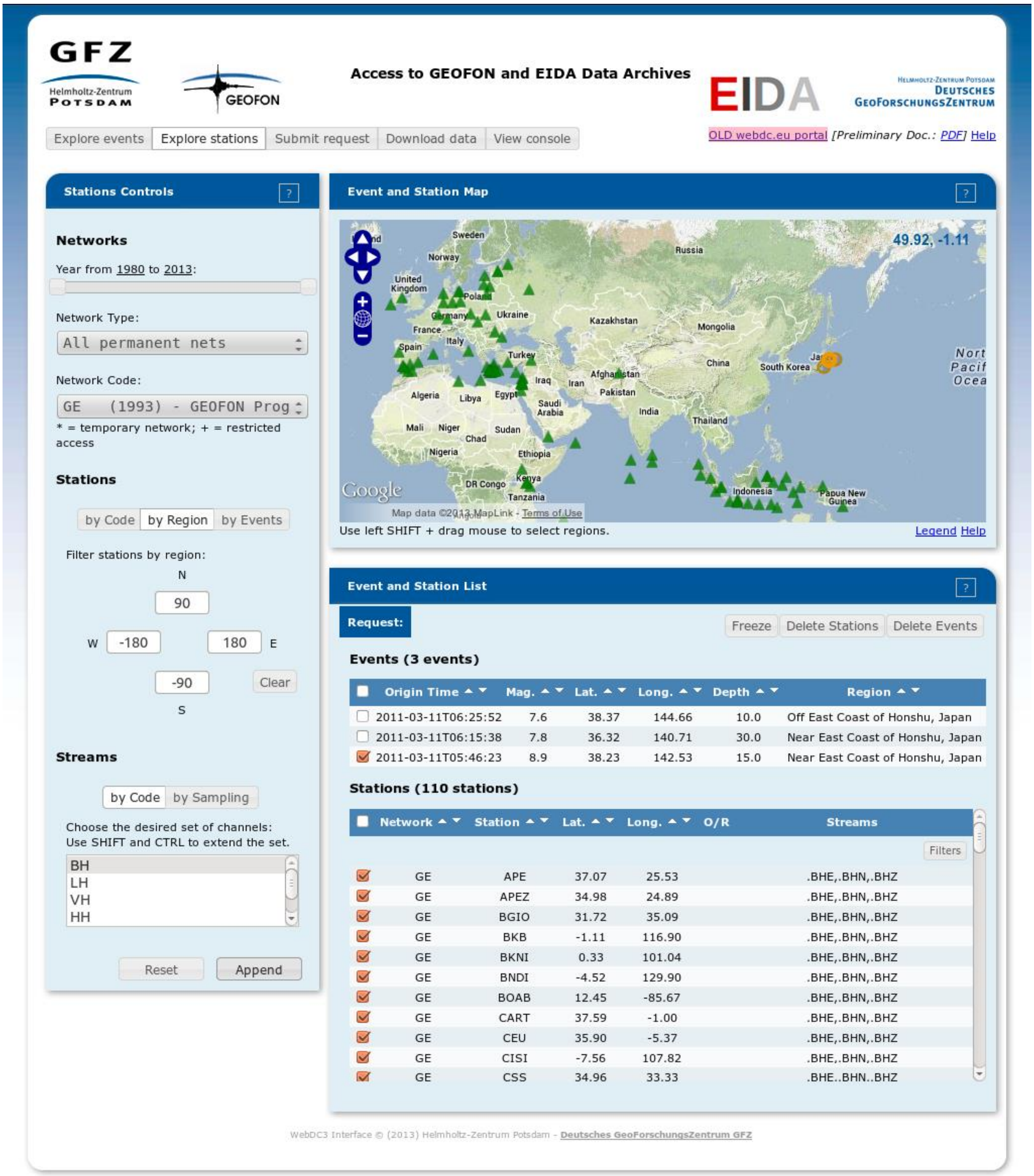

Figure 5 Explore stations using the WebDC interface

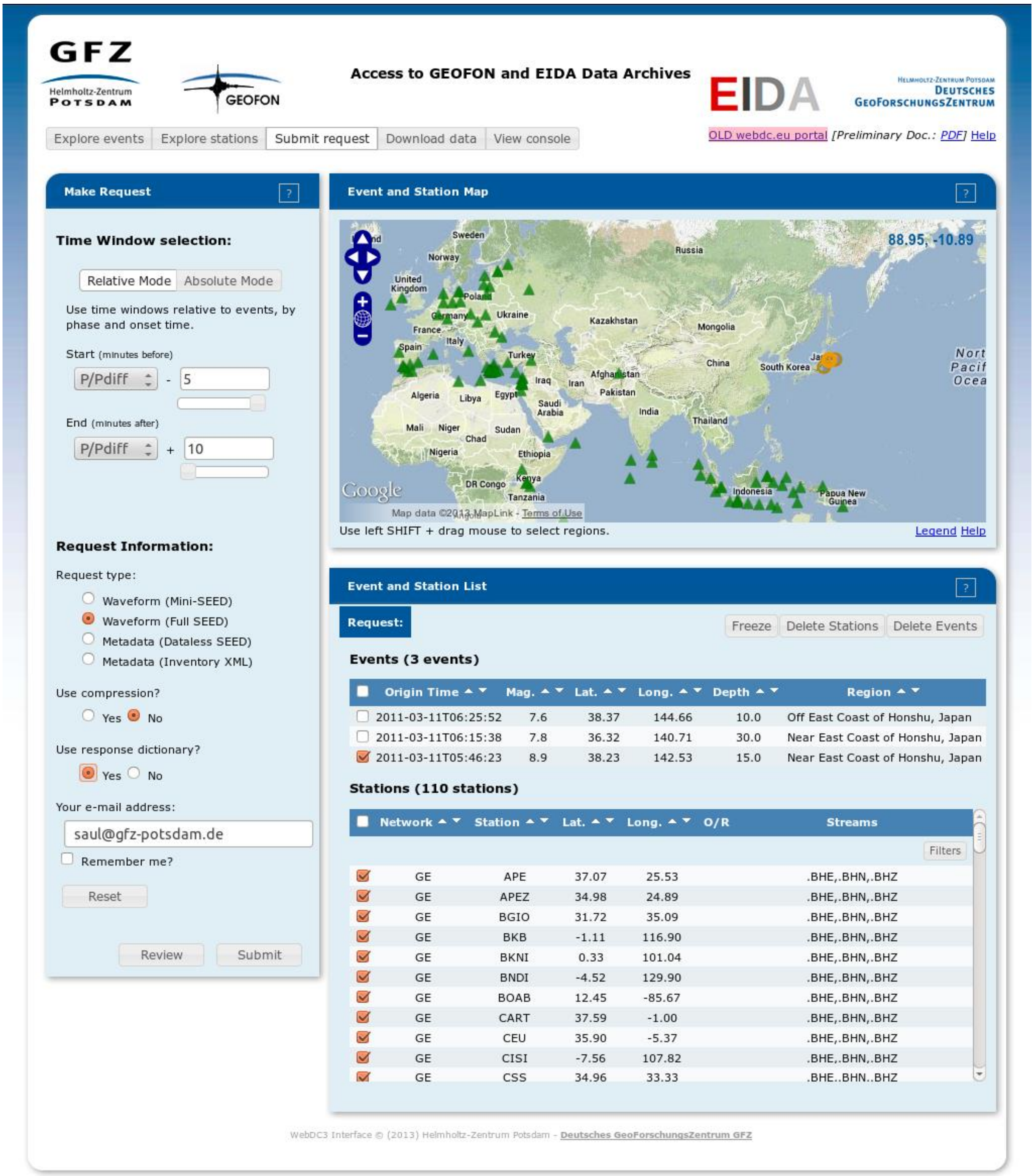

Figure 6 Data request submission

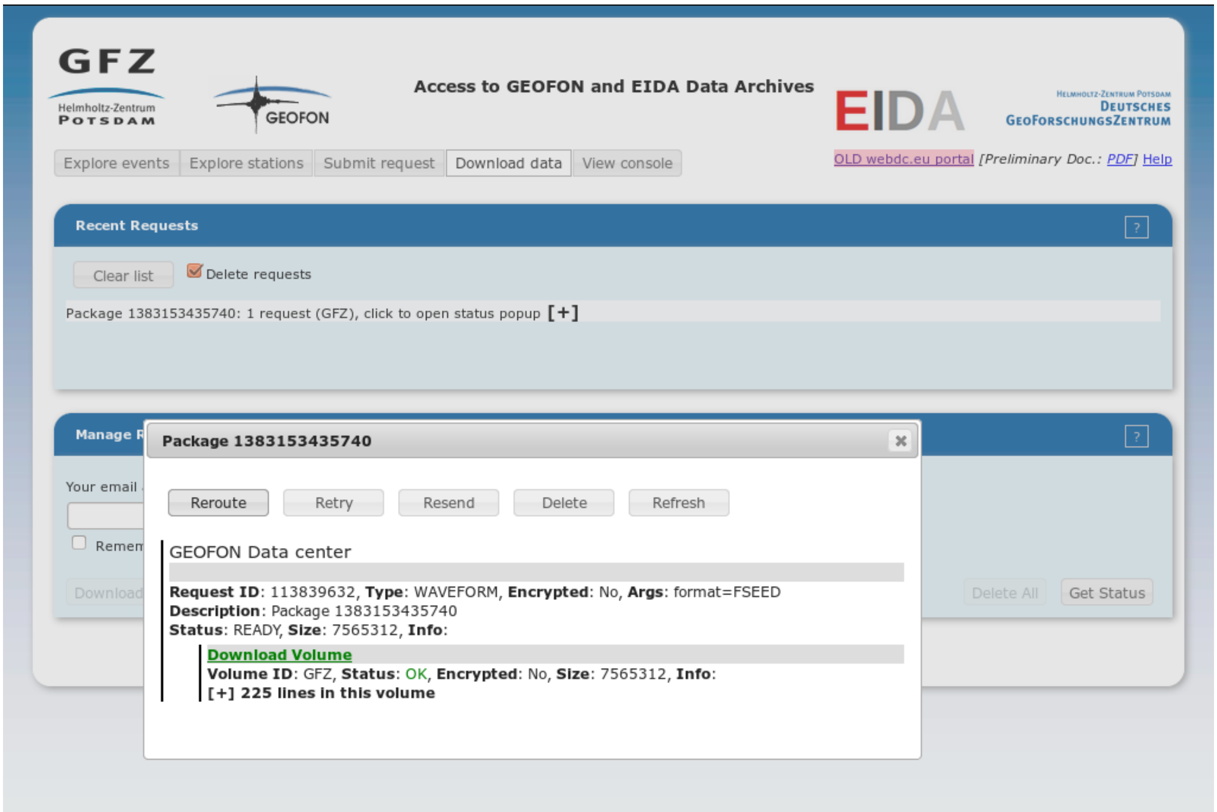

Figure 7 Finally, download the data as a SEED volume

3.1.3 Remarks

Both presented interactive request methods are similar in that they greatly facilitate the whole process of event selection, station discovery and retrieval. They are the methods of choice for the retrieval of a limited data set for one or few earthquakes, especially if it is unclear which stations shall be used. If large data sets need to be compiled, the interactive sessions may be too time consuming. Fully automated processing is not possible at all, at least not until the data have been downloaded. This is where the use of non-interactive methods is required, which we will look at next.

3.2 Non-interactive methods

Any user who wants to retrieve large quantities of data, like many stations for many events, will quickly realize that while interactive methods provide an excellent starting point for exploring data archives, often at the beginning of larger projects, they are normally too inefficient to use if hundreds or thousands of data sets need to be retrieved. This is where automated requests using non-interactive request methods come into play. Automated requests are also needed in automated seismic monitoring environments.

A variety of automatic request tools exist, some of which have been in existence for many years and are widely known and used, others are newer and reflect technical developments of the recent years.

Here we will provide an overview and a discussion about some of the more important request tools.

3.2.1 Email based methods

Being one of the earliest Internet communication protocols, email was quickly adopted as a means for requesting and transmitting seismic data. Several different techniques have been in use over the years, some of which are now obsolete but mentioned here for the sake of completeness:

AutoDRM

developed at the Swiss Seismological Service (SED) at ETH Zürich. AutoDRM was a purely email based request service in which both the request as well as the compiled data volume were transmitted as text messages through email in GSE format. AutoDRM pioneered seismological data access but is now obsolete due to the availability of much more efficient techniques. However, some data centers especially in Europe, continue to provide this service.

BREQ_FAST

is another email based tool and was developed by IRIS. Data requests are sent in an email message to a data request manager, where they are processed. Upon completion, the data are delivered via a medium as specified in the request. Nowadays the FTP Internet protocol is the most common medium, but for large data volumes magnetic tapes might occasionally still be the preferred media type.

A BREQ_FAST request message consists of a header, in which the user provides some personal information like name, postal and email address. The user can specify a preferred media type for the data delivery and a request label, under which the request can be tracked.

Following the header, the request message contains an arbitrary number of lines, each of which consists of a combination of station/network/channel codes and associated time windows to be extracted. Each of the textual fields may contain wild cards, e.g., for matching more than one channel code to get all available components of a sensor.

For more information about BREQ_FAST, see
http://www.iris.edu/dms/nodes/dmc/manuals/breq_fast/

A detailed example will be shown below.

NetDC - Networked Data Center Protocol

Like BREQ_FAST, requests are email based, whereas data are delivered to the user via either the FTP Internet protocol or magnetic tapes (Exabyte). NetDC is in principle superior to BREQ_FAST because it allows the user to submit a data request without knowing where the data are stored physically. It introduced the concept of "networked data centers", where requests are automatically routed to the data center responsible for the data.

When making the data request, the user can choose between either receiving data from individual data centers or receiving the data as a single merged product from the data center originally contacted. Either way a user interested in data will have to contact only one site when making a request, which is certainly a big advantage.

Due to the availability of more modern request techniques, NetDC is now considered obsolete. While it is still in use at a few data centers (e.g. GEOSCOPE) it has already been shut down at the IRIS DMC in 2013.

Out of these three most popular email based request techniques, nowadays only BREQ_FAST continues to be of importance and is supported by GEOFON, IRIS and EIDA. We will therefore briefly explain the basics of a data request using BREQ_FAST.

Example BREQ_FAST waveform data request

In the following example, we want to request a recording of the 2011 Tohoku earthquake recorded at the GEOFON station Wanagama, Indonesia, which has the station code UGM. The GEOFON network code is GE. We want to retrieve all three components in a time window starting 10 minutes before and ending one hour after the earthquake origin time, which is 2011-03-11 05:46:23 UTC.

```
.NAME Joe Seismologist
.INST GFZ Potsdam
.MAIL Telegrafenberg, Potsdam, Germany
.EMAIL joe.seismologist@seismology.org
.MEDIA: Electronic (FTP)
.LABEL Tohoku-UGM
.END

UGM GE 2011 03 11 05 36 23.0 2011 03 11 06 46 23.0 1 BH?
```

As data for this particular station are archived both at GEOFON and IRIS data centers, the request may be submitted to either the GEOFON/WebDC (email: breq_fast@webdc.eu), the ORFEUS DC (email: breq_fast@knmi.nl) or the IRIS DMC (email: breq_fast@iris.washington.edu). In general, however, when using BREQ_FAST for requesting data, it must be known beforehand which data center provides what data, because the BREQ_FAST service itself provides no functionality for station discovery. This is often not a problem, especially if the same set of stations (or “virtual network”) is always used for a certain analysis and if the number of stations is small enough to be easily manageable using e.g. home grown station inventories like station lists. When BREQ_FAST was developed more than 20 years ago, this was possible simply because the digital seismic networks consisted of much fewer stations than today.

But things have changed. What if there are many stations in our region of interest and we want to use just as many of them as possible? What if the station/network configuration changes nearly on a daily basis like in the case of the EarthScope network? It then becomes increasingly tedious and time consuming to keep a local station inventory up to date. Fortunately, there are now services that make this task easier. We will explore these in more detail below.

3.2.2 ArcLink

ArcLink is a data request protocol developed within EIDA especially to facilitate access to distributed, de-centralized data archives. High-quality, digital seismic waveform data are nowadays openly available from many institutions and data centers world-wide. Data are therefore physically stored in different places. This requires tools that facilitate the discovery and access to such distributed archives. In particular, data requests need to be routed from a client to often several server nodes. ArcLink can therefore be considered an enhancement of the now discontinued NetDC.

ArcLink uses its own protocol language between an ArcLink server and an ArcLink client program. ArcLink is also the architecture behind the EIDA WebDC data portal. It consists of dedicated server and client software. For data retrieval, a client software is needed. There are several options, with the most popular being

arclink_fetch

is a stand-alone download client, which is distributed as part of the SeisComP software package (Hanka et al., 2010). It is invoked either from the command line prompt or from within shell scripts:

```
$ echo '2011,3,11,5,36,23 2011,3,11,6,46,23 GE UGM BHZ *' | ¥
arclink_fetch -u "joe@seismology.org" -o GE.UGM..BHZ.mseed
```

The resulting MiniSEED file GE.UGM..BHZ.mseed can then be saved and viewed by using any seismogram viewer or be further processed.

obspy.arclink

is a library of client functions that are invoked from programs written in the Python programming language. Example:

```
$ python
Python 2.7.3 (default, Feb 27 2014, 19:58:35)
[GCC 4.6.3] on linux2
Type "help", "copyright", "credits" or "license" for more information.
>>> from obspy import UTCDateTime
>>> from obspy.arclink.client import Client
>>> client = Client(user='joe@seismology.org')
>>> t = UTCDateTime("2011-03-11 05:46:23")
>>> st = client.getWaveform("GE", "UGM", "", "BHZ", t-600, t+3600)
>>> st.plot()
```

Instead of plotting the data, it is of course also possible to write them to a file:

```
>>> client.saveWaveform("GE.UGM..BHZ.mseed",
... "GE", "UGM", "", "BHZ",
... t-600, t+3600, format='MSEED')
```

This will save the raw MiniSEED data to a file GE.UGM..BHZ.mseed, which can then be processed. Also other formats commonly used by data analysis packages are supported (e.g., SAC, SEISAN, GSE etc.)

SeisComP3

is software for seismological data archival, exchange and processing. It has built-in ArcLink client support. In fact, ArcLink is the default archive access protocol of SeisComP (Hanka et al., 2010).

NOTE: ArcLink is a popular protocol in Europe within the EIDA initiative and within the SeisComP ecosystem. It is not supported by major data centers like IRIS DMC, however, and has therefore not become a world-wide standard. A standard archive access mechanism is provided by web services.

3.2.3 Web Services

Web services in general are a method of communications over the World Wide Web, hence the name. The fact that web services use the HTTP Internet protocol allows access to resources offered by web services usually even in tightly security controlled environments (e.g., behind network firewalls, etc.), because the Internet ports used by the web services are the same as for interactively browsing the WWW. Many Internet technologies nowadays are based on web services and as a natural consequence web services also have been adopted in seismology as a means for exchange of waveform and parametric data.

Web services for seismic waveform data retrieval have been developed over the last approximately 10 years in several institutions in Europe and the United States, resulting in initially different and incompatible solutions. Fortunately in 2013, an agreement on a common web service standard could be achieved by the members of the FDSN. The set of web services defined in this standard are referred to as "FDSN web services"' (FDSN, 2013):

`fdsnws-station` - For access to station metadata in FDSN Station XML format
`fdsnws-dataselect` - For access to time series data in MiniSEED format
`fdsnws-event` - For access to event parameters in QuakeML format

The access to the FDSN web services is very simple and there exists a variety of techniques to chose from. In fact, the HTTP protocol allows the use of any web browser to access data from an FDSN web service, which is very convenient especially for the novice user to explore the web service features. Non-interactive HTTP client programs like "wget" and "curl" may easily be used as FDSN web service clients. In addition, specialized client software exists to make the data retrieval simple and transparent.

The data request can be specified in two different ways. One is simply via a URL, which essentially must consist of the stream to load (possibly containing wild cards matching several streams at once) and the time window. The stream is specified by the network, station, location and channel codes. See the SEED manual for more details about these fields.

For example, if we want to request the same data as in the above example BREQ_FAST waveform data request (70 minutes of data for the 2011 Tohoku earthquake recorded at GEOFON station UGM), the URL to fetch these data would be http://geofon.gfz-potsdam.de/fdsnws/dataselect/1/query?sta=UGM&net=GE&start=2011-03-11T05:36:23.0&end=2011-03-

11T06:46:23.0&cha=BH? This URL may simply be entered in any web browser and the resulting data file may be immediately viewed or processed.

Using the same URL, the data can also be fetched with command line utilities like “wget” or “curl”, which allows partial automation of data requests, e.g. using shell scripts. Since the URLs are very easy to generate, this is a very convenient method for requesting and immediately retrieving seismic data.

When many different stations or time windows shall be retrieved at one time, these cannot all be specified on a single URL. Here an alternative flavor of the web service comes into play, which uses the HTTP-POST technique to upload a request file to the web service. This request file contains lines with individual combinations of streams and time windows, thus similar to the BREQ_FAST email requests. The text file containing the request is uploaded to the web service and the data are retrieved immediately.

NOTE that the data retrieved using FDSN web services are formatted as MiniSEED, which means that unlike full SEED volumes, they lack metadata such as instrument responses which, if needed, have to be requested separately using the fdsnws-station web service.

wget

is a general-purpose HTTP client that can also be used to request seismic data using the FDSN web services. Invocation simply as

```
$ wget -O data.mseed "http://..."
```

curl

is an alternative to “wget”. Invocation equally simple as

```
$ curl -o data.mseed "http://..."
```

obspy.fdsn

is a client library that may be used from programs written in the Python programming language. Example:

```
$ python
Python 2.7.3 (default, Feb 27 2014, 19:58:35)
[GCC 4.6.3] on linux2
Type "help", "copyright", "credits" or "license" for more information.
>>> from obspy import UTCDateTime
>>> from obspy.fdsn import Client
>>> client = Client("GFZ")
>>> t = UTCDateTime("2011-03-11 05:46:23")
>>> st = client.get_waveforms("GE", "UGM", "", "BHZ",
... t-600, t+3600)
>>> st.plot()
```

Note the great similarity with the obspy.arclink example provided above. Instead of plotting the data, it is of course also possible to write them to file:

```
>>> client.get_waveforms("GE", "UGM", "", "BHZ", t-600, t+3600,
... filename="GE.UGM..BHZ.mseed")
```

This will save the raw MiniSEED data to a file GE.UGM..BHZ.mseed, which can then be processed. Note that unlike in the above obspy.arclink example, the data format is not specified because here the only available format is MiniSEED, as it is not possible to download full SEED using the FDSN web services. Note that the above call only retrieves raw waveform data. In order to retrieve metadata, the above ObsPy client provides a "get_stations" method. Please refer to the ObsPy documentation for more details (e.g., Beyreuther et al. 2010, and the ObsPy home page).

IRIS Fetch scripts

can be used on the command line or from shell scripts. Invocation example:

```
$ FetchData -N GE -S UGM -C BHZ -s 2011-03-11T05:36:23 ¥
    -e 2011-03-11T06:46:23 -o UGM.mseed -m UGM.metadata
```

In the above call both raw waveform data as well as basic metadata like sensor coordinates and gain are downloaded.

Data for GEOFON station UGM happens to be available from both GEOFON and IRIS. By default, the FetchData script connects to the IRIS fdsnws server. In order to connect to the GEOFON fdsnws server, this default can be overridden using so-called environment variables. The above call then becomes:

```
$ export SERVICEBASE=http://geofon.gfz-potsdam.de
$ FetchData -N GE -S UGM -C BHZ -s 2011-03-11T05:36:23 ¥
    -e 2011-03-11T06:46:23 -o UGM.mseed -m UGM.metadata
```

SeisComP3

comes with a fdsnws server implementation included, which can be configured to access the data through ArcLink or directly from a mounted file system. Additionally, SeisComP has built-in support for fdsnws-dataselect as client. All SeisComP programs, which use waveform data as fixed time windows can be configured to retrieve waveforms through fdsnws-dataselect.

Other web service clients

for MatLab and Java are available for download under http://service.iris.edu/clients

As the number of data centers supporting FDSN web services increases, it should be possible to develop clients that can discover data holdings across the entire system of centers. For instance, IRIS is currently developing a web service that will look at the holdings of all data centers running FDSN web services, and return information to a client application that will enable parallel recovery of waveforms from multiple data centers. Again the data will be sent from multiple data centers directly to the client application allowing each data center to understand its customer base. A similar "routing" facility is also being developed in Europe within EIDA. Any such solution will require special "smart client" software to make this data center integration transparent to the user.

3.2.4 Data discovery

In the previous section about web services, we have assumed that the user knows which data streams to request. In general, however, this may not always be the case and we may want to make a selection of stations/streams based on ad-hoc criteria such as a geographical region or epicentral distance range around an earthquake. This "discovery" process becomes actually quite simple using the "fdsnws-station" web service. If for instance we want to request data from IRIS for all stations within 50 degrees around the 2011 Tohoku earthquake. Given the epicenter coordinates and the time window it is very simple to request a list of stations within that distance range:
http://service.iris.edu/fdsnws/station/1/query?level=channel&format=text&channel=BHZ&network=I*&includerestricted=false&start=2011-03-11&end=2011-03-11&longitude=142.5&latitude=38.2&maxradius=50

This is a long URL, but the format is actually quite simple and almost self-explanatory. `level=channel&format=text&channel=BHZ` requests a text list containing BHZ channels. Stations without BHZ channels will therefore not be listed, allowing a crude pre-selection of broadband sensors. Here we want to list only stations from the IRIS networks (II,IU,IC).

The full specification for the FDSN web services can be found under http://www.fdsn.org/webservices/.

3.2.5 Outlook

Nowadays automated data retrieval is possible using email based techniques as BREQ_FAST, ArcLink and web services. All of these techniques have their specific advantages and will continue to be available for years to come. BREQ_FAST and ArcLink allow data retrieval as full SEED volumes, which the web services don't. Instead, the web services promote a stronger separation of waveforms (MiniSEED format) and metadata (FDSN StationXML). This may appear like a disadvantage and for many users used to working with full SEED it currently is. Work flows need to be adopted, which is not always trivial. However, the data separation also has advantages. Erroneous metadata may more easily be corrected if the metadata are loaded "on the fly" during processing. Use of XML as metadata format will allow extensions to the metadata schema more easily.

It is likely that in the future the web services become by far the most important request mechanism, as they rely not only on standard data formats but are also based on standardized protocols. The web services are therefore being adopted and implemented worldwide at a growing number of institutions. What the web services are currently lacking is a possibility for routing requests as in ArcLink. It is therefore still necessary to know where the data are physically located. Also it is not yet possible to request data depending on quality metrics. But at the time of this writing, there are several initiatives working in that direction and working solutions will likely become available within the year 2015.

4 Conclusions

We have presented some of the main methods for requesting seismic data. The interactive web portals provide an excellent way to conveniently and quickly retrieve the data for a limited data set in a nearly self-explanatory way. If you have done it once you know how it works. Very little learning is required.

The non-interactive methods - web services or `arclink_fetch` but also old-fashioned email - are much more powerful whenever the requests can be automated, e.g. if very large and/or numerous data sets have to be assembled. Embedding these mechanisms into a work flow may require a little more work at the beginning. Client software like ObsPy helps to minimize the work, which will anyway soon pay off due to the greatly reduced work to retrieve the data.

Acknowledgment

The very careful and constructive reviews by W. Hanka, T. Ahern and J. Wassermann are acknowledged with thanks, as well as the reviews and discussions at several stages of the developing draft with the Manual editor, who also cared for the final formatting.

References

Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr Y. and Wassermann, J. (2010). ObsPy: A Python Toolbox for Seismology SRL, 81(3), 530-533 DOI: 10.1785/gssrl.81.3.530

Hanka, W., Saul, J., Weber, B., Becker, J., Harjadi, P., Fauzi, GITEWS Seismology Group (2010). Real-time earthquake monitoring for tsunami warning in the Indian Ocean and beyond. - Natural Hazards and Earth System Sciences (NHESS), 10, 12, p. 2611-2622. DOI: 10.5194/nhess-10-2611-2010

FDSN (2012). SEED reference manual, Standard for the Exchange of Earthquake Data, SEED format version 2.4, August 2012, http://www.fdsn.org/seed_manual/SEEDManual_V2.4.pdf

FDSN (2013). Web Services Version 1.0 Specification, available as online document under http://www.fdsn.org/webservices/FDSN-WS-Specifications-1.0.pdf

Links to software packages:

ObsPy home page: http://www.obspy.org

SeisComP home page: http://www.seiscomp3.org

IRIS rdseed manual: http://www.iris.edu/ds/nodes/dmc/manuals/rdseed