IRIS Data Management Center Data Access Tutorial

skip to: Types of Data Data Request Tools Receiving the Data

The purpose of this document is to introduce users to the types of data archived at the IRIS DMC and how to request it. There is an Adobe Acrobat® version that can be downloaded as a tar file that includes all of the manuals for the tools mentioned in the tutorial or, you can download just this tutorial as a pdf. *Note: the Web version of this document is the most up-to-date version of the tutorial and should be used in preference to the printable version!*

Types of Data at the DMC

Before making a data request, it is important to understand what types of data we have at the IRIS Data Management Center (DMC). There are basically two types of data (i.e. waveform data) - continuous and event-related - but multiple data products available for each type. Below are explanations of the different types of data available from the DMC.

Continuous Data
For users interested in time windowing their own data

MiniSEED data is the *data only* file format. For more about miniSEED, look here.

BUD: near real-time, not quality controlled (unQC'd), partial data sets

Some data streams into the DMC in near real-time into an online buffer called BUD (Buffer of Uniform Data). Data in BUD is not quality controlled. The data in BUD is continuous miniSEED data organized by channel day. BUD data should be useful to the person interested in doing their own time windowing of near real-time data. Here is more information about the BUD system.

Archived Data: delayed, quality controlled (QC'd), full data sets

A customized data set from the archive is defined by the user. The following sections of this tutorial will cover how to make a customized SEED data request and how to read the data you receive. The DMC has several tens of terabytes of waveform data. Customized requests allow anyone to request any subset of this large archive.

Event Oriented Data
For data users interested in predefined event-oriented time windows

SPYDER® is an acronym for **S**ystem to **P**rovide **Y**ou **D**ata from **E**arthquakes **R**apidly.

SPYDER® data: near real-time, unQC'd, partial data sets

SPYDER® data are event-oriented data products created shortly after an earthquake occurs. The data is unchecked and should be considered to be of unknown quality. This data comes from a variety of sources, primarily from the IRIS DMC's BUD system However, autoDRMs around the world and direct station dial-up augment the data in SPYDER®. SPYDER® data are accessed via the WILBER II interface.

The FARM: delayed, QC'd, full data set

The DMC recognizes the fact that the most interesting and most frequently requested data come from major earthquakes. The most notable events are larger than a magnitude of 5.5 (Mw). For this reason, the DMC routinely pre-assembles data from earthquakes that exceed magnitude 5.7 at any depth, and for events down to magnitude 5.5 if the depth is greater than 100 km. These pre-assembled data sets are called FARM Products and consist of data collected from stations all over the world and from many different networks.

FARM is an acronym for **F**ast **A**rchive **R**ecovery **M**ethod.

Recognizing the fact that most seismic phase arrivals occur within 60 minutes of the initial phase arrival at a seismic station, we extract a 60 minute window for broadband, mid period, and short period channels. Long period windows are a function of magnitude, tuned to record the maximum number of Rayleigh phases for larger events. FARM products become available 6 weeks to 3 months after the event and are most easily accessed using WILBER II (described later in this tutorial).

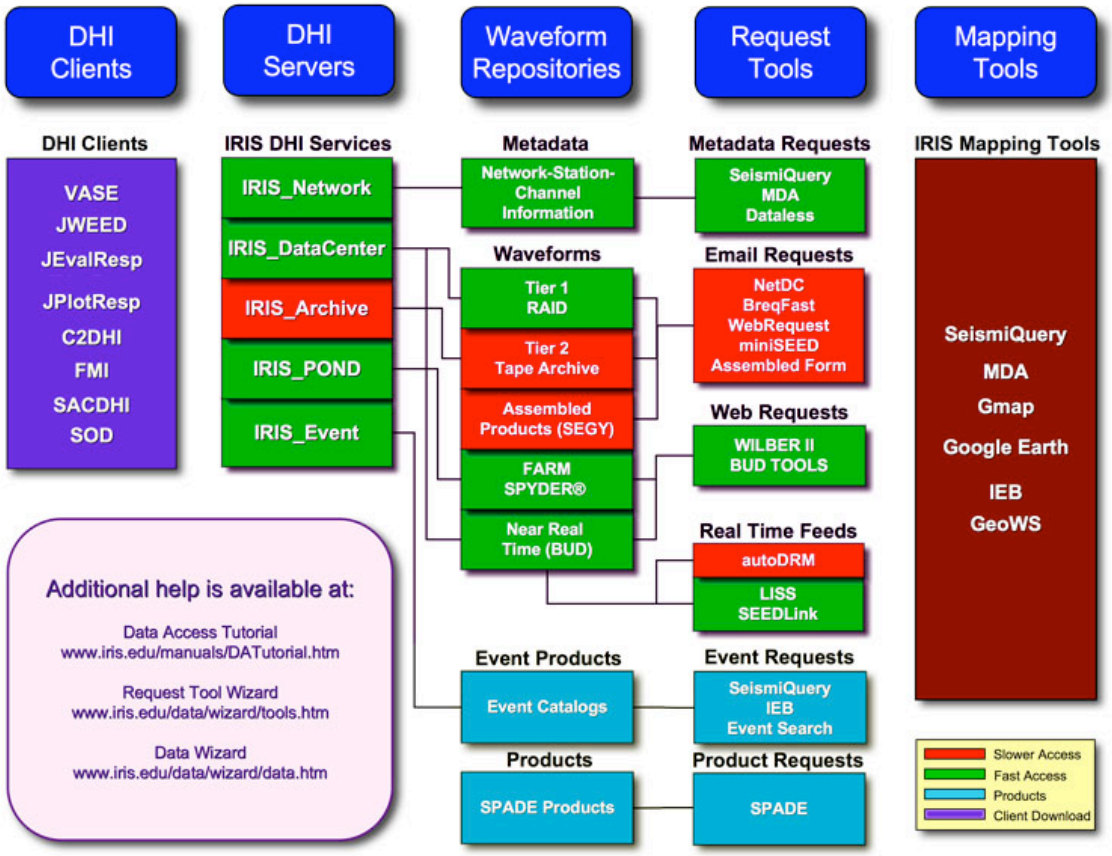
Assembled Data

PASSCAL data in SEED format is requested using the same tools used for other network data.

The DMC distributes several pre-assembled data sets in non-SEED format and from diverse programs. A data set is considered "assembled" if it is in a format other than SEED (except for FARM data) and is therefore not accessible using our standard SEED data request methods. Assembled data sets, for the most part, come "as-is." Some sets can be broken up into smaller sets but most come as a complete package. Sources of assembled data sets include:

- **PASSCAL**
  These products are made up of data collected from portable seismic instruments recording active source reflection, active source refraction or natural source recordings of earthquakes. Formats vary.
- **Other Sources**
  This category includes assembled sets that contain data or information collected from sources other than GSN or PASSCAL including several USGS data sets, NASA data, IDAC sets, nonproliferation experiments, gravimeter data, etc.

Data Request Tools



85% of *all* DMC data requests are made using the BREQ_FAST format.

The DMC offers several tools for making customized data requests. The tool you choose to use will depend on what type of request you wish to make and/or the complexity of your request. The diagram above categorizes the available tools by request type and interface.

The shapes in the diagram above represent the how and where of data storage. The request tools used for the different data types are listed to the side. The tools listed are linked to their corresponding manual or help page (for HTML tools). Below is a brief description of each tool. For complete information, please review the manuals.

Assembled Data Web Form
A Web form for requesting assembled data sets.

BREQ_FAST
An e-mail request format. Breq_fast is the most common format for making a request because no special tool is needed to format a request.

WILBER is the DMC's most popular *Web-based* request tool.

BUD Interface
A Web interface for searching and requesting data from the near real-time BUD system. Data is delivered in miniSEED format. BUD data can be accessed by the BUD Web interface, LISS client, or DHI client. One can itemize primary functions, view waveforms, check latency, continuity, request data, etc.

JWEED
JWeed is a map-based seismogram request tool. It enables the user to select events and stations and then download the data to their computer. JWeed is platform-independent and allows the user to access information from remote data centers, using the Data Handling Interface (DHI), and from local file sources that the user has generated and downloaded. Download a copy of JWeed >>

NetDC
Another e-mail based request format. Similar to a BREQ_FAST request in format, NetDC allows users to query or request data from more than one networked data center with only one e-mail. Request routing is handled internally by every participating network data center.

SeismiQuery
A Web interface into our Oracle database of data holdings including all timeseries information and all meta data, like responses, sensor locations, etc. Users are encouraged to search the database before making a request. SeismiQuery is also used to extract channel response files and can be used to make data requests (like BREQ_FAST).

VASE
VASE is a Java-based client application designed for viewing and extracting seismic waveforms from the DHI waveform repository via BUD. Download a copy of VASE >>

WebRequest
A Web form used to submit BREQ_FAST-style requests directly to the DMC. This tool does NOT query the database. There is a BREQ_FAST request generator within SeismiQuery that accesses the database before generating a request. This way, the user knows exactly what they will be getting when their request is fulfilled.

WILBER II
A Web interface for searching and requesting data from our SPYDER® and FARM archive. Data is delivered in SEED, miniSEED, SAC binary or SAC ASCII files.

Receiving the Data

SEED data

There is a SEED manual available in hard copy or Adobe Acrobat® format. The PDF version is the most up-to-date. Download the SEED Manual.

All customized data requests (i.e. continuous data from the archive) and some BUD and WILBER II requests will result in the user receiving a SEED data volume either on tape or via anonymous FTP. A SEED volume consists of a control header section, describing the seismic stations and instrumentation, and a waveform data section. These two parts of a SEED volume can live separately as a dataless SEED file (just the headers) and a miniSEED file (just the data).

Dataless SEED volumes can be downloaded (by network) via ftp from ftp://ftp.iris.washington.edu/pub/RESPONSES/DATALESS_SEEDS/ or by sending a BREQ_FAST request to DATALESS@iris.washington.edu

The DMC request processing system does not "slice" time series at the requested time window; instead, complete data blocks in which the start and end times fall are extracted from the archives and assembled into the SEED volume you receive. To slice the data to meet specific time window parameters you must do one of two things:

1] use JWEED to create a summary file to use as input to rdseed.
2] use rdseed and re-enter the desired window start and end times.

JWEED

If you used JWEED to generate your original data request, you can use the resulting .summary file to extract data from the SEED volume. The .summary file contains all the time parameters (and other parameters) you set when making your request. Use the .summary file in conjunction with rdseed (described below) to extract the time windowed data from your SEED volume. To use your .summary file with rdseed, select the "d" option (when prompted) and then name your .summary file.

rdseed

Rdseed (pronounced "read seed") is a UNIX tool that "reads" SEED volumes and converts data into one of several common analysis formats (SAC ASCII and binary, AH, CSS, miniSEED, and SEED). If you choose not to use JWEED to generate a .summary file, you will need to enter your time parameters when prompted (within rdseed) to extract your data (or get all the data). Rdseed is also used to extract response information, and to "suture" dataless SEED and miniSEED volumes together to make full SEED volumes. Because users can request to receive just miniSEED volumes, rdseed can be used in conjunction with a separate dataless SEED volume to process the data.

SEED volumes written to tape are written as individual files, with an End of File (EOF) marker between each volume. Although rdseed is typically used to read data from local disk, it is capable of extracting SEED directly from tape media. To access a specific volume on a tape, you can either specify the volume number in rdeed or use the UNIX "mt" command (check your system's man page) to fast-forward to the desired volume on tape before running rdseed.

Download a copy of rdseed >>

Jrdseed NEW - works with Windows

Announcing Jrdseed – a Java version of rdseed currently under development. This evaluation supports core rdseed functionality, including the reading and writing of SEED volumes, as well as SAC (binary) output. Run this application as a self-running jar: java -jar JrdseedVer0.06.jar. Java 1.4 or later required.
Download Jrdseed v0.06 >>

Other Tools used with SEED data:

evalresp

An IRIS developed tool, evalresp uses response files generated by rdseed to calculate the response of a time series to ground motion at a given frequency. You can download evalresp here: ftp://ftp.iris.washington.edu/pub/programs/sel/sun

JEvalResp [a DHI client]

Java version of evalresp that is platform independent

JPlotResp [a DHI client]

A graphically-oriented Java program for processing and plotting the response information (also is platform independent)

SEG-Y Data

If your request results in a volume of data in SEG-Y format, you may need to refer to the Digital Tape Standards manual for details on your data set. We cannot provide you with this manual but you can purchase one from:

Society of Exploration Geophysicists
P.O. Box 3098, Tulsa, Oklahoma 74101 USA

Some PASSCAL data sets are not in "true" SEG-Y format; they do not include complete reel identification headers and instead just have SEG-Y traces. This can cause problems if you try to view these files with some commercial analysis packages. The PASSCAL Software release includes tools that read, plot and gather PASSCAL SEG-Y data. For information about these tools, look here: http://www.passcal.nmt.edu/

Other data formats

Some of our older data sets are not in SEED or SEG-Y. They may be UNIX tar files of traces in SAC, AH, or other formats. (Also, WILBER II offers SAC as an output option.) At times these tar files are compressed using standard UNIX compress. The name of these files generally reflects the data format in the suffix (example: filename.AH.tar.Z).

This ends the tutorial. If you have any questions, please contact us:

E-mail: comments@iris.washington.edu