

The Status of the FDSN Archive for Continuous Data at the IRIS DMC

by Tim Ahern
Program Manager, IRIS Data Management System

Prepared May 17, 2002

FDSN Meeting
June 12, 2002
Waikoloa Beach, Hawaii

Summary

The IRIS DMC continues several activities within the FDSN. The more significant FDSN related activities during the past year include the following:

1. Archiving and distribution of FDSN data at the IRIS DMC
2. Participation in the NetDC effort to support distributed data centers. NetDC has now been installed at
 - IRIS DMC
 - ORFEUS
 - NCEDC
 - GEOSCOPE
 - GEOFON
 - *OHP/ Pacific 21*
 - *NCDSN Data Center*
 - *MEDNET*
 - *SCEDC*

Those centers listed in italicized type have installed NetDC but it is not yet available for general release.

3. Major effort in expanding the FARM system of data hi-grading. Data from most FDSN networks and stations are now being routinely placed in FARM and SPYDER event volumes.
4. Development of the SEED classes for Object-Oriented representation

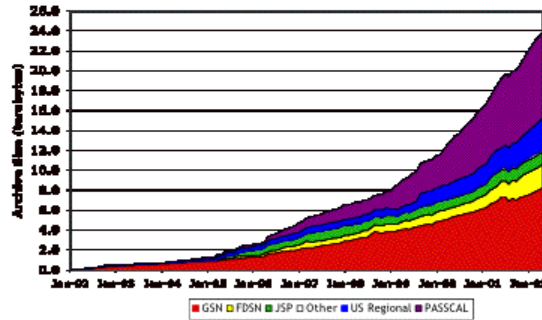
FDSN Data Holdings at the IRIS DMC as of May 16, 2002

Network	Network Code	Starting Date	Ending Date
CDSN	CD or IC	May 1986	April, 2002
CNSN	CN	April 1982	March, 2002
Czech Republic (DPC)	CZ	January, 1973	April, 2002
GEOFON	GE	September 1993	January, 2002
GEOSCOPE	G	January 1991	March, 2000
Graefenberg	GR	January 1990	January 2002
IRIS	IU, II, GT	January 1988	April, 2002
MEDNET	MN	January 1990	March 2001
Netherlands	NL	January 1994	May 2002
Pacific 21	PS	January 1990	April, 2002
USNSN	US / LB	June 1996	March 2002

This is very good progress in terms of decreasing the latency of archiving data from all FDSN networks at the FDSN Archive. Most FDSN networks have contributed data to the FDSN Archive during 2002. Additionally all but two of the FDSN data centers have data for the current year at the FDSN Archive. This is a very noticeable improvement.

The FDSN ARCHIVE AT THE IRIS DMC

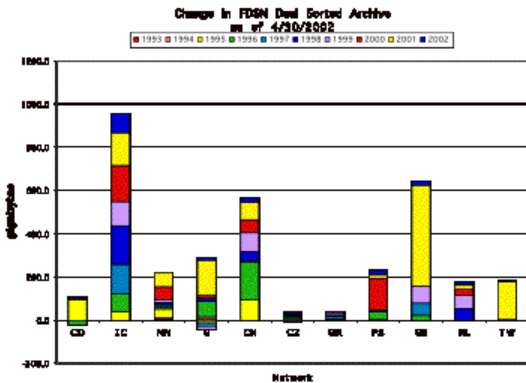
IRIS DMC Archive
(Dual Sort)
as of 4/27/02



The Amount of Data in the IRIS Archive.

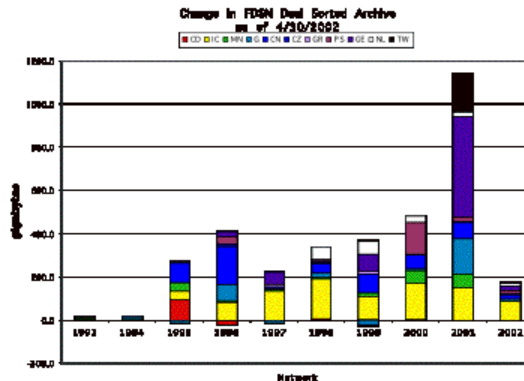
The bottom portion of the above figure represents IRIS GSN data, the next areas are FDSN data, and the data from JSP arrays, then US Regional Network data, and the top area is data from IRIS PASSCAL.

The IRIS archive now contains 23.9 terabytes of waveform data in 5.4 million files. These figures represent the dual sorted values where we store data once in a time sort and once in a station sort order. FDSN data in the FDSN archive consists of 2.3 terabytes of waveform data in 412,756 files, roughly 10% of the entire DMC archive. Since we have sufficient capacity in the mass storage system we store a redundant copy of the time sort and the station sort as well as creating one copy on DLT for out-of-area storage. Therefore the total amount of data we manage is 59.75 terabytes in 13.5 million files.



FDSN Data in the IRIS Archive. Of the 24 terabytes of data in the archive, there are roughly 2.3 terabytes (dual sort) of FDSN data archived. The above figure shows the annual change in the amount of data by FDSN Network code. During the past year GEOFON (GE), Taiwan BATS (TW) and Geoscope (G) were the largest contributors to the FDSN archive. Data that show negative values represent data that were removed from the archive during a given year for a variety of reasons

Annual Change to the FDSN Archive by Year. Considerable increases in the data from GEOFON, GEOSCOPE and Taiwan BATS are most notable. As this diagram shows, a very large increase to the FDSN archive occurred during the past year, much of it a result of the increased data from GEOFON.



SOFTWARE

The IRIS DMC has continued to develop and enhance software for use by the IRIS and FDSN communities. Most of this software is

available by anonymous ftp at

ftp iris.washington.edu cd pub/programs

or through the WWW at

http://www.iris.edu/NEW/HTM/dms_software.htm

RDSEED

The current version level is Version 4.18.1. Now supports LINUX. A new version of RDSEED will be available shortly to support the recently adopted ability to represent waveform quality in the data records (D/R/Q).

VERSEED

The current release is 2.12.2. No new release of verseed was made since the last FDSN meeting.

EVALRESP

The current release version of EVALRESP is 3.2.20.

JEvalRESP

A new version of EvalRESP in JAVA is being beta tested. We find that JEvalRESP executes faster than the C version of EVALRESP.

RELISH

The current release version is 3.2.6. No new release of RELISH has been made.

A replacement program for RELISH is in the final testing stages. JPlotRESP is totally written in Java and is compatible with JEvalRESP. Both JEvalRESP and JPlotRESP were written and are supported by ISTI.

POD

The latest release is 4.1.

WEED

The latest release is 2.9.1

PDCC Portable Data Collection Center

Work continues in the recoding of the PDCC in Java. It is being implemented to be FISSURES compatible. One of the significant benefits is that the SEED class definitions have been created and circulated.

NetDC

As mentioned earlier, NetDC is now operating at IRIS DMC, NCEDC, ORFEUS, Geoscope, and GEOFON. It is installed, but operational status is not known at Pacific-21, China, SCEDC, and MEDNET. We believe MEDNET will be releasing NetDC access very soon.

FISSURES/DHI

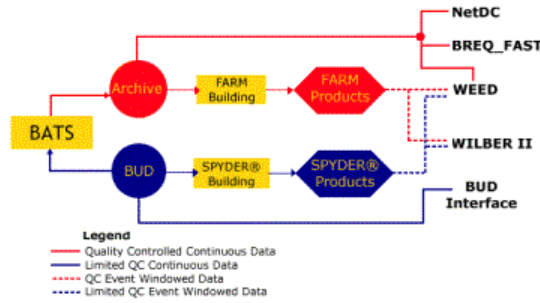
Several new Data Center Services and Clients are now available, although they have not been thoroughly tested. The Data Center Services include

- Event Service - hypocenters
- Network Service, Stations, their characteristics, and responses
- Waveform Service - access to data in BUD, Archive, FARM and SPYDER.

Additionally, several clients are in various stages of development. These include

- VSN Explorer
- DHI WEED
- DHI Standing Order for Data, an event based SOD
- DHI to MATLAB Interface
- VASE - a waveform Visualization and Seismogram Extraction Application
- DHI to JEvalRESP

Data Organization and Request Tools



As the above diagram shows, the DMC has simplified its request tools to a certain degree. We distribute SEED volumes from four different data repositories. These are

1. The continuous waveform archive in the near-line mass storage system or RAID.
2. The BUD real time buffer
3. The FARM on-line dataset or the
4. The SPYDER on-line dataset.

We support four primary tools (there are several other tools that no longer receive support but are available for use). These are NetDC, and Breq_Fast, WEED and WILBER II. We also have a variety of Web-based tools and real time tools that allow access to the BUD. We are currently developing a series of FISSURES/Data Handling Interface Tools that will provide additional programmatic access techniques to data and metadata. We also actively support the SeismiQuery Web-based interface into the Oracle database at the IRIS DMC.

We have also created a new method of creating customized dataless SEED volumes. If a user send an email to dataless@iris.washington.edu and the format of the request is a Breq_FAST file, then a dataless SEED volume for the information requested in the Breq_FAST formatted file will be returned. This is of significant value for users of the new FARM and SPYDER data products who choose not to use the WILBER II interface

Data Shipments

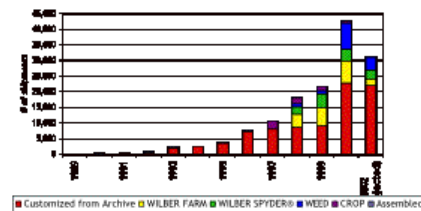
Ignoring the FARM shipments that we make by routine tape copy, the number of customized data shipments remains quite large. Most of the growth is in the area of requests that are serviced automatically from our on-line data collections. We have been attempting to limit the number of requests that must be serviced from our mass storage system. We have done this by placing the most often requested data on-line in the FARM and SPYDER archives. FARM products are now produced as soon as the CMT catalogue for a given month is available, typically 6 weeks behind real time. We have also put considerable effort into developing the WILBER II and WEED request tools that can directly access the on-line data sets.

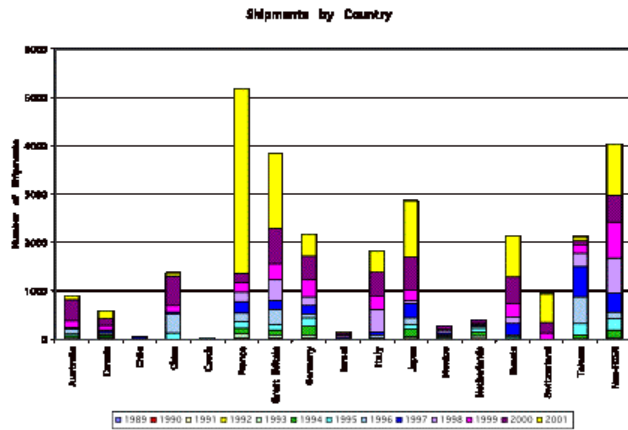
The following figure shows the growth in data shipments from the DMC over time. The difficult requests for the DMC to service are those shown at the bottom of each column, the so called "customized" requests. Most of the growth in customized data shipments now comes from the on-line data sets of the FARM and SPYDER.

Data Shipments Made by the IRIS DMC. The bottom sections are for customized requests using one of the previously indicated request tools, followed by WILBER accesses of the FARM, WILBER accesses of SPYDER, WEED requests and finally CROP.

The number of requests sent outside the United States continues to increase. The following figure summarizes the number of shipments of customized data that went outside the United States starting when the DMC was established in 1989.

Customized Data Shipments
(Excluding Physical FARM Shipments)





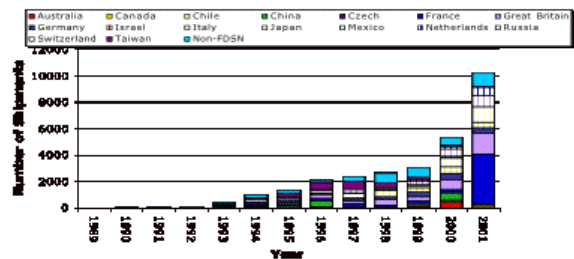
Customized Shipments Sent Outside the US by Country. This chart shows the number of customized requests processed by the IRIS DMC that were shipped outside the United States. In 2000 5,324 shipments were made outside the United States, while in the year 2001 there were 10,299. There was a very large increase in shipments to France this past year.

The greatest number of shipments in 2001 was made to France (3,798) followed by Great Britain (1,536) and Japan (1,172).

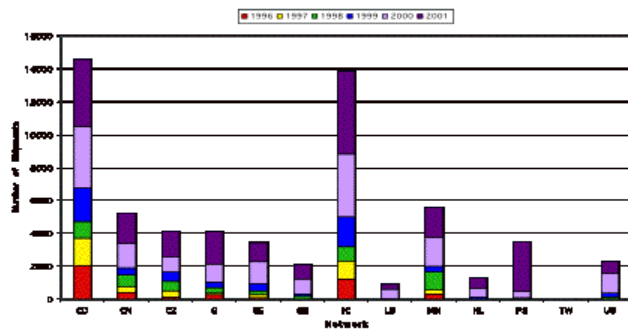
Number of Data Shipments Sent Outside the US by Year. We serviced 10,299 requests for data from outside the United States during 2001. This was nearly double the number in the previous year (5,324).

The amount of FDSN data requested by users of the IRIS DMC is displayed in the following chart. In general, there is an increase in FDSN data being shipped by the DMC in 2000 over previous years.

Non-US IRIS DMC Data Shipments



Shipments of FDSN Network Data



FDSN Data Shipped from the IRIS DMC. The number of requests for data from various FDSN networks is shown here for the years 1996 - 2001. Although dominated by the data from the two digital networks in China, several thousand requests from other networks are being serviced by the DMC as well. Data from several FDSN networks were sent to users more than 1000 times last year including PS (3,000), G (1,947), CN (1,857), MN (1,800), CZ (1,548), and GE (1,177).