

# SOUTHERN CALIFORNIA EARTHQUAKE DATA CENTER OPERATIONS

Katrin Hafner and Robert W. Clayton

Seismological Laboratory  
California Institute of Technology

## 1999 SCEC Annual Progress Report

January 28, 2000

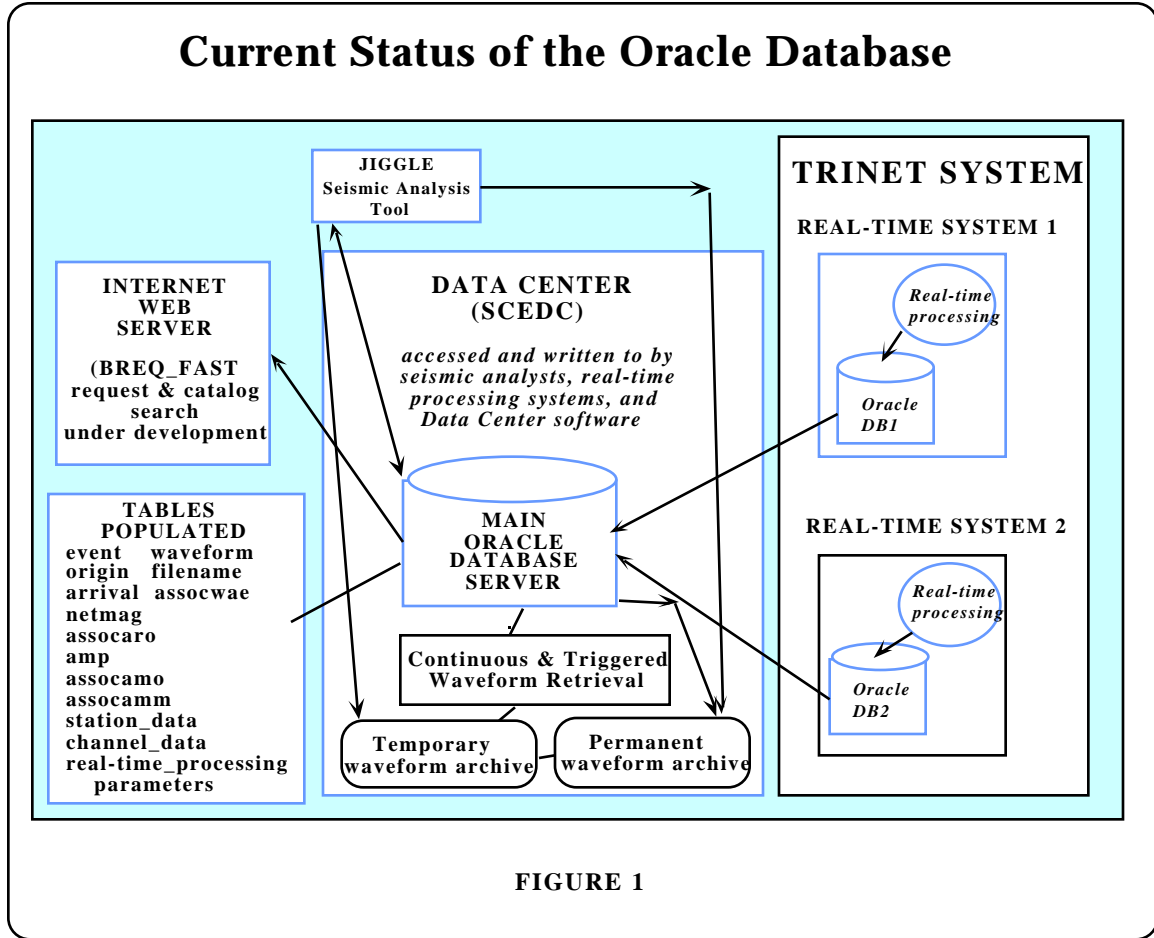
During 1999, activities at the SCEC Data Center (SCEDC) were focused primarily on moving the Data Center towards a new phase in its role as the primary on-line, and near "real time" archive of seismological data for Southern California. These efforts involved working closely with the TriNet development group at Caltech/USGS in implementing a system of ORACLE databases, which will archive all parametric earthquake data, as well as pointers to the waveforms on the mass storage system. In addition, the Data Center transferred all of the existing waveform data from the old 0.6 Tbyte SONY jukeboxes to the new 5 Tbyte DISC mass storage system. The SCEDC continued improving its ability to serve as an outreach facility to the SCEC community, the public, and educational and government agencies. A more detailed description of the Data Center's efforts this past year are outlined below:

**1) UPGRADE OF ORACLE DATABASE SYSTEMS:** The Data Center upgraded the existing ORACLE databases to version 8i from (version 8.0.4) to store seismic data generated by the prototype Trinet seismic system. The database system is designed so that the data center database receives data from the two real-time database systems in near "real-time" via one-way data replication. This allows for quick access to the earthquake data, as well as providing a means to continuously archive parametric earthquake data on the real-time systems should the data center be temporarily unavailable (Figure 1). Although the data is not officially available, the ORACLE database on the SCEDC has been receiving and archiving parametric data from the real-time system since October 1, 1999, as well as archiving the pointers to both the triggered and continuous TriNet waveform archives. The parametric and waveform archival database schemas used at the SCEDC have been developed in conjunction with the NCEDC, and have been implemented at both Data Centers, facilitating the exchange of data in the future.

**2) IMPLEMENTATION OF EVENT TRIGGERED WAVEFORM RETRIEVAL:** In conjunction with the implementation of the newest ORACLE database, the Data Center has implemented a waveform retrieval agent which reads waveform requests for events from the database, and retrieves the appropriate waveforms for the event from a waveform pool. The data are retrieved and archived in miniseed format. In addition, the retrieval agent then populates waveform tables in the ORACLE database with the pointers to the locations of these data in the archive, as well as their attributes. Waveform analysis programs developed for the seismic analysts by the USGS to retrieve and plot waveforms, as well as to re-analyze events use the information in these database tables. These same tables will be used, as we develop the tools needed for users of the SCEDC to retrieve these waveforms.

**3) IMPLEMENTATION OF CONTINUOUS WAVEFORM RETRIEVAL:** As of October of 1999, the SCEDC has begun retrieving and archiving 20 sps or less continuous waveform data channels from the TriNet system. A program has been developed which combines hourly channel files into daily continuous files, which make up the official archives on the Jukebox

system. A BREQ\_FAST type system of data retrieval is under development to retrieve these continuous data in complete SEED volumes ([www.scecdc.scec.org/broadband-retrieve.html](http://www.scecdc.scec.org/broadband-retrieve.html)). We were fortunate to have the new system running during the Oct 16, 1999 Hector Mine earthquake, and were able to continuously archive the entire network for the 22 hours preceding the event to 14 hours following it.



**4) DEVELOPMENT OF A STATION DATABASE IN THE ORACLE DATABASE:** In order to provide the best possible instrument response information, as well as up to date station information to the real-time systems, data analysts, and data center users, the existing USGS DBASE-IV station database is being converted to station database schemas developed by the NCEDC, and implemented by the SCEDC. Currently, station/channel tables, containing information about locations, on-off dates, sample rates, etc. have been implemented in the ORACLE database on the data center. The data center will host the "authoritative" version of the station history database, and this database is being used by the TriNet real-time and event analysis systems for their processing now.

**5) COMPLETED DATA TRANSFER OF WAVEFORM ARCHIVE TO NEW MASS STORAGE SYSTEM:** The installation of the new 5-Terrabyte DISC jukebox system was completed at the beginning of December 1998. New SCSN and GPS earthquake data has been archived to the new Jukebox system since January 1999. All of the historic earthquake and GPS data, previously stored on the 0.6 Tbyte SONY mass-storage systems, was transferred to the new mass storage system by the end of summer 1999. After the data was transferred, each of the

over 600,000 files was checked to verify that the copy of the data was complete. Users of the Data Center have been retrieving data from the new mass storage system since the fall of 1999. Outstanding issues of slow NFS transfers are currently being addressed by the Jukebox software provider.

**6) UPGRADE OF THE DATA CENTER WEB SERVER:** During the fall of 1999, the data center web server was upgraded from a Sparc II machine with 128 Mbytes of Memory, to an Ultra 60, with 500 Mbytes of memory. The web pages for [www.scecdc.scec.org](http://www.scecdc.scec.org) were reorganized, redesigned and updated. In addition to upgrading the hardware for the webserver, and redesigning the web pages, we implemented a system designed to handle the expected loads of 300 hits/sec immediately after an earthquake. This involved installing a reverse proxy web server called "SQUID. In this system, the SQUID server is a front end to the standard Apache Web Server which runs the httpd daemon. Squid takes in all of the incoming web requests and forwards them to the Apache web server. It caches the data returned to it, so that it can use it for future requests. Because it caches data both in memory and on disk, and it can serve requests much faster than the conventional Apache web server alone. We have seen a three-fold increase in the speed of our web server because of this modification to our web server design, and we anticipate that the new webserver will be able handle requests as needed.

**7) SYNTHETIC APERTURE RADAR (SAR) DATA:** As part of the WInSAR consortium, we have archived over 220 SAR scenes on the mass-storage system for a total online data volume of 40 Gbytes. These data are available to authorized WInSAR users on a password protected basis. We have developed a simple Web interface to show what data is available and to initiate FTP transfers of the data. The Web address is <http://www.winsar.scec.org>. This facility is being operated in cooperation with similar operations at UCSD and UCB.

**8) PROCESSING OF BACKLOGGED DATA:** The Data Center has continued its efforts to eliminate the backlog of earthquake data. Data from January 1979 through February 1981 were retrieved from 9-track archive tapes, and converted to CUSP format mem and gram files, which can now be processed by the seismic analysts. Backlogged data from March and April of 1983 were also processed by the analysts. A small backlog of events from the Hector Mine Earthquake sequence should be completed by the end of February 2000.

### **SUPPORT TO THE DATA CENTER ACTIVITIES FROM OTHER SOURCES**

The operations of the Data Center benefited from support beyond that provided by SCEC. An award from NSF of \$228K (matched by a \$114K grant from SCEC) allowed the purchase of the new mass-storage system. TriNet contributed both equipment and support for an Oracle database administrator. The major equipment items of this support are:

|                            |       |
|----------------------------|-------|
| SUN Enterprise 450 Server  | \$30K |
| 300 Gbyte RAID disk system | \$15K |

In addition, the Data Center depends on the staff of Southern California Seismic Network to provide the bulk of the parametric data that is archived, and on the TriNet and Seismo Lab computer staff to keep the computer and network infrastructure going.

## **PUBLICATIONS AND CONFERENCES PARTICIPATED IN**

Hafner, K. and Clayton, R.W., Recent Developments at the Southern California Earthquake Data Center, (Abstract), March 1999 SSA meeting, Seattle, Washington

Hafner, K. and Clayton, R.W., The Southern California Earthquake Center Data Center (SCEC\_DC), (abstract), 1999 IUGG meeting, Birmingham, England

Hafner, K. and Clayton, R.W., Recent Changes at the Southern California Earthquake Data Center (SCEDC), (abstract), Fall 1999 Annual SCEC meeting)

Hafner, K. and Clayton, R.W., Recent Changes at the Southern California Earthquake Data Center (SCEDC), (abstract), Fall 1999 AGU meeting)