2005 SCEC Annual Report

Southern California Earthquake Data Center (SCEDC) 2005 Accomplishments

Principal Investigators:
Robert W. Clayton and Vikki L. Appel

Institution:
California Institute of Technology
Seismological Laboratory
Pasadena, CA 91125

Proposal Category:
Data Gathering and Products
Disciplinary Committee:
Seismology

Start Date:
February 1, 2005

2005 SCEDC Annual Report

The following are the Southern California Earthquake Data Center's major accomplishments for 2005:

1. Continued our key data-acquisition and archiving functions by maintaining and updating the primary online, near real-time searchable archive of seismological data for southern California. Added 87,600 station-days of continuous data and parametric and waveform data for 16,765 local events and 810 teleseismic earthquakes.

2. In 2005, the SCEDC undertook a significant project to develop an improved Station Information System (SIS) for the Southern California Seismic Network (SCSN). The goal of this project was to develop a simplified database-driven system that interacted with a single database source to enter, update and retrieve station metadata easily and efficiently.

    Problem Statement: *The current methods of keeping track of all station information at the SCMC are not providing the functionality required to meet the needs of all of the users who need complete, accurate station information from a single source.*

    The scope of the system was to develop and implement a simplified metadata information system with the following capabilities:

    *   Provide accurate station/channel information for active stations to the SCSN real-time processing system.
    *   Provide accurate station/channel information for active and historic stations that have parametric data at the SCEDC e.g., for users retrieving data via STP from the SCEDC.
    *   Provide all necessary information to generate dataless SEED volumes for active and historic stations that have data at the SCEDC.
    *   Provide all necessary information to generate COSMOS V0 metadata information.
    *   Be updated through a graphical interface that is designed to minimize editing mistakes.
    *   Allow stations to be added to the system with a minimum, but incomplete set of information using predefined defaults that can be easily updated as more information becomes available. This aspect of the system becomes increasingly important with historic data when some aspects of the metadata are simply not known
    *   Facilitate statewide metadata exchange for both real-time processing and provide a common approach to CISN historic station metadata.

    The development was guided by two fundamental design requirements:
    *   A well-designed database that is the single source of information, i.e., the authoritative information is what is contained in the database, not in flat-files or any other databases.
    *   Applications that interact with the data will interact directly with the database, i.e., they will not check information out of the database, manipulate it in another environment, and repopulate the database with the modified information.

The SIS became the production station database system for the SCEDC on November 1, 2005 and all station updates to data are done through SIS GUI (graphical user interface) in parallel with the legacy database because the SCSN Real-Time system has not modified their code to use the SIS database yet. However, to accommodate this lag, views have been created against the SIS schema, which are prefixed with "sis_" and are the same in name, structure and attribute names as the legacy station database's tables, so the only change to the RT system code that is required is to prefix their current table calls with "sis_".

The DC staff spent a tremendous amount of time working with the SCSN/USGS field staff to determine their needs, test the system with them, and work through scenarios to determine if any modifications were necessary to make the system work better for users and to familiarize the users with all of the panels on the interface.

The SIS can produce dataless SEED volumes for broadband stations and for USGS short period telemetered stations, a functionality that did not exist in the previous systems. We are currently putting effort into finalizing production of V0 format metadata from the SIS.

The Data Center developed stored procedures for the SIS as an API between applications and data model. By using stored procedures, changes can be made to the data model without requiring changes to the applications that access the database. Stored procedures include writing SEED volumes, inserting logger, sensor information into the database and modifying existing instruments.

The station-hardware information contained in the SIS can be accessed via the Internet via a PHP page at: http://www.data.scec.org/stations/views/sta_hardware.php. More information about the SIS is available from the project's homepage at: http://www.data.scec.org/stations/SIS/, which includes project documentation and progress reports.

3. The SCEDC developed an XML schema for distributing earthquake information. The eXtensible Markup Language has become the standard format for transferring data between computers and organizations because it is a clearly-defined way to structure, describe, and exchange data. The rapid proliferation of XML and the large number of XML parsers that can be freely integrated into applications has made XML a good format option for the Data Center to distribute seismic data and we put effort into XML distribution of seismic data because we anticipate tremendous benefits associated with this method of flexible data exchange between distributed systems.

    The SCEDC currently provides event and phase data in XML format from the catalog search pages on the Internet, and via the STP client. The schema and documentation are available from: http://www.data.scec.org/xml/.

4. The SCEDC, in cooperation with the Caltech Seismological Lab, recently completed a project to scan pre-digital analog recordings of major earthquakes recorded in Southern California. We have scanned records for M>3.5 earthquakes between 1962 and 1992 and other significant teleseisms. These scans are now available for download through our new search page at http://www.data.scec.org/research/scans/. Search features include the ability to search by date, station, instrument, and orientation; the option to sort search results by date, and the option to download multiple files as a single zipped archive.

5. The SCEDC is currently archiving Moment Magnitudes and Moment Tensor Solutions (MTS) produced by the SCSN in real-time and post-processing solutions for events spanning back to 1999. These solutions are in the SCEDC searchable database and are available for distribution from the consolidated catalog search page (Moment Tensors tab) at: http://www.data.scec.org/catalog_search/CMTsearch.php.

   The automatic MTS runs on all local events with Ml>3.0, and all regional events with Ml>=3.5 identified by the SCSN real-time system. The solution is emailed to SCSN personnel about 10 minutes after an event. If the quality of waveform fits is good enough, and the event is within the SCSN reporting region, it is automatically distributed to the outside world. The distributed solution automatically creates links from all USGS Simpson Maps to a text e-mail summary solution, creates a .gif image of the solution, and updates the moment tensor database tables at the SCEDC. The solution can also be modified using an interactive web interface, and re-distributed. The SCSN Moment Tensor Real Time Solution is based on the method developed at UC Berkeley by Doug Dreger.

6. **Catalogs**:
- The Data Center modified the catalog search page to a new URL:
  http://www.data.scec.org/catalog_search/index_date_mag_loc.php
  The new page has a new look where most of the text is below the search form and also moved the Caltech and USGS logos to be more visible on the page. There are five tabs that will allow users to search by:
    1. Location, magnitude and time (default)
    2. Event ID (can be multiple comma-separated events)
    3. 4-point polygon
    4. Radius
    5. Multiple magnitude types for single events
- We added an option to the catalog search page to "Download to File" to accommodate XML output, as well as other catalog output formats.
- Posted Hardebeck's focal mechanisms from P-wave polarity and S/P amplitude ratios catalog at http://www.data.scec.org/research/socal_focal_JLH.html. The dataset includes more than 24,000 focal mechanisms for earthquakes occurring 1984 through 2003.
- Updated the alternate location catalog with new SHLK_1.01 and SHLK_IMP_1.0 data and updated http://www.data.scec.org/ftp/catalogs/SHLK/. When users query for all information from all catalogs from STP, they will get up to five data sources:
STP> altloc -e 9875225
  9875225 2002/12/31,21:59:57.670 33.1812 -115.6160 3.15 2.18 l SHLK_IMP2005
  9875225 2002/12/31,21:59:58.030 33.1753 -115.6120 3.65 2.18 l HAUK2003
  9875225 2002/12/31,21:59:58.540 33.1803 -115.6150 3.02 2.18 l SHLK2005
  9875225 2002/12/31,21:59:58.540 33.1960 -115.6208 1.77 2.18 l HAUK2004
  9875225 2002/12/31,21:59:58.920 33.1862 -115.6110 1.63 2.18 l SCSN
# Number of events= 5

7. The Data Center organized a town-hall meeting of the SCEDC users at the 2005 SCEC Annual Meeting in Palm Springs, CA to gather feedback and identify the needs of the SCEC research community. User recommendations addressed SCEDC priorities including science, data-management, operations, data-products and funding.

Contribution to the SCEC Community

The Data Center is a central resource of SCEC and continues to be an integral part of the Center. In 2005, the SCEDC continued to contribute to the SCEC scientific community by providing online access to a stable and permanent archive of seismic waveforms and earthquake parametric data. The seismological data archive held at the SCEDC has contributed significantly to the publication of many scientific papers pertinent to the region, most of which have SCEC publication numbers. The Caltech/USGS catalog archived by the SCEDC is the most complete archive of seismic data for any region in the United States.

The SCEDC has allowed the data to be distributed to a much broader community of scientists, engineers, technologists, and educators than was previously feasible. The electronic distribution of data allows researchers in the world-wide scientific community to analyze the seismic data collected and archived in southern California and contribute their results to the SCEC community.

Archive Size and Contents
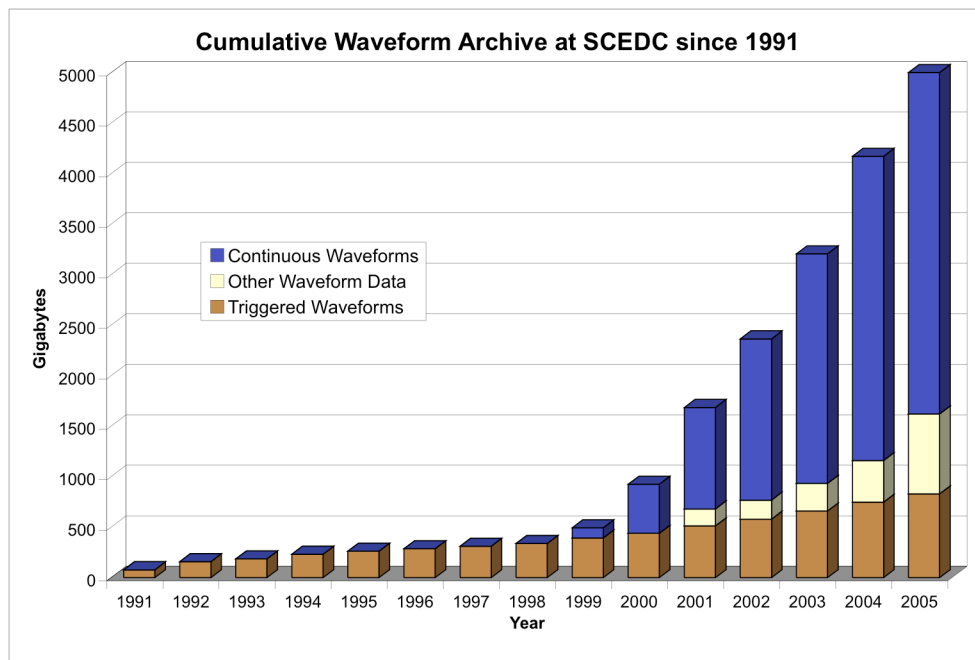
The archive at the SCEDC currently has the following holdings:
- The Caltech/USGS catalog of over 521,500 earthquakes spanning 1932-present.
- 3.65 terabytes of continuous and triggered waveforms.
- 11.3 million phase picks.
- 51.7 million triggered waveform segments.
- Nearly 6 years of continuous broadband recording of representing more than 250,000 station-day records, accumulating at 50,000 station-days per year (for the current 155-station network).
- 20.7 million amplitudes available for electronic distribution.
- Triggered data for more than 19,100 significant teleseismic events.

Number of Events at the SCEDC: 1932 - 1976
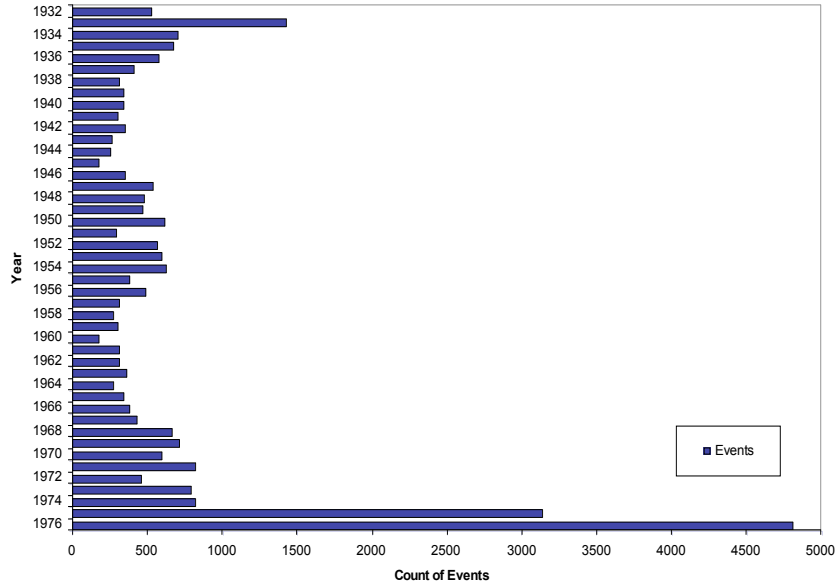
YEAR	# EVENTS	YEAR	# EVENTS
1932	520	1955	374
1933	1428	1956	485
1934	699	1957	313
1935	668	1958	266
1936	574	1959	300
1937	404	1960	171
1938	311	1961	308
1939	341	1962	312
1940	336	1963	363
1941	303	1964	268
1942	344	1965	339
1943	257	1966	377
1944	254	1967	430
1945	170	1968	663
1946	347	1969	707
1947	533	1970	597
1948	478	1971	822
1949	469	1972	457
1950	615	1973	787
1951	291	1974	821
1952	561	1975	3134
1953	590	1976	4801
1954	626		

Event parametric data for this era is available electronically

Number of Events and Waveforms at the SCEDC: 1981 - present
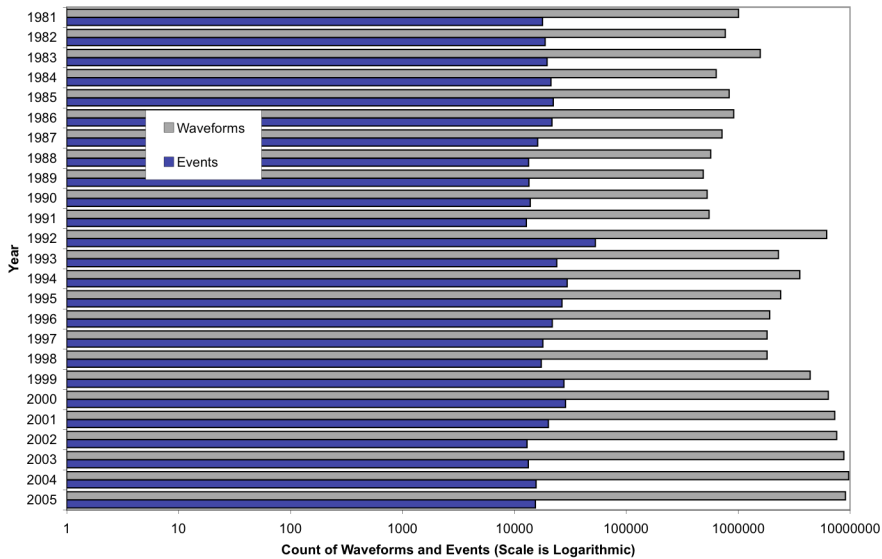
YEAR	# EVENTS	# WFs
1981	17740	1005322
1982	18717	765445
1983	19453	1563899
1984	21078	633208
1985	22196	825298
1986	21496	908084
1987	16037	714975
1988	13350	564993
1989	13400	485335
1990	13755	525263
1991	12723	547003
1992	52643	6136719
1993	23834	2269606
1994	29586	3532232
1995	26530	2377957
1996	21661	1904673
1997	17950	1799151
1998	17316	1797010
1999	27639	4379859
2000	28599	6343495
2001	20101	7274854
2002	12920	7561069
2003	13265	8730187
2004	15594	9701721
2005	15375	9045634

All event parametric and waveform data for this era are available electronically.

Data Volume Served and Archived

Number of earthquakes in the 1932-present Caltech/USGS catalog   623,872 earthquakes
Total size of the waveform archive: 6,893 GB
Size of SCEDC parametric and waveform database: 239,552,775 rows

Data transferred via STP in 2004-2005:

2004 Q4: October 1-December 31:
- 5,461,804 waveforms = average of 59,367 waveforms daily = 41.2 waveforms per minute!
- 746 gigabytes of waveform data = average of 8,177 megabytes daily = 94 kilobytes per second.

2005 Q1: January 1-March 31:
- 17,325,347 waveforms = average of 191,667 waveforms daily = 2.22 waveforms per second!
- 532 gigabytes of waveform data = average of 6,051 megabytes bytes daily = 70 kilobytes per second

2005 Q2: April 1-June 30:
- 3,156,771 waveforms = average of 34,690 waveforms daily
- 465 gigabytes of waveform data = average of 5,111 megabytes daily = 59 kilobytes per second.

2005 Q3: July 1-September 30:
- 1,671,318 waveforms = average of 18,570 waveforms daily
- 271 gigabytes of waveform data = average of 3,018 megabytes daily = 35 kilobytes per second.

From October 1, 2004 – Sept 30, 2005, the SCEDC archived:

- 20,423 events
- 4,692,692 waveforms
- 370,138 arrivals
- 1,329,124 amplitudes

Magnitude	Number of local events (le):
0-1	5,206
1-2	9,403
2-3	1,902
3-4	214
4-5	33
5-6	7

# events:	Event type
16,765	le (local event)
809	qb (quarry blast)
2,452	re (regional event)
274	sn (sonic blast)
810	ts (teleseism)
21,110	Total