FINAL TECHNICAL REPORT
September 25, 2003 - January 31, 2007

Operation of the Northern California Earthquake Data Center

AWARD NUMBER: 03HQAG0171

Barbara Romanowicz, P.I., and Doug Neuhauser
Berkeley Seismological Laboratory, UC Berkeley, CA 94720-4760
Telephone: (510) 643-5690, FAX: (510) 643-5811, barbara@seismo.berkeley.edu
Telephone: (510) 642-0931, FAX: (510) 643-5811, doug@seismo.berkeley.edu

NEHRP Elements: I & II

Research supported by the U.S. Geological Survey (USGS), Department of Interior, under USGS award 03HQAG0171. The views and conclusions contained in this document are those of the authors and shold not be interpreted as necessarily the official policies, either expressed or implied, of the U.S. Government.

FINAL TECHNICAL REPORT
September 25, 2003 - January 31, 2007

Operation of the Northern California Earthquake Data Center

AWARD NUMBER: 03HQAG0171

Barbara Romanowicz, P.I., and Doug Neuhauser
Berkeley Seismological Laboratory, UC Berkeley, CA 94720-4760
Telephone: (510) 643-5690, FAX: (510) 643-5811, barbara@seismo.berkeley.edu
Telephone: (510) 642-0931, FAX: (510) 643-5811, doug@seismo.berkeley.edu

NEHRP Elements: I & II

Keywords: Seismology; Geodesy; Database

TECHNICAL ABSTRACT

The Northern California Earthquake Data Center, a joint project of the Berkeley Seismological Laboratory (BSL) and the U.S. Geological Survey at Menlo Park, is an online archive for various types of digital data relating to earthquakes in central and northern California. The NCEDC is located at the Berkeley Seismological Laboratory, and has been operational since 1992.

The primary goal of the NCEDC is to provide a stable and permanent archival and distribution center of digital geophysical data for networks in northern and central California. These data include seismic waveforms, electromagnetic data, GPS data, strain, creep, and earthquake parameters. The seismic data comes principally from the Berkeley Digital Seismic Network (BDSN), the Northern California Seismic Network (NCSN) operated by the USGS, the Berkeley High Resolution Seismic Network (HRSN) at Parkfield, the EarthScope USArray TransportableArray stations in northern California, the various Geysers networks, and selected stations from adjacent networks such as the University of Reno, Nevada network and the Southern California Seismic Network (SCSN). GPS data are primarily from the Bay Area Regional Deformation (BARD) GPS network and the USGS/Menlo Park GPS surveys. The collection of NCSN digital waveforms dates from 1984 to the present, the BDSN digital waveforms date from 1987 to the present, and the BARD GPS data date from 1993 to the present. The BDSN includes stations that form the specialized Northern Hayward Fault Network (NHFN) and the MiniPBO (MPBO) borehole seismic and strain stations in the SF Bay Region. Additional related NCEDC datasets include EarthScope seismic data from the San Andreas Fault Observatory at Depth (SAFOD) and strain data from the Plate Boundary Observatory (PBO). The EarthScope projects provide additional funding to the NCEDC to archive and distribute these data.

The NCEDC also provides support for earthquake processing and archiving activities of the Northern California Earthquake Management Center (NCEMC), a component of the California Integrated Seismic Network (CISN). The CISN is the California regional organization of the Advanced National Seismic Network (ANSS).

Data are made freely available to all users over the Internet through a variety of web interfaces, email requests methods, and client/server interfaces. Many of these data distribution methods are standardized and supported by multiple data centers.

NEW ACTIVITIES

By its nature, data archiving and data distribution are ongoing activities. During 2003-2006, the NCEDC continued to expand its data holdings and enhance access to the data. Projects and activities of particular note include:

- Upgraded NCEDC computer hardware and software to support the 5x increase in waveform data archived at the NCEDC.
- Entered station hardware information and SEED instrument responses for all NCSN operated stations into the NCEDC database.
- Converted NCSN CUSP waveforms from their original CUSP format to standard MiniSEED using standard SEED channel names.
- Associated NCSN event waveforms with NCSN eventids in the NCEDC database to allow SEED data retrieval by eventid.
- Began continuous archiving of the NCSN broadband stations, and later the entire NCSN network, using standard SEED channel names. This includes data from all USGS NC stations, northern California USGS NSMP continuous telemetry stations, and continuous data from contributing networks.
- Developed and implemented FISSURES/DHI servers as a data distribution service for NCEDC data.
- Enhanced and installed SCSN's STP server as a data distrbution service for the NCEDC.
- Developed and implemented the NCEDC DART (Data Available in Real Time), a system to provide access to real-time timeseries data.
- Acquired hardware and developed software and procedures to read and archive continuous NCSN seismograms from tapes for 2001-2005.
- Expanded datasets to include EarthScope seismic data from northern California USArray stations, PBO strain data, and SAFOD seismic data.
- Supported NCEMC database efforts to supply the NCEDC with real-time earthquake parameters.
- Relocated the NCEDC computer and storage servers to a new state-of-the-art co-location data center in a new seismically braced building on the Berkeley campus.

These activities and projects are described in more detail below.

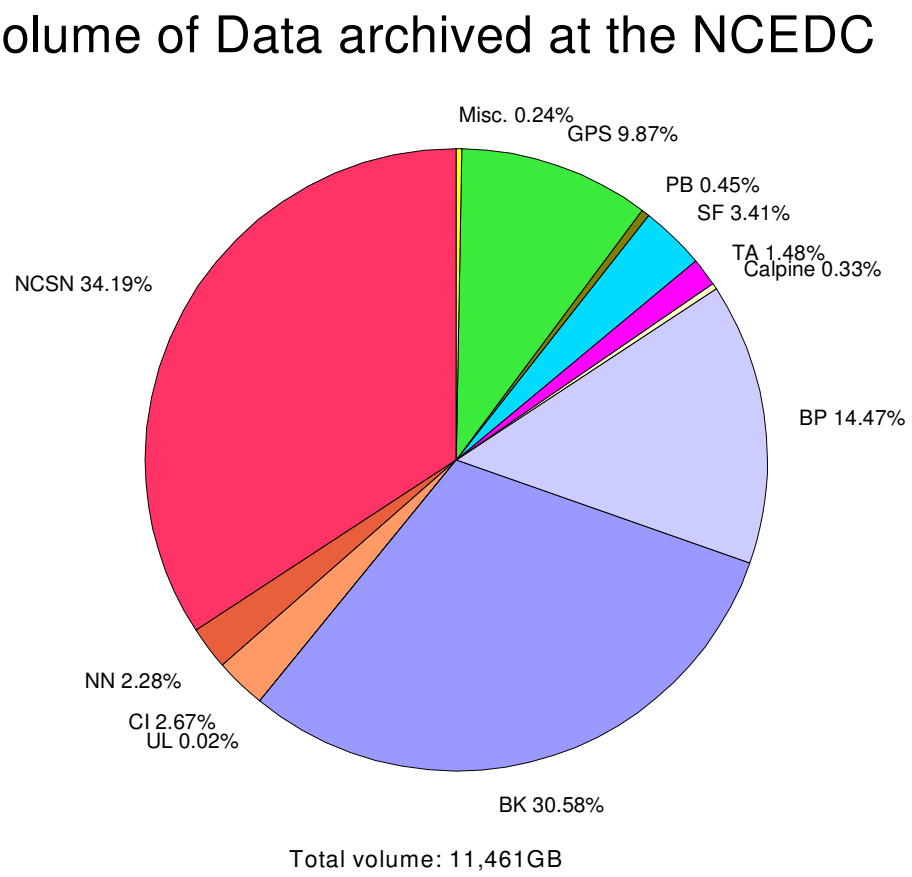

Figure 1: Chart showing the relative proportion of each data set at the NCEDC.

Data Type	GBytes
BDSN/NHFN/MPBO (broadband, electric and magnetic field, strain) waveforms	3,505
NCSN seismograms	3,918
Parkfield HRSN seismograms	1,659
BARD GPS (RINEX and raw data)	1,132
UNR Nevada seismograms	260
SCSN seismograms	306
Calpine/Unocal Geysers region seismograms	38
EarthScope SAFOD seismograms	391
EarthScope USArray seismograms	170
EarthScope PBO strain waveforms	52
USGS Low frequency geophysical waveforms	2
Misc data	28
Total size of archived data	11,461

Table 1: Volume of Data Archived at the NCEDC by network

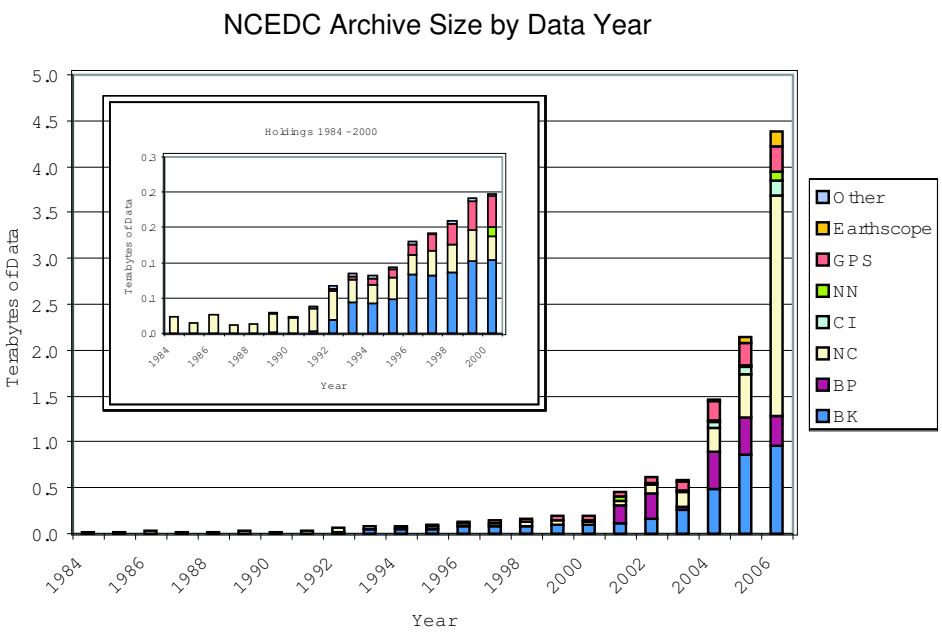

Figure 2: Chart showing the amount of data per year archived at the NCEDC.

NCEDC ACTIVITIES AND STATUS

Datasets

The bulk of the data at the NCEDC consists of waveform and GPS data from northern California. Figure 1 shows the relative proportion of each data set at the NCEDC. The total size of the datasets archived at the NCEDC is shown in Table 1.

Figure 2 shows the NCEDC archive size for each data year. During the 3.5 years of this award, the total data archive has grown from 2.5 TB to 11.4 TB with no significant increase in budget.

Figure 3 shows the geographic distribution of data archived by the NCEDC.

BDSN/NHFN/MPBO Seismic Data

Archiving current BDSN, NHFN, and Mini-PBO (all stations using the network code BK) seismic data is an ongoing task. These data are telemetered from 47 seismic data loggers in real-time to the BSL, where they are written to disk files, used for CISN real-time earthquake processing, and delivered in real-time to the DART (Data Available in Real Time) system on the NCEDC, where they immediately available to anyone on the internet. Since the inception of the NCEDC, data channels sampled at 20 Hz and below have been archived continuously, and higher frequency data were archived only for events. In September 2004, the NCEDC began archiving continuous high frequency data (80 Hz and 100 Hz) from all of the BDSN broadband, strong motion, and strainmeter sensors. In early 2006, the NCEDC began receiving all of the BK stations in real-time and making them available to users through the DART. All timeseries data from the Berkeley networks continue to be processed and archived by an NCEDC analyst using *calqc* in order to archive the highest quality and complete data stream at the NCEDC.

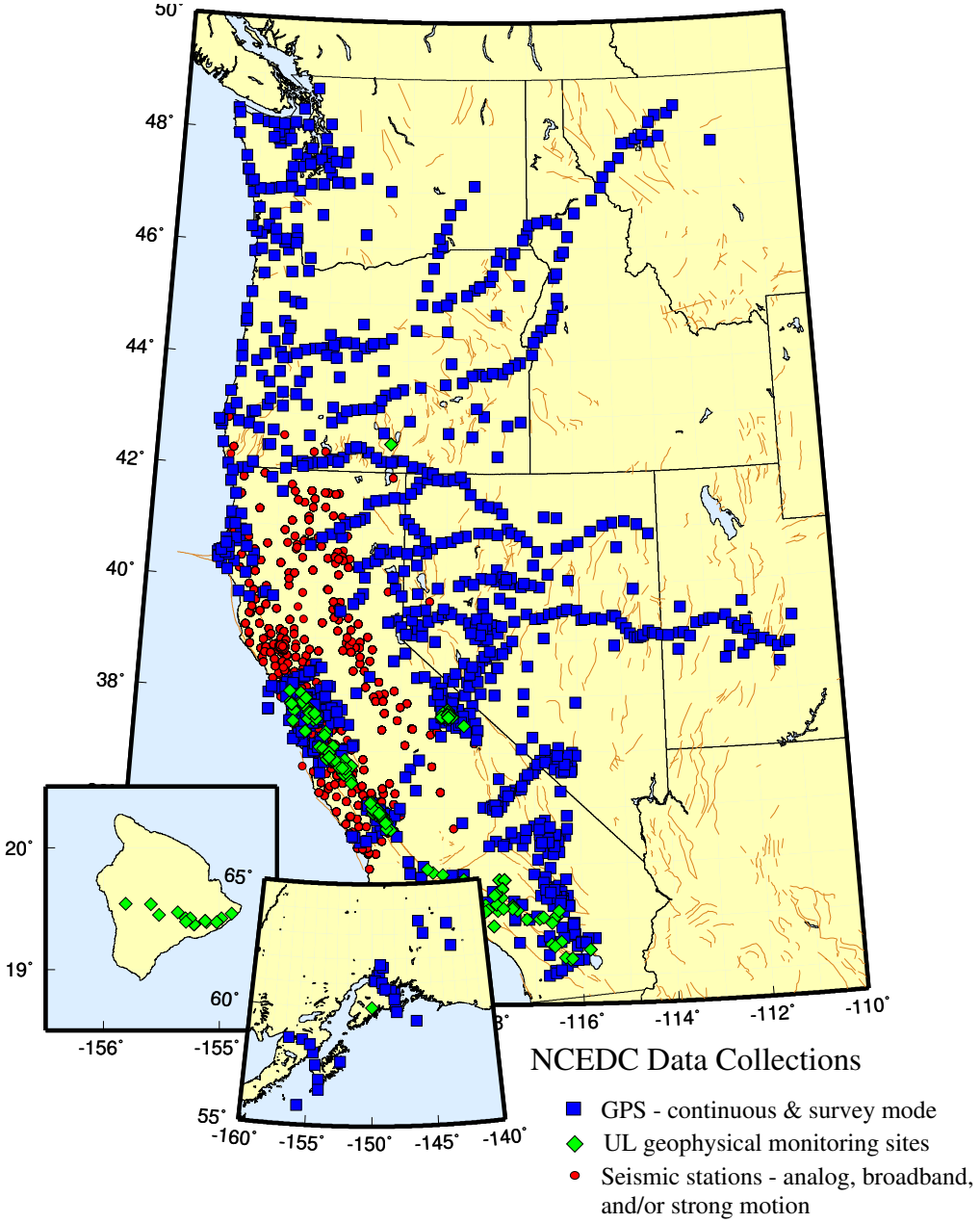

Figure 3: Map showing the location of stations whose data are archived at the NCEDC. Circles are seismic sites; squares are GPS sites, and diamonds are the locations of USGS Low-frequency experiments.

NCSN Seismic Data

NCSN continuous and event waveform data are sent to the NCEDC via the internet and/or private IP network. The NCSN event waveform files are currently assembled and analyzed at Menlo Park, and are then delivered to the NCEDC, where they are automatically converted to MiniSEED and archived.

The NCEDC maintains a list of teleseismic events recorded by the NCSN, which is updated automatically whenever a new NCSN event file is received at the NCEDC, since these events do not appear in the NCSN catalog.

Since 2002 the NCEDC has archived continuous data from the 15 continuously telemetered digital NCSN broadband stations: 11 stations in northwest California and southwest Oregon in support of the USGS/NOAA Consolidated Reporting of EarthquakeS and Tsunamis (CREST) system, two digital broadband stations in the Mammoth region, and two digital broadband stations in the Parkfield region. At the USGS's request, we also continuously archived the three component 500 Hz data from the Mammoth Deep Hole.

In January 2005, in response to interest in non-volcanic tremors detected in northern and central California, the NCEDC began archiving continuous high frequency data from 21 additional NCSN stations in selected regions of northern California and Parkfield. In response to requests for additional continuous NCSN data, we received approval to rebudget funds to purchase disk and tape systems to support the reading and archiving of continuous waveforms from the entire NCSN from 2001 to the present and to establish procedures continuous archiving of current NCSN data. We purchased the require hardware, and developed procedures to read, convert, and archive the continuous data from the NCSN tapes.

In December 2005, the NCEDC began archiving all available continuous data from the NCSN continuously telemetered stations. We initially started with the stations owned and operated by the USGS/MP NCSN (network code NC), and in early 2006, and after discussions with the other cooperating networks that supply data to the NCSN, we expanded the continuous archive to include data from all stations that are contributed to the NCSN.

The NCEDC installed a freeorb server at the USGS in Menlo Park to acquire and buffer the NCSN data for delivery to the NCEDC over the internet. The orbserver currently provides 2 hours of storage in the memory-mapped ring buffer file at Menlo Park. The NCEDC developed an Earthworm-to-Orb-MiniSEED acquisition program to acquire all NCSN waveform data from an Earthworm ring on a USGS/MP, convert the data to MiniSEED format, and insert the data into the observer's ring buffer. We created an orbserver client that runs on the NCEDC computer that connects to the orbserver in Menlo Park, retrieves the the NCSN waveform data records, and writes them to daily channel files in the NCEDC DART.

Most of the NCSN data are automatically archived at the NCEDC, but data from the NCSN broadband stations and Mammoth Deep Hole, most of which can deliver out-of-order data through the USGS Nanometrics satellite system, are processed by an NCEDC analyst using *calqc* to ensure that the most complete data set available from these sites are archived.

Parkfield High Resolution Seismic Network Data

Event seismograms from the Parkfield High Resolution Seismic Network (HRSN) from 1987 through June 1998 are available in their raw SEGY format via NCEDC research accounts. A number of events have faulty timing due to the lack or failure of a precision timesource for the network.

Due to funding limitations, there is currently no ongoing work to correct the timing problems in the older events or to create MiniSEED volumes for these events. However, a preliminary catalog for a significant number of these events has been constructed, and the catalog is available via the web at the NCEDC.

The original HRSN acquisition system died in late 1998, and an interim system of portable RefTek recorders were installed at some of the sites. Data from this interim system are not currently available online.

Starting in 2000, the HRSN was upgraded with Quanterra Q730 data loggers and digital telemetry, and 3 new borehole stations were added to the network. In 2000-2003 the PASO array, a temporary IRIS PASSCAL broadband network with real-time telemetry, was installed in the Parkfield area and its recording system was housed at the HRSN recording site in Parkfield. During this time, the HRSN collected event data from both the HRSN and PASO array and provided this integrated data set to researchers in near-real-time. The HRSN detected triggers using the HRSN stations and delivered triggered high-rate data from the HRSN and the PASO stations in real-time to the NCEDC, where they were made available to the research community via anonymous ftp until they are reviewed and permanently archived. In addition, the HRSN 20 Hz (BP) and state-of-health channels were archived continuously at the NCEDC. As an interim measure, the NCEDC also archived the continuous 250 Hz (DP) data channels through late 2002 in order to help researchers retrieve events that were not detected during the network upgrade.

The increased seismic activity related to the magnitude 6.5 earthquake in nearby San Simeon on December 22, 2003 drastically increased the number of triggers by the HRSN network. From December 2003 through August 2004, the HRSN had over 70,000 triggers. The 56Kb frame relay connection from Parkfield to UC Berkeley, which was installed to transmit continuous 20 Hz data, selected 250 Hz channels, and event triggered 250 Hz waveforms from the network, was saturated from the increased activity. The HRSN stopped telemetering the event-triggered waveforms, and the NCEDC started to archive continuous 20 and 250 Hz data from the entire network from tapes created at the HRSN operations center in Parkfield in order to preserve this unique dataset. The seismicity again increased after the magnitude 6.0 Parkfield earthquake on September 28, 2004.

In early 2006 the NCEDC started to receive the HRSN 20 Hz data and a subset of the 250 Hz data in real-time for distribution through the DART. The NCEDC continues to archive continuous 250 Hz and 20 Hz data streams from the HRSN tapes written in Parkfield and processed at the NCEDC. Plans are underway to provide complete and continuous real-time HRSN data to the NCEDC over the USGS and NCEMC T1 circuits starting in early 2007.

EarthScope USArray Transportable Array

EarthScope began installing broadband stations for the Transportable Array component of USArray in California in 2005. The NCEDC started acquiring telemetered continuous data from the northern California and surrounding stations as they were installed, and is archiving these data to support users working with northern California seismic data. These data are made available to users using the same data request methods as all other continuous data waveform data at the NCEDC. The Transportable Array stations have a limited operational timespan of 18 to 24 months, after which they will be relocated to new sites across the country. Data from these stations are delivered to the NCEDC as they are received by the BSL for distribution through the DART, and for subsequent permanent archiving.

EarthScope Plate Boundary Observatory (PBO) strain data

The NCEDC has been designated by EarthScope as one of two archives for PBO borehole and laser strain data. Strain data are collected from all of the PBO strain sites and are processed by UNAVCO. MiniSEED data are delivered to the NCEDC using SeedLink, and raw and XML processed data are delivered to the NCEDC using Unidata's Local Data Manager (LDM). The MiniSEED data are inserted into the NCEDC DART, and are subsequently archived from the DART. UNAVCO provides EarthScope funding to the NCEDC to cover the processing, archiving, and distribution costs for these data.

EarthScope SAFOD

The NCEDC is designated as the primary archive center for the SAFOD event data, and will also process the continuous SAFOD data . Starting in July 2002, scientists from Duke University successfully installed a three component 32 level downhole-seismic array in the pilot hole at the EarthScope SAFOD site in collaboration with Steve Hickman (USGS), Mark Zoback (Stanford University) and the Oyo Geospace Engineering Resources International (GERI) Corporation. High frequency event recordings from this array have been provide by Duke University for archiving at the NCEDC. We converted data from the original SEG-2 format data files to MiniSEED, and have developed the SEED instrument responses for this data set. We are currently converting continous high frequency data from the SAFOD Main Hole and Pilot Hole into MiniSEED, assembling the SEED instrument responses for these channels, and will maintain and distribute a buffer of "recent" continuous SAFOD data as well as a permanent archive of SAFOD event data. SAFOD will provide EarthScope funding to the NCEDC to cover the processing, archiving, and distribution costs for these data. A small subset of the continuous SAFOD data channels are also incorporated into the NCSN, and are available in real-time from the NCEDC DART.

UNR Broadband data

The University of Reno in Nevada (UNR) operates several broadband stations in western Nevada and eastern California that are important for northern California earthquake processing and analysis. Starting in August 2000, the NCEDC has been receiving and archiving continuous broadband data from four UNR stations. The data are transmitted in real-time from UNR to UC Berkeley, where they are made available for CISN real-time earthquake processing and for archiving. Initially, some of the stations were sampled at 20 Hz, but all stations are now sampled and archived continuously at 100 Hz.

The NCEDC installed Simple Wave Server (SWS) software at UNR, which provides an interface to UNR's recent collection of waveforms. The SWS is used by the NCEDC to retrieve waveforms from UNR that were missing at the NCEDC due to real-time telemetry outages between UNR and UC Berkeley.

In early 2006 the NCEDC started to archive continuous data from the UNR short-period stations that are contributed to the NCSN. Both the broadband and short-period UNR stations contributed to the CISN are also available in real-time through the NCEDC DART.

Electro-Magnetic Data

The NCEDC continues to archive and process electric and magnetic field data acquired at several UC Berkeley sites. Data loggers at PKD, SAO, and JRSC acquire data from 3 components of magnetic field and 2 or 4 components of electric field at 40 Hz, 1 Hz, and 0.1 Hz, and are telemetered in real-time along with seismic data to the Berkeley Seismological Laboratory, where they are processed and archived at the NCEDC in a similar fashion to the seismic data.

Using programs developed by Dr. Martin Fullerkrug at the Stanford University STAR Laboratory (now at the University of Bath), the NCEDC has computed and archived magnetic activity and Schumann resonance analysis using the 40 Hz data from this dataset. The magnetic activity and Schumann resonance data can be accessed from the Web.

The NCEDC also archived data from a low-frequency, long-baseline electric field project operated by Dr. Steve Park of UC Riverside at site PKD2. These data are acquired and archived in an identical manner to the other electric field data at the NCEDC.

GPS Data

The NCEDC continues to archive GPS data through the BARD (Bay Area Regional Deformation) network of continuously monitored GPS receivers in northern California. Prior to the establishment of the PBO GPS network, the NCEDC GPS archive included 67 continuous sites in northern California. There are approximately 50 core BARD sites owned and operated by UC Berkeley, USGS (Menlo Park and Cascade Volcano Observatory), LLNL, UC Davis, UC Santa Cruz, Trimble Navigation, and Stanford. Data are also archived from sites operated by other agencies including East Bay Municipal Utilities District, the City of Modesto, the National Geodetic Survey, and the Jet Propulsion Laboratory. Some of these sites have assimilated into PBO, and the NCEDC now archives approximately 40 continuous GPS sites.

In addition to the standard 15 second or 30 second continuous GPS datastream, the NCEDC is now archiving and distributing high-rate 1 Hz continuous GPS data from the 13 stations in Parkfield, and is providing these data to PBO. These high-rate data are available via anonymous FTP from the NCEDC but are currently not included in the GPS Seamless Archive (GSAC), since the GSAC does not currently handle both high-rate and low-rate data from the same site and day.

The NCEDC continues to archive non-continuous survey GPS data. The initial dataset archived is the survey GPS data collected by the USGS Menlo Park for northern California and other locations. The NCEDC is the principal archive for this dataset. Significant quality control efforts were implemented by the NCEDC to ensure that the raw data, scanned site log sheets, and RINEX data are archived for each survey. All of the USGS MP GPS data has been transferred to the NCEDC and virtually all of the data from 1992 to the present have been archived and are available for distribution.

Geysers Seismic Data

The Calpine Corporation currently operates a micro-seismic monitoring network in the Geysers region of northern California. Prior to 1999 this network was operated by Unocal. Through various agreements, both Unocal and Calpine have released triggered event waveform data from 1989 through 2000 along with preliminary event catalogs for the same time period for archiving and

distribution through the NCEDC. This dataset represents over 296,000 events that were recorded by Calpine/Unocal Geysers network, and are available via research accounts at the NCEDC.

The Lawrence Berkeley Laboratory (LBL), with funding from the California Energy Commission, operates a 22 station network in the Geysers region with an emphasis on monitoring seismicity related to well water injection. The earthquake locations and waveforms from this network are sent to the NCEDC, and the locations are forwarded to the NCSN so that they can be merged into the NCSN earthquake catalog. The LBL Geysers waveforms will be available at the NCEDC after the NCSN catalog has been migrated from flat files to the database.

USGS Low Frequency Data

Over the last 30 years, the USGS at Menlo Park, in collaboration with other principal investigators, has collected an extensive low-frequency geophysical data set that contains over 1300 channels of tilt, tensor strain, dilatational strain, creep, magnetic field, water level, and auxiliary channels such as temperature, pore pressure, rain and snow accumulation, and wind speed. In collaboration with the USGS, we assembled the requisite information for the hardware representation of the stations and the instrument responses for many channels of this diverse dataset, and developed the required programs to populate and update the hardware database and generate the instrument responses. We developed the programs and procedures to automate the process of importing the raw waveform data and convert it to MiniSEED format. Since these data are delivered to the NCEDC on a daily basis and immediately archived, these data are not inserted into the NCEDC DART.

We have currently archived timeseries data from 887 data channels from 167 sites, and have instrument response information for 542 channels at 139 sites. The waveform archive is updated on a daily basis with data from approximately 275 currently operating data channels. We will augment the raw data archive as additional instrument response information is assembled by the USGS for the channels, and will work with the USGS to clearly define the attributes of the ”processed” data channels.

SCSN/Statewide seismic data

In 2004 the NCEDC started to archive broadband and strong motion data from 15 SCSN (network CI) stations that are telemetered to the Northern California Earthquake Management Center (NCEMC) of the California Integrated Seismic Network (CISN). These data are used in the prototype real-time state-wide earthquake processing system and also provide increased coverage for northern California events. Since the data are telemetered directly from the stations in real-time to both the SCSN and to the NCEMC, the NCEDC archives the NCEMC’s copy of the data to ensure that at least one copy of the data will be preserved.

In early 2006 the NCEDC started to continuously archive all of the selected SCSN short-period stations that are contributed to the NCSN. All of these data are available in real-time from the NCEDC DART.

Northern California Seismicity Project

The objective of Northern California Seismicity Project is to characterize the spatial and temporal evolution of the northern and Central California seismicity during the initial part of the earthquake

cycle as the region emerges from the stress shadow of the great 1906 San Francisco earthquake. Although the current BSL catalog of earthquakes for the region appears to be a simple list of events, one must remember that it really is a very complex data set. The existing catalog is inhomogeneous in that it suffers from the three types of man-made seismicity changes: namely detection changes, reporting changes, and magnitude shifts. The inherent catalog inhomogeneity exists because the location and magnitude determination methodologies have changed as the instrumentation and computational capabilities improved over the past century. It is easy to misinterpret observed variations in seismicity if we do not understand these inherent limitations of the catalog. As a result, the northern and central California seismicity since 1906 is poorly understood.

Creation of a northern and central California catalog of seismicity that is homogeneous, that spans as many years as possible, and that includes formal estimates of the parameters and their uncertainty is a fundamental prerequisite for probabilistic studies of the seismicity. The existence of the invaluable BSL seismological archive, containing the original seismograms as well as the original reading/analysis sheets allows the application of modern analytical algorithms towards the problem of determining the source parameters of the historical earthquakes.

Our approach is to systematically re-analyze the data acquired from the reading/analysis sheet archive to develop a homogeneous catalog of earthquake location and local magnitude (ML) including formal uncertainties on all parameters which extends as far back in time as the instrumental records allow and which is complete above appropriate threshold magnitudes. We anticipate being able to compile a new catalog of location and ML which spans 1930 to the present and is which complete at approximately the ML 3.0 threshold.

We have completed the transcription of the original reading/analysis sheets to computer readable flat files for all ML 3.0 and larger earthquakes which have occurred in northern and central California and vicinity back to January 1, 1951. We started with the events that occurred in 1983 and worked backwards in time. The events from January 1, 1984 onwards were already in computer readable form. Data were transcribed from the original reading/analysis sheets for 5204 earthquakes and preliminary locations and local magnitudes have been calculated. We plan to transcribe reading/analysis sheet data back to at least 1932 but the process is more complicated and time consuming since we will have to pull the original Wood-Anderson seismograms from the archive to read the maximum trace amplitudes in order to calculate the local magnitude of the events.

Earthquake Catalogs

Northern California

The NCEDC provides searchable access to both the USGS and BSL earthquake catalogs for northern and central California. The ”official” UC Berkeley earthquake catalog begins in 1910 and runs through 2003, and the ”official” USGS catalog begins in 1966. Both of these catalogs are archived and available through the NCEDC, but the existence of 2 catalogs has caused confusion among both researchers and the public.

In early 2007, the NCEMC will begin providing a single unified northern California earthquake catalog in real-time to the NCEDC through database replication from the NCEMC’s real-time systems. The NCEDC has developed and is testing the required programs that will be used to enter all previous NCSN catalog data into the NCEDC database. We will then merge the the BSL catalog with the NCEMC catalog to form a single unified northern California catalog from 1910 to

the present. The BSL and the USGS have spent considerable effort over the past years to define procedures for merging the data from the two catalogs into a single northern and central California earthquake catalog in order to present a unified view of northern California seismicity. The differences in time period, variations in data availability, and mismatches in regions of coverage all complicate the task.

Worldwide

The NCEDC, in conjunction with the Council of the National Seismic System (CNSS), produced and distributed a world-wide composite catalog of earthquakes based on the catalogs of the national and various U.S. regional networks for several years. Each network updates their earthquake catalog on a daily basis at the NCEDC, and the NCEDC constructs a composite world-wide earthquake catalog by combining the data, removing duplicate entries that may occur from multiple networks recording an event, and giving priority to the data from each network's *authoritative region*. The catalog, which includes data from 14 regional and national networks, is searchable using a Web interface at the NCEDC. The catalog is also freely available to anyone via ftp over the internet.

With the demise of the CNSS and the development of the Advanced National Seismic System (ANSS), the NCEDC was asked to update its Web pages to present the composite catalog as a product of the ANSS. This conversion was completed in the fall of 2002. We continue to create, house, distribute, and provide a searchable web interface to the ANSS composite catalog, and to aid the regional networks in submitting data to the catalog.

Hardware and Software Systems

In 2005, the NCEDC relocated its archive and distribution system from McCone Hall to a new state-of-the-art computer facility in a new seismically braced co-location data center on the Berkeley campus. The facility provides seismically braced equipment racks, gigabit ethernet network, air condioning and power conditioning. The entire facility is powered by a UPS with generator backup.

During this award period, the NCEDC upgraded some of its computer and data storage hardware with a new tape library, RAID storage systems, and computers. The currently installed NCEDC facilities consist of a mass storage environment hosted by a Sun V240 host computer, a 100 slot LTO-2 tape library with two tape drives and a 20 TByte capacity, and over 30 TBytes of RAID storage. The storage and data backups are managed through the SAM-FS hierarchical storage management (HSM) software. A dual processor Sun Ultra 60 provides Web services and research account access to the NCEDC, a dual Sun 280R processor provide data import and export services, and a Sun Ultra 450 computer is used for quality control procedures. Two AIT tape libraries will be used to read NCSN continuous data tapes. An 64-bit Linux system hosts a database dedicated to providing data to external users. A new Sun Opteron processor has recently been purchased to upgrade the NCEDC web server.

The hardware and software system is configured to automatically create multiple copies of each timeseries file. The NCEDC creates an online copy of each file on online RAID, a second copy on LTO-2 tape which is stored online in the tape libraray, and a third copy on LTO-2 tape which is stored offline and offsite. All NCEDC data are online and rapidly accessible by users.

The NCEDC acquired an additional unlimited user Oracle license, and now operates two instances of its Oracle database; one for internal operations, and one for external use for user data queries and data distribution programs. The databases are synchronized using multi-master replication. In 2007, an additional NCEDC database will be installed and replicated at USGS Menlo Park as part of the NCEDC/NCEMC database replication environment.

Data Quality Control

The NCEDC developed a GUI-based state-driven system *calqc* to facilitate the quality control processing that is applied to the continuously archived data sets at the NCEDC.

The quality control procedures for these datasets include the following tasks:

- data extraction of a full day of data,
- quickcheck program to summarize the quality and stability of the stations' clocks,
- determine if there is missing data for any data channel,
- provided procedures to retrieve missing data from the stations and incorporate it into the day's data,
- optional creation of multi-day timeseries plots for state-of-health data channels,
- optional timing corrections for data,
- optional extraction of event-based waveforms from continuous data channels,
- optional repacking of MiniSEED data,
- creating waveform inventory entries in the NCEDC database,
- publishing the data for remote access on the NCEDC.

Calqc uses previously developed programs to perform each function, but it provides a graphical point-and-click interface to automate these procedures, and to provide the analyst with a record of when each process was started, whether it executed correctly, and whether the analyst has indicated that a step has been completed. *Calqc* is used to process all data from the BDSN network, and all continuous broadband data from the NCSN, UNR, SCSN, and HRSN networks that are archived by the NCEDC. The remainder of the continuously archived data are automatically archived without any analyst interaction.

Database Development

Due to restrictions imposed by the USGS/MP NCSN CUSP event analysis system, the NCEDC has traditionally stored the the official NCSN earthquake catalog, phase, amplitude, and coda readings in flat text files. The NCEDC worked closely with the Northern California Management Center (NCEMC) of the CISN to develop and test procedures that will allow the USGS/MP to replace the CUSP analysis system with *jiggle*, the analysis tool developed by the SCSN and to deliver earthquake parametric data in real-time to the NCEDC database. We developed the database tools

to insert the NCEMC earthquake parametric information into databases in the real-time earthquake analysis systems, and have extensively tested database replication between the NCEMC databases and the NCEDC database. We developed the programs necessary to migrate the NCSN catalog into the CISN parametric schema and to search and retrieve earthquake data from the database. In December 2006, the NCEMC retired the CUSP earthquake analysis software, an is now writing earthquake parameters directly ito the replicated database environment.

During 2002-2004, the NCEDC and NCSN jointly developed a system consisting of an extensive spreadsheet containing per-channel information that describes the hardware of each NCSN data channel and provides each channel with a SEED-compliant channel name. This spreadsheet, combined with a limited number of of files that describe the central-site analog digitizer, FIR decimation filters, and general characteristics of digital acquisition systems, allow the NCSN to assemble its station history in a format that the NCEDC can use to populate the hardware tracking and instrument response database tables for the NCSN.

The NCEDC instrument response schema represents full multi-stage instrument responses (including filter coefficients) for the broadband data loggers. The hardware tracking schema represents the interconnection of instruments, amplifiers, filters, and data loggers over time, and is used to describe all of the UC Berkeley and USGS stations and channels archived at the NCEDC. All NCSN event waveform and continuous timeseries data has been converted from CUSP and Earthworm format to MiniSEED, and are available along with the UC Berkeley data and data from the other networks archived at the NCEDC in full SEED format.

The NCEDC has developed XML import and export procedures to provide better maintenance of the hardware tracking information and resulting instrument responses for stations in our database. When changes are made to either existing hardware or to station configurations, we export the current view in XML format, use a GUI-based XML editor to easily update the information, and import the changes back into the database. When adding new stations or hardware, we can easily use information from existing hardware or stations as templates for the new information. This allows us to treat the database as the authoritative source of information, and to use off-the-shelf tools such as the XML editor and XML differencing programs as part of our database maintenance procedures.

Additional details on the joint catalog effort and database schema development may be found at `http://www.ncedc.org/db`

Data Distribution

The NCEDC continues to use the World Wide Web as a principal interface for users to request, search, and receive data from the NCEDC. In fall 2005 the NCEDC acquired the domain name *ncedc.org*. The NCEDC's Web address is now `http://www.ncedc.org/`

Earthquake Catalogs

The NCEDC provides users with searchable access to northern California earthquake catalogs and to the ANSS world-wide catalog via the web. Users can search the catalogs by time, magnitude, and geographic region, and can retrieve either hypocenter and magnitude information or a full set of earthquake parameters including phase readings, amplitudes, and codas.

Station Metadata

In addition to the metadata returned through the various data request methods, the NCECD provides dataless SEED volumes and SEED RESP file for all data channels archived at the NCEDC. The NCEDC currently has full SEED instrument responses for 9615 data channels from 1536 stations in 15 networks. This includes stations from the California Geological Survey (CGS) strong motion stations that will contribute seismic waveform data for significant earthquake to the NCEDC and Southern California Earthquake Data Center (SCEDC). We provide a generic ”digital counts” response for any channel for which we do not yet have an accurate instrument response.

SeismiQuery

We have ported and installed the IRIS *SeismiQuery* program at the NCEDC, which provides a common web interface for users to query network, station, and channel attributes, and to query the availability of archived timeseries data. We have provided both IRIS Data Management Center (DMC) and the SCEDC with our modified version of *SeismiQuery*.

DART (Data Available in Real Time)

The DART (Data Available in Real Time) represents the first step in NCEDC’s effort to make current and recent timeseries data from all networks, stations, and channels available to users in real-time. The NCEDC developed DART in December 2005 to provide a mechanism for users to obtain access to real-time data from the NCEDC. All real-time timeseries data stream delivered to the NCEDC are placed in MiniSEED files in a web-accessible directory structure. The DART waveforms can be accessed by web browsers or http command-line programs such as *wget*, a *FISSURES* waveform server, and a Berkelely-developed Simple Wave Server (SWS) which provides programmatic access to the DART data by specified SEED channel and time interval. We will be providing users with a client program to retrieve data from the SWS in the near future. The DART currently provide assess to the most recent 30 days of data.

We are using the Freeorb software, an enhanced version of the open-source orb software developed by the IRIS-funded Joint Seismic Project (JSP), as the primary method for delivering real-time data to the NCEDC and into the DART. The freeorb package implements an object ring buffer (ORB) and orbserver, which provides a reliable storage ringbuffer and an interface for orb client programs to read, write, and query the orbserver. Orbserver clients running at the NCEDC computer connect to remote orbservers at the BSL and USGS/Menlo Park, retrieve the MiniSEED timeseries data records, and write them to daily channel files in the NCEDC DART. Strain data from the EarthScope PBO network are delivered to the NCEDC using SeedLink, and are inserted into the DART using a similar SeedLink client program.

The NCEDC developed an automated data archiving system to archive data from the DART on a daily basis. It allows us to specify which stations should be automatically archived, and which stations should be handled by the NCEDC’s Quality Control program *calqc*, which allow an analyst to review the waveforms, retrieve missing data from stations or waveservers that may have contain late arriving out-of-order data, and perform timing corrections on the waveform data. The majority of data channels are currently archived automatically from the DART.

NetDC

In a collaborative project with the IRIS DMC and other worldwide datacenters during a previous award period, the NCEDC helped develop and implement *NetDC*, a protocol which will provide a seamless user interface to multiple datacenters for geophysical network and station inventory, instrument responses, and data retrieval requests. The *NetDC* builds upon the foundation and concepts of the IRIS *BREQ_FAST* data request system. The *NetDC* system was put into production in January 2000, and is currently operational at serveral datacenters worldwide, including NCEDC, IRIS DMC, ORFEUS, Geoscope, and SCEDC. The *NetDC* system receives user requests via email, automatically routes the appropriate portion of the requests to the appropriate datacenter, optionally aggregates the responses from the various datacenters, and delivers the data (or ftp pointers to the data) to the users via email.

STP

In 2002, the NCEDC wrote a collaborative proposal with the SCEDC to the Southern California Earthquake Center, with the goal of unifying data access between the two data centers. As part of this project, the NCEDC and SCEDC are working to support a common set of 3 tools for accessing waveform and parametric data: *SeismiQuery*, *NetDC*, and *STP*.

The *Seismogram Transfer Program* or *STP* is a simple client-server program, developed at the SCEDC. Access to *STP* is either through a simple direct interface that is available for Sun or Linux platforms, or through a GUI Web interface. With the direct interface, the data are placed directly on a user's computer in several possible formats, with the byte-swap conversion performed automatically. With the Web interface, the selected and converted data are retrieved with a single ftp command. The *STP* interface also allows rapid access to parametric data such as hypocenters and phases.

The NCEDC has continued work on *STP*, working with the SCEDC on extensions and needed additions. We added support for the full SEED channel name (Station, Network, Channel, and Location), and are now able to return event-associated waveforms from the NCSN waveform archive.

EVT_FAST

In order to provide Web access to the NCSN waveform before the SEED conversion and instrument response for the NCSN has been completed, the NCEDC implemented *EVT_FAST*, an interim email-based waveform request system similar to the *BREQ_FAST* email request system. Users email *EVT_FAST* requests to the NCEDC and request NCSN waveform data based on the NCSN event id. Initially the NCSN waveform data was converted to either SAC ASCII, SAC binary, or AH format, and placed in the anonymous ftp directory for retrieval by the users. *EVT_FAST* event waveforms can now also be provided in MiniSEED format, and are now named with their SEED channel names.

FISSURES

The *FISSURES* project developed from an initiative by IRIS to improve earth scientists' efficiency by developing a unified environment that can provide interactive or programmatic access to waveform data and the corresponding metadata for instrument response, as well as station and

channel inventory information. *FISSURES* was developed using CORBA (Common Object Request Broker Architecture) as the architecture to implement a system-independent method for the exchange of this binary data. The IRIS DMC developed a series of services, referred to as the *Data Handling Interface (DHI)*, using the *FISSURES* architecture to provide waveform and metadata from the IRIS DMC.

The NCEDC has implemented the *FISSURES Data Handling Interface (DHI)* services at the NCEDC, which involves interfacing the DHI servers with the NCEDC database schema. These services interact with the NCEDC database and data storage system, and can deliver NCEDC channel metadata as well as waveforms using the *FISSURES* interfaces. We have separate *FISSURES DHI* waveform servers to serve archived and DART data stream. Our *FISSURES* servers are registed with the IRIS *FISSURES naming services*, which ensures that all *FISSURES* users have transparent access to data from the NCEDC.

GSAC

Since 1997, the NCEDC has collaborated with UNAVCO and other members of the GPS community on the development of the *GPS Seamless Archive Centers (GSAC)* project. This project allows a user to access the most current version of GPS data and metadata from distributed archive locations. The NCEDC is participating at several levels in the *GSAC* project: as a primary provider of data collected from core BARD stations and USGS MP surveys, and as a wholesale collection point for other data collected in northern California. We helped to define database schema and file formats for the *GSAC* project, and have produced complete and incremental monumentation and data holdings files describing the data sets that are produced by the BARD project or archived at the NCEDC so that other members of the *GSAC* community can provide up-to-date information about our holdings. Currently, the NCEDC is the primary provider for over 138,000 data files from over 1400 continuous and survey-mode monuments. The data holdings records for these data have been incorporated into the *GSAC* retailer system, which became publicly available in late 2002.

In addition, the NCEDC is archiving and distributing high-rate 1 Hz GPS data from 13 Parkfield GPS stations. These high-rate data are available by FTP from the NCEDC, but are not available through GSAC due to GSAC's inability to distinguish multiple data streams with different sample rates for the same day and station.

PUBLICATIONS AND PRESENTATIONS

Neuhauser, D., Northern California Earthquake Data Center: Long term data archive and distribution center for geophysical data with a focus on northern and central California, invited talk at the SCEC Workshop: Converting Advances in Seismology into Earthquake Science, Pasadena, CA, September 22-23, 2003.

Neuhauser, D., F. Klein, S. Zuzlewski, E. G. Jensen, D. Oppenheimer, L. Gee, and B. Romanowicz, Open Access to Decades of NCSN Waveforms at the Northern California Earthquake Data Center, Eos Trans. AGU, 84(46), Fall Meet. Suppl., Abstract S52C-0150, 2003; poster presented at the American Geophysical Union, 2003 Fall Meeting, San Francisco, CA, December 8-12, 2003

Neuhauser, D., F. Klein, S. Zuzlewski, L. Gee, D. Oppenheimer, and B. Romanowicz, The Northern California Earthquake Data Center: Seismic and Geophysical Data for Northern California and Beyond Eos Trans. AGU, 85(47), Fall Meet. Suppl., Abstract S21A-0256, 2004; poster

presented at the American Geophysical Union, 2004 Fall Meeting, San Francisco, CA, December 13-17, 2004

Neuhauser, D. S., L. Gee, S. Zuzlewski, M. H. Murray, B. Romanowicz, F. W. Klein, H. Macbeth, and D. Oppenheimer, ANSS and Other Earthquake Data at the NCEDC, Seismological Research Letters, 76(2), p230, 2005; poster presented at the Seismological Society of America, 2005 Annual Meeting, Incline Village, NV, April 27-19, 2005

Neuhauser, D., L. Dietz, S. Zuzlewski, W. Kohler, L. Gee, D. Oppenheimer, B Romanowicz, New Continuous Timeseries Data at the Northern California Earthquake Data Center, Eos Trans., AGU, 86(52), Fall Meet. Suppl, Abstract S33B-0326, 2005; poster presented at the American Geophysical Union, 2005 Fall Meeting, San Francisco, CA, December 5-9, 2005

Neuhauser, D. S., F. Klein, S. Zuzlewski, M. Murray, L. Dietz, N. Houlié, D. Oppenheimer, B. Romanowicz, Diverse Continuous Seismic, Geophysical, and Geodetic Data at the Northern California Earthquake Data Center (NCEDC), Seismological Research Letters, 76(2), p252, 2006; poster presented at the Seismological Society of America, 2006 Annual Meeting, San Francisco, CA, April 18-22, 2006

Neuhauser, D., L. Dietz, P. Lombard, F. Klein, S. Zuzlewski, W. Kohler, M. Hellweg, J. Luetgert, D. Oppenheimer, B. Romanowicz, The Northern California Earthquake Management System: A Unified System From Realtime Monitoring to Data Distribution, Eos Trans. AGU, 87(52), Fall Meet. Suppl., Abstract S11A-0198; poster presented at the American Geophysical Union, 2005 Fall Meeting, San Francisco, CA, December 11-15, 2006

DATA AVAILABILITY

Data are available from the Northern California Earthquake Data Center via the internet at `http://www.ncedc.org`. For additional information, contact Douglas Neuhauser at (510) 642-0931 or doug@seismo.berkeley.edu.

NON-TECHNICAL SUMMARY

Operation of the Northern California Earthquake Data Center

Award Number: 03HQAG0171

Barbara Romanowicz, P.I., and Doug Neuhauser
Berkeley Seismological Laboratory, UC Berkeley, CA 94720-4760
Telephone: (510) 643-5690, FAX: (510) 643-5811, barbara@seismo.berkeley.edu
Telephone: (510) 642-0931, FAX: (510) 643-5811, doug@seismo.berkeley.edu

NEHRP Elements: I & II

Keywords: Seismology; Geodesy; Database

The Northern California Earthquake Data Center (NCEDC) is an long-term online archive and distribution facility for waveform and catalog data for several regional networks including: The USGS Northern California Seismic Network (NCSN), the Berkeley Digital Seismic Network (BDSN) and Parkfield High Resolution Seismic Network (HRSN) operated by the UC Berkeley Seismological Laboratory; the USGS Low Frequency Geophysical Data set; the Bay Area Deformation Array (BARD) GPS network; and portions of the EarthScope strain and seismic networks. These data serve as a basis for many research projects relevant to NEHRP goals in the Bay Area and central and northern California.