

LAWRENCE LIVERMORE NATIONAL LABORATORY'S MIDDLE EAST AND NORTH AFRICA RESEARCH DATABASE

Stanley D. Ruppert, Teresa F. Hauk, Jennifer L. O'Boyle, Douglas A. Dodge, R. Miki Moore
Geophysics and Global Security, Lawrence Livermore National Laboratory
University of California

Sponsored by U.S. Department of Energy
Office of Nonproliferation and National Security
Office of Research and Development
Contract No. W-7405-ENG-48

ABSTRACT:

The Lawrence Livermore National Laboratory (LLNL) Comprehensive Nuclear-Test-Ban Treaty Research and Development (CTBT R&D) program has made significant progress populating a comprehensive seismic research database (RDB) for seismic events and derived research products in the Middle East and North Africa (ME/NA). Our original ME/NA study region has enlarged and is now defined as an area including the Middle East, Africa, Europe, Southwest Asia, the Former Soviet Union and the Scandinavian/Arctic region. The LLNL RDB will facilitate calibration of all International Monitoring System (IMS) stations (primary and auxiliary) or their surrogates (if not yet installed) as well as a variety of gamma stations. The RDB provides not only a coherent framework in which to store and organize large volumes of collected seismic waveforms and associated event parameter information, but also provides an efficient data processing/research environment for deriving location and discrimination correction surfaces and capabilities. In order to accommodate large volumes of data from many sources with diverse formats the RDB is designed to be flexible and extensible in addition to maintaining detailed quality control information and associated metadata. Station parameters, instrument responses, phase pick information, and event bulletins were compiled and made available through the RDB. For seismic events in the ME/NA region occurring between 1976 and 1999, we have systematically assembled, quality checked and organized event waveforms; continuous seismic data from 1990 to present are archived for many stations. Currently, over 11,400 seismic events and 1.2 million waveforms are maintained in the RDB and made readily available to researchers. In addition to open sources of seismic data, we have established collaborative relationships with several ME/NA countries that have yielded additional ground truth and broadband waveform data essential for regional calibration and capability studies. Additional data and ground truth from other countries are also currently being sought. Research results, along with descriptive metadata are stored and organized within the LLNL RDB and prepared for delivery and integration into the Department of Energy (DOE) Knowledge Base (KB). Deliverables consist of primary data products (raw materials for calibration) and derived products (distilled from the organized raw seismological data). By combining travel-time observations, event characterization studies, and regional wave-propagation studies of the LLNL CTBT research team for ground truth events and regional events, we have assembled a library of ground truth information, event location correction surfaces, tomographic models and mine explosion histories required to support the ME/NA regionalization program. Corrections and parameters distilled from the LLNL RDB provide needed contributions to the KB for the ME/NA region and will enable the United States National Data Center (NDC) to effectively verify CTBT compliance. The LLNL portion of the DOE KB supports critical NDC pipeline functions in detection, location, feature extraction, discrimination, and analyst review in the Middle East and North Africa.

Key Words: seismic, waveform, database, metadata

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48 for the Office of Research and Development, NN-20, within the Office of Nonproliferation and National Security, NN-1.

OBJECTIVE

The primary objective of the Lawrence Livermore National Laboratory (LLNL) comprehensive seismic research database (RDB) is to help coordinate the LLNL Comprehensive Nuclear-Test-Ban Treaty (CTBT) Middle East and North Africa (ME/NA) regionalization program. Corrections and parameters distilled from the LLNL RDB provide needed contributions to the Department of Energy (DOE) Knowledge Base (KB) for the ME/NA region and will enable the United States National Data Center (NDC) to effectively verify CTBT compliance. The LLNL portion of the DOE KB supports critical NDC pipeline functions in detection, location, feature extraction, discrimination, and analyst review in the Middle East and North Africa. The LLNL RDB provides efficient access to, and organization of, thousands of seismic events and associated waveforms, while also providing the framework to store, organize, and disseminate research results for delivery into the DOE KB (Figure 1). Reference event libraries and ground truth datasets showing space and time clustering of natural earthquake and mine events, phase blockage maps, and event characterization parameters are compiled from the RDB. Sufficient metadata (including measurement procedures, codes, comments and measurement errors) are stored at each step in the analysis process to allow recreation or verification of results at any stage in the processing flow.

DOE Knowledge Base deliverables created using the RDB may be grouped under two major categories: primary data products and derived products. The primary products are those developed in the process of collecting the raw materials for calibration: ground truth data, waveform data, event catalogs, phase pick information, regional station information and instrument responses. The derived products (distilled from the organized raw seismological data) are models and corrections that improve detection, location and discrimination functions. In order for LLNL to calibrate all International Monitoring System (IMS) stations (primary and auxiliary), as well as a variety of gamma stations that experience has shown to be useful, the LLNL RDB must incorporate and organize the following categories of primary and derived measurements, data and metadata:

- 1) Waveforms and phase picks on high-priority data
- 2) Waveform metadata (trace quality, properties and processing history)
- 3) Station and channel parameters
- 4) Global, regional, and local earthquake catalogs
- 5) Travel-time models, regional velocity models, kriged corrections surfaces
- 6) Surface wave group velocities
- 7) Detection filter bands to optimize signal-to-noise
- 8) Regional discriminants, phase amplitude and magnitude calibrations
- 9) Reference archives of waveform and parameter data for special regions of study

RESEARCH ACCOMPLISHED

Data Acquisition

Data collection for the LLNL RDB began in 1996 and continues today. The LLNL study area surrounding the Middle East and North Africa to include Europe, part of the Former Soviet Union, and Central Africa, has been termed the ME/NA region. The region of study spans the area between latitudes 10° South and 60° North and between longitudes 35° West and 85° East. Most of the archived waveforms in the LLNL RDB are from events located within this region and occurring between 1976 and 1999, but we are now expanding our event coverage into the Former Soviet Union and the Scandinavian/Arctic region (Figure 3). The new region, designated ME/NA/FSU, surrounds the mid-ocean ridges and spans 50° West to 90° East and 55° South to the North Pole and also includes an additional region of 65° North to the North Pole and 90° to 145° East to encompass the remainder of the Arctic ridge events. Special datasets are compiled for specific areas of interest or events, such as the Novaya Zemlya test site and the recent India and Pakistan nuclear tests.

Waveforms

We are collecting seismic data from IMS primary and auxiliary stations, as well as surrogate stations (for IMS stations not yet installed) and other stations needed to support calibration in the region of study (Figure 2). We have obtained up to 10 years of continuous data for important ME/NA stations from the Incorporated Research Institute for Seismology (IRIS), Institut de Physique du Globe de Paris GEOSCOPE

program, and the GEO-Forschungs Zentrum/Potsdam, Germany GEOFON program. Current data is being received on a regular basis since June 1998 from several arrays and stations supplied via the Air Force

Technical Applications Center (AFTAC) through cooperation with Sandia National Laboratories. Current data is also being supplied by an LLNL joint project with the Jordan Natural Resources Authority from two seismometers deployed in Jordan. We are establishing a similar collaboration with Kuwait. Data for particular events has been obtained from the prototype International Data Center (PIDC). AFTAC and the Center for Monitoring Research (CMR) have provided waveforms for special regions/events, such as the Novaya Zemlya test site.

Data requests are usually made through automated systems provided by networks and seismic data centers, which supply data by email, ftp, or tape. Our current emphasis is on events recorded between 1990 and present, except for IMS surrogate stations no longer in operation or for special events such as nuclear tests. All available channels and components are requested from each station or array. Stations in operation prior to 1990 are typically limited to long period (1 sps) or very long period (0.1 sps) continuous data and short period (40-80 sps), vertical component, triggered data; broadband (20 sps) data exist from very few locations before 1990. Data from IRIS, GEOSCOPE, and GEOFON are provided in SEED format, waveforms from AFTAC and CMR are organized in CSS format, and data from LLNL deployments are recorded in REFTEK format.

The RDB is organized in CSS3.0 (Center for Seismic Studies Version 3.0) format with NDC and LLNL extensions. All data is converted to CSS3.0 format before it is added to the RDB. Waveforms are stored as **.w** files and are referenced through **wfdisc** and **wftag** tables. Additionally, although the continuous data remains archived on tapes, seismic events are extracted from the continuous waveforms. Events used to populate the database are a subset of the (United States Geological Survey) USGS Monthly (Final) Preliminary Determination of Epicenters (PDE) bulletin and consist of events in the ME/NA region from 1990 to early 1999. Originally, the RDB was populated with events of magnitude 4.0 and greater, but the selection has now been enlarged to include all events magnitude 3.5 and above. The list of events has also been altered to include the expanded ME/NA/FSU region. Waveforms are extracted from continuous data based on consistent time window criteria: the waveform begins at the event origin time and ends at the arrival time of seismic waves traveling at 1 km/sec. Due to the size of the region of study, source-receiver distances range from local to teleseismic.

The number of waveforms in the RDB now exceeds 1.2 million, which corresponds to over 11,400 seismic events (Figure 4). There are over 30 stations in the ME/NA region for which we have over 10,000 waveforms for up to 6000 events (Figure 2). We have nearly mined the available historic data supply for open stations in the ME/NA region, but will continue to collect current data from these stations as it becomes available. The data collection process is now being focused on stations in the expanded ME/NA/FSU region. Also, we are augmenting the open historic station data with seismic data from LLNL field deployments and collaborative research efforts established in ME/NA countries.

In addition to individual event waveform segments and continuous data traces managed by the RDB, we also maintain an archive of active and passive seismic data from various field deployments. One example is the archive of very long FSU PNE refraction profiles collected under contract to LLNL by the USGS (Figure 5). Other deployments include IRIS experiments in Tanzania, Pakistan, Caspian, and Geyokcha. These profiles (in SEG-Y format) collected throughout the ME/NA/FSU region provide critical information for independently determining spatial correlation distances and developing the regional geophysical models necessary to calibrate the many aseismic areas in our study region. Both LLNL and the USGS are jointly analyzing and modeling the refraction profiles to provide both travel time and amplitude information throughout the FSU.

Phase Information

Analyst Flori Ryall and LLNL researchers have made phase picks for over 3000 events to yield over 10,000 travel-time observations available to the LLNL research team for location and discrimination projects. Phase analysis is an ongoing effort since new events and waveforms are continually being added to the RDB. Phase information is recorded in the **arrival** and **assoc** tables. Augmenting the LLNL picks, we added ~5.5 million USGS Earthquake Data Report (EDR) catalog phase pick observations to the RDB

to be used for travel-time correction studies and correction surface generation. Since the EDR picks are available only for 1990-1997, we supplemented the EDR picks with the complete Bulletin of the International Seismological Centre (ISC) phase arrival measurements spanning from 1964-1995. Phase picks have also been entered into the RDB from the Joint Seismic Observing Program (JSOP) and Bob Engdahl's reviewed subset of the ISC bulletin.

Station Parameters

Development of travel-time corrections requires not only accurate source locations and origin times, and quality-controlled waveform data, but also accurate knowledge of station locations. Seismic station information is a metadata requirement needed to support all stages of seismic waveform analysis. This metadata includes such parameters as station operation dates, location and elevation, type of channels and instruments, sampling rates, and instrument responses. Our main source of this information is IRIS "dataless" SEED files, which are provided by each of the networks affiliated with IRIS. These files contain all station parameters and response information and are updated periodically by the networks. Other station information has been obtained through internet station books and AutoDRM systems. **Site** and **sitechan** table entries (listing station location, available channels, sensor orientations, operation dates, etc.) were created for almost all IRIS affiliated networks. This will provide worldwide station information to accommodate any future changes in research needs or the LLNL study region and greatly minimize the need to add or change station entries in the future.

We reviewed all new and existing station and channel information for completeness and coherency before updating the RDB tables. Over 1100 station and array element table entries have been updated, but we still have not located parameter data for many stations (Figure 3). Minimal or inconsistent information has restricted the reliability of certain entries. Often discrepancies arise between multiple information sources for the same station, but IRIS or the network operator is assumed to offer the most complete information in most cases. Network operators have been contacted to resolve significant problems. We maintain a list of sources for each item of station information (**sitesrc**, **sitechansrc**), since each parameter may be obtained from a different source. **Network** and **affiliation** tables have been created to track the network(s) to which each station is associated. **Instrument** and **sensor** tables are used to document instrument type and response for each station and channel. The IRIS "dataless" SEED files are used to generate instrument response (RESP) files for each station/network/channel/time combination delineated by calibration periods. The RDB **instrument** table contains pointers to the online flatfiles. Frequency-amplitude-phase (FAP) files have been provided by AFTAC and are similarly stored with RDB pointers.

Event Bulletins

Reference event locations and origin time information is necessary in most stages of our seismic processing and research. Bulletin information from many global, local and regional earthquake catalogs has been incorporated into the LLNL RDB. The global catalogs include: USGS Monthly (Final) Preliminary Determination of Epicenters (PDE) catalog, USGS Earthquake Data Report (EDR) catalog with phase arrival information, Bulletin of the International Seismological Centre (ISC) with phase arrival information, Harvard Centroid Moment Tensor (CMT) catalog, and AFTAC PREACH unclassified global catalog. We have also compiled numerous regional and local catalogs from countries in the ME/NA region, including Jordan, Israel, Morocco, and the Joint Seismic Observing Program (JSOP). We have established a number of collaborative agreements with countries and institutes in our study region that are yielding both local seismic catalogs and ground truth information as well as seismic waveform data. We compiled, assembled, and cross-checked several global seismicity catalogs and bulletins of phase arrival measurements in order to create CSS3.0 tables and custom tables in the RDB. Since the catalogs are in a variety of formats, we use a combination of custom filter programs and modified (fixed) DATASCOPE programs to convert all of the bulletins to CSS3.0 table format. In the case of the PDE and CMT catalogs, original bulletin formats are archived in order to retain information not having a CSS3.0 table entry (such as source reference, source parameter information, moment, focal mechanism, etc.). We also reconcile and compare the same bulletin information obtained from different sources in order to create a final quality checked catalog.

Database Organization

Database Format

The LLNL RDB is designed to be flexible and extensible in order to accommodate large volumes of data in diverse formats from many sources in addition to maintaining detailed quality control and metadata. The RDB is comprised of ORACLE relational database software running on a SUN Server accessible from researcher workstations (Figure 6). Data are stored in CSS3.0 format (Center for Seismic Studies Version 3.0 database structure) with NDC and LLNL extensions. These formats provide parameter defined tables for different elements of seismic data, such as event and station information, as well as allowing for customized tables to be developed for specific research needs or results. The CSS3.0 format offers the ability to organize data in a standard format, store metadata necessary for documenting research methods and deliver compatible data to DOE and other researchers. Derived location and amplitude correction surfaces are stored in the standardized LibKBI format for efficient access by location and discrimination software programs.

Table Population

In order to efficiently generate new database entries and update existing tables, it was necessary to develop software to automate these procedures. The following software reduces the time involved with the upkeep of the large amounts of incoming data:

- 1) UpdateMrg – inserts waveform pointers in **wfdisc** and **wftag** tables for new waveforms, verifies waveforms are not “flat line,” and joins all waveform segments related to a single event
- 2) UpdateResp – determines displacement amplitude responses, calculates nominal calibration values, populates instrument and sensor tables based on calibration epochs in RESP and FAP files
- 3) UpdateArrival – populates **arrival** and **assoc** tables with phase information obtained from different bulletin formats or analyst picks
- 4) Catalog Parser – extracts event information from various bulletin formats in order to generate entries for an **origin** table

We are also in the process of developing custom tables (where existing CSS3.0 or NDC tables are inadequate) in order to organize and store research results in a manner that will allow these results to be accessible to the research team and to track metadata related to table entries. Examples of necessary custom tables are analyst comments for phase picks and events, mine explosion statistics and surface wave measurements.

Waveform Quality Assessment

The organization of waveform data into a database with a measure of data quality is fundamental to calibration activity, estimation of errors within other KB products based on waveform analysis, and to future detailed evaluation of special events by analysts at the NDC and elsewhere. The assembly of seismic waveform quality measurements is major task since it involves assessing large quantities of data from numerous waveform sources with several incompatible formats of varying quality. RDB waveforms may contain dropouts, glitches, timing errors or other problems affecting waveform analysis. We have developed an algorithm that analyzes and reports on 18 separate problems in the categories of timing errors, zero slope detection, discontinuity detection, and median filtering. This waveform quality information can be generated while focusing on all or a part of a waveform, examples of which include pre-event noise, the first arrival or event coda. This automated procedure will measure waveform quality in a consistent manner and generate three custom database tables: **wavequal**, **wavequalspt**, **wavequaldisc**. These tables contain the types of problems encountered and pointers to detailed files with the location and frequency of problems encountered. The development of these tables allows researchers and analysts to quickly identify sets of reviewed waveforms, thus reducing the laborious process of previewing each trace.

Data Access

Different researcher needs for data and metadata require that subsets of data must be provided in a format easily accessible to many diverse types of software and analyses tools. Therefore, the RDB access tools have been designed to utilize the power of the relational database to facilitate efficient queries and data retrieval (Figure 6). The Seismic Analysis Code (SAC) software used by LLNL researchers provides direct access to database table information and waveforms and uses the response files to perform instrument response corrections. PL/SQL language can be used to make database queries on contents of any of the available CSS3.0 tables. We are currently developing a GUI interface to the ORACLE database to support

easier access to RDB information; a web tool (Web Application Server) is also available for the ORACLE system. For spatial queries and organization, we have adopted the ESRI product ArcView (Figure 6). Arcview has been linked with the ORACLE database to provide joint spatial and relational queries. Given the large quantity of data now managed by the RDB, emphasis has shifted to produce the efficient “production” level seismic data selection and processing tools necessary to meet programmatic and KB delivery schedules. In addition, data browsers are under development in coordination with Sandia National Laboratory to allow visualization and quick access to both data and delivered research products.

CONCLUSIONS AND RECOMMENDATIONS

Research Product Delivery

Derived Research Products

Corrections and parameters distilled from the LLNL RDB provide needed contributions to the DOE Knowledge Base for the ME/NA region and will enable the USNDC to effectively verify CTBT compliance. The LLNL portion of the DOE KB supports critical NDC pipeline functions in detection, location, feature extraction, discrimination, and analyst review in the Middle East and North Africa. A wide range of research projects required to support the ME/NA regionalization program are being derived from waveforms, station parameters, and bulletin information contained in the LLNL RDB. Current discrimination and location research includes:

- 1) Mine Atlas: ground truth mine explosions and mine explosion statistics
- 2) Ground truth depth and mechanism estimates based on waveform modeling
- 3) Regional phase amplitude measurements and correction surfaces
- 4) Surface wave group velocities and correction surfaces
- 5) Regional coda magnitude algorithms
- 6) Location calibration and travel-time correction surfaces
- 7) Phase Amplitude correction surfaces and phase amplitude ratio discriminant results for specific regions
- 8) Important primary data products, reviewed station parameters, supplement these research products
- 9) Regional assessment & validation maps (preliminary capability studies used to guide LLNL research)

By combining travel-time observations, event characterization studies, and regional wave-propagation studies of the LLNL CTBT research team for ground truth and regional events, we have assembled a library of ground truth information (origin times, locations, depths, magnitudes), event location correction surfaces, tomographic models and mine explosion statistics required to support the ME/NA regionalization program. Future efforts will involve creating customized database tables to store all of the derived data products and related metadata. This will enhance the efficiency in compiling research results for dissemination to the DOE KB.

Knowledge Base Deliveries

LLNL research product submissions during the past year to the DOE KB have included updates for many of the above research products (Specific details of LLNL research products delivered are described in the Research Product Delivery Documents). The research products delivered in the version 2.0 release [Ruppert et al., 1999] represent a significant advance over the initial research products delivery in Spring 1998. In the version 2.0 release, we delivered both reference ground-truth parameters (reference mining event waveforms and Mine Parameter Atlas) and derived location (travel-time) and discrimination (amplitude) correction surfaces for key selected stations within our ME/NA research region. The travel-time correction surfaces are delivered as a framework that allows initial location and performance assessment to be completed within the ME/NA region while also allowing each individual station surface to be easily updated as additional data becomes available. We also delivered completely revised and expanded lookup tables for critical station parameter information (location, response, and other operational characteristics). Surface wave measurements delivered in version 2.0 are completely revised from the version 1.0 spring release, and are integrated with amplitude measurements for two stations that form an integrated package of measurements, discriminants and correction surfaces to support source identification efforts of CTBT monitoring. The algorithms developed and used to create these research products are currently being adapted to efficiently process the additional volume of data necessary to expand these techniques to additional stations necessary to support the next major DOE Knowledge Base delivery scheduled for mid FY 2000.

REFERENCES

Ruppert, S., CTBT R&D Staff, LLNL Middle East and North Africa Research Product Delivery Version 2.0, LLNL, June 1999

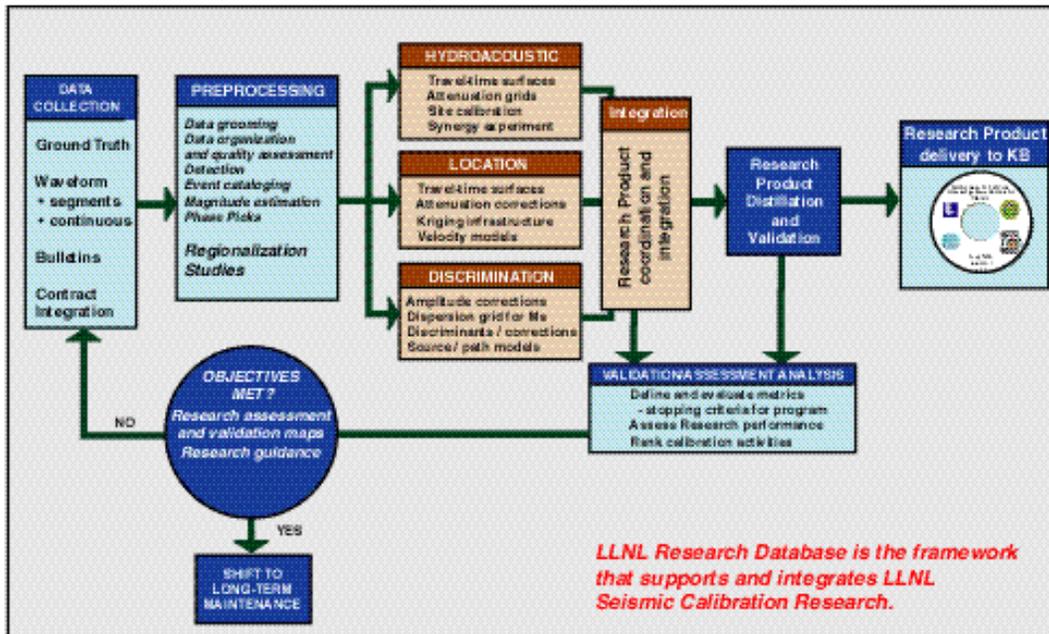


Figure 1: Overview of LLNL research database framework and product production flow.

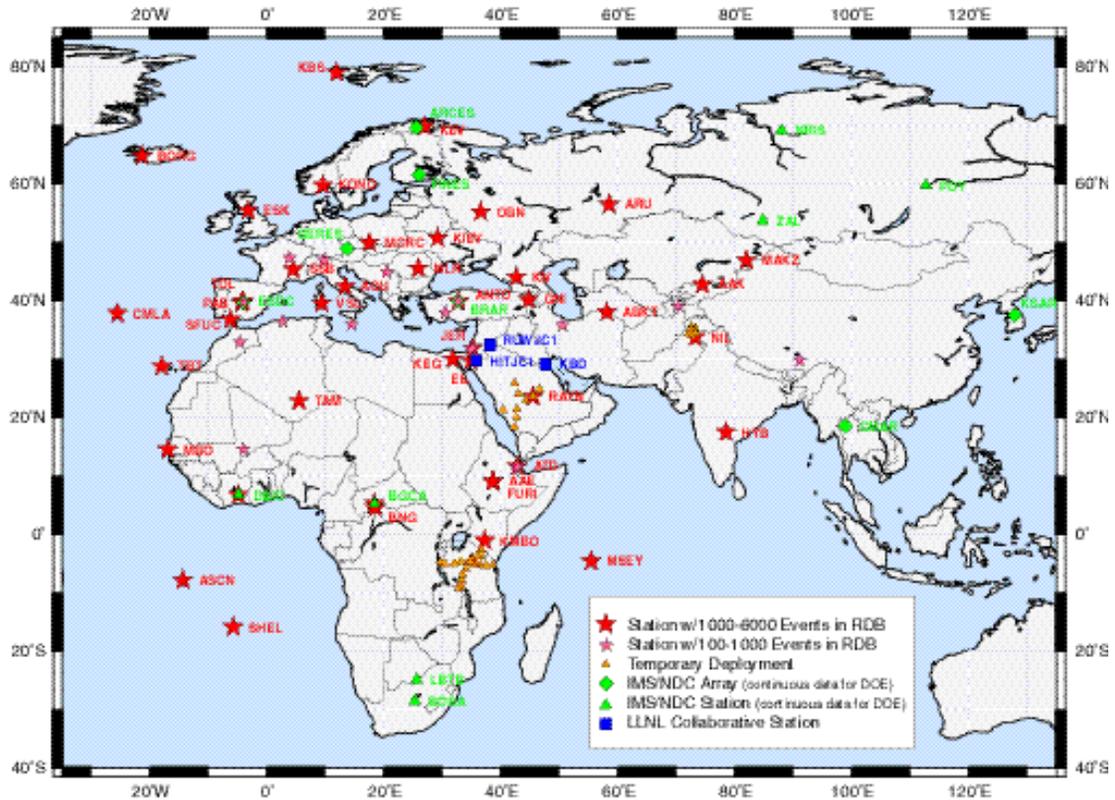


Figure 2: Seismic stations supplying large quantities of seismic event data to the LLNL research database. Data from additional stations is actively being acquired for future addition to the database.

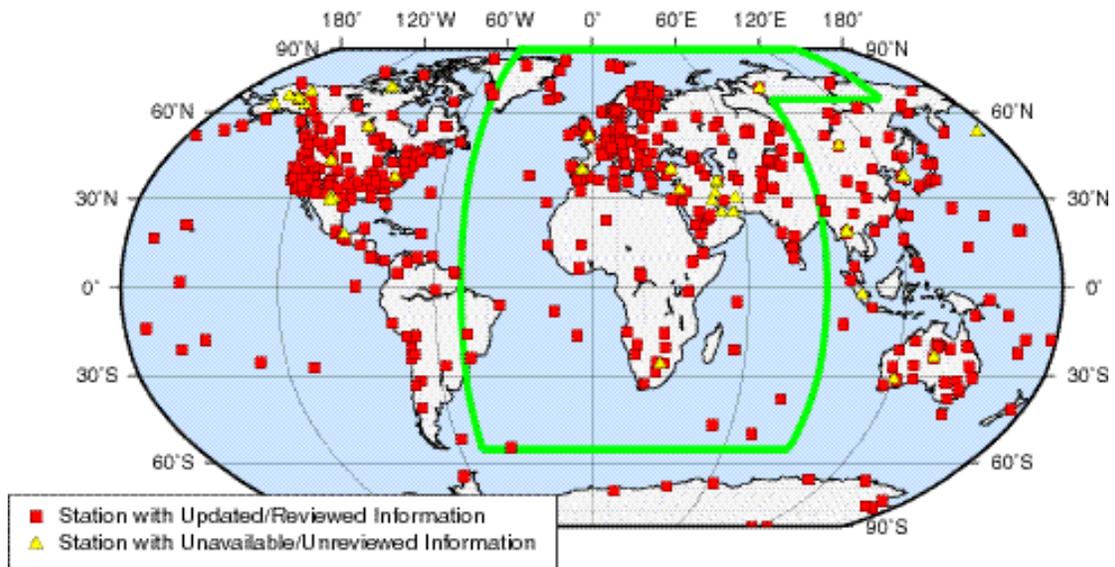


Figure 3: Distribution of all stations contained in the LLNL research database station parameter tables. Station information has been updated and reviewed for completeness and coherency for 1100 station and array element table entries. The outlined region defines the expanded ME/NA/PSU study area, incorporating the Former Soviet Union, Southwest Asia and the Scandinavian/Arctic region, within which seismic events will be collected.

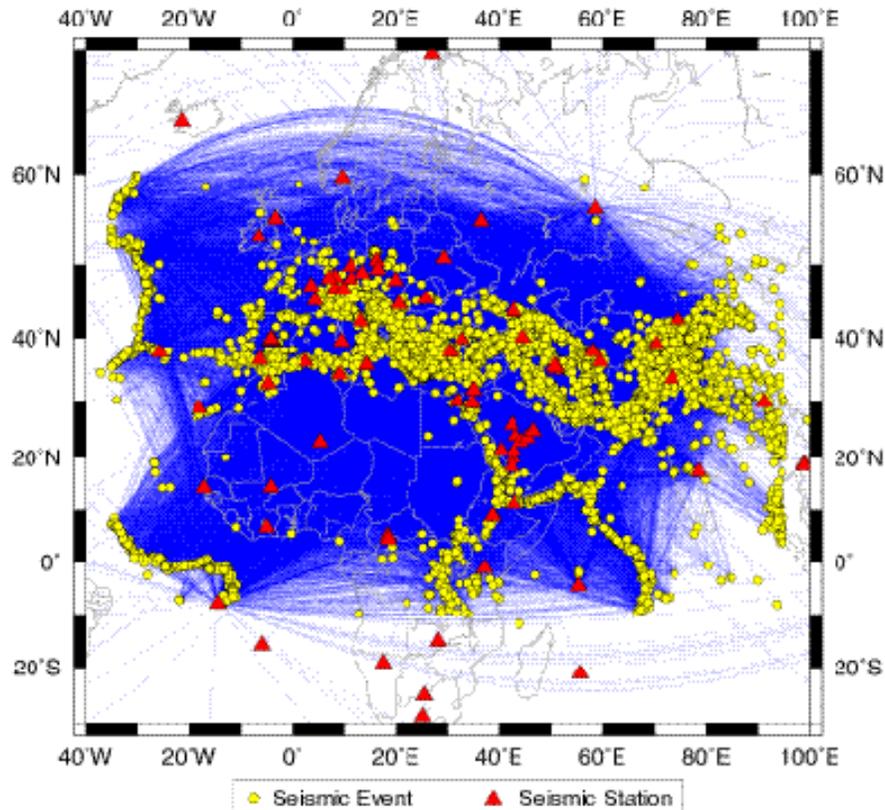


Figure 4: Seismic data available in the LLNL research database. Raypaths show source to receiver coverage for 1.2 million waveforms in the ME/NA region, corresponding to over 11,400 seismic events available to LLNL researchers.

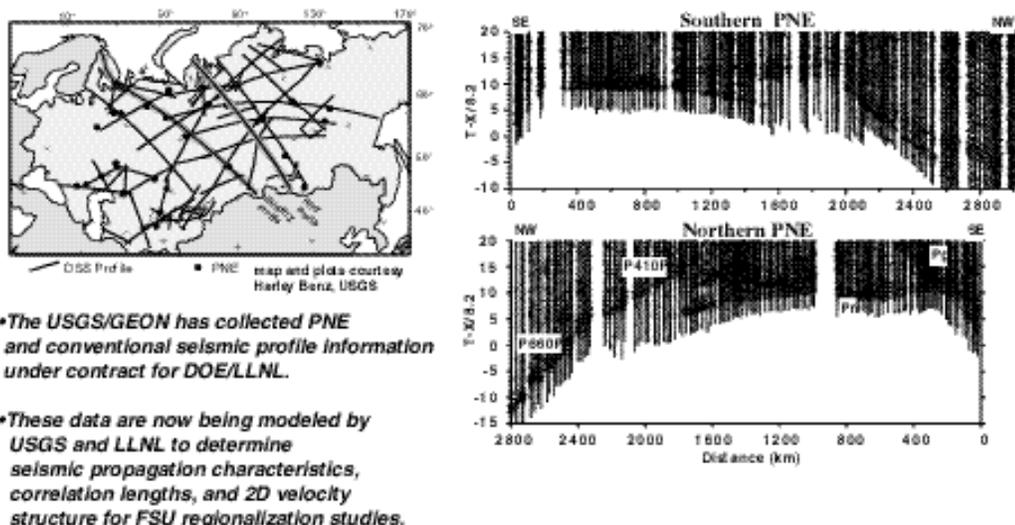


Figure 5: Example of seismic refraction data from FSU QUARTZ profile using PNE's.

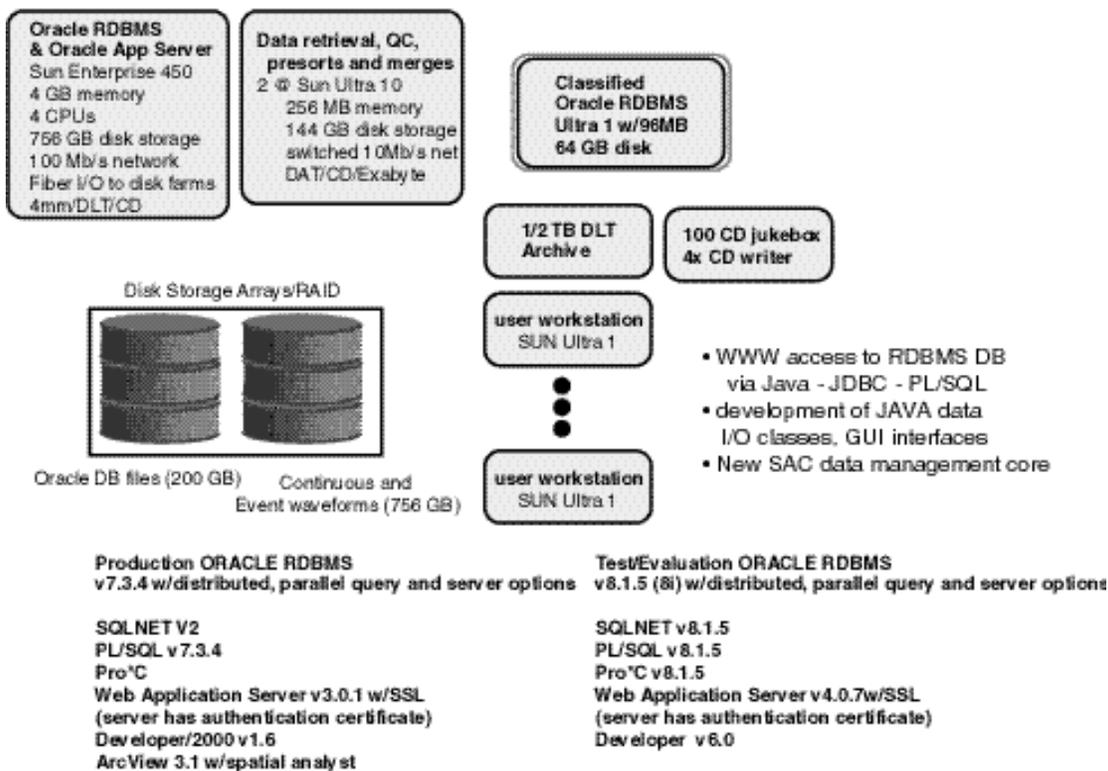


Figure 6: LLNL research database architecture and access tools.