DATA RESOURCES TO SUPPORT NUCLEAR EXPLOSION MONITORING RESEARCH AND DEVELOPMENT

Robert L. Woodward, Manochehr Bahavar, and Robert G. North

Science Applications International Corporation

Sponsored by Army Space and Missile Defense Command

Contract No. DASG60-03-C-0009

ABSTRACT

A wide range of raw data, unique research databases, and state-of-the-art data access tools are available to the nuclear explosion monitoring research and development (R&D) community through the Research and Development Support Services (RDSS) project operated by the United States (U.S.) Army Space and Missile Defense Command (SMDC) Monitoring Research Program (MRP).

The waveform data archive component of the RDSS has recently been re-engineered to facilitate a new class of computational experiments and to vastly improve data access. All waveforms in the archive (roughly 12 Tb of seismic, hydroacoustic and infrasound waveforms from over 300 locations world-wide) are stored in an analysis-ready format (CSS3.0) on a high-speed spinning disk mass-storage system. Every waveform in the archive can be directly and instantly accessed, making it possible to perform experiments on any desired cross-section of the archive, with no need to "stage" data or make special modifications to scientific software. Further, we have developed a new interactive, web-based tool which supports direct downloads of desired data. The web tool also supports browsing of event bulletins—when the desired event is found in a bulletin the user simply clicks through a simple on-line form to download data associated with the event. Traditional data request tools, such as the e-mail based *AutoDRM* software, not only provide backwards compatibility for users but also benefit from the new mass-storage system by providing greatly improved response times.

The RDSS provides a range of value-added R&D databases that are a significant resource for the U.S. R&D community. For example, the infrasound database draws on a unique collection of waveforms (many of which are not archived anywhere else) from infrasound arrays operated by the Department of Energy, the International Monitoring System, and other organizations. These data go back to 1995, and include data from sites on every continent. The database includes acoustic recordings of a variety of natural and man-made events, including atmospheric nuclear explosions. We are actively adding data from recent infrasound events to the database and we provide examples of a number of such events. The seismic R&D databases include: the Nuclear Explosion Database, an exhaustive collection of metadata and waveforms from nuclear tests; the Ground Truth Database, containing a wide range of carefully selected and quality-controlled events for ground truth levels of 0 to 15 km; and region-specific databases for the Lop Nor, China region, the arctic region, and the Korean Peninsula (currently in preparation). The region-specific databases are prime examples of value-added databases, as they bring together a wide range of open-source data into a single, well-organized package. All of the R&D databases are accessible through interactive web-based tools.

A multi-technology hypocenter location service (HLS) is also provided by the RDSS. This service will locate event hypocenters or epicenters based on any combination of seismic, hydroacoustic and infrasonic arrivals. The service supports both default and user-defined travel-time tables for each technology, and supports multiple input/output formats. The service is based on the *EvLoc* hypocenter location software (the software currently in operational use at the International Data Centre) and is accessed through an e-mail-based interface.

In this paper we provide a summary of the major RDSS data assets and resources. We provide examples and descriptions of the types of data and metadata in each database, and provide information on how these resources can be accessed.

OBJECTIVE

The objective of the RDSS is to support the nuclear explosion monitoring research and development community with a wide range of data, state-of-the-art data access tools, and value-added datasets.

RESEARCH ACCOMPLISHED

During the current contract we have pursued initiatives in three primary areas with the goal of extending existing resources or developing new resources which will provide direct benefit to the nuclear explosion monitoring R&D community. First, we have expanded and improved the waveform archive, concentrating especially on creating a high-speed archive and new access tools which fully exploit the capabilities of the new system. Second, we have improved our existing, value-added R&D databases and have created several new databases. Third, we have developed new tools, using an application service provider model, to provide remote users with access to sophisticated R&D software. In the following sections we describe these developments in greater detail.

Waveform Archive and Data Access

The RDSS maintains a large archive of waveform data from seismic, hydroacoustic, and infrasound stations. Data in the archive go back to 1995 (earlier for some stations) for stations distributed worldwide (Figure 1).

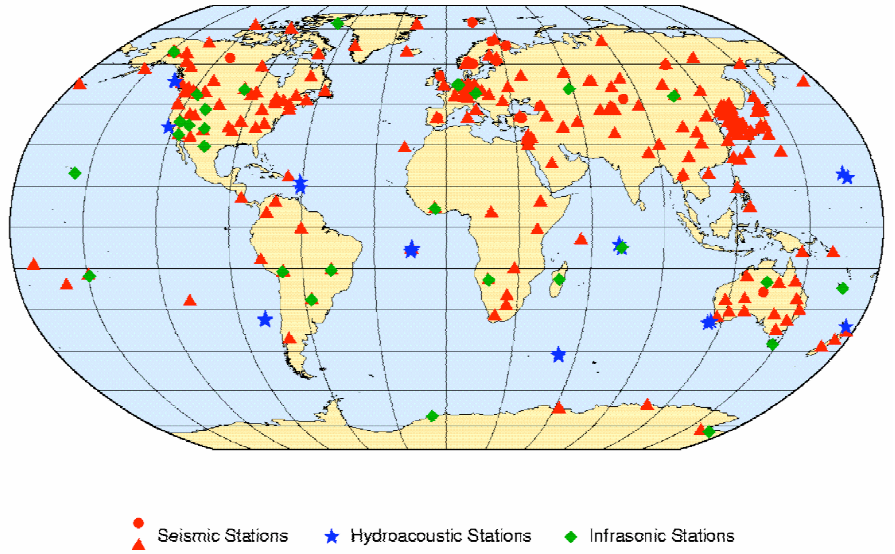

Figure 1. (right) Seismic arrays (circles) and three-component stations (triangles), and hydroacoustic and infrasonic arrays (stars and diamonds, respectively) for which waveform data are available from the RDSS archive. For many of these stations and arrays the data are continuous over periods of several years.

The archive currently contains over 12 Tb of data. To facilitate a new class of computational projects, we have migrated our data archive from a robotic mass-storage system to a system based on conventional spinning-disk technology (RAID disk farm). The current system (Figure 2) provides access to the entire archive that is at least two orders of magnitude faster than its predecessors, and makes very large scale experiments and data visualizations possible. Data on the mass-store are stored in an analysis ready format (CSS3.0) so that they may be accessed directly by application software -- making it possible to perform experiments on any desired cross-section of data in the archive, with no need to "stage" data or make special modifications to scientific software.

Figure 2. (left) This 15 Tb mass-storage system provides direct access to over 12 Tb of waveform data.

Figure 3 provides an example of the type of computations the new mass-storage system facilitates. We show the results of computing root-mean-square (RMS) pressure noise using an entire year of 20 samples per second data (~600 million samples per year per channel) from the Mongolian (I34MN) infrasound array. The results of these calculations show fluctuations in pressure noise on daily to seasonal time scales, with many of the shorter duration daily effects being directly related to wind. The software which computes the running RMS values accesses data directly from the mass-storage system on-the-fly as the computations proceed. These computations are only limited by the CPU requirements of the applications and the spinning-disk input/output speeds. Similarly, the same mass-storage system has been used for a large-scale project to compute ambient infrasonic noise for International Monitoring System (IMS) and other infrasound arrays. The noise computations are based on data samples from multiple time windows per day, for one year (or longer) time spans (Bowman et al, 2004). While the ambient noise computations are relatively straightforward, they require access to a massive quantity of data spread across both stations and time and would pose a significant challenge to traditional robotic mass-storage systems which cannot readily support direct, random access to data.

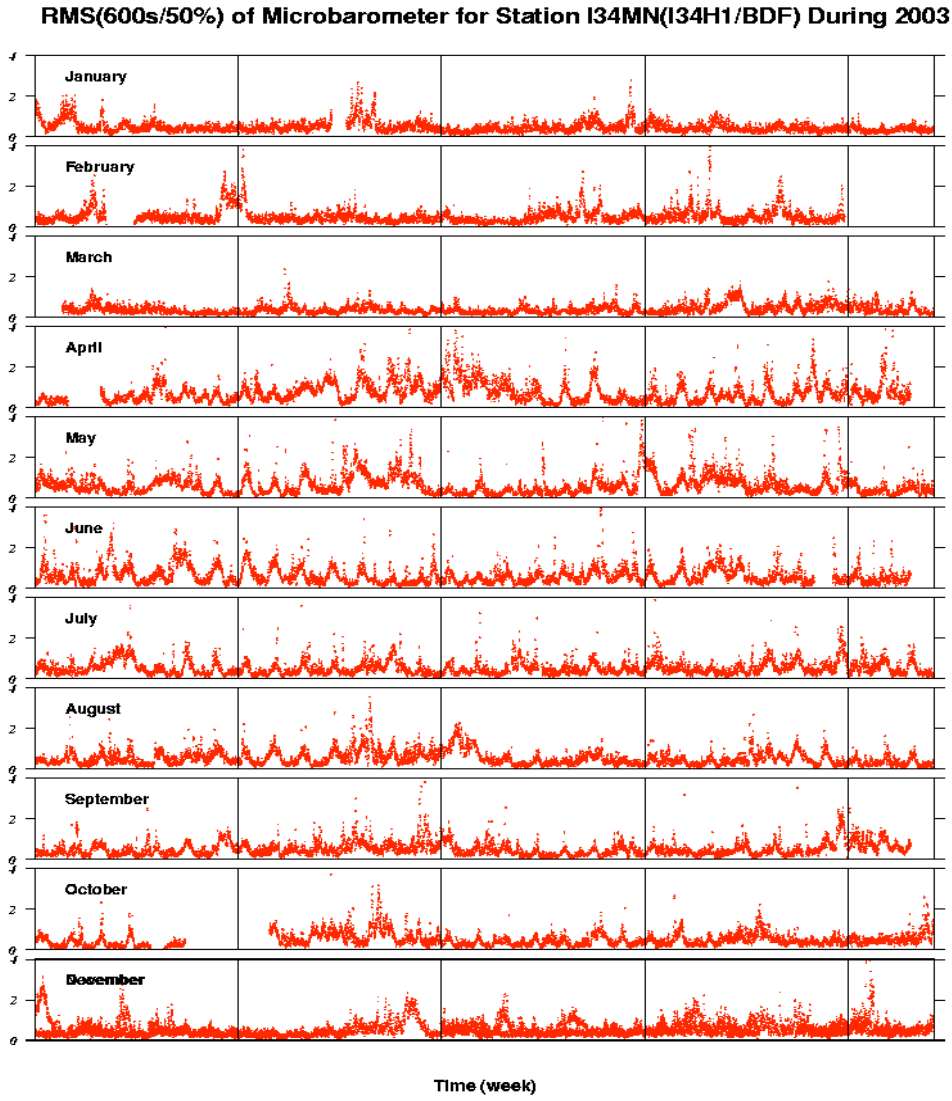

Figure 2. RMS pressure at station I34MN (Mongolia) during 2003 (from a single microbarograph channel). RMS pressure is calculated using a 600 s running window, with 50% window overlap. The pressure noise tracks wind speed fluctuations measured at the same site (and also available from the archive), with a few notable exceptions. An entire year's worth of data (600 million samples) was processed for this figure.

To enhance the value of the mass-store system for the research community we have initiated an effort to upgrade the data access tools. First, we have developed a new interactive, web-based tool which supports direct downloads of desired data. The web tool supports browsing of event bulletins—when the desired event is found in a bulletin the user simply clicks through a simple on-line form to obtain an immediate download of the data associated with the event (Figure 4). The data time windows can be of fixed length or can be adjusted relative to predicted phase arrival times or group velocities. This interactive tool also supports the download of custom data windows – based on arbitrary time windows or based on user-supplied event information. Data are downloaded immediately upon completion of the specification of the desired stations/channels/time windows and are in analysis ready formats (CSS3.0 is currently supported, SAC is being implemented). As a second major enhancement to the mass-store system, we are pre-rendering waveform plots for all events in the Reviewed Event Bulletins (REBs) produced by the Prototype International Data Center (1995-2000) and the International Data Centre (2000-present). These plots allow the user to review the waveforms utilized for each REB event, and can further assist the user in identifying waveforms which the user may want to download. Finally, we are actively developing a mechanism for users to dynamically view any waveform in the data archive, as a means of providing a "browse before you download" capability. Finally, traditional data request tools, such as the e-mail based *AutoDRM* software, provide backward compatibility for users and benefit from the new mass-storage system by providing greatly improved response times.

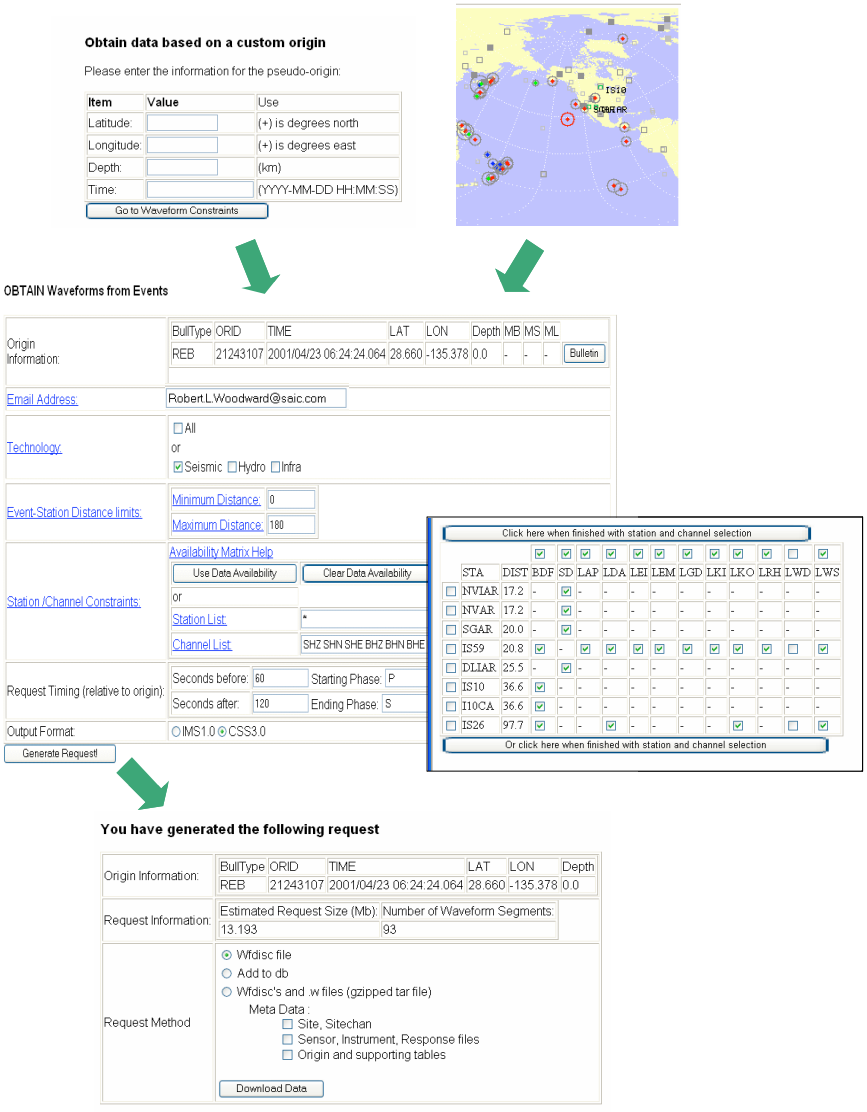

Figure 4. The web-based data access tool provides easy access to data, through both custom query and map-based interfaces. A station-channel selection matrix allows fine-tuning of the data request and data may be downloaded immediately.

Value Added Databases

The RDSS provides a range of value-added R&D databases that are a significant resource for the U.S. R&D community (Table 1). The data (waveforms and/or arrivals) which are included in these databases are extracted from the general RDSS waveform archive (described above) as well as from a wide variety of other open sources. The careful compilation of data, along with relevant metadata, is typically a challenging and time consuming task. However, the resultant data sets provide unique, value-added resources which can support studies in nuclear explosion detection, location, and identification. We describe several of the value added databases below.

Table 1. Summary of value-added research databases which are available as part of the RDSS.

Database	Description
Infrasound	A wide range of acoustic recordings from natural and man-made events, including atmospheric nuclear explosions. Data suitable for detection and source classification studies.
Nuclear Explosion	A vast compilation of seismic and infrasound recordings and arrival data from underground, underwater, and atmospheric nuclear explosions.
Ground Truth	Seismic phase arrival data for GT0 through GT15 events.
Lop Nor	Seismic waveforms (~100 GB) and phase arrival data for events and stations in the Lop Nor, China region. Includes recordings of larger nuclear explosions which have been scaled down to smaller yields and embedded in background noise. Provides multiple source location estimates obtained from different open sources. See Kohl et al (2002).
Arctic Region	Seismic waveforms and phase arrival data for events and stations in the Arctic region. Provides multiple source location estimates obtained from different open sources.
Korean Peninsula	Includes seismic waveforms, ground truth locations, and event characterization information for hundreds of events on and near the Korean Peninsula.

The infrasound database draws on a unique collection of waveforms (many of which are not archived anywhere else) from infrasound arrays operated by the Department of Energy, the International Monitoring System, and other organizations. The 27 infrasound arrays for which continuous data are currently in the database are shown in Figure 5. These data holdings go back to 1995, and include data from sites on every continent. The database includes acoustic recordings of a variety of natural and man-made events, including atmospheric nuclear explosions. We are actively adding data from recent infrasound events to the database. For example, Figure 6 shows the location of several recent events which occurred during a nine hour interval on a single day and which were recorded by the IMS infrasound station in Kazakhstan (I31KZ). The signals in the infrasound database are suitable for a variety of signal detection and source characterization studies, including the examination of such issues as seasonal propagation variability, etc.

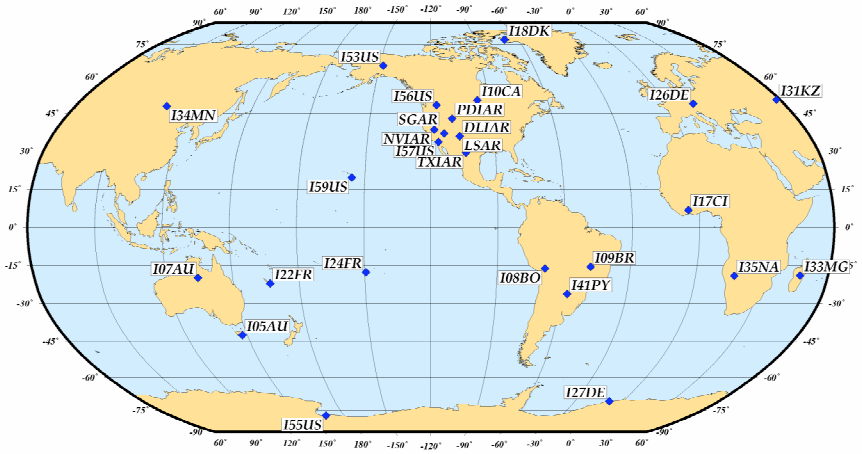

Figure 5. The infrasound database contains continuous data from the 27 infrasound arrays shown on this map.

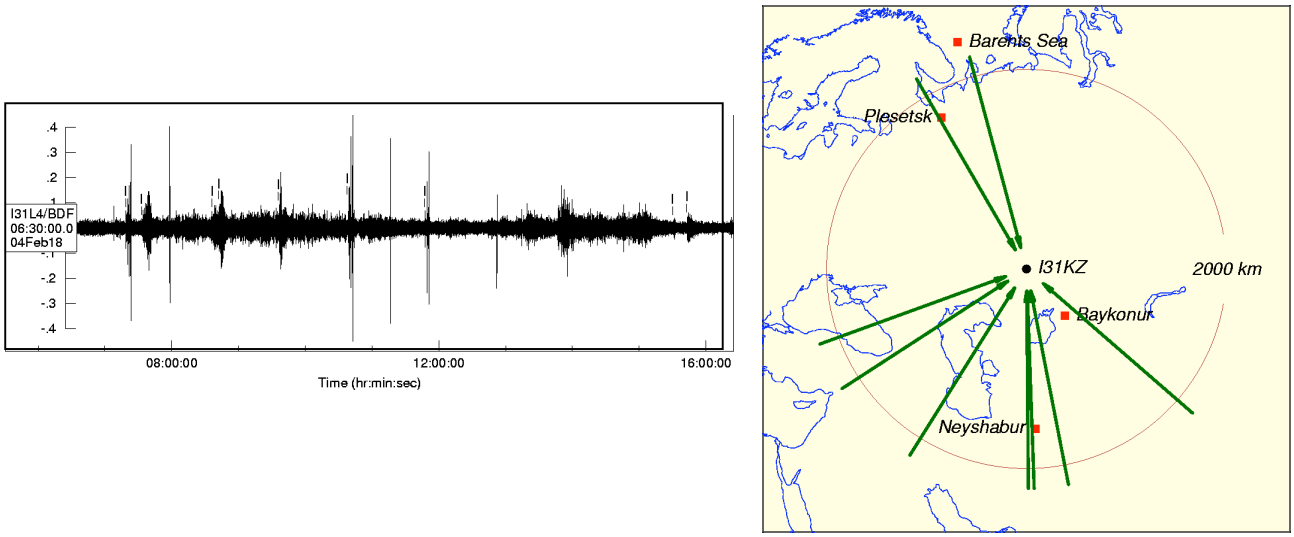

Figure 6. Multiple signals were observed at station I31KZ (Kazakhstan) during a nine hour period on February 18, 2004. The left panel shows a nine hour record recorded at I31KZ, with numerous clear signals. The right panel shows likely sources and azimuths for each of the observed signals. Signals were observed from missile launches at Baykonur, Kazakhstan and Plesetsk, Russia and the Barents Sea, as well as from a train explosion in Neyshabur, Iran.

The Nuclear Explosion Database contains information on all reported nuclear explosions in the atmosphere, underground and underwater since the first nuclear explosive device was set off in 1945 (Figure 7). The database gives the most accurate and complete unclassified information on time and place of both announced and presumed nuclear explosions and provides, whenever available, information on explosion yield, depth and shot medium (Table 2). In addition, the database contains a large archive of seismic (and some infrasound) recordings from about one third of the explosions (Figure 8).

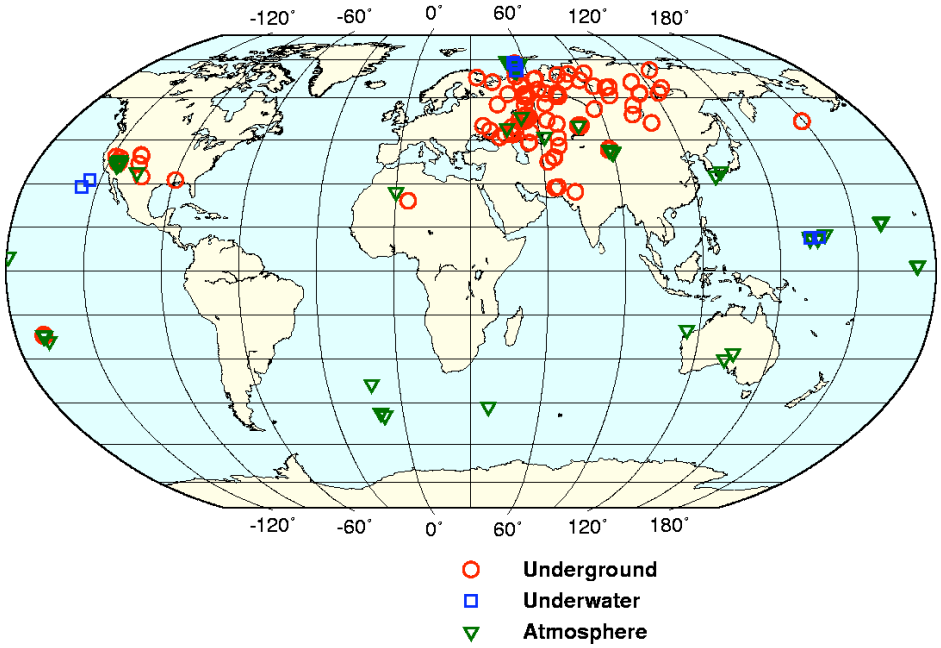

Figure 7. Locations of nuclear explosions for which information is included in the Nuclear Explosion Database.

The explosion database was originally compiled in 1997 and has subsequently been maintained and updated and it currently holds data for 2041 explosions. The list of explosions in the data base was derived from more than 40

different data sources with information reported by both official, such as Department of Energy (DOE/NV-209, 2000), and non-official organizations and appearing in a variety of publications. A detailed description of the database has been compiled by Yang et al. (2000, 2003) and Bondar et al. (2001).

Table 2. Locations of stations for which nuclear explosion waveforms (seismic or infrasound) or arrival data are included in the Nuclear Explosion Database.

Explosion Environment	Total Number of Explosions	Number of Explosions with:				
		Shot Depth/ Height	Yield	Shot Medium	Arrival data	Waveforms
Underground	1516	632	1035	457	608	739
Underwater	10	7	9			1
Atmosphere	515	110	452		7	25

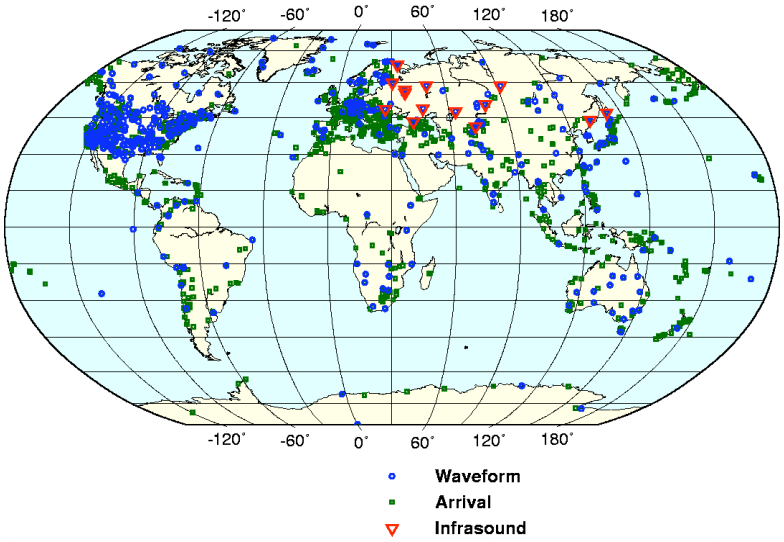

Figure 8. Locations of stations for which nuclear explosion waveforms (seismic or infrasound) or arrival data are included in the Nuclear Explosion Database.

The Ground Truth Database contains a collection of events of ground truth (GT) quality to support location calibration studies. The database includes nuclear and chemical explosions, mine blasts, rock bursts and earthquakes of global coverage. The earthquakes in the database all have shallow focus (less than 40 km focal depth), since deep earthquakes are of lesser interest from a nuclear monitoring point of view. Every event is accompanied with arrival data as well as references to document the source of information.

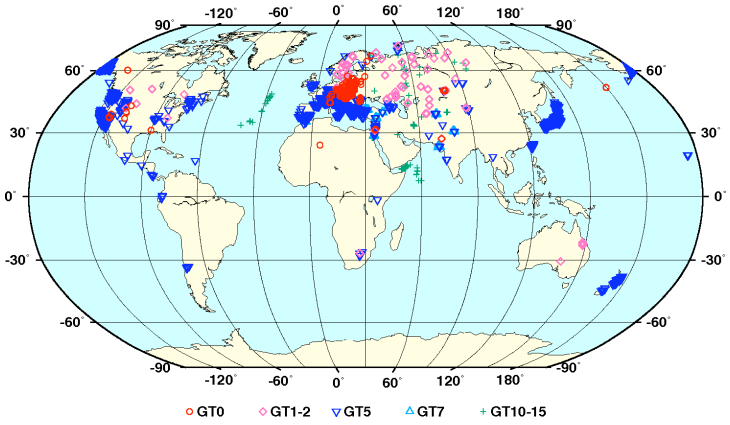

Figure 9. Locations of events in the Ground Truth Database.

The events in the Ground Truth Database are distributed globally, though they are dominantly in the northern hemisphere (Figure 9). The database currently holds 13,379 GT0-15 events (Table 3) with some 900,000 associated phases (of which some 810,000 are defining) recorded at 3,771 stations. Residuals in the bulletins refer to IASP91 (Kennett and Engdahl, 1991) predicted travel times with the source fixed to the GT locations. Note that the categories GT7 and GT11 stand for events promoted to ground truth status after performing

multiple event location on an event cluster. For details see Bondár et al. (2004), Engdahl and Bergman (2001), and Engdahl et al. (2002). The GT15 events (Pan et al., 2002) are only included to provide coverage on mid-oceanic ridge events in the mid-Atlantic ridge, the Carlsberg ridge and the Gulf of Aden.

Table 3. Summary of Ground Truth Database holdings, by source type and GT level.

Source	GT0	GT1-2	GT5	GT7	GT10	GT11	GT15	Total
Nuclear explosion	428	359	23	-	26	-	-	836
Chemical explosion	155	44	1	-	-	-	-	200
Mine blast, rock burst	-	141	58	-	-	-	-	199
Earthquake	-	-	11596	324	12	183	29	12144
Total	583	544	11678	324	38	183	29	13379

Hypocenter Location Server

The RDSS has developed the Hypocenter Location Server (HLS) to provide a remotely accessible interface to the powerful hypocenter location program *EvLoc*. The *EvLoc* software is the same software used by the International Data Centre (and by the former Prototype International Data Center) to produce Reviewed Event Bulletins. *EvLoc* is capable of using any combination of seismic, hydroacoustic, and infrasound arrivals to determine hypocenters and (seismic only) to compute magnitudes. The HLS is accessed by sending formatted e-mail messages to the server (in this sense it looks much like an *AutoDRM* server), as illustrated in Figure 10. The user provides data files, configuration files, and custom parameter files via FTP (using a flat-file structure or using XML). The HLS processes requests, with their associated input files, and then e-mails the results to the user.

EvLoc is a complex software component that relies heavily on a carefully configured database environment (e.g. Oracle) to function properly. As such, *EvLoc* is not readily portable. By providing access to *EvLoc* functionality through the Application Service Provider model as described here, researchers can readily use *EvLoc* in their analysis without worrying about software portability or configurability. Further, the capability provided by the HLS completely supercedes that provided by the stand-alone hypocenter determination program *LocSAT*. Note that the HLS is fully backward compatible with *LocSAT* input files.

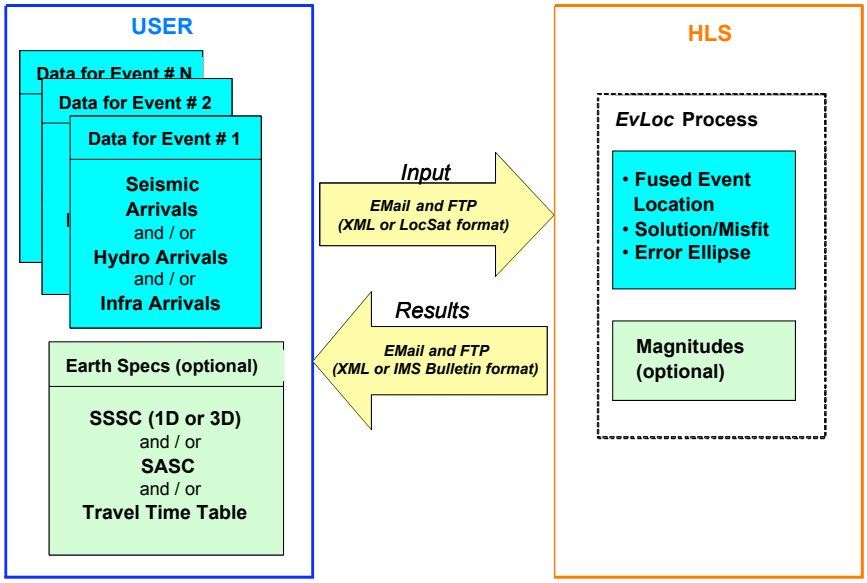

Figure 10. Illustration of the relationship between user provided inputs and HLS-produced processing results.

CONCLUSIONS AND RECOMMENDATIONS

We have described a variety of resources that are available to the U.S. nuclear explosion monitoring research and development community.

A new, state-of-the-art mass-storage system makes a vast quantity of data available for direct computations and rapid remote access. With this new system it is possible to perform computations in which data are accessed directly from the mass-store and to select and instantly access data across any combination of stations and time windows. Sophisticated new tools are provided to enable remote users to browse the data holdings and obtain instant downloads of data in analysis-ready formats.

A wide range of value added databases have been produced. These databases represent unique assemblies of waveform data, arrival and event data, and metadata. The databases are all available on-line, and provide valuable resources for use in detection, location, and identification studies.

To support remote users who require event location capability, we provide an application service provider interface to the powerful event location software *EvLoc*. *EvLoc* can provide fused event location solutions (using any combination of seismic, hydroacoustic, and infrasonic arrivals). The interface to *EvLoc* can support the user with either small, occasional event location requirements, or the user performing bulk relocations as part of a calibration project. In either case, there is no need for the user to install and configure the complex database environment that software like EvLoc requires.

We recommend that researchers in the nuclear explosion monitoring R&D community visit the MRP web sites to learn more about the MRP's RDSS resources. The MRP maintains an open web site and a restricted access web site, limited to U.S. government-authorized researchers. As an additional convenience to authorized users, the MRP web site may also be accessed using the SMDC channel of the *Army Knowledge Online* web portal at https://www.us.army.mil.

ACKNOWLEDGEMENTS

We thank Rod Whitaker of Los Alamos National Laboratory (LANL) for making available the data from the infrasound arrays that LANL has operated in the western U.S. for many years. We thank Hans Israelsson and Istvan Bondar for their contributions to this paper.

REFERENCES

Bowman, J.R., E. Baker, and M. Bahavar (2004), Infrasound Station Ambient Noise Estimates (2004), *Proceedings of the 26th Seismic Research Review: Trends in Nuclear Explosion Monitoring*, Orlando, FL (Sept. 21-23, 2004).

Bondár, I., E.R. Engdahl, X. Yang, H.A.A. Ghalib, A. Hofstetter, V. Kirichenko, R. Wagner, I. Gupta, G. Ekström, E. Bergman, H. Israelsson and K. McLaughlin, (2004), Collection of a reference event set for regional and teleseismic location calibration, *Bull. Seism. Soc. Am.*, in press.

Bondár, I., X. Yang, R.G. North and C. Romney (2001), Location Calibration Data for CTBT Monitoring at the Prototype International Data Center, *Pure Appl. Geophys*., 158, 19-34.

DOE/NV-209 (2000), United States nuclear tests, July 1945 through September 1992, Rev. 15.

Engdahl, E.R. and E.A. Bergman (2001), Validation and generation of reference events by cluster analysis, in *Proceedings of the 23rd Seismic Research Review: Worldwide Monitoring of Nuclear Explosions*, LA-UR-01-4454, Vol. I, pp. 205-214.

Engdahl, E., E. Bergman, M. Ritzwoller, N. Shapiro and A. Levshin (2002), A reference event data set for validating 3-D models, in *Proceedings of the 24th Seismic Research Review - Nuclear Explosion Monitoring: Innovation and Integration*, LA-UR-02-5048, Vol. I, pp. 261-270.

Kennett, B. and E.R. Engdahl (1991), Travel times for global earthquake location and phase identification, *Geophys. J. Int*., 105, 429-465.

Kohl, B., R. North, J.R. Murphy, M. Fisk, and G. Beall, (2002) Demonstration of Advanced Concepts For Nuclear Test Monitoring Applied to the Nuclear Test Site at Lop Nor, China, in *Proceedings of the 24th Seismic Research Review - Nuclear Explosion Monitoring: Innovation and Integration*, LA-UR-02-5048, Vol. I, pp. 302-312.

Pan, J., M. Antolik and A. Dziewonski (2002), Locations of mid-oceanic earthquakes constrained by sea- floor bathymetry, *J. Geophys. Res.*, 107, 2310, doi:10.1029/2001JB001588.

Yang, X., R. North, and C. Romney (2000), CMR Nuclear Explosion Database (Revision 3), *CMR Technical Report CMR-00/16*.

Yang, X., R. North, C. Romney, and P. Richards (2003), Worldwide nuclear explosions, in *International Handbook of Earthquake and Engineering Seismology*, Vol. 81B, eds. W.H. Lee, H. Kanamori, P. Jennings, and C. Kisslinger, *Academic Press*.