# DATA AND TOOLS TO SUPPORT NUCLEAR EXPLOSION MONITORING RESEARCH AND DEVELOPMENT

Robert Woodward, Michael Skov, Manochehr Bahavar, Geoff Davis, and Yu-Long Kung

Science Applications International Corporation

Sponsored by Army Space and Missile Defense Command

Contract No. DASG60-03-C-0009

## ABSTRACT

The nuclear explosion monitoring research and development (R&D) community can access raw data and unique research databases using state-of-the-art data browsing and selection tools as part of the Research and Development Support Services (RDSS) project operated by the United States (US) Army Space and Missile Defense Command Monitoring Research Program.

The waveform data archive maintained by the RDSS provides direct and immediate access to 100% of the data in the archive. A high-speed spinning disk mass-storage system is used to store all the waveforms in the archive (roughly 12 Tb of seismic, hydroacoustic and infrasound waveforms from over 300 locations worldwide). Every waveform in the archive can be directly and instantly accessed, making it possible to provide data users with a range of new tools and services that go beyond traditional data center functions. User interaction with the data archive is provided through several mechanisms, including a web-based, “e-commerce-like,” interface. This web-based interface allows users to visually browse data (waveforms) or data products (e.g., bulletins), select data or data products to be placed in a “shopping cart,” and then provides the capability to manage and download the selected objects. A waveform viewer runs under a standard web browser and provides the capability to display multiple traces, perform filtering, zooming, and scrolling. Any waveform in the entire data archive can be viewed using this tool. The user can select waveforms from the display for subsequent download. A download manager then provides the ability to download the various products a user has selected during a session (or sessions—the download manager can manage product selections made during multiple sessions). Any data in the archive can be downloaded in either CSS3.0 or SAC formats.

The waveform data archive also supports data intensive computing by providing direct access to all data files in the archive via RDSS servers. Users who load software on these servers can perform experiments on any desired cross-section of the waveform archive, with no need to “stage” data. Further, it is not necessary to make special modifications to scientific software, since the data are stored in the archive in a common, analysis ready format. An example of a recent large-scale, data intensive experiment was the analysis of ambient noise at infrasound stations. This experiment required computing noise values at multiple infrasound stations, at multiple times per day, for every day of a 2-year period. Over a million discrete (noncontiguous) time series were analyzed for this study.

The RDSS also provides a range of value-added R&D databases, which are a significant resource for the US R&D community. These databases are accessible through interactive, web-based tools and bring together a wide range of open-source data into individual, well-organized packages. For example, the recently expanded 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, as well as historical recordings of atmospheric nuclear explosions. 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 (GT) levels of 0 to 15 km; and region-specific databases for the Lop Nor, China, region and the Arctic region.

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).

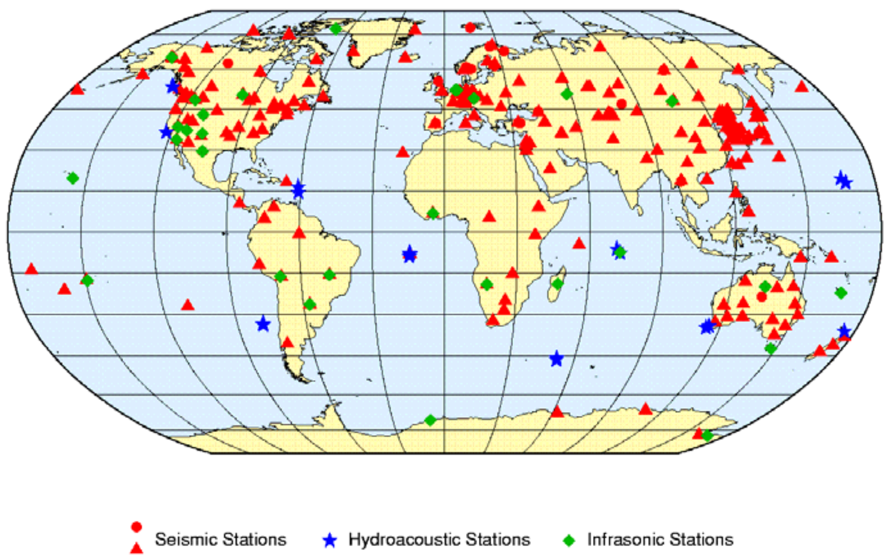


**Figure 1. Seismic arrays (circles) and three-component stations (triangles) and hydroacoustic and infrasonic arrays (stars and diamonds, respectively) for which waveform data are available from the RDSS archive. For many of these stations and arrays the data are continuous over periods of several years.**

The waveform archive now exceeds 14 Tb of data. The entire data archive is hosted on a spinning-disk mass storage system (RAID disk farm). This system 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.

The RDSS website is continuously enhanced to improve data access for the research community. During the past year we have initiated work to integrate the suite of tools and interfaces available on the web site into an "e-commerce-like" interface. The goal of this effort is to provide a unifying theme for all user interactions (i.e., data selection, data downloads, etc.) with the site that is based on concepts familiar from e-commerce (on-line shopping) websites. The goal of this web model is to allow users to visually browse data (waveforms) or data products (e.g., bulletins), select data or data products to be placed in a "shopping cart," and then provide the capability to manage and download the selected objects.

We have substantially improved the web-based waveform viewing capability. Last year we introduced an online waveform viewing capability which allowed users to view any waveform in the entire data archive. This capability has been extended to provide mult-itrace support and enhanced waveform display control. The updated waveform viewer (Figure 2) allows users to simultaneously view multiple channels of data from a single station or to view data channels from different stations. The waveform tool is not intended as an analysis tool, but rather as a waveform browser to streamline the process of locating, previewing, selecting, and acquiring waveforms of interest. However, basic scroll, zoom and filter functions are available. The waveform viewer is accessed through a standard web browser—no special installations or modifications on the user's (client's) machine or web browser are required.

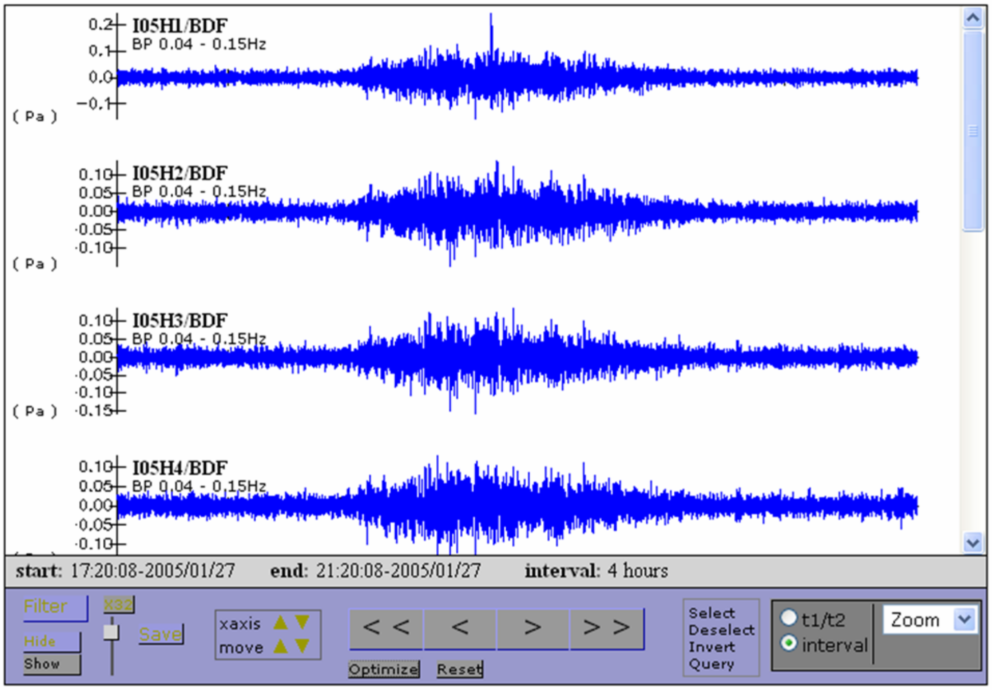


**Figure 2. The web-based data viewing tool provides easy access to all data in the archive. Data on display are from station I05AU in Australia and show signals from the eruption of a volcano on Manam Island at approximately 38° distance. The data have been bandpass filtered between 0.04 and 0.15 Hz.**

Once the desired data have been identified, the user "selects" these data (equivalent to placing items in a shopping cart on an e-commerce website). Data selections are stored in a private folder for later review, revision, and download (Figure 3). User preferences for data format and delivery method are recalled automatically, greatly reducing the effort required to obtain waveform data. For example, both CSS3.0 and SAC formats are available for waveform data. Multiple data selections may be packaged together into a single download and once the content of a particular data package is defined, the user may return to browsing while the data are collected and converted to the user's preferred format. A given data package may be downloaded more than once and prior downloads are recorded in a history folder, freeing researchers from the need to maintain permanent data warehouses on their computer.

An enhanced waveform data location service is in development as a replacement for the existing General Data Extractor (GDE) web service. Like the existing GDE, this tool will permit users to locate waveform data based on event information and other constraints such as group velocity limits and preferred station and channel selections. The replacement tool is enhanced to provide integration with both the new private folder infrastructure and the improved waveform browser—enabling the user to preview GDE-based waveform selections as well as manage selections via the user's private data folder.

A major goal in the development of these new tools is to support collaborative research by distributed research teams. Research teams might identify and define relevant data sets to be shared by members of the team. These data selections would then be available to any team member at any time without each individual requiring a personal copy of all data. Researchers would need only to download to their personal workstations those data subsets relevant to the current task. These downloaded data could be removed and replaced with other research team data as needed, freeing each team member from managing individual data sets.

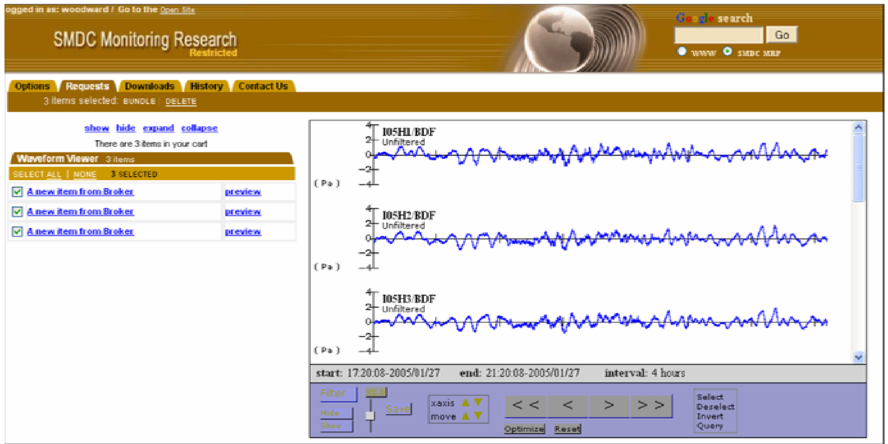


**Figure 3. Using the data download manager, the user can review data in a "shopping cart." In this example, the data selected with the waveform viewing tool (e.g., the data shown in Figure 2) are being reviewed prior to downloading (note that in this view the data are shown unfiltered).**

In addition to the new web-based data access tools described above, we provide backward compatibility for existing tools, such as the e-mail based *AutoDRM* software. In many cases the performance of these existing tools is greatly improved because of the quick response times of the spinning-disk mass-storage system.

## Data-Intensive Computing

It is possible to perform massive, data-intensive experiments directly on the data in the archive. This is possible since the archived data are written on the mass-storage system in an analysis ready format (CSS3.0). The data can be read 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 in advance.

Recent studies of infrasound ambient noise (Bowman et al., 2005a, b) provide an excellent example of the types of data-intensive computations facilitated by the high-speed mass-storage system. To characterize ambient infrasound noise, power spectral density was measured for 28 of the 34 infrasound stations shown in Figure 4. Data were analyzed from January 20, 2003, through December 31, 2004, from 21 consecutive 3-minute segments of 20 sample per second data taken four times daily, beginning at 06:00, 12:00, 18:00, and 24:00 local time. This required computing power spectra for 1,476,309 data segments (Figure 5), comprising approximately 5.3 billion data samples. In a companion study (Bowman, 2005a, b), average meteorological conditions (wind speed, wind direction, temperature, and absolute pressure) were calculated for the same 3-minute windows at 21 of the IMS stations. This computation utilized approximately three million data segments.

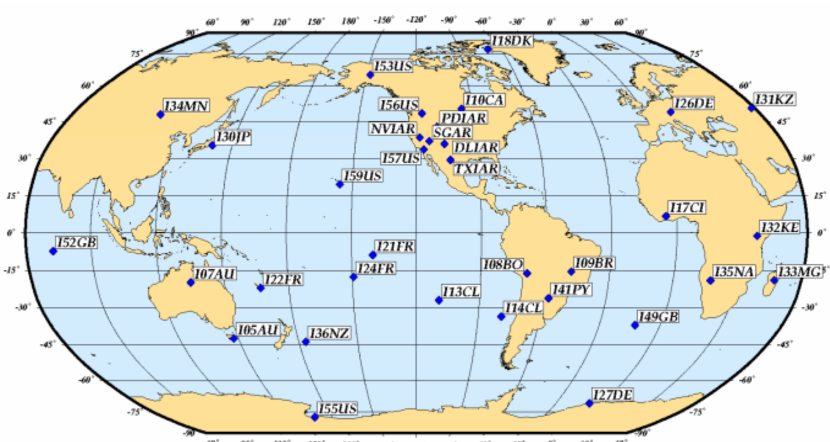


**Figure 4. Data from 34 current infrasound stations are in the RDSS data archive.**

The RDSS has a special server configuration to provide data intensive computing capability to the nuclear explosion monitoring R&D community. Remote users access a server which has direct (read-only) file access to the entire spinning-disk mass-storage system. Users essentially upload their code to the server rather than download data. The UNIX server is equipped with compilers, database access, and local storage for writing the results of computations. The user's code runs on the server and can read any waveform data in the archive (all data files are in CSS3.0 format). Researchers interested in utilizing this service should contact the RDSS.

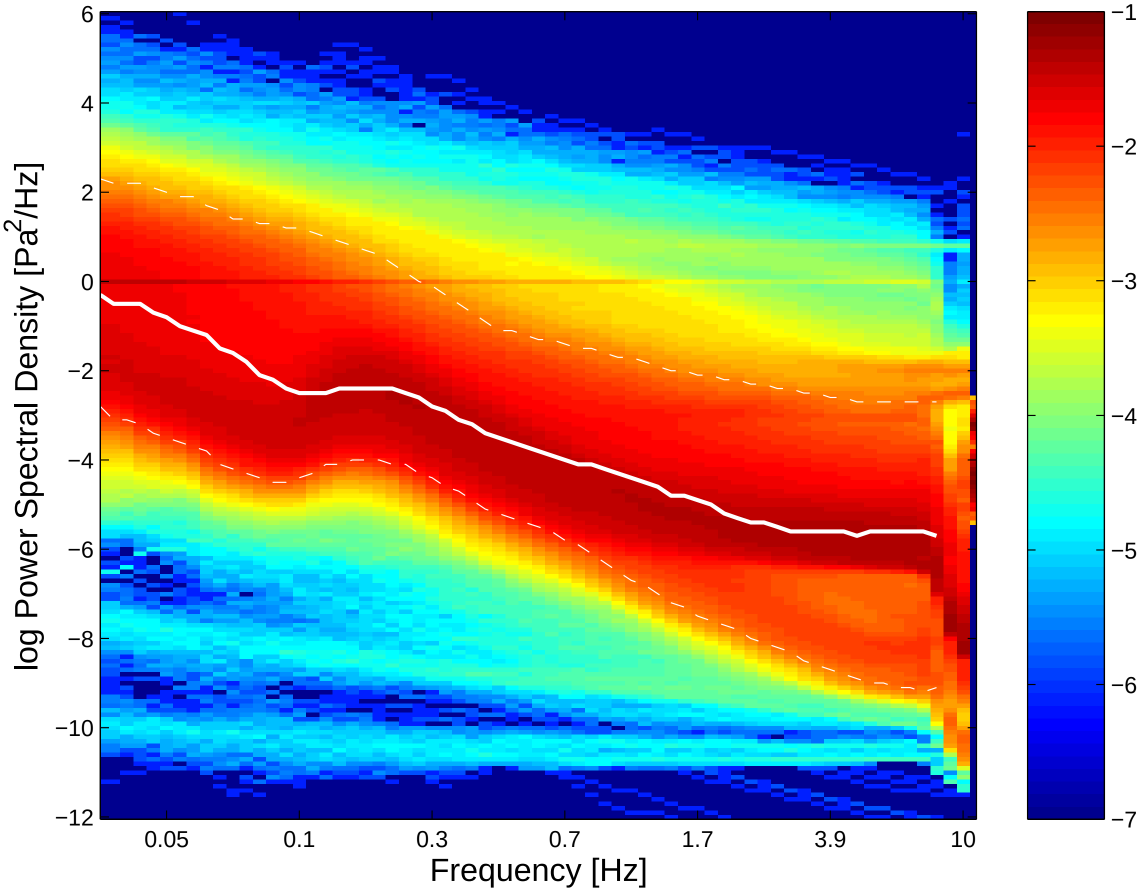


**Figure 5. Probability density function of Power Spectral Density (PSD) for all time intervals at 25 infrasound stations. About 1.4 million spectra are binned in intervals of 0.1 log PSD. The color scale is proportional to the log of the number of spectra that fall within each bin. The solid white line shows the all-time network median for 15 stations having data for 1 year or longer, and the lower and upper dashed white lines show the low- and high-noise models, respectively (figure from Bowman et al., 2005b).**

## Value-Added Databases

The RDSS provides value-added R&D databases which are a significant resource for the US R&D community (Table 1). The data (waveforms and/or arrivals) 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. Users are encouraged to visit the Monitoring Research Program's RDSS web pages (http://www.rdss.info) for more details on all of the value-added databases described in Table 1.

**Table 1. Summary of value-added research databases available through the RDSS.**

| Database | Description |
|---|---|
| Infrasound | 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. |
| Network Data Set | Database of event and waveform data for seismic and infrasound stations to support evaluation of detection and network processing performance. |

The infrasound database has been significantly expanded during the past year. The database contains waveforms from infrasound arrays operated by the Department of Energy, the International Monitoring System, and other organizations (many of which are not archived anywhere else). The 34 infrasound arrays for which continuous data are currently in the database are shown in Figure 4. These data holdings go back to 1995 and include data from sites on every continent. The database also includes a research component, which contains recent and historic acoustic recordings of a variety of natural and man-made events, including atmospheric nuclear explosions (Table 2 and Figure 6). The signals in the infrasound database are suitable for a variety of signal detection and source characterization studies, such as the examination of such issues as seasonal propagation variability.

**Table 2. Summary of events in the Infrasound Research Database. Numbers in black indicate current database holdings. Numbers in red represent data recently added, and numbers in blue indicate events to be added (i.e., in the queue).**

| Event Source Type | Events | Events with Arrivals | Arrivals | Station Waveforms |
|---|---|---|---|---|
| Nuclear explosion | 63 (31) (8) | | | 250 (110) |
| Chemical explosion | 14 (4) (2) | 10 (1) | 17 (2) | 26 (6) |
| Gas pipe explosion | 3 (1) | 2 | 3 | 5 (2) |
| Mine explosion | 5 (5) (13) | 5 (5) | 7 (7) | 5 (5) |
| Rocket launch | 7 | 7 | 10 | 10 |
| Bolide | 19 (7) (8) | 10 | 18 | 43 (19) |
| Rocket reentry | 4 | 4 | 7 | 7 |
| Volcano | 29 (3) | 1 | 1 | 29 |
| Earthquake | 3 (15) | 3 | 4 | 4 |
| Aircraft (not sonic boom) | 1 | 1 | 1 | 1 |
| Auroral wave | 48 | – | – | 48 |
| Gravity wave | 5 | – | – | 5 |
| Microbarom | 29 | – | – | 29 |
| Mountain wave | 49 | – | – | 49 |
| Unknown | 18 | 14 | 14 | 18 |
| **Total** | 297 (48) (49) | 57 (6) | 82 (9) | 529 (142) |

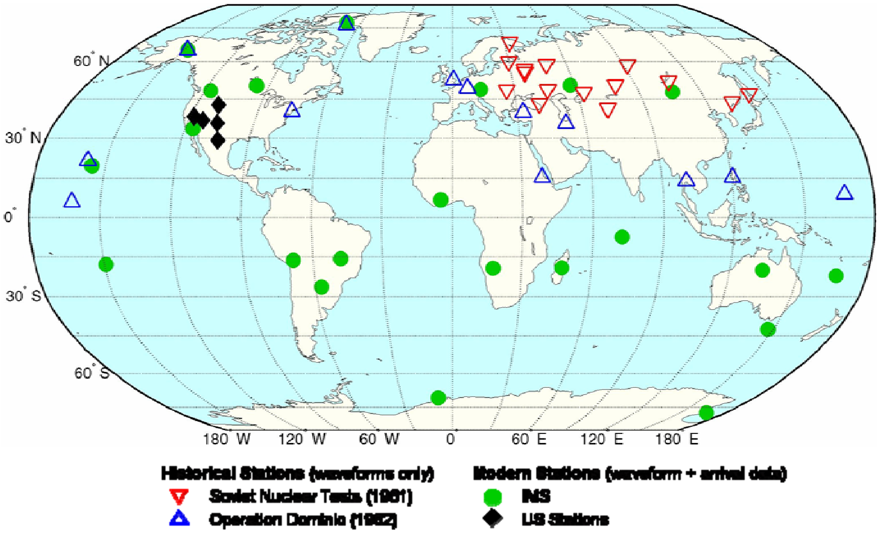


**Figure 6. The Infrasound Research Database contains waveform data from a variety of historic and current station locations.**

The Network Data Set (Kohl et al., 2005) provides a comprehensive database for evaluating detection and network processing performance. This "synthesized" network data set provides the means to evaluate signal processing algorithms on a large data set which has completely known characteristics. The data set utilizes seismic and infrasound signals from actual events. The signals are scaled to various sizes and embedded in clean background noise. Figure 7 shows the locations of the stations and events currently in the Network Data Set.

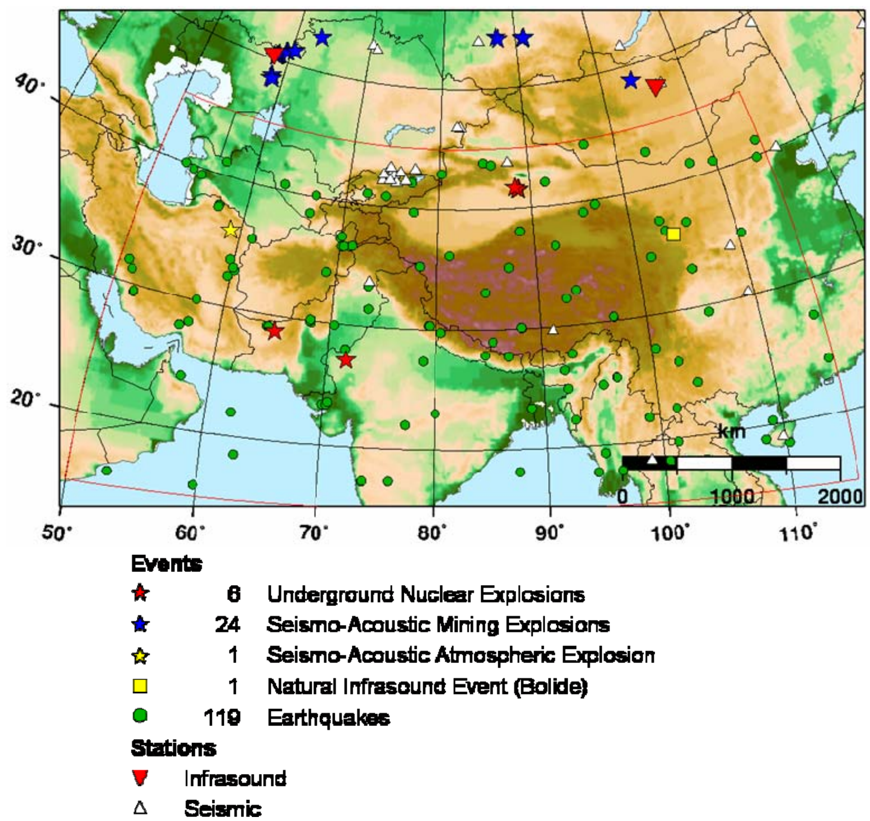


**Figure 7. Locations of events and stations currently in the Network Data Set.**

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. 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 3). In addition, the database contains a large archive of seismic (and some infrasound) recordings from about one-third of the explosions (Figure 8).

The explosion database was originally compiled in 1997 and has subsequently been maintained and updated. 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 Bondár et al. (2001).

**Table 3. 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 |

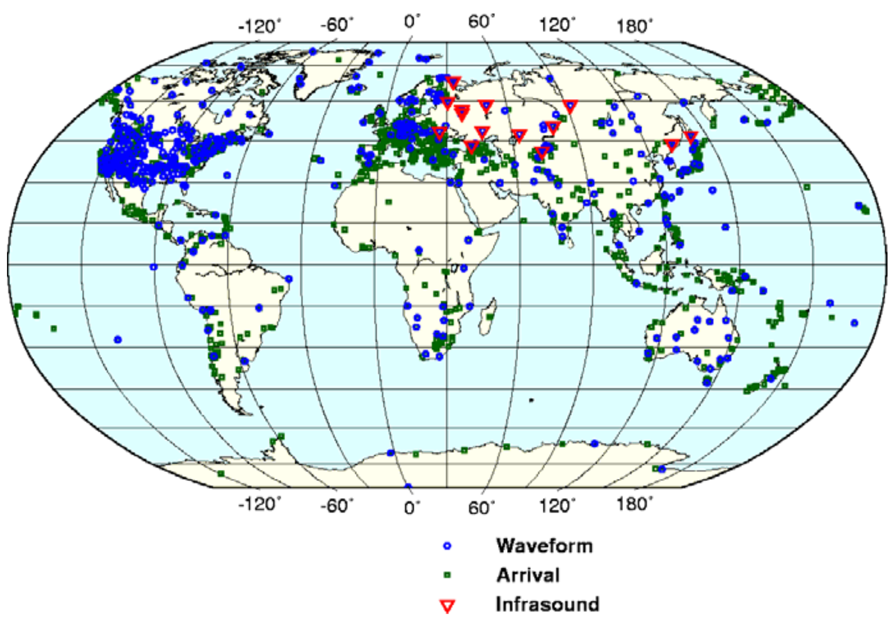


**Figure 8. Locations of stations for which nuclear explosion waveforms (seismic or infrasound) or arrival data are included in the Nuclear Explosion Database.**

The Ground Truth Database contains a collection of events of 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.

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 4) 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 GT 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 included only to provide coverage on mid-oceanic ridge events in the mid-Atlantic ridge, the Carlsberg ridge and the Gulf of Aden.

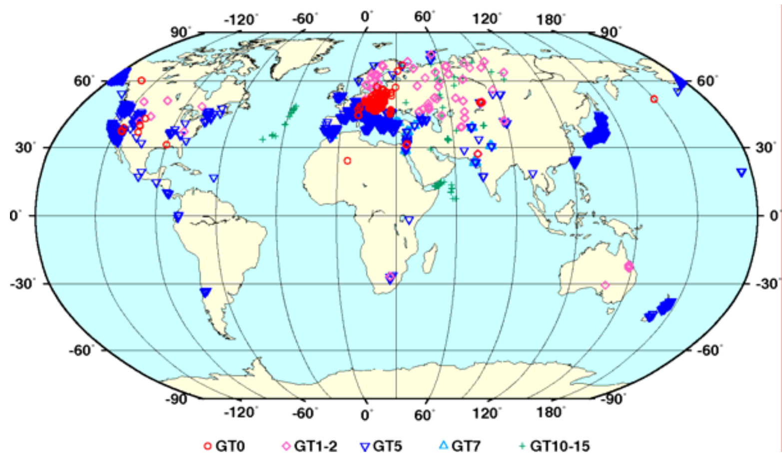


**Figure 9. Locations of events in the Ground Truth Database.**

**Table 4. 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 |

## CONCLUSIONS AND RECOMMENDATIONS

We have described a variety of resources which are available to the US nuclear explosion monitoring research and development community.

A vast quantity of data is available to users via the spinning disk mass-storage system. New tools have been developed which make it possible to visually browse any waveform in the entire data archive. The waveform viewer runs under any standard web browser and provides the ability to view multiple channels of data from one or more stations. Data can be selected via the waveform viewer or with other tools. Selected data can then be downloaded via the data download manager, which provides user-specific configuration and download management.

The mass-storage system also supports data intensive computing experiments. Application software can access data directly from the mass-store—facilitating experiments which require large volumes of data from any combination of stations and time windows. Researchers interested in remotely accessing the data intensive computing infrastructure should contact the RDSS (e.g., via the website).

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.

We recommend that researchers in the nuclear explosion monitoring R&D community visit the MRP websites to learn more about the MRP's RDSS resources. The MRP maintains both an open website (http://www.rdss.info) and a restricted access website (at https://www.rdss.info; access limited to US government-authorized researchers).

## ACKNOWLEDGEMENTS

We thank the members of the US Infrasound Team for making available high-quality data from the research stations and International Monitoring System stations which they operate.

## REFERENCES

Bowman, J. R. (2005a), Meteorological Conditions at Infrasound Stations, *Inframatics,* 9.

Bowman, J. R. (2005b), Meteorological Conditions at Infrasound Stations, SAIC Technical Report SAIC-05/3000.

Bowman, J. R., G. E. Baker, and M. Bahavar (2005a), Ambient Infrasound Noise, *Geophys. Res. Lett.*, 32: L09803, doi:10.1029/2005GL022486.

Bowman, J. R., G. Shields, G. E. Baker, and M. Bahavar (2005b), Infrasound Station Ambient Noise Estimates and Models, SAIC Technical Report SAIC-05/3001.

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.* 94: 1528–1545.

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. 1, 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: 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. 1, pp. 302–312.

Kohl, B., T. J. Bennett, I. Bondár, B. Barker, W. Nagy, C. Reasoner, H. Israelsson, P. Piraino (2005), Development of a Network Data Set for Evaluating Detection and Network Processing Performance, in current Proceedings.

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, W. H. Lee, H. Kanamori, P. Jennings, and C. Kisslinger, Eds., Academic Press.