LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-CONF-440051

Enhancing Seismic Calibration Research Through Sortware Automation and Scientific Information Management

D. A. Dodge, M. D. Ganzberger, T. F. Hauk and S. D. Ruppert

July 1, 2010

Monitoring Research Review
Cocoa Beach, Florida, United States
September 21-23, 2010

Disclaimer

ENHANCING SEISMIC CALIBRATION RESEARCH THROUGH SOFTWARE AUTOMATION AND SCIENTIFIC INFORMATION MANAGEMENT

Douglas A. Dodge, Michael D. Ganzberger, Teresa F. Hauk, Stanley D. Ruppert
Lawrence Livermore National Laboratory

Sponsored by National Nuclear Security Administration
Office of Nonproliferation Research and Development
Office of Defense Nuclear Nonproliferation

Contract DE-AC52-07NA27344

ABSTRACT

The National Nuclear Security Administration (NNSA) Ground-Based Nuclear Explosion Monitoring Research and Development (GNEMRD) Program at LLNL continues to make significant progress enhancing the process of deriving seismic calibrations and performing scientific integration, analysis, and information management with software automation tools. Our tool efforts address the problematic issues of very large datasets and varied formats encountered during seismic calibration research. New information management and analysis tools have resulted in demonstrated gains in efficiency of producing scientific data products and improved accuracy of derived seismic calibrations.

This year most of our development work has been directed toward additional automation of our data ingestion capabilities. Our specific automation methodology and tools improve the researchers' ability to assemble quality-controlled research products for delivery into the NNSA Knowledge Base (KB). The software and scientific automation tasks provide the robust foundation upon which synergistic and efficient development of GNEM R&D Program seismic calibration research may be built. Three new Java programs, BulletinLoader (seismic bulletins), SegmentLoader (seismic waveforms), and MTLoader (moment tensor information) have been developed that replace older C/C++ codes and extend the capabilities of those codes to reduce the amount of required human supervision. This allows resources to be focused on analysis and production of calibration products.

The BulletinLoader tool is a Java replacement for the C++ program ORLOADER which we have been using for many years to load earthquake catalogs. Currently, BulletinLoader supports 16 bulletin formats. More importantly, it runs autonomously, retrieving bulletins as they become available on the Internet. BulletinLoader also can retrieve data from remote databases, and can be configured to automatically retrieve such data as they are added to the remote database. The SegmentLoader tool is a new Java application that replaces existing (C/C++) applications DDLOAD and UPDATEMRG. SegmentLoader loads waveform data from various flat file formats as well as from remote databases. It has the capability to cut segments from continuous sources or to load segments that have been created by other means. SegmentLoader leverages part of its code base with the WaveformRetriever and extends its functionality. In addition, we are creating a generic application layer that will allow clients to identify and retrieve needed bulletin and waveform data, load data as required into our database, and then display the new data in the client. The moment tensor loading tool (MTLoader) utilizes the recently-implemented NNSA schema (G.Ichinose, personal communication, 30 June 2010) for moment tensors. The tool allows loading of moment tensor information in a number of formats and can be run autonomously to retrieve moment tensor information from external network sources as soon as data become available. The Surface Wave Amplitude Processor (SWAP) tool produces high quality amplitude measurements of surface wave amplitudes. It leverages the NNSA moment tensor schema and is our first client software that uses the NASA World Wind software (http://worldwind.arc.nasa.gov/java/) for interactive mapping capability that produces fully interactive, 3-D imagery-based maps with excellent resolution.

These new applications comprise a significant re-engineering of our data management software suite designed to leverage new data distribution technologies that have been developed by data providers in the last few years. Exploitation of the new methods has resulted in demonstrated gains in efficiency and reduction in time required to identify, acquire, process raw data essential for calibration research.

OBJECTIVES

As the GNEM database and supporting software have matured, we have continually found more opportunities for automating pieces of our processing flow. Initially, our efforts were devoted mainly to replacing data-loading scripts with software specialized for the task of loading waveforms and bulletins. Later we developed automated procedures for identifying and publishing ground-truth information in newly ingested data.

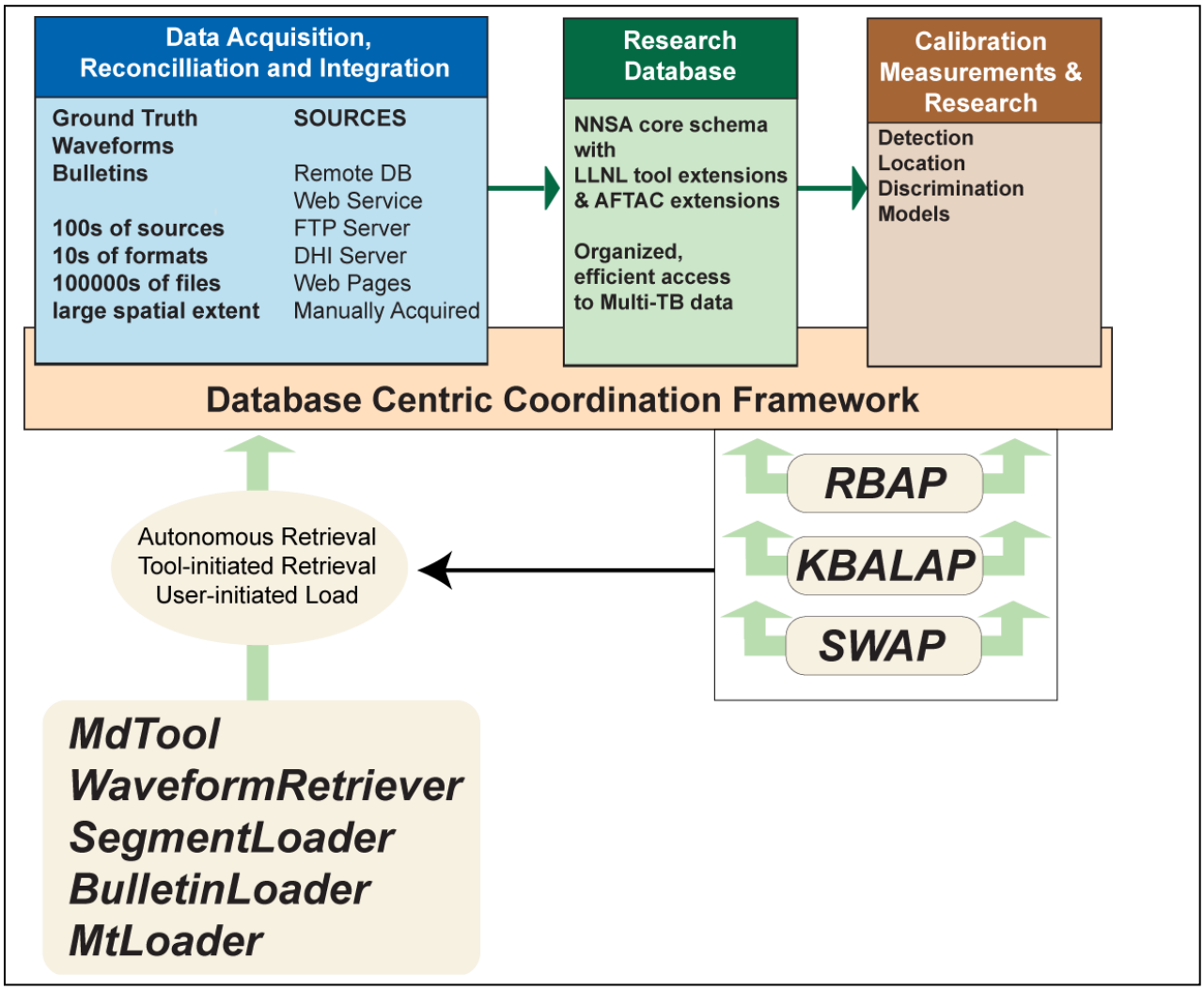

Figure 1. Software support for automated ingestion and integration of seismic data in bulk mode and as requested by analysis tools

Still later, we developed analysis tools such as RBAP for amplitude measurement and KBALAP for travel-time measurement to help researchers produce calibration information from the steadily-growing raw data stored in the database.

Over the last two years we have shifted our focus back to the beginning of our data flow and worked on codes that help automate both the retrieval and ingestion of raw data. The bulk of such efforts this year have been on the development of the BulletinLoader, SegmentLoader, and MtLoader tools. In addition, we have begun development on a new client application for measuring surface wave amplitudes.

RESEARCH ACCOMPLISHED

BulletinLoader

The BulletinLoader program is a Java program used for loading earthquake bulletin data into the LLNL schema. It is a replacement for the ORLOADER program, but it also includes a number of new capabilities. These include

- Raw bulletin (text) data is now archived in the database for all single-file source types. These data may be accessed on a per-orid basis using the ArchiveReader program.
- Bulletins may be retrieved and loaded directly from a number of public internet sources.
- The program may be run in an autonomous mode where it continually checks for new content.
- Events may be loaded directly from an external database account, possibly residing on a remote instance.
- Events may be loaded automatically from a remote instance based on entries in the LLNL.REMOTE_DB_EVID_MAP table. The map table is populated by a DBMS job, and automatically gets updated as new events are added in the remote database. This option allows the LLNL schema to be synchronized with the remote instance.
- Extensive QC checks are applied to new bulletin data prior to insertion into the LLNL schema. Where possible, BulletinLoader will modify invalid data as required to make it compatible and consistent. Where this is not possible, the program will pause, inform the user of the problems requiring attention, and resume when the problems have been addressed.
- BulletinLoader can be set to notify selected email addresses when data meeting certain criteria are loaded.

Loading Data from Internet Sources
BulletinLoader is able to retrieve and load bulletin data from several Internet sources. The currently-supported sources are:

- NEIC
    - Weekly and Monthly PDE
    - Weekly and Monthly EDR
    - M2.5+ earthquake data feed
    - Hourly and 7-day latest earthquakes
    - US Mines bulletins
- ISC Monthly bulletins
- NORSAR NORSAR Reviewed Bulletins
- University of Helsinki Regional Bulletins

BulletinLoader remembers what it has retrieved previously from each of these sources, so when you direct BulletinLoader to retrieve data from Internet sources, it only brings back new data. In addition, BulletinLoader keeps track of the last time it visited a source, and for those that only update occasionally, if enough time has not elapsed since the last visit, the source is skipped.

QC Operations
BulletinLoader attempts to ensure that all the data it loads are consistent and meet the constraints in the LLNL schema while at the same time being robust to inconsistent or unexpected data values. Rather than loading data directly into the LLNL schema, BulletinLoader first loads into stage tables in the ORLOADER schema. These tables have no column constraints, so unexpected data will not cause a constraint violation to occur. Next, the program checks the staged arrivals against the SITE table in the LLNL schema to identify any arrivals that would not join to the SITE table. Any problems are presented to the user, and must be addressed either by removing the offending arrivals or by modifications to the LLNL SITE table before execution can continue.

Next the program identifies all the column constraints currently in effect by querying the USER_CONSTRAINTS view LLNL and then checks each table-column affected by the constraints. Progress cannot continue until any problems are addressed. The principal advantage of this approach compared to hard-coding the tests, is that it isolates the source code from changes in the database and allows the same code to be used with different databases having different constraints.

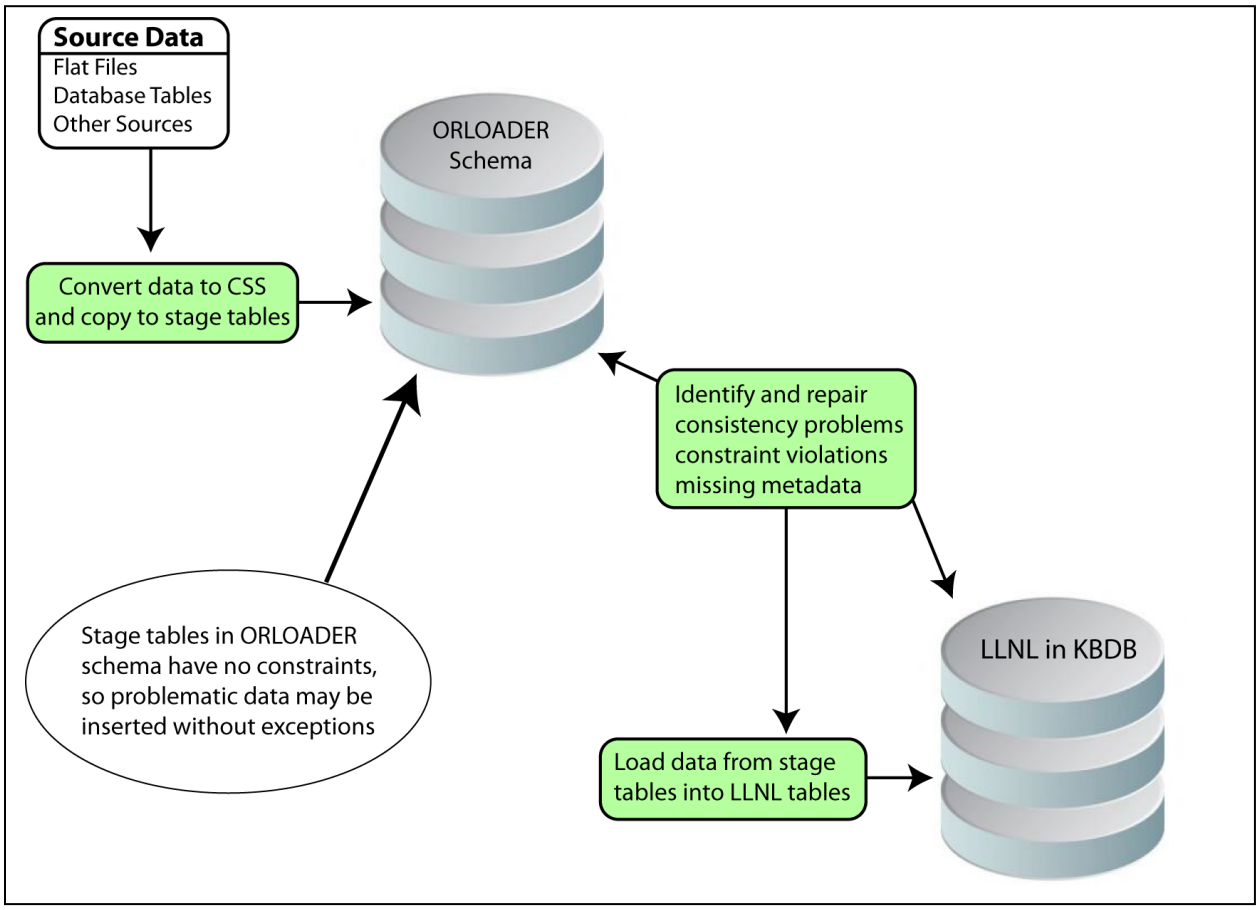

Figure 2. BulletinLoader data flow

Autonomous Mode

In autonomous mode, BulletinLoader runs continuously. Most of the time the program is sleeping, but periodically, it wakes up and checks sources to see if new data are available. If so, it retrieves the data and loads it. In autonomous mode, BulletinLoader can retrieve from both Internet sources and from remote database instances. Retrieval from internet sources is governed by the mechanism discussed in a previous section. Retrieval from a remote database instance is governed by entries in the LLNL.REMOTE_DB_EVID_MAP table.

This table is populated by a DBMS_JOB that runs periodically in the LLNL account. Each time the job runs it queries each database listed in the REMOTE_DB_REMAP_BASE table to find events that have been added since the last job execution and that meet certain user criteria. For each such event an entry is created in the REMOTE_DB_EVID_MAP table. If any origins in the event correlate with an origin already in the LLNL schema, then the LLNL EVID is entered along with the correlation information. Otherwise the information is left null.

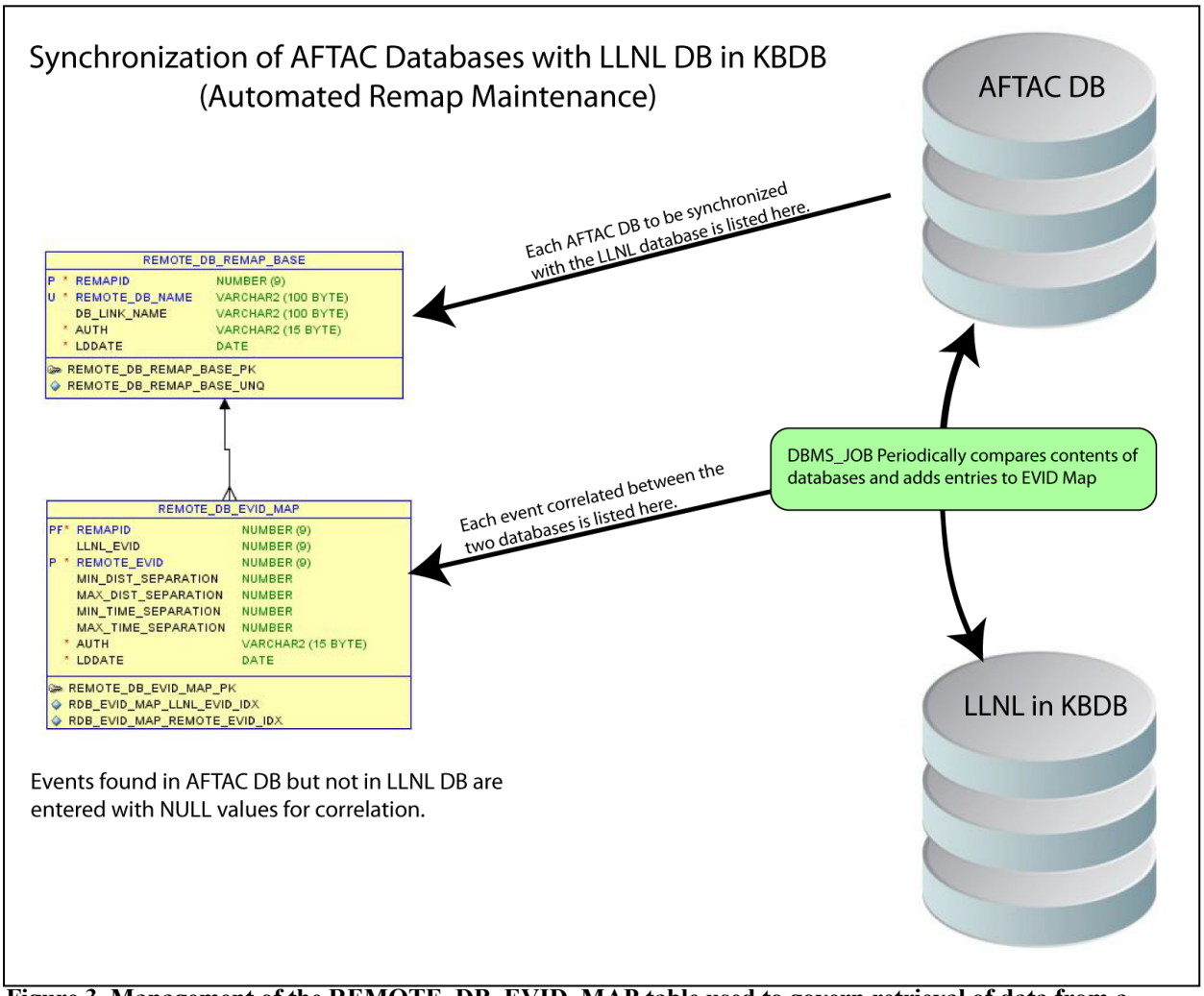

Figure 3. Management of the REMOTE_DB_EVID_MAP table used to govern retrieval of data from a remote instance.

BulletinLoader identifies new events to load from remote instance(s) by choosing rows in the REMOTE_DB_EVID_MAP that have a null LLNL_EVID. For those rows, BulletinLoader retrieves the data from the remote instance and loads it into the LLNL schema. After the data are loaded, BulletinLoader updates the REMOTE_DB_EVID_MAP with the correlation information.

SegmentLoader

The SegmentLoader program is a utility for adding data to the LLNL.WFDISC table. It combines the functions of the DDLOAD (waveform segment production and staging) and UPDATEMRG (updating the WFDISC table with new waveforms) programs. SegmentLoader currently supports three input sources:

1. Wfdisc flat files (and associated .w files)
2. Wfdisc tables (and associated .w files)
3. Wfdisc/Wftag pre-segmented data

In cases 1 and 2 the data are considered to be un-segmented. For these, SegmentLoader will identify time segments that are appropriate for events that exist in the LLNL.EVENT table, extract the corresponding waveforms, and merge them into the LLNL.WFDISC table. For case 3, the program assumes that the data are pre-segmented and that the WFTAG evid tagid values are correct for the LLNL schema. In this mode, the program functions like the old UPDATEMRG software. Continuous WFDISC data may be loaded from a table in the current instance or from a

table in a different instance. Any user with an account in the instance containing the LLNL account may run the program provided that they have been granted the WAVEFORM_LOADER role and have appropriate file system permissions in the target directory.

Event Screening

When loading data from a continuous WFDISC source, SegmentLoader identifies all the events in the LLNL.EVENT table that fall within the time frame of the source WFDISC table. However, in general only a few of the possible events will have a usable signal for many of the channels in the source. To avoid loading a lot of noise, SegmentLoader applies a screen to each event-station combination, to determine whether to extract the data.

Segment Merge details

In the simplest case, there is no segment in LLNL.WFDISC for a new source segment. In that case, SegmentLoader simply writes the new segment, and creates WFDISC and SEARCH_LINK rows. Often, though, there is pre-existing data for the EVID-STA-CHAN. In these cases, SegmentLoader must merge the new data with the existing.

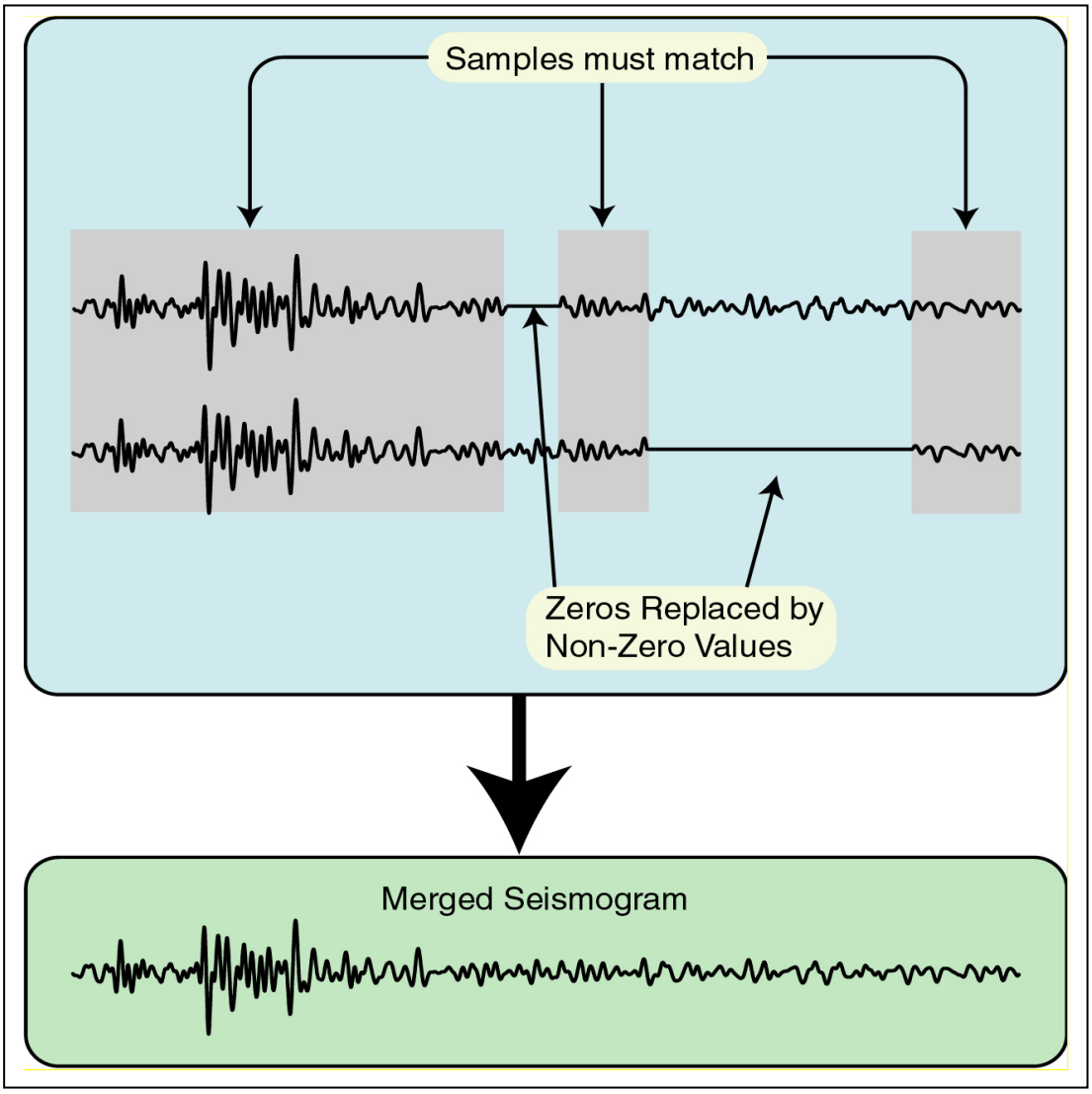

Figure 4. Schematic illustration of the process by which SegmentLoader merges waveforms.

The process is illustrated in Figure 4. The program starts the merge by reading the existing segment from disk and comparing the samples to those of the new segment. If all samples match, the program reports that the data are an exact match, and moves on. Otherwise, the program verifies that the sample rates are comparable and then aligns the data on common time samples. Then over the range of samples in the union of the two traces the following algorithm is applied:

- If one trace has a value and the other does not, then the value is copied to the merged trace.
- If both traces have a value then
    - If the values are the same, the common value is copied to the merged trace.
    - If one value is zero and the other is non-zero, then the non-zero value is copied to the merged trace.
    - If both values are non-zero, but do not match, then a MergeException is thrown. This exception will be reported in the log, and the new data will not be merged into the target, but SegmentLoader will continue processing the remaining waveform segments.

MtLoader

The moment tensor loading tool (MtLoader) was developed to load data into the recently-implemented NNSA schema (G.Ichinose, personal communication, 30 June 2010) for moment tensors. The tool allows manual loading of moment tensor information in a number of formats. It can also run autonomously and retrieve moment tensor information from various Internet sources as the data become available. So far, we have loaded ~57,000 moment tensor solutions with this tool. The primary database tables populated by this tool are:

- MOMENT
- FOCAL_PLANE
- MOMENT_VERSION

The tool also populates the ORIGIN and NETMAG tables as required, and by means of an insert trigger on NETMAG, populates the PREFERRED_MAGNITUDE table as well.

Input formats read by MTLOADER vary in their completeness of the moment tensor description. At one extreme, the NDK format supplies data for most of the columns in the MOMENT and FOCAL_PLANE tables. At the other extreme, a solution may be expressed as a (strike, dip, rake, scalar moment) tuple. Where the tensor elements are specified, the units may be CGS or MKS and the coordinate systems may vary as well. MTLOADER always translates moment tensor data as required into MKS and (X,Y,Z) reference frame.

For incompletely specified solutions, MTLOADER will calculate a number of additional columns as well. These can include the eigenvalues and eigenvectors, PISO, PCLVD,PDC, EPSILON, and the auxiliary plane parameters. Therefore in using the values in these tables, it is important to know that not all values are necessarily reported by the author of a given row.

We are currently extending MTLOADER so that it can accept as a source a MOMENT table and FOCAL_PLANE table from a different schema, possibly residing on a different instance. This will facilitate importing solutions produced by other organizations using the NNSA tables.

SWAP

The Surface Wave Amplitude Processor (SWAP) tool is intended to aid in the production of high quality amplitude measurements of surface wave amplitudes. In addition to being our first client to use the NNSA moment tensor schema, it is our first client software that uses the NASA World Wind software (http://worldwind.arc.nasa.gov/java/) for interactive mapping capability. With this software, applications can produce fully interactive, 3-D imagery-based maps with resolution in some places as high as 0.2 meters per pixel. Display of World Wind imagery requires a connection to one or more WMS servers for image and geographic data. By default, the imagery is served from a NASA server. However, for installations in which it is not possible to have a connection to the public Internet, the imagery can be stored locally in a cache pack or can be served from a locally-administered WMS server.

SWAP uses World Wind to display stations for which we have both moment tensor solutions and 3-component waveform data. (Both are required for producing the measurements.) For selected stations, World Wind is used to display the events for which we have both moment tensor solutions and 3-component waveform data.

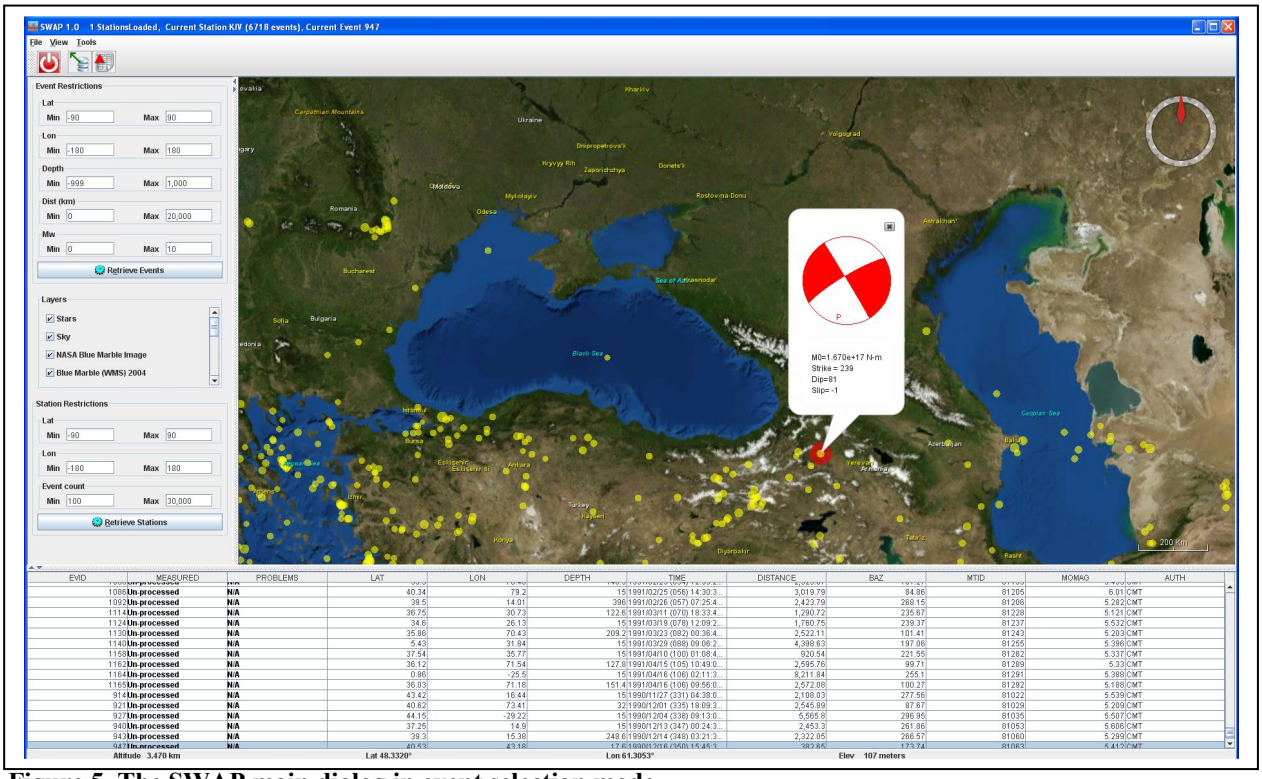

Figure 5. The SWAP main dialog in event selection mode.

Figure 5 shows the main dialog after a station has been selected and a set of events recorded at the station have also been selected. The events are shown on the map as colored circles scaled by magnitude. The color of the circle indicates the processing status. In the example, all stations are unprocessed. Events may be selected for processing by clicking on them in the map or alternatively, by selecting them in the table underneath the map.

Selecting an event loads it into the waveform dialog (Figure 6). In this dialog, all available 3-component trace sets are displayed on the left-hand-side, each separated from the others by a splitter. By making all sets available next to each other, the user has the opportunity to select the best (largest bandwidth, fewest data problems) set for measurement.

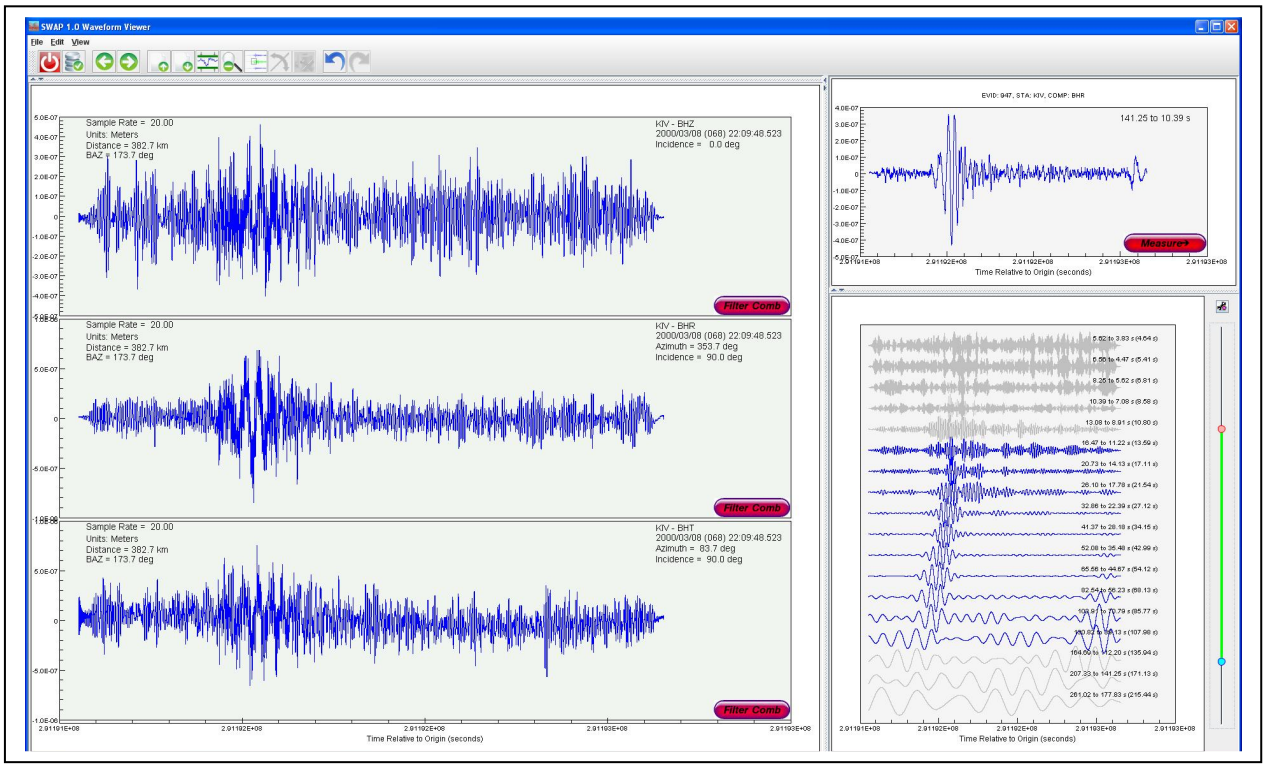

Figure 6 The SWAP waveform dialog showing the 3-component seismograms (left), the band selection control (lower-right), and the band-pass filtered component ready for measurement (upper-right).

After selecting the trace set to be measured, the user removes the instrument response and rotates the horizontal components into radial and transverse orientation.

After the user selects a trace it is narrow-band filtered and presented in the filter comb view shown in the lower right of Figure 6. This view provides controls to easily select the band of frequencies over which surface wave measurements are to be performed. The resultant filtered seismogram is displayed in the upper-right of Figure 6.

This is as far as we have developed SWAP to date. The next steps include adding a module for managing velocity models and adding the actual measurement code.

CONCLUSIONS AND RECOMMENDATIONS

This year most of our development work has been directed toward additional automation of our data ingestion capabilities. Our specific automation methodology and tools improve the researchers' ability to assemble quality-controlled research products for delivery into the NNSA Knowledge Base (KB). The software and scientific automation tasks provide the robust foundation upon which synergistic and efficient development of GNEM R&D Program seismic calibration research may be built. Three new Java programs, BulletinLoader (seismic bulletins), SegmentLoader (seismic waveforms), and MTLoader (moment tensor information) have been developed that replace older C/C++ codes and extend the capabilities of those codes to reduce the amount of required human supervision. This allows resources to be focused on analysis and production of calibration products.

These new applications comprise a significant re-engineering of our data management software suite designed to leverage new data distribution technologies that have been developed by data providers in the last few years. Exploitation of the new methods has resulted in demonstrated gains in efficiency and reduction in time required to identify, acquire, process raw data essential for calibration research.

ACKNOWLEDGEMENTS

We acknowledge the assistance of the LLNL computer support unit, Stan Ruppert and Laura Long, in implementing and managing our computational infrastructure. Prepared by LLNL under Contract DE-AC52-07NA27344.

REFERENCES

Ruppert, S. D., D. A. Dodge, T. F. Hauk, M. D. Ganzberger, D. B. Harris, and E. M. Matzel (2009). Enhancing seismic calibration research through software automation and scientific information management, in *Proceedings of the 2009 Monitoring Research Review: Ground-Based Nuclear Explosion Monitoring Technologies*, LA-UR-09-05276, pp. 808–817.

World Wind JAVA SDK 15.75K. NASA. 8 October 2009. <http://worldwind.arc.nasa.gov/java/>