EMERALD: A Flexible Framework for Managing Seismic Data

John D. West, Matthew J. Fouch & J Ramón Arrowsmith

School of Earth and Space Exploration, Arizona State University, Tempe, AZ 85287

AGU Fall Meeting 2010 S13C-2022

Data Management

The seismic community has open access to massive volumes of high quality broadband seismic data from the IRIS Data Management Center and from other regional and national data centers. Data from the hugely successful EarthScope USArray Transportable Array program, regional networks, and hundereds of PI-driven experiments threaten to overwhelm most individual seismic data processing methodologies simply because data sets containing millions of individual seismograms are now possible and are becoming increasingly common.

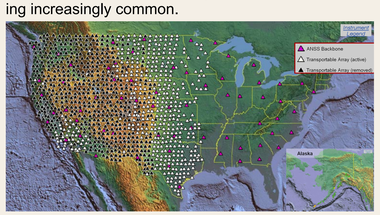


Figure 1: The EarthScope USArray network (as of 11/24/2010, www.earthscope.org). The huge volumes of data from this and other networks pose new data processing challenges and exacerbate existing issues.

Researchers have long faced the challenge of station metadata changes, i.e. changes in recorded location, orientation, instrument response, etc. over the lifetime of the station. The advent of large datasets makes it extremely difficult to manually search for updated metadata, and few if any existing systems provide methods for automating such updates.

Existing command-line driven seismic pre-processing methodologies require a significant learning curve, which can act as a barrier to exposing new and undergraduate students to seismic data processing.

Sharing of processing methods between researchers is difficult within the seismic community, as each individual researcher typically develops a personal naming convention and file organization structure. This, along with differences in computing hardware and operating systems, makes it difficult for one researcher to easily test and adapt the methods of another.

To address these issues, we have developed EMERALD, an integrated software framework for processing seismic event data. EMERALD is capable of easily handling many millions of seismic waveforms, while allowing researchers to easily add and share processing methods. EMERALD can automatically check for metadata updates and alert the user to changes, and the web browser interface is intuitive and easy for beginning researchers to learn.

EMERALD Features

Explore, Manage, Edit, Reduce, & Analyze Large Datasets

Metadata Updates

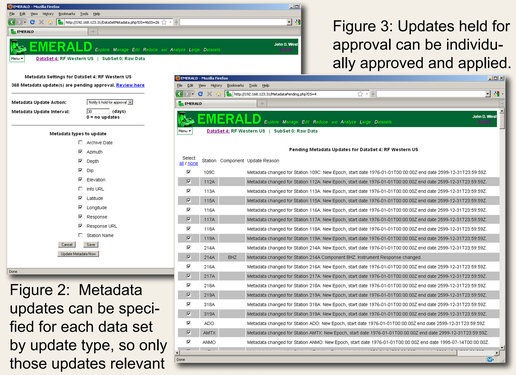


Figure 2: Metadata updates can be specified for each data set by update type, so only those updates relevant to that experiment are checked. Updates can be immediately applied or held for review, and regular automatic update checks can be scheduled.

Figure 3: Updates held for approval can be individually approved and applied.

E-Mail Notifications

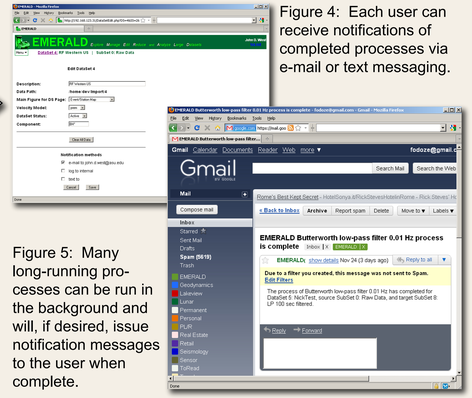


Figure 4: Each user can receive notifications of completed processes via e-mail or text messaging.

Figure 5: Many long-running processes can be run in the background and will, if desired, issue notification messages to the user when complete.

Trace Editing

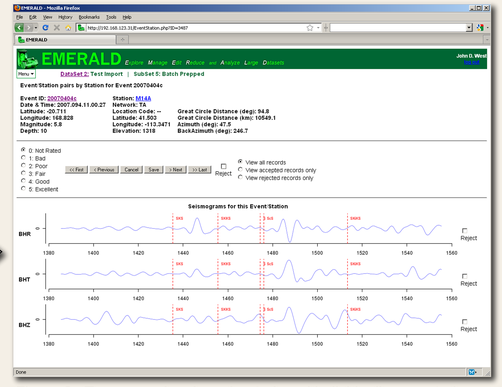


Figure 6: Traces can be viewed, rated, and individually accepted or rejected for individual event/station pairs, either reviewing all stations for a given event or all events for a given station. Hyperlinks allow easy toggling between event-centric and station-centric views. Ratings and navigation can accomplished with keyboard shortcuts or with a mouse.

Figure 7: For fast trace reviewing, all traces for a given event/component or station/component are shown on a single screen. This allows the user to quickly and easily compare seismograms and choose which to retain in the working data set.

Batch Process Automation

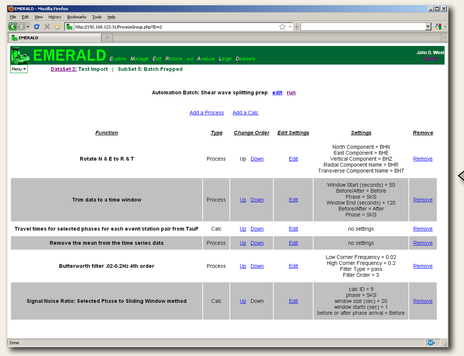


Figure 10: Users can define their own automation batches containing any combination of calculations and processes. This allows users to develop, customize, and execute a series of methods as their preferred workflow. The user can choose to be notified at the completion of each step and/or at the end of the batch.

Calculated Values

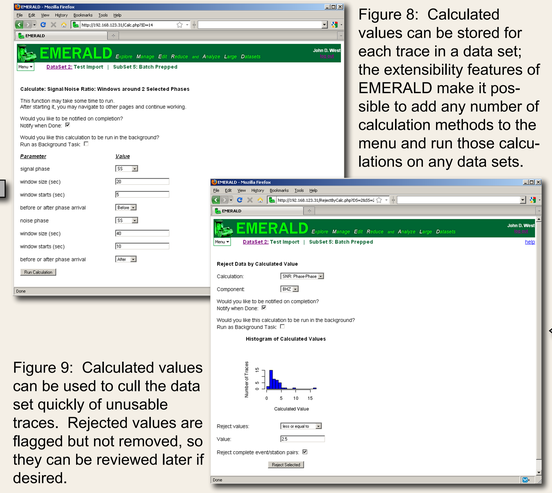


Figure 8: Calculated values can be stored for each trace in a data set; the extensibility features of EMERALD make it possible to add any number of calculation methods to the menu and run those calculations on any data sets.

Figure 9: Calculated values can be used to cull the data set quickly of unusable traces. Rejected values are flagged but not removed, so they can be reviewed later if desired.

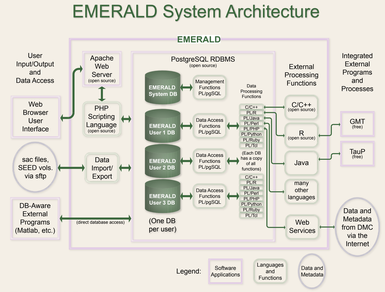
Beta Testing

Beta testing of EMERALD began in early October, 2010 and is ongoing. Here are some comments from our early testers:

- "I've been asking for this for years!" -- Nick Schmerr, Goddard Space Flight Center.
- "Seriously - I sort of can't believe that I'm working on seismograms via web interface on the phone. Totally brilliant." -- name withheld by request.
- "I tried it for the first time, and it is great!" -- Panxu Zhang, USC.
- "Finally, someone with both the realization of what tools are essential within our community and the capability to produce them!" -- Michael Thorne, University of Utah.

I will be available all week for live EMERALD demonstrations. Email John.D.West@asu.edu if you would like a demo or to discuss EMERALD.

System Architecture



EMERALD is an open-source client-server database application intended for use by an individual researcher or small workgroup, and is based on the PostgreSQL open-souce Relational DataBase Management System (RDBMS).

The user interface is a web browser; web pages are served to the user via the Apache web server and PHP scripts. Users can move seamlessly between clients, including smart phones and tablets.

EMERALD is designed to be easily extended, and incorporates many existing utilites such as TauP, GMT, and the DHI/Fissures framework. New functionality can be added and shared between researchers, using any of at least 12 programming languages.

EMERALD is packaged as a virtual machine appliance which can be easily installed under Mac, Windows, or Linux without concern about complex configuration or missing drivers or libraries.