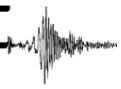
SKATE



USER MANUAL V1.0

Table of Contents

Introduction

USER NOTE:

SKATE is a work in progress, and the current version reflects research and development that was done up to the completion of current DOE SBIR Phase II funding. It is currently not complete, and users should understand that the purpose of making this software publicly available is to allow qualified researchers to evaluate the work to date, and to make detailed comments and suggestions as to how it can be improved. Known issues include improved meanline detection, significant reduction in spurious features, and most importantly incorporating advanced connection algorithms that we are developing.

IMPORTANT NOTE ON UPDATES:

Because SKATE is continually being updated, users need to ensure that their browser has the latest software. SKATE downloads and runs an app in the user's browser; in order to update this software, the browser's cache needs to be cleared. In Chrome, go to History -> Clear Browsing Data, and select Cached Images and Files. A restart of the browser might be required to run the latest version of the app.

SKATE (Seismogram Kit for Automatic Trace Extraction) is a web-based software tool for the digitization of seismic traces in historic analog seismograms. This work was performed under DOE contract DE-SC0008219.

SKATE is designed to address the need to digitize[1] the millions of historic seismograms that are currently unavailable for analysis. While there exist other digitizing programs, they are typically limited by the need for significant user interaction, slow speed, and/or proprietary software requirements. The design of SKATE specifically addresses these limitations by:

- Web-based architecture. To operate SKATE, the user simply needs to direct a browser to the software's web location at seismo.redfish.com. No software needs to be installed on the user's machine; modern browser capabilities allow the entire process to be run on a remote server.
- Fast and parallelized processing. Tens of thousands of seismograms have already been processed using Amazon EC2 virtual computers, running in parallel. Processing time per seismogram is currently 15 minutes. We anticipate that this will be speeded significantly in future versions.

[1] 'Digitize' refers to the creation of (x,y) timeseries data for seismic traces. 'Scanning' refers to the creation of a digital image from an original paper or film record. A scanned image has not yet been digitized.

- Reduced user interaction. We have designed robust and detailed algorithms to extract and separate real trace information from background noise and other spurious features. Segment connection algorithms automatically track traces across crossings when there is significant seismic activity. User interaction is still recommended and required in order to eliminate spurious features, but easy to use tools and intelligent algorithm design reduces this effort significantly.

This is the alpha version of SKATE, and users should recognize that testing of the product is ongoing. Hence, this manual will contain phrases like 'should' and 'likely,' which indicate that while we believe most features will behave as described, uncertainties such as variabilities in browsers and versions and end user machines can affect the behavior. The user should understand that SKATE is constantly being upgraded, and as such features will be added and subtracted without warning. Please contact andy@retrievertech.com with questions and comments about operation, updates and suggestions.

Current Status and Available Images

Retriever Technology has scanned approximately 150,000 WWSSN images under several contracts, including but not limited to USGS Awards G09PC00116, G09PD02064, and G10PD02236. These images were saved as high resolution, uncompressed .tif files, and were delivered to the USGS. Retriever Technology has used these images to develop SKATE, and has stored them for this project as lossless .png files on an Amazon S3 server. As of now, we have completed two types of digitizing of WWSSN images:

- What we call 'raw data,' meaning that we have processed approximately 30,000 randomly selected images up to the point of automatic meanline assignment, and segment and centerline identification. These data have not been run through the final segment assignment algorithm. This assignment algorithm can be run at any time, but we feel that absent some user interaction and editing, the final result will have less than ideal results. These images were limited to WWSSN long period seismograms from 1970 and on.
- The second type of processed images are listed as 'has edited data.' This is a small subset of < 20 images that we have hand edited to remove spurious features and to ensure that the meanlines are properly placed. They were then run through the segment assignment algorithm and have data that is ready to be downloaded. Please refer to later sections in this manual that discuss editing and downloading of data.
- There is a third category available through the UI called 'no data.' There are over 150,000 of these high resolution scans available for viewing only. Budget limitations due to Amazon's EC2 operational costs prevent these from being processed or downloaded at this time.

Currently, users cannot upload and process their own images, nor create 'raw data' on any of the 'no data' seismograms. What users can do is edit any of the 'raw data' seismograms,

and run the assignment editor on them. Please note that there are severely limited funds to maintain and operate the web server. As such, we ask that users design their experiments to maximize results while minimizing computational time. Downloading data incurs costs as well, so only download data when editing and all processing is complete.

While SKATE has been optimized for project development purposes for long period WWSSN images, it can readily be configured to digitize short period images, of which there is an example in the 'edited data.' In addition, we have successfully digitized data from the Caltech archives, see the following reference for information on this archive: http://ds.iris.edu/seismo-archives/projects/caltech_archive/Caltech_Seismograms_v2.pdf. We are working to expand the types of seismograms that can be digitized. This includes partial images that contain only an event of interest. This effort is limited only by the need to understand and address particulars of resolution, size and other general features found in these images. SKATE's algorithms are very general for finding seismic features as lines traces. We have even digitized historic mareograms (tide gauge charts).

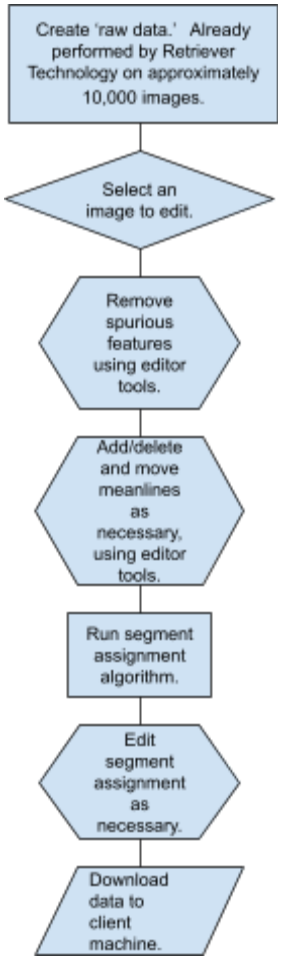
Requirements

The SKATE seismogram processing package is hosted on a remote web server. There is no need to install any software. Computer and browser requirements are:

- Browser. SKATE has been tested in Chrome only. The most current version tested is 45.0.2454.101 m. It is recommended that the user update Chrome to the latest version. There are no known settings or permissions required. Internet Explorer 11 has not been tested but should work. Mozilla Firefox version 45.0.2454.101 m has not been tested extensively, but should work as well. It is highly recommended that Chrome is used. If other browsers are used, it is highly recommended that the user update to the latest version.
- System requirements. Because SKATE runs in the browser, client system requirements are light. Reasonable and standard RAM and CPU performance should suffice. Users might want to consider an upgraded graphics card if the display does not respond quickly to commands. A desktop or laptop computer is recommended. Though we have tested SKATE on Android tablets and phones, these devices have not yet been fully optimized.

How to use SKATE

This version of SKATE has been designed to allow users to explore and evaluate its current capabilities by viewing, editing and downloading seismogram timeseries data that have been solved up to the point of having the above described 'raw data.' The flowchart outlining the process is shown below.



Operating Instructions

When referring to a command or other information that appears on the UI, those words will appear quoted in italics. When referring to a keyboard action, the actual keystroke will appear in angle brackets, e.g. <+> indicates use of the plus sign on the keyboard. Do not include the angle brackets in the action.

Definitions

Terms used in this manual are defined here.

meanlines. This is the zero-energy line that a trace would follow if there was no seismic activity. Meanlines are the image-wide line segments that we calculate for this zero-energy path, and which are used for subsequent trace assignment and connection algorithms

trace. In a WWSSN image, this is the image path that the incident light beam records on the recording drum's photographic paper. It is the seismic data from which we extract timeseries data.

segments. A seismogram image file consists of foreground and background. Foreground features are the traces themselves. Our task is to identify and separate them from the background. Background includes all non trace features such as noise, and additional features such as hand written notes, etc. Because of timing marks, trace crossings, and the multiple hourly lines found in a WWSSN seismogram, the traces are broken up into pieces. We call these pieces of a seismic trace a segment.

centerlines. Centerlines represent the actual path that a trace is following. It is a single pixel wide object that tracks the middle of a segment.

ROI. The ROI is the Region of Interest in the seismogram. It is the portion of the image that contains the actual seismic traces. In the image below, the ROI is shaded in light blue.

scanning. Seismograms that are scanned are simple image files, usually .tif or .png.

digitize. Digitizing a seismogram takes the scanned image file and extracts the time series that represents the seismic trace's amplitude as a function of time. We output the time series as either a .json file or a .csv. Other formats such as mSEED will be created in future versions.

S3. Amazon S3 (Simple Storage Service) is an online file storage web service offered by Amazon Web Services. Amazon S3 provides storage through web services interfaces (REST, SOAP, and BitTorrent)[2]. Retriever Technology stores all of its scanned seismograms on S3. In addition, the UI is hosted on S3. Monthly costs accrue for this service.

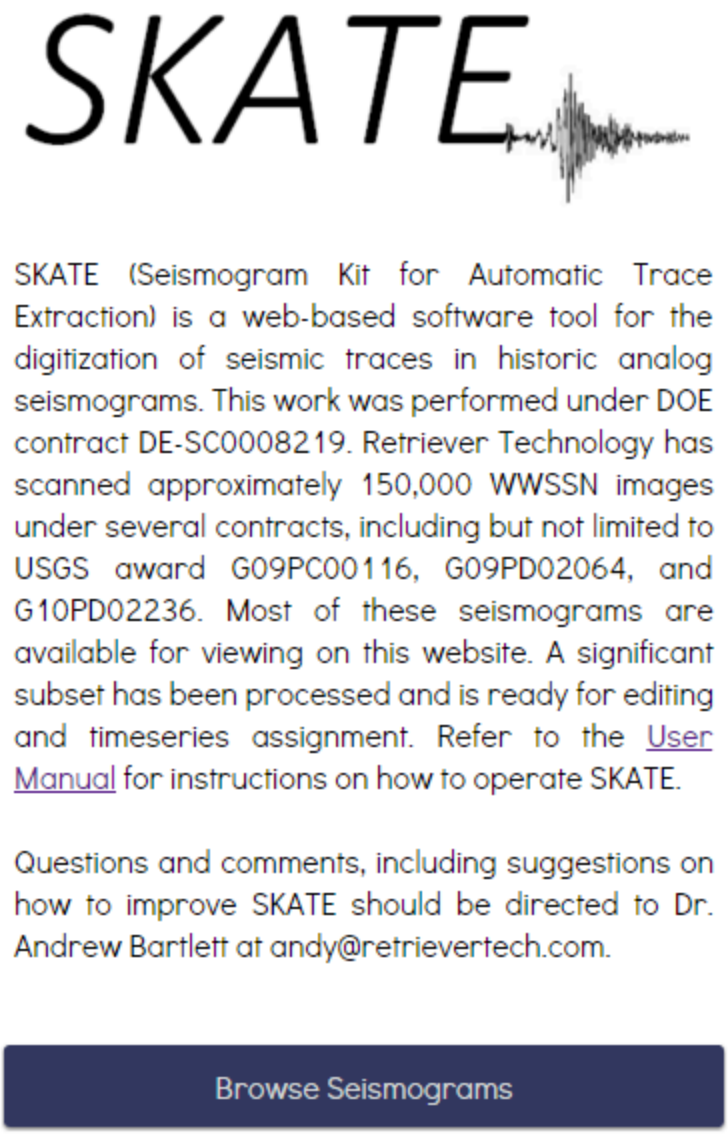
Login

Login requirements have changed from earlier releases. Now, there are no login requirements for browsing, editing and downloading timeseries data. The only time a login is required is

[2] https://en.wikipedia.org/wiki/Amazon_S3

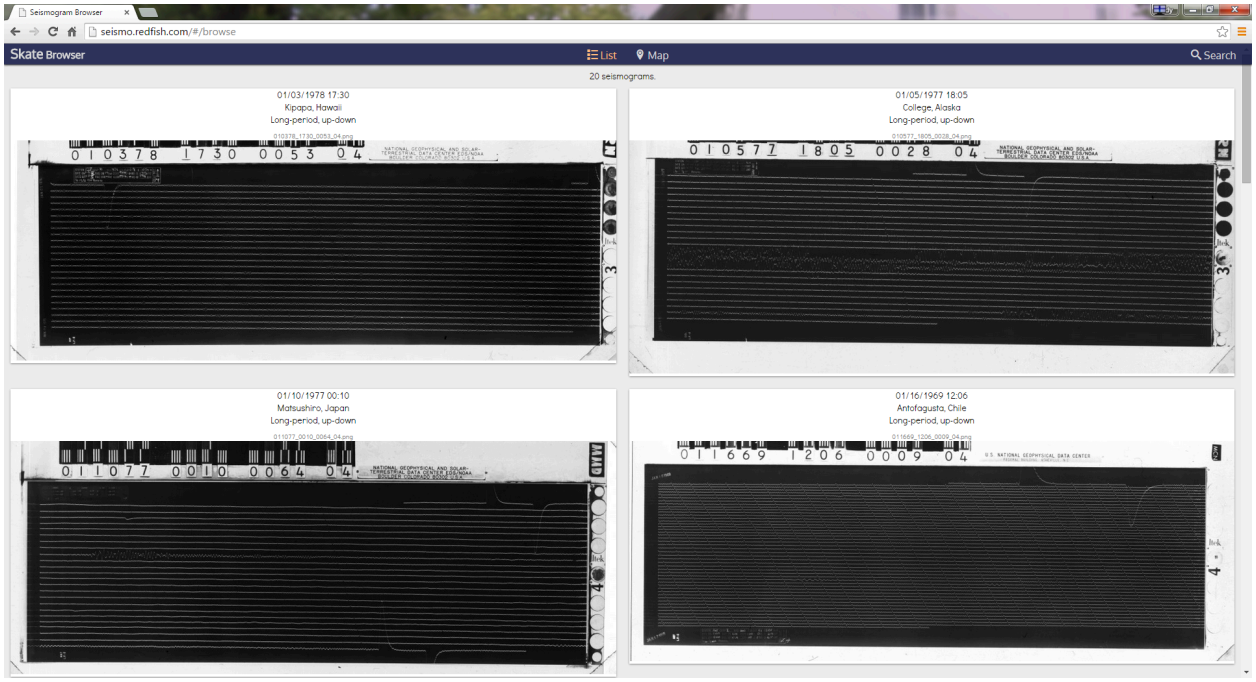
when saving data; because this permanently alters the metadata on S3 we are limiting this functionality to approved users. Refer to the section on saving data for login instructions.

To begin using SKATE,,navigate to seismo.redfish.com The following welcome page appears.



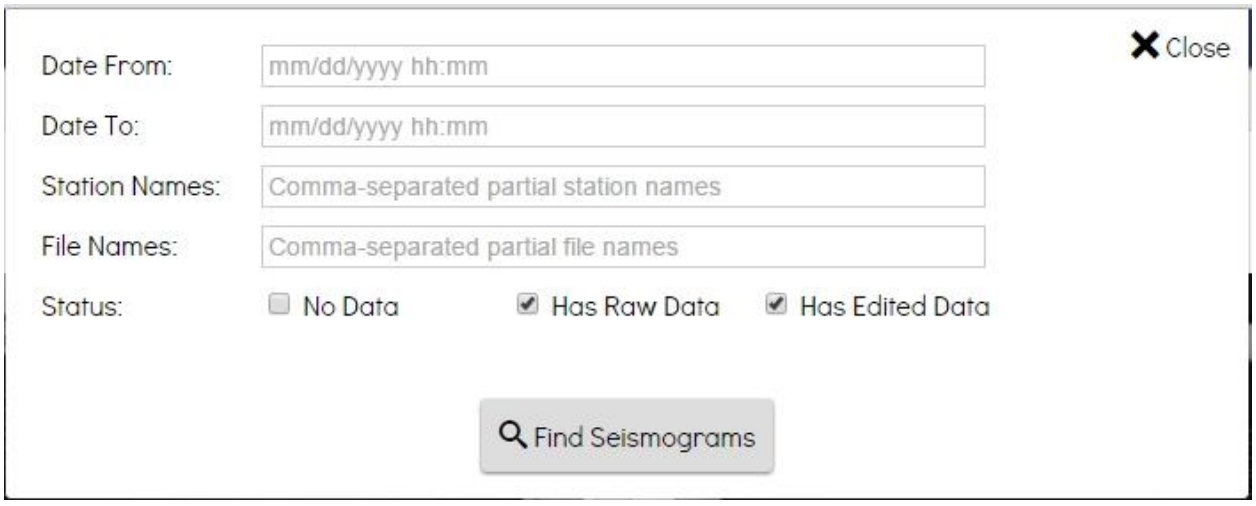
Click on '*Browse Seismograms*.'

At this point a thumbnail list of seismograms will likely be displayed:



Searching

To search for a particular seismogram, click on the "*Search*" link. The following menu item will appear.



Dates. The Search window allows specific dates and times to be searched. A complete date can be entered. Alternatively, the four digit year can be inputted in the '*Date From*' and '*Date To*' fields. Searching by only putting in a month or a day or a time will not work.

Station names. This allows for searching of one or more stations by name. Type in the three character station name, or a portion of the name, or the name or portion of the actual location (i.e. city and country name). Multiple stations can be searched by separating station names by a comma. Do not search by station number, as this option is not available. Quotes and wildcard operators do not work.
Station names and other WWSSN information can be found in the WWSSN User's Guide at http://pubs.usgs.gov/of/2014/1218/pdf/ofr2014-1218.pdf.

No Data. Searching can include/be limited to seismograms with *'No Data*.' This indicates a seismogram that is accessible from the UI but has not been processed in any way. There are currently 154,658 of these seismogram which represent Retriever Technology's database on S3. Users may search and view these images, but due to cost constraints tied to Amazon's data download fees and EC2 operational fees, these are not available for download or processing. Contact Retriever Technology if there is a particular seismogram of interest.

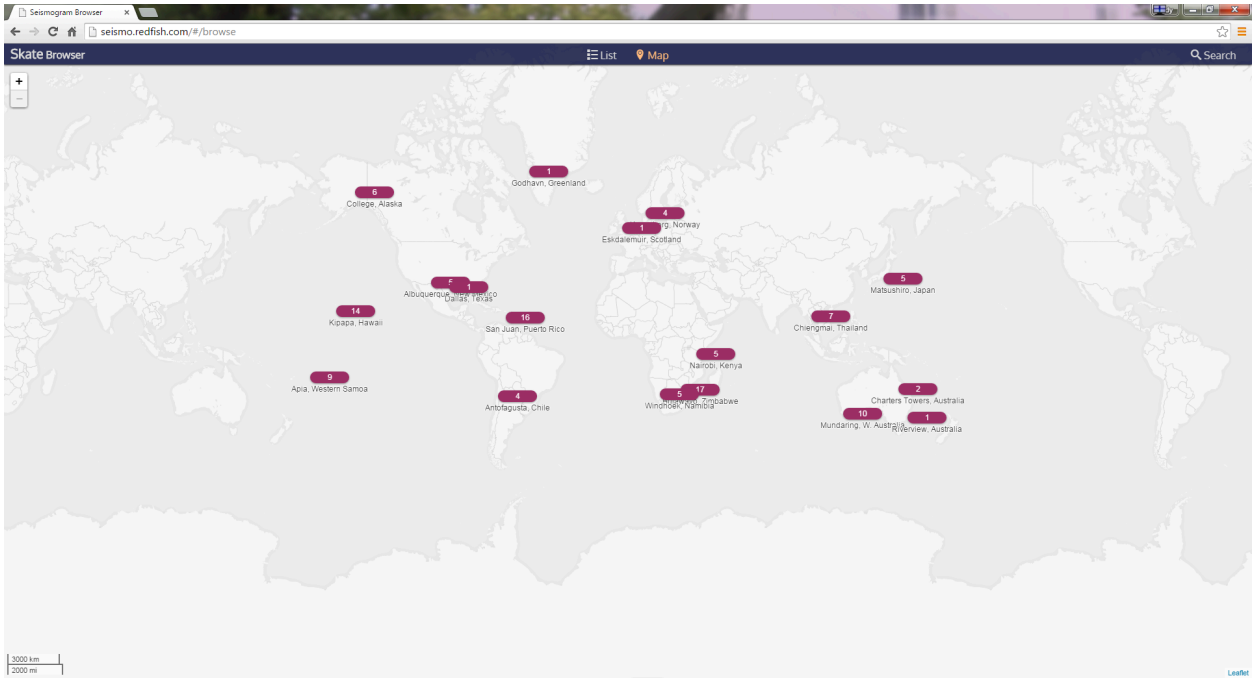
Has Raw Data. Searching can include/be limited to seismograms with '*Raw Data*.' This indicates seismograms that have been processed up to the point where they have: ROI, meanline, segment and centerline information. They do not have segment assignment information. There are currently over 11,000 of these file, with more being processed on 40 EC2 instances. There will be approximately 50,000 available shortly.

Has Edited Data. Searching can include/be limited to seismograms with '*Has Edited Data*.' These seismograms have the same information as in '*Has Raw Data*', and in addition have been edited to improve meanlines and remove spurious segments as necessary, and then have been run through the assignment editor. Segment assignment means that segments are sequentially attached to previous and subsequent segments. In other words, the seismogram has been solved.

This is a small subset of images that have been hand-edited by summer interns at Retriever Technology. They are likely very good, though not perfect, in their editing.

Browsing and Viewing

Once a search is complete, a thumbnail list of found seismograms will be displayed. The default "*List*" display is active and will display 20 thumbnails per page. Scrolling down will continue to pull up more thumbnails. Alternatively, clicking on the *"Map*" icon will display a world map with location and counts by station of the searched-for seismograms.



To view seismograms from a particular station, click on the location. The available seismograms will display in the "*List*" view.

Once the seismogram of interest is located, click on the thumbnail to display the image. Navigating the image is done as follows.

Zooming in and out

Using the scroll wheel on a mouse zooms in and out on the displayed image. Also, the "+/-" button on the screen allows zooming.

Absent a scroll wheel, the <= > keyboard key zooms in. The <- > keyboard key zooms out. If the "*Seismogram Data*" or "*Seismogram Actions*" menus have been used, it might be necessary to first place the mouse over the image and click once in order for the keyboard zoom function to become active.

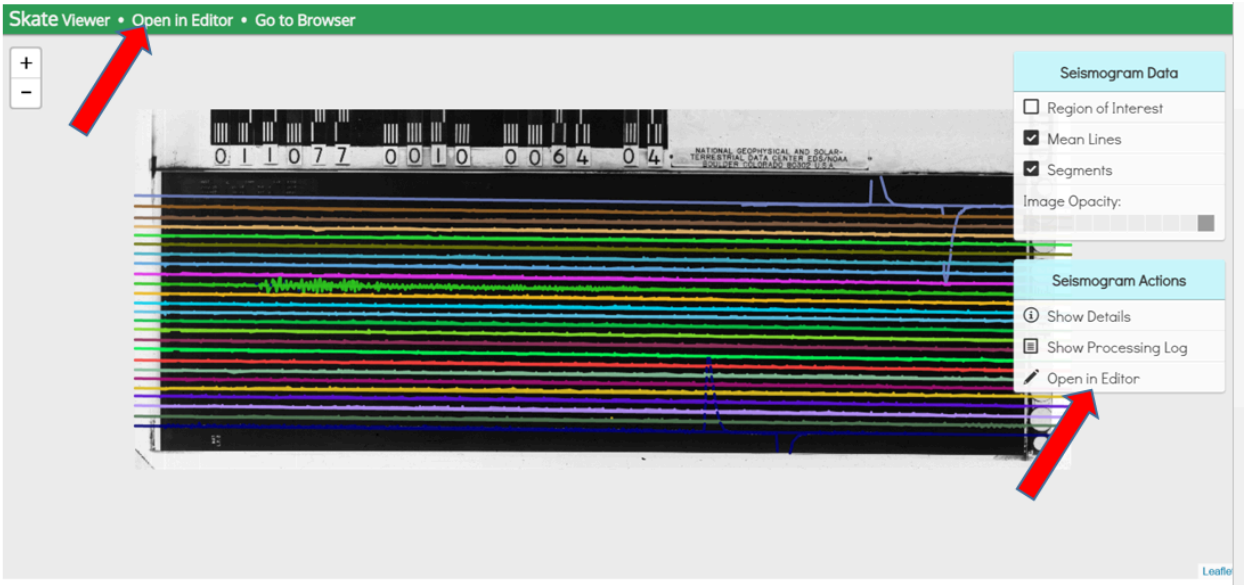
Panning. Left click and hold while dragging the mouse. When zoomed in, the up/down, left/right arrows on the keyboard also should allow panning.

Saving Images

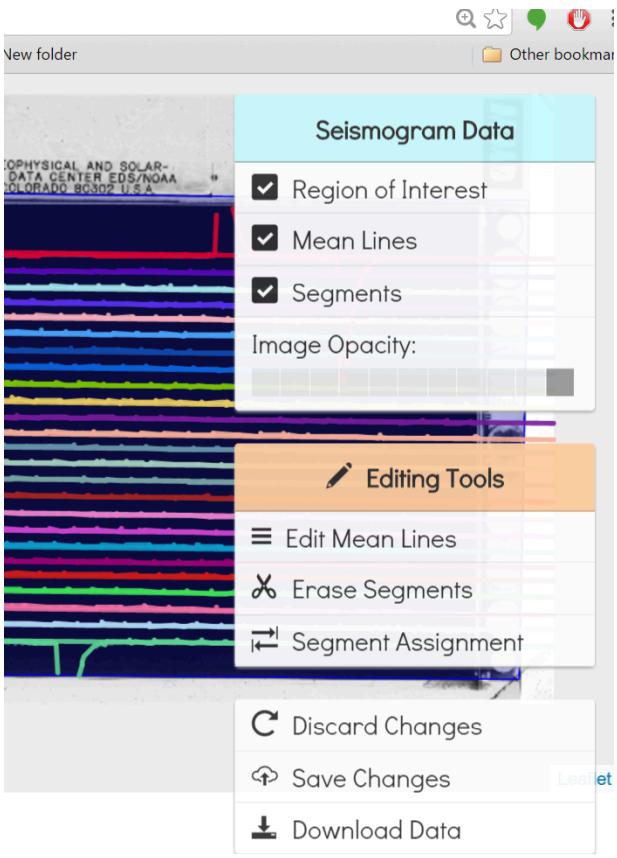
Any full resolution .png image that is available on the UI can be downloaded. Open the desired image and click on "Download Original." Please limit the number of downloads to only those necessary.

Editing

Editing can be done both before and after segment assignment. To switch to editing mode, click on either of the links shown by red arrows below.



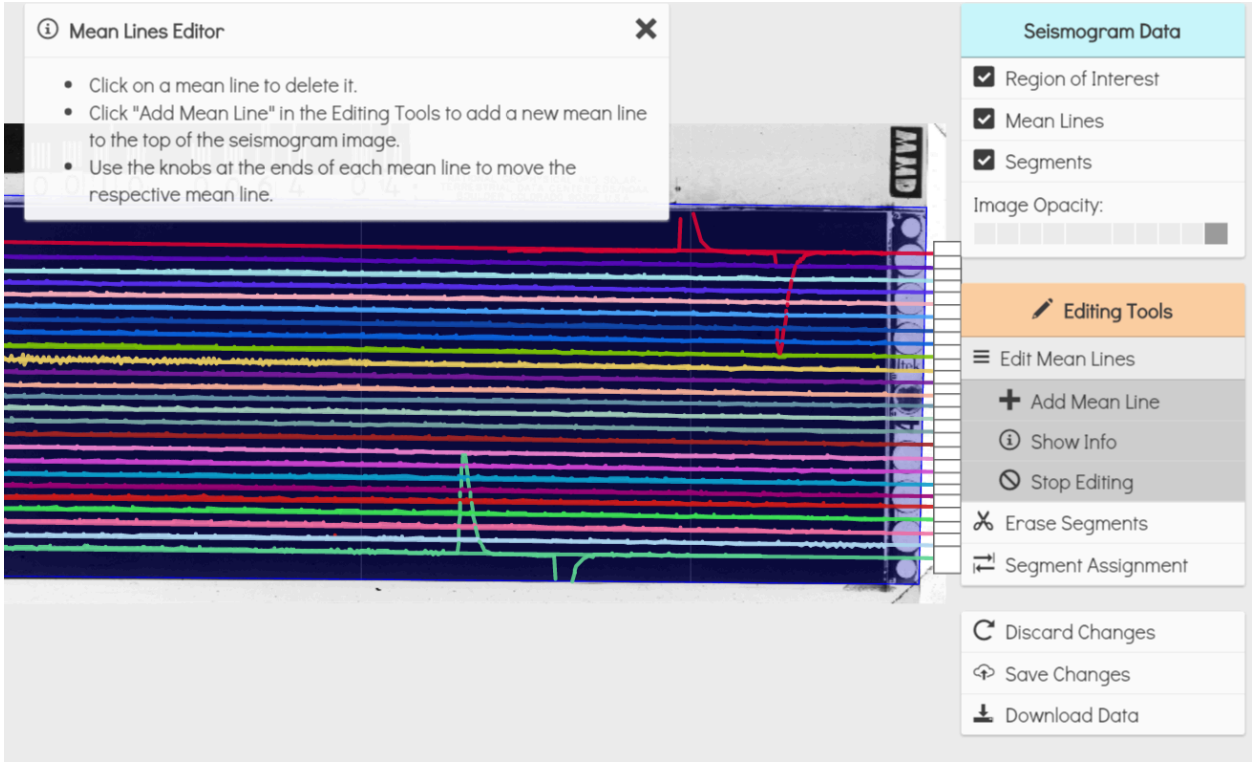
Once in Editor mode, the following screen with menu items becomes available.



Editing Tools

Edit Mean Lines

Clicking on "*Edit Mean Lines*" pulls up the following menus. It might be necessary to click on "*Show Info*" in order to view the "*Mean Lines Editor*" menu box:

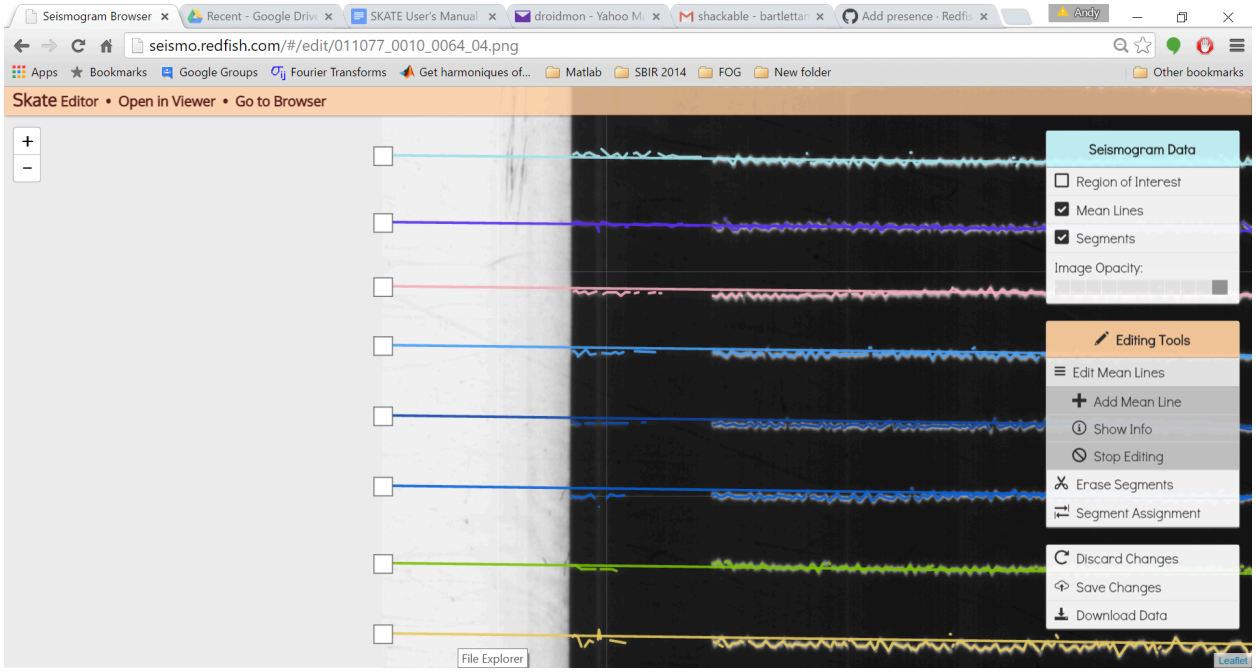


Meanlines are the zero-energy line of the seismogram, and are used for Segment Assignment. However, they are not used as data normalization lines, so it is not critical to exactly align them to the zero-energy line.[3] Images that are listed as "*Has Raw Data*" are those in which meanlines have been automatically calculated. The algorithm is not 100% accurate, so the user should add, delete, or move meanlines in order to cover all lines and partial lines with a meanline.

Moving a meanline

Meanlines have handles on the end of them, and extend beyond the ROI and out to the edge of the entire image in order to make it easy to visualize and manipulate. Zooming in while in the meanline editor looks like this:

[3] Image distortions such as lens distortions, scanning errors, paper shrinkage, etc., make a linear fit to a zero-energy line less than exact. The issue of zero-energy line fit will be handled in future iterations when manipulating the trace information in data space, as opposed to image space.

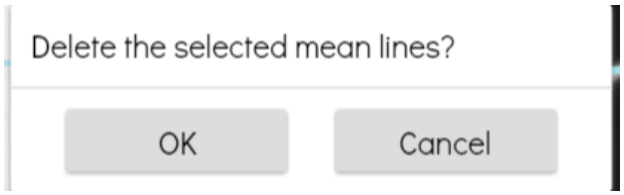


Note that the meanlines in this example are not perfectly aligned with the zero-energy line, but represent positioning that is sufficiently close for subsequent segment assignment.

To reposition a meanline, click on the handle and drag the mouse to the desired position. Repeat for any line that needs to be repositioned. Changing one meanline does not change any others.

Deleting a meanline.

To delete a meanline, click anywhere on the meanline besides its handle. A dialog box will appear:



Select "*OK*" or "*Cancel*" as appropriate.

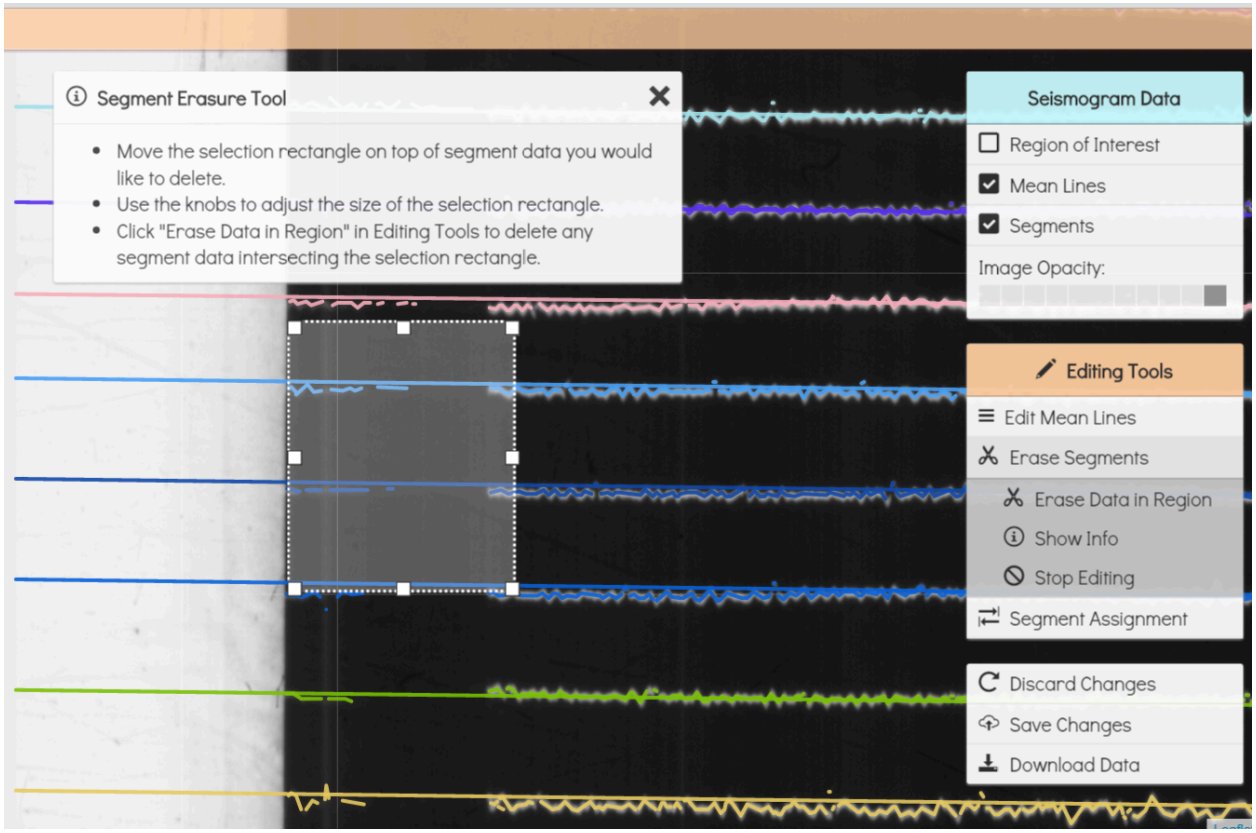
Adding a meanline

To add a meanline, click on "*Add Mean Line*." The new meanline will appear at the top of the image. (Zooming out might be required to see the new meanline.) Multiple meanlines can be added. Note that clicking on "*Add Mean Line*" multiple times will overlay the new meanlines on top of one another. Each needs to be dragged out of the way in order to view the new meanline that is below it. Place meanlines where desired using the procedure to move meanlines described above.

Erase Segments

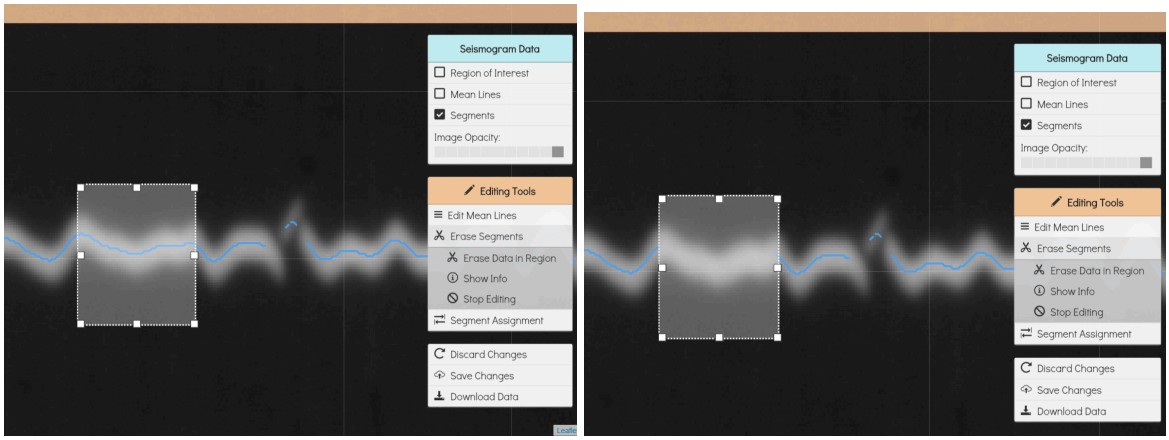
Images that are listed as "*Has Raw Data*" are those in which segments and centerlines have been automatically calculated. As it is with meanlines, the algorithm to calculate segments and centerlines is not 100% accurate, and typically errs by identifying noise and other superfluous features as segments. The editor allows the user to selectively delete these features[4].

To get into the segment editor, click on "*Erase Segments*." The following menu items appear. Clicking on "*Show Info*" will pull up the "*Segment Erasure Tool*" info box.



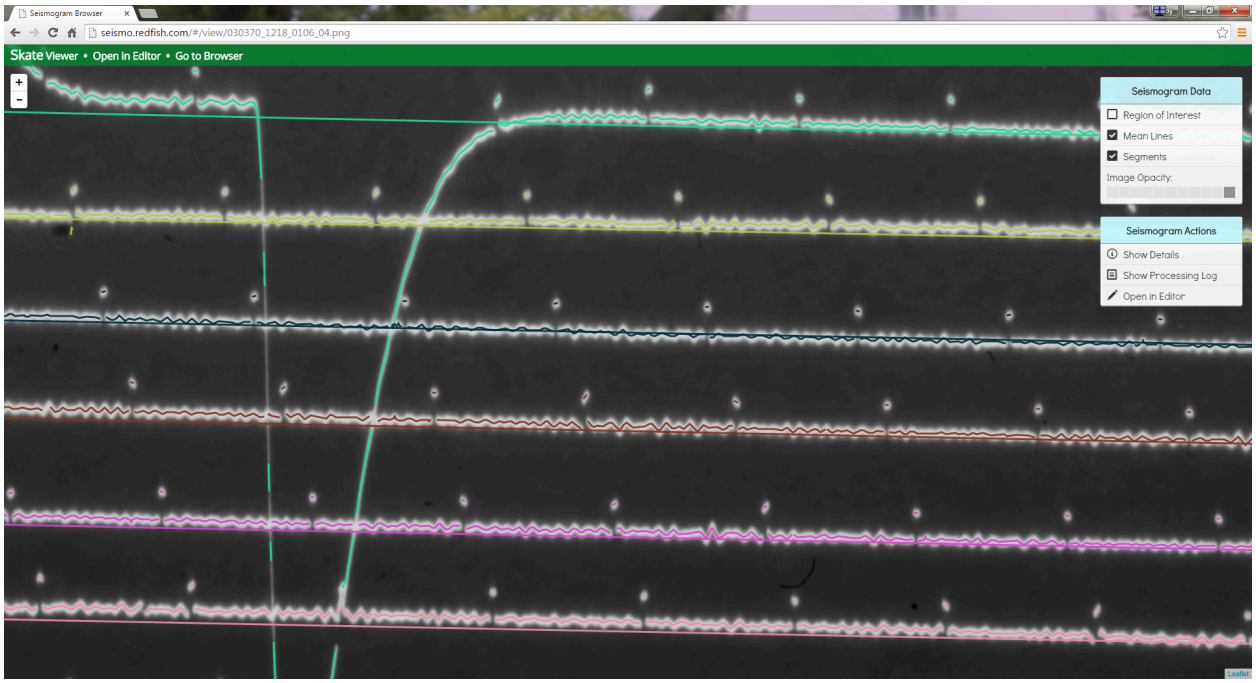
Once "*Erase Segments*" is clicked, an 8-handled rectangle selection tool will appear, as shown below. Zoom and pan as required to locate those segments to be deleted. Click and drag in the middle of the selection tool to move it around. Click and drag the selection tool's handles to adjust its size. When the segment to be deleted is enclosed by the selection rectangle, click "*Erase Data in Region*." The images below show before and after segment deletion.

[4] At this time there is no way to *add* segments and centerlines. Future improvements will likely allow this feature to be performed in data space rather than image space.

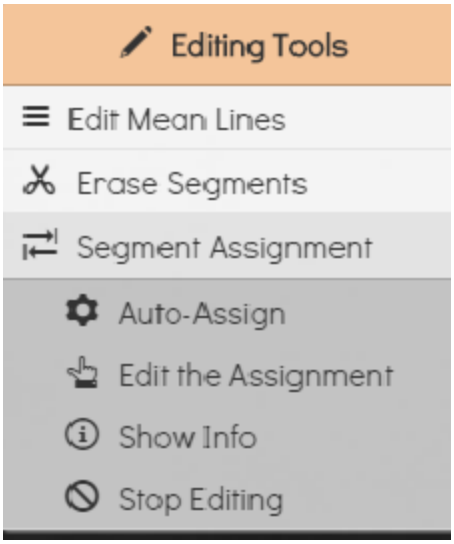


Segment Assignment

Once editing is complete, the Segment Assignment algorithm can be run by clicking on "*Segment Assignment.*" The algorithm uses the meanlines as a fundamental organizing feature. Therefore, after running the assignment algorithm, the segments will be colored according to the meanline with which they are associated, as shown in the image below.

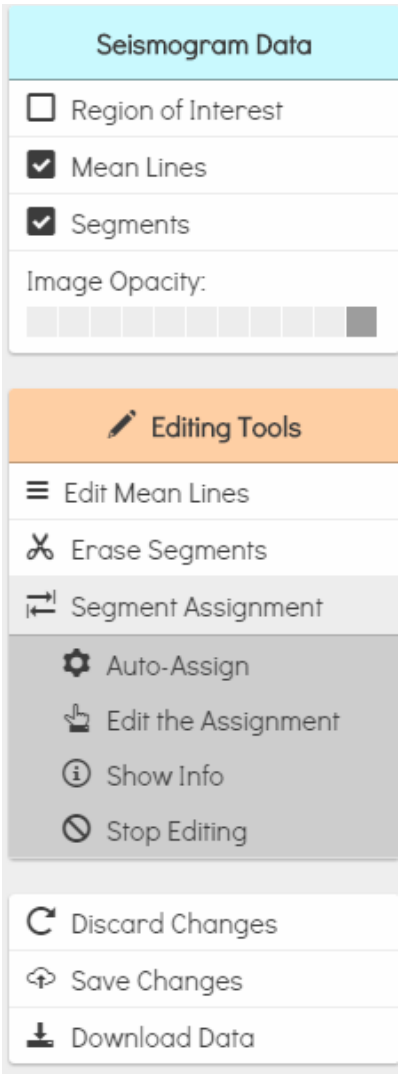


To run the assignment algorithm, click on "*Segment Assignment*" and then "*Auto-Assign*." While the time to run the algorithm varies, expect it to take 1-2 minutes. Do not click any other buttons while waiting for the auto assignment to complete.

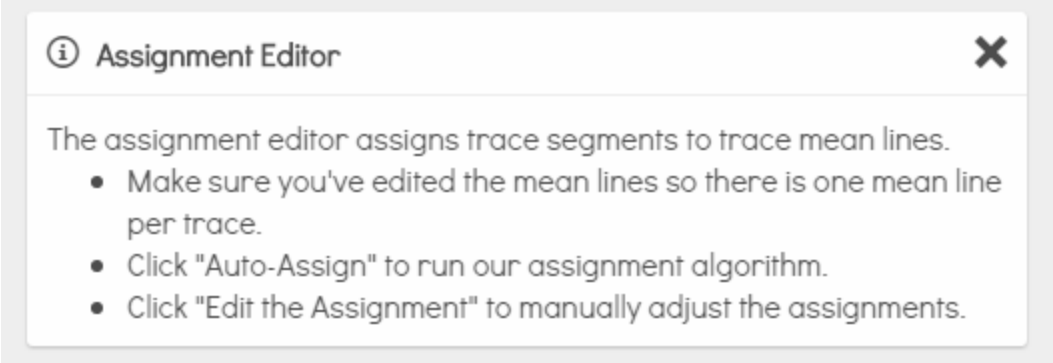


Editing Segment Assignment

Once the segment assignment algorithm has been run, the image can be inspected for accuracy, and edited again. Click on "*Segment Assignment*" and the following menu will appear.

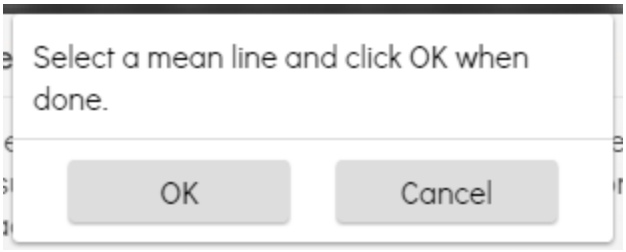


And clicking on "*Show Info*" will pull up the explanatory menu:



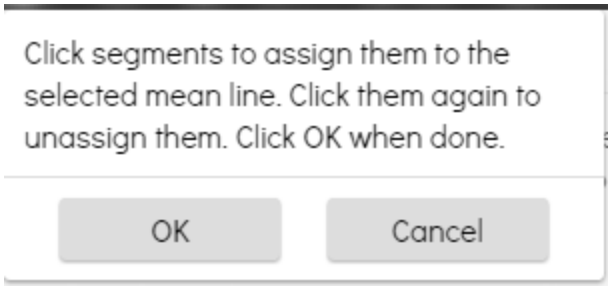
As the menu explains, it is important that one meanline is present and drawn to reasonable accuracy for each hour or partial hour of the seismogram[5]. This step should have already been performed during previous editing.

To change the assignment of a segment to a meanline, click on "*Edit the Assignment*." The following menu appears.



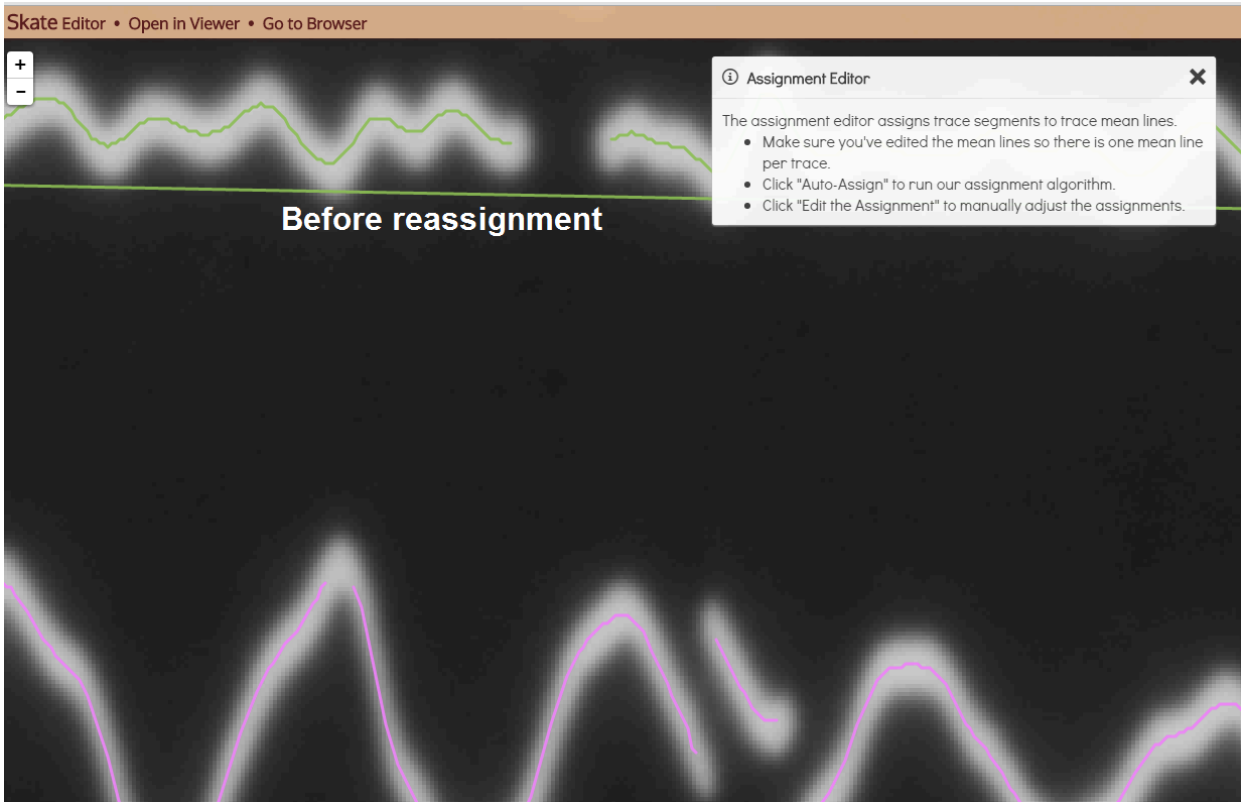
Click on the meanline to which you which to assign a misassigned segment, then click OK.

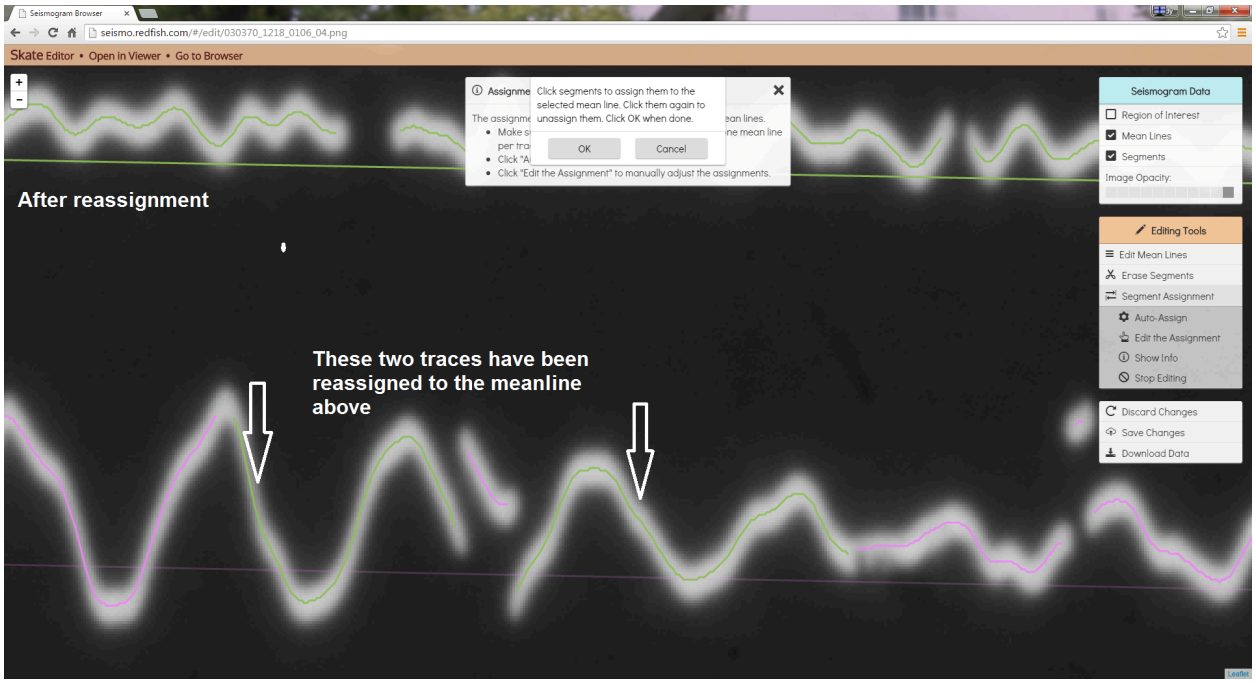
A new menu will appear.



Next, click on the segment to be reassigned, then click "*OK"*. More than one segment can be selected; continue to click on segments that will be reassigned. They will change color to match the meanline to which they will be reassigned as demonstrated in the image below (noting that the image shows an improper assignment, but nevertheless is shown here for instructional purposes). When selection is completed, click OK.

[5] Noting that we are currently only processing later date WWSSN long period seismograms, in which the length of a single line corresponds to one hour of recording time. Refer to the WWSSN Users Guide previously referenced for more information on seismogram characteristics.





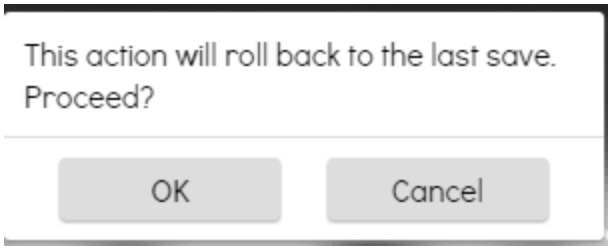
When reassigning is complete, the data can be saved or discarded. Also, the segment assignment algorithm can be run again in case the reassignment of segments changes the assignment of other segments. At this time we do not recommend running the assignment algorithm after editing the assignment.

Saving or discarding changes. Downloading data.

Important note on saving changes. SKATE is currently set up to store the raw data on S3; it will not be changed by any editing. However, changes to edited data will be permanently changed on S3; therefore the user needs to be very careful to make only those changes that are correct and necessary. As currently configured, the philosophy behind editing and saving is that multiple users can correct and improve any given seismogram; it is the availability of input from the broader seismic community that allows for the best result to be ultimately obtained. In addition, each save requires a transfer of data to S3, which incurs data transfer costs. Because there is a very limited operational budget, please limit your saving activities. Note: Only users with login credentials are able to save changes to S3.

Discard Changes

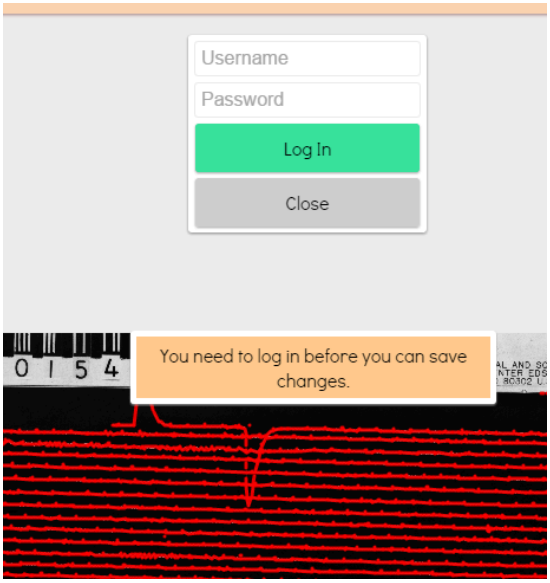
At any time any changes made can be discarded. This will revert the data back to the last saved value. Select "*Discard Changes*" and the following menu will appear.



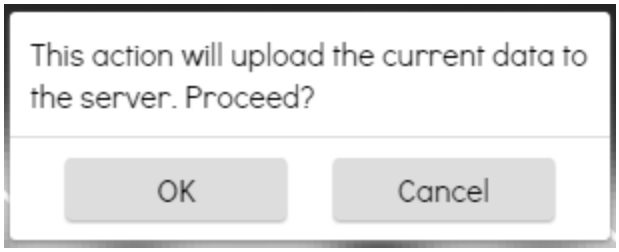
Click "*OK*" to discard changes.

Save Changes

To save the changes, click on "*Save Changes*" and the following menu will appear.



If you have login information, enter it now in order to proceed. Once logged in, the following menu item will appear.



Select "*OK*". This uploads the data to the S3 server.

Download Data

Once the seismogram is complete, the results can be downloaded to the user's machine by clicking on "*Download Data*". The data will be saved as a .zip file with the name of the seismogram record. You must be in Editor mode in order to access the download menu.

All of the data will be saved in a JSON format. Assignments will also be saved as a .csv file. Refer to www.json.org for details on the JSON format.

There are four JSON data files that are downloadable after analysis: roi.json; meanlines.json; segments.json; assignment.json. Their descriptions follow.

roi.json

The .json file looks like the following.

{"type":"FeatureCollection","features":[{"geometry":{"type":"Polygon","coordinates":[[[184,813],[14442,938],[14381,5555],[144,5431],[184,813]]]},"type":"Feature","id":null,"properties":{}}]}

The coordinates start in the upper left hand corner and follow a clockwise path around the image. Note that there are five coordinate pairs; the fifth is identical to the first, closing the polygon that defines the ROI.

meanlines.json

Meanlines are reported with an ID, and start and finish coordinates in an (x,y) format. The first few lines of the .json collection is shown below.

{"type":"FeatureCollection","features":[{"geometry":{"type":"LineString","coordinates":[[0,4356],[15456,4716]]},"type":"Feature","id":0,"properties":{}},{"geometry":{"type":"LineString","coordinates":[[0,4041],[15454,4378]]},"type":"Feature","id":1…}

Note that there may be gaps in the meanline ID numbering, and that they are not necessarily in a top to bottom order. This is a result of editing. Moving, adding or deleting lines will affect only the ID of the meanline being operated on, but will not affect other IDs.

The ID is an important identifier, as later it is used as the binding element to which all segments' centerline values are tied.

segments.json

Segments.json presents as:
{"type":"FeatureCollection","features":[{"geometry":{"type":"LineString","coordinates":[[10529,4713],[10530,4713]]},"type":"Feature","id":6769},{“geometry”…}

The “coordinates” are the (x,y) pairs of the centerlines for each segment. The coordinate system starts at the upper left of the image, with +y in the downward direction, and +x to the right. The “id” is the segment ID. This segment ID is used in conjunction with the meanline ID in the subsequently described assignment.json file.

assignment.json

The assignment.json file describes those segments that are associated with each meanline. The file looks like:

{"0":[segment ID 0-1, segment ID 0- 2,…segment ID 0-n],"1":[segment ID 1-1, segment ID 1-2,… segment ID 1-n], “N”:[segment ID N-1…segment ID N-n]}

The values in quotes are the meanline IDs. The values in brackets are the IDs of the segments belonging to that ID. The segment IDs are not necessarily sorted from left to right on the image.

assignment.csv

To download a .csv file of the assigned (i.e. connected) segments, choose ‘Open in Editor’ and select ‘Download CSV.’ The data is organized with column headers as: x0, y0; x1,y1;...;xN,yN. Each (xN,yN) pair corresponds to a single meanline, which in the case of long period WWSSN seismograms represents a one hour line of data. Note that typical spreadsheet plots, such as in Microsoft Excel, display positive Y values in Quadrant I of cartesian coordinate systems. However, the data in SKATE was obtained using Python, which plots positive Y values in Quadrant IV. Hence, Excel plotted data will be flipped about the horizontal axis with respect to the original image. Again, note that the meanline IDs will not necessarily be sorted from top to bottom, as editing and other actions might alter this sequence. Nevertheless, the data is easily parsed and represents the true output of the program.

mSEED

Not currently available, the data will be made available in mSEED format in the future.

Logout

Logging out is no longer required.

Final

As discussed, SKATE is still undergoing significant development, and is not yet a complete digitization package. Future work will address the most important needs including, but not limited to: better meanline identification, improved rejection of spurious features, and the incorporation of advanced segment connection algorithms. Bug reports and comments on SKATE should be directed to Dr. Andrew Bartlett at andy@retrievertech.com.

Acknowledgements

Retriever Technology would like to acknowledge the Department of Energy for supporting this work under contract DE-SC0008219. The efforts of our program manager, Leslie Casey, are greatly appreciated.

The New Mexico Small Business and Assistance Program (http://www.nmsbaprogram.org) allowed us to partner with Sandia National Laboratories to investigate promising technologies. Bill O'Rourke directed Sandia's efforts. Thanks, Bill.

In Santa Fe, there is a very strong community of researchers, software developers and business professionals that we were fortunate to be able to tap into. Our partnership with the Redfish Group was invaluable in developing many of the technologies used in SKATE. Stephen Guerin led the effort, with important input from Frank Wimberly, Owen Densmore, and others.

The core of the website and the underlying algorithms were created by an amazingly talented group of developers. Marius Nita and Ben Lichtner developed the impressive user interface and its underlying code; if you like this website and what it does, thank them. Benamy Yashar took a wide range of image processing tools and molded them into a processing pipeline that is impressive in so many ways. And our summer intern, Lowell Bartlett, did a superb job creating segment connection algorithms and automatic feature detection code.