



## Centaur Data Processing

Congratulations, you have offloaded a service run and have data from each Centaur. Follow the steps in this document to reorganize the data into station/channel/day volumes. Then, create a stationXML for your experiment using Nexus (see step 6) before submitting data to the EarthScope Primary Instrument Center (EPIC). **Unix commands (bold print) and any command line arguments are highlighted in yellow.**

Additional documentation can be found on the EPIC website:

<https://www.passcal.nmt.edu/content/passive-source-seed-archiving-documentation>

**1. Create an organized directory structure for your data.** Start by creating a main directory for the project. Under your main project directory, make a first level directory "SVC1" for service run number 1. For each subsequent service run create a new directory, e.g. SVC2, SVC3. Create a directory within the SVC1 directory for your raw data. For example: **mkdir RAW**

**2. Copy your data.** Copy the data folder from each Centaur into the RAW directory. Each Centaur should have a separate folder in the RAW directory with subfolders for each day of data.

**3. Reorganize the miniSEED data into station/channel/day volumes.** **Dataselect** is an EarthScope program that allows for the extracting and sorting of miniSEED data (<https://github.com/iris-edu/dataselect>). This will read data from the Centaur folders in the RAW directory and reorganize them into day volumes with the required naming format. You will need to run **dataselect** twice; first, to reorganize the waveform data; and second, to de-multiplex the SOH data.

The "Archive directory pattern" on the Centaur and naming of files are programmable. The example below is for an experiment with the Archive directory pattern set to "YEAR/Julian day" and filename extension set to ".miniseed".

For the waveform data:

```
dataselect -A DAYS/%s/%s.%n.%l.%c.%Y.%j RAW/*/20??/??*/*.miniseed
```

For the SOH data:

```
dataselect -A DAYS/%s/%s.%n.%l.%c.%Y.%j RAW/*/20??/??*/soh/*.miniseed
```

The (-A) flag writes file names in the specified custom format. The format flags are (s) for station, (n) for netcode, (l) for location, (c) for channel name, (Y) for year, and (j) for Julian date. See the help menu for more details on options (**dataselect -h**).

Depending on how much data you have, you may need to run **dataselect** in a loop that runs over the different days or stations in your experiment to avoid memory overflow errors. If you receive a message about “too many open files” you will need to run dataselect in a loop. A very simple example is given below.

*EXAMPLE: Run dataselect in a loop*

- a. Generate a list of miniseed files using the following command:

```
ls -D $PWD/RAW/*/*20??/???/*miniseed > mseed-file-list.txt
```

- b. Run dataselect in a loop with the following command:

```
cat mseed-file-list.txt | while read line; do dataselect -A DAYS/%s/%s.%n.%l.%c.%Y.%j  
$line; done
```

**4. Confirm your station and channel names.** In the DAYS folder just created by **dataselect**, check to see if you have folders for each of your stations. The data should be organized into those folders in station/channel/day volumes named STA.NET.LOC.CHAN.YEAR.JULDAY. For example: BA01.XR..HHZ.2018.039 (The .. after XR is where the location code would be if needed).

Since the station/channel names are programmed into the Centaur, you should not need to modify any of the headers in the miniSEED files unless a mistake was made when programming the Centaur at the site. To change any miniSEED headers to correct a station name, network code, etc., see the **fixhdr** doc on the EPIC website (see link on previous page). After you have modified the headers with **fixhdr**, rename the files so that the station-network-location-channel codes in the miniSEED file names match the corrected headers.

Note: Run fixhdr on the miniseed files in DAYS directory.

Note: To rename the files after any changes to the headers, rerun dataselect by pointing it to the recently updated files: **dataselect -A DAYS\_UPDATED/%s/%s.%n.%l.%c.%Y.%j DAYS/\*/\***

**5. Perform quality control of waveforms and SOH.** Verify the data quality by reviewing the traces and SOH files using **sohviewer** and **SQLX** (or other waveform quality control software of your choice). Obvious signs of trouble include loss of GPS timing, overlaps, gaps, corrupted files, etc. Make a note of any problems. Use **fixhdr** to mark timing issues, and/or to convert the files to big endian if they are not already. For more information on how to use these tools, refer to the appropriate documentation on the EPIC website (see link on the previous page). Note: If you make any changes to endianness or mark timing issues using **fixhdr** but DO NOT alter any header values, you do not need to rerun dataselect.

**6. Create metadata for your experiment.** Use Nexus to generate a stationXML file for your experiment metadata. See the “Metadata Generation with Nexus in a Nutshell” document on the EPIC website (see link on previous page).

**7. Send miniSEED data to PASSCAL.** Please drop a note, with your experiment name in the subject, to [epic.data.group@earthscope.org](mailto:epic.data.group@earthscope.org) before sending the data so that we can set up a receiving area. Attach the stationXML created with Nexus to this email unless it is larger than 5Mb. Use our tool **data2passcal** to send the data:

```
data2passcal DAYS/
```

**data2passcal** will scan all subdirectories of the DAYS folder and send any miniSEED files that have the correct file names.