Generating a PH5 Archive from Texan (RT125) and/or RT130 data

Created and last revised July 21, 2026

Table of Contents

Scope of this document:

This is a guide to building a PH5 from Texan (RT125/A) raw data, RT130 raw data, or both. Because Texans and RT130s have been superseded in PH5 by nodal instrumentation, information about building a PH5 using these dataloggers will not be updated after the date on this guide.

Note that regarding input data formats compatible with PH5:

- The data must be in the relevant proprietary RefTek format
- If you have SEG2 data, see Tips/Essential Notes below
- Other data formats (e.g., miniSEED, SAC, SEG-Y) *cannot* be ingested into PH5
- Building a PH5 from SEGD data (Fairfield, SmartSolo) is not covered in this guide; see 'Generating a PH5 Archive from SmartSolo and/or Fairfield Data'

If you do not already have it on your machine, install PH5 using the directions at https://github.com/PIC-IRIS/PH5/wiki/Installation-Instructions, and install with full GUIs (environment-gui.yml).

In this document, scripts and commands are **bolded**, commands are highlighted yellow, and GUI items are highlighted gray.

Essential terminology

- PH5: a version of the HDF5 hierarchical data format.
- master.ph5: The PH5 file with metadata. A PH5 can only have one master.ph5.
- miniPH5_xxxxx.ph5: The PH5 files with data, where xxxxx is a number (e.g., 00001).
- Texan .TRD: raw Texan (RT125 or RT125A) data file.
- RT130 .ZIP or .cf or .ref: raw RT130 file or folder.
- Array table(s): Metadata table(s) with .kef extension describing the stations in the PH5.
- Event table(s): Metadata table(s) with .kef extension describing any events (sources) in the PH5.
- DAS tables: Data tables identified by datalogger serial number, with .kef extension; do not modify except as directed.

Tips/Essential Notes Prior to Starting

- Do NOT try to load RT130 data with the **pforma** GUI because it is highly likely to result in an invalid PH5 volume. This functionality is disabled in PH5 installed or updated after December 2024. If you have only RT130 data and did not use active sources, PH5 is not the recommended archive format; archive in miniSEED instead.
- In general, data recorded on the Geometrics Geode or Stratavisor datalogger (SEG2 .dat files) does not go into PH5; only do so if the data are in UTC time and have some level of GPS-provided timing. If you do add SEG2 data, load it with the seg2toph5 utility (do NOT use **pforma**) and be advised you will need to create your own datalogger response file.
- If you installed a station with the RT130 datalogger but no data was recorded, do not include that station in the metadata because it results in a validation error.

- Do NOT build multiple PH5s for the same experiment. If you have a need to build a PH5 partway through, for example, a 2 year RT130 deployment, contact data-help@earthscope.org to discuss your needs.
- Verify your PH5 install is reasonably up to date. Run **ph5_validate -h** and if the version is less than 2022.066, you need to update your PH5 install. If you installed PH5 before May 2021, you will need to do a fresh installation of PH5 because the GUI package dependencies changed.
- Check the raw data size and verify that you have at least two times that space available on your machine.
- PH5 loads all data files fed to it in an input list if it is able to do so: if you offloaded raw files to more than one directory, it is strongly recommended that you check for duplicate files in the input list. Duplicated data in PH5 is not desirable because it unnecessarily increases the size of the archive and can lead to downstream issues.
- PH5 automatically retains all metadata tables loaded, so if you need to update a table, follow directions carefully.
- You need an FDSN network code and DMC assembled ID to complete the PH5. If you are building a PH5 for a PASSCAL experiment, log in to PASSCAL and fill out the mobilization form. If you are building a PH5 for a non-PASSCAL experiment, go to the FDSN temporary network request and specify IRISPH5 as the web service when you complete the form.

0. Preparatory steps

First, set up the directory structure, such as

mkdir exp_name
mkdir exp_name/Archive-PH5

The Archive-PH5 (or PH5v1, ph5v1, etc.) is the directory in which the PH5 will be built. Then, make a metadata directory:

mkdir exp_name/metadata

Optionally, also make a directory to hold all of the raw data:

mkdir exp_name/RAW

Go to step 1.

1. Input lists

Make your input file lists. If you have Texan data, go to step 1.a.; if you have RT130 data, go to step 1.b. The commands in this step assume the RAW directory is a subdirectory in the main experiment directory; modify commands as needed if this is not the case. If you have raw data without associated metadata (e.g., station, coordinates, deploy and pickup time, etc.), it is strongly recommended that you remove those files from the input list.

1.a. (Texan only)

Make a list of the TRD files. For example:

ls -d $PWD/RAW/*/*.TRD > trdlist.txt

where the first wildcard searches different offload directories. This list should include all of the TRD files you will input. Make sure all of the TRD files are named in the format INNNNDETAIL.TRD, where I is capital letter i, NNNN is the four-digit Texan serial number, and DETAIL is optional, can be longer than 6 characters, and can contain information such as offload date or a note about the location. If you also have RT130 data, continue to step 1.b.; if not, go to step 2.

1.b. (RT130 only)

If you have RT130 data, make a file list. For example, if you offloaded data in ZIP files:

ls -d $PWD/RAW/*.ZIP > ziplist.txt

If you offloaded data in CF card directories:

ls -d $PWD/RAW/*.cf > cflist.txt

Modify the command as needed if you used different offload directories. Make sure the ZIP files are named in the format YYYYJJJ.AAAA.ZIP, where YYYY is the four-digit year, JJJ is the day of year & nominally the day recording started, and AAAA is the 4-character serial number of the RT130. Make sure the CF card directories are named in the format AAAA.cf, where AAAA is the 4-character serial number of the RT130. If you have RT130 files offloaded with .ref extension, they also need to be named with the AAAA prefix.

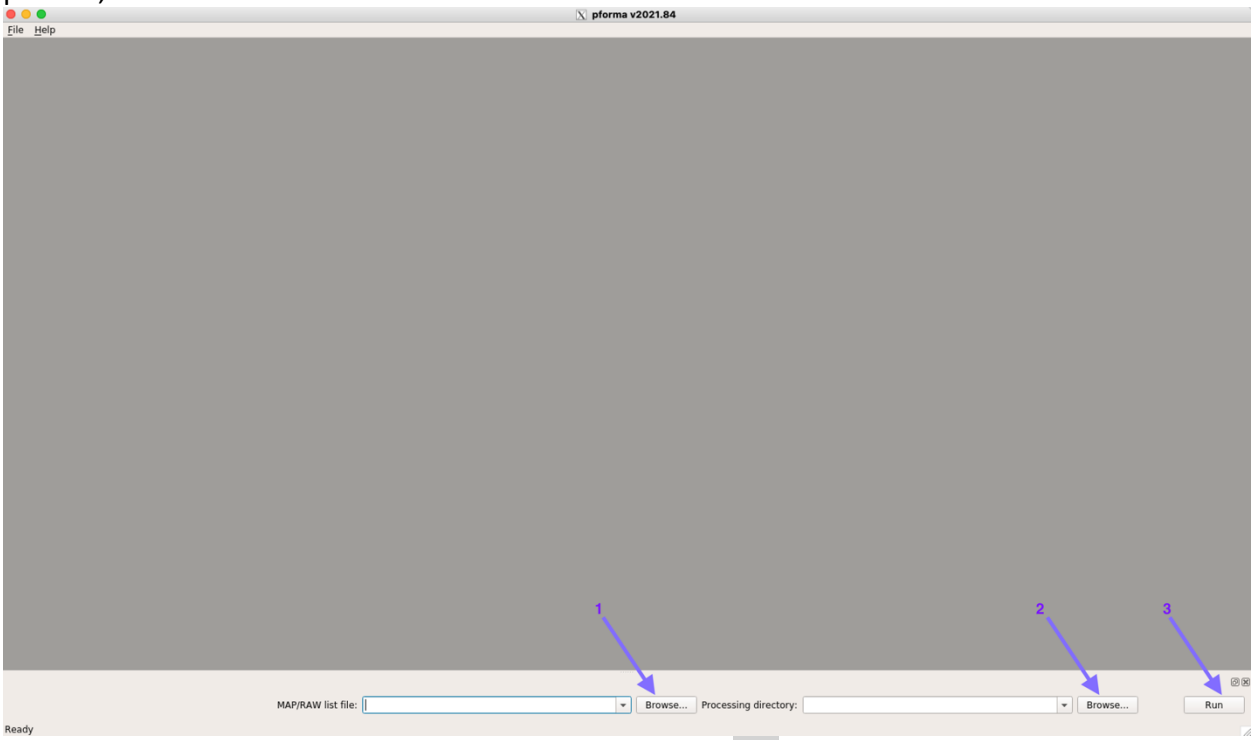
If you have only RT130 data, go to step 3. If you have both Texan and RT130 data, go to step 2 and then step 3.

2. Load Texan data files

Now you will load files into PH5 using **pforma**, a GUI that multithreads the process for faster building of the PH5 archive. Start **pforma** going to the experiment processing directory and running:

pforma

Click Browse to the right of MAP/RAW list file and select the trdlist.txt file you made earlier. Next, click Browse and select the PH5 directory (e.g., Archive-PH5) you made earlier. Then, click Run. It may take a few minutes for the processing directories appear in pforma. This does not mean pforma is frozen: let the program continue to run. See other notes on pforma, below.



Wait until all processing directories in the GUI show the Log button and the lower left corner of **pforma** shows the message 'All processes finished. Wrote pforma.cfg'. Then, click File -> Merge. After both the Merge Progress and Completed popups appear in **pforma**, you can close the popups and **pforma**. If you also have RT130 data, go to step 3. Otherwise, go to step 4.

Notes on **pforma:**

-If a progress bar turns orange, this is a failure to ingest data. Do not proceed with any further steps to build the archive; instead, click Log to see where it failed and details on the problem; there is also a menu to save the log file. You will need to exit pforma and after the problem file(s) is/are resolved (contact epic.data.group@earthscope.org if you are unable to resolve after three attempts), make a new PH5 directory (e.g., mkdir PH5_v2) and start again. If you are low on disk space, remove the original directory or back it up on external storage before starting again.

-Progress bar scrolling: this is a known bug and tends to occur when only one file has been assigned to the processing directory. It is not cause for concern.
-Progress bar does not reach 100%: this is known to occur when a large number of files (few hundred or more) are assigned to a processing directory. It is not cause for concern.

3. Load RT130 data files

See the Notes in this section before proceeding. If the experiment consists solely of RT130 data, go to step 3.a. If it also has Texan data and you followed step 2, go to step 3.b. If you have Texan data and did not do step 2 yet, go back and complete it before this step.

3.a. Experiment has RT130 data only

Go into the Archive-PH5 directory you created earlier.
Start the PH5 volume:
initialize_ph5 -n master.ph5

This command sets up the initial required file and a few basic metadata tables within the file.
Next, add (ingest) the data to the PH5:
130toph5 -n master.ph5 -f ../<filelist>

Where <filelist> is either the ziplist.txt or cflist.txt file you made in step 1.b. and is assumed to be one directory above the Archive-PH5 directory. Progress ingesting the data will be printed to screen and to 1302ph5.log. If any data fails to go into the PH5, please see the guidance in step 3.b under Notes. When loading data is complete, go to step 4.

3.b. Experiment also has Texan data

Notes: before proceeding to add RT130 data to the PH5 in the Sigma directory, you may want to make a test PH5 (e.g., mkdir exp_name/test1-PH5) using the process in step 3.a. to verify all RT130 data will go into the PH5 without error before running step 3.b. If there are power issues at any of the RT130 stations in your experiment, doing this is strongly recommended. If any RT130 data fails to go into the PH5, use the logged information and PASSOFT **logpeek** (installations prior to late 2023) or PASSOFT **sohviewer** (installations after late 2023) to check for any problems around the relevant time(s). Be sure to check the entirety of the log messages in logpeek or sohviewer for the relevant time(s), including ET errors. Do this for all affected RT130 zip files or cf card directories, then contact epic.data.group@earthscope.org for guidance before proceeding with any further steps in the PH5.

If you followed step 2 because you have Texan data, descend into the Sigma directory (e.g., cd Archive-PH5/Sigma). Then, add (ingest) the RT130 data to the PH5:
130toph5 -n master.ph5 -f ../../<filelist>

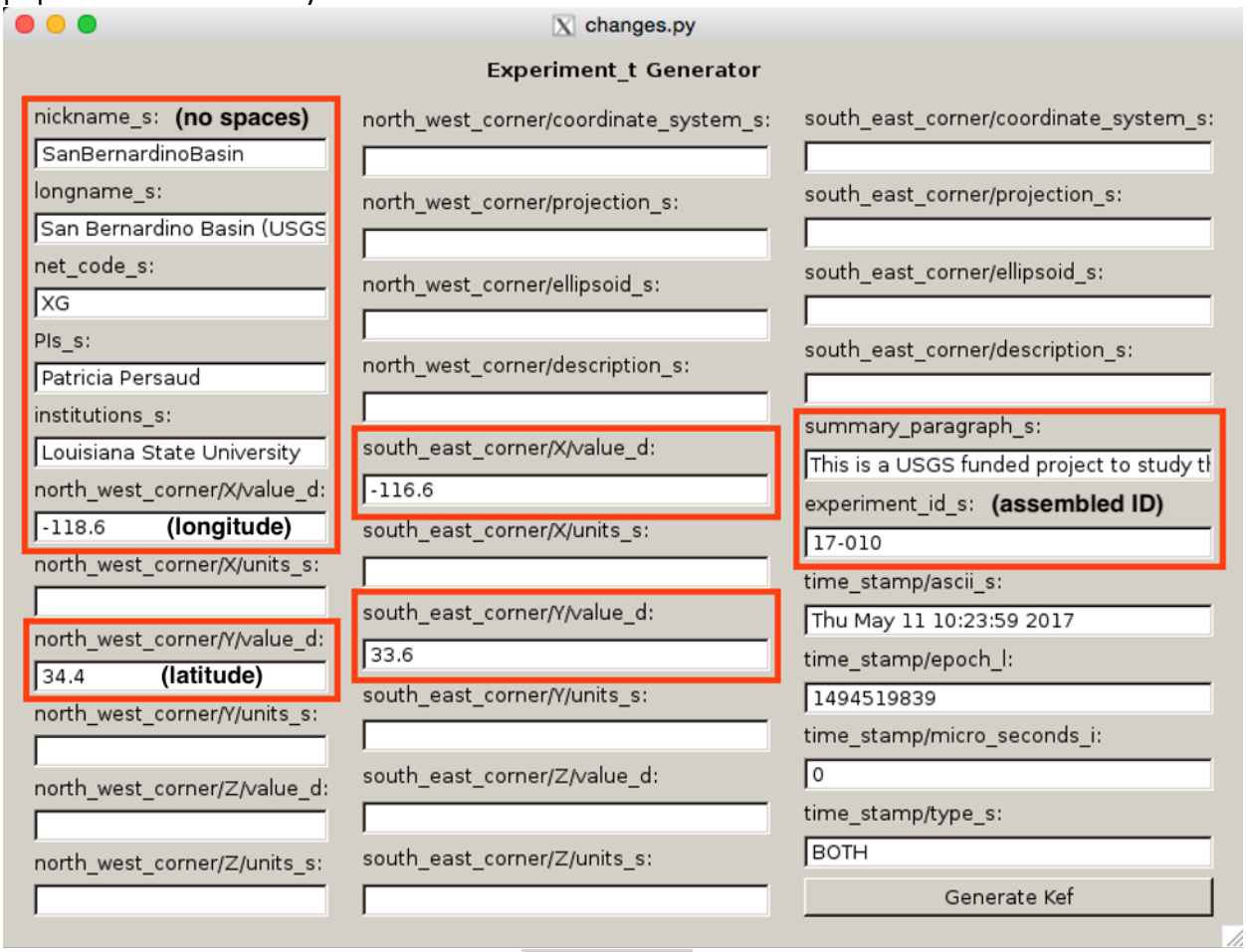
Where <filelist> is the either the ziplist.txt or cflist.txt file you made in step 1.b. Progress ingesting the data will be printed to screen and to 1302ph5.log. When it is complete, go to step 4.

4. Experiment table

Generate and load the experiment table. Note: you need to have a network code and assembled ID before you can complete this step.
Move into the combined directory:
cd PH5_Archive/Sigma

Start the experiment table generation GUI:
experiment_t_gen

The required fields are noted in red boxes, below. The last four fields on the right column are populated automatically and these fields should not be modified.



After filling out the required fields, click Generate Kef and save the file as experiment.kef. After exiting, load the kef:
keftoph5 -n master.ph5 -k experiment.kef

There is a kef2ph5.log file that tracks the filenames loaded, the timestamp, and the last time the file was modified before loading.
Go to step 5.

5. Load station metadata

5.a. Preparing station metadata

Make a .csv with the station metadata for the experiment. Each channel requires its own line in the .csv. If you have more than one array, you may make one .csv per array or use a single .csv that includes all arrays. Review the required fields guide and accompanying notes below. An example of a .csv is available on the EPIC website.

Required fields	**Description**	**Limitations**
Array	Array or line number for stations	Numeric, equal to or less than 999
Station number (id_s)	SEG-Y compatible station number	Numeric, less than 32,767
SEED station name	SEED compatible station name	Alphanumeric, 1-5 characters, all capital letters, may be the same as station number
DAS serial number	Full datalogger serial number	DAS serial number as marked in raw data headers, varies by manufacturer
Channel number	Channel number	Numeric, normally 1 for single channel dataloggers; 1, 2, or 3 for 3-channel; must match channel number in data headers (See Notes)
Sample rate	Sample rate for each recorded channel	Integer, must be greater than 1. For sample rates less than 1, adjust the sample rate multiplier (See Notes)
Sample rate multiplier	Multiplier for sample rate	Normally 1 unless sample rates less than 1sps are recorded (See Notes)
SEED channel name	SEED compatible channel name	3 characters, all capital letters (See Notes)
Latitude (location/Y)	Latitude in degrees	Positive or negative number; must be the same for all channels of a station
Longitude (location/X)	Longitude in degrees	Positive or negative number; must be the same for all channels of a station
Elevation (location/Z)	Elevation in meters	Numeric; must be the same for all channels of a station
Deploy time	Deploy time in appropriate date format	*YYYY:JJJ:HH:MM:SS.ssssss* where JJJ is day of year
Pickup time	Pickup time in appropriate date format	*YYYY:JJJ:HH:MM:SS.ssssss* where JJJ is day of year
DAS manufacturer	Datalogger manufacturer	Enter Reftek or REFTEK
DAS model	Datalogger model	If Texan, enter rt125; if RT130, enter rt130; all capital letters is acceptable
Sensor manufacturer	Sensor manufacturer	Enter the relevant manufacturer, such as Geospace, Sercel, MarkProducts, etc.

Sensor model	Sensor model	Enter the relevant model, such as GS11V, L28, L22, etc.

Optional fields	**Description**	**Limitations**
Sensor serial number	Sensor serial number	Varies by manufacturer
Description	Station description/comments	Optional station comments in ~80 ASCII characters or less; if used, will appear in Site column/field in EarthScope DMC metadata queries

Important Station Metadata Notes

- A given datalogger serial number can only be assigned to one station at a time for the experiment; review the station metadata now to be sure this is the case, to prevent errors later in the PH5 process.
    - If you find this is not the case and the datalogger in question is assigned to two or more stations at the same coordinates, keep one station.
    - If you find this is not the case and the datalogger in question is assigned to two or more stations at different coordinates, you must remove all stations for the datalogger in question unless you can determine the datalogger's true location.
    - Be aware that if you have RT130 data you can use the PASSOFT **logpeek** (prior to late 2023) or **sohviewer** (after late 2023) utility to plot coordinates recorded by the unit's GPS clock. However, Texans do not record any information about their coordinates. Therefore, you may be able to determine the true location of a RT130 without field notes, but cannot do so for a Texan.
- Texan serial numbers in PH5 are the unit serial number plus 10000
- If you opted to record more than one sample rate on the same station, place each sample rate in its own array, such as stations 1001, 1002, and 1003 at 500 Hz in array 1 and stations 1001, 1002, and 1003 at 100 Hz in array 2.
- If the sample rate is less than 1, increase the multiplier; for example, mass position recording at 0.1 Hz has a sample rate multiplier of 10.
- Channel numbers should be assigned as follows for both RT130 and Texan stations:
    - Orientation Z (or 3 if sensor was not vertical) is channel number 1
    - Orientation 1 (nominally N-S component) is channel number 2
    - Orientation 2 (nominally E-W component) is channel number 3
- SEED channel names if using geophones (e.g., GS11), where ? represents the orientation 1, 2, or Z:
    - GP? if 1000 Hz sample rate
    - DP? if 250 or 500 Hz sample rate
    - EP? if 100 or 200 Hz sample rate
- SEED channel names if using a high frequency/short period sensor (e.g., L-28LB, L-22) where ? represents the orientation 1, 2, or Z:
    - GH? if 1000 Hz sample rate
    - DH? if 250 or 500 Hz sample rate
    - EH? if 100 or 200 Hz sample rate

- For a more extensive guide to SEED channel names as well as information on sensor references in PH5, see Appendix A.

5.b. Check and load station metadata

Once your .csv file is ready and you have verified all columns are correct according to your field notes and documentation, start the noven GUI:

noven

Select File -> Open and select the event .csv file.
Next, use the dropdown above each column to specify which column is the station array, station ID, etc. A brief description of the column shows in the lower left corner when you hover over each option in the dropdown. When you are finished, select File -> Change Parameters, and for Skip Lines enter 1. Click Apply, and then click Close.
Next, select File -> Check input. Noven will verify all fields in the open .csv file. If the file passes verification, a popup with the message 'Nothing funky found!' will appear. Click OK, then select File -> Save As to save as a kef (e.g., arrays.kef) that can be loaded into PH5. If the file does not pass verification, a message window will appear with information about the affected line(s) and column(s).
If you have multiple station arrays and opted to make one .csv per station array, repeat these steps as needed (e.g., array_1.kef, array_2.kef, etc.). Close **noven** when you are done.
Next, load the station array kef(s) into the PH5. Make sure you are in the Sigma directory.

keftoph5 -n master.ph5 -k /path/to/arrays.kef

If your experiment used active sources and you are loading events, go to step 6. If not, go to step 7.

6. Load event metadata

6.a. Preparing event metadata

Events in the PH5 can include explosive sources, hammer strikes, local earthquakes, or other sources of interest during the experiment. Prepare a .csv file with the following required fields: event array number, event ID, latitude, longitude, elevation, and UTC time. Optional fields include description, event size & units, and depth. You can have one .csv file per event array, or include all event arrays in the same .csv file. If you made a metadata directory earlier, place the .csv file(s) there. Tables describing the required and optional fields are below.

Required fields	**Description**
Array	Event line/array number. Numeric, $1 \leq n \leq 999$.
Event ID (id_s)	SEG-Y compatible event number, less than 2,147,483,647. Unique for experiment.
Latitude	Latitude in degrees
Longitude	Longitude in degrees
Elevation	Elevation in meters
Event time	Time formatted as *YYYY:JJJ:HH:MM:SS.ssssss* where JJJ is day of year, or *YYYY-MM-DDTHH:MM:SS.ssssss*

Optional fields	**Description**
Event size	Numeric source size if available.
Event size units	Units for source size if available: kg, g, lb, magnitude, etc.
Depth	Source depth below surface in meters. Below surface is positive, above is negative.
Description	ASCII text field for providing additional information about a source. Limit to 80 characters.

Notes: having more than 20 event arrays and more than 5 station arrays, or any equivalent combination, is considered an edge case and may not be supported by PH5. Station array numbers and event array numbers can be re-used, e.g., station array 2 and event array 2 can exist in the same PH5, and a station ID 2002 and event ID 2002 can also exist in the same PH5. However, a PH5 cannot have two event IDs 2002, nor can event ID 2002 be assigned to event array 2 and event array 4, only one event array.

6.b. Check and load event metadata

Once your .csv file is ready, start the noven GUI:

noven

Select File -> Open and select the event .csv file.

Select File -> Change Parameters. In the menu that comes up, click on the Input Type dropdown, select 'event', click Apply, and then click Close.
Next, use the dropdown above each column to specify which column is the event array, event ID, etc. A brief description of the column shows in the lower left corner when you hover over each option in the dropdown. When you are finished, select File -> Change Parameters, and for Skip Lines enter 1. Click Apply, and then click Close.
Next, select File -> Check input. Noven will verify all fields in the open .csv file. If the file passes verification, a popup with the message 'Nothing funky found!' will appear. Click OK, then select File -> Save As to save as a kef (e.g., events.kef) that can be loaded into PH5. If the file does not pass verification, a message window will appear with information about the affected line(s) and column(s).
If you have multiple event arrays and opted to make one .csv per event array, repeat these steps as needed (e.g., events_1.kef, events_2.kef, etc.). Close **noven** when you are done.
Next, load the event kef(s) into the PH5. Make sure you are in the Sigma directory.
keftoph5 -n master.ph5 -k /path/to/events.kef

Go to step 7.

7. Load responses

Request from the Nominal Response Library (NRL) version 1, compile your own RESP file in the same format as NRL version 1, or contact epic.data.group@earthscope.org to request the response file(s).
Place the RESP file(s) in the metadata directory if you made one, or in the Sigma directory.
Do NOT use response files from the new NRL web service because they are incompatible with the ph5tostationxml code.
Note: Make sure that stations are separated into different arrays if they have different datalogger models, sample rates, or gains.

Make sure you are in the Sigma directory, then run
resp_load -n master.ph5 -a <comma separated list of arrays>

where the array list is non-zero-padded array numbers, such as 1,5,9.
If this command finishes and produces an input.csv without any warnings about no DAS table found, proceed. The input.csv contains the framework of essential response information for all dataloggers, sensors, & settings (sample rate, gain) used in the PH5. You complete the Das RESP path and the Sensor RESP path. If the resp_load command finishes with warnings about no DAS table found, do not proceed further and contact epic.data.group@earthscope.org for assistance. If the input.csv only has the column description row after this step, do not proceed further and contact the EPIC data group at the above email for assistance. An example of an input.csv file generated by resp_load for a Texan experiment (top) and RT130 experiment (bottom):

```
Das Model, Sensor Model, Sample Rate, sample rate multiplier,Gain, Das RESP path, Sensor RESP path
rt125,GS11V,250,1,32,,
Das Model, Sensor Model, Sample Rate, sample rate multiplier,Gain, Das RESP path, Sensor RESP path
RT130,L28,250,1,32,,
```

Below is an example of an input.csv file after the RESP file names have been added for a Texan experiment (top) and RT130 experiment (bottom). In this example, the RESP files exist in the Sigma directory. The RESP files may exist elsewhere on the machine and can be referenced by relative or absolute filepath. Save the updated input.csv file.

```
Das Model, Sensor Model, Sample Rate, sample rate multiplier,Gain, Das RESP path, Sensor RESP path
rt125,GS11V,250,1,32,../../metadata/RESP.XX.NR337..DLZ.125A.32.250,../../metadata/RESP.XX.NS390..SHZ.GS11D.4_5.4000.56160.97_4
Das Model, Sensor Model, Sample Rate, sample rate multiplier,Gain, Das RESP path, Sensor RESP path
RT130,L28,250,1,32,../../metadata/RESP.XX.NR024..HHZ.130.32.250,../../metadata/RESP.XX.NS361..SHZ.L28LB.395.2490.4_5.27
```

Then, run:
resp_load -n master.ph5 -a <comma separated list of arrays> -i input.csv

This will load the RESP files, update response references in the array table, and also update receiver table references in the array table using default orientations. Go to step 8.

8. Initial validation

This involves validating the PH5 and the stationXML. PH5 validation is included with PH5. The stationXML validator can be downloaded from the EarthScope github, and it is included with PASSOFT installed after 2021. It is not included with PH5.
To validate the PH5:
ph5_validate -n master.ph5

Progress information will be printed to screen, and the validation information will be saved in ph5_validate.log.
To extract and validate the stationXML:
ph5tostationxml -n master.ph5 -o exp-resp1.xml --level=response
stationxml-validator --input exp-resp1.xml --ignore-warnings

On some systems, you may have to run the validator by:
java -jar ${CONDA_PREFIX}/jar/stationxml-validator.jar --input exp-resp1.xml --ignore-warnings

Make note of errors and go to step 9.

9. Updating tables

Check that the station coordinates, start & stop times, array numbers, station IDs, and channel orientations are as intended. It is recommended to view the array tables as a csv. Run
ph5tokef -n master.ph5 --all_arrays > arrays_all.kef
keftocsv -f arrays_all.kef -o arrays_all.csv

Examine the array tables with preferred viewer. Recommended checks when viewing array tables:

- Station coordinates: verify from field notes
- Datalogger only assigned to one station at a time: verify from field notes and ph5_validate.log
- Start and stop times: verify from field notes; if a station has more than one epoch, the stationXML validator will flag if there is an overlap
- Verify other fields (sample rate, channel number, orientation, etc.) as needed based on ph5_validate and stationXML validator output

Based on your checks, field notes, and initial validation results, then:

- If you need to update station coordinates or channel orientation, and/or change array number, station ID, start or stop times, or other station metadata from what you originally listed, go to step 9.a.
- If you only need to change channel orientation from the default (Z +90 dip, 1/N 0 azimuth, 2/E 90 azimuth) go to step 9.b.
- If you only need to correct errors in the event table, go to step 9.c.
- If the ph5_validate.log contains one or more error messages about Das_t channel, time order and there are no other updates, go to step 9.d.
- If you used Texans with 2 or 3 channels at a station (i.e. usual Texan vertical channel plus one or two horizontal channels) and there are no other updates, go to step 9.e.
- If you need to correct errors in the experiment or response table, go to Appendix B.
- If all station coordinates and other metadata are as intended and there are no errors from ph5_validate or the stationXML validator, go to step 10.

Note: if you prefer to edit csvs in Microsoft Excel or any other software that detects column format and assigns attributes, proofreading for un-commanded or undesirable changes in a plain text editor is strongly recommended. This is particularly important if you are using RT130s.

9.a. Update array table

Open csv in preferred editor. Update the relevant metadata. Coordinates must be in decimal degrees (latitude & longitude) and meters (elevation). Note that location/X is longitude and location/Y is latitude; see Appendix C for a table fully describing station metadata requirements for array tables extracted from PH5. Please note that if you update start and/or stop times, you need to modify both the ascii and epoch timestamp columns. Also correct any station errors

(e.g., station ID too large error from ph5_validate, overlapping station epochs from stationxml validator) from step 8. If ph5_validate issued an overlap time for station error, see step 5's Important Station Metadata Notes for how to proceed with removing metadata. For each station requiring updates e.g. to coordinates, change all channels. Do not modify columns unless needed or modify the format used for columns (such as the ASCII date), as doing so can invalidate the PH5. Save the csv under a new filename and convert to a kef:
csvtokef -f arrays_updatedv1.csv -o arrays_updatedv1.kef

Note: the channel orientation convention 1,2 (e.g., GP1, GP2) is recommended to follow GSN convention and is required if sensors were installed more than 5 degrees from geographic north. GP1 is nominally N-S and GP2 is nominally E-W.
If some sensors were installed not oriented to geographic north, go to step 9.b. If all sensors were installed oriented to geographic north but you followed step 6 and have one or more errors in the event table, go to step 9.c. If sensors were installed oriented to geographic north and you have no errors in the event table, but the ph5_validate.log contains a channel, time order error, go to step 9.d. If it does not and you used Texans with 2 or 3 channels at a station, go to step 9.e. Otherwise, go to step 9.f.i.

9.b. Update receiver and array table

The receiver table will need to be edited for this step, as well as the array tables.
ph5tokef -n master.ph5 -C > receiver_original.kef
keftocsv -f receiver_original.kef -o receiver_original.csv

Open this csv in preferred editor.
Use field notes to populate orientations in addition to the default 0,90. Important: do NOT change the first row. Below is an example of an updated receiver table csv with notes:

	A	B	C	D	E	F	G	H	I	J	K	L
1	table	orientation/azimu	orientation/az	orientation/dip	orientation/di	orientation/descri	orientation/channel_number_i		index (implicit)		Stations	
2	/Experiment_g/Receivers_g/Receiver_t	0	degrees	90	degrees	Z	0		0			
3	/Experiment_g/Receivers_g/Receiver_t	0	degrees	0	degrees	N	0		1		1030, 1045, 1062, 1071, 1079	
4	/Experiment_g/Receivers_g/Receiver_t	90	degrees	0	degrees	E	0		2			
5	/Experiment_g/Receivers_g/Receiver_t	9	degrees	0	degrees	N	0		3		1000, 1001, 1016, 1020, 1027	
6	/Experiment_g/Receivers_g/Receiver_t	99	degrees	0	degrees	E	0		4			
7	/Experiment_g/Receivers_g/Receiver_t	356	degrees	0	degrees	N	0		5		1002	
8	/Experiment_g/Receivers_g/Receiver_t	86	degrees	0	degrees	E	0		6			
9	/Experiment_g/Receivers_g/Receiver_t	23	degrees	0	degrees	N	0		7		1004, 1092	
10	/Experiment_g/Receivers_g/Receiver_t	113	degrees	0	degrees	E	0		8			

The index(implicit) and stations columns are for in-progress tracking only and if used, must be removed prior to saving the finalized version and converting to a kef, as shown below:

	A	B	C	D	E	F	G	H
1	table	orientation/azimut	orientation/az	orientation/dip/	orientation/di	orientation/descri	orientation/channel_number_i	
2	/Experiment_g/Receivers_g/Receiver_t	0	degrees	90	degrees	Z	0	
3	/Experiment_g/Receivers_g/Receiver_t	0	degrees	0	degrees	N	0	
4	/Experiment_g/Receivers_g/Receiver_t	90	degrees	0	degrees	E	0	
5	/Experiment_g/Receivers_g/Receiver_t	9	degrees	0	degrees	N	0	
6	/Experiment_g/Receivers_g/Receiver_t	99	degrees	0	degrees	E	0	
7	/Experiment_g/Receivers_g/Receiver_t	356	degrees	0	degrees	N	0	
8	/Experiment_g/Receivers_g/Receiver_t	86	degrees	0	degrees	E	0	
9	/Experiment_g/Receivers_g/Receiver_t	23	degrees	0	degrees	N	0	
10	/Experiment_g/Receivers_g/Receiver_t	113	degrees	0	degrees	E	0	

Open the array csv; if you needed step 9.a., use that csv. Go to the receiver_table_n_i column and enter the relevant receiver table indices for each station. Note the receiver table starts with an implied index of 0, which is the Z channel dip lookup.
Save the csvs under new filenames and then convert them to kefs:
csvtokef -f arrays_updatedv2.csv -o arrays_updatedv2.kef
csvtokef -f receiver_updated.csv -o receiver_updated.kef

If you followed step 6 and have one or more errors in the event table, go to step 9.c. If you have no errors in the event table, but the ph5_validate.log contains a channel, time order error, go to step 9.d. If it does not and you used Texans with 2 or 3 channels at a station, go to step 9.e. Otherwise, go to step 9.f.i.

9.c. Update event table

If you need to correct any errors in the event table(s), run
ph5tokef -n master.ph5 --all_events > events_original.kef
keftocsv -f events_original.kef -o events_original.csv

Open this csv in a preferred editor, and make the needed changes. Save under a new filename and convert to a kef:
csvtokef -f events_updatedv1.csv -o events_updated.kef

If the ph5_validate.log contains a channel, time order error, go to step 9.d. If it does not and you used Texans with 2 or 3 channels at a station, go to step 9.e. Otherwise, go to step 9.f.i.

9.d. Fixing validate channel, time order error

If the ph5_validate.log contains any error messages that 'Das_t isn't in channel, time order. Run fix_das_t_order to fix that.', run
fix_das_t_order -n master.ph5

Depending on the size of the archive and number of DAS tables, the command will take a few minutes to a few hours to complete. If you used Texans with 2 or 3 channels at a station, go to step 9.e. Otherwise, go to step 9.f.i.

9.e. Texan with 2 or 3 channels at a station

This case last occurred on an EPIC experiment in 2019 and is generally considered deprecated, but the required tools to handle it may still work. Making a backup of the entire PH5 is recommended before attempting this step. An example of this case is an experiment that installed 3 Texans at one station and connected them to a single L-28 sensor with specially designed cables. In this case, each channel gets assigned the channel number as described in step 5's Important Station Metadata Notes. However, DAS tables for Texans has channel

number 1 by default and PH5 will not be able to 'see' the channel 2 & 3 (horizontal) data by default, which results in numerous no data found errors in ph5_validate. To prepare DAS tables with the correct channel number, run:
fix_3chan_texan master.ph5

This will create a new DAS table with updated channel numbers, e.g., Das_t_10665.kef for each DAS in the PH5, as well as a backup of the original DAS table, e.g., Das_t_10665orig.kef. You will then need to load the new DAS tables with **keftoph5**; a script is recommended. Do not attempt DAS table deletion in this step, or you risk invalidating the PH5. Go to step 9.f.i.

9.f.i. Remove the metadata tables being replaced

PH5 automatically retains what is loaded, so tables being replaced first need the old version deleted from the PH5.
If you followed step 9.a. and/or step 9.b.:
delete_table -n master.ph5 --all_arrays

If you followed step 9.b., also run:
delete_table -n master.ph5 -C

If you followed step 9.c. to update the event table(s), run:
delete_table -n master.ph5 -V n

Where n is the event line number for each event line in the PH5. If you installed or updated PH5 after mid-February 2025 and plan to replace all event lines, you can run:
delete_table -n master.ph5 --all_events

After you have finished deleting the table(s) you will replace, go to step 9.f.ii.

9.f.ii. Load new metadata tables

If you followed step 9.a. and/or step 9.b., run:
keftoph5 -n master.ph5 -k arrays_updatedv2.kef

If you followed step 9.b., also run:
keftoph5 -n master.ph5 -k receiver_updated.kef

If you followed step 9.c. to update the event table(s), run:
keftoph5 -n master.ph5 -k events_updated.kef

Go to step 10.

10. Make ancillary tables

Generate ancillary tables for PH5 to use. Run:
sort_kef_gen -n master.ph5 -a > sort_t.kef
This will generate a sort table to optimize searches.

If the experiment has any events (you used step 6), also create the event-station offset table by running:
geo_kef_gen -n master.ph5 > offset.kef

If the experiment used Texans, additionally create the time correction table by running:
time_kef_gen -n master.ph5 > time.kef

Go to step 11.

11. Final validation & loading ancillary tables

This involves validating the PH5 and the stationXML again. PH5 validation is included with PH5. The stationXML validator can be downloaded from the EarthScope github, and it is included with PASSOFT installed after 2021. It is not included with PH5.
To validate the PH5:
ph5_validate -n master.ph5

Progress information will be printed to screen, and the validation information will be saved in ph5_validate.log. Note that your ph5_validate.log from step 8 will be overwritten unless you change the filename.
To extract and validate the stationXML:
ph5tostationxml -n master.ph5 -o exp-resp1.xml --level=response
stationxml-validator --input exp-resp1.xml --ignore-warnings

On some systems, you may have to run the validator by:
java -jar ${CONDA_PREFIX}/jar/stationxml-validator.jar --input exp-resp1.xml --ignore-warnings

If there are any remaining errors, go to step 9 and follow as needed to fix the errors. If you need to go through step 9.a. and/or step 9.c. again, repeat step 10.
If you did step 9.b., check that azimuth and dip values are as expected in the stationXML. If any are incorrect, go to step 9 and follow as needed to correct.
Now, load the table(s) from step 10.
keftoph5 -n master.ph5 -k sort_t.kef

If applicable, also run:
keftoph5 -n master.ph5 -k offset.kef

If applicable, additionally run:
keftoph5 -n master.ph5 -k time.kef

Go to step 12.

12. Test data extraction

Output miniseed, such as the following run from the Sigma directory:

ph5toms -n master.ph5 -a 009 -s 2024-04-10T00:00:00 -t 2024-04-11T23:59:59 -o ms_test -F MSEED

where -a 009 is the station array (leading zeroes required), -s and -t define start and end time respectively, -o defines the output directory, and -F defines the format. See **ph5toms -h** for details of the available flags.

If you followed step 6, also test SEGY output, such as the following run from the Sigma directory:

ph5toevt -n master.ph5 --use_deploy_pickup -o gather_test -l 10 -A 9 -e 12002 --shot_line 5

where -l defines length in seconds, -A is the station array, and -e defines the event number. See **ph5toevt -h** for details of the available flags.

Appendix A. Tables for SEED channels & instrument lookups

Below are tables that you may need to reference in step 5 (station metadata).

Sample rate and SEED channel name lookup by sensor type:

Instrument	Sample Rates (Hz)							
	1000 ≤ SPS < 5000	250 ≤ SPS < 1000	80 ≤ SPS < 250	10 ≤ SPS < 80	1 < SPS < 10	1	0.1	0.01
STS-2	FH?	CH?	HH?	BH?	MH?	LH?	VH?	UH?
CMG-3T	FH?	CH?	HH?	BH?	MH?	LH?	VH?	UH?
CMG3ESP	FH?	CH?	HH?	BH?	MH?	LH?	VH?	UH?
TR-240	FH?	CH?	HH?	BH?	MH?	LH?	VH?	UH?
TR-120	FH?	CH?	HH?	BH?	MH?	LH?	VH?	UH?
TR-40	FH?	CH?	HH?	BH?	MH?	LH?	VH?	UH?
CMG-40T 30s	FH?	CH?	HH?	BH?	MH?	LH?	VH?	UH?
CMG-40T 1s	GH?	DH?	EH?	SH?	MH?	LH?	VH?	UH?
S-13	GH?	DH?	EH?	SH?	MH?	LH?	VH?	UH?
HS-10	GH?	DH?	EH?	SH?	MH?	LH?	VH?	UH?
L-4C*	GL?	DL?	EL?	SL?	ML?	LL?	VL?	UL?
L-22*	GL?	DL?	EL?	SL?	ML?	LL?	VL?	UL?
L-28	GL?	DL?	EL?	SL?	ML?	LL?	VL?	UL?
L-28LB (4.5Hz geophone)*	GL?	DL?	EL?	SL?	ML?	LL?	VL?	UL?
L-40A (40 Hz geophone)	GP?	DP?	EP?	SP?	MP?	LP?	VP?	UP?
FBA ES-T	FN?	CN?	HN?	BN?	MN?	LN?	VN?	UN?

* The use of "H", denoting high gain, assumes that the dataloggers are programmed using a gain of 32, which is the recommended gain setting for a typical PASSCAL experiment. In the event the gain is set to 1 at the datalogger, then the second character of the channel name should be set to "L".

Sensor manufacturer & model to list in PH5 by sensor type for some commonly used sensors:

Sensor	PH5-Manufacturer	PH5-Model
CMG-3T	Guralp	cmg3t
STS-2 Gen 1	Streckeisen	sts2_g1
STS-2 Gen 2	Streckeisen	sts2_g2
STS-2 Gen 3	Streckeisen	sts2_g3
L-22	Sercel	l22

L-28	Sercel	l28
PIC Geophone (4.5Hz 1C)	Geospace	GS11V
Trillium 120	Nanometrics	trillium_120
Trillium 240	Nanometrics	trillium_240
Trillium 120PH	Nanometrics	trillium_120PH
Trillium Compact 120s	Nanometrics	trillium_comp
CMG-40T	Guralp	cmg40t
Hyperion	Hyperion	hyperion
Trillium 40	Nanometrics	trillium_40
CMG-40T 1sec	Guralp	cmg40t_1

Appendix B. Updating the experiment and/or response table

If you need to correct any errors in the experiment table, run
ph5tokef -n master.ph5 -E > experiment_original.kef

The experiment table is short enough that you can open it as-is in a text editor with the .kef extension. Make the changes required to fix the errors, then save with a new filename (experiment_updated.kef).
Then delete the old table:
delete_table -n master.ph5 -E

Load the new version:
keftoph5 -n master.ph5 -k experiment_updated.kef

If you discovered you loaded the incorrect response file(s) and need to correct the response table, make a copy of the original input.csv if desired, then run (following command will overwrite existing input.csv):
resp_load -n master.ph5 -a <comma separated list of arrays>

Complete the Das RESP path and the Sensor RESP path, as you did in step 7. Then, run
resp_load -n master.ph5 -a <comma separated list of arrays> -i input.csv -r

Appendix C. Station metadata requirements table

This table is for use ONLY with array tables extracted from the PH5; if you are still preparing your station metadata, see step 5.
The fields are listed in the order in which they appear in the array table(s).

Field name	Description	Directions and Limits
table	Array or line number for stations (e.g., /Experiment_g/Sorts_g/Array_t_001)	Numeric, equal to or less than 999; do not change non-numeric parts of the field
id_s	SEG-Y compatible station number	Numeric, less than 32,767
location/X/value_d	Station longitude in decimal degrees	Positive or negative number, must be the same for all channels of a station
location/X/unit_s	Specify longitude units	Default is degrees, do not change
location/Y/value_d	Station latitude in decimal degrees	Positive or negative number, must be the same for all channels of a station
location/Y/unit_s	Specify latitude units	Default is degrees, do not change
location/Z/value_d	Station elevation in meters	Numeric, must be the same for all channels of a station.
location/Z/unit_s	Specify elevation units	Default is m; do not change
location/coordinate_system_s	Optionally provide a coordinate system	Default is blank, okay to leave as default
location/projection_s	Optionally provide a datum name	Default is blank, okay to leave as default
location/ellipsoid_s	Optionally provide a geoid name	Default is blank, can be left blank
location/description_s	List information about the station location	Blank unless you entered optional description in step 5 (see step 5 tables); this field is limited to ~80 ASCII characters and the contents will show in the Site column/field in EarthScope DMC metadata queries (MDA, station query, etc.)

deploy_time/ascii_s	Deploy time in ASCII format	Date and time in whole seconds, such as Wed Apr 24 16:26:25 2024; if day is less than 10, space-pad (e.g., Wed Apr 3 16:26:25 2024); if epoch deploy time is updated, this field must be updated to match and it must be the same for all channels of a station
deploy_time/epoch_l	Deploy time in epoch seconds	Must be whole seconds in reference to standard epoch time; if ASCII deploy time is updated, this field must be updated to match and it must be the same for all channels of a station
deploy_time/micro_seconds_i	If applicable, decimal seconds	Must be a number equal to or less than 999999; in most cases can be left as default and must be the same for all channels of a station
deploy_time/type_s	None	Default is BOTH, do not change
pickup_time/ascii_s	Pickup time in ASCII format	Date and time in whole seconds, such as Wed May 22 06:00:00 2024; if day is less than 10, space-pad (e.g., Wed May 1 06:00:00 2024); if epoch pickup time is updated, this field must be updated to match and it must be the same for all channels of a station
pickup_time/epoch_l	Pickup time in epoch seconds	Must be whole seconds in reference to standard epoch time; if ASCII deploy time is updated, this field must be updated to match and it must be the same for all channels of a station

pickup_time/micro_seconds_i	If applicable, decimal seconds	Must be a number equal to or less than 999999; in most cases can be left as default and must be the same for all channels of a station
pickup_time/type_s	None	Default is BOTH, do not change
das/serial_number_s	Full datalogger serial number	DAS serial number as marked in raw data headers, varies by manufacturer; do not change
das/model_s	Datalogger model	Text string; see step 5
das/manufacturer_s	Datalogger manufacturer	Text string; see step 5
das/note_s	None	Can leave as default
sensor/serial_number_s	Sensor serial number	Varies by manufacturer
sensor/model_s	Sensor model	Text string; see step 5
sensor/manufacturer_s	Sensor manufacturer	Text string; see step 5
sensor/note_s	None	Blank by default, can leave as default
description_s	General information, not usually used in Texan/RT130 experiments	Blank by default, can leave as default
seed_band_code_s	SEED-defined character for band (sample rate-instrument corner)	1 letter, if not correct see step 5
sample_rate_i	Sample rate for each recorded channels	Integer, must be greater than 1, do not change
sample_rate_multiplier_i	Multiplier for sample rate	1 for most cases, see step 5 for exceptions
seed_instrument_code_s	SEED-defined character for instrument	1 letter, if not correct see step 5
seed_orientation_code_s	SEED-defined character for channel orientation	1 letter or number; use of 1,2,Z (1 corresponds to N, 2 to E) is strongly recommended to align with GSN and EPIC SEED convention
seed_location_code_s	Optionally, 2-character SEED-defined location code	Default is blank, leaving as default is strongly recommended; cannot use

		different location codes within a station
seed_station_name_s	SEED compatible station name	Alphanumeric, 1-5 characters, usually the same as id_s; if alphabetical must be all capital letters
channel_number_i	Channel number of each component	Do not change
receiver_table_n_i	Index to look up in the receiver table for channel's numeric dip and azimuth	Automatically populated during resp_load; do not change unless following step 9.b.
response_table_n_i	Index to look up in the response table for response information	Automatically populated during resp_load; if arrays are set up correctly for different parameters, leave as-is