Metadata Tips

1. Texan serial numbers in the metadata must have 10,000 added to their value for the csv file. So, Texan 2345 should be listed as 12345 and Texan 643 would be 10643. This reflects the internal serial number for the unit, which is represented as I2345 and I0643 in the file name.
2. SEG-Y station IDs must be numbers <= 65536; no letters, punctuation or special characters.
3. Consider using 4 or 5 digit station and shot IDs where the left most digit is the number of the line (array).
4. Lines (Arrays) must be numbers. Consider starting the line values at 1 for receiver arrays. An array or line is a logical grouping of stations.
5. In multi-component data: Z=1, N=2, E=3
6. When station sites are re-occupied (e.g. installed Texans are swapped with fresh Texans to allow continuous recording at a site) build a separate csv and receiver kef file for each deployment. In other words, no station and component combination should have more than one deployment and pickup cycle in any given csv or receiver kef. This makes it easier to replace metadata for an array if errors are discovered.
7. Note the format of each field in the example receiver and event kefs. Values using decimal places must use decimal places. Do not stray from these formats. The columns may be in a different order in the csv, however the format used in each field is not subject to debate!

Processing Tips and How to Nuke Files

If any metadata in the csv files is found to be incorrect after loading the subsequent receiver or event kefs into the PH5 file, delete the associated tables containing the metadata or tables with calculations based upon the metadata with **nuke-table**. Correct the errors in the csv file, rebuild appropriate receiver/event kef and reload the corrected kef into the PH5 file. Repopulated any other affected tables.

For example, if a receiver location is in error after loading the array_1.kef into the PH5 file:

A. Delete affected receiver tables and any related tables: **nuke-table** –n master.ph5 –A 1 followed by **nuke-table** –n master.ph5 –O (removes the source-receiver offset information). If all receivers/arrays are contained in one csv, you must remove all receiver tables
B. Correct the receiver location in the csv file; rebuild the array_1.kef with **noven** (steps 4-5).
C. Load array kefs: (step 7).
D. Re-calculate the offsets by building a new offset kef and then load it (step 9).

Last revised: January 12, 2017

PH5 Data Processing In a Nutshell

PH5 is the recommended archiving format for controlled-source experiments. This abbreviated guide will walk you through the process. Suggested command line input is highlighted yellow. For command usage use the "-h" option. Unix PH5-related commands are bold.

If you are using pforma to load data, follow the instructions in the appropriate document and then skip steps 1 and 3 below. Start from step 2 and then continue to step 4 through 13.

1. Create a PH5 directory and go there. Initialize the PH5 file. Name it "master":

initialize-ph5 –n master

2. Kitchen exchange format files, or kefs, are used to import metadata into the PH5 file. Create and go to a directory called metadata. Here we build the experiment summary kef with the GUI:

experiment_t-gen

You will need information such as the project name, PIs, and general location to complete this step. Afterwards, return to the PH5 directory and load the resulting kef (in this example it was saved as experiment_t.kef) into the PH5 file:

kef2ph5 –n master.ph5 –k metadata/experiment_t.kef

*note: Every time you load a kef using **kef2ph5**, an entry is made in a kef2ph5.log file. This can be useful for checking the progress of your steps!

3. Add Texan raw data to the PH5 file with **125a2ph5**. The "trd_list" should list the absolute or relative path and file names as shown to the right. Use the "M" option, specifying 1/6th the data volume to be loaded, in integer GB (you can use **du -sh** ../RAW to estimate data volume). The "M" option causes the data to be uploaded into a number of "data-only" ph5 files (called 'mini' files) linked to the master ph5 file, optimizing subsequent data processing and extraction.

```
# Example trd_list file
#
../RAW/I1303RAW309.TRD
../RAW/I1308RAW309.TRD
../RAW/I1313RAW309.TRD
```

Here we load ~100 GB of data: **125a2ph5** –f trd_list –n master.ph5 –M 17 >& 125a2ph5.out

Add RT130 data with **1302ph5**, setting the "M" option to 1/2th the RT130 data volume to be loaded, in integer GB. A list of RT130 raw files may be referenced as input to **1302ph5** similar to the **125a2ph5** example above.

4. Return to the metadata directory to prepare the receiver (station) metadata. See the **Metadata Tips** section on the back page for details. Ensure the comma-delimited csv/txt files, documenting the receiver metadata, contain the necessary information and the values provided follow the required format. The example single-array csv file below shows some lines from a deployment utilizing a 3-component Texan configuration at each site.

The minimum fields required for the csv file are: array, station, component, DAS serial number, deploy and pickup times, latitude, and longitude.

```
Array,Station,Component,Texan S/N,Deploy Time,Pickup Time,Lat.,Long.,Elev.,
1, 1001,1,12422,2014:240:18:10:00,2014:243:16:39:00,38.33821,-122.34804,15.0,
1, 1001,2,10076,2014:240:18:10:00,2014:243:16:39:00,38.33821,-122.34804,15.0,
1, 1001,3,13688,2014:240:18:10:00,2014:243:16:39:00,38.33821,-122.34804,15.0,
1, 1002,1,10248,2014:240:20:15:00,2014:243:16:50:00,38.33794,-122.34917,14.0,
```

Follow the steps below to build the receiver (and shot if applicable) kef(s).

a. Open **noven**; under the File menu, select and open the receiver or shot csv file.
b. Under the File/Configure menu, set as appropriate: "Input Type", "Column Separator" and "View Lines" (lines of the csv file visible in the GUI; set to 1 to view the csv headers).
c. In the main GUI window, assign the appropriate field name from each column drop down menu (check the Help menu for meanings of the different field names); repeat as necessary to ensure all required columns are defined with a field name. Under the File/Configuration menu set "Skip Lines" to the number of header line(s) in the file.
d. Select "Check input" under the File/Configuration menu; act on errors reported in pop-up.
e. Save the output, placing the suffix ".kef" at the end of the file name (e.g. array_1.kef).

5. Build a Google Earth kml file to view the experiment's geometry via the File->Map locations menu in **noven**.
This file only covers the currently loaded receiver or shot geometry. Open the resulting kml file with Google Earth to review your receiver geometry. If necessary, correct any errors in the csv file and rebuild the kef before proceeding.

6. Prepare and review the event metadata, if applicable to your data set, as done for the receiver metadata in Steps 4-5. The minimum metadata fields required in the event csv file are: shot line, shot/event ID, latitude, longitude, and shot time. Event size and depth are recommended but not required. Below are several lines serving as an example of an event csv file.

```
shotLine,Shot-ID,latitude,longitude,Elevation(m),Time[year:day:hr(Z):mm:ss.s],size,size units,depth(m)
1,5001,32.68681867,-84.98627767,177.49,2014:079:03:05:00.02,1000.0,lbs,100.0
1,5002,32.30797034,-84.59777903,195.2,2014:079:03:10:00.02,2000.0,lbs,100.0
1,5003,32.16043585,-84.46409245,174.68,2014:078:03,00:00.02,2000.0,lbs,100.0
```

7. Return to the PH5 directory and load all of your receiver and event (if applicable) kef(s):
kef2ph5 –n master.ph5 –k metadata/array_1.kef

8. **kmz-builder is currently not working. Use the .kml files created by noven to view geometry.**
Build a Google Earth kmz file to view the complete experiment's geometry:
kmz-builder –n master.ph5
Open the resulting kmz file with Google Earth to review your experiment map. If necessary, correct any errors in the csv file and rebuild the appropriate kef. Nuke (see back page) the already-loaded old metadata and then load the new kef.

9. The contents of your PH5 file may be viewed with *hdfview.sh*; browse the GUI to find master.ph5. **Open in read-only mode** as careless editing may corrupt the file.

10. Calculate the source-to-receiver offset for each station and shot and load the resulting kef:
geod2kef –n master.ph5 > offset_t.kef then run **kef2ph5** –n master.ph5 –k offset_t.kef

11. Texan data may need timing-drift corrections; calculate as shown below. **Skip this step if you only have RT130 data**. (RT130s are GPS-timed during data acquisition.)
time-kef-gen –n master.ph5 > time_t.kef then run **kef2ph5** –n master.ph5 –k time_t.kef

12. Run **sort-kef-gen**, which produces a kef containing information that optimizes data searches, and load the resulting kef:
sort-kef-gen –n master.ph5 -a > sort_t.kef then run **kef2ph5** –n master.ph5 –k sort_t.kef
*If you add new raw data, nuke the time_t.kef and sort_t.kef tables, and rerun steps 12 and 13 to include the new data.

13. Build and view SEG-Y gathers to test your metadata. Here is a gather-building example:
ph5toevt -n master.ph5 --use_deploy_pickup -o Gathers -N -l 10 –A 1 -x U -e 5012 --shot_line 1

The SEG-Y gather in this example is written to the Gathers directory and contains 10 second long un-time-corrected traces of data from array 1 receivers (all components) starting at the time of event 5012 and using a typical extended header format. A log of **ph5toevt** activity is also saved in the specified output directory. If you are ready to produce all of the gathers for the PI, you can save time by running **ph5tosevt** with the '-E' flag to create gathers for all of the events at once. '-A' can create all arrays at once. Run
ph5toevt –h
for many more gather-creating options.

Use **segyVista**, an easy-to-use gather viewer, to view the SEG-Y file.

More notes on PH5ViewerwVispy.py sometime in the future...

This step should only be followed in situations where a sensor's horizontal components are not oriented with N as True North and E as 90° East of True North, or where a sensor's vertical component provides positive voltage output with upward ground motion. **Skip this step for standard installations of PASSCAL GS-11, L-4C, L-22 and L-28 sensors.**

For the aforementioned situations:

a. Dump the receiver table to a kef: **tabletokef** –n master.ph5 –C > receiver_t.kef

b. Delete all lines from the kef corresponding to rows in the table not requiring a change to the azimuth or dip, otherwise those rows will be duplicated in the table upon upload of the modified kef.

c. Add the string ':Update:description_s' to the row description of the remaining rows (Remember the rows in need of revision should be the only rows left in the kef.)

   In the example below we: removed the lines corresponding to the N and E rows of the table, added ":Update:orientation/description_s" and changed the dip value from "90" to "-90".

```
#  Table row 1
        /Experiment_g/Receivers_g/Receiver_t:Update:orientation/description_s
        orientation/azimuth/value_f = 0
        orientation/azimuth/units_s = degrees
        orientation/dip/value_f = -90
        orientation/dip/units_s = degrees
        orientation/description_s = Z
```

d. Load the receiver_t.kef into the PH5 file with **kef2ph5**. Note the "-c" option is used as a check of kef syntax (necessary for hand-edited kefs).

   kef2ph5 –n master.ph5 –k receiver_t.kef –c >& receiver_kef2ph5.out

e. Inspect the output file for errors. Try: **egrep** –i "err|warn|prob|fail|fatal" receiver_kef2ph5.out to search for these strings in the output file. Correct any issues discovered and re-check.

f. Repeat the **kef2ph5** command without the "-c" option for the changes to be written to the PH5 file.