Managing Onshore-Offshore Data with PH5

Steve Azevedo, Derick Hess, & Bruce Beaudoin
IRIS PASSCAL Instrument Center, New Mexico Tech
Socorro New Mexico, USA (bruce@passcal.nmt.edu)

Seismic experiments recording offshore sources have onshore seismographs that have been used to illuminate continental margins. Test experiments supported by IRIS/PASSCAL have archived these data as receiver gather data, cutting the aquisition and recording the data at a remote site. With the successful application of PASSCAL-developed PH5’s HDF5 format (PH5), to land-based controlled source seismic experiment, PH5’s next logical extension of PH5 was accommodate the onshore-offshore seismic experiments. The 2014 onshore-offshore portion of the Eastern North American Margin Community Seismic Experiment (ENAM 380) provided an excellent opportunity to broaden PH5 utility for the seismic community. The 74,587 airgun shots were recorded on 3 arrays comprised of Rottik BT110 data loggers and 4.5Hz geophones with 3-Channels at 250sps for 45 days. Archiving these data in PH5 offers several advantages including access to all of the data collected (i.e., not just the airgun shots) and flexible data requests.

HDF5 was chosen for the following desirable characteristics: 1) Freely available and open source; 2) Well supported by the HDF Group (part of NCSA); 3) Support for numerous data models including time series data; 4) Self-describing hierarchically data format; 5) Large number of data tools already available; 6) Elements can be accessed generally more quickly than if they were in an SQL database; 7) Large cal data format; 8) Easy to edit elements in-place; and 9) Ability to accommodate extremely large data sets. The PIC KITCHEN is used to organize all of the experiment’s data and metadata into PH5. The extensibility and portability of HDF5 allows the PH5 format to evolve and operate on a variety of platforms and interfaces. Storing data in PH5 also facilitates the interactive pro-
duction of SEG-Y gathers, as well as other common seismic data formats based on varying and flexible data-request parameters.

Some Advantages of PH5
- Waveforms are stored separately from metadata
- Data is compressed
- Customized data requests
- Data is in a standard format that is interoperable
- Data visualization tool to help quickly perform field QC
- Dataless output (StationXML/PZ etc.)
- Field metadata QC improvements
- Filter to allow miniSEED input
- Programs API to allow quick development of tools that use PH5
- Changes to PH5 file organization

During the process of working with the IRNMA PIs to archive the onshore-offshore data in PH5, PASSCAL became familiar with some of the nuances associated with marine sources. To better support marine seismic data with PH5 will require a close collaboration between IRIS/PASSCAL and the marine community. If PH5 is recognized as a useful tool for marine data, we propose the following broad goals and outline the development identified to support these goals.

Goals:
- Develop strong ties with stakeholders; efficient data collection and field data handling; quick access to data for PIs in requested format with desired header mappings; ability to efficiently and quickly move data to DMC archive; useful and easy access to data at the DMC in desired format and organization; ability to efficiently ingest metadata and trace data in miniSEED or SEG-D formats; quick and efficient turn around of data sets.

Development (new and ongoing)
- Changes to PH5 file organization
  - Maps groups: – Map PH5 to various data formats on a per-experiment basis, e.g., SEG-Y trace headers
  - Events: – Map PH5 to various data formats on a per-experiment basis, e.g., SEG-Y or SEG-D formats;
- Quick and efficient turn around of data sets.
- Support for non-seismic channels such as pressure
- Programmers API to allow quick development of tools that use PH5
- Filter to allow miniSEED input
- Calculate output (StationXML, PZ etc.)
- Data visualization tool to help quickly perform field QC
- Field heading of raw data in parallel to process field data handling
- Field metadata QC improvements
- Coordinate with the IRIS DMC to create data request tools to better serve the marine controlled source community