



PH5 for integrating and archiving different data types

Steve Azevedo, Derick Hess, and Bruce Beaudoin

IRIS PASSCAL, New Mexico Tech, Socorro, NM, United States (bruce@passcal.nmt.edu)

PH5 is IRIS PASSCAL's file organization of HDF5 used for seismic data. The extensibility and portability of HDF5 allows the PH5 format to evolve and operate on a variety of platforms and interfaces. To make PH5 even more flexible, the seismic metadata is separated from the time series data in order to achieve gains in performance as well as ease of use and to simplify user interaction. This separation affords easy updates to metadata after the data are archived without having to access waveform data. To date, PH5 is currently used for integrating and archiving active source, passive source, and onshore-offshore seismic data sets with the IRIS Data Management Center (DMC). Active development to make PH5 fully compatible with FDSN web services and deliver StationXML is near completion. We are also exploring the feasibility of utilizing QuakeML for active seismic source representation.

The PH5 software suite, PIC KITCHEN, comprises in-field tools that include data ingestion (e.g. RefTek format, SEG-Y, and SEG-D), meta-data management tools including QC, and a waveform review tool. These tools enable building archive ready data in-field during active source experiments greatly decreasing the time to produce research ready data sets. Once archived, our online request page generates a unique web form and pre-populates much of it based on the metadata provided to it from the PH5 file. The data requester then can intuitively select the extraction parameters as well as data subsets they wish to receive (current output formats include SEG-Y, SAC, mseed). The web interface then passes this on to the PH5 processing tools to generate the requested seismic data, and e-mail the requester a link to the data set automatically as soon as the data are ready.

PH5 file organization was originally designed to hold seismic time series data and meta-data from controlled source experiments using RefTek data loggers. The flexibility of HDF5 has enabled us to extend the use of PH5 in several areas one of which is using PH5 to handle very large data sets. PH5 is also good at integrating data from various types of seismic experiments such as OBS, onshore-offshore, controlled source, and passive recording. HDF5 is capable of holding practically any type of digital data so integrating GPS data with seismic data is possible. Since PH5 is a common format and data contained in HDF5 is accessible randomly it has been easy to extend to include new input and output data formats as community needs arise.