A New Approach to Archiving and Translating Seismic Data using HDF5

Steve Azevedo, Derick Hess, and Bruce Beaudoin

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

The PASSCAL Instrument Center at New Mexico Tech has developed an alternative to SEG-Y for archiving and translating controlled source seismic data. This alternate method (PH5) is based on the Hierarchical Data Format (HDF5) and has the advantage that the waveform data is decoupled from the meta-data. The primary motivator for developing this new format was to mitigate the need to handle waveforms multiple times in the archiving process if meta-data corrections are discovered. All PASSCAL data sets are archived at the IRIS Data Management Center (DMC). When SEG-Y data are in the DMC archive, meta-data changes require that the entire data volume be resubmitted. In contrast, if the data are archived in PH5, the separation of the waveform data from the meta-data afford an efficient method to update meta-data without re-handling waveforms.

The extensibility and portability of HDF5 allows the PH5 to evolve and operate on a variety of platforms and interfaces. Early implementations of PH5 saw use in the field, in the lab, and at data centers. However, usage was largely limited to command-line tools, multistep procedures and static data-request webpages. Recent modifications to the PH5 architecture have allowed for a simplified user interface and have achieved significant performance gains. Additional efforts focused on increasing the ease and efficiency of the meta-data collection and processing by field teams. Meta-data extraction tools were simplified by creating a unified, single interface that manages a suite of processes heretofore run individually and on the command-line. Storing data in PH5 facilitates the interactive production of SEG-Y gathers, as well as other common seismic data formats, based on varying and flexible data request parameters. A recently developed web interface dynamically generates a unique web form and pre-populates much of it based on the meta-data provided from the PH5 file. The data requester can then intuitively select the extraction parameters as well as data subsets. The web interface then passes the user's selection on to the PH5 processing tools to generate the requested SEG-Y data.