Contribution ID: **15** Type: **”Standard talk”**

# Easy Columnar File Conversion with ’hepconvert’

*Wednesday, 3 July 2024 16:30 (30 minutes)*

Though columnar file formats are popular among HEP users, the process to convert between file formats has multiple steps, and generally requires the use of one I/O package per file format. Often users need to customize the process as well, either due to memory constraints or to modify the data before writing it to a new file. This entails both more lines of code and experience with I/O packages, and in some cases knowledge about each data format.
To streamline this process and save user’s time, we are developing the Python package ‘hepconvert.’This package aims to simplify columnar file conversions and common customizations down to single function between file formats Parquet, ROOT, and HDF5 files. It uses pre-existing functions from reputable columnar I/O packages such as Uproot, Awkward Array, and h5py, with additional builtin features for common customizations. The customizations are added at user request and include automatic reading and writing in batches, compression setting, branch skimming and slimming, histogram summing, and more. In addition to making the features in hepconvert, we are also adding relevant functionality to Uproot that will eventually be included in hepconvert; adding new TBranches to existing TTrees.

**Primary authors:** PIVARSKI, Jim (Princeton University); BILODEAU, Zoë (Princeton University (US))

**Presenter:** BILODEAU, Zoë (Princeton University (US))

**Session Classification:** Plenary Session Wednesday