



## ASDF - A Modern Data Format for Seismology

Lion Krischer (1), James Smith (2), Wenjie Lei (2), Matthieu Lefebvre (2), Youyi Ruan (2), Elliot Sales de Andrade (3), Norbert Podhorszki (4), Ebru Bozdogan (5), and Jeroen Tromp (2)

(1) ETH Zürich, Geophysik, Switzerland, (2) Princeton University, Department of Geosciences, NJ, USA, (3) Department of Physics, University of Toronto, Canada, (4) Oak Ridge National Laboratory, Oak Ridge, Tennessee, United States, (5) Université de Nice Sophia Antipolis, Géoazur, Valbonne, France

Seismology as a science is driven by observing and understanding data and it is thus vital to make this as easy and accessible as possible. The growing volume of freely available data coupled with ever expanding computational power enables scientists to take on new and bigger problems. This evolution is to some part hindered as existing data formats have not been designed with it in mind.

We present ASDF (<http://seismic-data.org>), the Adaptable Seismic Data Format, a novel, modern, and especially practical data format for all branches of seismology with particular focus on how it is incorporated into seismic full waveform inversion workflows. The format aims to solve five key issues:

1. **Efficiency:** Fast I/O operations especially in high performance computing environments, especially limiting the total number of files.
2. **Data organization:** Different types of data are needed for a variety of tasks. This results in ad hoc data organization and formats that are hard to maintain, integrate, reproduce, and exchange.
3. **Data exchange:** We want to exchange complex and complete data sets.
4. **Reproducibility:** Oftentimes just not existing but crucial to advance our science.
5. **Mining, visualization, and understanding of data:** As data volumes grow, more complex, new techniques to query and visualize large datasets are needed.

ASDF tackles these by defining a structure on top of *HDF5* reusing as many existing standards (*QuakeML*, *StationXML*, *PROV*) as possible. An essential trait of ASDF is that it empowers the construction of completely self-describing data sets including waveform, station, and event data together with non-waveform data and a provenance description of everything. This for example for the first time enables the proper archival and exchange of processed or synthetic waveforms.

To aid community adoption we developed mature tools in Python as well as in C and Fortran. Additionally we provide a formal definition of the format, a validation tool, and integration into widely used tools like *ObsPy* (<http://obspy.org>), *SPECFEM GLOBE* ([https://geodynamics.org/cig/software/specfem3d\\_globe/](https://geodynamics.org/cig/software/specfem3d_globe/)), and *Salvus* (<http://salvus.io>).