

ObsPlus: A Pandas-centric ObsPy expansion pack

Derrick J.A. Chambers¹, M. Shawn Boltz¹, and Calum J. Chamberlain²

¹ National Institute for Occupational Safety and Health, Spokane Mining Research Division ² School of Geography, Environment and Earth Sciences, Victoria University of Wellington, New Zealand

DOI: [10.21105/joss.02236](https://doi.org/10.21105/joss.02236)

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Editor: [Pending Editor](#) ↗

Submitted: 21 May 2020

Published: 23 May 2020

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC-BY](#)).

Summary

Over the past decade, ObsPy, a python framework for seismology (Krischer & Wassermann, 2015), has become an integral part of many seismology research workflows. ObsPy provides parsers for most seismological data formats, clients for accessing data-centers, common signal processing routines, and event, station, and waveform data models.

ObsPlus significantly expands ObsPy's functionality by providing simple data management abstractions and conversions between ObsPy classes and the ubiquitous pandas DataFrame.

Functionality and Features

1. **A data retrieval interface** ObsPlus provides a unified data retrieval interface for in-memory, on disk, and remote seismological data. This is enabled by in-process databases which provide a simple mechanism to index and access local seismological data stored in directories of arbitrary organization. Importantly, the classes implement a superset of the interface already provided by ObsPy's remote clients, making it straight-forward to write data-source agnostic code.
2. **Alternative data structures** While ObsPy's data structures are quite powerful, they are not always the most convenient. For example, the canonical event data representation is a deeply nested tree-like structure based on the QuakeML standard (Schorlemmer & Saul, 2011). Working with events in ObsPy often necessitates deeply nested recursive code which can become difficult to understand and maintain. ObsPlus provides functionality to flatten desired components of these tree structures into DataFrames which are simpler and more efficient when the full complexity of QuakeML isn't merited for the task at hand.
3. **Datasets** ObsPlus provides a simple mechanism to bundle, distribute, download, and interact with complete seismological datasets. This is done by creating a simple python package (for which we provide a cookie cutter template) which is published to PyPI. Each package includes small files and instructions to download large files. Optionally, a list of files and their corresponding hashes can be used to validate downloaded data. Datasets are then discovered and loaded through python's plugin system, and downloaded when needed.
4. **Utilities** ObsPlus's list of utilities is quite long and more are being added regularly. Many are focused around validating and manipulating event data.

ObsPlus has become an important part of the National Institute for Occupational Safety and Health (NIOSH)'s data processing and management workflows and has enabled rapid prototyping of new ideas while managing complexity through appropriate abstractions. It is our hope that ObsPlus will provide similar benefits to the broader seismology community.

References

- Krischer, M., L., & Wassermann, J. (2015). ObsPy: A bridge for seismology into the scientific Python ecosystem. *Computational Science & Discovery*.
- Schorlemmer, E., D., & Saul, J. (2011). QuakeML: status of the XML-based seismological data exchange format. *Annals of Geophysics*.