

# BESSIG Meeting Wed, Nov 20, 4 - 6 PM

Our meeting this month is a special event for several reasons. Copies of Andrew's book will be available to the first 50 attendees, and the HDF Group will be providing refreshments for us. Also, this may be our last meeting at the [Boulder Outlook Hotel](#), as the hotel has been sold. So, please join us **in the Crown Rock room (not our usual room)** at the Outlook for:

## Improving Science with Open Formats and High-Level Languages: Python and HDF5

**Andrew Collette**, Laboratory for Atmospheric and Space Physics (LASP)

This talk explores how researchers can use the scalable, self-describing HDF5 data format together with the Python programming language to improve the analysis pipeline, easily archive and share large datasets, and improve confidence in scientific results. The discussion will focus on real-world applications of HDF5 in experimental physics at two multimillion-dollar research facilities: the Large Plasma Device at UCLA, and the NASA-funded hypervelocity dust accelerator at CU Boulder. This event coincides with the launch of a new O'Reilly book, *Python and HDF5: Unlocking Scientific Data*, complimentary copies of which will be available for attendees.

As scientific datasets grow from gigabytes to terabytes and beyond, the use of standard formats for data storage and communication becomes critical. HDF5, the most recent version of the Hierarchical Data Format originally developed at the National Center for Supercomputing Applications (NCSA), has rapidly emerged as the mechanism of choice for storing and sharing large datasets. At the same time, many researchers who routinely deal with large numerical datasets have been drawn to the Python by its ease of use and rapid development capabilities.

Over the past several years, Python has emerged as a credible alternative to scientific analysis environments like IDL or MATLAB. In addition to stable core packages for handling numerical arrays, analysis, and plotting, the Python ecosystem provides a huge selection of more specialized software, reducing the amount of work necessary to write scientific code while also increasing the quality of results. Python's excellent support for standard data formats allows scientists to interact seamlessly with colleagues using other platforms.

[Schedule \(more or less\)](#)

4:00 - 5:00 presentation

5:00 - 6:00 social