Fusion Data Framework

A data API for magnetic fusion experiments

J. Schmitt, D. Smith, K. Tritz and H. Yuh

10/9/2015

Key findings from April survey

- Many colleagues want more functionality in their data tools (e.g. scopes, logbook, efitviewer, reviewplus), but feel extending functionality is difficult
- Many colleagues feel IDL code is difficult to find, difficult to reuse, and that their own codes are not reusable
- Broad agreement of significant code duplication

Motivation

- Benefits of a data API
  - Reduce barriers to entry for new students/colleagues
  - A platform for collaborative software development and distribution
  - Flexible and extensible (in contrast to stand-alone tools)
  - Improve software development efficiency by reducing duplicative efforts and promoting user-driven solutions
- We feel a data framework/API can address many of the findings from the April survey

FDF is a data access, management, and visualization API for magnetic fusion experiments

- Data access without node names
- Object-oriented

  >>> nstx.s140001.mpts.te
- Shot/results database integration

  >>> nstx.s140001.logbook()

  >>> nstx.addshot(xp=1032)
- Data is cached
- Built-in methods for plotting, splines, ffts, despiking, flux mapping, archival, save/restore, etc

  >>> nstx.s140001.chers.vt.plot()

  >>> nstx.s140001.xray.hup.despike().plot()
- Built on the Anaconda distribution of Python
  - free, multi-platform
- Flexible usage
  - command-line interface, in your routines, or GUIs

Many data-intensive scientific fields develop and leverage APIs

- Root (high energy, 115 devs)
  - https://root.cern.ch/
- Astropy (astronomy, 142 devs)
  - https://github.com/astropy/astropy
- Sunpy (solar physics, 48 devs)
  - https://github.com/sunpy/sunpy
- EUTelescope (astronomy, 17 devs)
  - http://eutelescope.web.cern.ch/
- Thunder (neuroscience, 25 devs)
  - https://github.com/thunder-project/thunder
- Neo (neuroscience, 18 devs)
  - http://neo.readthedocs.org/en/latest/
- Schemas (genomics, 32 devs)
  - https://github.com/ga4gh/schemas
- Gemini (genomics, 23 devs)
  - https://gemini.readthedocs.org/en/latest/
- ObsPy (seismology, 39 devs)
  - https://github.com/obspy/obspy
- UV-CDAT (climate science, 29 devs)
  - https://github.com/UV-CDAT/uvcdat
- Galaxy (biology, 84 devs)
  - https://github.com/galaxyproject/galaxy
- QIIME (ecology, 48 devs)
  - https://github.com/biocore/qiime
- Scikit-bio (bioinformatics, 36 devs)
  - https://github.com/biocore/scikit-bio
- And many federal agencies

FDF code repository and documentation

- Code repository
  - https://github.com/Fusion-Data-Framework/fdf
  - Also, submit bug reports and feature requests
- Documentation
  - http://fusion-data-framework.github.io/fdf/
  - Includes user and developer guides