Project Jupyter: From interactive Python to Open Science

Fernando Pérez (@fperez_org)

UC Berkeley & LBNL
Why?

Open Source in Science?
Technical: better tools
2001: physics grad student
Python FTW!
Interactivity: Humans in the loop!

Thanks to a blog post by D. Beazley about sys.displayhook in Python 2
Others were thinking about the same…

Janko Hauser: Interactive Python Prompt (IPP)

Nathan Gray: Lazy Python/deep_reload
Dec 10, 2001, on comp.lang.python

From the get-go, a collaborative effort!

AUTHORS: Fernando Pérez (current), Janko Hauser, Nathan Gray (past)
Ethical:
access & collaboration
Medellín, Colombia

Image Credits:
latinorebels.com
wired.com
IPython sprint @ SciPy India 2010

Prabhu Ramachandran,
IIT Mumbai, Enthought
Scientific:
black boxes don’t belong in science

What I cannot create, I do not understand

Richard Feynman
Most important in the long run: Community!
SciPy: Scientists collaborating across disciplinary boundaries for the fun of building better tools, rather than competing for papers.
IPython: life-long partnerships

SciPy 2008, Caltech

PyCon 2012, Santa Clara
Team today: where *all the credit* goes

And over 500 contributors!
Brett Cannon said it best
What:
Human-centered, interactive, computing and science
At the terminal or the notebook
Jupyter: beyond Python

~90 different kernels: https://github.com/jupyter/jupyter/wiki/Jupyter-kernels
What do we build?

❖ Low-level standards:
  ❖ Messaging protocol, notebook format
  ❖ Reusable libraries that implement them
  ❖ User-facing applications
    ❖ IPython, Jupyter Notebook/Lab, JupyterHub, …
  ❖ Services that make them accessible
    ❖ nbviewer, try.jupyter

All in support of an open ecosystem
protocol and format

- Core foundations
- Capture lasting, important elements of human-centered computational processes
- But meant for machines to digest well!
- Archival, reusable, multi-purpose

100% open: work with us, don’t reinvent at this level!!
Where next
Exciting developments

JupyterLab

nbdime

binder

Turn a GitHub repo into a collection of interactive notebooks

Build a repository

... lots more!
Executable books
Berkeley’s Foundations of Data Science

- Interactive textbook is Jupyter Notebooks
- Course deployment is JupyterHub
- Based on original effort by Jess Hamrick (2017 PhD)
- Ryan Lovett, Yuvi Panda, Vinitra Swamy, Gunjan Baid, Chris Hench, Claudia von Vacano...

data.berkeley.edu, data8.org

inferentialthinking.com
Fastest growth in Berkeley history

We estimate ~1200 students this semester.

First class was yesterday!
How do we keep doing this?
A sustainable future for Jupyter & friends?

When [platforms] try to extract too much value for the platform owner, they ultimately fail… Platforms must create more value than they capture if their ecosystem is to survive.

Tim O’Reilly -  WTF? What’s the Future and Why It’s Up To Us
Industry platforms
NumFOCUS is a 501(c)(3) nonprofit that supports and promotes world-class, innovative, open source scientific computing.
Funding and partnerships
A personal note

- IPython was born out of deep, hard failure
- The SciPy community saved me!
- Impostor syndrome is still very real
- Scientist, doing software, crossing disciplinary boundaries...
I trust this community to continue to be as welcoming and open as it can be innovative.

Let’s keep building tools that contribute to an open, collaborative, accessible and fair society.
Thank You!