Hyper*Spy*

Francisco de la Peña

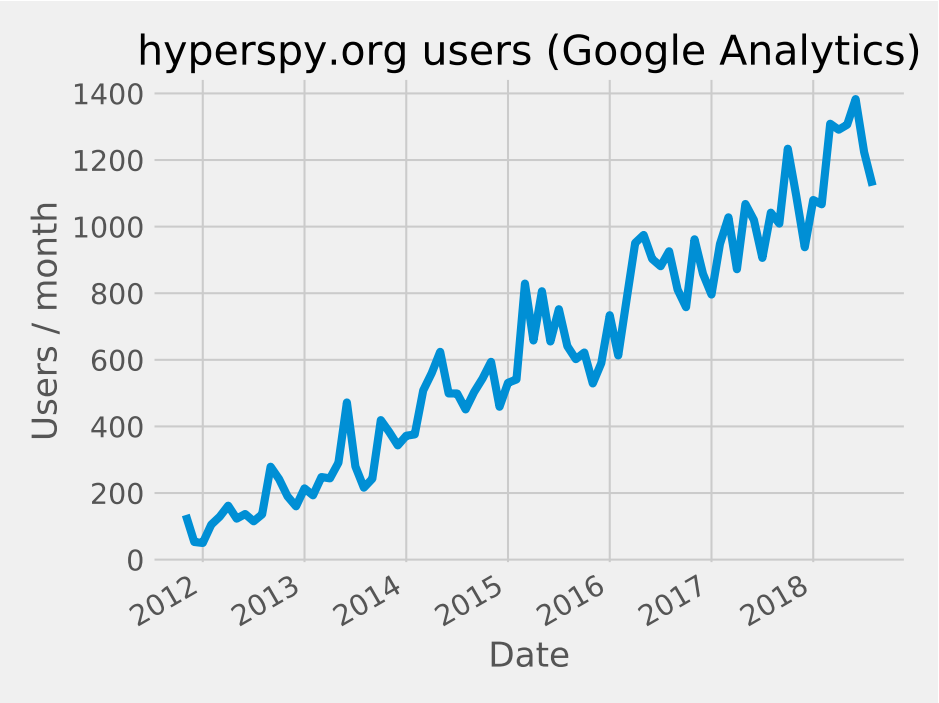
Diamond Light Source
2nd of March 2020

Outline

1. HyperSpy Today
2. How did we get here?

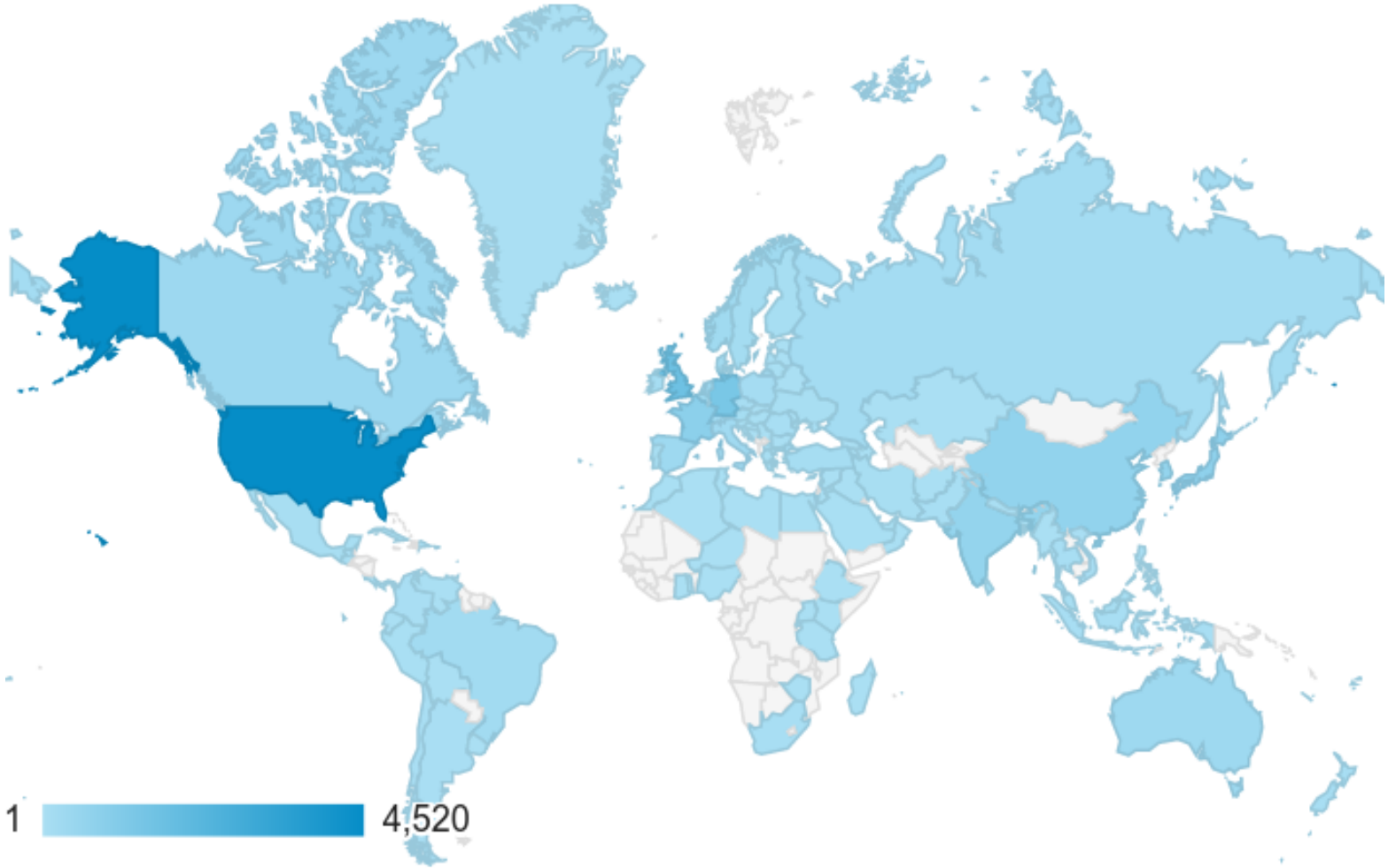
HyperSpy today–some stats

- $> 3000\,\mathrm{downloads/month}$
- $> 100k$ lines of code
- > 4000 citations
- 40 contributors
- Used by more than 40 packages in GitHub



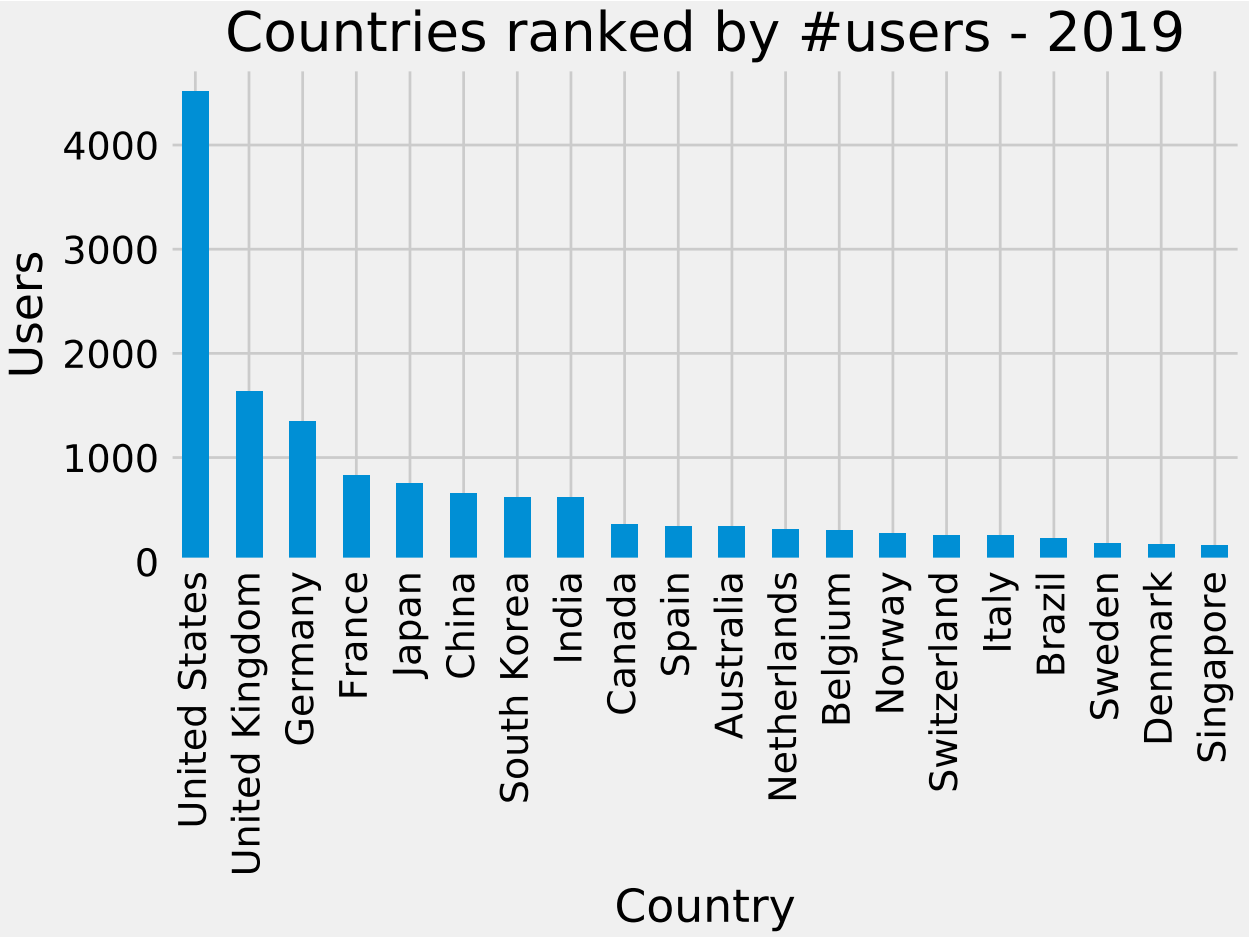
HyperSpy users by country

hyperspy.org users from January-September 2018



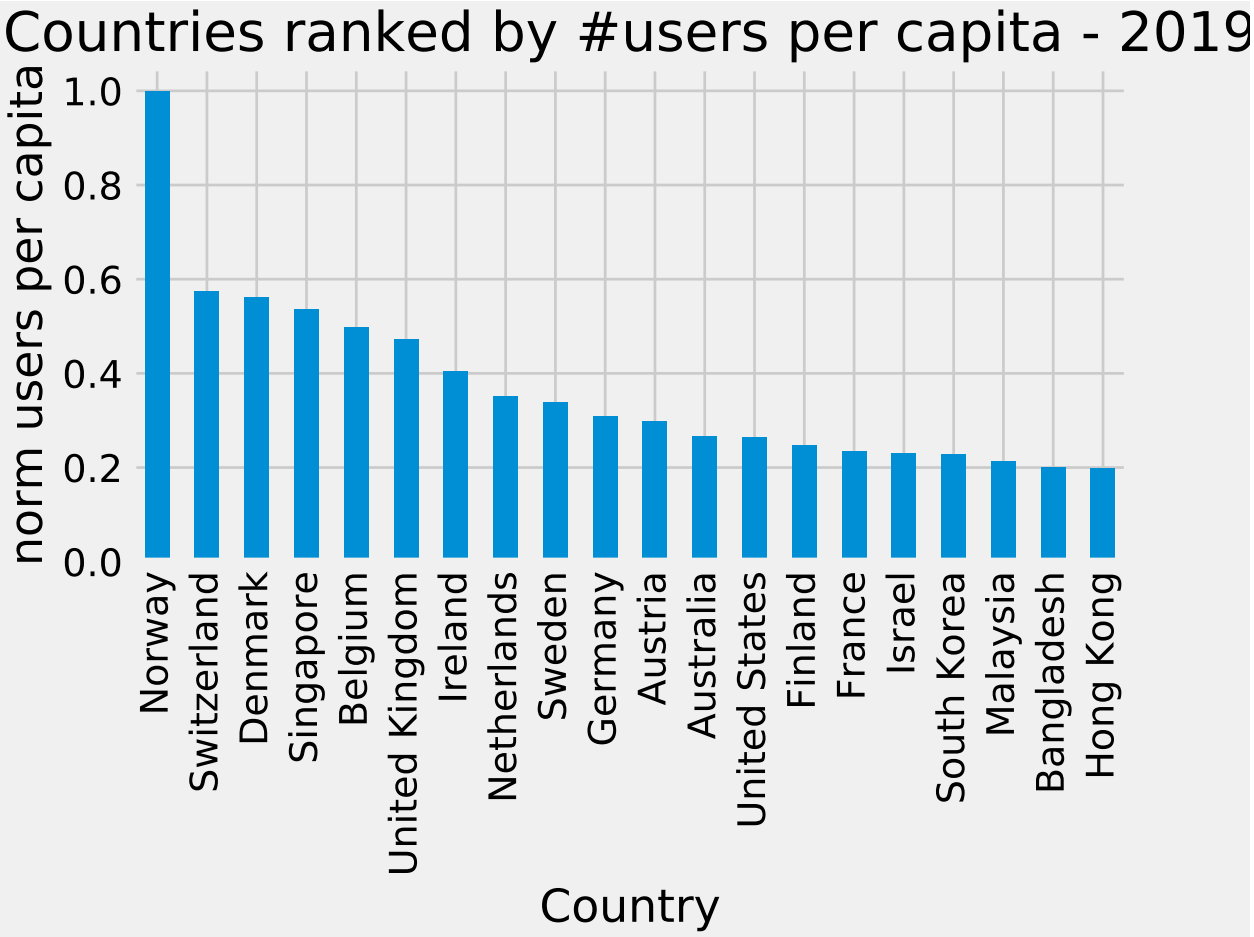
HyperSpy users by country

hyperspy.org users from January-September 2018



HyperSpy users by country

hyperspy.org users from January-September 2018



Outline

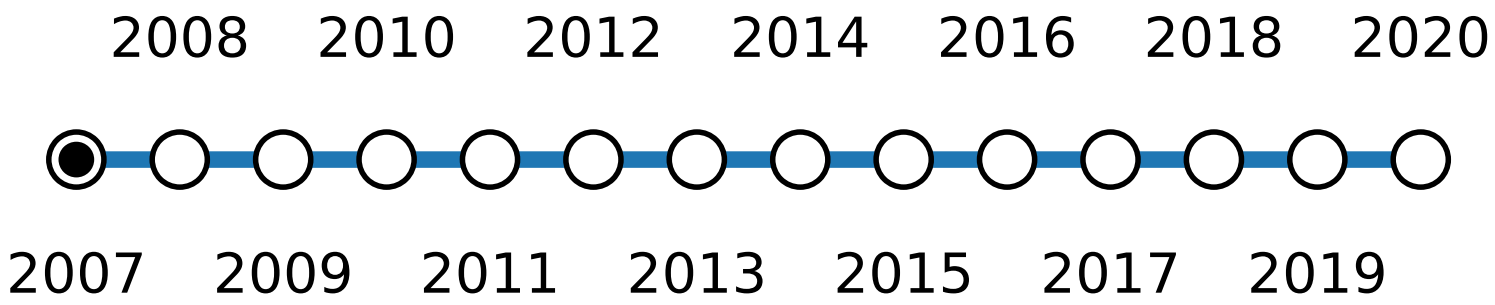
1 HyperSpy Today

2 How did we get here?

What is HyperSpy? Take 1.

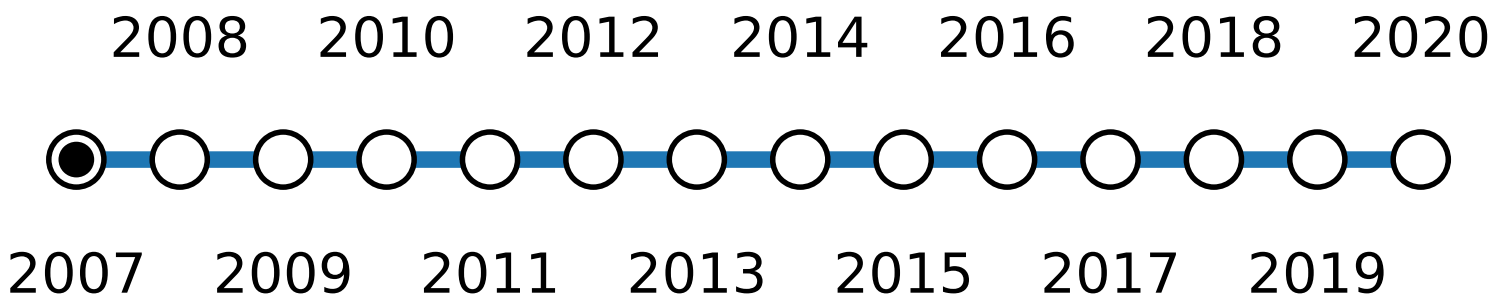
HyperSpy is a computer program for electron microscopy data analysis.

What is HyperSpy? (2007)



- During my PhD at the Paris-Sud University (LPS-Orsay), I started using Python for EELS data analysis.
- By placing the functions in classes a structure started to emerge.

What is HyperSpy? (2007)

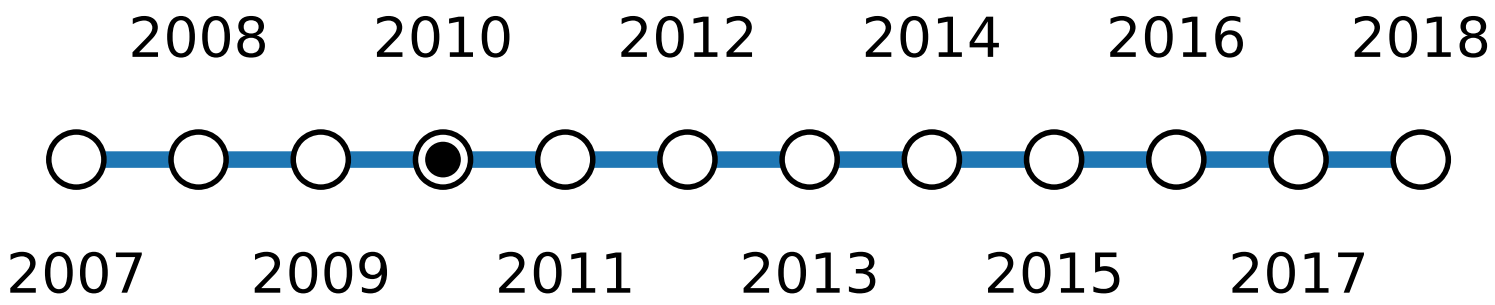


- During my PhD at the Paris-Sud University (LPS-Orsay), I started using Python for EELS data analysis.
- By placing the functions in classes a structure started to emerge.

HyperSpy is . . .

My personal set of Python routines for EELS data analysis of spectrum images.

What is HyperSpy? (2010)

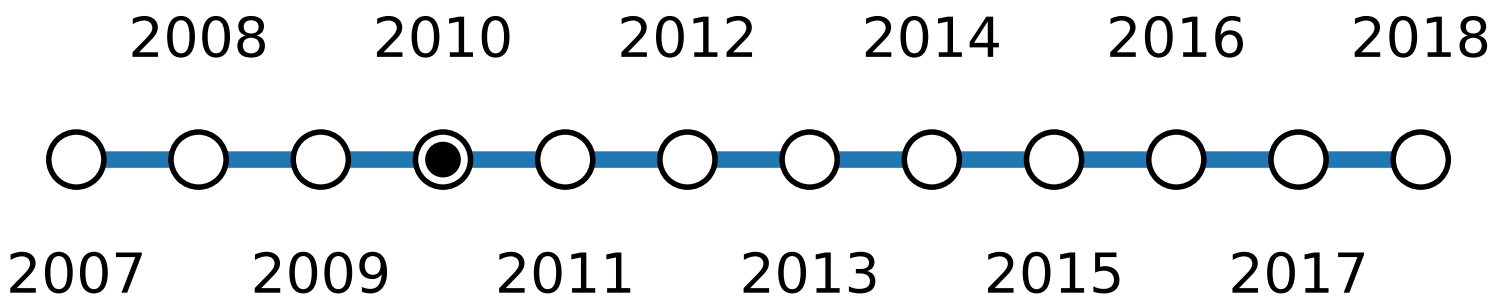


- Multivariate analysis routines implemented.
- EELSLab released under GPLv2 license.
- First users. First trainings in Paris and Oxford.

HyperSpy is . . .

My personal set of Python routines for EELS data analysis of spectrum images.

What is HyperSpy? (2010)

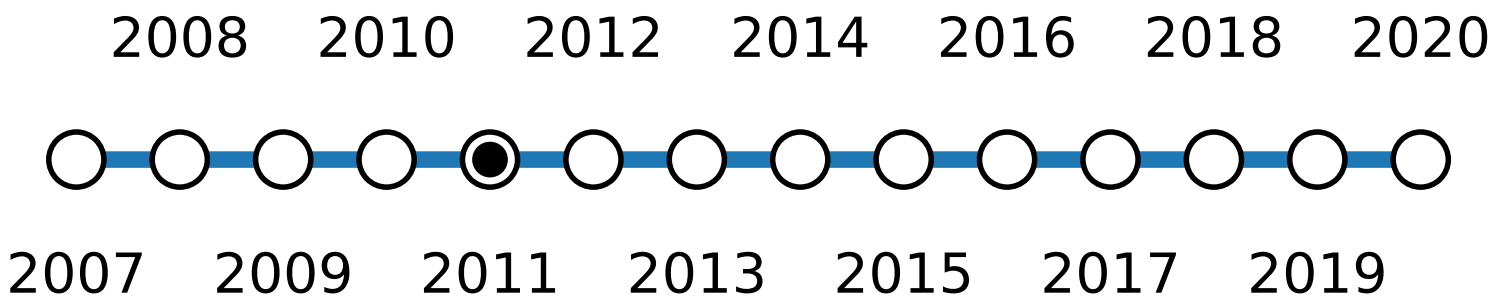


- Multivariate analysis routines implemented.
- EELSLab released under GPLv2 license.
- First users. First trainings in Paris and Oxford.

HyperSpy is . . .

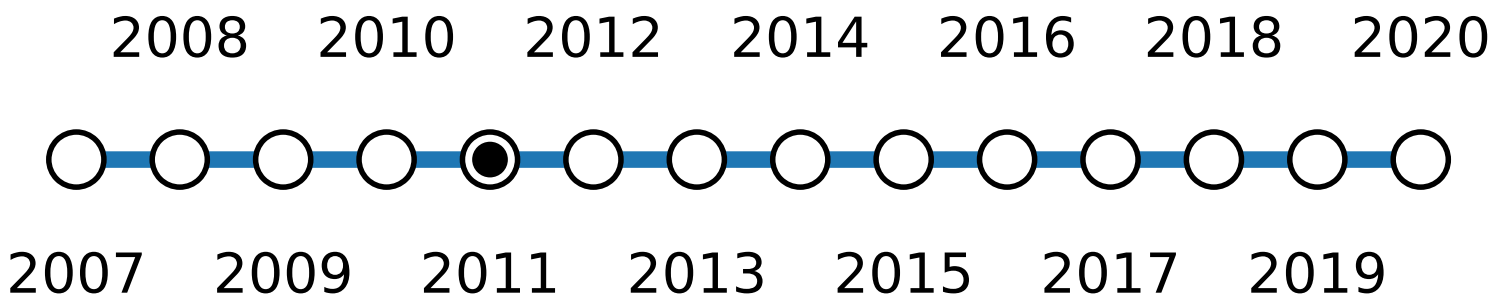
an *open-source* Python package for *data analysis of EELS multi-dimensional datasets*.

What is HyperSpy? (2011)



- Michael Sarahan (SuperSTEM) and Stefano Mazzucco (NIST) join the development team.
- We refactor the code to make it multi-dimensional.
- We rename it to “HyperSpy”.

What is HyperSpy? (2011)



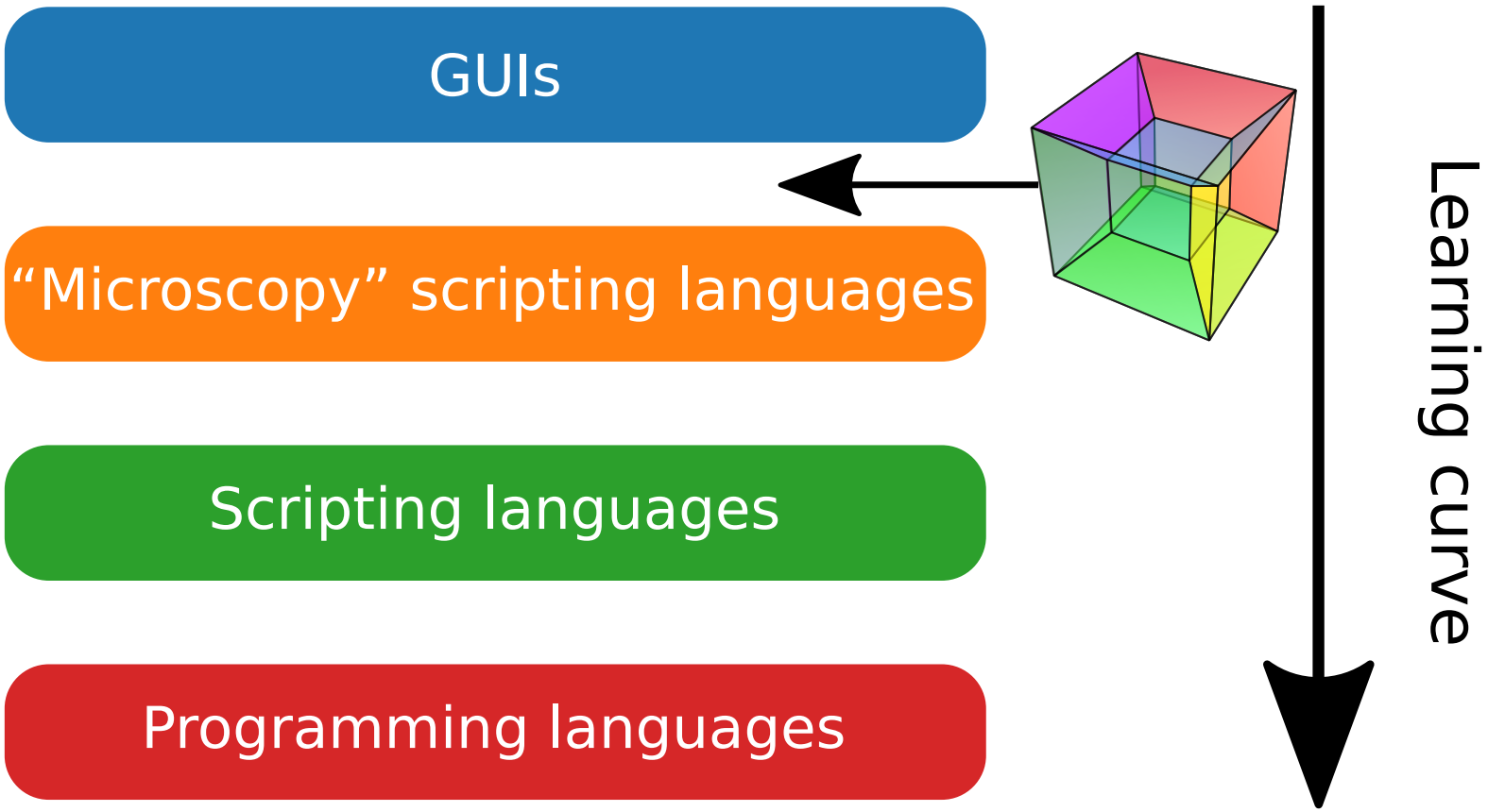
- Michael Sarahan (SuperSTEM) and Stefano Mazzucco (NIST) join the development team.
- We refactor the code to make it multi-dimensional.
- We rename it to "HyperSpy".

HyperSpy is . . .

an open-source *open-development* Python package for *data analysis of multi-dimensional datasets*.

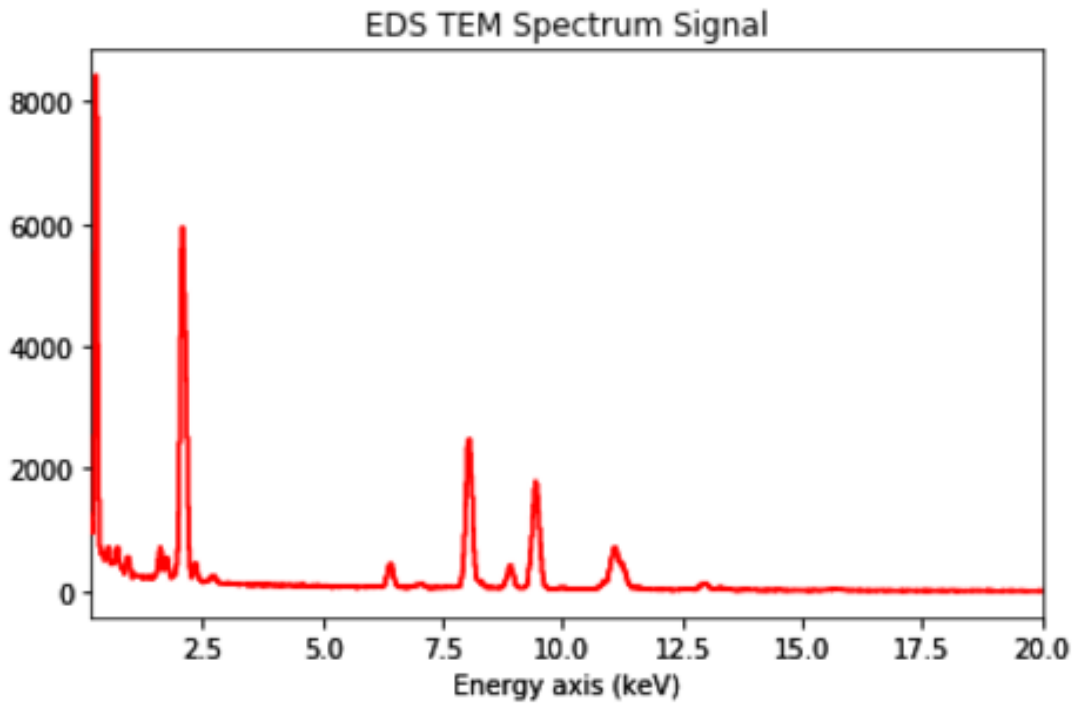
Why writing another software package?

Why writing another software package?



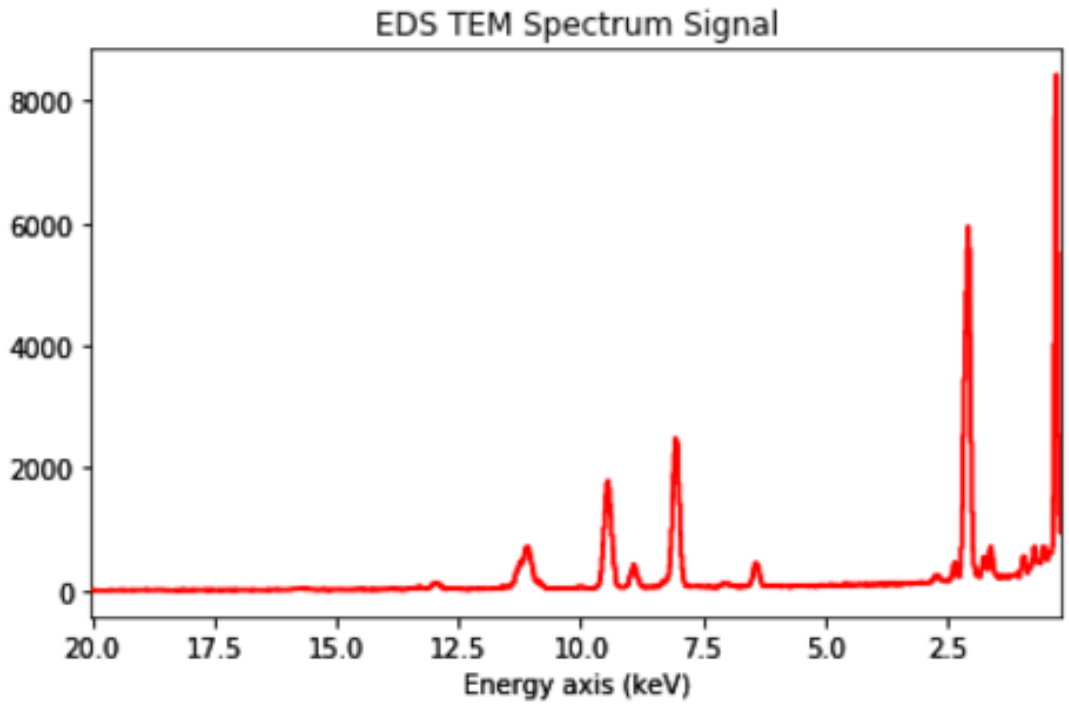
Example of HyperSpy vs Digital Micrograph syntax

```
In [4]: s.plot()
```



Example of HyperSpy vs Digital Micrograph syntax

```
In [5]: s.isig[::-1].plot()
```

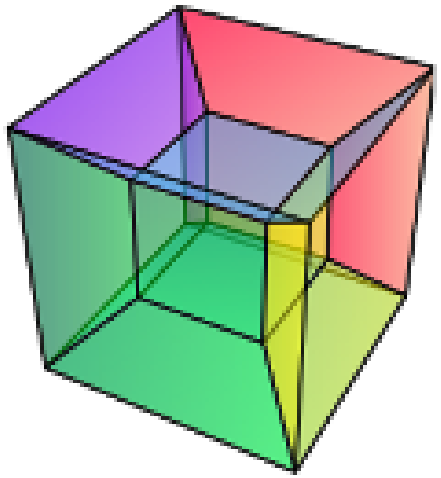


Example of HyperSpy vs Digital Micrograph syntax

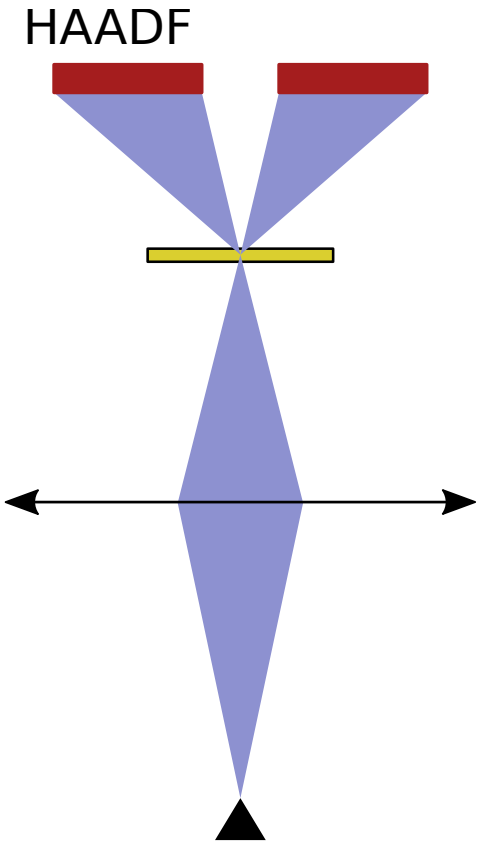
```
image img := GetFrontImage()
if (img.ImageGetNumDimensions() == 3)
{
        number sx = img.ImageGetDimensionSize(0)
        number sy = img.ImageGetDimensionSize(1)
        number sz = img.ImageGetDimensionSize(2)
        image res = Slice3(img, 0, 0, sz - 1, 0, sx, 1, 1, sy, 1, 2, sz, -1)
        res.ImageCopyCalibrationFrom(img)
        TagGroupCopyTagsFrom(ImageGetTagGroup(res),ImageGetTagGroup(img))
        res.ImageSetName(img.ImageGetName() + "_reversed")
        ShowImage(res)
}
```

HyperSpy design goals in 2011

- Cutting-edge features.
- Truly multi-dimensional.
- Scalable
  - Powerful, yet
    - easy to use.
    - easy to learn.
    - Easy to extend and contribute to.
    - Code fully available.

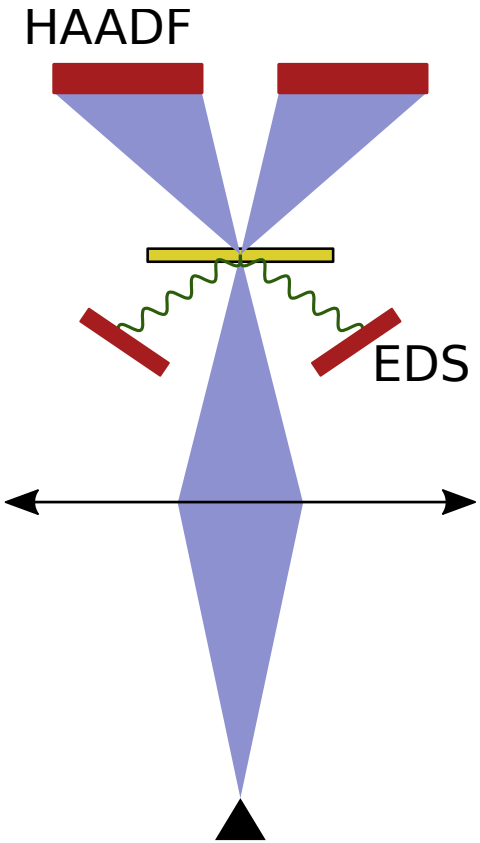


Why the emphasis on multi-dimensional datasets?



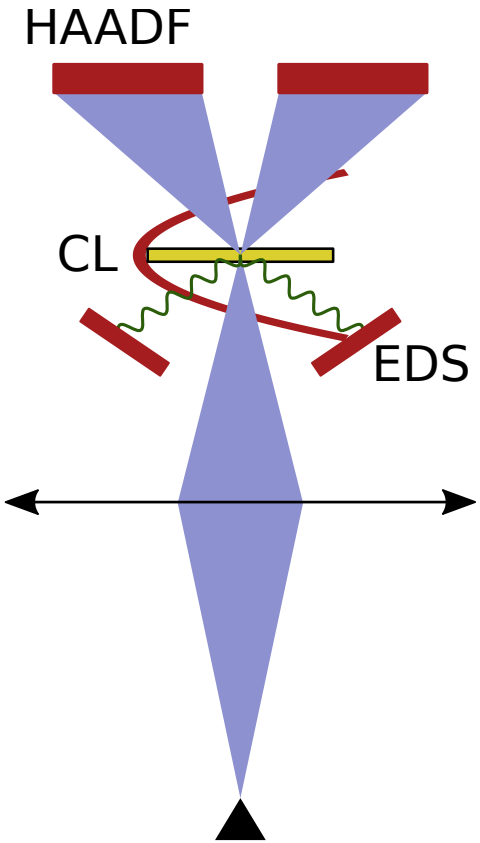
Detectors	Dimensions
HAADF	(x, y|)

Why the emphasis on multi-dimensional datasets?



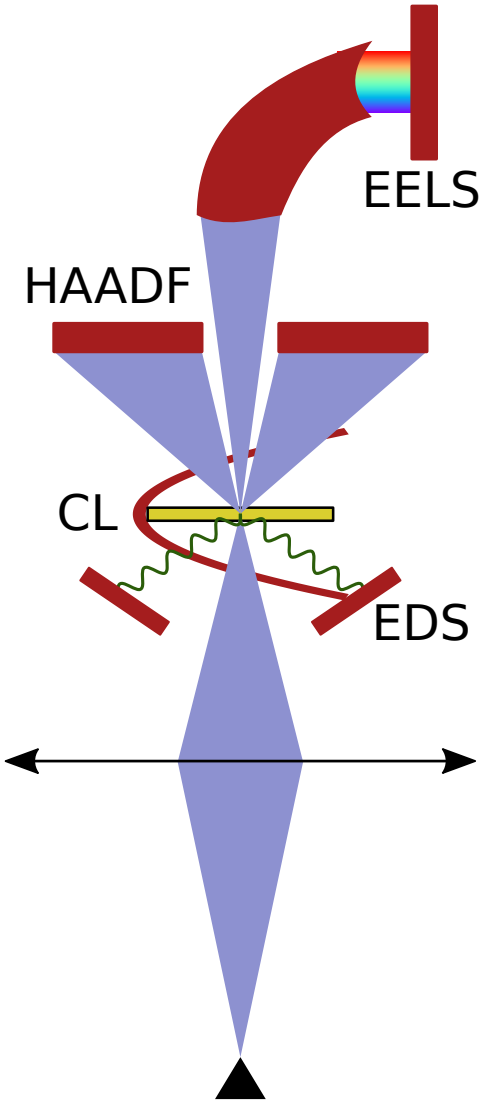
Detectors	Dimensions
HAADF	(x, y|)
EDX	(x, y | E)

Why the emphasis on multi-dimensional datasets?



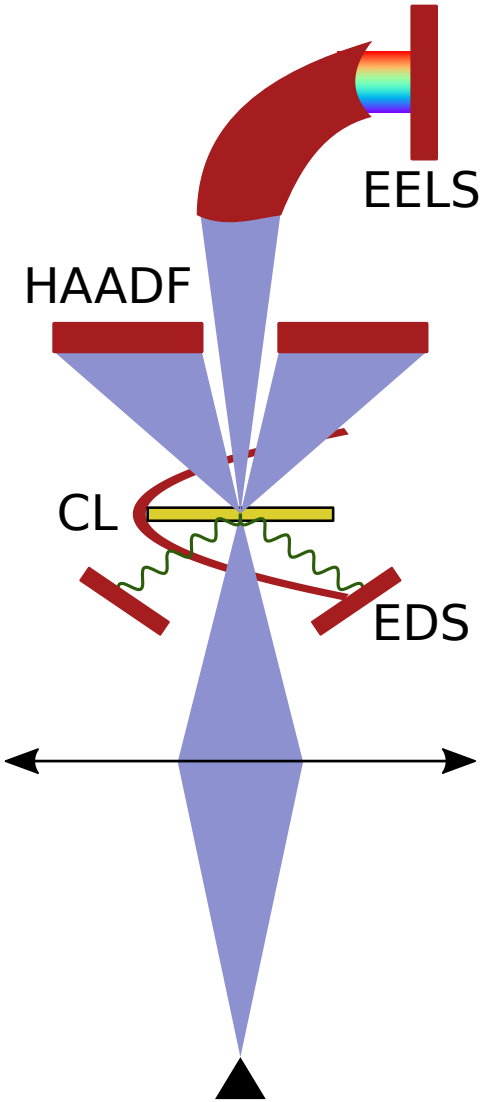
Detectors	Dimensions
HAADF	(x, y|)
EDX	(x, y | E)
CL	(x, y | ω)

Why the emphasis on multi-dimensional datasets?



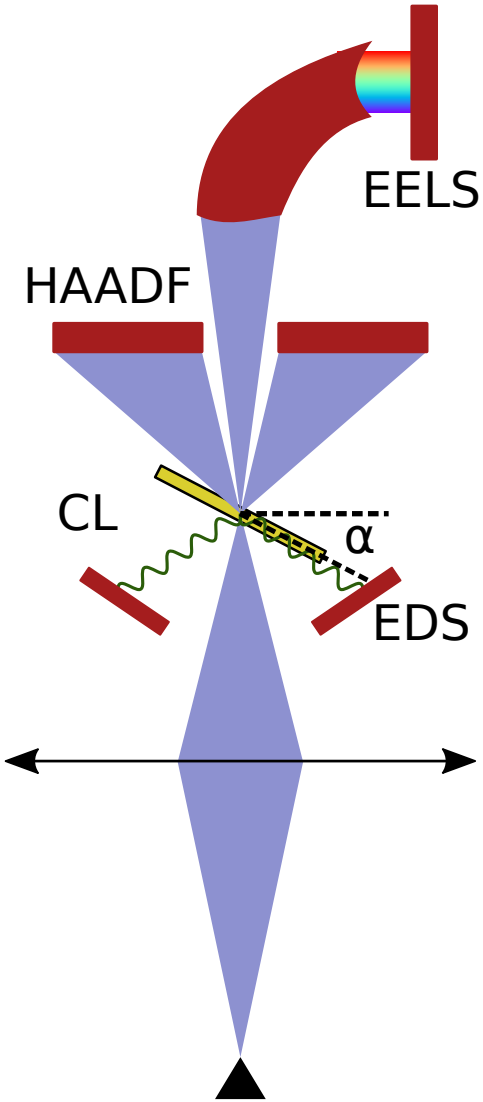
Detectors	Dimensions
HAADF	(x, y|)
EDX	(x, y | E)
CL	(x, y | ω)
EELS	(x, y | E)

Why the emphasis on multi-dimensional datasets?



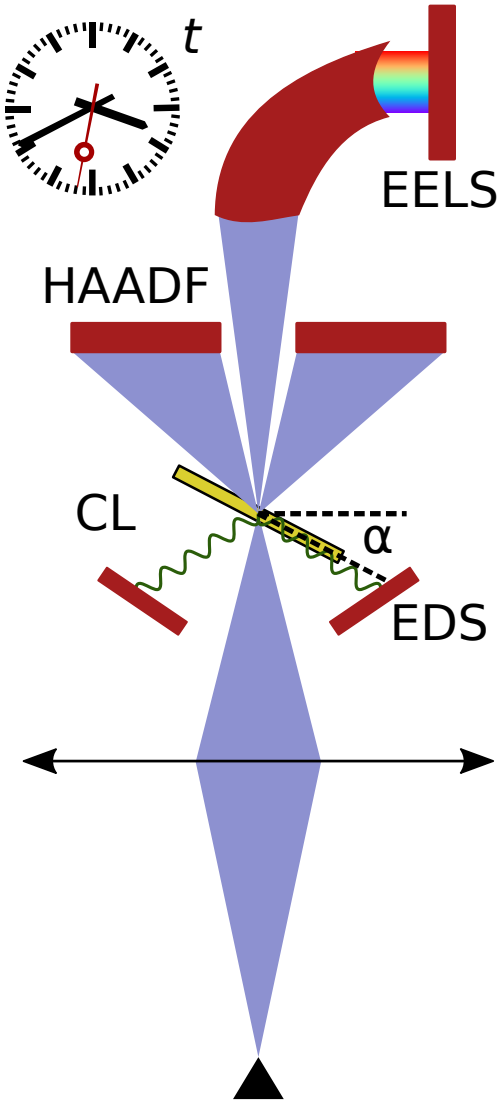
Detectors	Dimensions
HAADF	(x, y|)
EDX	(x, y | E)
CL	(x, y | ω)
EELS	(x, y | E)
Diffraction	(x, y | x^*,y^*)

Why the emphasis on multi-dimensional datasets?



Detectors	Dimensions
HAADF	(x, y|)
EDX	(x, y, α | E)
CL	(x, y | ω)
EELS	(x, y, α | E)
Diffraction	(x, y, α | x^*,y^*)

Why the emphasis on multi-dimensional datasets?



Detectors	Dimensions
HAADF	(x, y|)
EDX	(x, y, α, t| E)
CL	(x, y, t| ω)
EELS	(x, y, α, t| E)
Diffraction	(x, y, α, t| x^*,y^*)

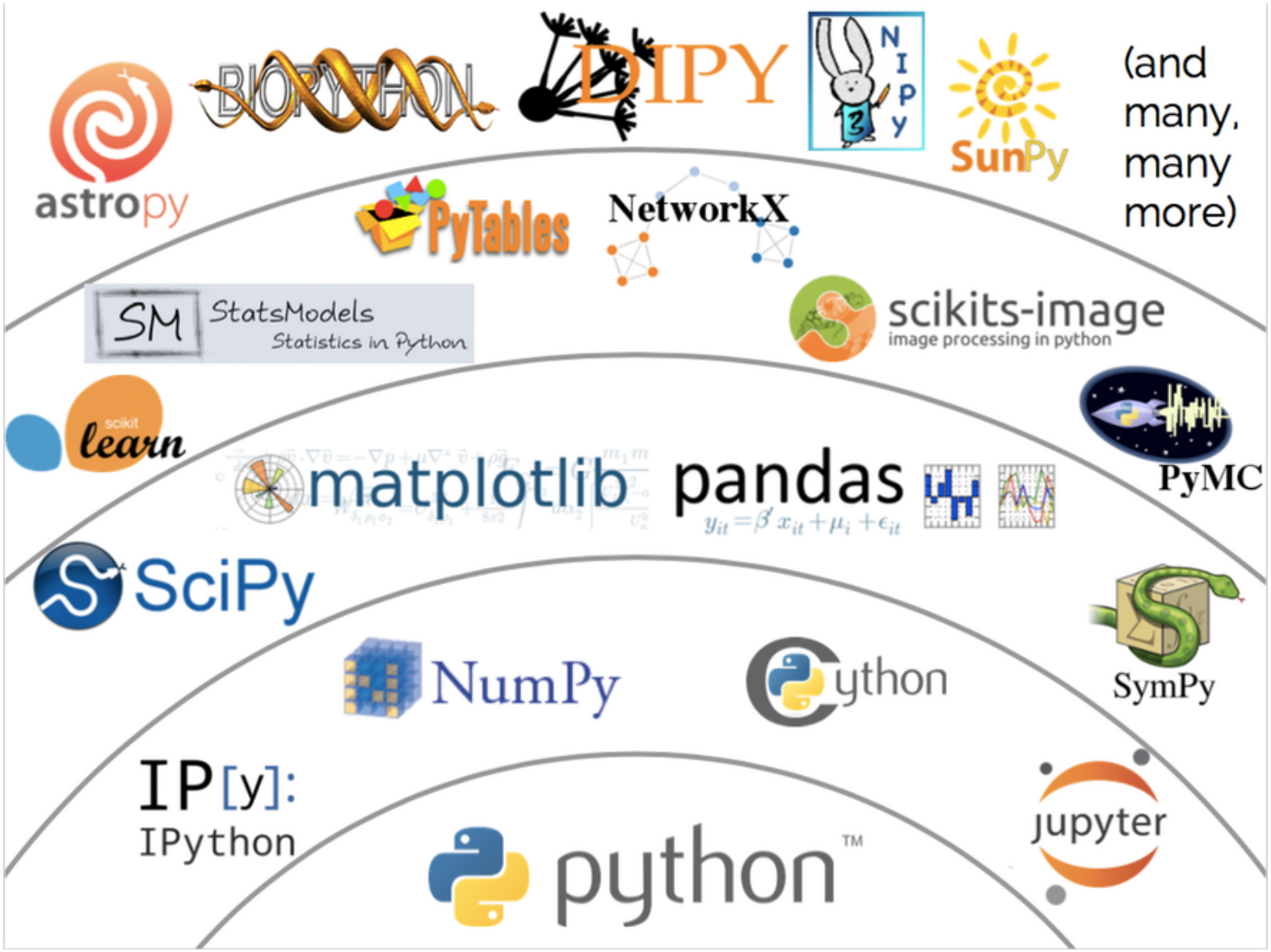
Why Python?

- Scripting language i.e. no need to compile ⇒fast development.
- Genuine programming language.
- Readability: no need to know Python to understand what the code does ⇒Low entry barrier.
- Runs natively on Windows, Mac OS, Linux.
- Open-source: not a black box and is free.

Why Python?

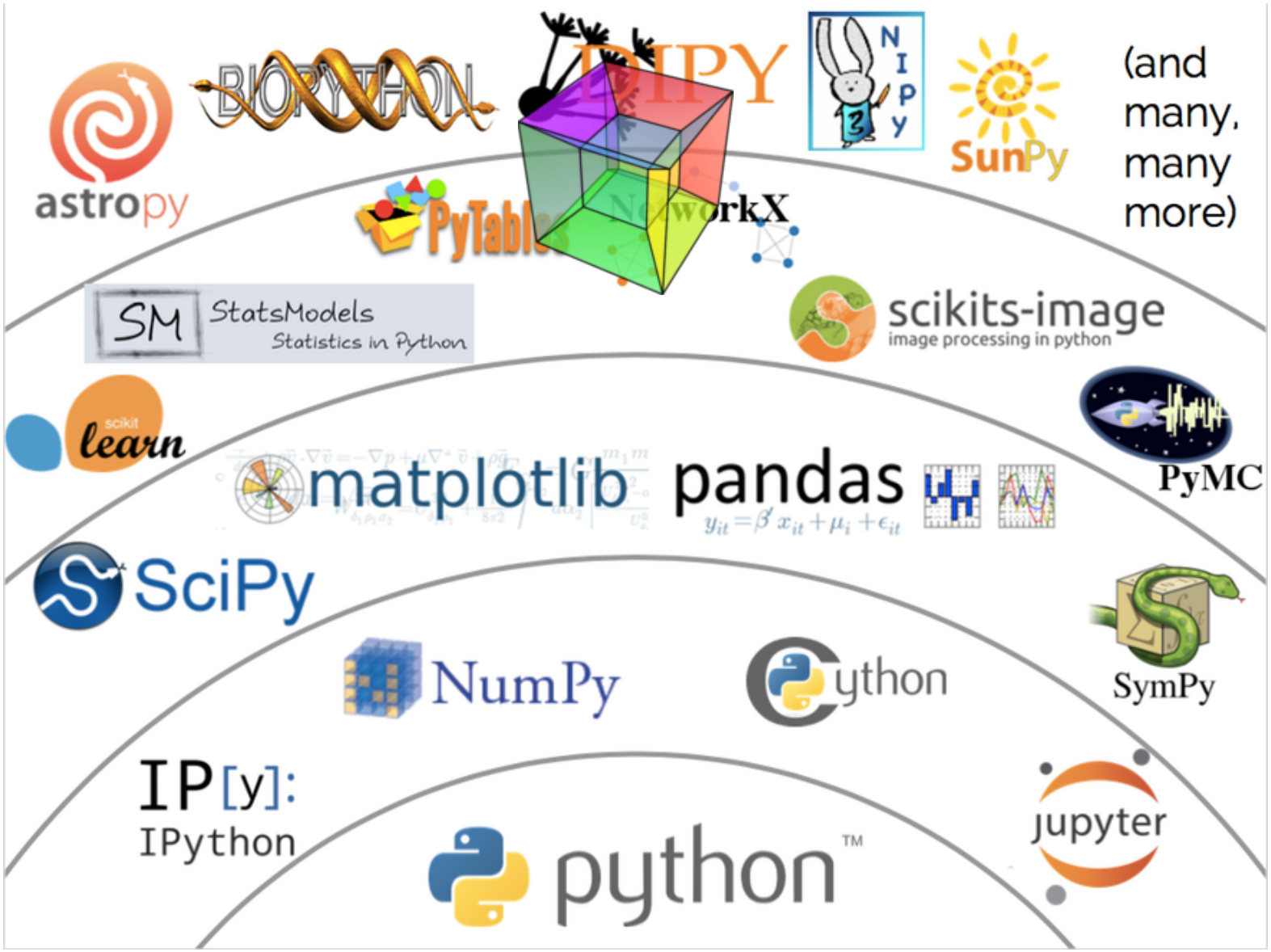
- Scripting language i.e. no need to compile ⇒fast development.
- Genuine programming language.
- Readability: no need to know Python to understand what the code does ⇒Low entry barrier.
- Runs natively on Windows, Mac OS, Linux.
- Open-source: not a black box and is free.
- Lingua franca for scientific computing.
- Unmatched environment for scientific computing:
  - Numpy+Scipy+matplotlib ⩾(Matlab + Toolkits replacement)
  - Jupyter
  - sklearn, skimage...

The Scientific Python Stack



State of the "Scientific Python Stack" circa 2015, Jake VanderPlas

The Scientific Python Stack



State of the "Scientific Python Stack" circa 2015, Jake VanderPlas

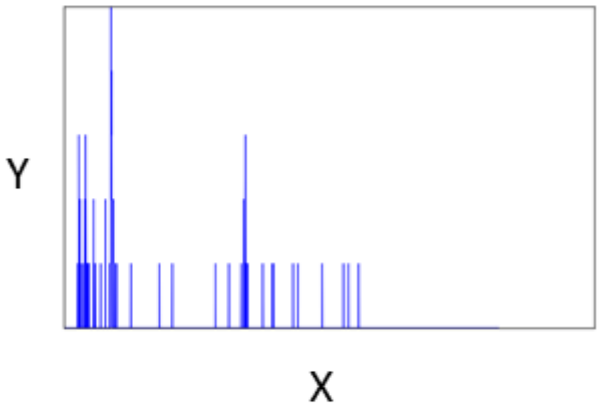
Why Python?

I came to Python lured by the language, but I stayed because of its [scientific] community.

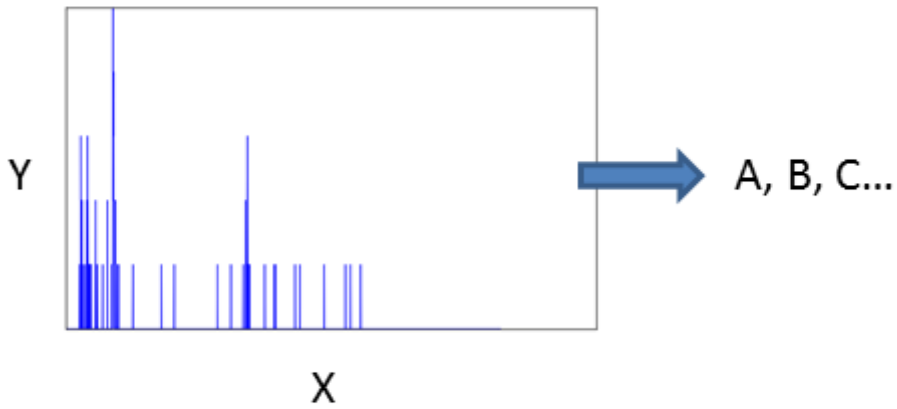
Fernando Pérez (UC Berkeley, creator of IPython)

Achieving sustainability: recycling

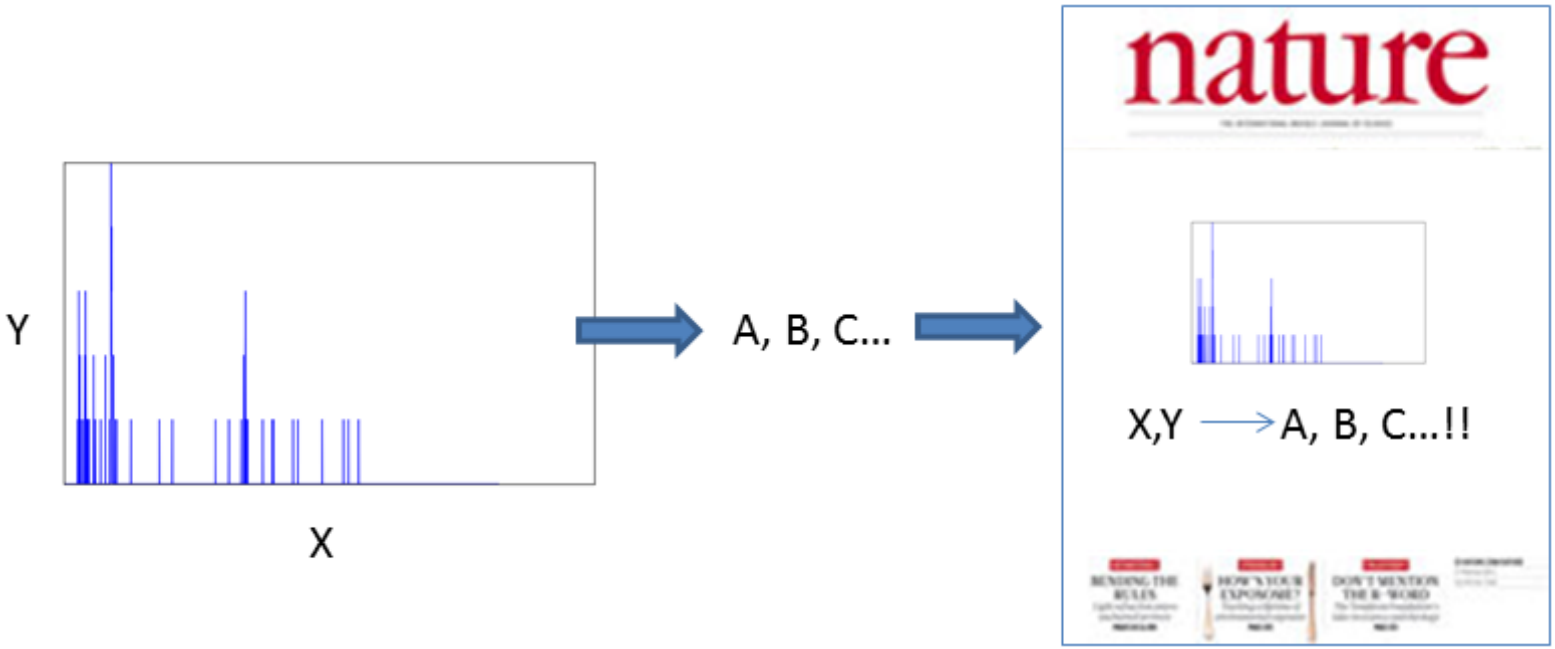
Achieving sustainability: recycling



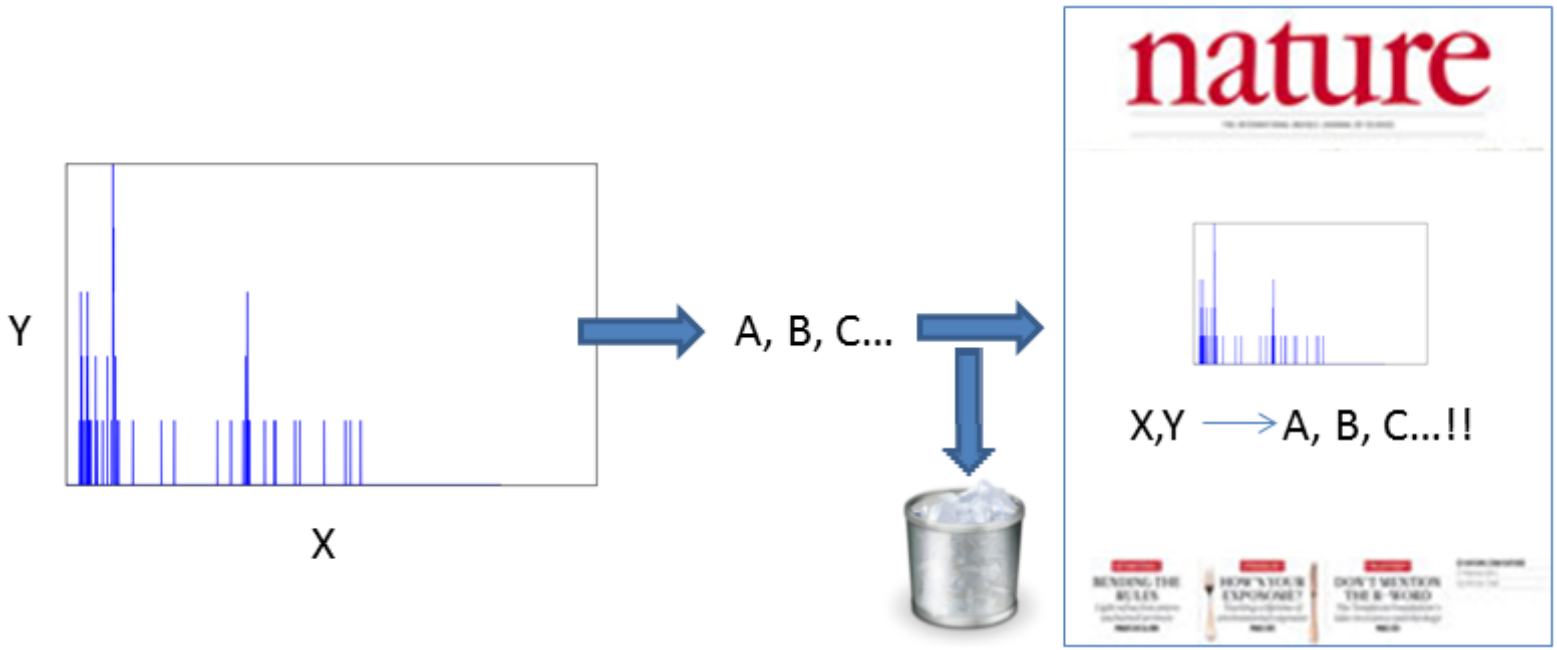
Achieving sustainability: recycling



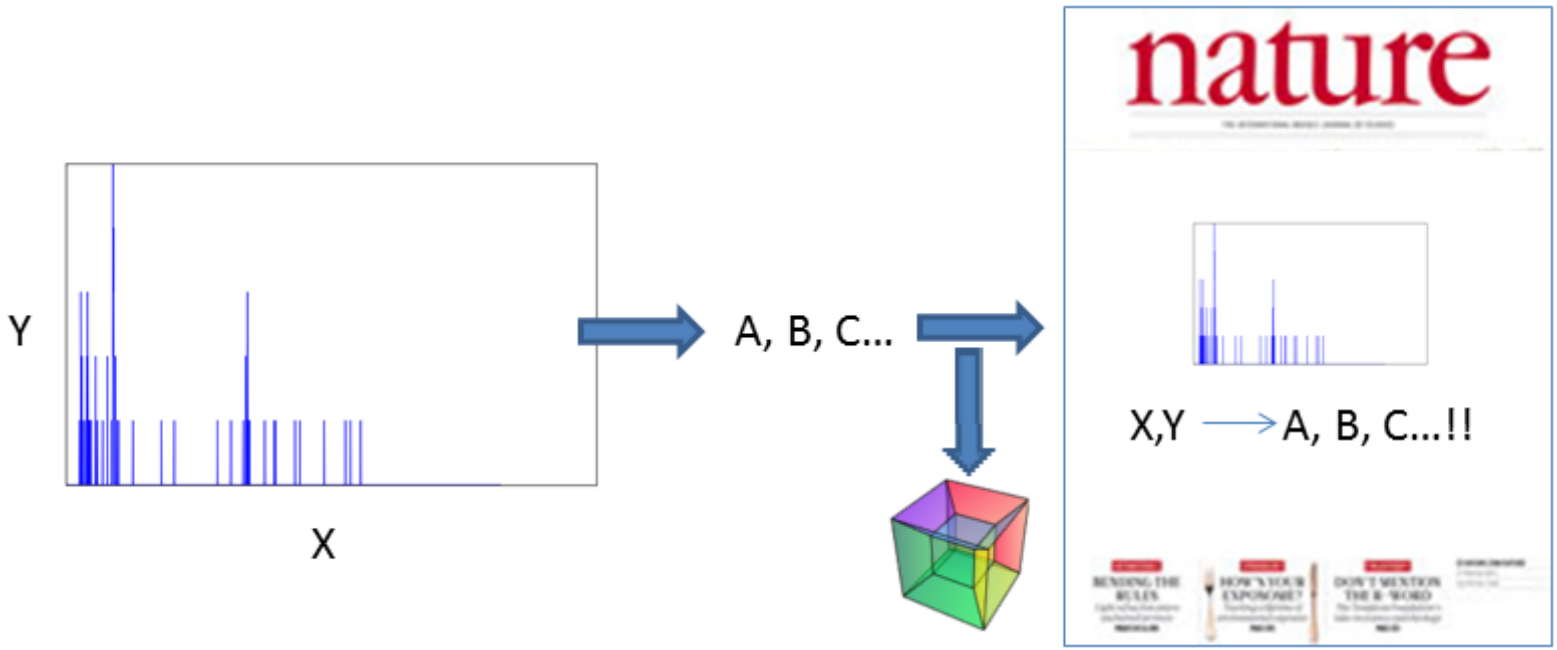
Achieving sustainability: recycling



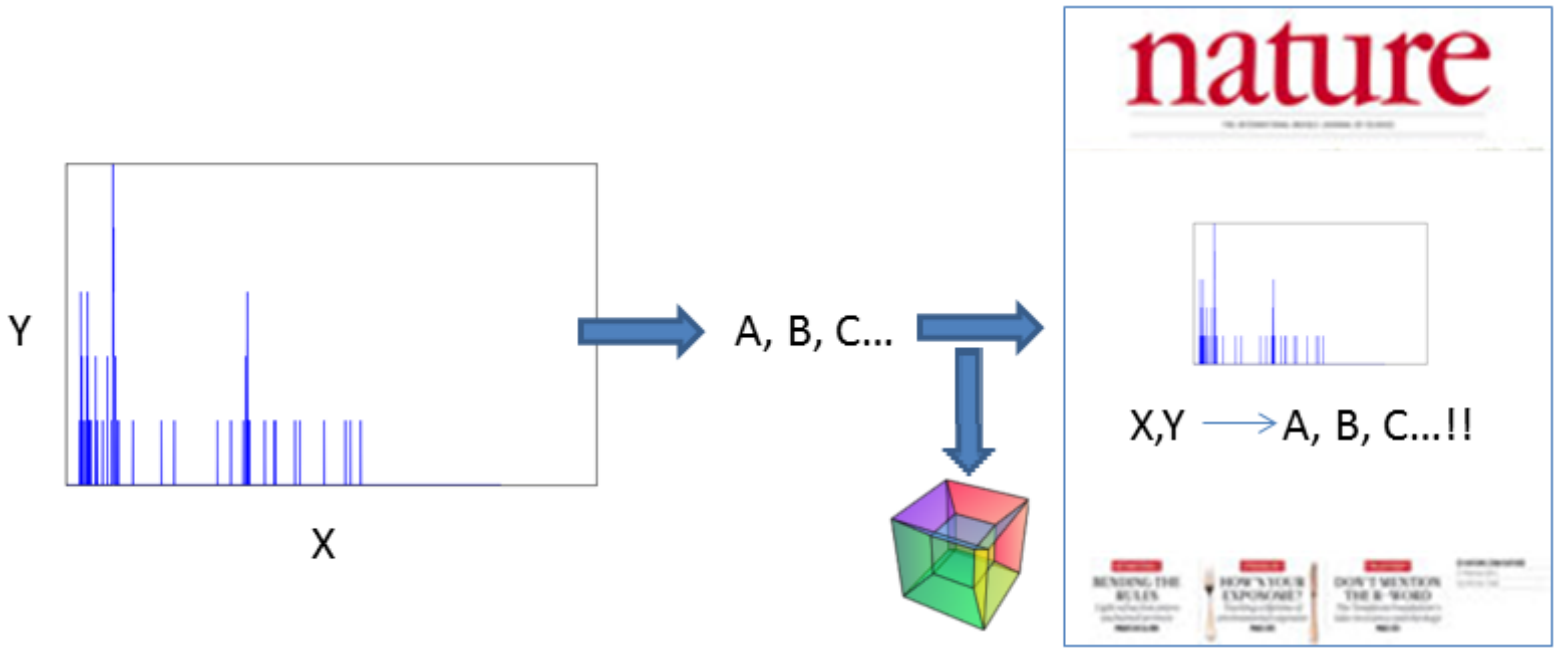
Achieving sustainability: recycling



Achieving sustainability: recycling



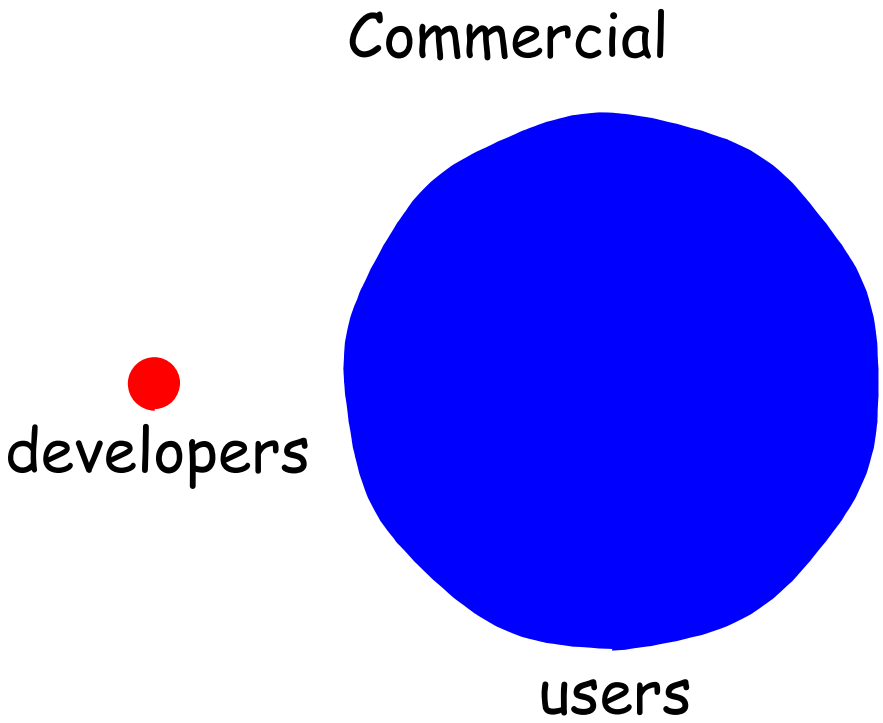
Achieving sustainability: recycling



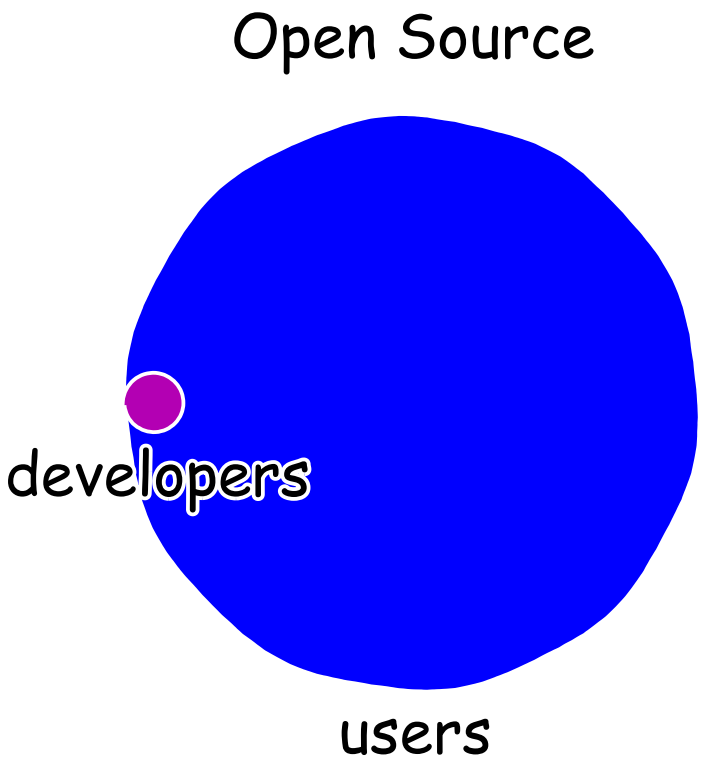
HyperSpy is . . .

a *peer-reviewed open-access journal* specialized in code for EM multi-dimensional data analysis.

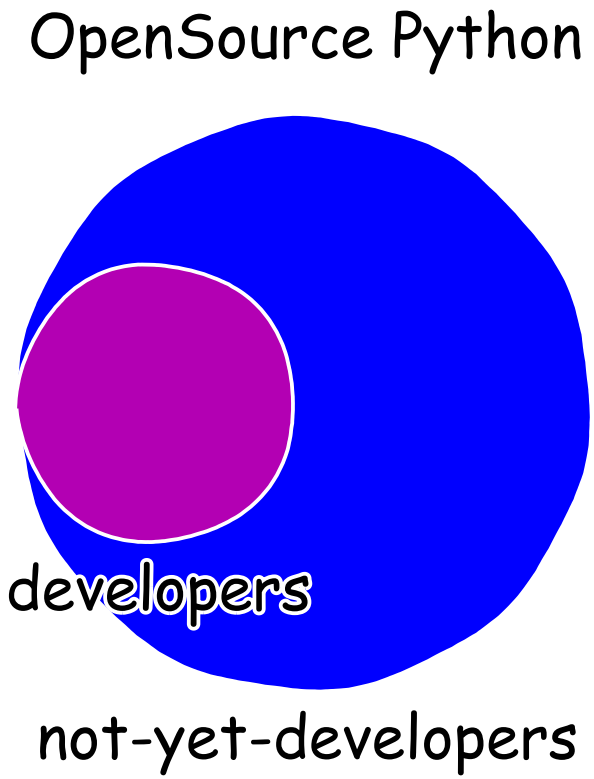
Sustainable opensource development



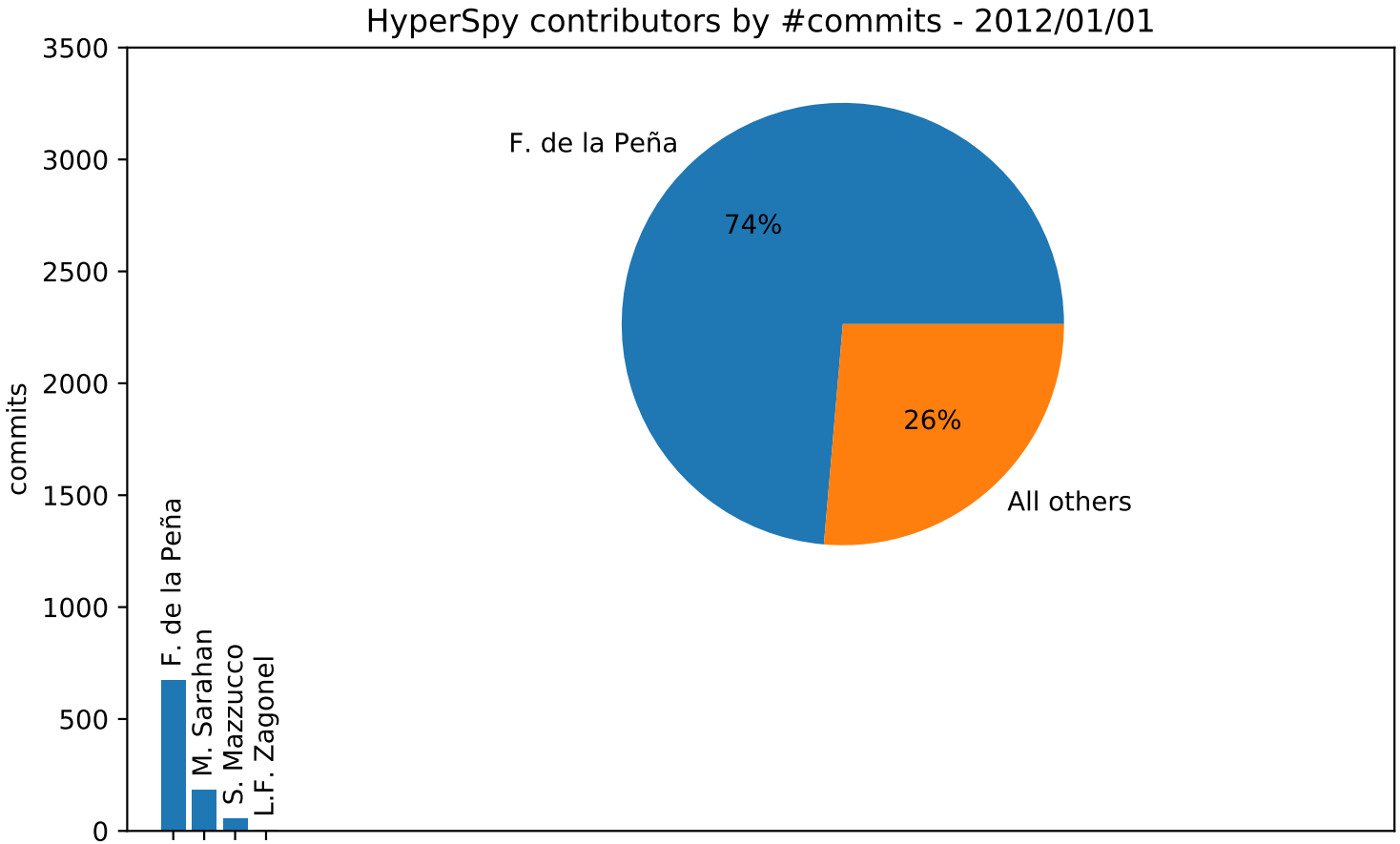
Sustainable opensource development



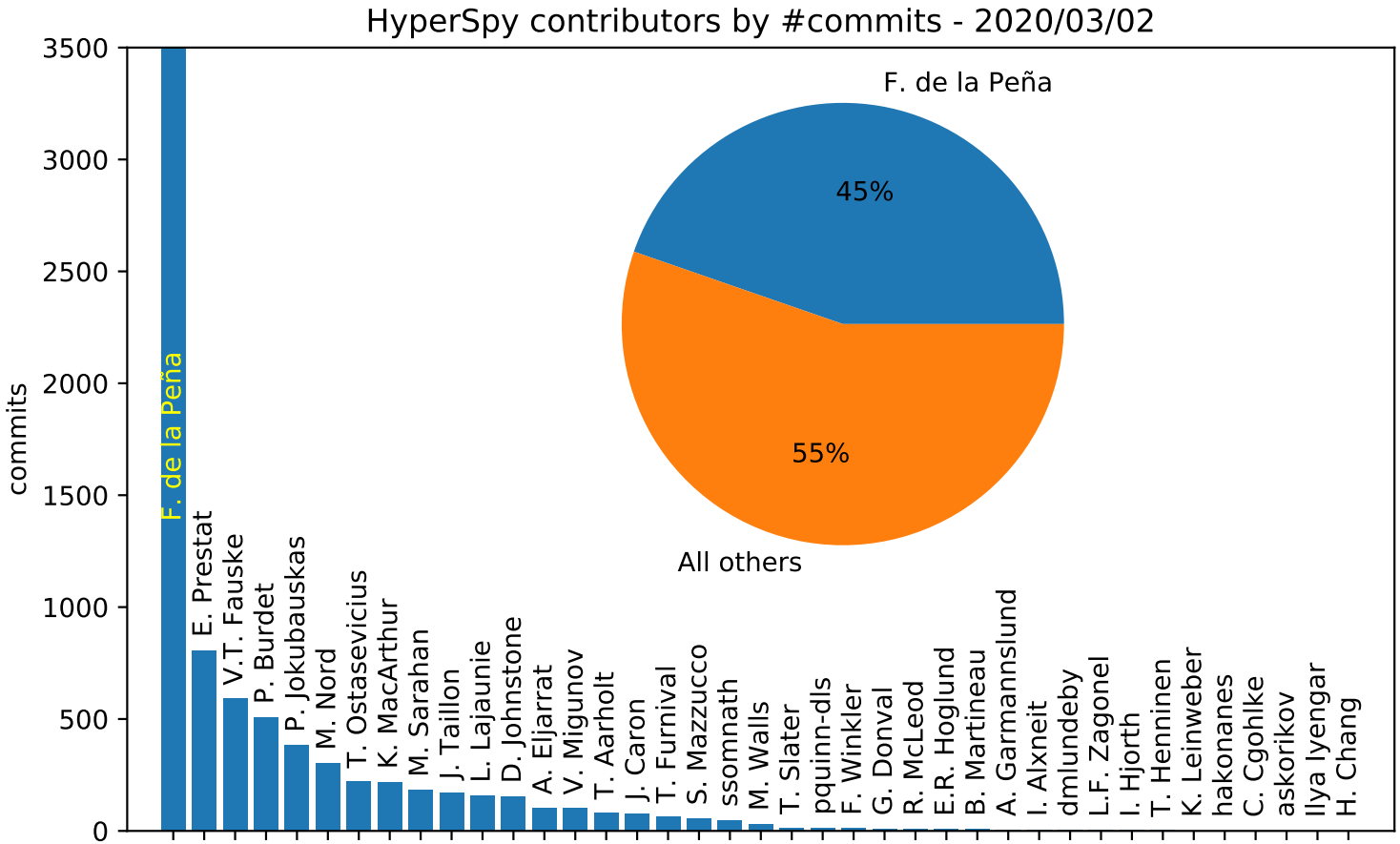
Sustainable opensource development



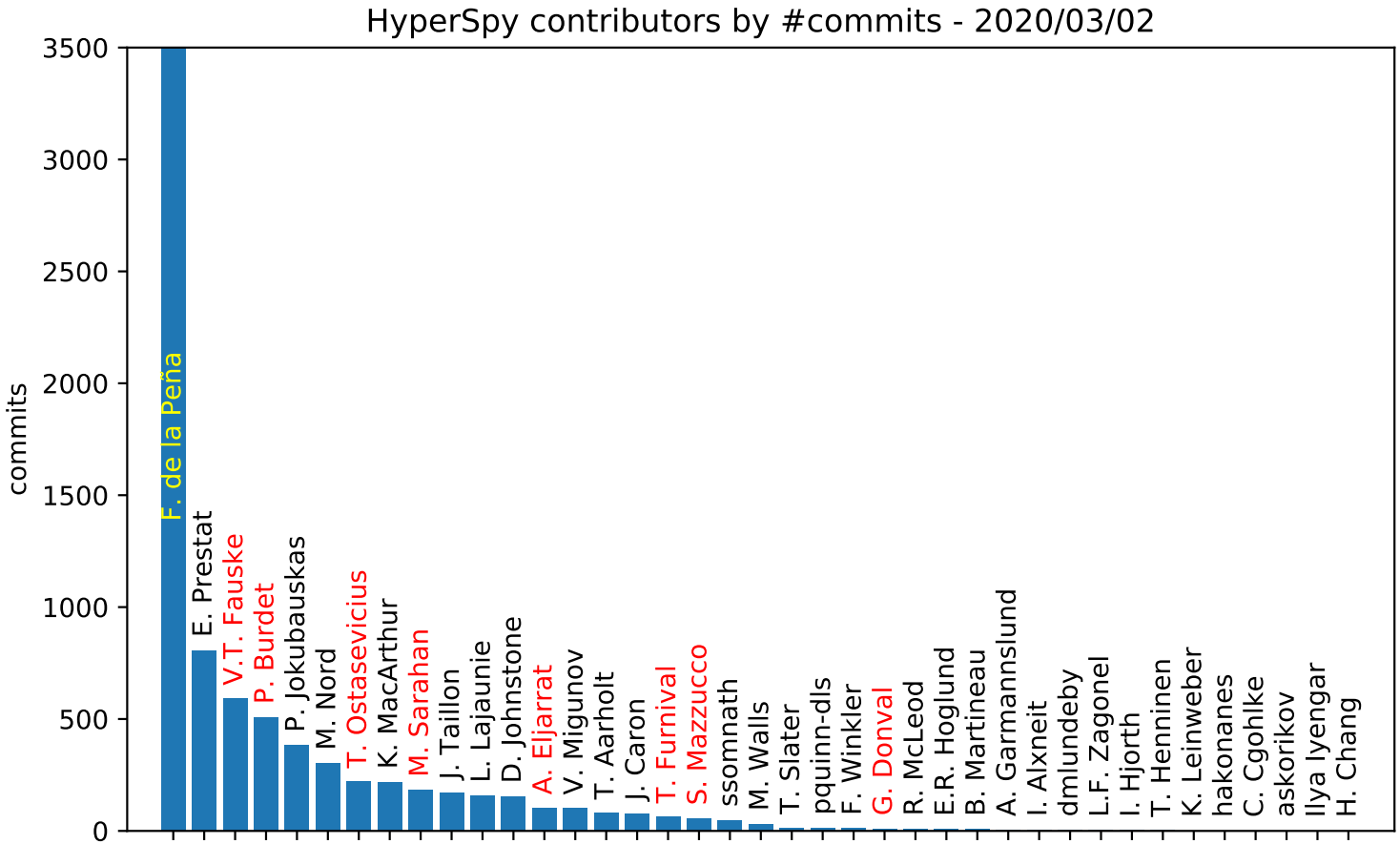
Is open development working for HyperSpy?



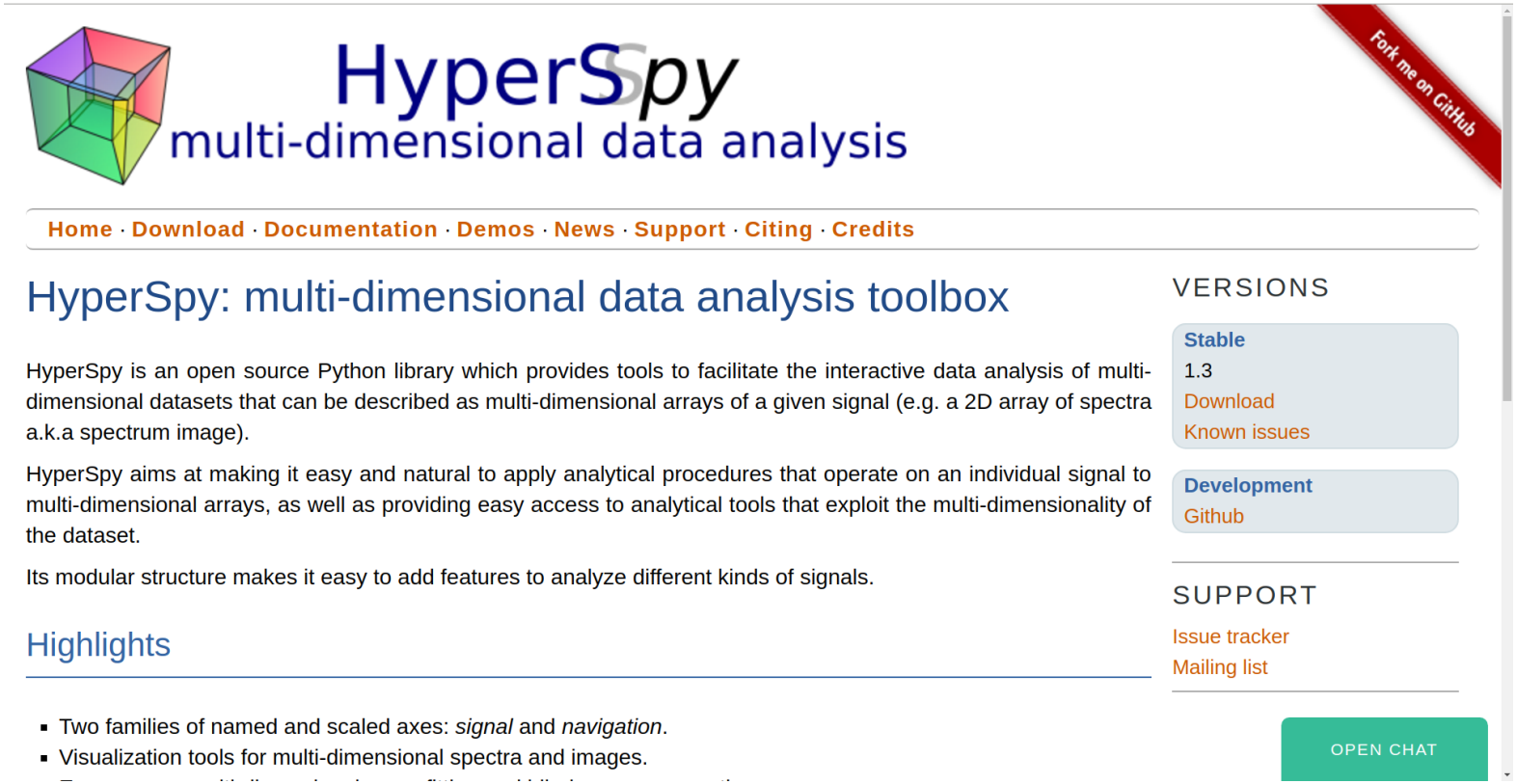
Is open development working for HyperSpy?



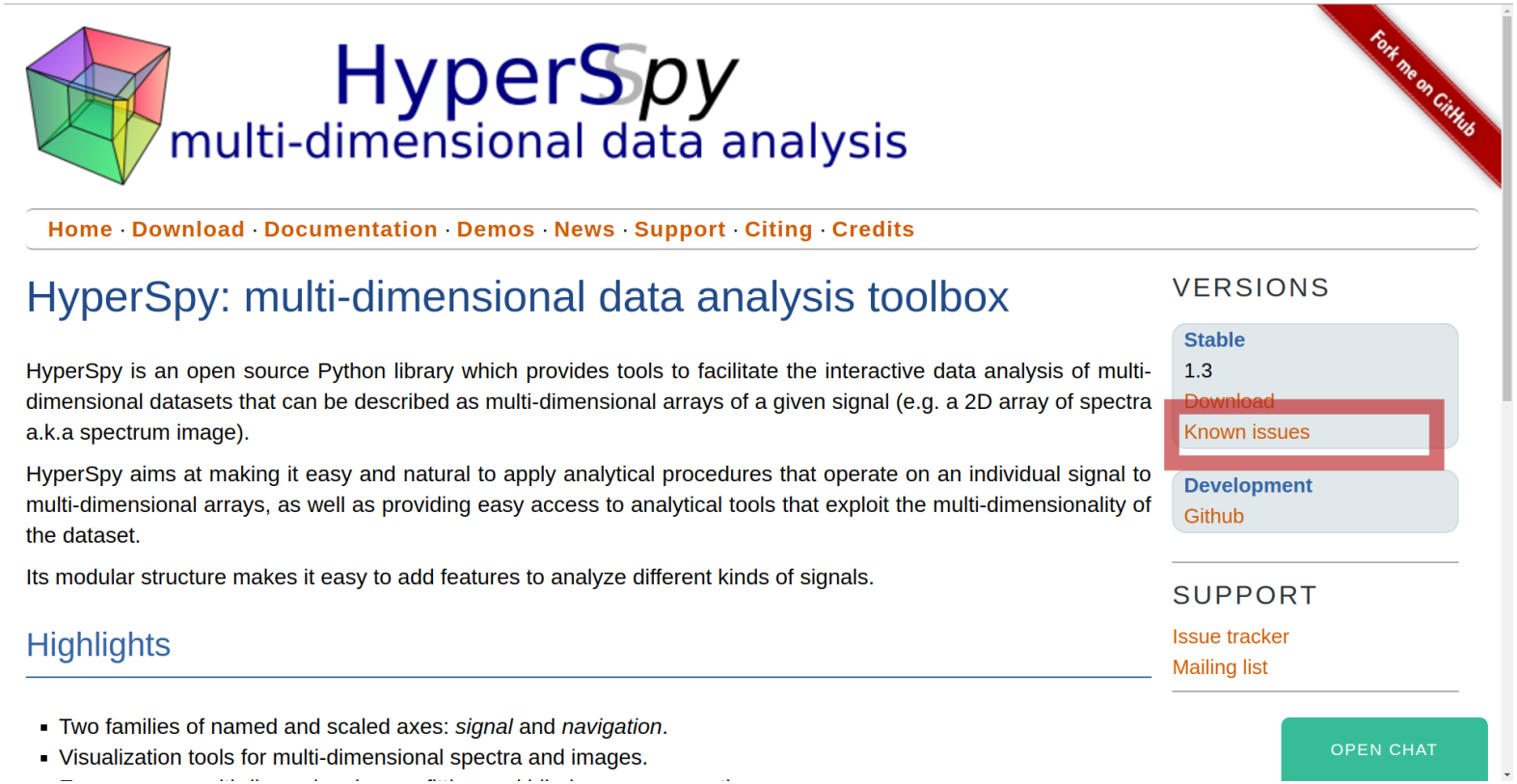
Is open development working for HyperSpy?



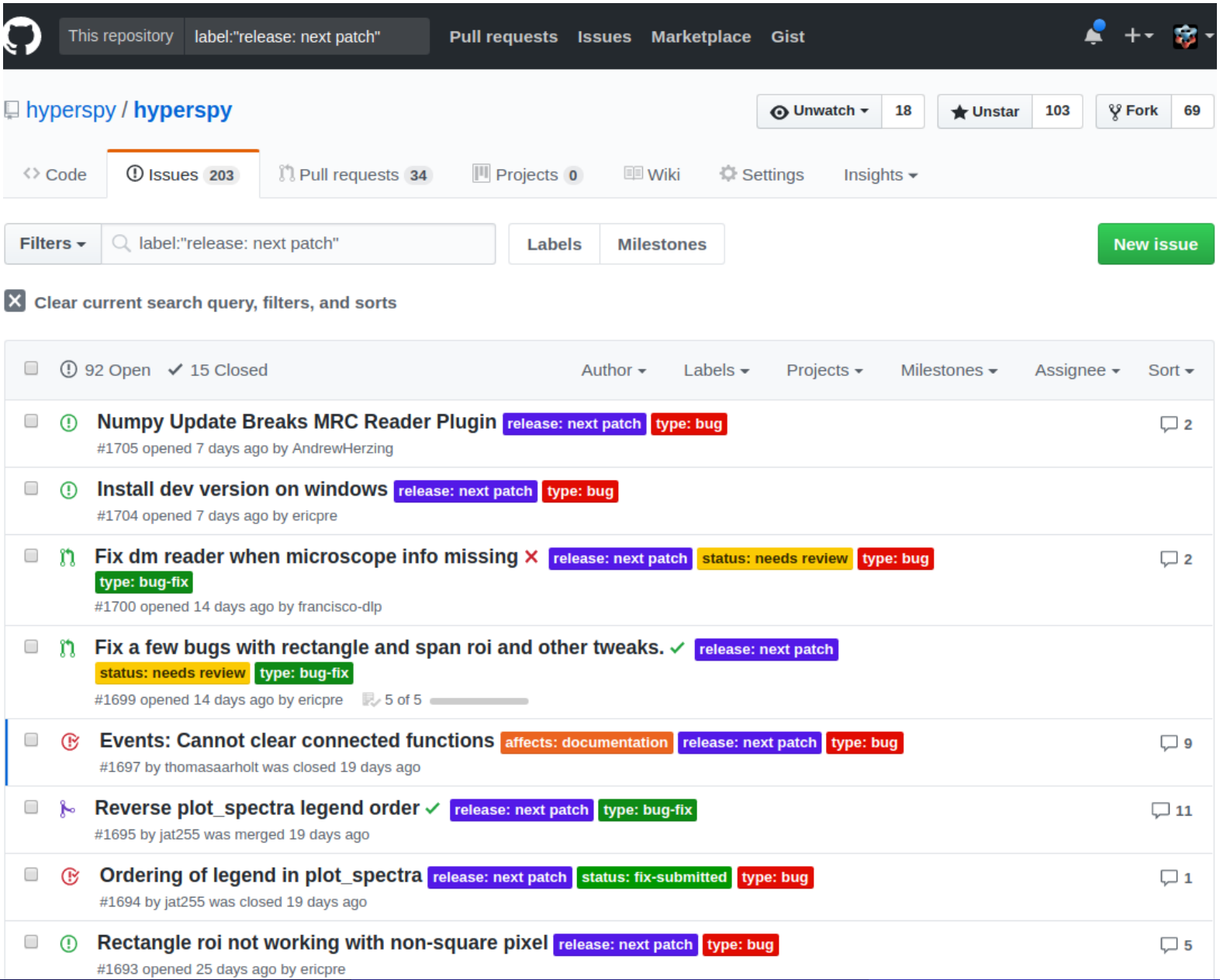
Community and support



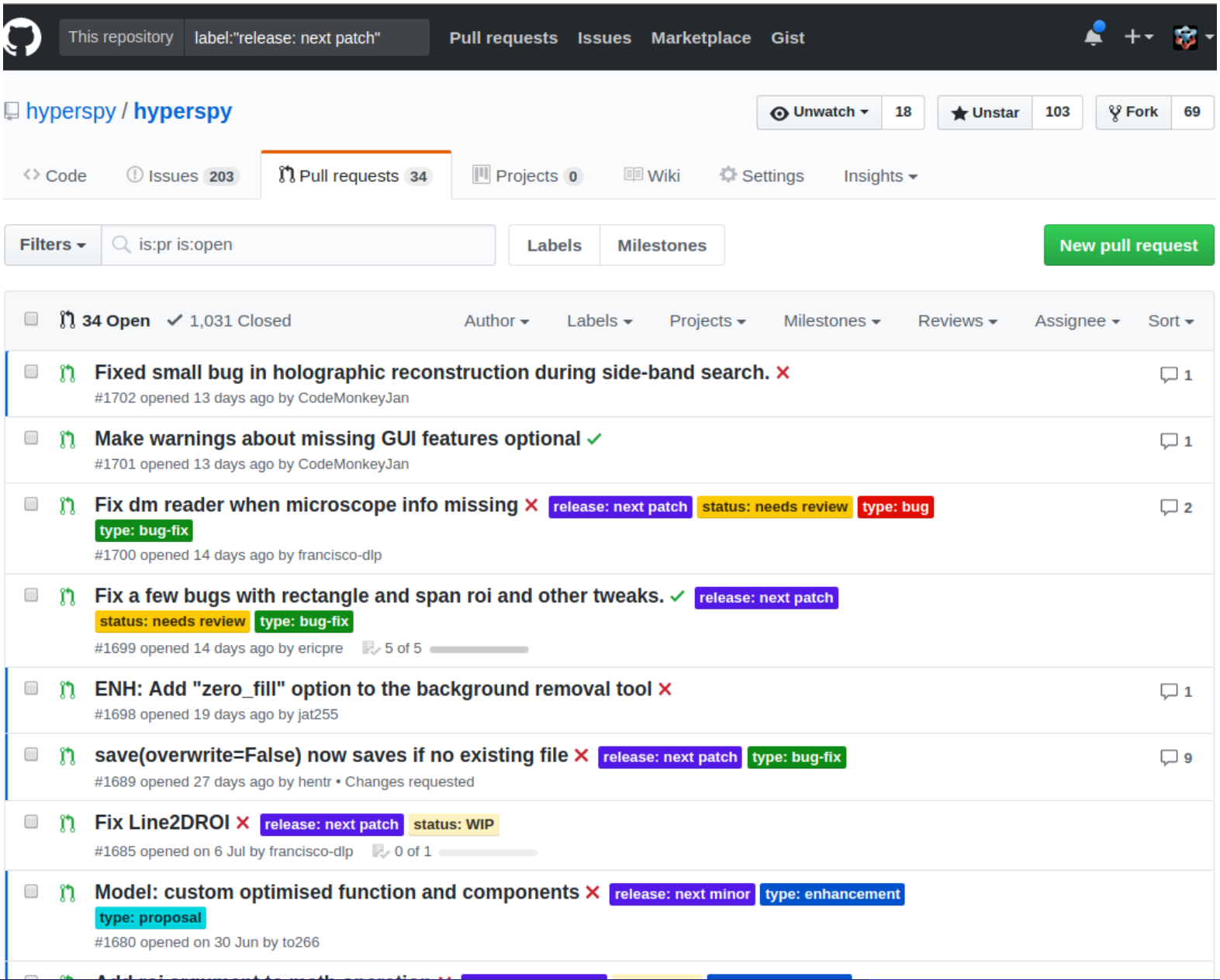
Community and support



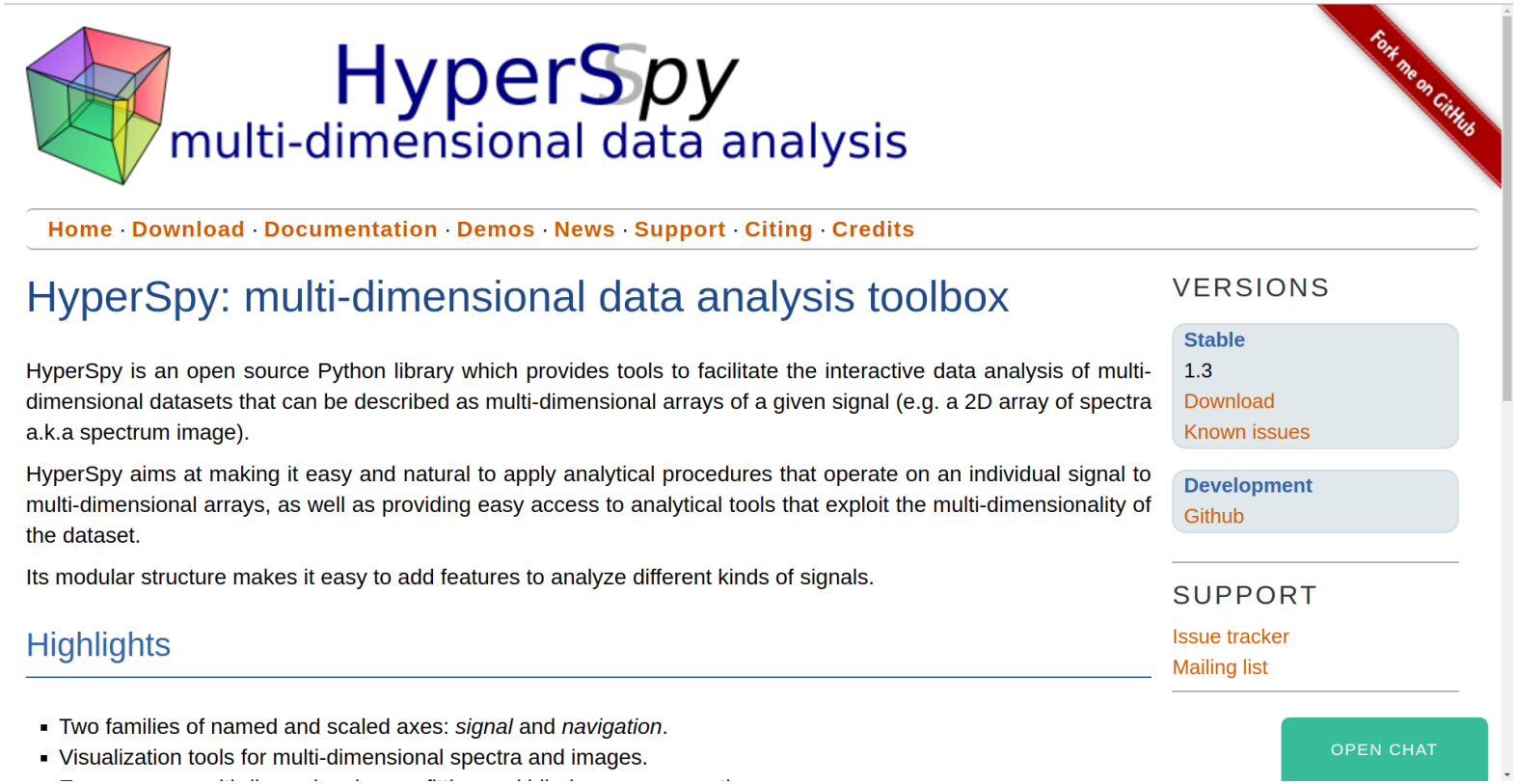
Community and support



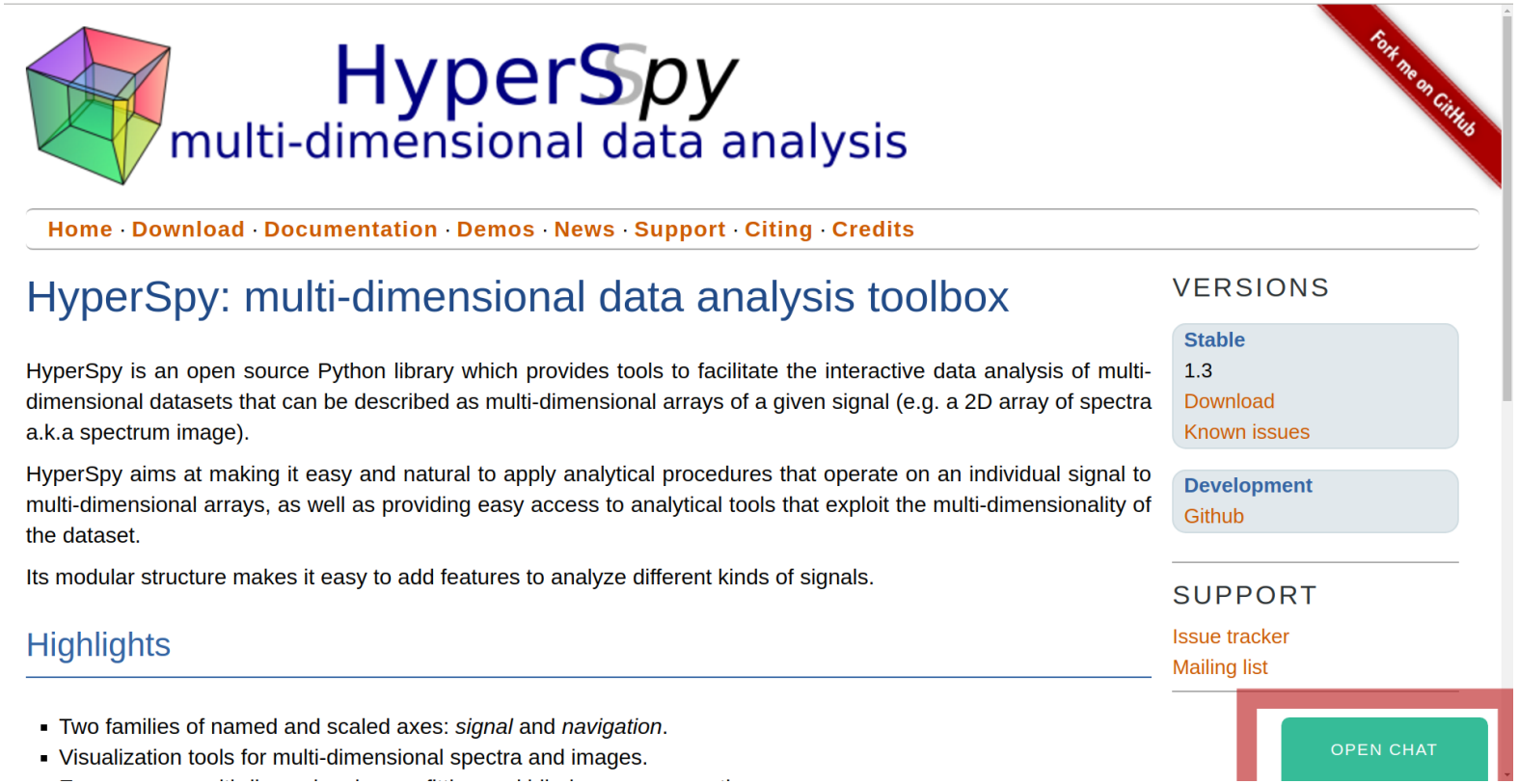
Community and support



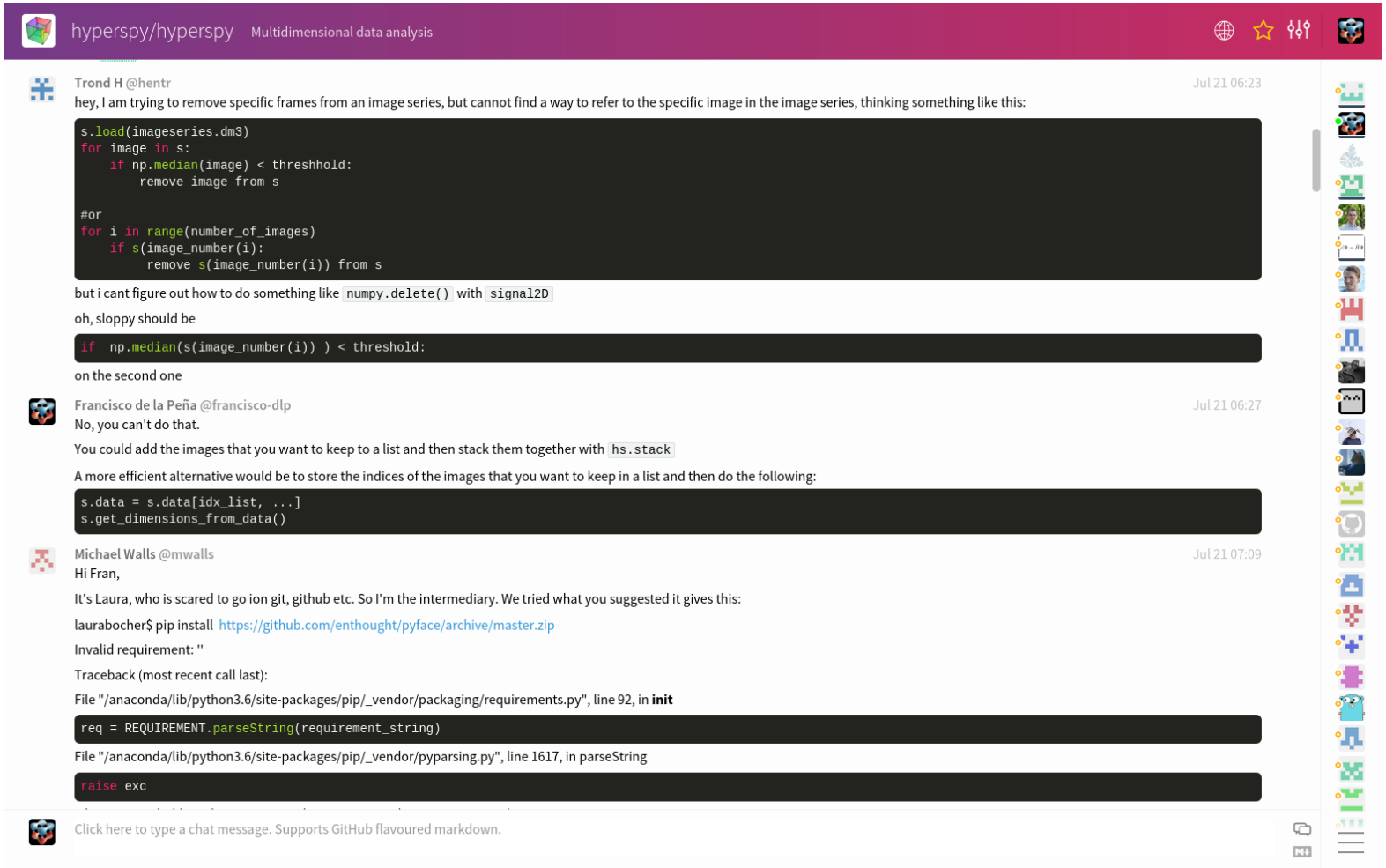
Community and support



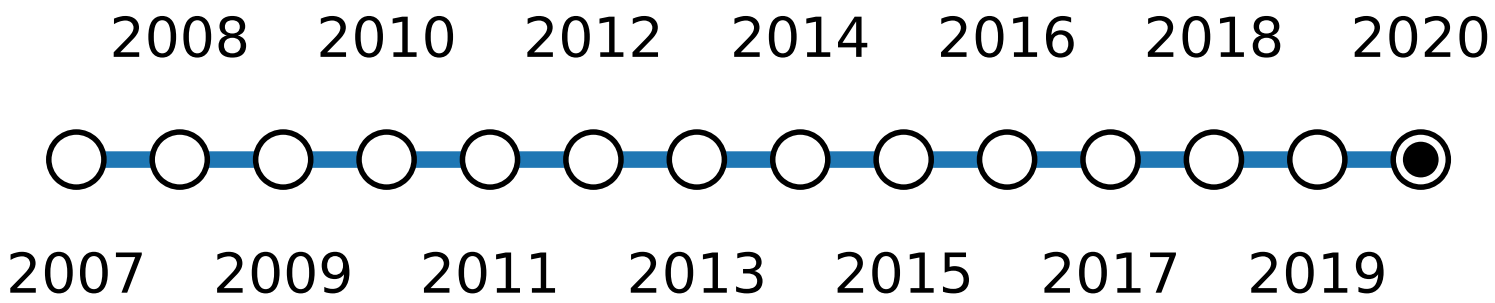
Community and support



Community and support

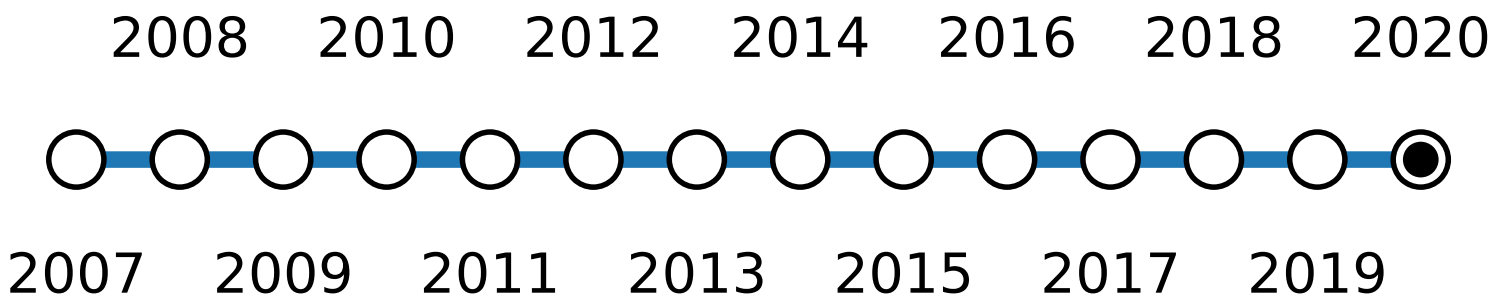


What is HyperSpy? (2020)



- Mature code base
- Used by a growing number of external packages e.g.: pyXem, atomap, ParticleSpy...

What is HyperSpy? (2020)

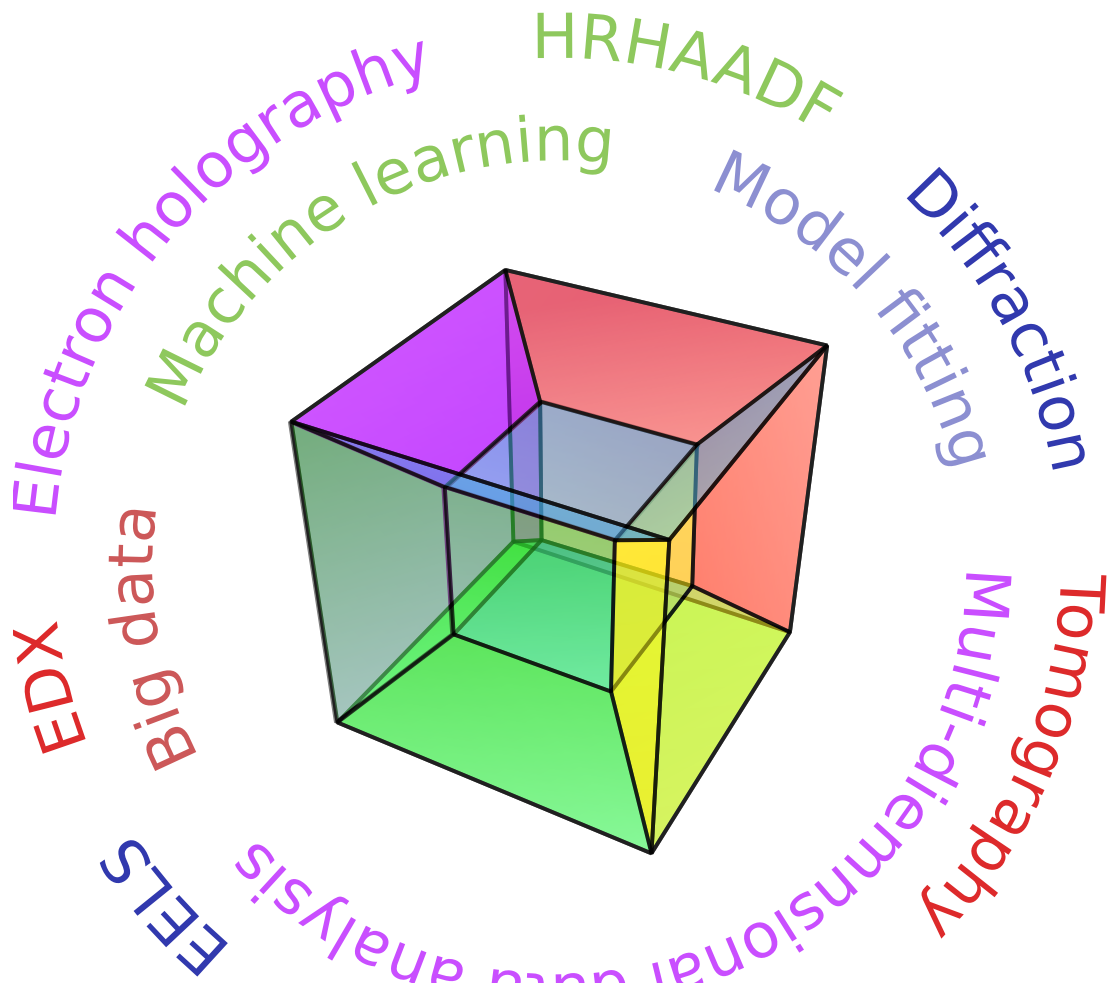


- Mature code base
- Used by a growing number of external packages e.g.: pyXem, atomap, ParticleSpy...

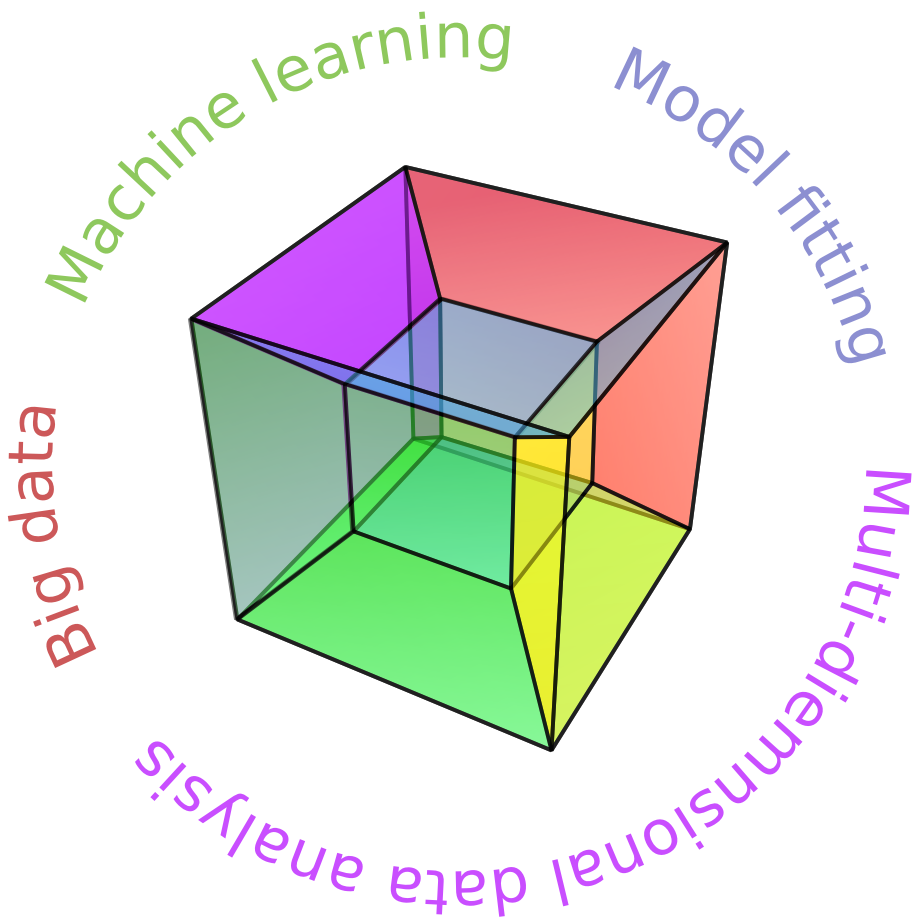
HyperSpy is . . .

A *community* that aims at *democratizing* advanced data processing and analysis of *multi-dimensional datasets* across scientific fields by developing the features and *syntax* of the Python package of the same name.

Less is more: splitting HyperSpy



Less is more: splitting HyperSpy



Conclusion

When you people have a new idea:

Conclusion

When you people have a new idea:

- In academia: I hope no one scoops me

Conclusion

When you people have a new idea:

- In academia: I hope no one scoops me
- In Open Source: thanks goodness someone already thought of this!

Elizabeth Seiver @tweetotaler in Twitter

Thank you all for you attention