

C. Troupin*, A. Barth*
C. Muñoz*, S. Watelet*, & J.-M. Beckers*



*GHER-University of Liège

*Balearic Islands Coastal Ocean
Observing and Forecasting System

*Notebooks for documenting
work-flows*

Motivation

Reproducibility

Notebooks: interactive computational environments

Notebooks combine:

- 1 code fragments that can be executed,
- 2 text for the description of the application and
- 3 figures illustrating the data or the results.

```
In [2]: import numpy as np
import matplotlib.pyplot as plt
```

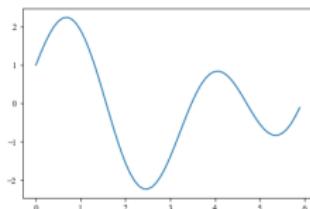
Data

Let's create a simple function.

```
In [6]: x = np.arange(0, 6, .1)
y = np.cos(x) + 1.5 * np.sin(2 * x)
```

Make a simple plot

```
In [7]: plt.plot(x, y)
plt.show()
```



Notebooks: interactive computational environments

Notebooks combine:

- 1 code fragments that can be executed,
- 2 text for the description of the application and
- 3 figures illustrating the data or the results.

"Digital Playground"

"Data Story Telling"

"Computational Narratives"

Notebooks: interactive computational environments

Notebooks combine:

- 1 code fragments that can be executed,
- 2 text for the description of the application and
- 3 figures illustrating the data or the results.

"Interactive notebooks: Sharing the code", Nature (2014)

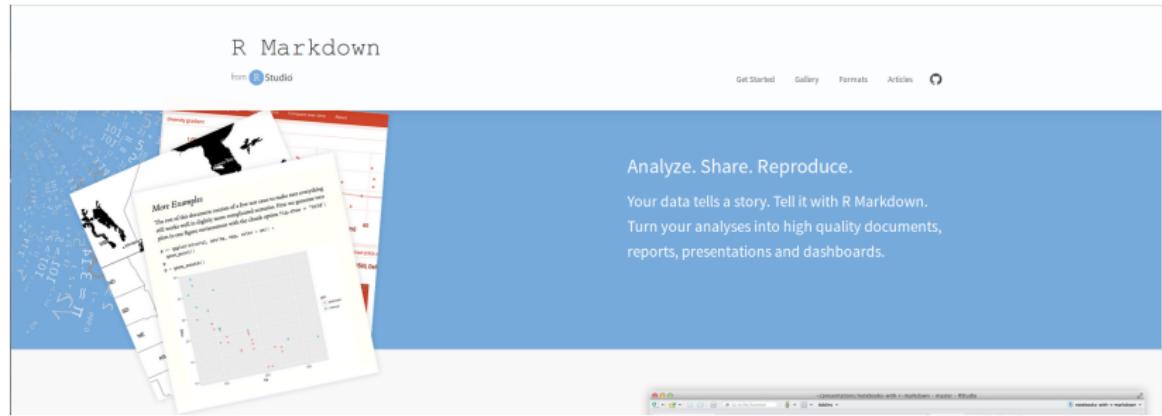
[http://www.nature.com/news/
interactive-notebooks-sharing-the-code-1.16261](http://www.nature.com/news/interactive-notebooks-sharing-the-code-1.16261)

Interactive environments:

what exists today?

R-Markdown

<http://rmarkdown.rstudio.com/>



R-Markdown

<http://rmarkdown.rstudio.com/>

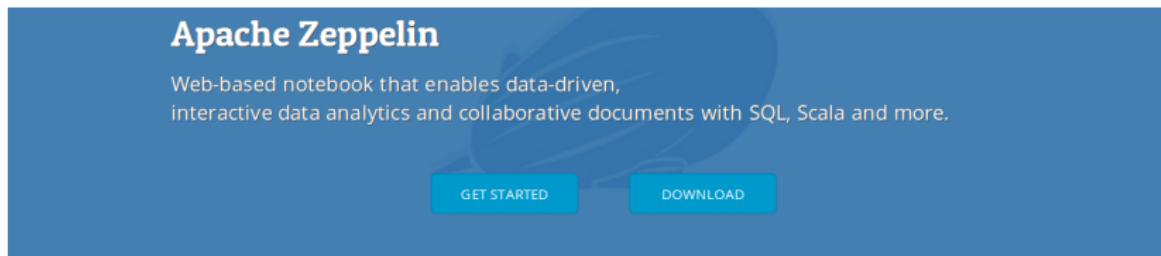
Creation of dynamic, self-contained documents
with embedded chunks of code.

Features of interest:

- ▶ Possible to export in journal
(<https://github.com/rstudio/rtitles>) or
presentation formats
- ▶ \LaTeX templates to ensure journal standards

Apache Zeppelin

<https://zeppelin.apache.org/>



The image shows the official Apache Zeppelin landing page. The header features the Apache logo and the text "Apache Zeppelin". Below the header, a sub-header reads "Web-based notebook that enables data-driven, interactive data analytics and collaborative documents with SQL, Scala and more." At the bottom of the main content area are two buttons: "GET STARTED" and "DOWNLOAD".

TECHNOLOGIES



Apache Zeppelin

<https://zeppelin.apache.org/>

Web-based notebook for
data-driven, interactive and collaborative documents.
Intended for *big data* and large scale projects.

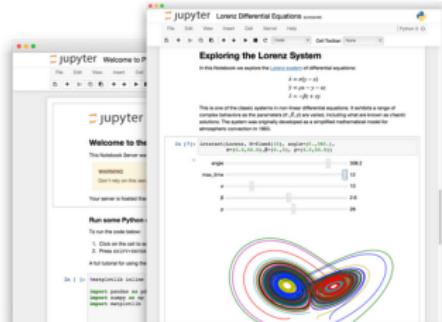
Features of interest:

- ▶ Languages can be mixed in the same notebook
- ▶ Users can write their own interpreter (*language backend*)

Jupyter

<http://jupyter.org/>

(stands for Julia - Python - R)



Language of choice



Share notebooks



Interactive widgets



Big data integration



The Jupyter Notebook

The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and explanatory text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, machine learning and much more.

Jupyter

<http://jupyter.org/> (stands for Julia - Python - R)

Web application for the creation and sharing of notebook-type documents.

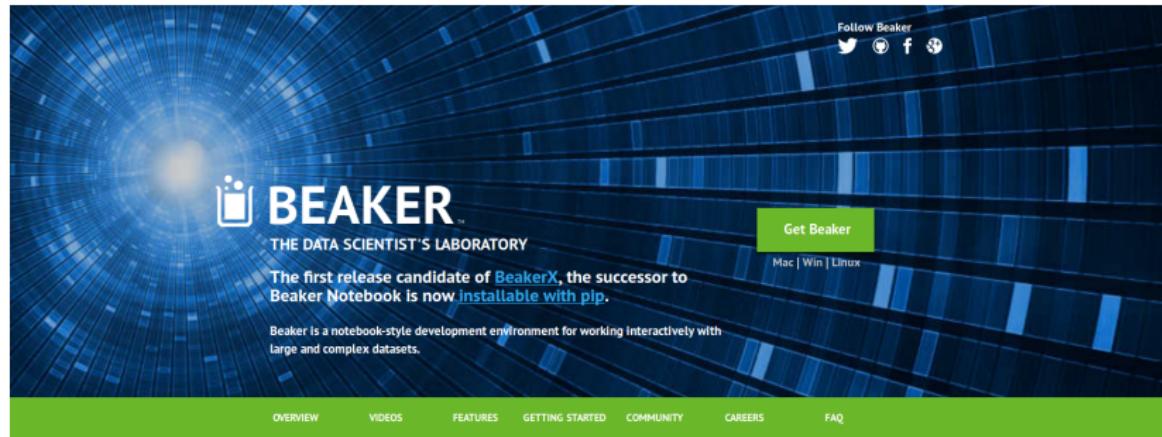
Evolved from IPython, a command shell for interactive computing (2001).

Features of interest:

- ▶ More than 40 language *kernels* available
- ▶ Can be used as a multi-user server (`jupyterhub`)
→ avoid installation steps on several users' machine

Beaker

<http://beakernotebook.com/>



The screenshot shows the Beaker homepage with a blue digital tunnel background. The Beaker logo (a stylized flask icon) and the word 'BEAKER' in large white letters are on the left. Below the logo is the tagline 'THE DATA SCIENTIST'S LABORATORY'. A green button labeled 'Get Beaker' is on the right, with 'Mac | Win | Linux' underneath. Social media links for Twitter, GitHub, Facebook, and LinkedIn are at the top right. A sub-headline says 'The first release candidate of BeakerX, the successor to Beaker Notebook is now installable with pip.' Below that is a description of Beaker as a notebook-style environment for large datasets. A green navigation bar at the bottom includes links for Overview, Videos, Features, Getting Started, Community, Careers, and FAQ.

The Perfect Tool for Iterative Exploration

Beaker

<http://beakernotebook.com/>

Notebook-style development environment for working interactively with large and complex datasets.

Features of interest:

- ▶ Usage of different languages in different cells, within the same notebook
- ▶ Language manager

<https://cocalc.com/>

"Collaborative Calculation in the Cloud"



Collaborative Calculation in the Cloud

[Run CoCalc](#)

or [sign in](#) with your account

Online computing environment

<https://cocalc.com/>

"Collaborative Calculation in the Cloud"

Web-based cloud computing platform, formerly called formerly called SageMathCloud.

Features of interest:

- ▶ Support of many languages
- ▶ Users to upload their file on the platform to be later read or processed

Comparison

Tool name	R- Markdown	jupyter	beaker	Cocalc	Zeppelin
-----------	----------------	---------	--------	--------	----------

Comparison

Tool name	R- Markdown	jupyter	beaker	Cocalc	Zeppelin
GitHub	rmarkdown	notebook	beakerx	cocalc	zeppelin

Comparison

Tool name	R-Markdown	jupyter	beaker	Cocalc	Zeppelin
GitHub	rmarkdown	notebook	beakerx	cocalc	zeppelin
Languages	R, Python, SQL, Bash, Rcpp, Stan, JavaScript	Julia, Python, R, Scala, Bash, Octave, Rubi, Fortran, PHP, ...	Julia, Python, R, Javascript, C++, Torch, Scala, Bash, Octave, Rubi, Fortran, ...	R, Python, Octave, Cython, Julia, Java, C/C++, Perl, Ruby	Scala, Python, SparkSQL, Hive, Markdown

Comparison

Tool name	R-Markdown	jupyter	beaker	Cocalc	Zeppelin
GitHub	rmarkdown	notebook	beakerx	cocalc	zeppelin
Languages	R, Python, SQL, Bash, Rcpp, Stan, JavaScript	Julia, Python, R, Scala, Bash, Octave, Rubi, Fortran, PHP, ...	Julia, Python, R, Javascript, C++, Torch, Scala, Bash, Octave, Rubi, Fortran, ...	R, Python, Octave, Cython, Julia, Java, C/C++, Perl, Ruby	Scala, Python, SparkSQL, Hive, Markdown
Export formats	HTML, PDF, MS Word, Beamer, HTML5 slides, ...	PDF, LaTeX, HTML, Markdown, reST	Beaker format		JSON

Comparison

Tool name	R-Markdown	jupyter	beaker	Cocalc	Zeppelin
GitHub	rmarkdown	notebook	beakerx	cocalc	zeppelin
Languages	R, Python, SQL, Bash, Rcpp, Stan, JavaScript	Julia, Python, R, Scala, Bash, Octave, Rubi, Fortran, PHP, ...	Julia, Python, R, Javascript, C++, Torch, Scala, Bash, Octave, Rubi, Fortran, ...	R, Python, Octave, Cython, Julia, Java, C/C++, Perl, Ruby	Scala, Python, SparkSQL, Hive, Markdown
Export formats	HTML, PDF, MS Word, Beamer, HTML5 slides, ...	PDF, LaTeX, HTML, Markdown, reST	Beaker format		JSON
Cloud deployment	-	JupyterHub	Beaker Lab (discontinued)	-	Yes

Summary

- 1 Most of the environments provides supports for many languages
- 2 Beaker is the only option allowing the mix of different languages
but its installation/utilisation are not trivial
- 3 JupyterHub is an option for the deployment on a server so that multiple users can work at the same time using the same infrastructure

Notebook example:

divaND interpolation

A quick example of how to document workflow

[Click here](#)

Conclusions

- 1 Notebooks are interactive computational environments combining code, text, results, figures...

Conclusions

- 1 Notebooks are interactive computational environments combining code, text, results, figures...
- 2 Notebooks are not Virtual Research Environment, but can be one of their components

Conclusions

- 1 Notebooks are interactive computational environments combining code, text, results, figures...
- 2 Notebooks are not Virtual Research Environment, but can be one of their components
- 3 Notebooks are not new (15 years) but their use has evolved

Conclusions

- 1 Notebooks are interactive computational environments combining code, text, results, figures...
- 2 Notebooks are not Virtual Research Environment, but can be one of their components
- 3 Notebooks are not new (15 years) but their use has evolved
- 4 Such a tool is great to document a workflow
Example: climatology production

Future work

- 1 Examples using SeaDataCloud data

Future work

- 1 Examples using SeaDataCloud data
- 2 Application with data API (SOCIB, NOAA OneStop)

Future work

- 1 Examples using SeaDataCloud data
- 2 Application with data API (SOCIB, NOAA OneStop)
- 3 Notebook citation See on Wednesday