Notebooks to document workflow: a possible useful component in Virtual Research Environments

Oceanography | Data analysis | Reproducibility | SeaDataCloud | ODIP | Virtual Research Environments | Python | Julia | Jupyter

0 What do we want to do?

Reproducibility

To use jupyter notebooks both as a **user guide** and as a **user interface** to describes the different steps to generate research products.

"*Interactive notebooks: Sharing the code*", Nature (2014) http://www.nature.com/news/interactive-notebooks-sharing-the-code-1.16261

1 A few definitions & acronyms...

Application Programming Interface (**API**): rules and specifications that software programs can follow to communicate with each other.

DIVAnd: a software tool designed for the interpolation of in situ measurements in *n* dimensions.

Notebook: web application for the creation and share of documents that include:

- code
- figures and animations
- text, including equations

Virtual research environment (**VRE**): online infrastructure which aims to help researchers to work collaboratively by managing

- the software tools, services and technologies,
- the access to data resources, public or private,
- the publication of the results obtained.

2 Notebooks: what exists today?

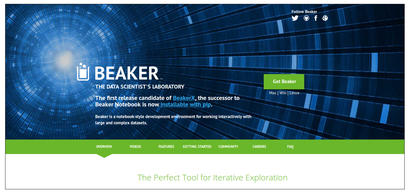

Figure 1: Beaker (http://beakernotebook.com/)
Notebook-style development environment for working interactively with large and complex datasets.

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

Figure 2: "Collaborative Calculation in the Cloud" (https://cocalc.com/)
Web-based cloud computing platform, formerly called formerly called `SageMathCloud`.

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

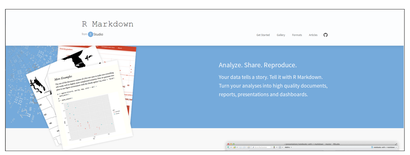

Figure 3: R-Markdown (http://rmarkdown.rstudio.com/)
Dynamic, self-contained documents with embedded chunks of code.

- ✓ *Possible to export in journal (https://github.com/rstudio/rticles) or presentation formats*
- ✓ *LaTeXtemplates to ensure journal standards*

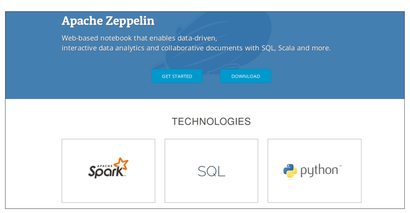

Figure 4: Apache Zeppelin (https://zeppelin.apache.org/)
Web-based notebook for data-driven, interactive and collaborative documents.
Intended for big data and large scale projects.

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

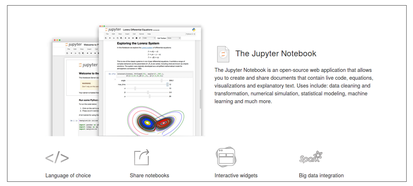

Figure 5: Jupyter (http://jupyter.org/)
Web application for the creation and sharing of notebook-type documents.
Evolved from IPython, a command shell for interactive computing (2001).

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

3 Why do we use Jupyter?

1. Open source... as the others
2. Many programming languages... as the others
3. Easy installation... as some of the others
4. A nice solution to deploy on a cloud: JupyterHub

Comparison

Tool name	R-Markdown	Jupyter	Beaker	Cocalc	Zeppelin
https://github.com/...	.../rstudio/rmarkdown	.../jupyter/notebook	.../twosigma/beakerx	.../sagemathinc/cocalc	.../apache/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, HMTL, Markdown, reST	Beaker format		JSON
Cloud deployment	–	JupyterHub	Beaker Lab (discontinued)	–	Yes

4 A few more words about jupyterhub

Multi-user Hub which { spawns / manages / proxies } multiple instances of the single-user Jupyter notebook server (https://github.com/jupyterhub/jupyterhub).

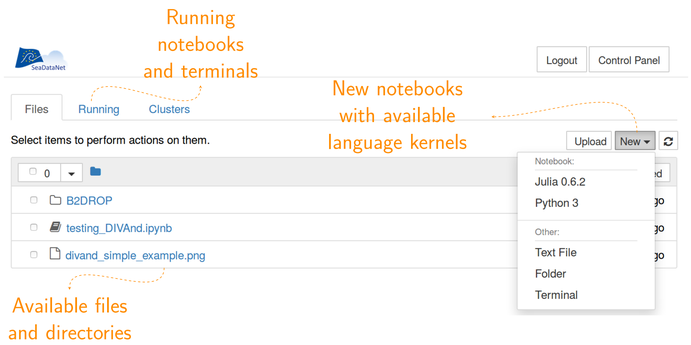

Figure 6: Test instance of jupyterhub deployed for the SeaDataCloud VRE.

Spawner: responsible for the start of the computer environment for the user, either directly on the server or on a cluster. Several spawners are available, among them DockerSpawner, which enables JupyterHub to spawn single user notebook servers in Docker containers (https://github.com/jupyterhub/dockerspawner).

5 How will we use notebooks in the VRE?

3 steps into 1

With DIVA { Read the doc / Compile and run the code / Document the execution: parameters, configuration, ... } → with DIVAnd Run the notebook!

6 Anatomy of a notebook

Notebooks contain:

1. Text cell to document the code
2. Cell codes that can be exectuted piece by piece
3. Results from the code execution
4. Figures or animations

Define the resolution. It is recommended to start with a low resolution and increase the resolution when the results are suitable. → Text cell for documenting (M↓)

```
In [9]: dx = 0.2; # longitude resolution in degrees
        dy = 0.2; # latitude resolution in degrees
        @show dx, dy;
        (dx, dy) = (0.2, 0.2)
```

→ Code cell to be executed

→ Result of the cell execution

Define bounding box spatial domain

```
In [8]: lonr = 27.:dx:42.; # the range of longitudes (start:step:end)
        latr = 40.:dy:47.; # the range of latitudes (start:step:end)
        @show lonr, latr;
        (lonr, latr) = (27.0:0.2:42.0, 40.0:0.2:47.0)
```

→ Order of execution

Figure obtained from the cell execution

```
In [23]: pcolormesh(lon, lat, transpose(salinityplot))
         colorbar()
         show()
```

→ Code cell to generate the plot

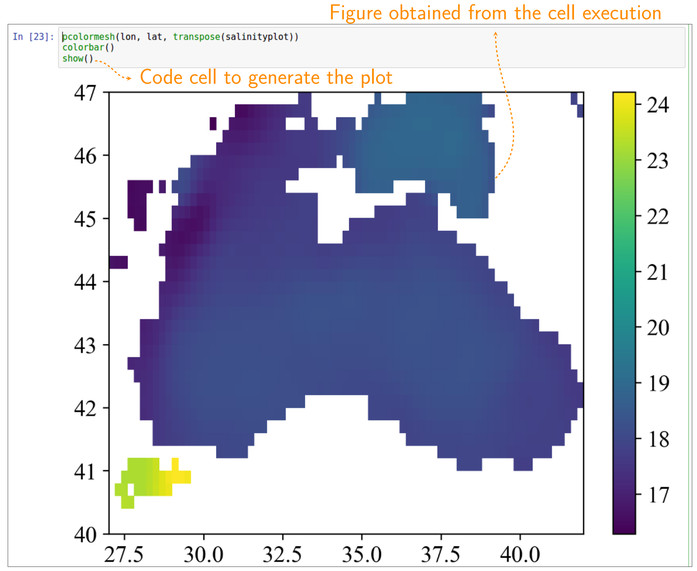

Figure 7: Examples of notebook cells.

7 Next steps

- ✓ Use the files produced by webODV as an input for DIVAnd
- ✓ Build a RESTful API to make easier the integration into the VRE
- ✓ Publish the notebooks along with the data and the products

Acknowledgements

The work presented in this poster has been developed in the frame of SeaDataCloud–*Further developing the pan-European infrastructure for marine and ocean data management*, Project ID 730960 and ODIP 2–*Extending the Ocean Data Interoperability Platform*, Project ID: 654310.

Authors: Charles Troupin, Alexander Barth, Sylvain Watelet and Jean-Marie Beckers

@GHER_ULiege | https://github.com/gher-ulg/ | http://labos.ulg.ac.be/gher/