

CARNET DE RECHERCHE SCIENTIFIQUE AVEC PYTHON

PyConFR 2019, Bordeaux @c24b DRISS

QUI PARLE ?

- c24b ^(elle)
- Python ~ [8-10] ans ^(a perdu le compte)
- Ex-ingé/corsaire de recherche ^(INRA, Mines, CNRS, BnF, PSE)
- Cheffe d'entreprise ^(DRISS)

DRISS (Digital Research In Science & Society)

Aider les chercheurs à :

DRISS (Digital Research In Science & Society)

Aider les chercheurs à :
utiliser des outils open source

DRISS (Digital Research In Science & Society)

Aider les chercheurs à :

utiliser des outils open source

faciliter leur processus

DRISS (Digital Research In Science & Society)

Aider les chercheurs à :

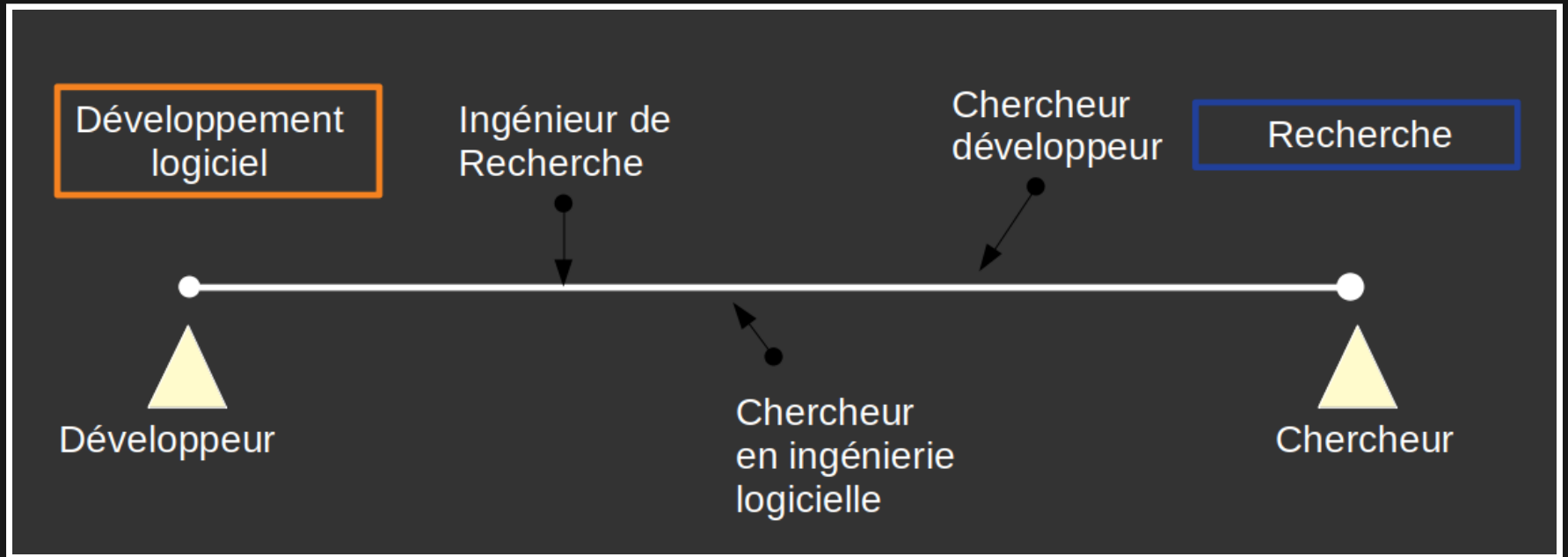
utiliser des outils open source

faciliter leur processus

faire de la recherche ouverte et reproductible

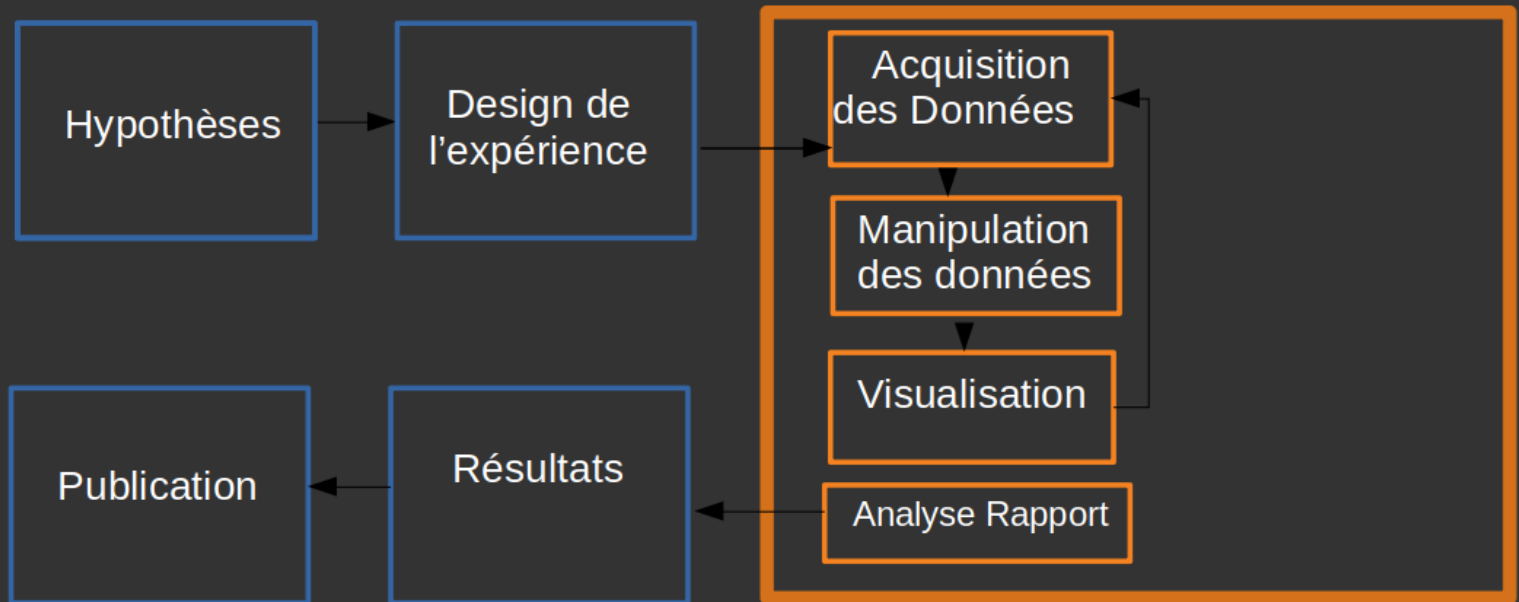
CARNET DE RECHERCHE

POURQUOI ?



Credits: @ixek

LE CYCLE MODÉLISÉ



Du côté des ingénieurs

*Outiller et documenter le travail autour
des données*

Du coté des ingénieurs

*Outiller et documenter le travail autour
des données*

- médiation, redevabilité et transparence

Du côté des ingénieurs

*Outiller et documenter le travail autour
des données*

- médiation, redevabilité et transparence
- projet maintenable, reproductible, réutilisable

Du côté des chercheurs

*Communiquer ses recherches et ses
résultats*

THE EVOLUTION OF ACADEMIA

PUBLISH



PUBLISH
OR
PERISH



PUBLISH
IN HIGH IMPACT
JOURNALS
OR
PERISH



PUBLISH
FREQUENTLY IN
HIGH IMPACT
JOURNALS
AND
MAYBE
YOU WON'T
PERISH



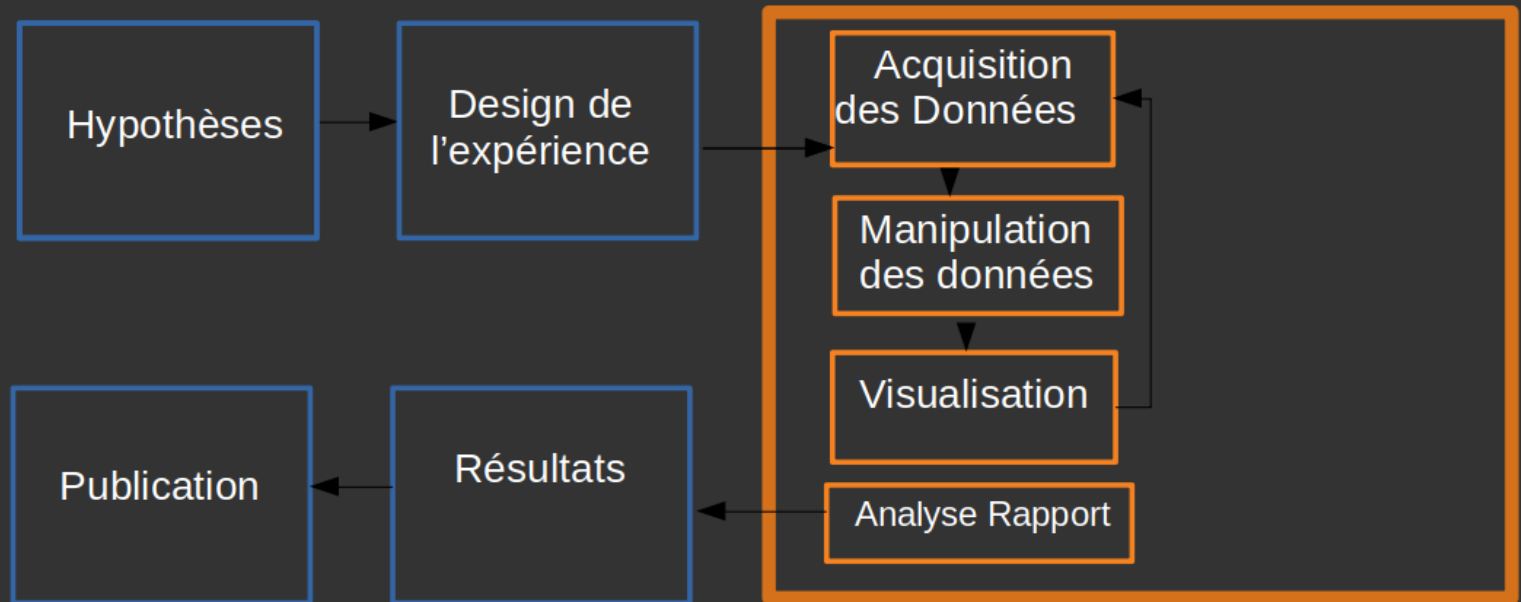
Plusieurs enjeux derrière la publication

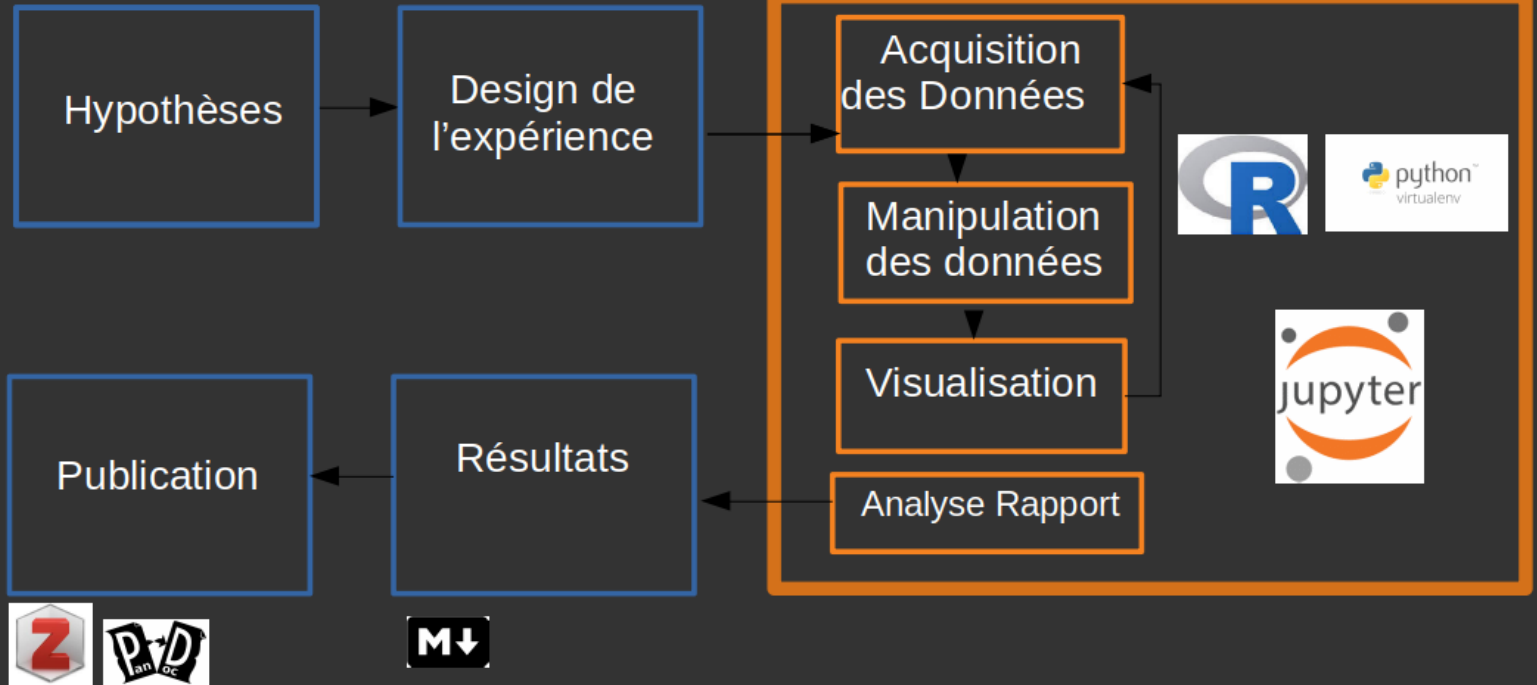
- Collaboration
- Expérimentation
- Evaluation par les pairs
- Publication (formats, contraintes, contenus)

Plusieurs enjeux derrière le développement

- Valider ses traitements
- Résultats répliquables
- Méthodes réutilisables


LE CYCLE DE RECHERCHE (MODÉLISÉ)









Gérer un projet de recherche comme du développement logiciel ?






- Tickets: hypothèses et plan d'analyse
- Branches: analyse
- Releases: resultats + rédaction de l'article à publier

 **GitLab**

Projects ▾ Groups ▾ Activity Milestones Snippets  


 ▾


Search or jump to... 


    ▾  ▾

K

kalulu_db

 Project

 Repository

 Issues


16

List


Board

Labels

Milestones

 Merge Requests

0

 Wiki

Team GIT Workflow

#18 · opened 1 week ago by Constance de Quatrebarbes

To read

updated 1 week ago

Demo Teacher Dashboard

3 of 4 tasks completed

#16 · opened 1 week ago by Constance de Quatrebarbes

~ Démo

updated 1 week ago

Jupyter Notebook load DB data

#15 · opened 1 week ago by Constance de Quatrebarbes

Admin Dashboard

API DB

updated 1 week ago

API raw endpoints

#14 · opened 1 week ago by Constance de Quatrebarbes

Admin Dashboard

API

updated 1 week ago

DB: Create tables for API Graphs

0 of 4 tasks completed

#13 · opened 1 week ago by Constance de Quatrebarbes

New Framework

DB

updated 1 week ago

Graphs

0 of 6 tasks completed

#11 · opened 1 week ago by Tristan

New Framework

updated 1 week ago

API: Endpoints for graphs

0 of 4 tasks completed

#10 · opened 1 week ago by Constance de Quatrebarbes

New Framework

Oct 25, 2019

API

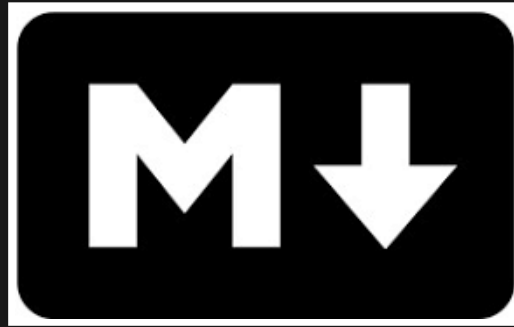
updated 1 week ago

COLLABORER ENTRE CHERCHEURS

- s'accorder sur un langage descriptif simple et minimal commun
- travailler à plusieurs chercheurs
- correction et reprise, historique et attribution

OUTILS POUR LA RÉDACTION COLLABORATIVE

markdown



scholarly
markdown

git



git for research

gitlab



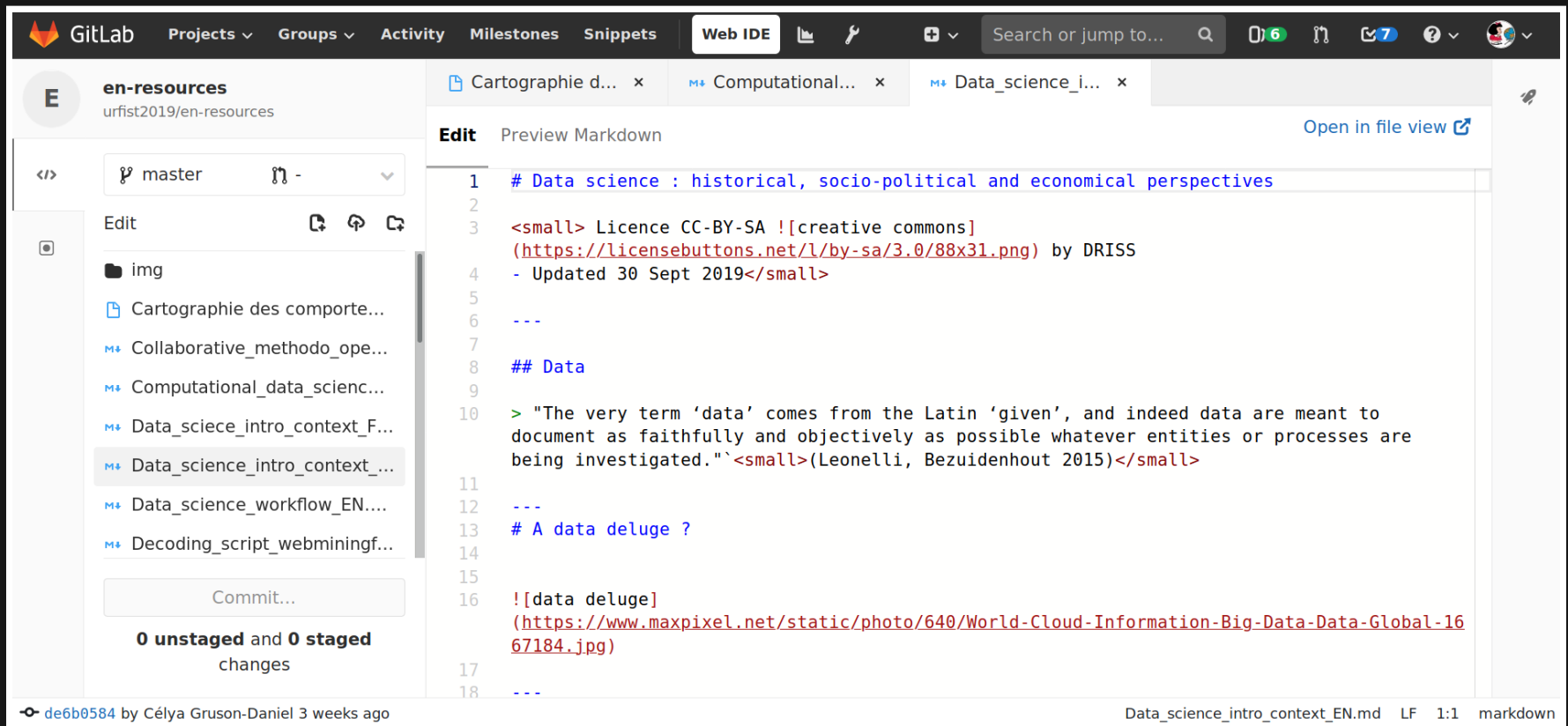
gitlab for
researcher

Ecrire en markdown des articles scientifiques:

- séparer le contenu de la présentation
- adapter le contenu à chaque format de présentation:
 - une présentation md2reveal
 - un article de blog md2html
 - un appel à proposition
 - un article de revue md2pdf
- gérer ses versions successives

Pour les débutants git

GitLab-CE offre une interface web d'édition en ligne de markdown





EXPERIMENTER

Travail entre l'ingénieur et le chercheur

Pour le dev, plusieurs préoccupations:

- ne pas se rajouter du boulot
- code compréhensible, modifiable, appropriable
- environnement répliquable deployable et testable en ligne

Pour le chercheur:

- documentation claire et compréhensible
- test et expérimentation clé en main
- graphiques
- ajout de citation et de style

Le carnet de recherche

- dans un environnement replicable
- versionné et historicisé
- partageable modifiable et exportable

CRÉER UN KERNEL SPÉCIFIQUE POUR JUPYTERLAB:

A partir du fichier requirements.txt dans le projet

```
$ virtualenv ~/.sci-env --p /usr/bin/python3  
$ source ~/.sci-env/bin/activate  
(sci-env)$ pip install -r requirements.txt
```


UN ENVIRONNEMENT OS-INDÉPENDANT



DOCUMENTER SON CODE

Choix d'utiliser `py2nb`

```
pip install py2nb
```

- Versionnable en l'état
- Documentation intégrée

```
#| # Testing ipython notebook
#| This is designed to demonstrate a simple script that conver
#| a jupyter notebook with a simple additional markdown format
#|
#| Code by default will be put into code cells
#|
#| * To make a markdown cell, prefix the comment line with wit
#| * To split a code cell, add a line beginning with '#-'

import numpy
import matplotlib.pyplot as plt
%matplotlib inline

#| Here is a markdown cell.
#| Maths is also possible:  $A=B$ 
```

TRANSFORMER SON CODE EXISTANT EN JUPYTER-NOTEBOOK

```
(sci-env)$ py2nb myscript.py
```

Crée un jupyter-notebook de test à destination du chercheur

TESTER LE CODE EN LIGNE



Turn a GitHub repo into a collection of
interactive notebooks

Have a repository full of Jupyter notebooks? With Binder, open those notebooks in an executable environment, making your code immediately reproducible by anyone, anywhere.

Build and launch a repository

GitHub repo or URL



Git branch, tag, or commit

Path to a notebook file (optional)

File ▼

launch

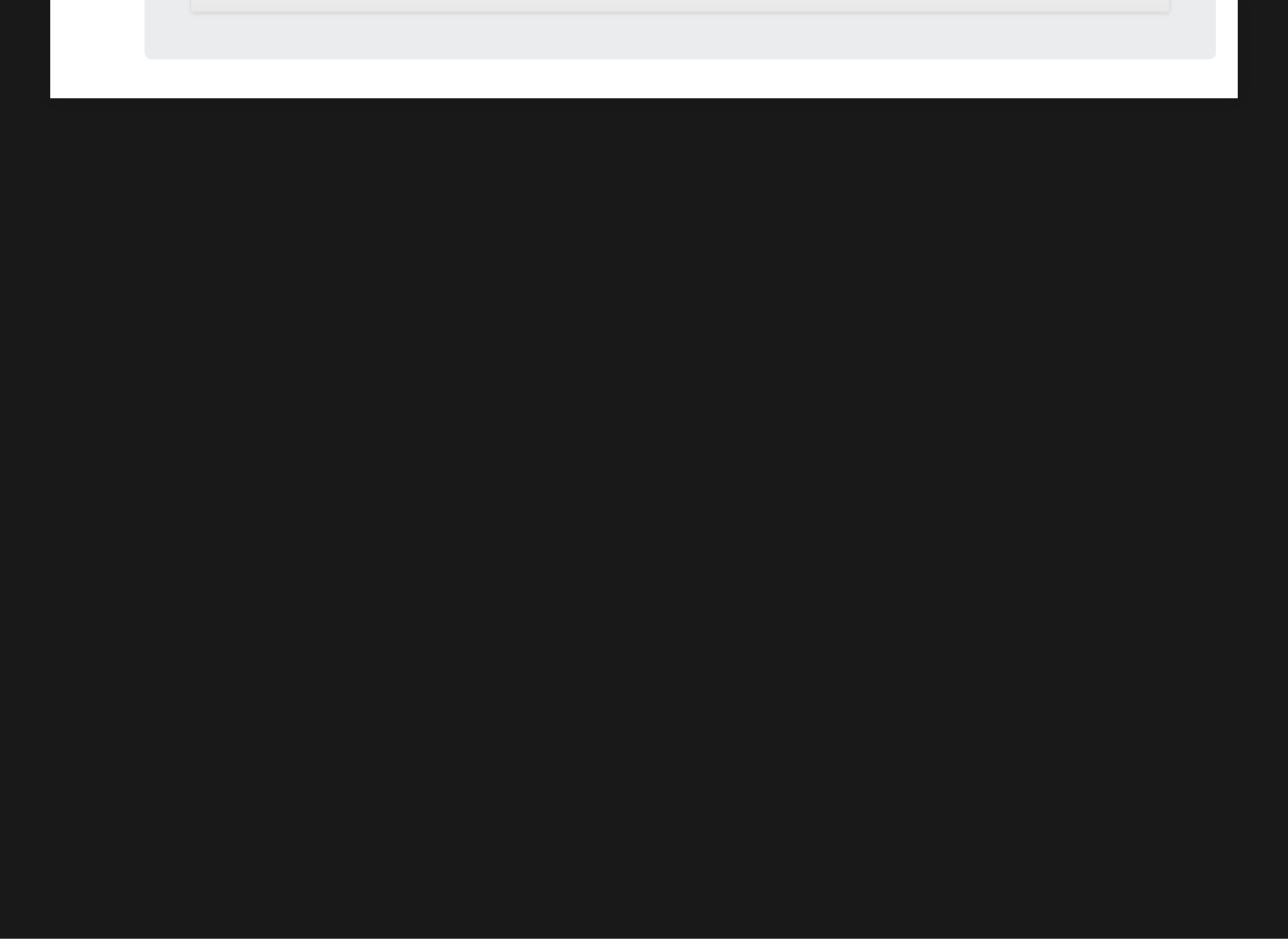
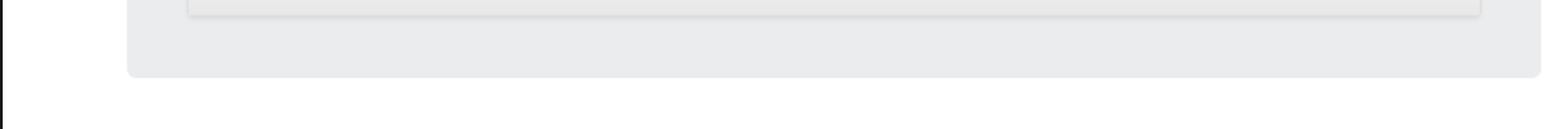
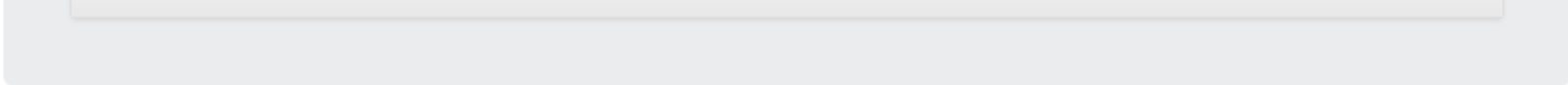
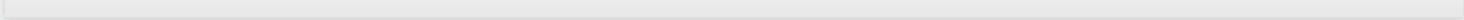


Already built!

Launching

Build logs

show



POST-TRAITEMENT

Convertir dans l'autre sens pour les modifications et versionning

- Nettoyage des cellules
- Ajout des modification dans le code source
- Ajout à la doc

```
jupyter-nb-convert --to python  
jupyter-nb-conert --to markdown
```

EXPORTER

```
jupyter-nbconvert myscript.ipynb --to pdf jupyter-  
nbconvert myscript.ipynb --to slides jupyter-nbconvert  
myscript.ipynb --to latex
```


EXPORTER AVEC CITATIONS

- On gère les sources grâce à un fichier BibTeX.
- Le style de citation peut être personnalisé grâce un fichier CSL.
- Le tout est ensuite combiné dans Pandoc pour générer un fichier de sortie (Word, PDF, LaTeX, etc.).
`pandoc pandoc-citeproc zotero-betterbibtext`

[Blog Zotero](#)

CREDITS

Largely inspired by [@ixek talk](#)