



HAL
open science

Jupyter, un carnet numérique au service des SHS

Marie-Laure Massot, Arnaud Picandet, Maxime Popineau, Émilien Schultz,
Agnès Tricoche

► **To cite this version:**

Marie-Laure Massot, Arnaud Picandet, Maxime Popineau, Émilien Schultz, Agnès Tricoche. Jupyter, un carnet numérique au service des SHS. Atelier Digit_Hum 2023. Manipuler des données en SHS : R, Python et les carnets interactifs, Oct 2023, Paris, France. hal-04216829

HAL Id: hal-04216829

<https://hal.science/hal-04216829v1>

Submitted on 27 Sep 2023

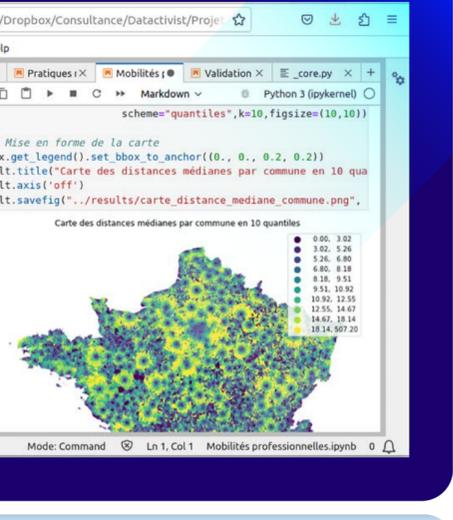
HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0
International License

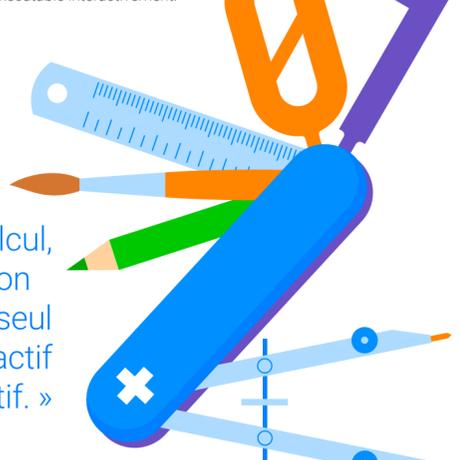
Jupyter, un carnet numérique au service des SHS



Un carnet. Oui, mais pour quel usage ?

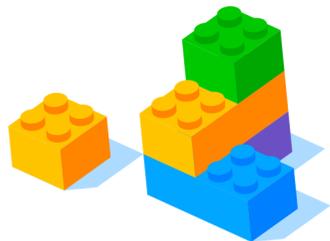
Les carnets numériques, aussi appelés notebooks ou encore calepins, sont notamment **utilisés en science des données pour l'exploration et l'analyse de ces dernières**. Ce sont des interfaces de programmation interactives permettant de combiner des sections en langage naturel et des sections en langage informatique.

Un carnet Jupyter est **un document manipulable depuis un navigateur, composé de plusieurs cellules, pouvant conjuguer du code, du texte simple ou enrichi ainsi que des visualisations graphiques**. Le code est modifiable et exécutable interactivement.



« Une solution flexible pour assembler de la narration, du calcul, de la visualisation, de l'interaction et de la programmation en un seul document, portable, interactif et collaboratif. »

Jupyter, en quatre mots



Interopérabilité

Ce type de carnet numérique supporte plus de 40 langages de programmation. Le document n'est pas juste limité au code : **il est conçu pour intégrer des images, des visualisations, des formules mathématiques, des explications textuelles et même des vidéos.**

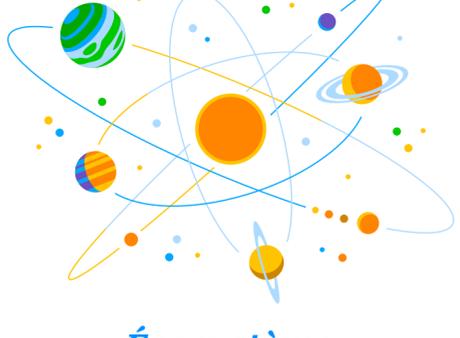
Partage

Les carnets Jupyter sont **d'une grande facilité d'appropriation et de réutilisation**. Ils peuvent être exportés au format HTML, PDF ou Markdown. Très pratiques dans le cadre de la recherche collaborative mais aussi dans un contexte d'enseignement.



Interactivité

Les utilisateurs peuvent directement exécuter le code dans le carnet numérique, voir instantanément le résultat et le modifier si besoin.



Écosystème

Les carnets Jupyter permettent d'interagir avec de nombreuses bibliothèques et profiter d'une communauté très active. Il est donc facile de trouver des ressources et de se faire aider pour résoudre des tâches spécifiques.

Les plus

« Un support unique pour penser et raconter, avec du code et des données »



Structuration à l'aide de cellules pour clarifier et scénariser le document.



Exploration fine notamment par des interactions riches grâce aux ipynbwidgets¹.



Exportation vers d'autres formats.



Simplicité du partage des données et de leur traitement.



Intégration naturelle avec des forges comme GitHub ou GitLab pour faciliter la traçabilité, la robustesse, le partage, la publication et l'archivage.



Permet le travail collaboratif à partir d'un même outil.



Forte maléabilité. L'utilisateur modèle l'outil selon ses besoins.



Interaction avec un grand nombre de langages de programmation.



Exploitation multi-plateforme (navigateurs et appareils).



100% libre.

Les moins

« Un environnement complexe »



La connaissance d'un langage informatique (Python ou autre) est fortement conseillée.



une philosophie qui n'est pas nécessairement intuitive dans l'exécution.



Difficulté de choisir quelles solutions exploiter (plugins, extensions...), face à un outil très personnalisable.



Le code peut être exécuté de manière non linéaire, provoquant des « états cachés ».



Faible pratique des tests de vérification des bugs.



Favorise des pratiques discutables des fichiers, version et modularisation du code...).



Pas de règles homogènes sur l'écriture des carnets Jupyter.



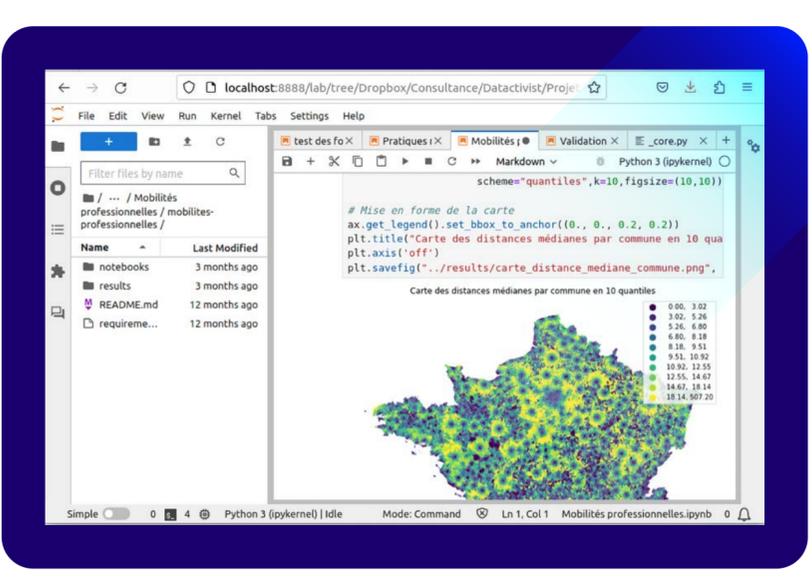
Possibles problèmes de reproductibilité (sur un million de carnets numériques sur GitHub, seulement 24% ont pu être exécutés sans erreur, 4% permettent de reproduire les résultats).

1. Les ipynbwidgets sont des éléments conçus pour les carnets Jupyter. Ils permettent d'ajouter des composants interactifs tels que des boutons, des curseurs, des listes déroulantes, des graphiques aux carnets Jupyter. Ces widgets facilitent l'interaction utilisateur en offrant des fonctionnalités interactives et en permettant la manipulation des données en temps réel. Ils sont très utilisés dans le cadre de l'enseignement et ils permettent aussi à des personnes qui ne savent pas programmer d'interagir avec le carnet numérique.

2. Il est possible d'exécuter le code de manière non linéaire et donc de provoquer des « hidden states ». Ces derniers peuvent entraîner la création de variables ou de résultats intermédiaires qui ne sont pas immédiatement visibles dans le carnet Jupyter et peuvent provoquer des changements dans les résultats si les cellules sont exécutées dans un ordre différent.

Naissance d'un projet ouvert

En 2014, le chercheur en physique Fernando Pérez annonce **un projet communautaire** issu de son environnement interactif Python facilitant la programmation pour les scientifiques : le projet Jupyter. Ce dernier se développe **autour d'un écosystème de standards, de logiciels libres et de services pour le calcul interactif** : Jupyter notebook, JupyterLab, JupyterHub, etc. Grâce à son interopérabilité et à sa communauté, Jupyter s'impose comme le format le plus répandu et adapté au traitement des données. **Jupyter a reçu, en 2017, le prix ACM Software System Award.**



Un outil plébiscité

Ces dernières années, les carnets numériques ont gagné en popularité dans le monde de l'industrie, de la recherche ou de l'éducation.

En France, c'est actuellement **l'un des outils les plus exploités par les enseignants en informatique en raison de sa facilité pédagogique**. Permettant aux élèves de progresser étape par étape, les cellules s'exécutent au fur et à mesure au sein d'une narration cohérente. **Le projet Capytale**, par exemple, est un service web proposé par l'Académie de Paris utilisant les carnets numériques pour créer et partager des activités pédagogiques de codage (500 000 utilisateurs).

Encore peu « quantités » dans le domaine des SHS, ce support est pourtant en mesure de faciliter la pratique computationnelle des chercheurs, en favorisant un standard permettant d'homogénéiser les réflexions intégrant du code et de partager facilement leur recherche. **Les carnets Jupyter peuvent, grâce à leur caractère versatile, faciliter une approche interdisciplinaire et être utilisés autant pour la recherche que pour l'enseignement.**

Le saviez-vous ?

Le nom de Jupyter fait référence aux trois premiers langages de programmation pris en charge : **Julia, Python et R.**

Aller plus loin

- **Jupyter** (site officiel, nombreuses extensions, partage de code) : jupyter.org/
- **Anaconda** (environnement pour ouvrir et créer des carnets Jupyter, avec de nombreux services comme des IDE (DataSpell, Spyder) et d'autres logiciels comme JupyterLab, Jupyter Notebook, RStudio) : www.anaconda.com/download
- **Programming Historian** (leçons en ligne en quatre langues dont le français, pour aider les chercheurs en SHS à acquérir des compétences en programmation et apprendre à créer et exécuter un carnet Jupyter) : programminghistorian.org/

Cette fiche pratique vous est proposée dans le cadre de l'Atelier Digit_Hum « Manipuler des données en SHS : R, Python et les carnets interactifs ». ENS, 12 octobre 2023

