

Callisto, un démonstrateur JupyterLab pour les sciences humaines et sociales

Stéphane Pouyllau (0000-0002-9619-1002)¹, Nicolas Sauret (0000-0001-7516-3427)², and Mélanie Bunel (0000-0002-7314-3892)¹

¹CNRS, Huma-Num

²Université de Paris Nanterre, Huma-Num

sept. 2021

Table des matières

1	Introduction	1
2	Qu’est-ce que JupyterLab ?	2
3	Un établi pour faire de la recherche ?	2
4	Objectifs de notre preuve de concept : Jupyter dans l’écosystème de la TGIR Huma-num	3
4.1	Démarche de conception	5
4.2	Ressources	5

1 Introduction

Ce document reprend, sous la forme d’un document d’étape, l’article évolutif publié¹ sur le *Log* du Huma-Num Lab (HN Lab) depuis mai 2021. Il présente la preuve de concept « Callisto » mise en œuvre par le HN Lab au sein de la TGIR Huma-Num.

Le démonstrateur Callisto, du nom de l’une des lunes de Jupiter, est une preuve de concept proposant la mise en œuvre du logiciel en ligne Jupyter et plus particulièrement sa version JupyterLab pour des pratiques de recherche en SHS et en lien avec l’écosystème de la TGIR Huma-Num. Ainsi, Callisto repose sur plusieurs briques disponibles au sein d’Huma-Num et fonctionne ainsi comme « hub » les interconnectant. Il s’agit de :

1. Voir sur Log du HN Lab : <https://hnlab.huma-num.fr/blog/2021/05/26/callisto-un-demonstrateur-jupyter/>

- JupyterLab
- JupyterHub
- HumanID (de dispositif de la TGIR Huma-Num pour l'authentification des utilisateurs)
- Gitlab (forge de dépôt permettant le versionnage de documents)

2 Qu'est-ce que JupyterLab ?

JupyterLab et son dispositif d'authentification sur serveur JupyterHub, sont des logiciels en ligne qui permettent d'écrire des documents (des *notebooks* ou carnets) contenant à la fois du texte (à lire, écrit en markdown) et du code informatique exécutable au sein du carnet. Ce code peut être en Python, R, Perl, Julia, etc. suivant les configurations installées dans le JupyterLab.

Les carnets Jupyter offrent aux chercheurs la possibilité d'étayer dans un même document : un propos, une démonstration en publiant à la fois les données, les traitements opérés sur les données, le ou les programme(s) exécuté(s), les résultats (données interprétées par exemple) et l'argumentaire/conclusion sous la forme d'un texte académique.

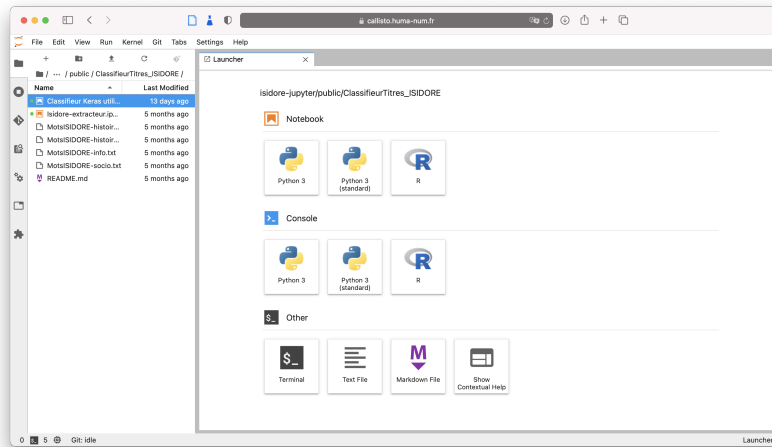


FIGURE 1 – Environnement de travail sous JupyterLab

3 Un établi pour faire de la recherche ?

Très utilisé pour l'enseignement et l'apprentissage du codage, JupyterLab, quand il est placé dans un écosystème permettant le travail collaboratif et la publication en ligne, offre un dispositif complet permettant de travailler les données, de les analyser et de les publier. L'exposition, didactique et argumentée,

du ou des processus de recherche offre aux chercheur·e·s la possibilité de partager leur savoir-faire et de présenter ensemble leurs méthodes de travail et les résultats sous la forme de publication réutilisable. Il s'agit de favoriser d'une part la réflexivité et la discussion partagée des méthodes, d'autre part d'en assurer la reproductibilité, conformément aux principes FAIR. Une fois stabilisé, un carnet peut être publié en ligne sous une forme statique, citable et versionnée. L'exemple ci-dessous présente un carnet Jupyter déposé dans NAKALA qui n'offre pas encore de visionneuse pour ce type de contenu.

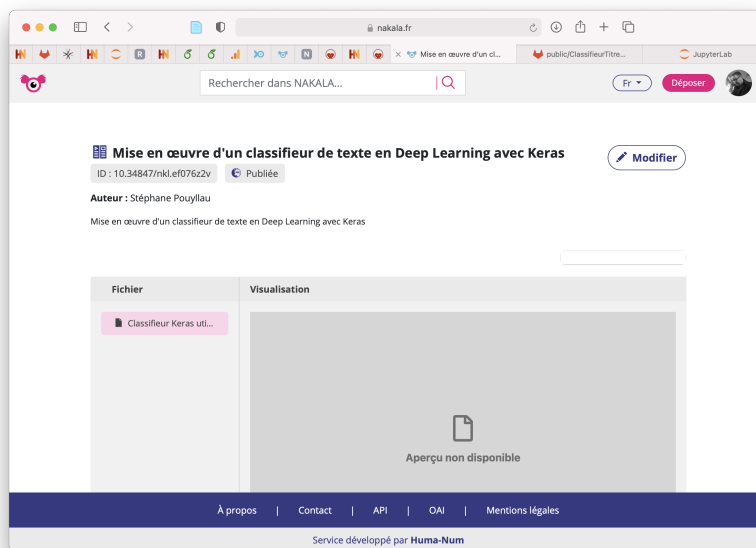


FIGURE 2 – Publication dans NAKALA d'un carnet Jupyter, la visualisation/exécution reste à construire

4 Objectifs de notre preuve de concept : Jupyter dans l'écosystème de la TGIR Huma-num

Il s'agit pour le HN Lab de faire une première proposition aux chercheur·e·s souhaitant utiliser ce type de dispositifs afin de définir avec eux ce que serait un environnement JupyterLab interconnecté aux autres briques de la TGIR Huma-Num : Gitlab, HumanID (ci-dessous), NAKALA, et à terme sans doute du calcul sur CPU et/ou GPU (en association avec des centres de calcul).

Dans un premier temps, des interconnexions de Callisto ont été installées pour cette preuve de concept avec HumanID via le protocole Oauth2, ainsi qu'avec l'instance Huma-Num de la plateforme GitLab mise à disposition des programmes de recherche. Il est ainsi possible, pour une communauté de cher-

cheur-e-s et d'ingénieur-e-s, de synchroniser et de versionner les carnets Jupyter produits dans Callisto dans un répertoire GitLab.

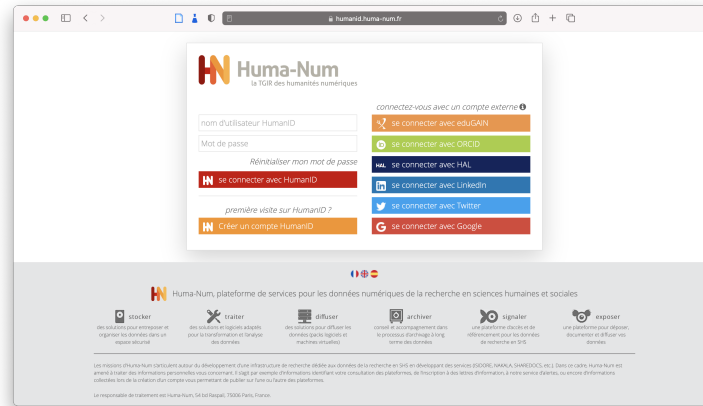


FIGURE 3 – HumanID : l'interface commune aux services Web de la TGIR Huma-Num

La publication d'un carnet JupyterLab dans un environnement Gitlab permet d'un côté une gestion versionnée et itérative (collective), ouvrant de nouvelles opportunités en terme de collaboration (co-développement, tickets et attribution de tâches). D'un autre côté, un répertoire Gitlab contenant des carnets Jupyter peut être couplé à des outils de traitements et de calcul tel que MyBinder, permettant au carnet d'être publiquement accessible dans une version exécutable.

Dans un second temps, une interconnexion avec NAKALA pourrait être imaginée facilitant le dépôt académique, la publication dans une revue ou un ouvrage, ou encore la préservation à long terme. L'inclusion dans NAKALA d'un service d'exécution de carnets, tel que MyBinder serait l'ultime aboutissement d'un tel dispositif. Cette configuration d'écriture collective, exécutable et réexécutable, peut être schématisée tel que ci-dessous :

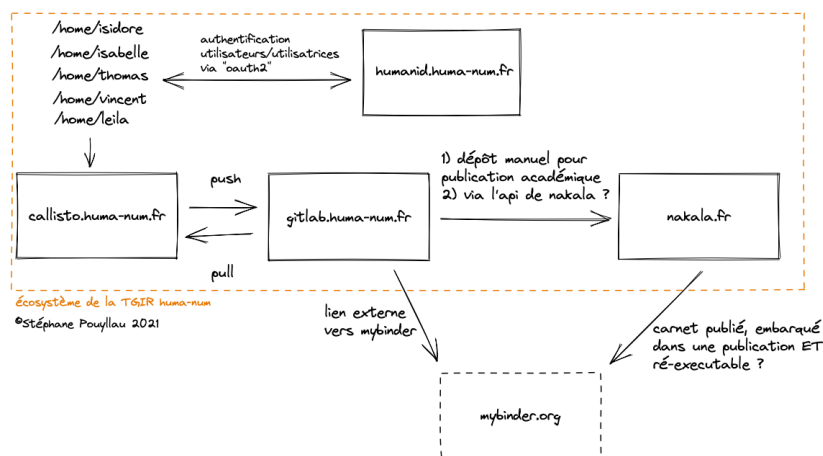


FIGURE 4 – Schéma conceptuel de l’inclusion de Callisto dans l’écosystème de la TGIR Huma-Num

4.1 Démarche de conception

Dans la démarche du HN Lab, il s’agit aussi de créer une communauté d’utilisateurs afin d’échanger sur les besoins².

Enfin, outre la preuve de concept évolutive, il s’agit de proposer une expérimentation d’instance JupyterLab conçue spécifiquement pour les besoins SHS. La preuve de concept Callisto a ainsi vocation à intégrer des éléments pertinents sur le plan scientifique pour la communauté SHS :

- Des modèles de carnets par disciplines (ou par familles de projets)³,
- Des bibliothèques correspondant aux pratiques et aux méthodes de traitements scientifiques en SHS, et facilitant l’interconnexion aux services de la TGIR Huma-Num, - Des environnements Python (ou autres) dédiés à une pratique SHS particulière et en lien avec les axes de l’HN Lab⁴.

4.2 Ressources

- Preuve de concept Callisto : <https://callisto.huma-num.fr>
- Documentation et notes d’installation de Callisto par l’équipe du HN Lab (sur le serveur gitlab d’Huma-Num, inscription HumanID est nécessaire)
- Gestion des kernel dans Callisto

2. Un groupe de travail sur Callisto et l’utilisation de Jupyter pour la recherche s’est constitué autour d’une liste de diffusion : <https://listes.huma-num.fr/wws/info/humanum-lab-gt-jupyter>.

3. Voir le projet MODOAP (Labex « les passés dans le présent »). Porté par Julien Schuh, ce projet propose une boîte-à-outil de carnets génériques à prendre en main et à adapter selon les projets.

4. Lire les axes du HN Lab sur le site Web d’Huma-Num : <https://www.huma-num.fr/hnlab/>

Table des figures

1	Environnement de travail sous JupyterLab	2
2	Publication dans NAKALA d'un carnet Jupyter, la visualisation/exécution reste à construire	3
3	HumanID : l'interface commune aux services Web de la TGIR Huma-Num	4
4	Schéma conceptuel de l'inclusion de Callisto dans l'écosystème de la TGIR Huma-Num	5