



Atelier Digit_Hum 2023

Manipuler des données en SHS :
R, Python et les carnets interactifs

Judi 12 octobre 2023 – 9h-17h

École normale supérieure | Salle Dussane



En 2023, la thématique « Programmer en SHS » a été retenue.

Il s'agit de dresser un panorama des possibilités offertes par les langages R ou Python et des environnements de programmation lettrée (literate programming) interactifs (Notebooks) pour se saisir des données et faciliter la reproductibilité des travaux de recherche. Les écosystèmes de programmation qui se sont développés dans le monde scientifique permettent de relier les données et d'accompagner l'utilisateur sur une variété de tâches, tels que les traitements statistiques, l'analyse de grands corpus historiques, littéraires, philosophiques, ethnographiques (...), la collecte de données sur les réseaux sociaux ou encore la réalisation d'interfaces interactives. Représentatifs de ces outils initiés dans le monde de la recherche, les carnets numériques bénéficient d'une communauté d'utilisateurs croissante et, au-delà de leurs différences, partagent de nombreuses caractéristiques communes. Cette journée d'étude à destination d'un public non spécialiste propose d'ouvrir des perspectives, donner la possibilité de se familiariser avec les principales notions pour se repérer dans ces univers et partager des exemples de traitements possibles de données en SHS, autour de projets en cours.

PROGRAMME

9h | Accueil autour d'un café

9h15 | Ouverture

Marc Bui (EPHE, AOROC, CNRS, ENS PSL - Université Paris 8)

MATINÉE : R, PYTHON ET LES USAGES INTERACTIFS DE LA PROGRAMMATION

Mots d'introduction par Marie-Laure Massot (CAPHÉS, CNRS, ENS PSL) & Agnès Tricoche (AOROC, CNRS, ENS PSL) : les chiffres marquants de l'enquête Digit_Hum issus du sondage réalisé en avril 2023.

9h30 | Quelle place pour la programmation scientifique en sciences humaines et sociales ? Petit panorama des usages de Python

Émilien Schultz (SESSTIM) - **Matthias Bussonnier** (Quansight)

Quand on parle de la programmation informatique en recherche, trois questions se posent : À quoi cela peut servir ? Est-ce que cela va m'être utile ? Comment s'y mettre ? A partir d'un retour d'expérience à la fois de pratique et de formation sur l'utilisation de Python en SHS, nous proposons de faire un petit état des lieux illustré de différents usages, allant de l'automatisation de petites tâches répétitives au développement d'outils en passant par la possibilité d'articles exécutables.

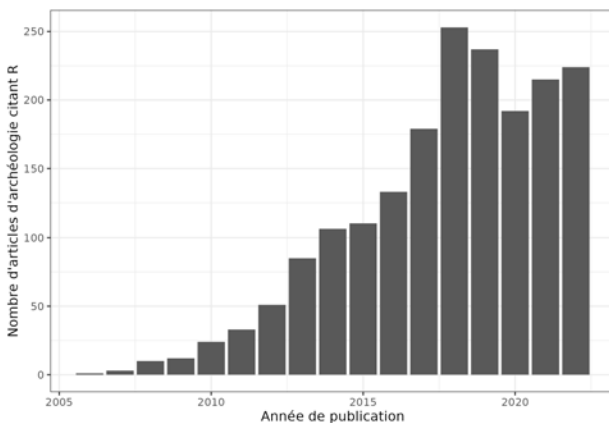
Nous défendons l'importance de constituer la programmation scientifique comme une composante spécifique de nos pratiques scientifiques sur les données, différente de celle des ingénieurs en informatique, qui permet non seulement d'introduire de la modularité et de l'ouverture à nos traitements, mais aussi de favoriser leur conceptualisation. Quant à la question si tout le monde devrait programmer, et en quel langage... discutons-en !



10h15 | R en archéologie : la révolution n'aura pas lieu

Nicolas Frerebeau (Archéosciences Bordeaux, CNRS, Univ. Bordeaux Montaigne)

Au regard des liens étroits qu'elle entretient avec les mathématiques depuis la fin du XIXe siècle, et avec l'informatique depuis le dernier quart du XXe siècle, l'archéologie pourrait apparaître comme un espace privilégié pour l'adoption de nouvelles pratiques numériques. Au début de la décennie, Schmidt et Marwick (2020) s'interrogeaient ainsi sur le rôle des méthodes numériques dans le développement de l'archéologie, notamment en termes de transparence et de reproductibilité de la recherche. Loin d'une révolution, les auteurs soulignent la lente transformation des pratiques, induite par l'usage des langages de script et de la programmation lettrée. Partant d'un rapide tour d'horizon du contexte français, en termes d'enseignement et de recherche, cette contribution propose un retour d'expérience sur le développement de packages R en archéologie. En s'appuyant sur le projet tesselle (<http://www.tesselle.org>), il s'agira de discuter les usages, les facteurs limitants et les évolutions possibles.



D'après Schmidt et Marwick 2020.
Source : Web of science (dernier accès le 26/06/2023).

10h45 | Pause

entremêlant narration, interaction, visualisation, calcul et programmation. Dans cet exposé, nous introduirons les carnets numériques et ferons un tour d'horizon de l'écosystème Jupyter, avec ses extensions et services permettant par exemple de réaliser et publier des livres, des diapos, des tableaux de bord, des mini-applications web interactives, etc.

14h15 | *Bonnes pratiques autour des carnets interactifs*

Maxime Popineau (CESR, Université de Tours) et **Emilien Schultz** (SESSTIM), résultat du stage Digit_Hum 2023

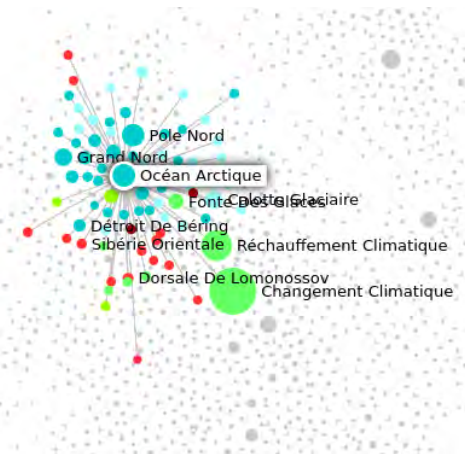
Que ce soit pour l'enseignement, l'exploration des données ou la rédaction d'articles exécutoires, les carnets Jupyter deviennent de plus en plus visibles dans la recherche notamment en SHS. Combinant du code, des résultats et du texte, ils sont pensés comme un support adapté à l'interaction avec les données et leur mise en récit. Cependant, s'ils provoquent l'enthousiasme de beaucoup en raison de leur flexibilité, ils amènent aussi de nouvelles contraintes. Cette communication présente un panorama des forces et des limites de l'utilisation des carnets Jupyter dans le cadre des SHS à travers les résultats d'un stage Digit_Hum réalisé d'avril à mai 2023 qui a conduit à une synthèse sur l'utilisation des carnets Jupyter en SHS et sa mise en oeuvre dans une étude des usages de Jupyter sur HAL.



14h45 | *Explorer des réseaux dans un carnet interactif avec ipysigma*

Béatrice Mazoyer et **Guillaume Plique** (médialab, Sciences Po)

Les notebooks ou carnets interactifs sont désormais un incontournable du travail d'exploration des données. Un de leurs principaux avantages est de permettre d'afficher facilement des schémas et des diagrammes. Toutefois, il reste encore difficile d'y visualiser des données en réseaux, ce qui force les chercheuses et chercheurs en sciences sociales à faire des aller-retours entre leurs carnets et des logiciels tels que Gephi ou Pajek. Ipysigma a été conçu afin de pallier ce manque : il s'agit d'un module Python capable d'afficher des réseaux



interactifs, et de modifier simplement des variables visuelles telles que la taille et la couleur des nœuds, ou l'épaisseur des arcs. L'outil permet aussi de montrer simultanément différentes vues du même réseau afin de faciliter les comparaisons. Notre intervention présentera différents exemples d'analyse visuelle de réseaux en sciences sociales et reviendra sur les raisons de la création d'ipysigma.

15h15 | Pause

15h30 | *Garder trace de son travail et le partager en traitement automatique des langues*

Caroline Koudoro-Parfait (ObTIC, STIH, SCAI, Sorbonne Université)

La constitution des corpus de textes anciens et leur exploitation informatique passe par l'élaboration d'une chaîne de traitement adéquate qui réunit plusieurs communautés de chercheurs : des chercheurs en traitement automatique des

langues (TAL), en humanités numériques (HN) et en sciences humaines et sociales (SHS).

La structure du notebook, qui supporte l'inclusion de paragraphes explicatifs, d'images mais aussi de vidéos autour du code, en fait un outil didactique permettant aux chercheurs des communautés TAL et HN d'une part de commenter les phases du développement du programme informatique, et d'autre part de l'expliquer et de mieux communiquer avec des chercheurs en SHS intéressés mais dont ce n'est pas le domaine de compétence.

Je vous propose, par le prisme de mon travail de thèse, qui porte sur l'évaluation des systèmes de reconnaissance d'entités nommées sur des textes anciens, d'observer la manière dont les notebooks servent en effet d'interface entre ces différentes communautés.

16h | *Le groupe de travail Notebooks d'Huma-Num : Présentation de l'initiative, retour d'expérience et perspectives du GT Notebook (2021-2024) par des membres du GT Notebook*

Membres du GT Notebook (https://gt-notebook.gitpages.huma-num.fr/site_quarto/)

Le groupe de travail Notebook s'est constitué autour d'un projet d'atelier documenté et reproductible réalisé le 3 décembre 2021 à l'IMEC. Il s'agit de présenter les enjeux et problématiques du Notebook, pris comme un objet complexe au sein du cycle des données de la recherche.

Le collectif émerge d'un appel à contribution lancé dans le cadre des Rencontres Huma-Num 2021 par le HN Lab sur la question de l'écriture exécutable, réinterprétée au sein du collectif comme problématique associée au literate programming.

En 2023, il réalise un cycle de 5 webinaires, accessibles sur la chaîne Canal U du GT (<https://www.ca->





nal-u.tv/chaines/gt-notebook/webinaire-20222023), traitant du Notebook Callisto, des questions éditoriales avec Quarto, de revues disciplinaires fondées sur des Notebooks ou de plateformes de flux permettant de les "jouer", de l'interactivité et de l'esthétique de ces "carnets numériques".

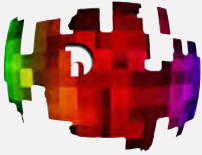
La démarche collective du GT est pluridisciplinaire, et engage un dialogue réflexif tant sur les usages que sur la technicité de ces objets numériques.

16h30 | Retours sur la journée et discussions

Avec **Emile Gayoso** (Guichet d'Assistance aux Projets Numériques, Bibliothèque Ulm-Lettres ENS PSL), président de séance

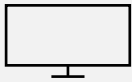
Organisation :

Marie-Laure Massot (CAPHÉS, UAR 3610, CNRS-ENS)
Agnès Tricoche (AOROC, UMR 8546, CNRS-ENS-EPHE)
Emilien Schultz (SESSTIM)



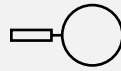
DIGIT_HUM

Humanités numériques
entre savoirs et savoir-faire



Vidéos

des ateliers
Digit_Hum



Enquête

sur les humanités
numériques à l'ENS



Portraits

d'acteurs en humanités
numériques



Ressources

pour s'informer,
se former, s'orienter

et un **portfolio** d'artistes dessinant les humanités numériques

Retrouvez toute notre activité sur
dighum.huma-num.fr

Digit_Hum est une initiative visant à échanger sur les différents usages qui sont faits des humanités numériques et à structurer ces nouvelles pratiques à l'échelle de l'École normale supérieure et de l'université Paris Sciences & Lettres.

Cette initiative est portée par le CAPHÉS (Centre d'Archives en Philosophie, Histoire et Édition des Sciences),
et AOROC (Archéologie et Philologie d'Orient et d'Occident).

Elle bénéficie du soutien de l'EUR Translitteræ et de l'Observatoire des humanités numériques de l'ENS.