

L'activité du HN Lab en 2022

Stéphane POUYLLAU, ingénieur de recherche hors classe CNRS, responsable du HN Lab.

1. Le HN Lab : une force de propositions pour Huma-Num

Initialement, le Huma-Num Lab (HN Lab) est le pôle d'accueil des chercheurs, des activités de recherche et développement, d'innovation et de valorisation d'Huma-Num. Le HN Lab s'est doté d'un espace et de moyens pour accueillir des chercheurs en résidence, dont la principale fonction est de permettre, par l'accueil et l'accompagnement de projets menés ensemble, d'alimenter la définition, l'évolution et la création de nouveaux services numériques génériques au sein de l'infrastructure Huma-Num.

Après deux ans de fonctionnement, un premier bilan permet de nuancer cette ambition et d'en réorienter les objectifs. À l'origine le HN Lab était composé de 3 personnels (1 titulaire et 2 contractuels). Cette équipe, très limitée en nombre afin d'être agile, s'est principalement appuyée sur plusieurs collaborations en dehors d'Huma-Num afin de mener son programme de travail fixé par le comité de pilotage et le conseil scientifique. Outre des prestations, importantes pour aller chercher de l'expertise qu'Huma-Num n'a pas en interne, le HN Lab fonctionne aussi avec la mise en place et l'animation de groupes de travail (GT) et des coopérations bilatérales issues d'initiatives extérieures (tel que les programmes du Labex « Les passés dans le présent », mais aussi au Canada avec l'entente avec Centre de recherche interdisciplinaire sur les humanités numériques de Université de Montréal).

En 2022, l'équipe a été réduite à une personne en raison du départ vers le privé et d'autres emplois de deux des trois personnels du HN Lab. Ces personnels n'ont pas été remplacés à ce jour. Depuis sept. 2022, le programme a été adapté à cette situation. Par ailleurs, le HN Lab a été engagé, quasiment à temps complet pour le responsable du HN Lab, entre juin et décembre 2022 dans le suivi et l'exploitation d'ISIDORE (ainsi que dans la rédaction d'une importante documentation technique disponible sur le GitLab interne) en raison du départ du responsable des développements informatiques d'Huma-Num. Ce travail s'est doublé d'un important travail de relecture documentaire et technique des sources de données (près de 10000) entre mai et sept. 2022 afin de permettre à ISIDORE d'être opérationnel dès sept. 2022. Cependant de multiples autres facteurs sur la mise en exploitation de la nouvelle chaîne et l'arrêt de l'ancienne n'ont pas permis cette mise en fonctionnement). Cette réduction de voilure et la surcharge de travail dû à l'exploitation d'ISIDORE, ont été un facteur de ralentissement des activités de recherche du HN Lab.

Malgré cela, plusieurs actions ont été menées sur les missions du HN Lab et montre l'utilité d'un « lab » au sein d'une infrastructure de recherche afin de défricher les problématiques avenir des outils de la recherche en SHS. En 2022, le HN Lab s'est engagé à la fois pour des publications qui sont en cours, dont un ouvrage au sein des Presses universitaires de Nanterre et une communication lors du colloque final du programme H2020 Triple mais aussi pour apporter une expertise à l'extérieur d'Huma-Num sur l'évolution que pourrait prendre les SHS à l'ère numérique. Les propositions faites par le HN Lab ont intéressé ainsi plusieurs communautés de recherche (voir ci-dessous la plateforme « Callisto » par exemple). En revanche, force est de constater que la mise en exploitation sous la forme de services au sein d'Huma-Num des propositions faites dans les publications du HN Lab ne fonctionne pas (hormis les travaux menés pour ISIDORE dans le cadre du programme HNSO, voir ci-dessous également). Une analyse plus fine des blocages devrait être faite afin d'en identifier les causes et surtout pour construire des solutions permettant de les dépasser.

2. Bilan 2022

Comme en 2021, les activités se déploient selon 4 axes de recherche¹ :

a. Écriture scientifique

Contenus : comment se transforme l'écriture scientifique ? Comment intégrer dans l'écriture les jeux de données, leurs traitements et l'outillage.

Analyse des corpus de revues dans le cadre du projet Revue 2.0

Ce programme, débuté en 2021, s'est achevé en oct. 2022 avec la restitution aux équipes de recherche des revues impliquées des résultats du programme. De cette restitution, un bilan a été publié sur le Log du HN Lab : <https://hnlab.huma-num.fr/blog/2022/01/20/experimentation-revue20/>

Rappel du bilan : l'expérimentation a offert de nombreux résultats à plusieurs niveaux et plusieurs échelles.

En premier lieu, le projet a permis de construire une chaîne complète et très détaillée de traitement et d'analyse des corpus étudiés, en expérimentant plusieurs algorithmes de prédictions, et en explorant de manière poussée des voies d'amélioration de ces algorithmes. La chaîne est composée de différents modules sous forme de notebooks Python, articulant ainsi l'algorithmique et sa documentation.

En second lieu, le projet a mis en production une infrastructure complète, déployée sous docker et comprenant différentes plateformes pour le traitement des données (Jupyter Hub/Lab), l'annotation de corpus (Doccano), la visualisation des données analysées (Kibana basé sur Elastic Search) et la gestion des flux de données (Apache Airflow). Un espace de stockage interconnecté à l'infrastructure (via RCLONE) a permis la gestion des données et des données intermédiaires en temps réel. À cette infrastructure s'ajoute un mini-site Web à destination des chercheur·e·s mais aussi éditeurs et éditrices ayant participé aux évaluations « métiers » des prédictions de l'IA.

Ce programme alimente les réflexions du HN Lab sur l'avenir d'ISIDORE et a été en parti précisé dans le cadre du travail du HN Lab sur le programme Huma-Num Science Ouverte (HNSO, voir ci-dessous). Cependant, l'infrastructure ne possède pas actuellement les équipes pour développer un tel moteur. Dans ce contexte, plusieurs publications du HN Lab dessine déjà les contours de l'avenir d'un moteur de recherche probabiliste pour les SHS (ces travaux sont disponibles sur https://zenodo.org/communities/hn_lab/).

Par ailleurs, le recrutement prévu d'un *Data-Scientist* au sein du HN Lab en mars 2023 devrait permettre de reprendre ce travail de recherche dans sa dimension technologique.

Co-développement et urbanisation de l'éditeur de texte Stylo

L'offre des services d'Huma-Num s'est enrichie depuis le 1er octobre 2021 du nouvel outil Stylo, un éditeur de texte scientifique conçu spécialement pour les SHS. Stylo est connecté à HumanID et accessible à tous les utilisateurs et utilisatrices ayant un compte à Huma-Num. Stylo a été conçu et développé par l'équipe de la Chaire de recherche du Canada sur les écritures numériques (CRC-EN), Arthur Juchereau, Thomas Parisot, Guillaume Grossetie avec le support d'Érudit et d'Huma-Num.

¹ Voir <https://www.huma-num.fr/hnlab>

Sa mise à disposition sur HumanID est l'aboutissement d'une première phase du travail en commun réalisé par les équipes de la CRC-EN et d'Huma-Num et qui se poursuit aujourd'hui au sein du HN Lab. À la suite de ces échanges, une entente franco-canadienne a été signée entre Huma-Num (CNRS) et le Centre de recherche interuniversitaire sur les humanités numériques (Université de Montréal).

En 2022, le travail s'est accéléré avec l'ajout de plusieurs innovations apporté par les travaux du HN Lab en particulier en matière d'IHM et de liaison avec les référentiels scientifique d'ISIDORE (via l'API d'ISIDORE).

Statistiques d'usages de Stylo (décembre 2022)

Les statistiques ont été arrêtées au 23 décembre 2022 (nous les proposons avec les comparaisons de 2021) :

- 4 571 comptes en tout (+34%)
- 2 758 comptes créés avec mots de passe (+161%)
- 1 813 comptes créés avec HumanID (+57%)
- 16 874 articles en tout (+36%)
- 3,7 articles par compte en moyenne

La croissance de Stylo (+36% d'articles scientifiques écrits avec Stylo et +34% d'utilisateurs sur un an est le fait de l'important travail de formation et de pédagogie réalisé par l'équipe de la Chaire de Recherche du Canada sur l'écriture numérique (direction Marcello V. Rosati, Professeur à l'Université de Montréal) qui a créé et développe Stylo et le place dans l'écosystème international de l'édition des données (en relation avec Érudit, Cairn, Métopes, etc.). édition des données (en relation avec Érudit, Cairn, Métopes, etc.). Il faut noter cependant que le manque de documentation et le fonctionnement complexe d'HumanID pour les utilisateurs explique la forte croissance des comptes créés avec mots de passe (en particulier les difficultés de renouvellement de mot de passe et la confusion entre services ouverts et ceux laissés à la validation de l'IR* Huma-Num). C'est un point de vigilance à l'heure où d'autres dispositifs se développent à la fois en France (HAL, OpenEdition) mais aussi est surtout à l'international (EDUGAIN, EGI, etc.). Une meilleure coordination doit être mise en place entre l'équipe de développement et l'IR* Huma-Num afin d'éviter par ailleurs les incidents de l'année 2022 sur les comptes utilisateurs. L'enjeu est d'améliorer la fiabilité des dispositifs distribués et leurs évolutions qui doivent être avant tout piloté par les utilisateurs finaux des outils.

Stylo, outil développé à l'extérieur d'Huma-Num avec l'apport en matière d'innovation du HN Lab ce qui permet de le positionner à l'avant-garde des outils d'édition de la science ouverte pour les SHS. Ainsi les apports identifiés l'an passé se sont renforcés :

- Stylo apporte une brique logicielle open-source en tant que service pour l'écriture et l'édition des textes scientifiques et de recherche en SHS (ayant déjà une large communauté d'utilisateurs) ;
- Stylo est compatible avec les chaînes d'édition de Métopes, Érudit et prochainement OpenEdition et en particulier avec les utilisateurs de la TEI-XML et LaTeX.
- Stylo permet de proposer, avec un couplage avec NAKALA et ISIDORE, un ensemble complet pour l'édition d'articles de données (ou *data-papers*). Plus largement apporte des éléments de veille et de recherche et développement pour le programme HNSO et le projet COMMONS. Lire sur le sujet la communication de l'équipe du HN Lab :
https://www.meshs.fr/page/vers_un_ecosysteme_d_ecriture_avec_les_donnees et
<https://project.memorekall.com/nicolas-sauret-stephane-pouyllau-melanie-bunel---vers-un-ecosysteme-decriture-et-dedition-avec-les-donnees/>

b. Images scientifiques

Contenus : quels sont les usages scientifiques de l'image ? Comment instituer son exploitation, tant en termes de circulation et de publication que de traitement ?

L'axe 2 aurait dû être développé en 2022. Mais les moyens humains affectés au HN Lab n'ont pas été suffisant, surtout à partir de septembre 2022, pour lancer les travaux préliminaires à la résidence de Mme Servanne Monjour. Le programme de travail consiste en la préparation d'un environnement numérique de travail soit sous VIVO Web ou Omeka S / GraphDB pour permettre le développement des travaux de cette résidence.

Cependant, le HN Lab depuis sept. 2022, en relation avec la DSI du CNRS a mis en place les premières briques d'infrastructure sur 3 machines virtuelles. Cette preuve de concept servira d'environnement de départ pour la partie VIVO Web. Un important travail a été fait pour développer une solution réutilisable avec ces outils d'une dimension plus institutionnelle que les outils déployés traditionnellement dans les projets de recherche SHS.

Par ailleurs, le recrutement prévu d'un *Data-Scientist* au sein du HN Lab en mars 2023 devrait permettre d'associer un travail réflexif et de traitement sur les données de la résidence de S. Monjour.

Apports pour Huma-Num : Pour le moment, les apports n'ont pas été identifiés. Huma-Num ne développe pas de service dans cette direction pour le moment. Cette résidence se tourne plus vers des partenaires mêlant gestion des collections (ex. GED Condorcet) et extraction des connaissances iconographiques (EPHE ou INRIA avec l'IA, etc.).

c. Modélisation des données

Contenus : comment répondre à la tension entre normalisation des données et autonomie dans les processus de recherche ? Comment améliorer la littératie numérique sur la modélisation de données ?

L'axe 3 n'a pas été développé par manque de personnel. Il sera développé en 2024 si les conditions le permettent.

Cependant, dans cet axe, un travail a été initié en oct. 2022 sur les questions de modélisation des données dans le contexte de l'obsolescence des outils de publication Web massivement utilisés par Huma-Num. Ce travail, disponible sur <https://hnlab.huma-num.fr/blog/2022/09/08/Web-statique/> peut être résumé tel que :

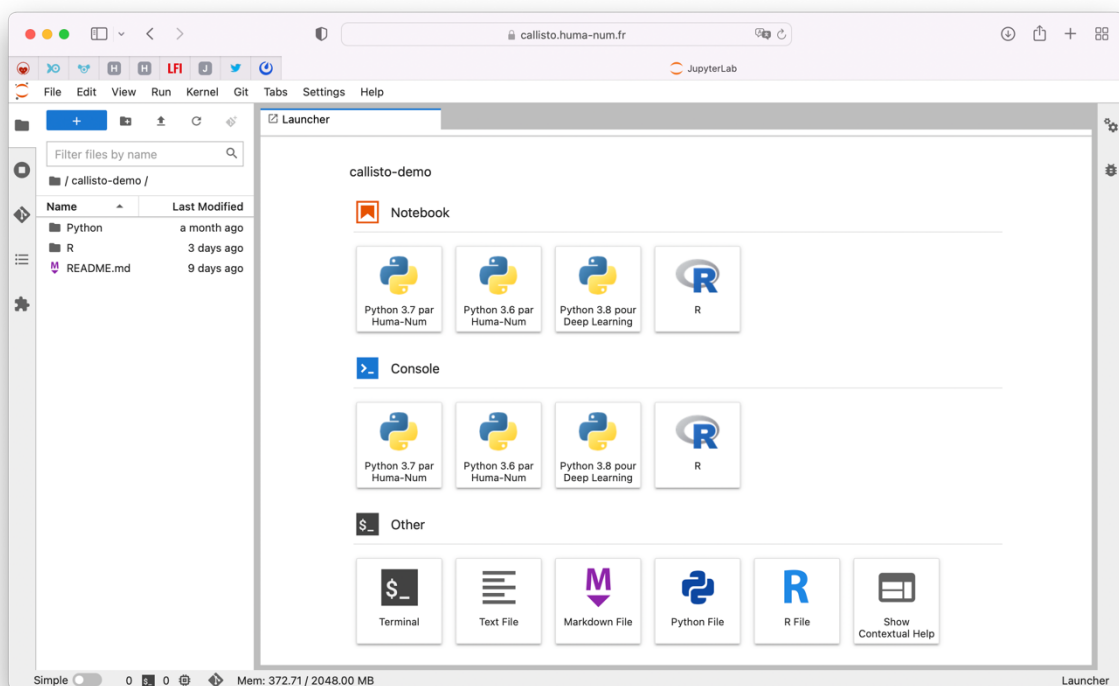
« Face à l'obsolescence rapide des systèmes de gestion de contenus Web s'appuyant sur les langages de programmation dynamiques traditionnels et les systèmes de gestion de base de données relationnelles (tel que WordPress, Drupal, etc. qui utilisent PHP et les SGBDR Mysql ou MariaDB), il est possible d'opter pour des systèmes de publication Web utilisant des pages HTML statiques, dépendant moins des mises à jour sous-jacentes de ces langages et outils dynamiques et s'inscrivant pleinement dans les principes fondamentaux du Web, en utilisant principalement le langage HTML. Si aucun système n'est bien sûr totalement parfait et il ne faut pas oublier que si l'on recherche la pérennité d'un dispositif sociotechnique, il faut plus regarder du côté de l'organisation (humaines, technologiques) qui le met en place que de l'espoir qu'il sera assez bien conçu et robuste pour traverser les années sans être dépendant au monde technologique qui l'entoure. Cependant, revenir aux fondamentaux du Web : un serveur Web, des pages HTML permettent dans la plupart des cas

d'atteindre l'objectif qu'on se fixe en souhaitant mettre à disposition des résultats de recherche scientifique. Pour une communication scientifique directe, le Web reste la première marche et il est important qu'elle soit accessible à toutes et tous, en particulier pour des étudiants de masters ou des doctorants par exemple, qui ne sont pas forcément accompagnés pour cela et pour lequel il n'est pas toujours évident, en master, d'aller déposer dans HAL ou DUMAS des travaux non encore aboutis. L'édition statique Web permet aussi de diffuser des travaux et d'horodater les premiers tours de roue d'une recherche (et ainsi débiter la dissémination d'un travail et ainsi contrer plus facilement les tentatives de plagiat). »

d. Analyses des données

Contenus : comment faciliter l'accès au *deep learning* pour les communautés SHS ? comment mutualiser la production de modèles d'usage spécifiques aux SHS ?

L'axe 4 a été principalement consacré à l'expérimentation Callisto (<https://callisto.huma-num.fr> et <https://hnlab.huma-num.fr/blog/2021/05/26/callisto-un-demonstrateur-jupyter/>) et aux GT sur les *data-papers* et plateforme de traitement des données en SHS. L'expérimentation Callisto est un succès auprès des communautés et le HN Lab a montré qu'il y avait un public SHS pour ce type de dispositif. Cependant sur le plan infrastructurel les choses sont plus contrastées.



Écran d'accueil de Callisto permettant aux chercheur-e-s d'écrire des articles de données à l'aide de plusieurs langages (Python, R, Librairie Python dédié à tel ou tel besoins SHS).

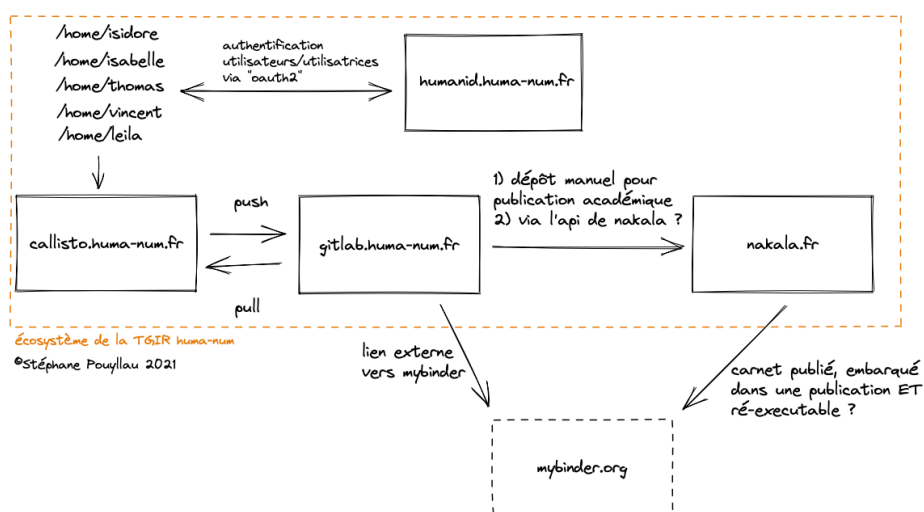
Le démonstrateur Callisto (du nom de l'une des lunes de Jupiter), est une preuve de concept proposant la mise en œuvre du logiciel en ligne Jupyter et plus particulièrement sa version JupyterLab pour des pratiques de recherche en SHS et en lien avec l'écosystème d'Huma-Num. Ainsi, Callisto repose sur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

- JupyterLab

- JupyterHub
- HumanID (le dispositif d'Huma-Num pour l'authentification)
- Gitlab (forge de dépôt permettant le versionnage de document)

Après un an et demi d'expérimentation Callisto dans son interconnexion à l'écosystème d'Huma-Num. L'expérience et le contexte actuel du HN Lab nous impose un premier bilan d'étape, à la suite duquel nous avons décidé de fermer la plateforme Callisto, dans sa forme actuelle, et cela pour plusieurs raisons :

- Malgré sa connexion avec HumanID et GitLab, Callisto n'a pas trouvé sa place dans l'écosystème d'Huma-Num qui a développé en parallèle une autre voie et a proposé un autre environnement JupyterLab (voir <https://jha-tools.huma-num.fr> et <https://documentation.huma-num.fr/calcul-scientifique/#le-service-de-calcul-scientifique>) ne s'appuyant pas sur les recommandations en séminaire de restitution et sur la proposition du HN Lab.
- Le développement des *kernels* dédiés aux besoins exprimés par les communautés autour de Callisto (principalement les membres du groupe de travail) fut un travail très intéressant, dont on a pu mesurer les besoins et le potentiel, mais qui est assez lourd à l'échelle du HN Lab. Nous remercions au passage les membres du GT et les collègues des communautés SHS qui ont participé et fait des propositions. Par ailleurs, nous avons déjà beaucoup de demandes d'utilisateurs et d'utilisatrices pour l'accès à Callisto dans le cadre leurs recherche ou de leurs doctorats et les limites que nous nous étions fixées pour cette preuve de concept sont déjà largement dépassées pour l'équipe du HN Lab qui se réduit en cette fin d'année (en particulier en matière d'exploitation et afin de ne pas proposer un service reposant sur une seule personne) ;
- Il nous semble important – si nous devons pérenniser Callisto comme un service, de le faire dans un cadre institutionnel et technique garantissant sa réussite, sa stabilité et son développement régulier. Dans cet esprit, une plateforme de « data science », adaptée aux besoins des SHS, telle que nous la concevions avec Callisto, doit nécessairement être exploitée et pilotée en relation avec les communautés (feuille de route, points d'étapes, etc.) et cela sort en partie du rôle du HN Lab en tant que tel. Du moins, le HN Lab ne peut faire ce travail seul.



Par ailleurs, suite aux ateliers menés avec la société coopérative et participative Dataactivist au printemps dernier, nous avons lancé un chantier éditorial pour produire 5 *notebooks* génériques et pédagogiques, centrés sur les besoins en SHS, qui seront prochainement mis à disposition de la communauté (ce travail, répondant au nom d'IO, est disponible sur <https://gitlab.huma-num.fr/io>). Les instruments créés par le HN Lab et Dataactivist sont centrés sur :

- **Les mobilités professionnelles des français - données spatiales** : il s'agit d'un traitement des données de l'INSEE sur les mobilités professionnelles des français et projections spatiales.
- **Les réponses du Grand Débat 2019 - textes et apprentissage automatique** : Analyse de la spatialisation des contributions au Grand Débat.
- **L'analyse des tweets sur le COVID-19 - réseaux et textes** : Analyse de tweets liés au COVID-19. Ce notebook présente les étapes de collecte, manipulation et analyse de données Twitter. A partir d'un jeu de données disponible, il analyse les tweets correspondant aux jours avant le confinement français de 2020. Données utilisées : liste des identifiants de tweets associés au COVID-19 multilingues collectées par une équipe de recherche et rendue disponible sur Zenodo : <https://zenodo.org/record/3723940>
- **La mise en place de questionnaire sur les pratiques numériques des chercheurs – tableaux et tableaux croisés** : Réanalyse des données d'un questionnaire sur les pratiques numériques des chercheurs en France conduite dans le cadre du projet State of Open Science Practices in France (SOSP).
- **L'analyse de recettes de cuisine dans l'art issues de Gallica (BnF) – images** : Analyse de l'évolution de l'iconographie dans un périodique de Gallica: l'art culinaire dans l'histoire.

Ainsi, malgré l'arrêt de Callisto, nous souhaitons poursuivre la réflexion pour porter à la fois la preuve de concept Callisto et ce chantier éditorial dans un contexte institutionnel ouvert aux besoins d'autres publics (doctorants, mais aussi l'enseignement supérieur). Les échanges au sein du GT Callisto ont donc vocation à continuer, notamment si une opportunité ou une initiative institutionnelle se présente pour reprendre le projet et surtout le développer dans le principe d'une plateforme communautaire. Une partie de la documentation du HN Lab est déjà disponible sur le GitLab d'Huma-Num comme retour d'expérience. En décembre 2022, dans le cadre des travaux menés dans le cadre du contrat doctoral HN Lab de Mme Léa Maronet (voir ci-dessous), une version améliorée de Callisto sera réactivée à des fins de recherche « à la demande » et exploité sans doute dans le cadre de l'infrastructure Calcul Québec.

3. Résidences

En 2022, le HN Lab a accueilli en résidence Mme Juliette De Maeyer, professeure agrégée à l'université de Montréal (Faculté des arts et des sciences - Département de communication). Sa résidence a été focalisée sur les bots (*Bots* au sens de robots) de chercheurs et ce qu'ils font aux archives et aux données de la recherche en SHS (<https://hnlab.huma-num.fr/blog/2021/02/01/une-nouvelle-residence-au-hnlab/>).

La résidence peut être résumée telle que :

Que font les bots à la recherche en SHS ? Dans notre démarche expérimentale, les bots ont agi à divers moments et à divers niveaux de la démarche de recherche. Ils constituent évidemment un outil de dissémination, en permettant la médiatisation de l'œuvre de Ford à différents publics, notamment sur les médias sociaux. Mais les bots ne sont pas simplement une façon de publiciser la recherche : leur comportement automatisé et aléatoire a parfois fonctionné de manière heuristique,

attirant notre attention sur certaines parties des archives négligées, ou proposant des juxtapositions inédites. Ils interrogent également les limites de l'archive et de son édition numérique : chaque action du bot, et les réactions qu'elle suscite, agissent comme tant d'annotations au corpus. Comment les capturer, les organiser et les publier au sein d'une édition numérique, qui serait nécessairement toujours en mouvement ? Enfin, les bots posent également des questions troublantes par rapport à l'écriture, aux sources et à leur autorité : notre chatbot génère du texte inédit en imitant le style de Ford. Écrivons-nous dès lors collaborativement avec un robot ? S'agit-il d'une source comme une autre ? La résidence HN Lab permettra de revenir sur toutes ces questions, et aussi d'identifier d'autres pouvoirs d'action de bots dans une démarche de recherche en SHS.

Faut-il négocier un passage du bricolage à l'industrialisation de ce genre d'expérimentation et si oui, comment ? L'expérimentation @franklinfordbot s'est faite en revendiquant une démarche artisanale et un certain goût pour le bricolage, et ces processus sont loin d'être "propres". Les transformations successives pour passer d'un ensemble de documents hétérogènes de la fin du 19^e siècle à des tweets créent, par exemple, des erreurs d'OCR et du bruit. Travailler avec ce matériau désordonné nécessite de devoir vivre avec des approximations, du non-sens, et des imperfections qui, plutôt que d'être considérés comme des défauts, témoignent d'une réflexion pratique sur la matérialité des médias. Dans quelle mesure ce désordre productif est-il possible si l'on change d'échelle ? La résidence HN Lab permettra d'interroger la transformation cette démarche de bricolage une fois confrontée aux préoccupations d'une infrastructure de recherche.

L'ensemble de la résidence a donné lieu à plusieurs publications encore sous presse et à plusieurs billets sur le log du HN Lab : <https://hnlab.huma-num.fr/blog/projets/bot-d-archives/> ainsi qu'une déclinaison pour ISIDORE d'un service de bots : <https://hnlab.huma-num.fr/blog/2022/01/08/ISIDOREBot/>

4. Contrat doctoral

Le HN Lab a accueilli son premier contrat doctoral en 2022 avec la thèse de Mme Léa Maronet (co-tutelle entre l'EPHE et l'Université de Montréal, financement CNRS). Son travail s'intègre dans le cadre de l'axe 2 du HN Lab. La thèse est présentée telle que :

« La sculpture en pierre dans le sous-continent indien, à son apparition, est d'abord caractéristique d'une période dite « formative », du iii^e siècle av. J.-C. au iii^e siècle ap. Difficiles à dater précisément en raison du caractère épars des sources, la majeure partie des monuments relevant de cette période formative sont au cœur de controverses dont nombre touchent à leur chronologie. Les motifs iconographiques du corpus aśokéen, figurés sur les colonnes appelés « piliers d'Aśoka », sont à la fois les premiers exemples de l'art monumental en pierre indien et sculptés sur l'ensemble des monuments de la période formative. Principalement employés comme motifs décoratifs, leur étude permettra de mieux appréhender la problématique de la datation des monuments, question centrale pour l'analyse des évolutions stylistiques et de la circulation des images dans le sous-continent. Parce qu'ils sont figurés en très grand nombre, leur étude sera rendue possible grâce à l'usage des humanités numériques et tout particulièrement de l'intelligence artificielle. Une méthode de reconnaissance automatique de motifs employant un réseau de neurones convolutif permettra de les cataloguer et de les classer. Celle-ci sera enrichie par une étude stylistique et iconographique afin de souligner la manière dont les motifs aśokéens évoluent dans les débuts de l'art monumental en Inde, à l'échelle du sous-continent. Une part de la réflexion se concentra sur l'intérêt de l'application d'une méthode computationnelle à l'étude de la tradition sculptée et sur les apports des humanités numériques à l'histoire de l'art. »

Le projet doctoral étudie un ensemble de motifs iconographiques issus d'une vingtaine de sites archéologiques indiens sur une période de six siècles. En raison de cet important corpus, la méthode proposée relève à la fois de l'histoire de l'art, à travers une analyse iconographique et stylistique de ces motifs, et des humanités numériques, avec l'usage de l'intelligence artificielle pour leur reconnaissance automatique et leur étude comparative.

L'usage d'un réseau de neurones convolutif (CNN) permettra la reconnaissance automatique des motifs et sera tout particulièrement essentiel des recherches : il s'agit de développer un algorithme et de l'entraîner sur les données à l'aide d'une méthode semi-supervisée – une partie des images sera pré-étiquetées par Léa Maronet. Au-delà de l'identification des motifs, le réseau de neurones permettra de classer ces derniers en sous-ensembles en fonction d'une typologie préalablement établie. À partir de ce travail, une étude iconographique pourra être menée afin de souligner quels éléments sont en commun et lesquels divergent dans chaque ensemble et afin d'établir une chronologie relative des différents motifs et des sites qui leur sont associés. Dans ce cadre, à l'automne 2022, un important travail a été mené afin de préparer la construction de la chaîne : tant sur le plan de l'infrastructure (échange avec Calcul Québec) que sur le plan de la conception et des briques logicielles nécessaires. Il s'agit aussi de tisser des liens avec des projets tels que MODOAP qui utilisent déjà ces dispositifs (en particulier Google Collab).

Si ces questions relèvent principalement de l'intelligence artificielle, le projet doctoral de Léa Maronet s'intéresse également à l'image numérique en elle-même car celle-ci est mon médium de travail principal. Il s'appliquera à questionner son traitement et à étudier, notamment, l'impact de la profondeur de la sculpture sur les performances du réseau de neurones. Il s'agit aussi d'étudier les questions de circulation des motifs – ou « contagions visuelles » à travers le sous-continent indien. L'objectif est, ici, d'identifier les routes de transmission, les influences inter-régionales, la prévalence – ou non – des différentes écoles de sculpture et si des apports extérieurs au sous-continent sont incorporés dans les représentations. Un tel projet s'inscrit assurément dans les thématiques « Images scientifiques » et « Intelligence artificielle et données de la recherche » proposées par le HN Lab.

5. Coordination du programme Huma-Num Science Ouverte (HNSO)

En 2021, le HN Lab avait proposé plusieurs réalisations dans le cadre du programme HNSO. La synthèse est disponible sur <https://hnlab.huma-num.fr/blog/2022/03/15/retrospective-hnso-2021/> dont les perspectives pour 2022 (disponible sur <https://hnlab.huma-num.fr/blog/2022/03/15/retrospective-hnso-2021/#perspective-pour-lannée-2022>). L'année 2022 devait être consacré au développement des fonctionnalités définies et décrites dans le cahier des charges d'HNSO (disponible sur <https://hnlab.huma-num.fr/blog/2022/03/15/hnso-cctp/>). Malheureusement ce travail n'a pas été entrepris dans ces fonctionnalités centrales. Le départ du responsable du pôle développement d'Huma-Num (non remplacé à ce jour), la résolution de la dette technique de NAKALA et le développement d'autres fonctionnalités, en particulier autour de NAKALA-PRESS, retarde ces travaux. En attendant, malgré la réduction du HN Lab, les actions de l'année 2022 ont été centrées sur trois principaux items :

- La question des référentiels dans ISIDORE et NAKALA (voir <https://hnlab.huma-num.fr/blog/2022/03/15/retrospective-hnso-2021/#les-référentiels>)
- La question des sources de données dans ISIDORE (voir <https://hnlab.huma-num.fr/blog/2022/03/15/retrospective-hnso-2021/#les-sources-de-données>)
- La question de la formation des utilisateurs et utilisatrices des plateformes ISIDORE et NAKALA (voir <https://hnlab.huma-num.fr/blog/2022/03/15/retrospective-hnso-2021/#la-formation-et-laccompagnement-utilisateurs-et-utilisatrices>)

La question des référentiels dans ISIDORE et NAKALA a été traitée sur le plan théorique et fonctionnel dans le cadre de l'ouvrage en édition continue disponible sur <https://hnlab.huma-num.fr/hnso/ouvrage-in/referentiels-isidore-nakala.html>.

a. ISIDORE

La question sur les sources de données dans ISIDORE a fait l'objet d'un important travail autour de la création d'une nouvelle application de gestion des sources. Dans cette évolution d'ISIDORE, qui entre dans HNSO (<https://gitlab.huma-num.fr/huma-num/hnso-cctp/-/issues/45>), une nouvelle interface de saisie et de gestion des sources de données — c'est à dire les sources à collecter et moissonner (en OAI-PMH ou à l'aide des flux de métadonnées Sitemap/XML et expression en RDFa de métadonnées) — a été créée et développée, elle est disponible sur <https://sources.isidores.science> :

Handle	Nom	Nom Court	Producteur	Isidore	Métrique	Date de Création	Date de Modification	Dernier Crawl	État	Action
pfv2nk	«_Centre-Entreprendre»	centreprenre	64	0	0%	2022-08-01	2022-08-01	2022-12-08	crawl	●
gfyvqe	«Musik – wohin?»	avantmusic	12	12	100%	2020-09-01	2020-09-01	2022-12-08	crawl	●
xdp5yk	«Wir sind das Volk!»	revolution1989	14	14	100%	2020-09-01	2020-09-01	2022-12-08	crawl	●
n290rw	« Au top ! »	unhiv	65	65	100%	2020-09-01	2020-09-01	2022-12-08	crawl	●
ayugqb	« Aux États-Unis de France et d'Irlande »	eufi	10	10	100%	2020-09-01	2020-09-01	2022-12-08	crawl	●
vq51rw	« Critiquables »	critiquabil	6	6	100%	2020-09-01	2020-09-01	2022-12-09	crawl	●
knusd4	« Des mots »	PUF_DEMOT	4	3	75%	2020-09-01	2020-09-01	2022-12-08	crawl	●
f9mqdx	« Devam edecek... »	devam	397	397	100%	2020-09-01	2020-09-01	2022-12-08	crawl	●
e69hiv	« Jardin » de poésie	versunjardin	37	40	108%	2020-09-01	2020-09-01	2022-12-08	crawl	●

Elle offrira aux producteurs de données un *back-office* complet leur permettant de configurer directement en ligne leur sources de données à ajouter, modifier, gérer pour ISIDORE. Elle facilitera la collaboration entre les producteurs de données et les équipes d'Huma-Num. Elle sera mise en service au passage en production de la nouvelle chaîne.

b. NAKALA

Concernant NAKALA, plusieurs fonctionnalités ont été engagées et ont été priorisées lors du comité projet n°2 du 25 mars 2022² : https://gitlab.huma-num.fr/huma-num/hnso-cctp/-/tree/master/comite-suivi-hnso/cr-reunions/25-03-2022_reunion-02. Fin 2022, on peut en tirer le bilan suivant :

L'amélioration du modèle « nakala:creator »

Lien : [hnso-cctp#7](#)

Objectif : Il s'agirait d'améliorer la notion de créateur mais aussi du rôle de contributeur en général, donc d'ajouter la notion de rôle (propriété ou attribut).

Avancée :

² Sur la formation des utilisateurs et utilisatrices, rien n'a été engagé pour le moment dans le cadre d'HNSO.

- semaines 51/52/1/2 : réalisation d'un benchmark de la gestion des profils auteurs/créateurs dans une sélection d'entrepôts généralistes et SHS : <https://sharedocs.huma-num.fr/wl/?id=BaX3PsWey4Jxp0QWV1wdTQDvAakImCpz> ;
- semaines 1/2/3/4 améliorations cosmétiques de la base actuelle : huma-num/nakala/nakala-data#13 et huma-num/nakala/nakala-data#14 ;
- premier trimestre 2023 : réalisation d'un sondage auprès des communautés sur l'utilisation des rôles dans le champ Auteur et scénarisation du nouveau fonctionnement.

Intégration de vocabulaires pour la description d'une donnée/collection

Lien : [hns0-cctp#2](#)

Objectif : L'objectif de cette fonctionnalité est de donner la possibilité aux chercheurs de décrire unitairement leurs données dans NAKALA grâce à des options d'ajout de vocabulaires lors de leur dépôt. Dans une première étape, une nouvelle étude des besoins en SHS et des conséquences possibles de l'introduction d'un ou n. nouveaux vocabulaires dans NAKALA (et aussi dans ISIDORE) est à faire.

Avancée : Cette fonctionnalité est prévue pour une livraison au premier trimestre 2023 : dans la continuité de l'issue sur le modèle « nakala:creator » qui prévoit de sonder les communautés sur l'usage du champ auteur, nous programmons d'élargir cette enquête aux usages des vocabulaires pour connaître les besoins.

Valideur de propriétés

Lien : [hns0-cctp#5](#)

Objectif : Contrôle de la conformité des propriétés.

Avancée : semaines 49,50,51,52 : prise en charge d'un premier contrôle de validation sur les dates (valideur sur les dates de types W3CDTF) : <https://gitlab.huma-num.fr/huma-num/nakala/nakala-api/-/issues/298>

Mise à jour du TripleStore RDF

Note : cette fonctionnalité n'existe pas en tant que tel dans le programme HNSO cependant elle pourrait être ajoutée comme celle qui concernait l'ajout d'un *endpoint graphql* : <https://gitlab.huma-num.fr/huma-num/hns0-cctp/-/issues/14>. En effet il s'agit d'assurer la mise à disposition d'une base de données RDF (*TripleStore*) et à le tenir à jour. Cette opération s'inscrit dans la politique de l'ouverture des métadonnées dans le *Linked Open Data* mondial auquel participe déjà ISIDORE depuis 2010.

Objectif : Mettre à jour le point d'accès normalisé RDF de NAKALA pour assurer une meilleure interopérabilité des données dans le *Linked Open Data*.

Avancée : semaines 48 à 52 : une première étape a consisté en l'ajout d'identifiants de type DOI aux données issues de NAKALA et ce avant décembre 2020 (NAKALA utilisait alors des identifiant de type *handle*) pour normaliser les URI de l'ensemble de la base RDF.

Passage à l'API platform

Lien : [hns0-cctp/-/issues/33](#)

Note : cette fonctionnalité n'existe pas en tant que tel dans le programme HNSO, cependant il s'agissait un prérequis indispensable pour mettre à niveau NAKALA avant d'engager d'autres travaux HNSO.

Objectif : passer sur API Platform (framework web associé à une communauté dynamique, utilisé pour générer des API REST et GraphQL).

Avancée : chantier en cours sur l'année 2022-2023

Ajout d'un endpoint GraphQL à l'API

Lien : hns0-cctp/-/issues/14

Objectif : permettre de requêter l'API par GraphQL (GraphQL est un langage de requêtes de données pour API). QL, comme dans SQL, veut dire *Query Language*. GraphQL permet de manipuler de la donnée de façon simple, flexible et très précise. GraphQL n'est pas un langage de programmation ou un *framework*. GraphQL est une spécification pour implémenter l'API. Les données doivent être en JSON-LD.

Avancée : réalisé (octobre 2022).