

# Ar(*abes*)ques

AVRIL - MAI - JUIN 2022

DOSSIER

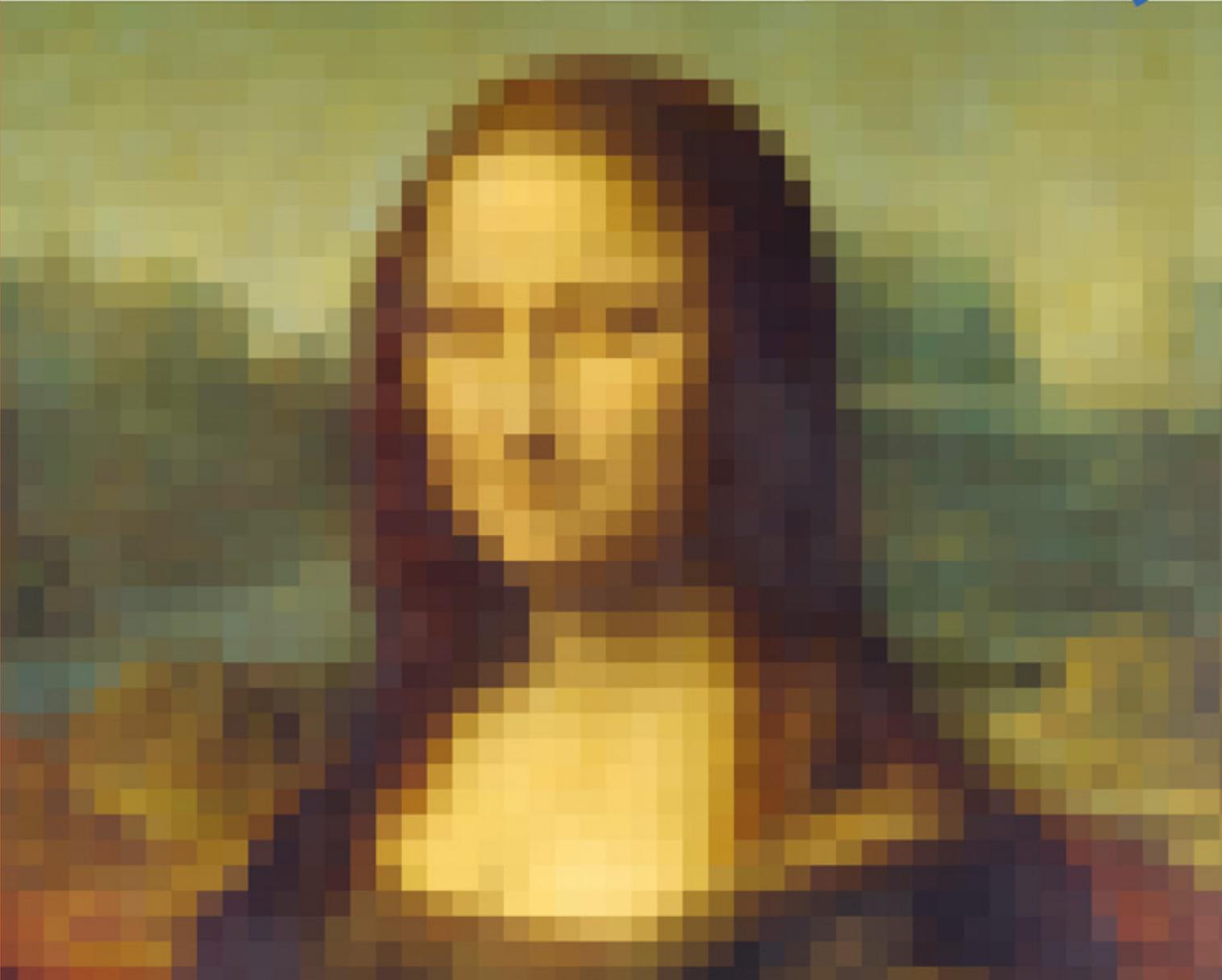
## Humanités numériques *Une renaissance 3.0*

**PLEINS FEUX SUR** • La médiathèque de l'Ifpo

**SYSTÈME D** • Decalog, une plateforme documentaire conçue pour vivre la mutation des catalogues

105

abes  
agence bibliographique  
de l'enseignement supérieur



# (Dossier) Humanités numériques

## Une renaissance 3.0

En appliquant les technologies et les outils les plus innovants de l'informatique et de l'intelligence artificielle aux sciences humaines et sociales, les humanités numériques bouleversent en profondeur le travail des chercheurs mais aussi celui des professionnels de l'information et de la documentation. Les bibliothèques, riches de leurs vastes corpus de ressources numérisées ou nativement numériques, constituent en effet des partenaires privilégiées des chercheurs et des praticiens des humanités numériques. Même si le champ des possibles en matière de collaboration entre ces différentes communautés professionnelles reste encore largement à inventer et à investir, les initiatives présentées dans ce nouveau dossier d'*Arabesques* illustrent des expérimentations déjà très abouties, comme la Boîte à Outils d'Intelligence Artificielle (BaOIA), fruit d'un travail entre les équipes de La contemporaine et un chercheur du Centre des sciences des littératures en langue française, ou encore les divers projets menés par l'université de Genève qui a élaboré une stratégie systématique en faveur des humanités numériques. Le récent DataLab de la Bibliothèque nationale de France, conçu comme un espace physique et virtuel de dialogue entre les différentes expertises mobilisées pour l'exploration des collections de la BnF, préfigure, quant à lui, une nouvelle génération de services en bibliothèque. Bonne lecture !

- 04 Humanités numériques : une présence dans le quotidien, une anticipation dans l'innovation et la préservation OLIVIER LE DEUFF
- 06 Huma-Num : une infrastructure au service des sciences humaines et sociales NICOLAS LARROUSSE, HÉLÈNE JOUGUET
- 08 La BnF et les services à la recherche à l'heure des Humanités numériques MARIE CARLIN, ARNAUD LABORDERIE
- 10 BaOIA : une « boîte à outils » pour explorer les corpus numérisés CÉCILE TARDY, CHLOÉ JEAN
- 12 Circulations des savoirs, de la recherche à l'enseignement : une approche genevoise BÉATRICE JOYEUX-PRUNEL, SIMON GABAY
- 14 L'enseignement des humanités numériques à l'École nationale des chartes et à Lyon INTERVIEW DE THIBAUT CLÉRICE, ET DE SABINE LOUDCHER
- 16 Des catalogues de bibliothèques aux projets en humanités numériques : les autorités IdRef font le lien FRANÇOIS MISTRAL
- 18 Un outil de comparaison visuelle des collections numérisées de Persée et de leurs équivalents imprimés
- 19 Entretien avec Aurélien Berra, rédacteur en chef de la revue *Humanités numériques* « Les humanités numériques développent des interrogations nouvelles »

## 20 (Pleins feux sur...)

**Des fonds patrimoniaux à l'actualité de la recherche : une médiathèque régionale au Proche-Orient**

TAOS BABOUR ET ELSA ZAKHIA

## 22 (Système D...) des outils pour vos données

Decalog, une plateforme documentaire conçue pour vivre la mutation des catalogues  
GUILLAUME MESSONNIER, DIDIER PIED ET FRANÇOIS RAFFIN

## 23 (Actus...)

Levée de voile sur le programme des Journées Abes 2022

## 24 (Portrait)

### Ar(abes)ques

REVUE TRIMESTRIELLE DE L'AGENCE BIBLIOGRAPHIQUE DE L'ENSEIGNEMENT SUPÉRIEUR,  
227, avenue du Professeur-Jean-Louis-Viala, CS 84308, 34193 Montpellier cedex 5.  
Tél. 04 67 54 84 10 / Fax 04 67 54 84 14 / <https://abes.fr>

Directeur de la publication : David Aymonin.

Coordination éditoriale et secrétariat de rédaction : Véronique Heurtematte.

Comité de rédaction : Christophe Arnaud, Aurélie Favre, Christine Fleury, Etienne Naddeo

Morgane Parra, Laurent Piquemal, Marie-Pierre Roux.

Iconographie rassemblée par Christophe Arnaud.

Conception graphique : Anne Ladevie ([anneladevie.com](http://anneladevie.com)). Impression : Pure Impression

Revue publiée sous licence Creative Commons CC BY-ND 2.0

(Paternité - Pas de modifications) sauf pour les images qui peuvent étre soumises à des licences différentes ou à des copyrights.

Les opinions exprimées dans Arabesques n'engagent que la responsabilité de leurs auteurs.

ISSN (papier) 1269-0589 / ISSN (web) 2108-7016



# Humanités numériques 2022 : les sciences sociales et humaines «reloaded»<sup>1</sup>

**A**vec le printemps qui pointait, l'équipe d'*Arabesques* s'est dit qu'était venu le temps de présenter un tableau coloré et sensible des activités nouvelles qui occupent les bibliothèques de l'ESR, leurs partenaires et leurs utilisateurs.

Le choix s'est porté sur les humanités numériques, dont le développement est spectaculaire en France, notamment depuis l'énoncé du « Manifeste des *Digital Humanities* »<sup>2</sup> en 2010. Ces portraits choisis ne veulent pas dresser un panorama exhaustif mais plutôt impressionniste, qui témoigne des acquis, des évolutions récentes et des perspectives à venir dans le domaine des humanités numériques en France. L'article inaugural d'Olivier Le Deuff permet de les replacer dans un contexte international. Comme le dit Marie Carlin dans son article consacré au BnF DataLab, « Si les humanités numériques transforment profondément le travail des chercheurs, elles modifient tout autant le métier des professionnels des bibliothèques. Introduire les humanités numériques dans la bibliothèque est un défi qui nécessite à la fois de nouvelles compétences (...) et de nouvelles infrastructures en termes d'espaces et d'équipements ».

De plus en plus de bibliothèques s'associent à des équipes de recherche, à des groupes d'étudiants avancés, ou recrutent des professionnels disposant des compétences nécessaires à ces nouvelles activités. Comme un courant inverse, de plus en plus d'équipes de recherche s'approprient les vastes corpus de documents et de données constitués autour des grands catalogues



et de réservoirs numériques de plus en plus variés, déployés par les centres documentaires. Le mouvement s'étend, et ce sont aujourd'hui les méta-données produites par les réseaux de l'Abes – notices bibliographiques et d'autorité,

qui deviennent fréquemment une matière première pour les humanités numériques.

Gageons que les articles de cette nouvelle livraison d'*Arabesques* vous donneront envie d'en savoir plus et de participer, si ce n'est pas encore le cas, au grand mouvement numérique qui transforme les sciences humaines et sociales. Ce numéro est aussi l'occasion de saluer nos collègues de la médiathèque de l'Institut français du Proche-Orient qui maintiennent leur activité dans des conditions très difficiles à Damas, Beyrouth, Amman et Erbil.

Bon courage à toutes et tous, là-bas, ici, ailleurs...

**DAVID AYMONIN**  
Directeur de l'Abes

[1] Clin d'œil à un billet de blog d'un de nos pionniers en HN, Etienne Cavalié, publié en 2015, et annonçant ce livre <https://books.openedition.org/pum/11091?lang=fr>

[2] « Manifeste des *Digital Humanities* », *Journal des anthropologues* [En ligne], 122-123 | 2010, mis en ligne le 01 décembre 2012, <http://journals.openedition.org/jda/3652> ; doi.org/10.4000/jda.3652

# Humanités numériques :

## *une présence dans le quotidien, une anticipation dans l'innovation et la préservation*

Les humanités numériques articulent les nouvelles potentialités de traitement informatisé pour améliorer la constitution et l'exploitation de bases de données et renouveler les possibilités en matière de visualisation et d'analyse.



L'expression d'humanités numériques ou digitales s'est imposée depuis une quinzaine d'années et s'accompagne d'une série d'avancées et de réalisations qui permettent de mieux saisir dorénavant cette évolution des sciences humaines et sociales. L'histoire longue des humanités numériques nous resitue dans celle de l'organisation des connaissances et de ses outils. L'histoire courte la place dans des initiatives de rapprochement entre lettres, humanités et informatique. Si le mythe tend à placer le projet d'*index thomisticum*<sup>[1]</sup> du prêtre Robert Busa en collaboration avec IBM comme précurseur, il est possible de citer parmi les pionniers bien d'autres projets universitaires, qu'ils soient sous l'appellation *Informatica Umanistica* ou bien celle d'*Humanities Computing* qui a longtemps précédé celle de *Digital Humanities*.

Concrètement, le mouvement articule les nouvelles potentialités de traitement informatisé pour améliorer la constitution de bases ou banques de données, leur interrogation, le partage des données et les possibilités renouvelées en matière de visualisation et d'analyse.

Les humanités numériques ou digitales sont donc des humanités augmentées autant dans leurs perspectives de recherche que d'enseignement. Elles requièrent des outils, des méthodes, des langages pour pouvoir bénéficier de nouvelles manières de voir et de nouvelles manières de faire. Les humanités numériques sont souvent mentionnées dans

les logiques de projet de recherche notamment pour obtenir des financements conséquents et plus ambitieux.

### RATIONALISER LES PROCESSUS

Innovantes et parfois en marge notamment avec les manifestations de type *THATCamp* qui avaient pour but de renouveler les colloques et journées d'étude pour inciter à davantage de participation, formation, échanges et productions, les humanités numériques se sont peu à peu institutionnalisées avec des reconnaissances au niveau des profils d'enseignants-chercheurs et d'ingénieurs. Des diplômes ainsi que des laboratoires affichent désormais la mention « humanités numériques » dans leurs intitulés. Depuis le manifeste de Paris 2010<sup>[2]</sup>, la reconnaissance s'est organisée aux niveaux universitaires, interuniversitaires et associatifs. Plusieurs associations existent comme ADHO<sup>[3]</sup> (*Alliance of Digital Humanities Organizations*), EADH (*European Association for Digital Humanities*)<sup>[4]</sup> et Humanistica<sup>[5]</sup>, association francophone créée lors du *THATCamp* de Saint Malo en 2013 et qui organise désormais son colloque et bénéficie d'une revue dédiée (cf p. 19).

La période est désormais à une rationalisation des processus de façon à éviter que les projets même potentiellement innovants finissent par périr faute de nouveaux financements. Il s'agit dorénavant d'accompagner chaque projet d'un plan de gestion de données, à des fins de conservation, mais également

[1] <https://www.corpusthomicum.org/it/index.age>

[2] « Manifeste des *Digital Humanities* ». *THATCamp* Paris. <http://tcp.hypotheses.org/318>

[3] *Alliance of Digital Humanities Organizations*

[4] <https://eadh.org>

[5] <http://www.humanistica.ca>

[6] <https://cahier.hypotheses.org>

[7] <https://documentation.huma-num.fr>



d'interopérabilité. Les développements informatiques et logiciels réalisés s'inscrivent dans des démarches *open source* de façon à faire bénéficier des avancées réalisées à l'ensemble de la communauté. Les évolutions de certains logiciels, comme Omeka et Omeka-S, ou Zotero produits par le *Roy Rosenzweig Center for History and New Media* font partie de ces exemples réussis. D'autres exemples existent dans le champ des bonnes pratiques autour de la numérisation et de l'indexation des corpus notamment en XML TEI avec des communautés dédiées. On peut citer par exemple le travail réalisé par le consortium Cahier (Corpus d'Auteurs pour les Humanités : Informatisation, Édition, Recherche)<sup>6</sup> ces dix dernières années.

### DES OUTILS DÉDIÉS INTÉGRÉS DANS LES PRATIQUES DE RECHERCHE

L'infrastructure Huma-Num (*cf p. 6-7*) participe également de cet esprit et facilite le travail des chercheurs dans leur gestion de projet en leur offrant une série de services et d'outils pour héberger un site Web, gérer leurs données avec Nakala, bénéficier d'un *drive* avec *Sharedoc* ou bien encore utiliser certains logiciels dédiés en ligne comme *Stylo* ou *Voyant.tools*<sup>7</sup>. L'infrastructure se développe à la fois à l'international, notamment au niveau européen avec Dariah (*Digital Research Infrastructure for the Arts and Humanities*)<sup>8</sup>, dans le cadre de partenariats avec le Québec, mais également au niveau local avec le développement de services disponibles au sein des Maisons des sciences de l'Homme.

Impossible de ne pas évoquer le rôle décisif d'*Open Edition* à plusieurs niveaux : celui de la valorisation et du rapprochement science-société avec les carnets de recherche Hypothèses, celui d'une proposition éditoriale riche au niveau des revues sous le logiciel Lodel et la mise à disposition d'ouvrages *via Open Books*. Les humanités numériques s'inscrivent désormais clairement dans le quotidien des chercheurs, ingénieurs et bibliothécaires avec des outils dédiés qui se sont peu à peu intégrés dans les pratiques. Malgré les quelques réticences au niveau des discours à l'égard des humanités numériques, force est de constater que la logique services et outils s'est montrée convaincante tout comme la volonté d'y puiser matière à innover et à renouveler des approches scientifiques et pédagogiques.

### IMAGINER DES FORMES HYBRIDES ENTRE RECHERCHE ET BIBLIOTHÈQUES

Il reste encore à accomplir la dimension d'ingénierie technique et d'ingénierie organisationnelle en imaginant des formes hybrides entre recherche et bibliothèques, entre documentation et production. Des pistes de « lab » ont été élaborées et mêlent différents types de profils pour parvenir à développer et maintenir des projets ambitieux. Ainsi, le Medialab de Sciences Po<sup>9</sup> présente un modèle intéressant en

matière de recherche et développement (les logiciels *Gephi*<sup>10</sup> et *Hyphe* y sont notamment développés) tandis que le récent BnFDataLab<sup>11</sup> (*cf p.8-9*) met l'accent sur la nécessité de prendre soin des données de la recherche, des plus récentes à celles issues de numérisation.

### MIEUX VALORISER LES COMPÉTENCES ET SAVOIRS ASSOCIÉS AUX HUMANITÉS NUMÉRIQUES

Quoi qu'il en soit, l'avenir des humanités numériques repose sur cette double perspective : celle de la pérennité des services et des données produites qui requièrent des processus mais aussi des recrutements dédiés ; celle de la recherche-développement et de l'innovation qui nécessitent des collaborations originales, des prises de risques, la capacité d'articuler concepts et logiciels et qui vont également avoir besoin de soutiens financiers pour le démarrage et d'assurance de pérennité en cas de succès. Cela implique en tout cas de mieux valoriser les compétences et savoirs associés aux humanités numériques dans les profils d'enseignants-chercheurs, d'ingénieurs et au sein des services de documentation.

OLIVIER LE DEUFF

Maître de conférences en sciences de l'information et de la communication, Université Bordeaux Montaigne

[8] <https://www.dariah.eu>

[9] <https://medialab.sciencespo.fr>

[10] <https://gephi.org>

[11] <https://www.bnf.fr/fr/bnf-datalab>



#### POUR EN SAVOIR PLUS

Burnard, Lou (2014) *What is the Text Encoding Initiative? : How to add intelligent markup to digital resources*. OpenEdition Press. <http://books.openedition.org/oepp/426>

Carlin, Marie, Laborderie, Arnaud (2021). « Le BnF DataLab, un service aux chercheurs en humanités numériques », *Humanités numériques*, 4, <http://journals.openedition.org/revuehn/2684>

Hockey, Susan (2008) « The histories of humanities computing », in S. Schreibman S., R. Siemens, J. Unsworth (Éd.), *A Companion to Digital Humanities*, Wiley-Blackwell

Idmhand, Fatiha (2021) *Dix ans avec CAHIER. Bilan du consortium CAHIER (2011-2021) de la TGIR Huma-Num*. [Rapport de recherche], Huma-Num; CAHIER - Consortium CAHIER. 2021. <https://halshs.archives-ouvertes.fr/halshs-03419128>

Le Deuff, Olivier (2014) « Bibliothèques et lieux de production de savoirs ». Dans *Le Temps des humanités digitales. La Mutation des sciences humaines et sociales*, sous la dir. d'Olivier Le Deuff. Limoges : FYP Éditions. [https://archivesic.ccsd.cnrs.fr/sic\\_01487073](https://archivesic.ccsd.cnrs.fr/sic_01487073)

Le Deuff, Olivier (2017) *Les humanités digitales. Historique et développements*. Iste éditions.

Mc Carty, Willard (2005) *Humanities Computing*. Palgrave Macmillan

# Huma-Num : une infrastructure au service des sciences humaines et sociales

Huma-Num propose des services dédiés à la production, la gestion et la réutilisation de données numériques et favorise l'émergence d'outils et de bonnes pratiques issus des communautés scientifiques des SHS.



Les projets de recherche en sciences humaines et sociales (SHS) produisent d'importants volumes de données numériques et utilisent des outils spécialisés pour les rendre accessibles, les traiter et les diffuser. Cela nécessite la mise en œuvre de moyens technologiques qui se doivent d'être stables, bien dimensionnés et en adéquation avec l'évolution des pratiques scientifiques. Pour accompagner la maîtrise par les projets de cet environnement complexe, il est nécessaire de lui associer des compétences diversifiées et en constante évolution. C'est la mission d'Huma-Num<sup>1</sup>, infrastructure de recherche (IR)<sup>2</sup> ayant pour tutelles principales le CNRS et le Campus Condorcet, et pour tutelle secondaire l'université d'Aix-Marseille. Huma-Num propose des services dédiés à la production, la gestion et la réutilisation de données numériques et favorise ainsi l'émergence d'outils et de bonnes pratiques issus des communautés scientifiques SHS. Huma-Num est également impliquée dans la construction d'infrastructures européennes.

## FOURNIR DES MOYENS ET OUTILS ADAPTÉS AUX BESOINS DES PROJETS

Huma-Num développe une grille de services qui répond aux principaux besoins de stockage, de traitement et d'exposition des données de la recherche en SHS. Celle-ci se caractérise par la production et l'utilisation de données ayant des typologies et des provenances très diverses selon la nature des projets. Par ailleurs, les formats, la volumétrie, les besoins de traitements, les possibilités de diffusion varient fortement d'un projet à l'autre. Dans ce contexte, Huma-Num maintient un ensemble de services informatiques mutualisés et sécurisés qui doivent pour autant répondre à des besoins hétérogènes. Pour se rapprocher au mieux des besoins de chaque chercheuse et chercheur, Huma-Num appuie son offre de services et d'outils sur l'expertise forgée par des communautés scientifiques réunies en consortiums. Ces consortiums Huma-Num sont constitués de chercheurs, d'informaticiens et de spécialistes de l'information et de la documentation qui disposent alors d'un cadre pour travailler collectivement sur des objets scientifiques issus d'une même discipline,

par exemple des cartes de géographes, ou transverse à plusieurs disciplines, comme les modèles 3D.

C'est en s'appuyant sur ses consortiums et sur le HNLab, pôle d'innovation qui développe des projets autour de nouveaux services en accueillant des chercheuses et des chercheurs en résidence, qu'Huma-Num structure ses services numériques autour du cycle de vie des données. Chaque étape de ce cycle peut être pris en charge par les services mis à disposition, que ce soit pour du stockage sécurisé et des outils pour le traitement des données sur la grille de calcul ; pour entreposer des données, les stocker et les exposer avec le service Nakala ; pour préserver des données sur le long terme sur la plateforme du Cines<sup>3</sup>. Huma-Num développe et maintient également le moteur de recherche et d'accès unifié Isidore<sup>4</sup>.

Dans la conception de ses services, Huma-Num porte une attention particulière aux enjeux actuels de la science ouverte et des aspects juridiques et éthiques des données de la recherche, tout en recherchant des solutions pratiques et fiables pour les utilisateurs.

## ACCOMPAGNER LES UTILISATEURS

Initialement intégré aux activités de chaque ingénieur, l'accompagnement des utilisateurs a été structuré en 2020 au sein d'un pôle interne dédié. Celui-ci s'appuie sur un réseau de correspondants relais au sein des Maisons des Sciences de l'Homme présentes sur l'ensemble du territoire national. En effet, l'accompagnement des utilisateurs est identifié comme un levier nécessaire et prioritaire pour améliorer la gestion des données et ainsi répondre pleinement aux enjeux du FAIR (*Findable, Accessible, Interoperable, Reusable*). Si les services numériques sont variés, les besoins en termes d'accompagnement sont également très différents selon que l'équipe projet dispose ou non de compétences informatiques et/ou documentaires et de temps nécessaire pour ces tâches sur l'ensemble de la démarche. Cet accompagnement se déploie également dans des structures partenaires, comme par exemple le BnFDataLab inauguré en octobre 2021 (*cf p. 8-9*).

En septembre 2021, l'équipe d'Huma-Num a orga-

[1] <https://www.huma-num.fr>

[2] Les IR sont des infrastructures de recherche qui relèvent d'une politique scientifique nationale du MESRI avec un financement dédié.

[3] <https://www.cines.fr/archivage-numerique-perenne-ouverture-collabore-avec-huma-num-et-le-cines>

[4] <https://isidore.science>

nisé une Action nationale de formation (ANF) intitulée « Gérer ses données en SHS avec les services et outils proposés par la TGIR Huma-Num »<sup>5</sup>. Cette formation, entièrement disponible en ligne, fournit aux projets de recherche les moyens de se préparer aux enjeux scientifiques qui nécessitent la « FAIRisation » des données produites. Composée de 6 modules, la formation aborde les différents aspects liés à la gestion du cycle de vie de l'information et articulés autour de l'entrepôt de données Nakala.

## DÉVELOPPER LA QUALITÉ DES DONNÉES

Si les services d'Huma-Num permettent la mise en oeuvre technique de la majorité des principes FAIR – gestion d'identifiants, gestion de métadonnées, protocoles ouverts et interopérables – il n'en reste pas moins nécessaire d'assister les utilisateurs dans le rôle qu'ils ont à jouer pour rendre intelligibles leurs données sur le long terme (organisation des données, transfert dans un entrepôt, description, réalisation d'une éditorialisation). Pour répondre à ces objectifs, l'équipe d'Huma-Num, en comité interne, analyse chaque demande d'utilisateurs selon plusieurs aspects : identification des besoins informatiques et documentaires, orientation vers d'autres sources d'informations pertinentes le cas échéant, détermination de la nature de l'aide à apporter. En outre, Huma-Num s'inscrit dans le mouvement collectif des infrastructures françaises – notamment Open Edition Center, Métopes, le CCSD, Persée – ayant vocation à développer l'interopérabilité entre les services de chaque partenaire pour contribuer à une meilleure diffusion des données de la recherche. Huma-Num accompagne l'évolution des pratiques de la recherche en SHS, induite par sa numérisation croissante. Il s'agit dorénavant de proposer un accompagnement dans l'utilisation des outils proposés afin d'optimiser la qualité des données et

## ● ● ● L'INFRASTRUCTURE DARIAH : LES HUMANITÉS NUMÉRIQUES À L'ÉCHELLE EUROPÉENNE

**Dariah** (*Digital Research Infrastructure for the Arts and Humanities*) est une infrastructure européenne, de type ERIC (*European Research Infrastructure Consortium*) pour les sciences humaines et sociales qui regroupe 20 pays membres à part entière et un observateur. Chaque pays contribue à Dariah, constituant ainsi un ensemble de services associés à des réseaux d'expertises qui sont à disposition des communautés nationales. En voici quelques exemples :

- Trouver et partager des outils, services et autres ressources sur le *SSH Open Market Place* développé dans le cadre du projet SSHOC<sup>7</sup>.
- Trouver et partager des ressources pédagogiques sur Dariah Campus.
- Trouver et partager des cursus universitaires en humanités numériques au niveau européen en consultant le *DH Course Registry* géré par les infrastructures Dariah Clarin.
- Participer aux activités des nombreux *Working Group* qui traitent de thématiques diversifiées et proposer d'en créer un le cas échéant.

Huma-Num, qui est l'institution coordinatrice pour la France, a pour mission de fournir les informations nécessaires sur Dariah et d'accompagner les interactions et échanges avec cette infrastructure.



métadonnées associées en s'appuyant sur l'expertise des communautés *via* le dispositif des consortiums Huma-Num. Ces activités et orientations s'inscrivent pleinement dans la dynamique nationale actuelle en faveur de la science ouverte et des initiatives portées par le programme « Recherche data gouv »<sup>6</sup>.

### NICOLAS LAROUSSE

Reponsable du pôle *Coordination nationale et internationale des communautés d'Huma-Num*  
nicolas.larousse@huma-num.fr

### HÉLÈNE JOUGUET

Reponsable du pôle *Données et coordination de l'accompagnement utilisateur d'Huma-Num*  
helene.jouguet@huma-num.fr

[5] <https://humanum.hypotheses.org/6607>

[6] <https://www.ouvrirelascience.fr/recherche-data-gouv-plateforme-nationale-federee-des-donnees-de-la-recherche>

[7] <https://sshopencloud.eu>

## HN DES SERVICES POUR LES DONNÉES EN SCIENCES HUMAINES ET SOCIALES



Par ses outils développés pour Gallica et son récent DataLab, la Bibliothèque nationale de France propose une nouvelle génération de services pour l'exploration des ressources numériques et la production de connaissances.



## La BnF et les services à la recherche à l'heure des humanités numériques

**Par la richesse et la particularité de ses collections, les liens qui unissent la Bibliothèque nationale de France (BnF) au monde de la recherche sont nombreux et anciens.** Cet attachement fort lui permet de suivre et d'accompagner les chercheurs qui, s'appuyant sur la disponibilité de vastes collections, renouvellent les pratiques de recherche en sciences humaines et sociales. L'augmentation massive des collections numériques (numérisations sur Gallica, archives de l'Internet, métadonnées du catalogue, notices d'autorité, etc.) a favorisé l'apparition d'un nouveau patrimoine dématérialisé et ouvert des pistes de recherche : fouille de textes et d'images, lecture distante, data visualisation, réutilisation et alignement de référentiels, entre autres.

Confrontée à des demandes récurrentes de mise à disposition de corpus massifs, la BnF a su au fil du temps adapter ses services pour répondre aux enjeux liés à ces nouvelles pratiques de recherche.

### FACILITER LA COMPRÉHENSION DES COLLECTIONS

Depuis sa création en 1994, la BnF s'est interrogée sur l'articulation entre accroissement exponentiel des collections et diffusion des savoirs : devant une masse sans cesse grandissante de documents, comment diffuser au mieux et au plus grand nombre ? Comment rendre cette masse non seulement disponible mais également exploitable pour la production des savoirs ?

La dématérialisation et le recours au réseau pour communiquer et valoriser les collections ont été mis en oeuvre dès 1997 avec la création de Gallica, la bibliothèque numérique de la BnF et de ses partenaires, qui comptait à son lancement 2 500 ouvrages et 10 000 images.

Depuis, les collections numériques de la BnF n'ont cessé de s'étoffer et représentent aujourd'hui environ 6 pétaoctets. Elles sont caractérisées par une variété considérable : documents numérisés (9 millions de documents disponibles dans Gallica), documents

nativement numériques, numériques sur support (CD, DVD, jeux vidéo), archives de l'Internet, métadonnées bibliographiques et données d'autorité. Autant d'ensembles de données hétérogènes, qui ont toutes des structures, formats, qualité, contextes de production, fonctions et contenus différents. La compréhension de ces collections exige des traitements spécifiques et par conséquent des compétences et expertises particulières, aussi bien pour les conserver ou les communiquer que pour les analyser. La BnF travaille non seulement à mettre à disposition ces collections mais aussi à en faciliter la compréhension et la manipulation par la mise en place d'une gamme d'outils dédiés aux opérations suivantes : aide à la constitution de corpus (interface de recherche par proximité, par occurrence, par similarité) ; extraction des documents (portail API et jeux de données, export csv du Catalogue général) ; recherche avancée dans les métadonnées (data.bnf.fr et le *sparql endpoint*).

### DES OUTILS POUR MIEUX APPRÉHENDER LES DOCUMENTS DANS GALLICA

D'abord numérisés en mode image, les documents textuels ont, depuis 2005, systématiquement été convertis en mode texte par un logiciel OCR, autorisant la recherche plein texte et l'extraction de corpus. À ce jour, une partie seulement de la collection numérique est accessible en mode texte. Pour y remédier, la BnF a engagé un ambitieux programme de rétroconversion.

De plus, toutes les collections textuelles ne sont pas éligibles à l'OCR. C'est le cas bien sûr des manuscrits, mais aussi d'une partie des collections de la presse dont les caractères, la mise en page, l'état de conservation rendent difficile la reconnaissance des caractères. L'un des axes de la feuille de route 2021-2026 de la BnF autour de l'intelligence artificielle se concentre sur la reconnaissance automatique des écritures manuscrites (HTR), dont les technologies et

outils arrivent aujourd'hui à maturité.

Au-delà de la transcription du texte, ces technologies permettront de naviguer à l'intérieur du document et de reconnaître les zones de contenu sur la page (segmentation), d'extraire la structure logique du document grâce à la technologie OLR (*Optical Layout Recognition*), d'identifier les entités nommées (REN). La catégorisation de ces objets dans des classes permettra d'améliorer les fonctionnalités de consultation et de présentation de Gallica.

L'interface de consultation de Gallica a fortement évolué depuis sa création. Décrits dans un format Dublin Core, les champs de la recherche depuis l'interface de Gallica sont moins riches que ceux du Catalogue général. Cependant, l'océrisation des documents permet une recherche plein texte, et même une recherche sémantique par proximité de termes. Les résultats ainsi obtenus peuvent être analysés, comparés et extraits grâce au rapport de recherche, dans des formats divers. Grâce aux API (IIIF, API documents, métadonnées), en service depuis 2017, les usagers peuvent extraire les contenus de Gallica à distance et lancer des requêtes sur les corpus. Si les images représentent un corpus important dans Gallica, il n'est pour l'instant pas encore possible de faire de la recherche iconographique à l'intérieur des documents, comme dans les corpus de presse, par exemple. Des prototypes, comme par exemple *GallicaPix* et le programme *GallicaSnoop*, développé en partenariat avec l'INA-Inria, travaillent à transformer la bibliothèque numérique Gallica en banque d'images grâce à des technologies de segmentation et de reconnaissance de forme. Ces outils, déjà opérationnels ou en phase de développement, ont été pensés pour aider l'utilisateur-chercheur, professionnel des bibliothèques ou simple lecteur - à naviguer au mieux dans les collections riches et complexes de la BnF. Cependant, la maîtrise et l'appropriation des collections numériques, soumises à la coordination de multiples expertises (sur les collections, les catalo-

gues, les outils d'extraction, les formats de données), imposent aux chercheurs comme aux professionnels des bibliothèques une évolution de leurs pratiques.

C'est pour répondre aux nouveaux usages et besoins des chercheurs et assurer une coordination des expertises à l'échelle de l'établissement que la BnF a mis en place un nouveau service, le BnF DataLab.

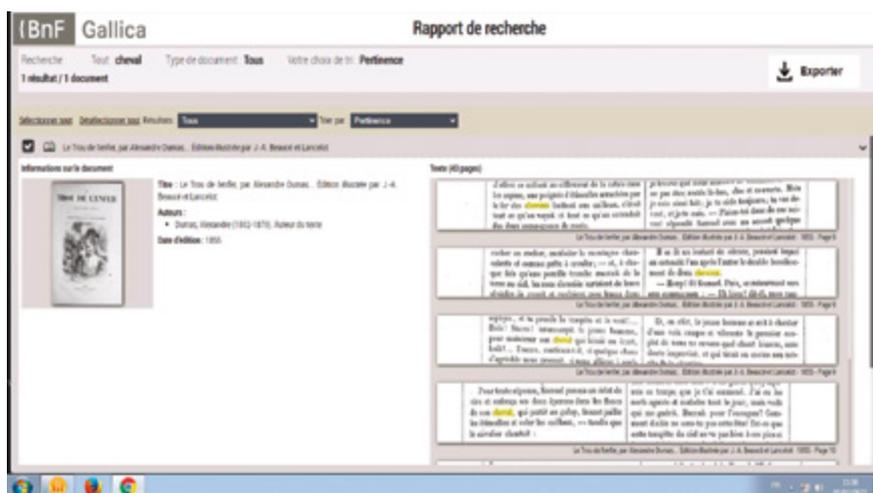
## LE BNF DATALAB, UN LABORATOIRE EN RECHERCHE ET DÉVELOPPEMENT

Inauguré le 18 octobre 2021, le BnF DataLab est à la fois un espace physique et un ensemble de services destinés à accompagner les chercheurs qui souhaitent travailler sur les collections numériques de la BnF. Plus qu'un simple point d'information ou une offre de services, le BnF DataLab a été imaginé comme un laboratoire, un lieu d'échange, d'interaction et de formation entre pairs, permettant des expérimentations et des collaborations entre équipes de recherche en résidence. Il a été pensé pour travailler en partenariat avec des chercheurs de différents profils et rattachés à différentes institutions scientifiques. Le partenariat privilégié avec Huma-Num, infrastructure de recherche dédiée aux humanités numériques, a pour objectif d'articuler l'offre de services d'Huma-Num avec celle du BnF DataLab afin de permettre un accompagnement complet autour de la recherche en humanités numériques. L'accueil d'un ingénieur d'étude ainsi que l'appel à projet conjoint Huma-Num/BnF lancé en juin 2021 sont les premières formes concrètes de ce partenariat.

Ce partenariat, comme ceux qui suivront, marque la volonté de valoriser les résultats de la recherche. Les outils développés dans le cadre de programmes de recherche ou d'appels à projets seront conservés dans une sorte de boîte à outils et réutilisables pour d'autres projets.

## CRÉER UNE COMMUNAUTÉ DE RECHERCHE AUTOUR DES COLLECTIONS DE LA BNF

L'enjeu est de créer une communauté de recherche, dans un lieu identifié au sein de la BnF et autour de ses collections, où chercheurs, bibliothécaires et ingénieurs peuvent échanger et porter ensemble des projets grâce à des expertises complémentaires. Une programmation de manifestations scientifiques (atelier de démonstration d'outils, colloques, séminaires, valorisation de travaux) a été mise en place dès 2020



➔ Le rapport de recherche dans Gallica

pour faire vivre cette communauté et favoriser les échanges entre les différentes cultures professionnelles.

En partant des cas d'usages, un catalogue d'une vingtaine de services, basé sur le cycle de vie des projets de recherche, a été élaboré pour accompagner les projets depuis la formulation du besoin, la constitution et le travail sur le corpus jusqu'à la valorisation des résultats.

Les services autour de l'accueil et de l'orientation constituent une phase cruciale : définir une problématique en termes de recherche bibliographique, déterminer ce qui est disponible, ce qui ne l'est pas, comment le rendre disponible et sous quel format sont autant d'étapes indispensables dans la construction du corpus. Les collections et les catalogues étant complexes, établir un corpus requiert souvent des recherches croisées et nécessite déjà plusieurs niveaux d'expertise. Il importe donc de centraliser les demandes pour construire un parcours cohérent et mobiliser les bons services. Le BnF DataLab est donc avant tout un service de coordination, qui doit permettre de coconstruire un parcours avec les chercheurs et les services internes : aide à la recherche, formation aux outils d'extraction des documents ou des données, à la manipulation des formats.

Une fois le corpus constitué, il paraissait important d'offrir un environnement de travail adapté à la fouille de données, en particulier pour les corpus sous droits (archives du Web, Gallica Intramuros, jeux vidéo) qui ne peuvent être consultés que dans les espaces recherche de la BnF. Un portail dédié, le DataLab Center, a été conçu pour empêcher la fuite des données sous droit, tout en autorisant les usagers à accéder à une

machine virtuelle (VM), à stocker et travailler sur les corpus. Un accès multimode offre aux usagers du BnF DataLab la possibilité de basculer aisément d'une VM sans accès à Internet, lieu de stockage du corpus, à une VM connectée à l'internet, permettant ainsi la mise à jour des outils ou le chargement de bibliothèques de scripts.

Si les humanités numériques transforment profondément le travail des chercheurs, elles modifient tout autant le métier des professionnels des bibliothèques. Introduire les humanités numériques dans la bibliothèque est un défi qui nécessite à la fois de nouvelles compétences métiers (experts, ingénieurs, etc.) et de nouvelles infrastructures en termes d'espaces et d'équipements : serveurs, machines virtuelles, logiciels, déploiement d'API.

C'est une nouvelle génération de services en bibliothèques que préfigure le BnF DataLab, articulant espace physique et virtuel, accueillant des pratiques mixtes, individuelles et collectives, tournées à la fois vers l'exploration des ressources numériques et la production de connaissance et d'outils. Cette évolution des usages et des métiers, qui ne concernent encore qu'un nombre restreint de chercheurs et de bibliothèques, n'en change pas moins le rapport aux collections et à l'espace même de la bibliothèque qui, à l'instar du BnF DataLab, se réinvente au plus près des besoins des usagers.

**MARIE CARLIN**

Conservatrice, coordinatrice du BnF DataLab  
marie.carlin@bnf.fr

**ARNAUD LABORDERIE**

Chef de projet Gallica, département de la  
Coopération, Bibliothèque nationale de France  
arnaud.laborderie@bnf.fr

L'objectif de BaOIA est de développer, à partir de corpus numérisés de La contemporaine et de la BnF, des outils d'exploration et d'analyse de corpus massifs, de les documenter et de les mettre à disposition des chercheurs pour une réutilisation ultérieure.

## BaOIA : une « boîte à outils » pour explorer les corpus numérisés



**Le projet de recherche Boîte à Outils d'Intelligence Artificielle (BaOIA), financé par CollEx-Persée,** est né de la rencontre entre les intérêts d'un chercheur pour les humanités numériques et ceux d'un établissement documentaire disposant de nombreux corpus numérisés. Quand Julien Schuh du Centre des sciences des littératures en langue française (CSLF) de l'université Paris Nanterre lui a soumis son projet de développement d'outils d'analyse de données massives, La contemporaine a saisi l'opportunité de valoriser ses collections numérisées et de développer un axe de sa propre politique de recherche.

La contemporaine poursuit de très longue date une mission d'appui à la recherche. La mise à disposition de collections spécialisées permettant d'écrire l'histoire des relations internationales et l'histoire politique et sociale depuis le début du XX<sup>e</sup> siècle en constitue la clé de voûte : monographies, revues, presse généraliste ou à diffusion restreinte, affiches, brochures et tracts, photographies, archives privées et audiovisuelles, dessins ou peintures, sont ainsi depuis 1917 collectés et signalés, sans hiérarchie, en tant que *matériaux* pour l'écriture de l'histoire. La volonté de mettre ces sources à disposition du plus grand nombre et de faciliter leur confrontation ont présidé à la mise en ligne d'une bibliothèque numérique, l'Argonnaute<sup>1</sup>, devenue l'une des plus importantes de l'Enseignement supérieur. Ce même enjeu a guidé le projet de construction d'un nouveau bâtiment qui a ouvert ses portes en octobre 2021 : toutes les collections sont désormais consultables dans une même salle de lecture, quel que soit leur support. La contemporaine continue de placer la notion de *matériaux pour l'histoire* au centre de sa réflexion sur l'adaptation de sa politique documentaire aux évolutions induites par le numérique natif, et au centre de sa politique de recherche. Elle participe à des programmes de recherche, en suscite parfois, autour de ses collections et de ses expositions. En tant que membre fondateur du Labex « Les passés dans le présent », porté par l'université Paris Nanterre, elle participe à des projets de recherche dans le champ des humanités numériques : le projet ModOAP (Modèles et Outils d'Apprentissage Profond) poursuit des objectifs voisins de ceux de

BaOIA et associe l'équipe de recherche du CSLF et la BnF, également partenaires du projet BaOIA.

### RENOUVELER ET DIVERSIFIER L'EXPLOITATION DES COLLECTIONS NUMÉRISÉES

Le projet de recherche Boîte à Outils d'Intelligence Artificielle<sup>2</sup> a été soumis à CollEx-Persée début 2020. Il s'agit de développer, à partir de corpus numérisés de La contemporaine et de la BnF, des outils d'exploration et d'analyse de corpus massifs, avec une focale particulière sur l'exploitation de l'image, de les documenter et de les mettre à disposition des chercheurs pour une réutilisation ultérieure sur d'autres ensembles de données. Le projet offre l'opportunité de diversifier et de renouveler l'exploitation des collections numérisées, au-delà de leur mise à disposition du public dans les bibliothèques numériques et à travers des modalités de valorisation comme les expositions virtuelles, dans une dimension de recherche active. Les corpus utilisés sont de nature différente afin d'exploiter textes et images : des guides touristiques de la BnF ; un fonds d'affiches françaises et internationales de La contemporaine ; des dépêches d'agences de presse soviétiques provenant des fonds de La contemporaine ; un corpus de presse illustrée et d'estampes satiriques (BnF, La contemporaine, bibliothèque de Heidelberg).

À partir de ces corpus ont été développés des outils correspondant à plusieurs objectifs : constitution des corpus, étude du contenu des documents, enrichissement et visualisation des données. Les outils se présentent sous la forme de *notebooks* Python et s'utilisent *via* l'environnement Google Colab. Actuellement, les outils sont hébergés sur GitHub, et sont accessibles depuis le site du projet BaOIA, où l'on trouve des tutoriels qui précisent leur utilisation et des exemples de réalisations tirés des tests effectués sur les corpus.

### PLUSIEURS SÉRIES D'OUTILS COMPLÉMENTAIRES

La première série d'outils est consacrée à l'extraction des données et à leur transformation pour les rendre exploitables. Des scripts permettent ainsi d'extraire

[1] <https://argonnaute.parisnanterre.fr>

[2] <https://baويا.huma-num.fr>

depuis Gallica des documents au format texte (s'il s'agit d'un document ocrisé) ou image (JPEG ou TIFF). Un *scraper* développé pour la bibliothèque numérique d'Heidelberg est réutilisé pour récupérer le texte brut ocrisé, mais aussi les illustrations et les métadonnées. Un autre outil d'ocrisation de JPEG permet d'en extraire du texte brut. Grâce à ces différents outils d'extraction ou de transformation vers un autre format, développés ou adaptés pour le projet, les chercheurs pourront constituer des corpus de textes ou d'images exploitables par une deuxième série d'outils, développés quant à eux pour étudier le contenu d'un document et enrichir ses métadonnées. Un outil de reconnaissance des entités nommées permet ainsi, à partir d'un fichier texte, de repérer en plusieurs langues des personnes, lieux, organisations, événements ou œuvres d'art, et d'effectuer des calculs statistiques sur les résultats obtenus (totaux, calcul de proportions). L'enrichissement des métadonnées est rendu possible à l'aide d'un outil qui récupère des informations sur ces entités par une requête *via* Wikidata. Enfin, la dernière série d'outils permet une visualisation des données en les cartographiant. La première cartographie s'effectue par requête Wikidata pour trouver des coordonnées géographiques à partir d'une liste de lieux. Les outils permettent également de créer des cartes interactives à partir du type de lieu (monument, ville, etc.) et de tracer des parcours entre les différents lieux. Ces outils sont utilisables indépendamment les uns des autres, selon les objectifs et besoins des chercheurs.

## UNE OUVERTURE VERS DE NOMBREUSES DIMENSIONS EXPLORATOIRES

À l'heure actuelle, les outils sont développés mais leur exploitation sur les corpus et la documentation sur leur utilisation ne sont pas achevées : il s'agit de la deuxième phase du projet, toujours en cours. Toutefois, le site du projet donne déjà un aperçu des possibilités de leur utilisation combinée<sup>3</sup> : à partir du corpus des guides touristiques, il présente les différentes opérations réalisables, de l'extraction des documents à la visualisation des données par une cartographie. Les outils créés dans le cadre de BaOIA ont aussi été appliqués à un corpus de romans scolaires de la BnF, exploité par le programme ModOAP<sup>4</sup>. Ils ont permis de réaliser une cartographie des lieux évoqués dans ces manuels. Un autre exemple d'application est visible dans l'exposition *Elie Kagan, photographie indépendant*, en cours à La contemporaine : les outils d'analyse

ont permis d'explorer la publication des clichés du photographe dans la presse, leurs occurrences et leurs réutilisations, et de proposer aux visiteurs un outil de visualisation des résultats (cartographie des sujets photographiés, exploration des thèmes récurrents).

S'il est trop tôt pour dresser le bilan d'un projet en cours et de mesurer pleinement les capacités des outils développés et leurs apports à la recherche, BaOIA confirme une dimension exploratoire riche d'enseignements : exploration de corpus numérisés dont on perçoit des dimensions encore inexploitées ; exploration de dynamiques de travail renouvelées qui mettent à profit les compétences croisées de chercheurs, d'ingénieurs et de bibliothécaires autour de projets expérimentaux, dans des ateliers où les objectifs des uns et des autres sont discutés et mis en œuvre, les outils testés collectivement, adaptés et expliqués pour devenir aussi accessibles que possible ; exploration, enfin, de nouvelles modalités d'exposition et d'exploitation des données, pour la recherche et pour le grand public à travers des initiatives comme celle de la borne interactive qui clôt l'exposition *Elie Kagan*. Les potentialités d'exploitation par les humanités numériques des données conservées par les bibliothèques sont vastes, et nous n'en sommes sans doute encore qu'à une exploration de surface.

**CÉCILE TARDY**

Directrice-adjointe de La contemporaine  
cecile.tardy@lacontemporaine.fr

**CHLOÉ JEAN**

Responsable de l'informatique documentaire  
chloe.jean@lacontemporaine.fr

[3] <https://baويا.huma-num.fr/contact/tutoriel-complet-de-lextraction-documentaire-a-la-cartographie>

[4] <https://modoap.huma-num.fr>

← La Contemporaine - Analyse d'une Une de presse



À l'université de Genève, les projets en humanités numériques suivent trois axes principaux : l'étude numérique des circulations, une expertise sur le traitement des images, et l'accent sur la dimension multiscalaire de toute recherche.

## Circulations des savoirs, de la recherche à l'enseignement: *une approche genevoise*



**Si les humanités numériques sont parvenues à se constituer en discipline, leur futur reste à écrire :**

après une première phase qui a vu le texte placé au centre des préoccupations, vient le temps de diversifier les objets d'étude en s'ouvrant davantage à l'image, statique comme mouvante, ou au son. Il est temps aussi de démocratiser l'accès à la machine, donc de réfléchir sur la place de la technique dans les humanités numériques elles-mêmes. L'approche genevoise répond à ces défis. Elle s'inscrit dans une stratégie systématique en faveur du numérique à l'université de Genève, concrétisée, entre autres, par la création d'une chaire en humanités numériques en 2019. S'il s'agit d'ouvrir de nouveaux horizons de recherche et de développer des solutions techniques innovantes, un tel processus s'incarne dans des projets concrets, qui suivent à Genève trois axes principaux : l'étude numérique des circulations, une expertise sur le traitement des images, et l'accent sur la dimension multiscalaire de toute recherche.

### UNE APPROCHE DU FAIT CULTUREL PAR LES CIRCULATIONS

L'ampleur de la littérature scientifique sur les circulations culturelles témoigne d'un intérêt ancien pour les questions de réemploi, d'allusion et citation, de transmission culturelle ou de culture partagée. Les ordinateurs, par leur capacité à gérer le temps long et l'espace, ont rapidement été mobilisés pour analyser et visualiser les informations indispensables qui permettent d'identifier les origines de certaines circulations culturelles, leurs acteurs, leurs mutations, leurs routes et leurs impasses, leurs logiques structurelles. Les textes, les idées, les objets, les images, les musiques, les personnes, les motifs circulant abondamment, la question des circulations est un objet d'étude transversal parfait pour construire une communauté de travail.

Le projet Artl@s<sup>1</sup>, fondé en 2009 à l'École normale supérieure de Paris et désormais localisé à Genève, cartographie la circulation mondiale des œuvres d'art à partir de traces textuelles, en particulier celles des catalogues d'expositions. Quoique rarement reproduites, les œuvres y sont amplement décrites (noms des auteurs présumés, titres, dates, lieux

de naissance et adresses, médium, propriétaire ou prix). Autant d'informations qui ont permis de montrer l'absence de centres uniques dans la circulation internationale de l'art, donc le côté mythique (géopolitiquement efficace) du récit moderniste selon lequel Paris aurait été le centre mondial de l'innovation artistique jusqu'en 1945, puis New York après la guerre.

Des méthodes comparables sont appliquées pour le projet Katabase<sup>2</sup> sur l'histoire des manuscrits. Katabase et Artl@s, dont les bases de données sont le fruit d'un *crowdsourcing* de grande ampleur, partagent une réflexion et des méthodes communes pour semi-automatiser la récupération du contenu textuel de catalogues (de vente ou d'expositions), en s'appuyant sur la mise en page très codifiée de ces derniers.

L'intérêt d'une approche numérique des circulations n'est pas seulement de gérer de grands corpus dans l'espace et le temps ; il est aussi de dénationaliser les propos, de déhiérarchiser les objets d'étude, donc de faciliter la sortie de réflexes nationalisateurs et de jugements de valeurs fréquents dans l'étude des circulations culturelles.

### UNE PLACE NOUVELLE POUR LES IMAGES

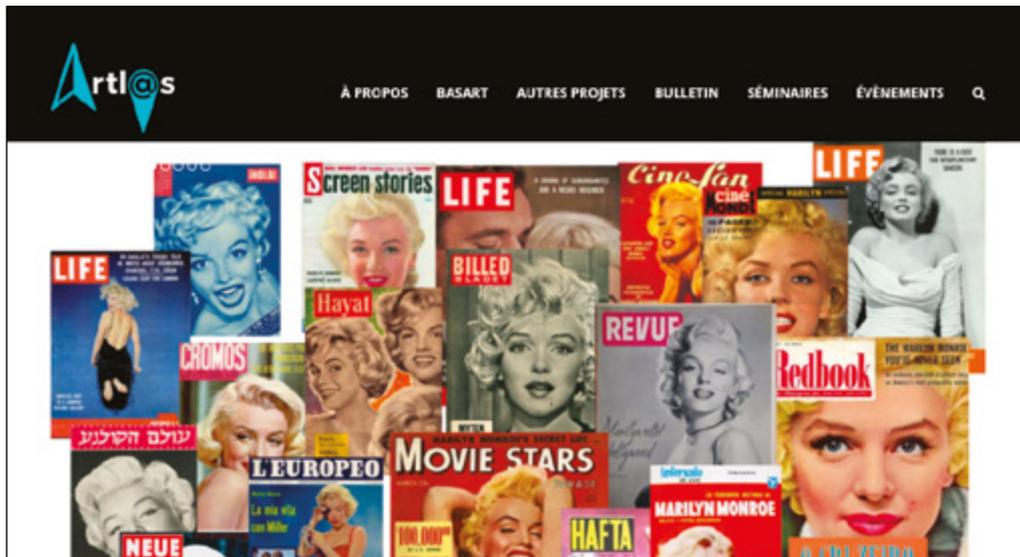
La fameuse vision artificielle a ouvert de nouvelles opportunités à la recherche, de même que la disponibilité de machines et de techniques plus puissantes pour le traitement de données. Ainsi peut-on travailler désormais sur les images en grande quantité, en deux ou trois dimensions, illustrées ou représentant des textes, qu'elles relèvent des beaux-arts, de la culture visuelle d'une époque (illustrations), mais aussi des belles-lettres (figures de style).

Le projet Visual Contagions<sup>3</sup> utilise ainsi les images à grande échelle pour mieux comprendre la mondialisation. Ce projet s'inspire des méthodologies précédemment exposées sur les circulations pour étudier la circulation des images imprimées, sur la longue période et à l'échelle mondiale. Il tire parti de la disponibilité éphémère d'imprimés illustrés du passé accessibles gratuitement en version numérisée. Grâce à un outil *ad hoc*, les illustrations de chaque page sont extraites puis comparées entre

[1] <https://artlas.huma-num.fr/fr>

[2] <https://katabase.huma-num.fr>

[3] <https://www.unige.ch/visualcontagions>



➤ Projet Artlas, capture de la page d'accueil du site <https://artlas.huma-num.fr/fr/>

elles et regroupées par proximités visuelles. Les métadonnées de la source originelle de chaque image (titres, dates, lieux de publication), permettent ensuite de mettre en évidence certaines circulations, avant des approfondissements selon d'autres échelles et méthodes.

L'image contient encore souvent du texte, un état de fait que le projet FONDUE<sup>4</sup> a pour ambition d'aborder techniquement comme historiquement. Il s'agit de déployer une infrastructure capable d'aider les chercheurs à transcrire, mais aussi à enrichir la transcription pour comprendre, par exemple, ce qui peut relier sémantiquement les contenus textuel et pictural, dans le temps et dans l'espace. Un tel outil, et les méthodes qu'il rend possibles, doivent permettre notamment de mieux comprendre l'histoire et l'esthétique visuelle des documents, partant d'une étude longitudinale d'éléments de mise en page des textes, un legs du passé relativement peu étudié à grande échelle.

## UNE INDISPENSABLE PLURIDIMENSIONNALITÉ

Le développement des études de corpus, textuels ou visuels, a entraîné une inflation de la quantité de données disponibles, elle-même alimentée par l'amélioration des techniques d'extraction. Si cette inflation justifie le tournant des humanités vers une science *data driven*, la réarticulation des méthodes quantitatives et computationnelles avec d'autres niveaux d'analyses devient plus cruciale que jamais. Les chiffres sont un mode d'accès au texte ou à l'image parmi d'autres - mode nouveau mais dont il s'agit de se détourner ensuite pour relire les textes et les images à nouveau frais selon des échelles et des méthodes plus traditionnelles. Cette lecture multiscalaire, du fait de sa complexité, se fait volontiers interdisciplinaire, d'abord par son recours à l'informatique, mais aussi aux concepts et méthodes d'analyses de disciplines connexes comme la stylistique ou la sociologie.

Pour le texte, si des analyses par « sac de mots » ont

montré leur pertinence (notamment pour l'attribution d'auteur), la lecture distante ne suffit pas toujours et doit permettre une redescende vers un niveau de lecture plus proche de l'œuvre. Dans ces grands ensembles textuels analysés, certains mots comptent plus que d'autres, que la machine peut repérer. Elle nous aide donc aussi dans la lecture proche du texte, sans la médiation des visualisations.

La « lecture » est ainsi au cœur de l'approche genevoise des humanités numériques. Une lecture outillée, pour laquelle l'informatique n'est pas une fin mais un moyen, et qui s'attache à donner sens aux artefacts du passé qui nous sont parvenus (tableaux, romans, archives, images). Cette lecture, prenant un sens différent chez les philologues, les historiens, les historiens de l'art, les linguistes, requiert des compétences disciplinaires qu'il s'agit de numériser, sans faire du degré de *maestria* informatique un critère de qualité : seule compte notre capacité à répondre de manière pertinente à une question de recherche pertinente.

Analystes des circulations artistiques au niveau scientifique, il nous paraît important de devenir acteurs enfin d'une autre circulation : celle, pédagogique, de la maîtrise de la machine par le plus grand nombre. Une large offre de cours est donc accessible à nos étudiants mais aussi aux chercheurs, jeunes et moins jeunes, *via* un certificat de spécialisation en humanités numériques d'un an, totalement intégré aux projets de recherche des participants. Cette autre circulation, entre recherche et enseignement, a depuis toujours été au cœur du projet de l'université : nous nous attachons à perpétuer cette tradition.

### BÉATRICE JOYEUX-PRUNEL

Professeur en humanités numériques  
à l'université de Genève  
[beatrice.joyeux-prunel@unige.ch](mailto:beatrice.joyeux-prunel@unige.ch)

### SIMON GABAY

Chargé d'enseignement  
à l'université de Genève  
[simon.gabay@unige.fr](mailto:simon.gabay@unige.fr)

[4] <https://www.unige.ch/lettres/humanites-numeriques/recherche/projets-de-la-chaire/fondue>

Répondant à un besoin croissant de gestion des données dans le domaine des sciences humaines, les formations en humanités numériques se sont multipliées au cours des dernières années. Illustration avec le master de l'École nationale des chartes et celui coaccrédité par quatre institutions d'enseignement supérieur de Lyon.

# L'enseignement des humanités numériques

## À l'École nationale des chartes

**THIBAUT CLÉRICE**, responsable pédagogique du master Technologies Numériques Appliquées à l'histoire à l'École nationale des chartes

### “ LA NUMÉRISATION PRODUIT UN DÉLUGE DE DONNÉES ”



© ENC.

#### Arabesques : quand et pourquoi a été créé ce master à l'École nationale des chartes ?

**THIBAUT CLÉRICE** : le master Technologies Numériques Appliquées à l'Histoire (TNAH) a été créé en 2006. L'École des chartes s'est alors dotée d'une équipe numérique formée d'ingénieurs ayant une expertise en informatique appliquée aux sources primaires. Le milieu des années 2000 voit se matérialiser une « révolution » numérique des pratiques en sciences humaines et dans le monde du patrimoine : une seconde version de Gallica voit le jour en 2000, *Google Books* apparaît en 2005, le Web commence à changer la manière de publier les produits de la recherche et on commence alors à parler « d'humanités numériques » en France. Face à cette révolution, l'École des chartes se rend rapidement compte d'un manque en termes de compétences sur le marché de l'emploi et donc de formation, en particulier initiale. Il existe alors très peu de personnes avec une double compétence en informatique et sciences humaines. Les institutions, de recherche et du patrimoine, se retrouvent très peu accompagnées dans cette course à la numérisation. La pérennisation des données numériques, la production de corpus, la valorisation et la standardisation des données et métadonnées requièrent des compétences rares.

#### À quels besoins répond ce master ?

**T. C.** : les besoins actuels ont évolué en même temps que les technologies. D'abord, les besoins d'aujourd'hui continuent de tourner autour de la standardisation et de la mise à disposition de données, qu'il s'agisse de données textuelles ou (audio)visuelles, nativement numériques ou non. Ensuite, après un premier temps de numérisation intensive, les besoins changent ou s'accumulent. Les institutions patrimoniales, les laboratoires de recherche ou encore les entreprises privées se retrouvent confrontés à un « déluge des données » produit par la numérisation. Enfin, toutes les institutions, privées ou publiques, commencent à prendre conscience de la valeur de leurs données, qu'il s'agisse de tableurs ou de textes : ces données, reformatées, « reparamétrées » puis exploitées peuvent devenir de véritables leviers pour elles. Notre master répond à tous ces besoins en formant de futurs professionnels pour lesquels la manipulation de données – qu'il s'agisse de valorisation ou de préparation à l'exploitation – est au cœur du métier.

#### Quel est le profil des étudiants ?

**T. C.** : les profils sont extrêmement divers mais les étudiants sont très majoritairement issus des cursus de sciences humaines, en particulier d'histoire, d'histoire de l'art et de lettres, avec sou-

vent une première formation en master. Les étudiants sont d'ailleurs souvent des étudiantes. Nous avons eu historiquement trois quarts de jeunes femmes dans nos promotions. Ils ont généralement une appétence préalable pour le numérique et cherchent une voie professionnelle autre que celles de la recherche, au sens de chercheurs, ou de l'enseignement. Dans le cadre de la formation continue, nous avons d'autres profils, notamment des personnes cherchant à changer de spécialité, venant parfois de milieux très éloignés comme celui de la banque, ou au contraire des personnes issues d'institutions patrimoniales qui veulent monter en compétence pour prendre en charge de nouvelles missions.

#### Quels types de poste trouvent-ils à l'issue de leur formation ?

**T. C.** : là encore, les postes sont très variés. Une partie de nos diplômés rejoignent les institutions patrimoniales comme responsables de mission numérique ou chargés de projet. Une autre se dirige vers l'ingénierie en appui à la recherche, dans les laboratoires comme dans les institutions patrimoniales. Quelques-uns vont vers le privé, parfois comme techniciens, plus souvent comme consultants. Au cœur de chacun de ces métiers divers, il existe une spécialité, celle de la connaissance des données et de leurs standards. ■

## À Lyon

**SABINE LOUDCHER**, responsable principale du master Humanités numériques de Lyon

### “L’ESSOR DU NUMÉRIQUE DANS NOTRE SOCIÉTÉ MODIFIE LES PRATIQUES PROFESSIONNELLES



#### **Arabesques: pourquoi a été créé le master Humanités numériques de Lyon ?**

**SABINE LOUDCHER** : l’essor du numérique dans notre société modifie les pratiques professionnelles, notamment dans le domaine des Arts, Lettres, Langues, Sciences Humaines ou Sciences Sociales (ALLSHS). Les institutions, les entreprises et les centres de recherche sont à la recherche de personnes qui sont à la fois compétentes dans l’une des disciplines relevant des ALLSHS et qui sont aussi capables de mobiliser les méthodes, les outils et les pratiques du numérique. Le master Humanités numériques de Lyon est coaccrédité par l’université Lumière Lyon 2, l’université Jean Moulin Lyon 3, l’École normale supérieure de Lyon et par l’École nationale supérieure des sciences de l’information et des bibliothèques (Enssib). Un seul parcours commun aux quatre établissements est proposé. Il est accessible uniquement en double-cursus : les étudiants sont inscrits principalement dans un master disciplinaire en Arts, Lettres, Langues, Sciences Humaines ou Sciences Sociales, ou ils sont diplômés d’un master disciplinaire, et ils sont également inscrits en humanités numériques en inscription secondaire pour obtenir une spécialisation complémentaire liée au numérique et un deuxième diplôme de master. Les cours se déroulent uniquement le lundi pour permettre aux étudiants de suivre le reste de la semaine les cours de leur master disciplinaire.

#### **À quels besoins répond ce master ?**

**S. L.** : l’objectif du master Humanités numériques est de préparer aux fonc-

tions qui nécessitent de comprendre de manière technique et critique les corpus numériques, corpus de documents, données ouvertes, données d’enquête, les outils qui les produisent, leurs représentations, leurs analyses et leurs usages. Il s’agit d’articuler les problématiques et les méthodes des ALLSHS, de la statistique et de l’informatique, pour élaborer, analyser, interpréter socialement et techniquement des corpus numériques. À l’issue du master, les étudiants sont capables de constituer un corpus de documents et de collecter des données (collecte, encodage, structuration, stockage), d’analyser et visualiser des données grâce à des méthodes de science des données, de valoriser, diffuser et pérenniser les corpus numériques, de conduire un projet de données numériques articulé avec des besoins sociétaux.

#### **Quel est le profil des étudiants ?**

**S. L.** : le master Humanités numériques s’adresse aux étudiants de licence ou de master en Arts, Lettres, Langues, Sciences Humaines ou Sciences Sociales. En première année, les cours obligatoires portent sur les bases des technologies numériques pour les humanités numériques (statistique, programmation, bases de données, technologies Web, constitution de corpus de documents textuels) ainsi que sur une dimension réflexive du numérique (histoire et écosystème des humanités numériques, anthropologie du numérique). Les étudiants doivent également choisir un cours de spécialité permettant soit d’approfondir la constitution de corpus

archivistiques soit l’analyse de données hétérogènes en ALLSHS. En deuxième année de master, les cours communs portent sur les enjeux épistémologiques, politiques, juridiques et éthiques du numérique ainsi que sur l’analyse des données textuelles, images ou en réseau. Ensuite, les étudiants ont le choix entre une spécialité permettant d’approfondir la constitution, l’encodage, l’édition et la mise en ligne de corpus archivistiques et une spécialité consacrée à l’analyse et la visualisation de données hétérogènes par des méthodes de science des données. La deuxième année comprend également une initiation à la recherche avec la réalisation d’un projet interdisciplinaire, la rédaction d’un mémoire ainsi qu’un stage obligatoire. Le stage et le mémoire sont à réaliser conjointement avec le master disciplinaire.

#### **Quels types de poste trouvent-ils à l’issue de leur formation ?**

**S. L.** : à l’interface entre les ALLSHS et les sciences du numérique, les jeunes diplômés du master en humanités numériques de Lyon travaillent dans les nouveaux métiers liés à la production, à l’analyse et à la valorisation des données numériques dans le domaine de la culture et du patrimoine, de la médiation de l’information ou du juridique. Le master peut également déboucher sur un doctorat dans une des disciplines des Arts, Langues, Lettres, Sciences Humaines ou Sciences Sociales. ■

# Des catalogues de bibliothèques aux projets en humanités numériques : les autorités IdRef font le lien

Les projets de recherche en humanités numériques font un recours croissant à IdRef, qui se déploie ainsi au-delà de son champ documentaire d'origine afin de contribuer à la production de données FAIR.



Les projets de recherche en humanités numériques font un recours croissant à l'utilisation d'IdRef, référentiel d'autorités destiné à décrire des objets documentaires. IdRef se déploie ainsi au-delà de son champ documentaire d'origine et s'incorpore avantageusement aux pratiques de recherche des sciences humaines et sociales, des arts et des lettres, afin de contribuer à la production de données FAIR (*Findable, Accessible, Interoperable, Reusable*).

## LES AUTORITÉS DOCUMENTAIRES

L'utilisation d'autorités pour une description qualitative des ressources documentaires des bibliothèques a une longue histoire. La Bibliothèque nationale de France y a eu recours pour mettre en œuvre son Catalogue général. De même, lors de sa création il y a plus de vingt ans, le Sudoc a fait le choix originel d'y recourir.

Les autorités font référence à un nombre restreint d'objets du monde réel dont le rôle est déterminant pour les documents constituant les collections des bibliothèques : Personnes, Collectivités, Familles, Lieux géographiques, Sujets et Œuvres. Plusieurs documents peuvent pointer vers une même autorité : ainsi s'opèrent simultanément économie d'échelle et fédération de contenus.

Pour illustrer ce propos, prenons l'ensemble des romans composant la « Comédie humaine » : ceux-ci sont naturellement liés à la notice de leur unique auteur : « Balzac, Honoré de (1799-1850) ». Entre autre richesse, une notice d'autorité recense les variantes de noms sous une forme normalisée (« H. de Balzac », « Honoré de Balzac »...), ce qui permet la constitution d'index de recherche à la fois riches et précis (ni silencieux, ni bruyants).

Par nature, une autorité a pour vocation l'identification sans ambiguïté de l'objet auquel elle réfère, que ce soit pour distinguer des homonymes ou pour rapprocher des pseudonymes. Au sein d'un référentiel, chaque autorité se distingue donc des autres : « Balzac, Honoré de (1799-1850) » = Lord R'Hoone ≠ Balzac, Henry de (1807-1856).

Rappelons enfin que toutes les données renseignées dans une autorité sont sourcées, établies par la citation de documents et que chaque notice

d'autorité IdRef est dotée d'un identifiant unique et pérenne qui assure la liaison avec les références bibliographiques, en s'appuyant notamment sur le format standard Unimarc.

## LA GRANDE FAMILLE DES HUMANITÉS NUMÉRIQUES

Ces caractéristiques rendent évident le fait que les disciplines académiques placées sous la bannière « humanités numériques » sont éligibles à l'utilisation d'autorités. Mieux, elles ont tout intérêt à y recourir. En effet, la documentation, en tant qu'activité spécialisée dans la description et l'accessibilité de ressources et corpus, se conçoit aisément comme une discipline inhérente à ce champ de recherche et de pratiques. Toutes ces disciplines partagent en effet une épistémologie commune fondée sur la citation des sources et le travail sur les objets documentaires, dont la quantité, la variété et la morphologie s'accroissent à l'ère numérique. Par ailleurs, les institutions culturelles telles que les bibliothèques, les musées et les archives élaborent Ides collections qui constituent des sources et données primaires pour les historiens, les chercheurs en littérature et pour toutes les disciplines des sciences humaines..

Même si la reconnaissance de cette situation n'est pas encore universellement partagée, elle est pourtant porteuse d'importants bénéfices collectifs.

## LA TRAJECTOIRE D'IDREF

Après que le Sudoc a fait le choix de cataloguer sous contrôle d'autorités, l'Abes a été chargée en 2005 de développer Calames, application spécifique au signalement des Manuscrits et Archives des bibliothèques de l'ESR, qui se devait d'intégrer cette bonne pratique. Pour ce faire, les autorités du Sudoc, jusqu'alors confinées en coulisses, se sont vues d'un seul coup exposées sur le « World Wide Web », donnant ainsi naissance à IdRef, application de réutilisation et d'édition mutualisée de ces autorités.

Ouvertes à Calames, les portes le sont depuis à quiconque souhaite réutiliser et, mieux encore, contribuer à l'enrichissement des référentiels d'autorité dont IdRef est le vecteur. C'est ce qu'a fait

Persée dès 2014<sup>1</sup>, date à laquelle la plateforme de ressources numérisées en SHS a migré sa base d'Auteurs et intégré dans son *workflow* le liage et la création d'autorités IdRef.

Au-delà du Sudoc, au-delà de l'Abes, un seuil est franchi et se profile alors la trajectoire idéale d'une utilisation d'IdRef, depuis les grands catalogues de bibliothèques en passant par les archives institutionnelles et les bibliothèques numériques jusqu'aux projets de recherche en humanités numériques.

## LES SOLLICITATIONS DES HUMANITÉS NUMÉRIQUES

En la matière, l'Abes commence à disposer d'un portefeuille de sollicitations assez garni : que ce soit spontanément ou par l'intermédiaire d'appels à projet (au premier rang desquels CollEx-Persée), leur nombre ne faiblit pas. Illustrations avec trois projets qui montrent une utilisation et une prise en main simples et efficaces des autorités IdRef à des fins de constitution d'index et d'identification fiable.

- Le projet SIPROJURIS<sup>2</sup>, Système d'information des professeurs de droit (1804-1950) a « idrefisé » son corpus de professeurs.
- Le projet PRELIB<sup>3</sup>, projet de recherche en littérature de langue bretonne, porté par le Centre de recherche bretonne et celtique (CRBC) et soutenu par la Maison des sciences de l'homme de Bretagne, a « idrefisé » l'ensemble des personnes, comme cela se voit dans l'interface d'IdRef :

Recherche sur Miossec, Yves idref 027031357

MIOSSEC, YVES (1907-2001)

CRBC Bibliographie dans PRELIB (Projet de recherche en littérature de langue bretonne)

7 références PRELIB : Tout voir dans PRELIB

(affichage limité aux 30 premières références)

- PRELIB 1916 Drest ar mor bras, Miossec Yves, Brud Nevez - Emgleo Breiz 1981
- PRELIB 1915 Eur veaj e Stadoù Unanet an Amerik, Miossec Yves, 1996
- PRELIB 1917 Lorh, kasoni ha poan, Miossec Yves, Brud Nevez - Emgleo Breiz 1983
- PRELIB 1918 Mond da Israel, Miossec Yves, Brud Nevez - Emgleo Breiz 1991
- PRELIB 1914 O haloupad war heñchou Breiz-Izel, Miossec Yves, 1987
- PRELIB 2353 Tamnou enverenou euz eun amzer griz, Miossec Yves, Brud Nevez - Emgleo Breiz 1994
- PRELIB 2254 Tamnou enverenou euz eun amzer griz, Miossec Yves, Emgleo Breiz - Fondation culturelle bretonne 2013

- Le projet Marc Michel Rey<sup>4</sup>, projet de publication de la correspondance de ce libraire du XVIII<sup>e</sup> siècle porté par l'IHRIM (Institut d'histoire des représentations et des idées dans les modernités), a « idrefisé » les expéditeurs et destinataires, donnant lieu à la génération d'environ 70 notices à partir des informations recueillies par les chercheurs, ce qui a permis de construire un index dédié au projet, comme par exemple :

INDEX DES PERSONNES

Accarias de Serionne, Jacques 073754684

Journaliste né à Die, rédacteur du Journal de commerce et d'agriculture (1759-1762).

L'intérêt croissant du recours au référentiel IdRef par les chercheurs et ingénieurs impliqués dans ce type de projet s'explique aisément par la nature même des données d'autorité. En effet, à l'aide des technologies déployées par IdRef en vue de leur « FAIRisation », chaque notice est dotée d'un identifiant unique et pérenne, pivot indispensable à l'interopérabilité entre différents systèmes d'information ; indexée, elle est aisément trouvable, requêtable et réutilisable, notamment du fait de sa mise à disposition sous licence ouverte dans les formats propres au web sémantique (RDF). De plus, les données peuvent être enrichies et sérialisées selon les besoins spécifiques de chaque projet.

## UN ACCOMPAGNEMENT RENFORCÉ

Cependant, données et outillage ne seraient pas si efficaces sans l'accompagnement que les chercheurs peuvent trouver auprès de l'Abes ou, localement, auprès des bibliothécaires et professionnels de l'information ayant développé une expertise de la structuration et de l'interopérabilité des données.

En éclaircur, l'Abes explore cette voie depuis 2019 en relation étroite avec des archéologues de la Maison de l'Orient et de la Méditerranée, pourvoyeurs de données hétérogènes, et des historiens, experts en ontologie et sémantisation des savoirs du Laboratoire de recherche historique Rhône Alpes dans le cadre du projet ANR HisArc-RDF<sup>5</sup> (pour HISToire et ARChéologie en triplets RDF) qui vise la création d'une chaîne de traitement menant à la production de données de recherche de qualité FAIR. Est récolté ainsi le fruit de l'expérience et de l'investissement dans les technologies du Web sémantique, qui s'incarne dans *data.idref*<sup>6</sup>, et pour ce projet même, dans un point d'accès *sparql* expérimental alliant accompagnement et exposition en RDF des jeux de données homogénéisées<sup>7</sup>, où se déploient dialogue, échange et acculturation réciproques et interdisciplinaires.

Continuité épistémologique, mutualisation des outils, des services et des savoir-faire, interdisciplinarité de l'effort heuristique : le chemin est encore long mais l'horizon est là qui nous guide pour que les communautés formant la grande famille des humanités numériques concourent à l'élaboration de leur grand corpus unifié de connaissances liées et ouvertes.

**FRANÇOIS MISTRAL**

Responsable IdRef-Autorités  
mistralf@abes.fr

[1] <http://info.persee.fr/lalignement-des-autorites-persee-au-referentiel-idref>

[2] <http://siprojuris.symogih.org>

[3] <https://mshb.huma-num.fr/prelib>

[4] <http://rey.huma-num.fr/presentation>

[5] <https://anr.fr/fr/lanr/engagements/la-science-ouverte/les-projets-laureats-de-lappel-flash-science-ouverte/projet-hisarc-rdf>

[6] <https://data.idref.fr>

[7] <https://dataforhumanities.abes.fr/sparql>

# Un outil de comparaison visuelle des collections numérisées de Persée et de leurs équivalents imprimés



Deux étudiants du master Humanités numériques de Lyon ont comparé le corpus de revues numérisées par Persée et les revues imprimées présentes dans les Plans de conservation partagée des périodiques.

Le 31 janvier 2022 avaient lieu les soutenance de projets du master en humanités numériques, parcours *Pratiques et méthodes en Humanités numériques*, porté conjointement par l'université Lumière Lyon 2, l'Enssib, l'université Jean Moulin Lyon 3 et l'École normale supérieure de Lyon (cf p. 15). Parmi les quatre projets présentés, celui d'Amélie Descollonges et Julien Colin propose une visualisation graphique comparant les collections numérisées par Persée (463 revues numérisées) et leurs équivalents imprimés présents dans les établissements membres du réseau Sudoc. Précisons que, dans le cadre de ce projet, le jeu de données (liste de PPN)<sup>1</sup> a été fourni par l'Abes et que le code écrit par les étudiants, prochainement disponible via le GitHub de l'Abes, est réutilisable et rejouable à volonté.

## UN TRAVAIL EN 2 VOILETS

En premier lieu, a été réalisée une rosace constituée de l'ensemble des titres Persée (1 barre = 1 titre), dans laquelle la hauteur de la barre représente le nombre de localisations de l'équivalent imprimé, et la couleur de la barre, la thématique disciplinaire de la revue.



Visualisation des localisations des périodiques numérisés par Persée répartis par thématiques

En second lieu, a été effectuée une comparaison titre à titre des états de collection de la revue papier et de la période numérisée par Persée.

Exemple avec le titre *Bulletin de l'Association française de recherche sur l'histoire du cinéma* : en vert les années 1986- 2000 auxquelles Persée donne accès ; en gris les états de collection de la revue imprimée dans les bibliothèques qui participent à un Plan de conservation partagée des périodiques. En glissant le curseur sur un des états de collection, on obtient plus d'informations sur le Plan concerné, les dates de début et de fin de l'état de collection et le nom de la bibliothèque de localisation.



Cette représentation sous forme de *timeline* se rapproche de celle de Périoscope<sup>2</sup>, avec comme valeur ajoutée la présence de l'accès Persée, absente de Périoscope : pour la même revue, on aura accès aux états de collection et aux localisations des autres établissements Sudoc possédant le titre, mais n'émargeant à aucun Plan. Le nombre de localisations pouvant avoisiner les 400 pour certains titres du corpus, cette représentation textuelle complète avantageusement la représentation graphique précédente. Un rapport d'erreur montrant les exemplaires papier sans état de collection est généré pour chaque titre.



## UN COUP D'ACCÉLÉRATEUR POUR LA QUALITÉ DES DONNÉES

Ce travail a permis à Persée et à l'Abes de collaborer étroitement à l'amélioration de la qualité des données descriptives du corpus Persée. À partir du fichier KBart Persée disponible *via* la base de connaissance Bacon<sup>3</sup>, des vérifications d'identifiants ont été effectuées en interrogeant le Sudoc : correspondances entre les PPN et ISSN de Persée, les PPN et ISSN imprimés. Ce corpus d'environ 450 titres a également été comparé avec le corpus Persée signalé dans Mir@bel.

Ainsi, certaines erreurs ont pu être identifiées et corrigées : demande d'attribution d'ISSN papier et électroniques manquants, rectification du fichier KBart Persée (suppression des doublons, introduction des manquants, alignements d'identifiants).

## LES PISTES DE TRAVAIL À VENIR ENTRE L'ABES ET PERSÉE

L'élan donné par la qualité du travail des étudiants devrait permettre de poursuivre l'amélioration réciproque de la qualité des données : établir davantage de passerelles entre les outils de l'Abes et Persée (qui utilise déjà IdRef) ; signaler dans le Sudoc les accès Persée ; améliorer la qualité des notices bibliographiques par des chantiers qualité sur le corpus de revues numérisées Persée (ajout d'indexation sujet, d'autorités collectivité ou auteurs, liens aux métarevues).

### L'ÉQUIPE PROJET<sup>4</sup>

[1] *Pica production number* identifiant univoque pour chaque titre

[2] <https://periscope.sudoc.fr>

[3] <https://bacon.abes.fr>

[4] Julie Mistral et Morgane Parra, Service Ressources continues de l'Abes ; Sitthida Samath, responsable Qualité et Interopérabilité des Données de Persée ; avec l'aide d'Amélie Descollonges et Julien Colin, étudiants en master Humanités numériques de Lyon.

**AURÉLIEN BERRA**, rédacteur  
en chef de la revue *Humanités numériques*



humanités  
numériques

# LES HUMANITÉS NUMÉRIQUES DÉVELOPPENT DES INTERROGATIONS NOUVELLES

Créée en 2020, la revue *Humanités numériques*<sup>1</sup> veut illustrer de manière concrète cette discipline en offrant aux praticiens un cadre d'expression pour relayer leurs expérimentations.

**Arabesques : quand a été créée la revue *Humanités numériques* et avec quelles finalités ?**

**AURÉLIEN BERRA :** *Humanités numériques*, première revue francophone dédiée à ce sujet, a été créée à l'été 2020, avec le soutien notamment du CNRS et d'OpenEdition. Elle est une émanation d'Humanistica, association francophone des humanités numériques, elle-même fondée en 2014. L'objectif était de contribuer à structurer plus fortement l'association – qui organise désormais son propre colloque – et d'illustrer très concrètement, par l'exemple, ce que sont les humanités numériques. En effet, cette étiquette apparaît de plus en plus fréquemment dans toutes sortes de contextes mais sans qu'on sache toujours très bien ce que cela recouvre et qui cela concerne. Le but principal est de donner un cadre d'expression aux personnes se considérant comme des praticiens des humanités numériques quels que soient leur appartenance et leur statut au sein des équipes de recherche, leur discipline de compétence, avec une forte dimension transdisciplinaire. Un historien peut se sentir concerné par une problématique développée dans un projet mené en linguistique ou en archéologie. Nous avons deux publications par an alternant les numéros thématiques, par exemple les humanités numériques spatialisées, et les numéros libres.

**Observez-vous des tendances à travers les articles que vous publiez ?**

**A. B. :** on voit une dominance de l'histoire et des études textuelles, ce qui est logique car ce sont des champs qui comptent beaucoup de chercheurs et d'étudiants et dans lesquels on observe un renou-

vement des méthodes de recherche. Pour les lecteurs, qui pour la plupart ne sont pas des spécialistes de la discipline abordée dans un article, il est intéressant de découvrir l'approche développée pour résoudre une problématique.

**Quels sont aujourd'hui les principaux enjeux autour des humanités numériques ?**

**A. B. :** je crois que le principal enjeu est, dans la poursuite de l'institutionnalisation et de la professionnalisation, de conserver un épicerie d'inventivité et des lieux de dialogue interdisciplinaire. Les humanités numériques peuvent donner lieu à des réalisations très cadrées, des éditions numériques en ligne pérennes, par exemple, mais aussi à des innovations avec des prototypes qui sont des propositions intellectuelles. Elles réinventent de l'intérieur les métiers, développent des interrogations nouvelles et de nouveaux moyens d'étudier des corpus, mais souvent dans un écosystème contraint par les problèmes de financement, la logique d'efficacité budgétaire et la précarité des personnels de l'enseignement supérieur. Or, pour mener un travail exploratoire, critique et collaboratif, il faut une certaine pérennité de financement et de personnel. C'est crucial, car un ingénieur qui a déjà suivi plusieurs projets en humanités numériques apporte une expérience et une compétence que ne pourra pas acquérir en quelques mois une personne qui arrive et qui découvre les problématiques.

**Quel rôle peuvent jouer les bibliothèques dans cette discipline ?**

**A. B. :** dans les pays anglophones, notamment, les bibliothèques sont souvent le noyau des recherches en humanités numériques, car elles disposent des compétences et appétences techniques ainsi que de la stabilité, notamment financière, nécessaires à ce type de projets. C'est moins le cas en France pour l'instant : il faudrait s'en donner les moyens ! Les bibliothèques peuvent être aussi des lieux de rencontre et de diffusion : les humanités numériques reposent sur un maniement de l'information, ce qui est justement le champ de compétences des professionnels des sciences de l'information et de la documentation. Elles contribuent également en offrant des formations et des espaces de pratique. De fait, contrairement à une idée reçue, les étudiants et jeunes chercheurs ne sont pas tous à l'aise avec les humanités numériques.

**Quelles sont les perspectives d'évolution de la revue ?**

**A. B. :** au-delà de l'élargissement progressif des thématiques, nous sommes curieux de voir comment évolueront certains axes déjà représentés. Par exemple, l'histoire des humanités numériques francophones, les présentations réflexives de projets ou encore les articles de présentation de données (*data papers*). En outre, il est important pour nous d'intégrer des liens vers d'autres articles, des blogs ou des plateformes de données. La revue ne se pense pas comme isolée mais ouverte, à travers le Web, à toutes sortes de modes de présentation de la recherche en train de se faire. ■

[1] <https://journals.openedition.org/revuehn>

Face aux défis de la diversité des publics et de son implantation dans quatre pays, la médiathèque de l'Institut français du Proche-Orient développe une stratégie de partenariats et de valorisation numérique de ses collections, couvrant les recherches en sciences humaines et sociales sur le Proche-Orient depuis le début du XX<sup>e</sup> siècle.

## Des fonds patrimoniaux à l'actualité de la recherche : une médiathèque régionale au Proche-Orient

**L'Institut français du Proche-Orient (Ifpo) est un institut de recherche en sciences humaines et sociales sous cotutelle du ministère de l'Europe et des Affaires étrangères et du CNRS.** Il accueille, au sein de trois départements scientifiques (archéologie et histoire de l'Antiquité, études arabes médiévales et modernes, études contemporaines) des chercheurs et doctorants travaillant sur les pays où est implanté l'Institut : le Liban, la Jordanie, l'Irak, les Territoires palestiniens et la Syrie.

À l'image de l'Ifpo, sa médiathèque est régionale et constituée de quatre bibliothèques localisées à Damas, Beyrouth, Amman et Erbil. Chacune bénéficie d'un fonds spécifique à sa localisation, mais aussi aux thématiques de recherche, plus ou moins durables ou récentes, des chercheurs en poste. Ainsi, le fonds de la bibliothèque d'Erbil est en cours de constitution sur le Kurdistan irakien, celui d'Amman est spécialisé sur la Jordanie avec une forte thématique sur les questions de migration et de réfugiés, tandis que la bibliothèque de Beyrouth comprend un fonds de référence sur l'archéologie et l'histoire antique de la région, un fonds sur le Liban contemporain spécialisé sur l'urbanisme et la reconstruction du Liban post guerre civile, auquel est associée une importante collection de cartes. Centenaire cette année et malheureusement fermée au public depuis 2011, la bibliothèque de Damas conserve le plus grand nombre de documents. Son fonds lié aux études arabes médiévales et modernes, ses fonds photographiques et sa cartothèque mandataire sont des collections patrimoniales importantes. L'ensemble des collections a reçu en 2019 le label CollEx-Persée.

Depuis 2011, Koha gère le catalogue en ligne<sup>1</sup>, mais seul le fonds de Beyrouth y est entièrement catalogué. Le déploiement dans le Sudoc s'est fait en 2013. Avec aujourd'hui 45 500 notices bibliographiques dans le Sudoc et 10 500 notices en caractères arabes,

une professionnalisation croissante de l'équipe orchestrée de Beyrouth, 8800 Unica majoritairement issus de la rétroconversion en cours du fonds de Damas, l'Ifpo fraie son chemin dans le réseau de l'Abes. En 2020, l'institut a adhéré à WorldCat afin de mieux faire connaître ses collections hors de France. Cette ouverture nous amène à programmer le rétrocatalogage dans le Sudoc du fonds urbanisme de la bibliothèque de Beyrouth. Le projet de passer toute mise à jour sur WinIBW, l'outil de catalogage du Sudoc, s'inscrit aussi dans l'objectif de dégager du temps pour la mise en valeur des documents.

### LES DÉFIS D'UNE MÉDIATHÈQUE RÉGIONALE

Dans une région à conflits réguliers, la configuration multisite a permis de s'adapter grâce à des déplacements de collections ou de direction : de Beyrouth vers Damas pendant la guerre civile libanaise, puis de Damas à Beyrouth en 2011 avec le début de la révolte syrienne. La réactivité des collègues de tous les sites et leur collaboration permettent, en dépit de conditions de travail de plus en plus difficiles, de suivre et mettre en valeur les fonds, y compris à

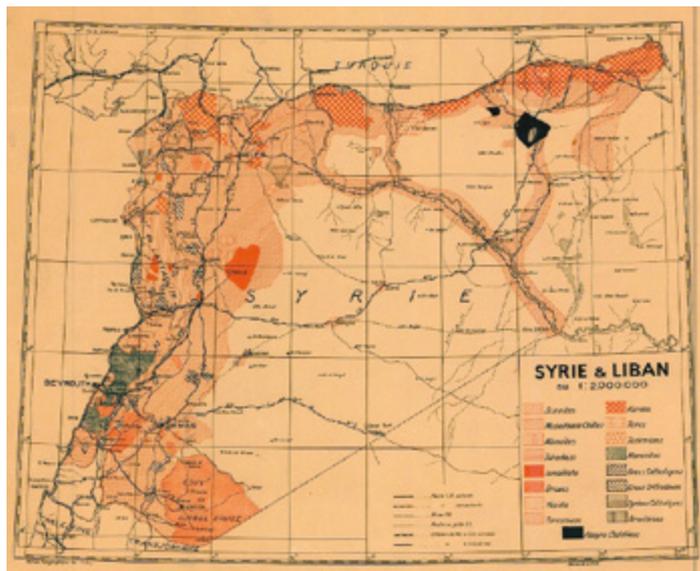
Damas où les collègues restés sur place participent à la rétroconversion et effectuent un travail de vérification précieux pour les correspondants Sudoc de l'Ifpo.

Les bibliothèques de l'Ifpo, dont l'accès est gratuit, servent des publics dont les thématiques de recherche sont variées et ne cessent d'évoluer. Dans sa politique documentaire et dans son fonctionnement, la médiathèque de l'Ifpo doit se prêter à un jeu d'équilibre pour répondre à ces différents besoins et continuer à actualiser ses fonds, cela en tenant compte d'une configuration multisite coûteuse en termes d'acquisitions et de circulation des documents entre les antennes. La répartition des fonds entre les antennes peut en effet s'avérer restrictive du point de vue de la pratique des usagers dont les recherches concernent souvent l'ensemble de la région. Depuis 2011, les recherches sur la Syrie se font par exemple aussi bien depuis le Liban que depuis la Jordanie. À ce titre, les partenariats et les échanges avec les bibliothèques locales ainsi que l'évolution vers des contenus numériques sont des solutions intéressantes pour accompagner au mieux notre public dans ses recherches.



➔ FIG.1 Syrie, Jebel Sem'an, Barad, ca 1960. Photothèque archéologique de l'Ifpo, <https://medihal.archives-ouvertes.fr/medihal-01860825v1>

FIG.2 Syrie, Jebel Barisha, Me'ez, ca 1960. Photothèque archéologique de l'Ifpo, <https://medihal.archives-ouvertes.fr/IFPOIMAGES/medihal-01846538v1>.



➤ Carte Syrie et Liban : [Répartition de la population par confessions], novembre 1935. Cartothèque de l'Ifpo. <https://gallica.bnf.fr/ark:/12148/bpt6k1290552k>



➤ Revue satirique Al-mudhik al-mubki, Damas, 1930. Médiathèque Ifpo. <https://gallica.bnf.fr/ark:/12148/bpt6k12905085>



Grâce aux publications des Presses de l'Ifpo, la médiathèque maintient par exemple des échanges alimentant une collection conséquente de périodiques en archéologie et histoire antique. Localement, un Prêt entre Bibliothèques avec des universités locales, comme l'université Saint-Joseph de Beyrouth et l'*American University of Beirut* au Liban, offre aux usagers une documentation plus large, comblant la fermeture de Damas. Depuis 2020, nos abonnements papier sont doublés d'un accès en ligne permettant à l'antenne d'Amman d'accéder aux périodiques achetés par Beyrouth. Dans le même but, le portail documentaire BibCnrs<sup>2</sup> est valorisé et les premiers *ebooks* sont en cours d'achat.

Aidée par la conjoncture, la médiathèque se concentre sur ses principales missions : l'accueil, la valorisation des fonds, la communication de la documentation *via* notamment le développement des services distants, le signalement des ressources numériques et leur numérisation et la veille documentaire.

### VERS UNE CONSERVATION ET UNE VALORISATION NUMÉRIQUE DES FONDS

Dans le cadre de la stratégie en faveur de la science ouverte menée à l'Ifpo depuis près de 15 ans<sup>3</sup>, les projets de conservation et de valorisation numérique des collections scientifiques tiennent une place importante. Grâce aux partenariats et aux outils d'accompagnement de la recherche, la médiathèque et le service des humanités numériques œuvrent ensemble à l'archivage numérique

et la mise en ligne, en accès ouvert, de ces collections.

Depuis 2016, l'Ifpo est partenaire de la Bibliothèque nationale de France dans le projet Bibliothèques d'Orient<sup>4</sup> qui vise, en s'appuyant sur Gallica, à créer une bibliothèque numérique collaborative rassemblant le patrimoine documentaire de bibliothèques patrimoniales au Moyen-Orient. La thématique et les bornes chronologiques du projet (les échanges entre la France, l'Europe et le Moyen-Orient de la fin du XVIII<sup>e</sup> au milieu du XX<sup>e</sup> siècle) permettent d'archiver et de rendre accessibles des fonds de l'époque mandataire dont l'Institut est dépositaire. C'est le cas de la revue satirique arabe *Al-Mudhik al-Mubki*<sup>5</sup>, ou bientôt d'une sélection de cartes de la région. Un partenariat avec le projet LiPoL, financé par l'Agence nationale de la recherche, permettra d'archiver les photographies de manuscrits de littérature populaire arabe, aujourd'hui disparus ou inaccessibles<sup>6</sup>. De tels projets sont d'autant plus riches qu'ils permettent l'implication des chercheurs dans la sélection documentaire. Les archives scientifiques produites par les chercheurs et les publications des Presses de l'Ifpo, qui constituent souvent l'aboutissement de leurs projets, sont des fonds patrimoniaux importants qui participent à la mémoire de l'Institut. Un bel exemple est celui des collections archéologiques : la *Bibliothèque archéologique et historique* (BAH), créée en 1921, rejoindra bientôt la revue *Syria* sur Persée ; et grâce au soutien de l'UE/Unesco puis de la fondation Aliph, la photothèque archéologique est archivée

et mise en ligne sur mediHAL<sup>7</sup>. En respectant les principes FAIR, ce projet permet de revaloriser des collections dont les originaux sont inaccessibles, et son succès nous encourage à poursuivre l'expérience en préparant l'archivage et la mise en ligne d'autres fonds tels que les photographies de l'Atelier du Vieux Damas et la cartothèque de l'Ifpo.

**TAOS BABOUR**

Responsable des humanités numériques  
t.babour@ifpoorient.org

**ELSA ZAKHIA**

Responsable des bibliothèques  
e.zakhia@ifpoorient.org

[1] <https://mediatheque.ifpoorient.org>

[2] <https://bib.cnrs.fr>

[3] Jean-Christophe Peyssard, « À l'Ifpo, les humanités numériques tiennent tête », Les aîlleurs de l'IST : Bibliothèques françaises à l'étranger, *Arabesques* n°86 (juillet 2017) <https://publications-prairial.fr/arabesques/index.php?id=431>

[4] <https://www.bnf.fr/fr/bibliotheques-dorient>

[5] <https://catalogue.bnf.fr/ark:/12148/cb458325361> ; Marion Slitine 2017, « Al-Mudhik al-Mubki, une revue satirique d'avant-garde en Syrie (1929-1939) », *Les Carnets de l'Ifpo. La recherche en train de se faire à l'Institut français du Proche-Orient* (Hypotheses.org), le 09 juin 2017, <http://ifpo.hypotheses.org/7517>

[6] Iyas Hassan 2021, « Édition de la Sira de Baybars. Les doigts de Francis et la clé à molette de Georges : quand la recherche en littérature populaire arabe était une affaire d'avant-garde », *LiPoL, Littératures Populaires du Levant. Archiver, analyser et conter le Roman de Baybars au XXI<sup>e</sup> siècle* (Hypotheses.org), le 02 juin 2021, <https://lipo.hypotheses.org/356>

[7] <https://medihal.archives-ouvertes.fr/IFPOIMAGE>

# DECALOG, une plateforme documentaire conçue pour vivre la mutation des catalogues

Depuis 10 ans, Decalog s'est engagé dans le processus d'évolution des catalogues vers le modèle IFLA-LRM avec la conception d'une plateforme documentaire sous forme de service en ligne.

**D**ecalog intervient dans le cadre de l'informatisation des bibliothèques depuis plus de 30 ans. Ses solutions équipent un large spectre de structures, d'une bibliothèque autonome jusqu'à des réseaux desservant une population de près de 1,5 million d'habitants.

Depuis 10 ans déjà, Decalog s'est engagé dans le long processus d'évolution des catalogues vers le modèle IFLA-LRM (harmonisation et consolidation de FRBR/FRAD/FRSAD). Une démarche construite autour de deux axes essentiels ; la conception d'une plateforme documentaire sous forme de service en ligne et un dialogue soutenu avec les agences bibliographiques.

Le modèle FRBR, de par son caractère tourné vers l'avenir, la richesse du modèle conceptuel et son adaptation aux mutations des catalogues, nous est apparu comme une évidence. Lorsque nous avons décidé de développer un nouveau SIGB, conscients que les formats évoluaient et que de simples adaptations seraient insuffisantes pour projeter une solution logicielle sur ces nouveaux enjeux, nous sommes repartis de la feuille blanche et avons structuré notre approche autour des modèles FRBR. Nous l'avons fait de manière pragmatique, sans imposer une rupture à laquelle ni les bibliothèques, ni les agences bibliographiques n'étaient encore prêtes. Nous avons avancé en soutenant un dialogue étroit avec nos utilisateurs, la BnF et l'Abes.

## UNE PLATEFORME DOCUMENTAIRE EN MODE SAAS

Nous avons retenu le modèle SaaS pour le développement de l'ensemble de nos applications. Il est à même de garantir la bonne gestion d'un parc utilisateurs dans les meilleures conditions de diffusion des évolutions. Il permet de proposer à tous, régulièrement et sans contrainte locale de déploiement, des nouvelles versions. Ainsi, nous proposons quatre fois par an des évolutions documentées,

accessibles immédiatement. Ce modèle permet d'absorber et de diffuser simplement les évolutions réglementaires, des innovations fonctionnelles et évidemment, les évolutions des formats.

Ce modèle est tout particulièrement adapté aux évolutions actuelles proposées par les agences bibliographiques. Il n'y aura pas de « grand soir » du catalogue. Les évolutions interviennent pas à pas, par étapes successives... Nous pouvons ainsi diffuser chacune de ces évolutions sans heurt, en accompagnant nos utilisateurs dans la compréhension des attentes et des possibilités associées.

## LA REPRÉSENTATION DES ÉDITEURS AU SEIN DU COMITÉ FRANÇAIS UNIMARC

Dès le début des évolutions des formats, la BnF et l'Abes ont souhaité associer les éditeurs de SIGB à la réflexion. Nous collaborons avec le Comité français Unimarc et le groupe « Transition bibliographique » pour un travail, au sens propre du terme, fondamental. De façon concrète et pragmatique, nous construisons ensemble ce modèle en devenir. Certains éléments qui nous semblaient pertinents il y a 10 ans sont aujourd'hui actés ou constituent une tendance forte. A titre d'exemple, la position de l'Œuvre au niveau bibliographique, évolution que nous avons proposée, est

aujourd'hui fortement envisagée au niveau international, comme le confirme le dernier *MARC Advisory Committee*.

## L'INNOVATION COMME MOTEUR

L'approche présentée ci-dessus traduit la présence de l'innovation dans l'ADN de Decalog. Au fil des ans, nous avons proposé des projets régulièrement primés. Nos derniers travaux en cours ont ainsi été retenus par le ministère de la Culture dans le cadre d'un projet d'innovation relatif à la conception d'un moteur de recherche adapté spécifiquement aux acquisitions des bibliothèques et médiathèques, basé sur l'apprentissage profond et un environnement de *Big Data*. Une innovation conçue pour l'usage des professionnels dans une approche concrète et pragmatique.

### GUILLAUME MESSONNIER

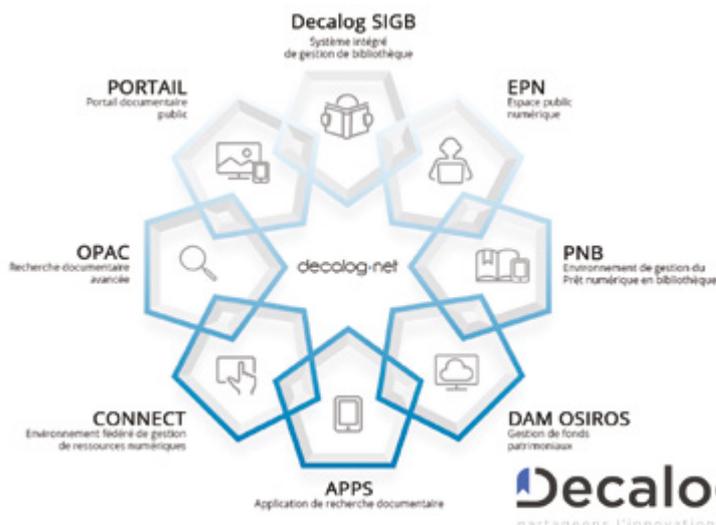
Directeur des études et développement  
chez Decalog  
g.messonnier@decalog.net

### DIDIER PIED

Directeur commercial  
d.pieddecalog.net

### FRANÇOIS RAFFIN

Directeur opérationnel Decalog  
Bibliothèques  
f.raffin@decalog.net



# Levée de voile sur le programme des Journées Abes 2022

Les inscriptions aux Journées Abes 2022 qui se dérouleront les 21 et 22 Juin au Corum, Palais des Congrès de Montpellier, ouvriront début mai<sup>1</sup>. D'ici là, chacun est invité à participer au concours de « Posters Journées Abes 2022 », contribution qui donne lieu à une invitation par établissement.<sup>2</sup>

## ABORDER LES RIVES DE LA RÉUTILISATION DES DONNÉES

Sujet crucial en temps de science ouverte, la question de la « réutilisation des données » sera au cœur des échanges programmés pendant la première journée. Produites par et pour les réseaux documentaires de l'ESR, les données de l'Abes, ouvertes et disponibles sous licence ouverte, sont largement réutilisées dans de nombreux environnements numériques. Comment optimiser l'exposition de ces données, gage de leur qualité selon les critères FAIR ? Comment, grâce à Science Plus ou Bacon, accroître leur visibilité sur le Web de données ou leur utilisation par les outils de découverte ? Quel rôle pivot jouent les données d'autorité et les identifiants IdRef dans le cadre de différents appels à projets et infrastructures de recherche ? Autant de sujets complexes qui feront l'objet de sessions plénières, d'une table ronde mais également de la conférence inaugurale, donnée par Nicolas Vigneron, spécialiste

international de Wikidata, actuellement en résidence au SCD de l'université Clermont Auvergne.

## EXPLORER LE CONTINENT DES RESSOURCES CONTINUES

Nouveau format expérimenté cette année, les Journées Abes 2022 intègrent la traditionnelle « Journée annuelle des Centres du Réseau Sudoc-PS (JCR) » afin de mettre en lumière, lors de la seconde journée, les problématiques spécifiques liées au signalement des ressources continues. Grâce à cette initiative, coordonnée avec brio par le service des Ressources Continues de l'Abes, nous irons de « cercles vertueux » en « bulles de retombées positives » avec, en matinée, une session riche de retours d'expériences stimulants et, l'après-midi, trois ateliers destinés principalement aux responsables CR selon le format habituel des JCR.

## LA TRAVERSÉE DES PARALLÈLES

Le mercredi après-midi, chacun sera invité à embarquer pour la session parallèle de son choix : un point d'étape sur la « transition bibliographique vue par le groupe de travail sur les ressources continues : quels changements dans les pratiques de signalement ? » ; un tour d'horizon, coordonné par l'équipe Calames, de « quelques collaborations réussies et inspirantes entre SCD et services d'Archives » ; et, pour les



amateurs de codes et astuces informatiques, une session en mode Hackaton qui abordera les territoires mal connus des scripts WinIBW et des requêtes croisées entre data.idref et Wikidata.

## ET LES AGAPES ?

La tradition a du bon puisqu'elle invite l'ensemble des participants à se retrouver lors du cocktail dînatoire autour des mets succulents préparés par le traiteur Germain. Cette année, rendez-vous est pris le 21 juin dans les Salons du Corum pour une ambiance, souhaitons-le, au diapason de la Fête de la Musique !

[1] Cette année, les inscriptions aux Journées Abes se feront via l'outil ScienceConf mis à disposition par le CCSD, ce qui nécessite la création au préalable d'un compte.

[2] Merci d'envoyer vos propositions de posters à : [jabes@abes.fr](mailto:jabes@abes.fr)

## LES PARRAINAGES

Cette année, les Journées sont parrainées par 15 organismes (édition scientifique, prestataire SGB, opérateurs) que nous remercions infiniment pour leur contribution à la qualité des Journées. Les sponsors seront tous présents pendant le congrès. N'hésitez pas à prendre rendez-vous et à aller à leur rencontre lors du salon professionnel programmé le 21 juin de 17h30 à 19h.

### PACK DIAMANT

OCLC et Gale Cengage, qui interviendront lors de 2 sessions plénières le 21 juin

### PACK OR

BibLibre, Cairn, De Gruyter, Ebsco, EDP Sciences, Ex Libris, IOP Publishing, Proquest, Springer, Tech Advantage, Wiley

### PACK CRISTAL / PARTICIPATION INSTITUTIONNELLE

INIST-CNRS et association KohaLa



## À NOTER

La Collection numérique, revue de l'AMUE, consacre son n°20 (à paraître en avril), au thème « Numérique universitaire dans les BU ». Ce numéro, que l'ADBU a codirigé, comporte plusieurs articles présentant la stratégie et les services de l'Abes à destination des établissements documentaires de l'ESR. Consulter : <https://www.amue.fr/systeme-dinformation/metier/la-collection-numerique>

(Portrait)

# Martine AUGOUVERNAIRE,

responsable de la BU de médecine et  
du pôle « Signalement et qualité des données »  
au SCD de Tours

## **Parlez-nous de vos fonctions actuelles...**

J'exerce la fonction transversale de responsable du pôle Signalement et contrôle qualité des données pour le service commun de la documentation de Tours et je suis également responsable de la BU de médecine de Tours. Coordinatrice Sudoc, j'organise le contrôle qualité des données de notre catalogue, avec l'aide de la correspondante catalogage. Je fais aussi partie du groupe des administrateurs du SIGB et je suis correspondante Calames.

## **Quelles sont les étapes qui vous semblent les plus importantes dans votre parcours professionnel ?**

J'ai réussi le concours de bibliothécaire en 1994, après quelques contrats dans plusieurs bibliothèques parisiennes. Affectée à Paris 12 (l'Upec aujourd'hui), j'ai participé au projet de réinformatisation. J'ai mesuré en travaillant sur le cahier des charges et sur la reprise des données l'importance de la qualité de ces données pour la réussite du projet. Je suis arrivée au SCD de Tours en 2001, l'année même où il intégrait le Sudoc. Adjointe de la responsable de la BU de médecine, j'ai peu à peu occupé des fonctions transversales, intégrant l'équipe du projet de réinformatisation en 2006, puis l'équipe des administrateurs du SIGB. J'ai également occupé le poste de correspondante formation pour le SCD pendant près d'une dizaine d'années, ce qui m'a permis d'appréhender les différents réseaux en œuvre dans les bibliothèques, mais aussi de mesurer à quel point la formation continue est primordiale dans nos métiers. Après un passage à l'Enssib en 2018, j'ai appris la fonction de coordinatrice Sudoc auprès de Véronique Lacan qui m'a précédée dans cette mission. La refonte de l'organigramme du SCD en 2020 a été l'occasion de faire évoluer le rôle de coordinatrice Sudoc en créant un véritable pôle signalement et d'intégrer davantage le contrôle qualité des données. La crise sanitaire et le développement du télétravail ont aussi fait évoluer les habitudes de travail en intégrant les chantiers de corrections dans les activités « télétravaillables ».

## **Quelles sont vos relations avec l'Abes ?**

L'Abes est pour moi un partenaire et un interlocuteur toujours à l'écoute. Mais nous sommes aussi sollicités pour faire évoluer les outils ou pour mettre en place des chantiers qualité. Ainsi, dans le cadre du dispositif CERCLES, nous avons pu contribuer à l'amélioration de la qualité des notices fournies par l'éditeur Open Edition. Quelles que soient nos questions, les équipes de l'Abes apportent toujours une réponse et travaillent avec nous sur nos propositions, comme cela a été le cas lors de notre chantier de vérification des constructions Rameau, par exemple.



## **Quels défis majeurs, d'après vous, aura à relever l'Abes dans les prochaines années ?**

La transition bibliographique est bien sûr le principal défi de l'Abes. Je pense qu'elle continuera aussi à jouer un rôle primordial pour aider les établissements dans le signalement de leurs ressources en ligne. L'Abes le fait déjà, en guidant les éditeurs pour qu'ils fournissent des métadonnées exploitables et en proposant aux établissements les outils pour les récupérer simplement, et son rôle pivot dans ce domaine ne fera sans doute que s'accroître. Un autre défi attend aussi l'Abes et les membres du réseau : l'alignement des identifiants auteurs.

## **Qu'appréciez-vous le plus dans votre métier ?**

Le travail en réseau et le partage des connaissances. C'est un métier qui évolue beaucoup et qui nécessite de se remettre souvent en question. C'est parfois fatigant mais c'est aussi stimulant !

## **Qu'est-ce qui vous énerve le plus ?**

Le manque de considération pour le travail de signalement et de description des données, qui est pourtant une des bases de notre métier. J'ai même entendu un intervenant à l'Enssib parler de « cataloguillage ». Or, si nos données ne sont pas précises et correctes, les facettes de nos outils de découvertes, par exemple, seront inopérantes... Et on mesure aujourd'hui l'importance du signalement pour la mise en valeur de nos collections numériques.

## **Quelle image donneriez-vous pour qualifier l'Abes ?**

Pour moi, l'Abes c'est un lien : un lien entre les membres du réseau, un lien avec les fournisseurs de métadonnées ou les fournisseurs de SIGB, un lien aussi avec les instances de normalisation dans la transition bibliographique. L'Abes fait du lien, et ce d'autant plus maintenant que nous sommes devenus des « cataliseurs »...

## **Votre expression favorite ?**

Le pire n'est jamais sûr... Je suis une incorrigible optimiste ! J'aime aussi beaucoup ce proverbe : tout seul on va plus vite, ensemble on va plus loin, qui illustre assez bien le travail en réseau.