



La science ouverte



À l'origine démarche citoyenne issue d'un mouvement de chercheurs, et désormais politique publique à part entière (nationale, européenne et internationale), l'écosystème de la science ouverte repose sur trois piliers : les données ouvertes (*Open Data*), le libre accès (*Open Access*) et des outils-logiciels libres et gratuits (*Open Source*).

Les acteurs et partenaires institutionnels du ministère de la Culture (institutions nationales, européennes et internationales, opérateurs et établissements de l'enseignement supérieur, structures et laboratoires de recherche, grandes infrastructures) s'en emparent et développent leurs propres usages.

Dans ce numéro, la recherche culturelle est scrutée sous l'angle de la science ouverte et de ses défis sociétaux. Car la tâche sous-tendue par les pratiques fédératives, collaboratives et inclusives de la science ouverte est immense : il s'agit non seulement de produire, publier et partager la recherche avec, par et pour le plus grand nombre, mais aussi de renouveler le lien entre science et société, réutiliser les données pour nourrir de nouvelles problématiques de recherche et incrémenter de nouvelles connaissances, diffuser la science dans toutes ses dimensions transdisciplinaires, créer des supports d'enseignement et de médiation, d'information et de formation à la science ouverte, enfin, d'évaluer la qualité des résultats de la « recherche culturelle ouverte » comme « bien commun ». Toutes ces actions constituent aussi aujourd'hui un levier dans les territoires pour la diversité des cultures et les systèmes de connaissances du monde.

Pour ce faire, il a fallu aux côtés des chercheurs l'engagement des ingénieurs, informaticiens, documentalistes, et de tous les acteurs de la culture. Dans ce contexte, les méthodes, outils, cadres juridiques, mesures incitatives conditionnant les financements publics ont été autant de « briques » apportées à la construction de la science ouverte. Ce changement de paradigme ne se fait pas sans débat. Mais, à chaque étape, l'enjeu reste de produire des données « aussi ouvertes que possible, aussi fermées que nécessaire ».

Au moment où le ministère de la Culture élabore sa stratégie numérique et décline les feuilles de route de sa stratégie de la recherche, ce numéro de *Culture & Recherche* témoigne de la façon dont les communautés scientifiques, les porteurs d'expertises du ministère de la Culture et le comité éditorial de la revue questionnent ensemble la science ouverte et sa production croissante, non sans exprimer freins et risques potentiels, notamment au regard des modes de transmission des connaissances et de la liberté académique, des droits d'auteur et de l'intégrité scientifique, de la fragilisation des modèles économiques et de production de la recherche.

Je les remercie chaleureusement. ■

NOËL CORBIN

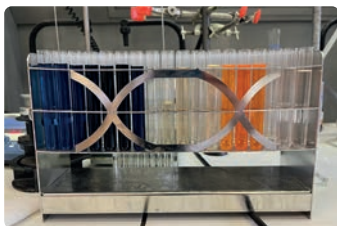
Délégué général à la transmission, aux territoires et à la démocratie culturelle



© MC-Didier Plovy

La science ouverte

- 1 Édito,
*Noël Corbin, Délégué général
à la transmission, aux territoires
et à la démocratie culturelle*
- 4 Préface,
C. Graindorge



6-57

**Un environnement
multidisciplinaire,
fédéré et ouvert**

8-25

S'engager et se positionner

- 8 Construire une édition scientifique ouverte : réflexion, structuration, financement, actions pour une bibliodiversité dans le contexte national,
O. Contat et A.-S. Gremillet
- 11 Dix ans d'évolution des pratiques de science ouverte en sciences humaines et sociales vus par l'infrastructure de recherche IR* Huma-Num,
O. Baude et N. Larrousse
- 14 Ouvrir et partager nos « Héritages » scientifiques,
J. Amiot-Guilouet et C. Laurière
- 18  *Bérose*, une expérience encyclopédique unique en humanités numériques,
C. Laurière
- 20 Apprendre à penser science ouverte dans un laboratoire de recherche en architecture : l'UMR CNRS Ambiances Architectures Urbanités (AAU),
F. Acquier, L. Bizien et L.-A. Charles
- 23 De la pratique quotidienne à une charte institutionnelle : la démarche vers la science ouverte de l'INHA,
F. Nurra

26-43


**Offrir des outils et services
à la recherche**

- 26 Comment mesurer l'efficacité d'une politique dans le domaine de la recherche ? Un exemple avec le Baromètre lorrain de la science ouverte,
L. Bracco
- 28 Un projet européen de *Cloud* collaboratif pour le patrimoine culturel,
L. De Luca et P. Liévaux
- 30 Ouvrir les données de la Bibliothèque nationale de France à la recherche,
A. Laborderie et F. Tibel
- 34  Le BnF DataLab : le laboratoire des humanités numériques de la Bibliothèque nationale de France,
M. Carlin
- 35 Reconstituer l'histoire de la France « par le bas » : un siècle de recensements de population décryptés par le projet Socface,
C. Kermorvant, L. Kesztenbaum et M. Restif
- 38 Le cercle vertueux des ressources documentaires du Centre de recherche du château de Versailles : l'exemple du projet VERSPERA,
M. da Vinha et B. Ringot
- 40 HALUR : des outils connectés à HAL au service des chercheurs,
L. Jonchère et O. Troccaz
- 42 OCdHAL, un outil de contrôle des publications déposées dans HAL,
P. Reynier et P. Bellemain

44-57

**Valoriser en réutilisant
des données**

- 44 Données et rapports d'opérations archéologiques à l'heure de la science ouverte,
J.-M. Conilleau, C. Cribellier, S. Deschamps, C. Domenech-Jaulneau, T. Lorho, Y. Menez, M. Monteil et M.-A. Virlovet
- 47 La science ouverte à l'INRAP,
M. Bouiron, E. Bryas et K. Salas Rossenbach
- 49 Valoriser en diffusant les travaux d'étudiants : le portail d'archives ouvertes DUMAS,
F. Gendre, C. Hugot, J. Schöpfel et A. Souchon

- 52  La mise en ligne des mémoires du Diplôme national supérieur d'expression plastique (DNSEP) dans le portail DUMAS, un atout pour la valorisation de la recherche en école d'art et de design,
M. Borios
- 54 Digital Library : le Plan national de numérisation du patrimoine culturel du ministère de la Culture italien,
A. Negri



58-89

**Un bien commun :
stocker
pour capitaliser**

60-70

**Mettre en œuvre les plans
de gestion des données :
administrer et conserver**

- 60 Données culturelles et science ouverte : pratiques et enjeux du ministère de la Culture,
R. Delassus et A. Conraux
- 63 Heurs et malheurs d'un programme européen : le cas du projet Rurland (*Rural Landscape in North Eastern Roman Gaul*),
M. Reddè
- 66 La Cellule Data Grenoble Alpes au service des communautés scientifiques de Grenoble pour la gestion de leurs données de recherche,
L. Albaret et V. Louvet
- 68 ESPADON : renforcer l'écosystème instrumental et numérique des sciences du patrimoine,
V. Detalle et R. Thomas

71-79

Mutualiser, agréger pour une meilleure répliquabilité et transportabilité

- 71 Interopérabilité des données culturelles et patrimoniales : le FAIR sans en avoir l'air, *E. Bermès*
- 73 Les données ouvertes au ministère de la Culture : mises à disposition et usages, *R. Aliacar*
- 75 La plateforme ouverte du patrimoine (POP) : un outil de mutualisation et de diffusion des données scientifiques des services culturels, *C. Duclert*
- 78 La Research Data Alliance (RDA) : pour des passerelles sociales et techniques afin de permettre le libre partage des données, leur réutilisation et leur interopérabilité scientifiques, *F. Genova*

80-89

Accélérer en innovant

- 80 Les archivistes, des professionnels au service de la diminution des impacts environnementaux du numérique : l'apport de l'outil Archifiltre, *C. Moser*
- 82 Compter les mots pour remonter le temps : Gallicagram et Gallicagrapher, deux outils d'exploration des archives numérisées de la BnF, *B. Azoulay, B. de Courson et W. Gleason*
- 85 Faites parler vos données ! Le *Text Mining* (la fouille de textes) en documentation, *D. Thebault*
- 88 BigScience, une collaboration internationale à l'origine du plus gros modèle de langue multilingue ouvert au monde, *L. Gianola*



90-137

De nouveaux modes de transmission et de diffusion

92-103

Publier et éditorialiser

- 92 Pourquoi le développement de la science ouverte a-t-il impérativement besoin d'éditeurs ?, *F. Gèze*
- 95 L'Observatoire de l'édition scientifique (OES), *C. Cherbut, B. Epron et A.-S. Gremillet*
- 97 La diffusion ouverte des travaux de recherche intégrant des images, *D. Pouchard*
- 99 La science ouverte et l'évaluation de la recherche, *T. Coulhon et B. Larroutourou*
- 101 L'utilisation des technologies IIF dans les publications scientifiques à forts contenus visuels : les expérimentations de la pépinière de revues DeVisu, membre du réseau Repères, *V. Cohoner et J. Hueber*

104-123

Promouvoir la découvrabilité et l'accessibilité

- 104 Les enjeux du libre accès pour les pratiques et expertises de la documentation : quels impacts de la science ouverte sur les activités documentaires en administration centrale ?, *S. Marguerin et M. Sayous*
- 107 La base EROS du C2RMF : mise en perspective de cette nouvelle ressource publique avec les programmes de recherche ESPADON, E-RHIS et IIF, *L. Bouiller*

- 110 « Lettres en Lumières » : l'intelligence artificielle appliquée aux décrets des États de Bourgogne, *E. Bouyé*
- 113 Des modèles tridimensionnels pour les grottes ornées : quels usages ?, *G. Pinçon, C. Montoya et S. Rassat*
- 116  La grotte Cosquer numérisée : un outil pour la connaissance, la conservation et la médiation, *X. Delestre, D. Lecouvreur et M. Olive*
- 118 De la bibliothèque numérique à un outil de gestion documentaire : l'Institut national du patrimoine (INP), *O. Blanc, C. Kermel, S. Le Conte, A. Méthivier et O. Zeder*
- 120 Archives, humanités numériques et études critiques en design : Problemata.org, une plateforme digitale *Open Source* pour le design..., *C. Geel*
- 122 La Plateforme éditoriale numérique Plara : recherches et pratiques artistiques en partage, *C. Diserens*

124-137

Créer, partager et intermédiaire

- 124 ArchiRès/Mir@bel : un partenariat qui valorise la science ouverte en architecture, *C. Avisse et Y. Bernardie*
- 126 Les recherches culturelles participatives dans la science ouverte, *A. Villarroel et F. Millet*
- 129 Plateformes culturelles contributives face au défi de l'ouverture de données, *M. Severo*
- 131 Les *Cahiers du Studio* – Théâtre de Vitry, *F. Krempfer*
- 134 Ossia Score, un logiciel de création interdisciplinaire au cœur de la plateforme logicielle libre du Studio de création et de recherche en informatique et musiques expérimentales (SCRIME), *P. Baltazar, J.-M. Celerier et M. Desainte-Catherine*
- 136 Le laboratoire Perception, représentations, image, son, musique (PRISM) ou la vie des archives entre recherche et création *R. Adjiman, C. Boë, P. Cesaro, C. Esclapez, J. Ferrando, M. Piqué et J. Simonnot*

Dossier coordonné par

CATHERINE GRAINDORGE

Rédactrice en chef, Délégation générale à la transmission, aux territoires et à la démocratie culturelle, Sous-direction des formations et de la recherche, Bureau de la recherche

En couverture



Tour d'observation à Bistumshöhe sur la rive sud-ouest du lac de Cospuden (région de Leipzig, Allemagne)

© Ansgar Koreng, Creative Commons BY 3.0 DE https://commons.wikimedia.org/wiki/File:Aussichtsturm_Bistumsh%C3%B6he,_1901201452_ako-2.jpg

Recherche culturelle et science ouverte un jardin partagé ou l'art du compost

CATHERINE GRAINDORGE

Rédactrice en chef, Délégation générale
à la transmission, aux territoires
et à la démocratie culturelle,
Sous-direction des formations
et de la recherche,
Bureau de la recherche

La science ouverte puise ses racines dans l'initiative citoyenne d'une communauté de chercheurs. Et d'aucuns s'accordent pour dire que tout commence en 1991, avec la création de la première archive ouverte : deux dépôts d'articles scientifiques (physique et mathématiques) sont réalisés par le physicien Paul Ginsparg. Il s'agissait alors de diffuser des articles non encore soumis ou acceptés par les revues scientifiques, sans être freiné par les délais, parfois longs, de relecture des comités scientifiques des revues.

« Appels », « Rapports », « Plans », dont le deuxième Plan national pour la science ouverte 2021-2024 du ministère de l'Enseignement supérieur et de la Recherche, stratégies nationales, européennes et internationales sont autant de jalons du périmètre annoncé dans la loi pour une République numérique du 7 octobre 2016, conduisant à la modification du Code de la recherche.

Et lorsque la recommandation de l'UNESCO sur une science ouverte paraît en 2021, puis la Déclaration du Conseil de l'Union européenne en juin 2022 sur la promotion des politiques européennes de science ouverte, la mise en culture est d'ores et déjà très avancée, ayant engrangé les 30 années de révolution numérique, pendant que les nombreux débats plantaient le décor d'un « jardin partagé ».

Ce premier enjeu historique qu'est la publication scientifique ouverte s'inscrit dans l'écosystème de la science ouverte avec ses trois

pilliers : les données ouvertes (*Open Data*), le libre accès (*Open Access*) et des outils-logiciels libres et gratuits (*Open Source*). Ces fondations soutiennent une architecture elle-même articulée autour de quatre principes FAIR : « facile à trouver, accessible, interopérable, réutilisable ».

Qui dit publication ouverte, dit données (et métadonnées). Qui dit données, dit logiciels. Pour mettre en culture ces premières friches, il a fallu des producteurs (publications par les chercheurs), des aménageurs (stockage) et des paysagistes (outils, logiciels et licences libres pour une culture enrichie). Très vite, cette approche citoyenne s'est inscrite comme un des modes de « faire la science », en portant aussi une recherche participative pour une innovation ouverte. Non sans que des risques connexes surgissent : prise en compte du rôle des éditeurs et de la fragilisation potentielle du modèle économique du livre, du pluralisme, de la liberté académique et des droits d'auteur des chercheurs.

Avec le lancement par le CNRS de l'archive ouverte Hyper Article en ligne (HAL) en 2001, la science ouverte amorce sa prise en compte par les politiques publiques, en questionnant d'emblée l'organisation d'accès aux productions de la recherche. L'évaluation de la recherche, la bibliodiversité et le multilinguisme pour améliorer la visibilité des résultats de la recherche européenne, la transparence et l'intégrité scientifique deviennent des enjeux prégnants qui vont nourrir de nombreux travaux des institutions (Assemblée nationale, Sénat, Observatoire de l'édition scientifique et rapports du médiateur du livre, Comité pour la science ouverte) pour une politique publique équilibrée et concertée de la science ouverte et de l'édition scientifique.

L'engagement, en vive accélération, des acteurs et partenaires de la recherche culturelle acte ce nouveau régime de production des savoirs qui s'adapte également à la participation d'un public plus large. Car c'est bien ce qui permet à de nouveaux acteurs de la société de participer aux processus scientifiques, à la montée en compétences sur la gestion des données, à la production de ressources pour les communautés de la recherche culturelle, aux innovations pédagogiques, à la diffusion sans entrave des résultats, des méthodes et des productions de la recherche. La forte progression, chaque année en France, du taux de publications en science ouverte en atteste. Enfin, la science ouverte soulève des questionnements complexes face à la multiplication des défis sociétaux en mobilisant de plus grandes communautés d'acteurs autour de plus grandes infrastructures.

Un environnement multidisciplinaire, fédéré et ouvert

La révolution numérique s'est imposée dans les grandes infrastructures de recherche, leviers de réseaux et lieux de partage entre les chercheurs et leurs savoirs. La transition vers l'édition ouverte, mais plus largement le stockage, la préservation de données et, désormais, les critères de traçabilité qui facilitent leur réutilisation s'y jouent pleinement (IR* Huma-Num). Autour de cette démarche qualitative et grâce aux utilisateurs, des services supports de la recherche et de soutien à la formation des chercheurs se développent.

Dans ce contexte, certains laboratoires mettent à disposition auprès de larges publics d'immenses corpus qui témoignent de la recherche en train de se faire. D'autres équipes

de recherche apprennent, ensemble, à « penser science ouverte ». Cela requiert non seulement l'engagement du laboratoire, mais aussi la formalisation de cette politique : groupes d'usagers et formations de sensibilisation, chartes d'intégrité scientifique et de reproductibilité, plans de gestion des données, etc.

En partant de ces pratiques de terrain et retours d'expérience, des grands établissements scientifiques poussent plus avant leur adhésion aux principes FAIR de la science ouverte en rédigeant leur propre charte d'établissement (Institut national d'histoire de l'art [INHA]). Cette structuration du « cultiver la science ouverte » repose sur des instruments et technologies consubstantiels à la recherche partagée : des indicateurs de ces pratiques sont déclinés depuis 2019 à l'échelle nationale dans le Baromètre de la science ouverte, *Cloud* européen collaboratif pour le patrimoine culturel, outils de l'immense chantier de la Bibliothèque nationale de France (BnF) pour mettre à disposition des chercheurs ses données, logiciels de reconnaissance automatique d'écriture manuscrite, outils facilitateurs de performance et de contrôle de l'archive HAL...

La montée en qualité des données semées porte intrinsèquement une valorisation des contenus : replanter et transplanter favorisent un essaimage, une diffusion performante. Qu'il s'agisse de rapports d'opérations archéologiques avec entrées géographiques et thématiques (rapports évolutifs, contextualisés et enrichis) ou de travaux d'étudiants de l'enseignement supérieur Culture (architecture, création artistique) sur HAL, la réutilisation des données engendre presque toujours une récolte à valeur ajoutée. Elle est aussi à même de susciter une réorganisation de la pensée institutionnelle (renforcement de valeurs sociétales) et de créer de nouvelles synergies entre structures européennes (Digital Library italienne).

Un bien commun : stocker pour capitaliser

Cette conscience est aujourd'hui portée dans les stratégies ministérielles : le développement exponentiel du nombre de données produites et le changement de pratiques de recherche issu d'outils et de technologies partagées et interopérables s'inscrivent dans les enjeux de recherche du ministère de la Culture. La plateforme de données ouvertes data.culture.gouv.fr et la licence ouverte Etalab sont emblématiques et répondent aux problématiques inscrites dans le deuxième Plan national pour la science ouverte (PNSO). La mise à disposition sous licence libre des codes sources des logiciels permet de les modifier et de les réutiliser, tout en soutenant la vérifiabilité et la réutilisation des résultats scientifiques.

Des démarches pionnières dans le domaine de l'archéologie, portées par des financements européens, se sont heurtées à la mise en place des premiers plans de gestion de données pour

administrer et conserver les productions de la recherche archéologique. Pour surseoir à ces difficultés, des « cellules d'accompagnement » et de proximité sont aujourd'hui mises en place, en s'appuyant sur des structures de recherche déjà engagées dans la science ouverte. Elles ont développé aujourd'hui une expertise solide pour la recherche culturelle. Tournées initialement vers les sciences humaines et sociales, elles ont l'ambition de servir tous les champs disciplinaires de la recherche (plateforme Recherche Data Gouv).

Stocker, capitaliser, administrer et conserver sont autant de sillons que labourent les sciences du patrimoine : elles portent une vision savante et innovante sur « l'objet patrimonial augmenté ». Pour ce faire, elles s'appuient sur des consortiums (ESPADON), à même de relever un défi instrumental (analyses multi-dimensionnelles, multi-échelles et multi-temporelles) et informationnel (données très hétérogènes relatives à l'objet).

L'enjeu du bien commun est aussi d'agréger les connaissances selon de bonnes pratiques FAIR, qui élèvent savoir-être et savoir-faire à hauteur des défis de répliquabilité et de transportabilité. Les plateformes du ministère de la Culture n'hésitent pas à y associer les citoyens pour améliorer ces qualités, et portent une stratégie de mutualisation (base POP). En outre, les passerelles internationales de la Research Data Alliance irriguent 150 pays, favorisant ainsi la diversité des cultures. Enfin, accélérer le traitement de gros volumes de données en développant des techniques d'analyse qualitative suscite des innovations, non seulement pour gagner du temps et sélectionner l'information pertinente (logiciels de lexicométrie, technique de fouille de textes), mais aussi pour diminuer l'impact environnemental (Archifiltre). La recherche intense en traitement automatique des langues (modèle BLOOM) avec les contributions de plus de 1 000 chercheurs internationaux et son corpus de 360 langues « ouvre » un nouvel horizon : richesse linguistique et souveraineté numérique y sont des enjeux à défendre.

De nouveaux modes de transmission et de diffusion

La science ouverte, levier puissant au service de la diffusion des connaissances, suscite des craintes relatives à l'avenir de différents « modèles » économiques, juridiques et scientifiques.

Le dernier *Avis du médiateur du livre sur l'édition scientifique dans le contexte des politiques en faveur de la science ouverte* (avril 2023) rappelle le « tissu riche et diversifié d'éditeurs scientifiques », en présente une analyse actualisée et formule des recommandations pour mieux équilibrer les rapports entre le monde de l'édition scientifique (imprimée) et celui de la science ouverte. La création, fin 2021, de l'Observatoire de l'édition scientifique permet d'interroger les conditions de pérennité de cet

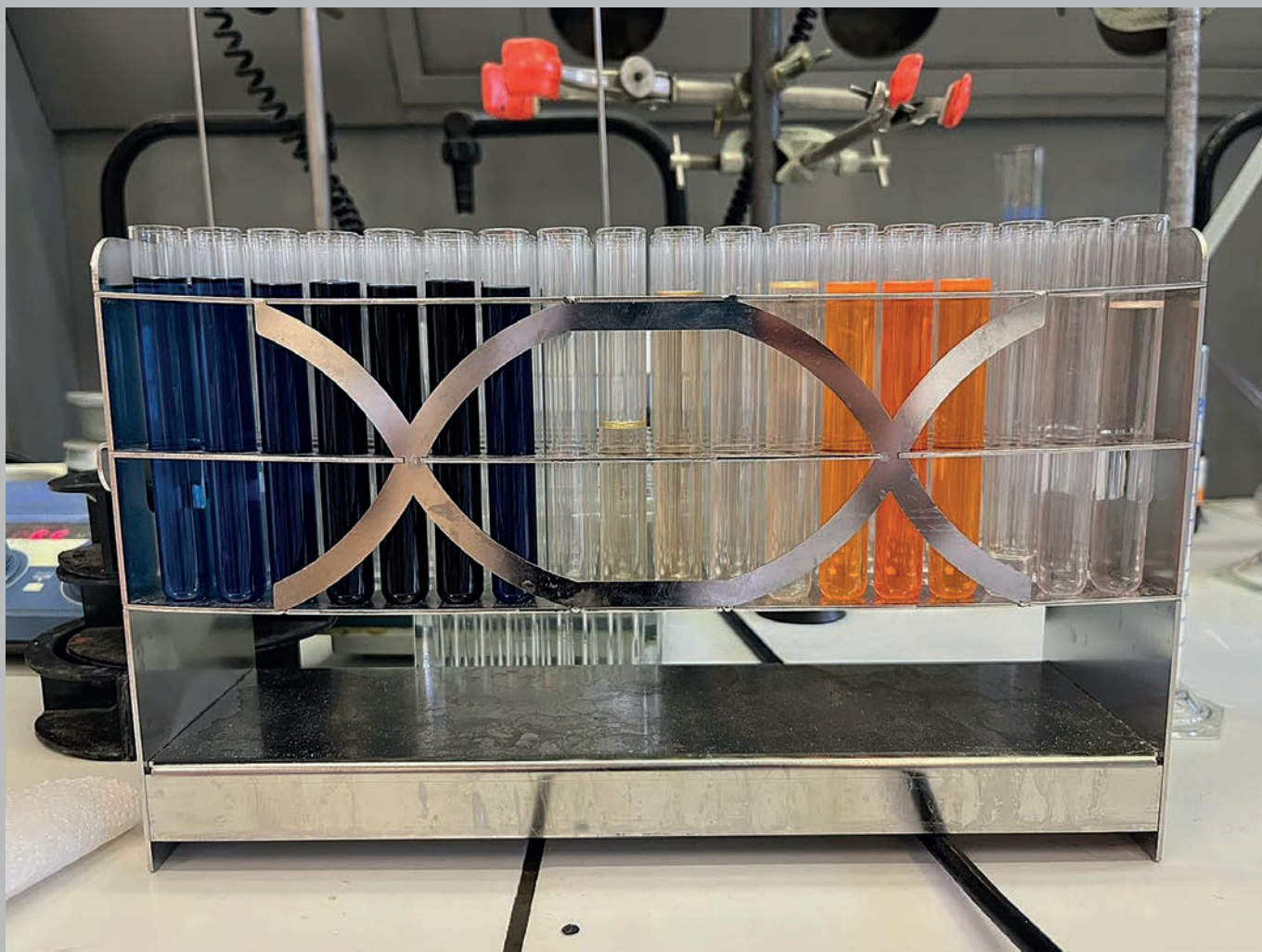
écosystème diversifié de revues et de publications, de conduire des études, d'être une aide à la décision, et si besoin de travailler avec le Comité pour la science ouverte.

La diffusion ouverte des travaux de recherche intégrant des images fait bouger les lignes du Code de la propriété intellectuelle. Les juristes cherchent des solutions équilibrées entre les enjeux croisés du droit d'auteur (arts visuels) et de la licence collective.

De même, le système d'évaluation des chercheurs est repensé pour prendre en compte, outre les publications scientifiques imprimées, les contributions en édition ouverte. Le Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur (Hcéres) porte, avec le CNRS et l'Agence nationale de la recherche (ANR), ces nouvelles évolutions.

Ces changements de paradigme induisent plus de découvrabilité et d'accessibilité : les pépinières de revues thématiques additionnent les qualités éditoriales de chaque revue en y adjoignant des techniques d'interopérabilité des visuels pour leur réutilisation par les chercheurs, et partagent cette production de ressources sur des plateformes ouvertes portées par des réseaux (Repères) ; les politiques documentaires institutionnelles apportent leur expertise en faisant évoluer leurs pratiques de collecte, de référencement et de valorisation des ressources ; les acteurs-chercheurs articulent de nouvelles bases de données publiques (EROS) autour de la recherche en conservation-restauration du patrimoine en Europe (méthodologie commune et Web sémantique) ; l'intelligence artificielle est sollicitée pour appliquer aux grands corpus de manuscrits anciens une transcription automatique des textes duplicable et transférable ; les nouveaux modèles tridimensionnels des grottes ornées dépassent désormais les enjeux essentiels de conservation de ce patrimoine, pour porter une exploitation de la connaissance interdisciplinaire et déclinée en usages partagés ; enfin, la mise en place de médiathèques numériques et de plateformes digitales articule leurs visées pédagogiques expérimentales et épistémologiques (INHA, Institut national du patrimoine, écoles d'art et de design) avec nos problématiques sociétales.

Ces dynamiques suscitent aussi des intermédiations. Les Écoles nationales supérieures d'architecture (ENSA), avec leurs réseaux Mir@bel et Archirès, et les acteurs-citoyens des sciences participatives avec leurs plateformes contributives sont au cœur d'une cartographie des territoires. La rencontre « ouverte » arts-science s'exprime pleinement : les *Cahiers du Studio* du studio-théâtre de Vitry ouvrent à tous la fabrication d'une recherche-crédation théâtrale en train de se faire. Les logiciels libres de création musicale et les bases d'archives ouvertes de sons et multimédias des laboratoires SCRIME et PRISM du CNRS témoignent encore, si besoin était, du caractère interactif, source de création et de médiation, de la science ouverte. ■



Sous le signe de l'infini
**(Séparation de molécules fluorescentes
par chromatographie sur colonne de silice)**

© Eléonore Tacke, Institut de chimie des substances naturelles (ICSN),
CNRS-Université Paris-Saclay, 2023

**Un environnement
multidisciplinaire,
fédéré et ouvert**

Construire une édition scientifique ouverte : réflexion, structuration, financement, actions pour une bibliodiversité dans le contexte national

Le ministère de l'Enseignement supérieur et de la Recherche (MESR) soutient et accompagne les changements vers une édition scientifique¹ française ouverte et multiple. Dans un contexte de publications de recherche européen et international dynamique, la France se positionne en pionnière avec ses infrastructures² rayonnant à l'international comme HAL, Métopes ou OpenEdition.

ODILE CONTAT

Cheffe du département Diffusion des connaissances et de la documentation, ministère de l'Enseignement supérieur et de la Recherche (MESR)

ANNE-SOLWEIG GREMILLET

Chargée d'études Édition scientifique et services à la recherche, MESR, secrétaire de l'Observatoire de l'édition scientifique (OES)

L'édition scientifique est aujourd'hui face à des transformations profondes dues à la conjonction entre puissance du numérique, internationalisation de la science, émergence des données de la recherche et implantation de la science ouverte.

En France, dès 2017, l'appel de Jussieu³ pour la science ouverte et la bibliodiversité affirme que « l'accès ouvert doit s'accompagner d'un soutien à la diversité des acteurs de la publication scientifique, [et d'un] développement de modèles innovants de publication scientifique ».

1. Anne-Solweig Gremillet, « Trois dispositifs nationaux pour une édition scientifique plus structurée, plus éthique et plus ouverte », [En ligne] *Arabesques*, n° 108, 2023, p. 6-7, mis en ligne le 16 janvier 2023, consulté le 25 mars 2023.

<https://publications-prairial.fr/arabesques/index.php?id=3132>

2. <https://www.enseignementsup-recherche.gouv.fr/fr/la-feuille-de-route-nationale-des-infrastructures-de-recherche-2021-84056>

3. <https://jussieucall.org/>

Jussieucall.org : appel cosigné par de nombreux acteurs de la recherche en faveur de la bibliodiversité.



Dans la même perspective, en 2018 puis en 2021, les deux Plans nationaux pour la science ouverte (PNOS)⁴ promeuvent la diffusion sans entrave des résultats, des méthodes et des produits de la recherche scientifique.

Le MESR accompagne cette transformation des pratiques, des structures et des fonctionnements du monde éditorial, et souhaite une implication de la communauté scientifique dans la construction d'un écosystème scientifique ouvert, éthique et transparent.

Contribuer à la définition et à la mise en œuvre d'un nouveau paradigme d'accès à la science

Afin de partager des connaissances communes sur l'édition scientifique dans le contexte des données⁵ et de la science ouverte, des études et réflexions sont menées principalement par deux instances : l'une relevant de la science ouverte, l'autre portant sur l'édition scientifique.

Les travaux du Comité pour la science ouverte participent de ce savoir commun :

- Fondée sur un jeu de données de plus d'un million d'articles publiés entre 2013 et 2020, une étude⁶ a permis de reconstituer les coûts pour les institutions de recherche françaises du modèle de publication en accès ouvert reposant sur le paiement « frais de publication » ou APC (*article processing charges*). Elle a produit plusieurs scénarios d'évolution de ces coûts à l'horizon 2030 et éclaire le cadre des négociations à mener.
- Appuyé sur une large démarche de consultation, le rapport de préfiguration pour la création d'une alliance des éditeurs scientifiques publics engagés dans la science ouverte⁷ dresse un état des lieux du secteur, présente des structures fédératives étrangères et propose un scénario de création d'une structure collective répondant aux besoins exprimés par ses futurs membres.

L'Observatoire de l'édition scientifique, créé conjointement par le ministère de l'Enseignement supérieur et de la Recherche et le ministère de la Culture, fait quant à lui dialoguer les acteurs de l'édition scientifique privée et publique. Partager un vocabulaire et des connaissances communes sur la base d'études et de travaux collectifs permettra ainsi d'aborder ensemble les enjeux de l'édition scientifique ouverte⁸ (voir l'article de Christine Cherbut, p. 95).

Faciliter la transition vers l'édition ouverte : nouveaux modèles, expérimentation et accompagnement

Au-delà de ce dialogue, le MESR a développé un plan de soutien à l'édition scientifique à destination des éditeurs français, privés et publics, tout en les accompagnant vers plus d'accès ouvert dans la diffusion des contenus scientifiques. Ce plan de soutien se concentre sur les principales plateformes de diffusion des revues scientifiques françaises, Cairn et OpenEdition en sciences humaines et sciences sociales (SHS) et Édition Diffusion Presse Sciences (EDP Sciences) en science, technologie et mathématiques



OpenEdition, infrastructure d'édition numérique au service de la communication scientifique en sciences humaines et sociales, est un acteur innovant, structurant et inspirant dans l'écosystème éditorial.



Des expérimentations vers des modèles économiques sont en cours : exemple de diffusion ouverte immédiate (*Subscribe to open*) EDP Sciences.

(STM). Il est porté en coordination par le MESR, le consortium Couperin⁹ et l'Abes¹⁰.

Un premier volet (2017-2021) du plan de soutien a eu pour effets l'extension du nombre d'établissements participant aux abonnements, l'accroissement du nombre de revues en accès ouvert, le travail en commun et le dialogue des trois plateformes. Le deuxième volet (2022-2026), doté d'environ 3 millions d'euros, articule plus fortement abonnement et ouverture, favorise le passage des revues à l'accès ouvert immédiat et construit un modèle de financement soutenable dans la durée.

4. <https://www.ouvrirlascience.fr/accueil/>

5. <https://www.numerique.gouv.fr/actualites/donnees-algorithmes-codes-sources-mobilisation-generale-sans-precedent-15-feuilles-de-route-ministerielles/>

6. Antoine Blanchard, Diane Thierry et Maurits van der Graaf, *Retrospective and prospective study of the evolution of APC costs and electronic subscriptions for French institutions*, Comité pour la science ouverte, 2022. hal-03909068

7. Caroline Dandurand, *Préfiguration d'une structuration collective des éditeurs scientifiques publics engagés dans la science ouverte* [Rapport de recherche], Comité pour la science ouverte, 2022, 86 p. hal-03713434

8. Voir dans ce numéro l'article de Christine Cherbut, p. 95.

9. Le consortium Couperin.org (Consortium unifié des établissements universitaires et de recherche pour l'accès aux publications numériques) est une association à but non lucratif financée par les cotisations des établissements membres et subventionnée par le ministère de l'Enseignement supérieur et de la Recherche.

10. L'Abes est l'Agence bibliographique de l'enseignement supérieur.

Une structure collective pour renforcer les éditeurs scientifiques publics engagés dans la science ouverte

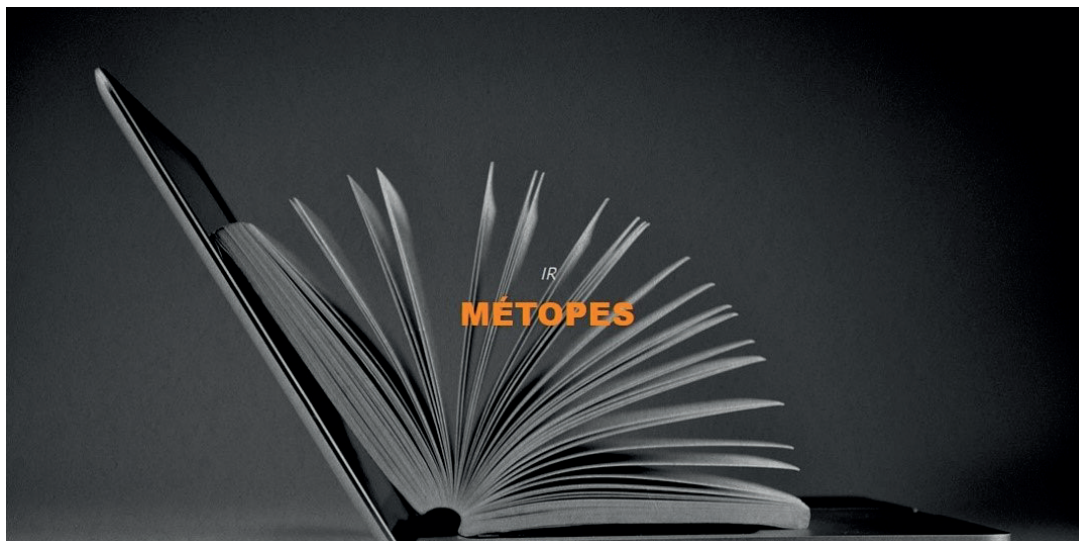
Il existe plus de 100 structures d'édition scientifique publiques en France (dont un tiers de presses universitaires), qui publient par an environ 1500 ouvrages et 280 revues dans toutes les disciplines¹. Cette activité éditoriale riche permet la validation et une diffusion importante de résultats de la recherche ainsi que l'avancée de la connaissance.

Prévue dans le deuxième Plan national pour la science ouverte², cette structure collective des éditeurs scientifiques publics a quatre objectifs principaux : valoriser et promouvoir, coordonner et mutualiser, fédérer

et représenter ses membres, ainsi que déployer la science ouverte dans l'édition scientifique publique. Chaque structure éditoriale publique, toute discipline confondue, est invitée à y participer pour plus de représentativité, de biodiversité et de mutualisation.

1. Caroline Dandurand, *Préfiguration d'une structuration collective des éditeurs scientifiques publics engagés dans la science ouverte*, op. cit.

2. <https://www.ouvrirlascience.fr/accueil/>



Métopes, infrastructure de recherche, pour une édition plus structurée.

En parallèle, le MESR, par l'intermédiaire du Comité pour la science ouverte, permet des expérimentations pratiques de l'édition scientifique ouverte financée par le Fonds national pour la science ouverte. Ce sont 5 millions d'euros qui ont ainsi été mobilisés pour les 49 lauréats des deux premiers appels à projets en 2020 et 2021. Un troisième appel à projets a été lancé en novembre 2022. L'objectif est de renforcer les capacités nationales en matière d'édition scientifique en accès ouvert en finançant des projets innovants et structurants : nouvelles formes de diffusion, d'écriture, d'évaluation ouverte, transition de revues scientifiques en accès ouvert, modèles ouverts pour les livres...

Permettre la consolidation et la pérennisation des infrastructures et plateformes structurantes pour l'édition ouverte

Enfin, les communautés scientifiques ont désormais accès à des infrastructures nationales et internationales qui garantissent la préservation et la circulation des connaissances. Les infrastructures de recherche nationales qui concourent à la publication scientifique – OpenEdition, HAL, Métopes¹¹ – assurent une fonction fondamentale de publication et de diffusion des savoirs qui complète et renforce la politique de soutien à l'édition scientifique. Elles portent ce principe essentiel qui considère la recherche scientifique comme un bien commun¹² partagé avec tous et sont des piliers de la science et de l'édition scientifique ouverte.

Ces actions de structuration, de financement et d'accompagnement permettent le renforcement d'une édition scientifique ouverte, diverse et durable, et la mise en œuvre d'une bibliodiversité qui s'inscrit dans un mouvement européen et international. ■

11. L'infrastructure Métopes est conçue à l'usage des éditeurs et au service de l'activité éditoriale de l'ensemble des établissements publics d'enseignement supérieur et de recherche.

12. Mélanie Dulong de Rosnay et Félix Stalder, « Digital commons », *Internet Policy Review*, vol. 9, n° 4, 2020. <https://doi.org/10.14763/2020.4.1530>

Un plan de soutien à l'édition pour sécuriser les modèles de financement des revues et des plateformes et construire des partenariats

Sont concernés :

- Cairn, un groupement de commande « abonnement » qui intègre une politique de soutien à la transition des revues candidates à une diffusion ouverte immédiate – modèle « *Subscribe to open* », qui sécurise les abonnements des bibliothèques en contrepartie d'un passage à l'accès ouvert ;
- EDP Sciences, un modèle de diffusion en accès ouvert financé par des frais de publication raisonnables. Modèle de type « accord dit transformant ».

EDP expérimente également sur d'autres revues le modèle « *Subscribe to open* » avec certaines communautés de recherche :

- OpenEdition, un soutien au modèle « accès ouvert *freemium* » et une incitation pour les revues à supprimer toute barrière à leur diffusion en accès ouvert – modèle « *freemium* » avec diffusion des contenus en libre accès et des services et formats réservés aux institutions abonnées.

Dix ans d'évolution des pratiques de science ouverte en sciences humaines et sociales vus par l'infrastructure de recherche IR* Huma-Num



L'infrastructure IR* Huma-Num a été créée il y a dix ans par la fusion de deux structures. L'une proposant des services technologiques, le très grand équipement (TGE) Adonis, et l'autre assurant la coordination de communautés organisées en consortiums autour de la création de jeux de données numériques, l'infrastructure de recherche (IR) Corpus. L'infrastructure est inscrite sur la feuille de route nationale des infrastructures du ministère de l'Enseignement supérieur et de la Recherche (MESR). Son label d'IR* (ex- très grande infrastructure de recherche [TGIR]) indique qu'elle relève d'une politique nationale et d'un fléchage budgétaire du MESR.

La notion d'infrastructure en sciences humaines et sociales (SHS) était à cette époque un concept qui restait largement à imaginer. L'organisation originale de IR* Huma-Num¹, issue de sa double origine qui associe un dispositif technologique aux communautés scientifiques, a permis de s'appuyer sur des compétences extérieures pour guider l'évolution de l'infrastructure au plus près des besoins des communautés. À cet égard, les consortiums-HN² labellisés et soutenus par IR* Huma-Num ont joué un rôle essentiel dans la construction de l'infrastructure.

Au commencement, le besoin principal concernant les données était surtout de ne pas les perdre, compte tenu du fait qu'elles se trouvaient stockées sur des supports fragiles comme des disques optiques, des clés USB, des disques durs portables, voire des supports plus exotiques (disquettes ou autres cartouches). Cependant, il est apparu rapidement que préserver simplement les données brutes n'était pas suffisant. En effet, il est inutile de conserver des données qui potentiellement ne seront plus lisibles car encodées avec des formats non pérennes, c'est-à-dire des formats qui ne sont pas documentés, qui ne sont pas outillés ou, plus généralement, dont les communautés utilisatrices ne sont plus actives. Les sciences humaines et sociales sont constituées de différentes disciplines qui utilisent des formats très diversifiés, et l'équipe technique de IR* Huma-Num ne peut disposer de l'expertise sur tous ces domaines. Afin d'accompagner les choix de formats effectués, IR* Huma-Num bénéficie des

compétences développées au sein de ses consortiums-HN, associées à celles de ses partenaires comme le Centre informatique national de l'enseignement supérieur (CINES) pour la préservation pérenne. Par exemple, le travail conjoint avec le consortium 3D a abouti à la création d'un jeu de métadonnées spécifiques adaptées aux objets 3D ainsi qu'au choix de formats de fichiers pour leur préservation sur le long terme. C'est ainsi que IR* Huma-Num a pu mettre progressivement en œuvre une chaîne de traitement qui couvre le cycle de vie des données de la recherche, de leur collecte à leur préservation.

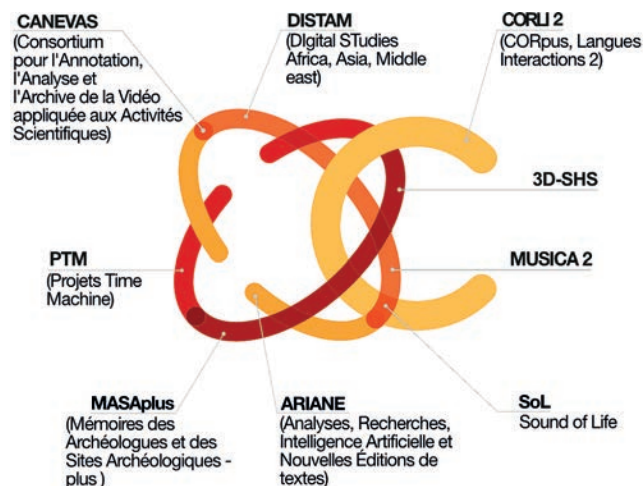
OLIVIER BAUDE

Directeur de IR* Huma-Num,
Unité d'appui et de recherche (UAR)
3598 CNRS

NICOLAS LARROUSSE

Directeur adjoint de IR* Huma-Num,
Responsable du pôle Coordination
nationale et internationale des
communautés (CONIC), Unité d'appui
et de recherche (UAR) 3598 CNRS

- <https://www.huma-num.fr/>
- <https://www.huma-num.fr/les-consortiums-hn/>



Le cycle de vie des données de la recherche et les services Huma-Num.



3. <https://nakala.fr/>
4. Les principes FAIR (*Findable, Accessible, Interoperable, Reusable* pour « facile à trouver, accessible, interopérable et réutilisable »). <https://www.go-fair.org/fair-principles/>
5. <https://isidore.science/>
6. <https://www.huma-num.fr/carte-des-relais-huma-num-dans-les-msh/>
7. Voir dans ce numéro l'article de Roselyne Aliacar, p. 73. <https://recherche.data.gouv.fr/>
8. <https://recherche.data.gouv.fr/fr/page/centres-de-referencethematiques-expertises-par-domaine-scientifique>

L'un des éléments clé de ce dispositif est l'entrepôt NAKALA³ qui a été conçu par IR* Huma-Num pour permettre de publier des données documentées, d'en assurer la lisibilité dans le temps et de les partager en vue de leur réutilisation. Ces caractéristiques ont permis de faciliter la mise en œuvre des principes FAIR⁴, constituants centraux de la science ouverte. La découvrabilité des données et productions scientifiques est ensuite assurée par le service ISIDORE⁵ qui est un moteur de recherche associé à un assistant scientifique fondé sur les technologies du web sémantique et de l'intelligence artificielle.

L'enjeu aujourd'hui n'est plus le risque de perdre les données, mais de pouvoir assister les projets de recherche dans l'amélioration de la qualité des données et des métadonnées qu'ils produisent. C'est dans cet objectif que IR* Huma-Num a créé un pôle d'accompagnement qui permet d'orienter les choix de formats, de référentiels et d'organisation en vue d'une publication des jeux de données respectant des critères de qualité permettant d'en assurer la traçabilité, la citation et la découvrabilité, afin d'en faciliter la réutilisation. De même que pour la création

de la chaîne de traitement de données mentionnée auparavant, IR* Huma-Num implique fortement les communautés dans ce processus. Cette implication se réalise notamment en s'appuyant sur le réseau de correspondants de l'infrastructure dans les Maisons des sciences de l'homme qui a été développé ces dernières années.

Ce réseau⁶ assure un relais naturel avec les projets de recherche soutenus par ces infrastructures régionales et complète les liens avec les communautés établis avec les consortiums de IR* Huma-Num.

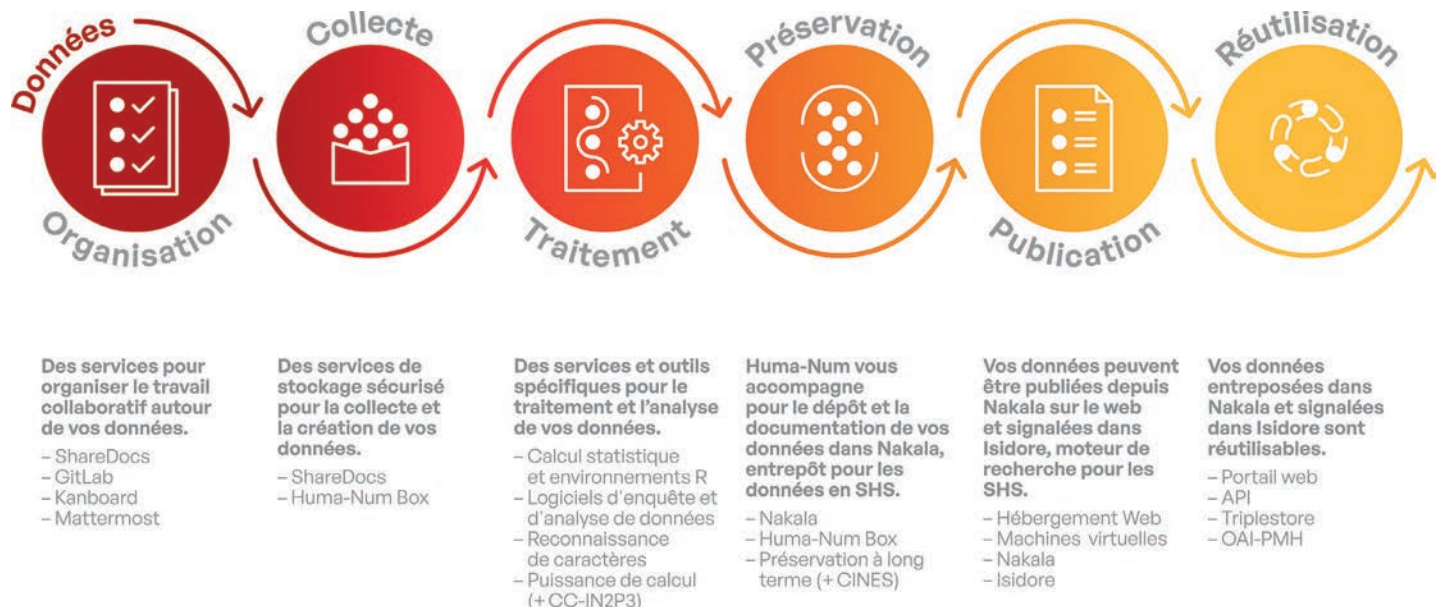
Durant cette décennie, le paysage national a lui aussi beaucoup évolué, en particulier récemment avec la création de l'initiative Recherche Data Gov⁷ (RDG) qui constitue un écosystème pour le partage et l'ouverture des données de recherche. Compte tenu de son expérience des dispositifs mis à disposition des communautés scientifiques ainsi que du maillage national constitué, IR* Huma-Num a été labellisé « Centre de référence thématique⁸ » pour les sciences humaines et sociales. Il permet à IR* Huma-Num de s'inscrire dans une initiative nationale et ainsi de favoriser la valorisation des données en sciences humaines et sociales stockées dans l'entrepôt NAKALA, qualifié d'« entrepôt de confiance », qui sera moissonné par le catalogue national. De même, la démarche d'amélioration de la qualité des données entreprise par



L'assistant de recherche ISIDORE.



L'entrepôt NAKALA.



Comme on peut le constater, ces dispositifs de soutien pour la mise en œuvre de la science ouverte ne peuvent se développer qu'en y associant étroitement les communautés utilisatrices.

IR* Huma-Num bénéficiera pleinement du dispositif d'accompagnement constitué par le réseau national des Ateliers de la donnée⁹ mis en place dans le cadre de l'initiative RDG. La maîtrise de la chaîne de traitement des données au cœur des pratiques scientifiques actuelles permet également à IR* Huma-Num de travailler avec différents partenaires, notamment avec la Bibliothèque nationale de France (BnF) dans le cadre du BnF DataLab¹⁰, un service d'accès outillé aux collections de la BnF pour les chercheurs.

La maturité du dispositif élaboré par IR* Huma-Num pour l'ouverture des données de recherche permet d'envisager de nouvelles possibilités de valorisation de ces productions essentielles à la démarche scientifique. Un exemple est l'établissement d'un lien permanent entre les publications et les données associées. Une première exploration de cette notion a été réalisée dans le cadre du projet

européen EOSC-Pillar¹¹. Une preuve de concept de réalisation du lien entre données et publications a été construite entre l'entrepôt de données NAKALA et l'outil de dépôt de publication HAL-SHS¹² géré par le Centre pour la communication scientifique directe (CCSD)¹³. Cette expérimentation a permis d'identifier les prérequis nécessaires à la création de liens respectifs entre les données et les publications : la création de métadonnées croisées sur les deux plateformes à la synchronisation nécessaire pour conserver la cohérence de ces liens, ainsi que leur publication standardisée pour en assurer la découvrabilité.

Les enseignements de ce projet seront mis à profit dans le cadre du projet d'EquipEx+ COMMONS¹⁴ qui associe trois infrastructures nationales autour de la création de liens entre les données et les publications : IR* Huma-Num pour les données, OpenEdition¹⁵ et Métopes¹⁶ pour les publications. Ce projet qui a débuté en 2022 se déroulera sur huit ans. Il vise à développer des passerelles entre les plateformes des différentes infrastructures pour faciliter l'établissement de liens entre les données et les publications pour les différentes catégories d'utilisateurs comme les éditeurs de revues ou les projets de recherche. Le développement de formations pour ces utilisateurs représente un volet important du projet ainsi que l'ajustement permanent des objectifs et des réalisations du projet par l'utilisation des résultats d'un observatoire des usages.

Comme on peut le constater, ces dispositifs de soutien pour la mise en œuvre de la science ouverte ne peuvent se développer qu'en y associant étroitement les communautés utilisatrices. Par ailleurs, en raison des évolutions très importantes et rapides des pratiques durant ces dernières années, il est nécessaire de disposer d'une vue prospective sur les dynamiques en cours afin de pouvoir anticiper les usages futurs. Dans cet objectif, IR* Huma-Num a créé un pôle qui permet d'alimenter la réflexion sur les nouveaux services à développer : le Huma-Num Lab¹⁷. ■

9. <https://recherche.data.gouv.fr/fr/page/ateliers-de-la-donnee-des-services-generalistes-sur-tout-le-territoire>

10. Voir dans ce numéro l'article d'Arnaud Laborderie et Florence Tfibel, p. 30 et le focus de Marie Carlin, p. 34. <https://www.bnf.fr/fr/bnf-datalab>

11. Le projet EOSC-Pillar (2019-2022) vise à coordonner les efforts nationaux en matière de science ouverte en Autriche, en Belgique, en France, en Allemagne et en Italie. Il assure leur contribution et leur préparation à la mise en œuvre du European Open Science Cloud (EOSC) : voir l'article de Pascal Liévaux et Livio De Luca dans ce numéro, p. 28. L'initiative propose de partager les meilleures pratiques et de susciter des possibilités d'approches interdisciplinaires au sein de l'EOSC. <https://www.eosc-pillar.eu/>

12. <https://shs.hal.science/>

13. <https://www.ccsd.cnrs.fr/>

14. <https://leo.hypotheses.org/17434>; <https://shs.hal.science/halshs-03881307>

15. <https://www.openedition.org/>

16. <http://www.metopes.fr/>

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

Ouvrir et partager nos « Héritages » scientifiques

Héritages est un laboratoire qui explore les articulations entre culture, patrimoine et création. Il a à cœur de diffuser les résultats de ses travaux au-delà des circuits académiques grâce à la mise en œuvre de programmes de science ouverte aux formats divers qui ont vocation à s'adresser à des publics pluriels. Plateformes, revue, enseignement innovant, documents, données et archives numérisés et mis à la disposition des internautes : autant d'outils qui permettent la visibilité et la circulation des résultats de recherches en train de se faire. Ces programmes de science ouverte sont fondés sur des productions textuelles, photographiques, audiovisuelles et multimédias formant de très riches corpus dont l'accessibilité est une des clés de voûte de la politique scientifique du laboratoire. Tous ces projets bénéficient du soutien du ministère de la Culture.

JULIE AMIOT-GUILLOUET

Professeure des universités,
CY Cergy Paris Université,
codirectrice de l'UMR 9022 Héritages :
Culture/s, Patrimoine/s, Création/s

CHRISTINE LAURIÈRE

Chargée de recherche au CNRS,
ex-codirectrice de l'UMR 9022 Héritages,
codirectrice de *Bérose – Encyclopédie
internationale des histoires de
l'anthropologie* et de la collection
des *Carnets de Bérose*

Héritages est une unité mixte de recherche (UMR) créée en janvier 2021 par le CNRS, le ministère de la Culture (MC) et CY Cergy Paris Université¹. Embrassant un large éventail de disciplines scientifiques dans les domaines des sciences humaines et sociales, elle explore les champs du patrimoine et de la création et leurs articulations, tout en réfléchissant

aux modalités d'instauration de la culture en tant que processus dynamique. Grâce à ses chercheurs inventifs et créatifs, Héritages porte plusieurs projets en science ouverte structurants pour son identité et son rayonnement scientifiques. Ils ont vocation à toucher des publics diversifiés, à restituer et faire circuler de façon moins académique le fruit de nos recherches.

1. <https://heritages.cyu.fr/>

Entourée de la famille Marquefave, Maguy Pichonnet-Andral enquête sur le jeu de l'accordéon diatonique auprès de Gabriel Piotton à Luxey, hameau du Chourd (Landes), 10 juin 1965.



Cliché Claudie Marcei-Dubois, Mucem Ph.1966.099.068, enquête Landes du MNATP-CEF

Il s'agit d'accroître la visibilité des recherches en train de se faire, afin que leurs méthodes et résultats soient davantage accessibles et appropriés. Les projets de science ouverte d'Héritages offrent un accès facilité aux savoirs et aux collections, aux contenus des bases de données scientifiques. C'est le cas de *Bérose*, encyclopédie internationale des histoires de l'anthropologie² (voir *infra* le focus), ou du site Les Réveillées³, plateforme de valorisation d'archives ethnomusicologiques multimédias qui offrent des exemples d'interface et de rencontre entre science et société civile, associations et communautés, artistes professionnels et amateurs. L'un des enjeux de ces projets, comme celui de la revue *Ethnographiques.org*⁴, est d'explorer de nouveaux modes de « faire savoir », au double sens de fabriquer et de diffuser, à la croisée de plusieurs médias (texte, son, image), d'élargir et d'internationaliser les publics concernés. Tous ces projets sont soutenus par le ministère de la Culture.

Notre-Dame, un chantier scientifique exceptionnel

Les recherches réalisées par le groupe de travail EMOBI (Émotions/Mobilisations), piloté par Claudie Voisenat, comme toutes celles qui sont produites dans le cadre du chantier scientifique Notre-Dame de Paris (CNRS-ministère de la Culture), sont destinées à intégrer un système d'information innovant, en données ouvertes (*Open Data*), véritable double virtuel de la cathédrale, qui permettra de rassembler et de partager toute la connaissance produite sur le monument actuellement en cours d'élaboration par le groupe de travail sur les données numériques.

Ce sont donc non seulement les résultats des recherches qui seront mis à disposition des chercheurs et du public, mais aussi l'ensemble des données qui ont permis de les réaliser, qu'il s'agisse des réponses au questionnaire « Venir à Notre-Dame de Paris. Expériences et perceptions » diffusé en 2020, comprenant plus de 500 variables, et qui a généré plus de 2 000 réponses actuellement en cours d'analyse sous la direction de Sylvie Sagnes, ou des séquences vidéo filmées par Jean-Christophe Monferran dans le cadre de l'ethnographie visuelle du chantier de restauration, en cours d'indexation.

Les Réveillées, une carte ethnomusicologique de la France inédite

Créé afin de nourrir la recherche comme la création musicale, le site Les Réveillées : ethnographies musicales des territoires français et francophones (1939-1984) a été mis en ligne en juin 2021. Son ambition est de redonner vie à des documents patrimoniaux historiques, reflets des pratiques musicales du milieu paysan au siècle dernier. Il donne à lire, à entendre et à regarder les archives des principales enquêtes ethnomusicologiques du Musée national des arts et traditions populaires (MNATP), menées en France métropolitaine ou ultramarine, entre 1939 et 1984. Il fait fond sur la numérisation des archives sonores (financée par la Délégation à la musique du ministère

Il s'agit d'accroître la visibilité des recherches en train de se faire, afin que leurs méthodes et résultats soient davantage accessibles et appropriés. Les projets de science ouverte d'Héritages offrent un accès facilité aux savoirs et aux collections, aux contenus des bases de données scientifiques.

de la Culture) et scientifiques du MNATP (financée par les Archives nationales).

Conçu et animé par François Gasnault et Marie-Barbara Le Gonidec, le projet Les Réveillées a permis l'appropriation de ce patrimoine convoité et réclamé, depuis le *revival* des années 1970, par les acteurs et actrices des musiques traditionnelles. Cette base de données comprend la totalité des archives des

2. <https://www.berose.fr/>
3. <http://les-reveillees.ehess.fr/> et <https://didomena.ehess.fr/collections/2z10wq61n>
4. <https://www.ethnographiques.org/>



« Roulée des œufs » du lundi de Pâques. À Courban, en Côte-d'Or, le 27 mars 1967, les enfants se présentent dans une maison, enregistrés par Maguy Pichonnet-Andral.



Page d'accueil du site Les Réveillées.



Florilège de couvertures de différents numéros d'*Ethnographiques.org*.

36 enquêtes retenues, soit environ 44 000 documents numériques répartis en fichiers sonores (plus de 10 000), photographies (plus de 12 000) et archives textuelles (plus de 22 000 scans).

***Ethnographiques.org*, une revue pionnière de la science ouverte**

Grâce, entre autres, au soutien institutionnel du CNRS et du ministère de la Culture, la revue internationale de sciences sociales en ligne, *Ethnographiques.org*, cofondée et codirigée par Sophie Chevalier et Thierry Wendling (CNRS, Héritages), fonctionne, depuis sa première livraison en 2002, sur le modèle (ultérieurement appelé) « diamant » : diffusion sans frais de publication pour les auteurs et

consultation gratuite pour les lecteurs. 44 numéros ont déjà paru, selon un rythme semestriel.

Comme son titre le suggère, la ligne éditoriale d'*Ethnographiques.org* est centrée sur des travaux (en ethnologie principalement, mais aussi en sociologie) qui développent leurs analyses des phénomènes sociaux et culturels sur la base d'études de terrain détaillées. Les titres des dossiers des cinq dernières années soulignent la diversité des thèmes abordés : la vieillesse, le jeu avec les animaux, le changement climatique, les univers du hip-hop, les rencontres entre ethnologues et artistes, les enquêtes auprès d'enfants, la question du genre, etc. Les prochains numéros traiteront de l'apprentissage, des institutions totales, des musées, de la dette.



Le chantier de Notre-Dame, 22 septembre 2020 : la fabrication de « berceaux » pour transporter les vestiges de la charpente calcinée.

© J.-Ch. Monterran

SELIN, une plateforme d'enseignement innovante

Depuis 2022, Héritages soutient SELIN (système d'*e-learning* inductif, ou *Self Induced Learning*), un dispositif pédagogique en ligne innovant, créé initialement par l'Institut d'ethnologie de l'Université de Neuchâtel pour entraîner les étudiants à l'observation ethnographique et à l'analyse de situations sociales et culturelles, leur apprendre à observer et approfondir leur réflexion. Thierry Wendling est membre du projet depuis ses débuts.

Développé par une équipe d'informaticiens, ce dispositif original d'enseignement à distance, accessible en ligne, offre de riches matériaux (vidéos, audios, photographies, textes) qui permettent de se faire une idée très concrète des éléments qu'un ethnologue peut découvrir, noter, photographier, enregistrer sur son propre terrain. Sur cette base, les étudiants sont invités, selon une progression pédagogique raisonnée, à décrire ce que ces matériaux donnent à voir.

Chaque « leçon » de SELIN est centrée sur un terrain particulier et correspond de fait à l'ethnographie d'un groupe social, d'une pratique. Une leçon représente pour chaque étudiant une quarantaine d'heures de travail. Plusieurs leçons ont été développées : les parties traditionnelles de loto dans les villages des cantons de Neuchâtel et de Fribourg ; la pratique du skate en Suisse et au Japon ; la permaculture dans le Jura suisse. Associé plus récemment au projet, le département d'anthropologie de l'Université d'Alaska à Fairbanks a développé une leçon sur le thème du *mushing* en Alaska.

Un atelier rassemblant des enseignants-chercheurs d'Héritages s'est, quant à lui, constitué pour valoriser des matériaux d'enquêtes financées par le ministère de la Culture (la mémoire de la *Retirada* dans le Sud-Ouest, l'inscription des fêtes de l'Ours au Patrimoine culturel immatériel – PCI, la mémoire des migrations, etc.). Certaines leçons s'adresseront au public étudiant (Master et Doctorat) de l'Université de Cergy, tandis que d'autres seront accessibles à tout internaute désireux de se familiariser avec la méthode ethnographique.

Héritages s'engage résolument dans des projets transversaux et des réseaux professionnels contribuant au développement des publications scientifiques ouvertes.



© J.-Ch. Monferran

Le chantier de Notre-Dame, 6 octobre 2020 : les « reliques » de l'incendie.

La science ouverte au cœur de la politique du laboratoire

Enfin, au sein d'Héritages, la science ouverte n'est pas uniquement l'affaire des chercheurs et enseignants-chercheurs.

Elle se construit aussi dans les actions et les missions du personnel de soutien à la recherche, notamment du service d'édition en la personne d'Anabel Vazquez (CNRS), qui assure le suivi éditorial de deux projets (*Bérose* et *Ethnographiques.org*) et appuie le déploiement des principes FAIR de la science ouverte pour ces deux supports. Anabel Vazquez assume ainsi un rôle de référente science ouverte pour un projet européen. Elle participe au co-pilotage du SPOC (*Small private online course*) QUÉRO⁵, parcours pédagogique à destination des équipes éditoriales de revues académiques, consacré aux bonnes pratiques en matière d'édition ouverte et structurée, porté par le réseau Repères⁶.

Ce réseau de pépinières de revues scientifiques en accès ouvert a bénéficié d'un premier financement du Fonds national pour la science ouverte (FNSO). Il a récemment soumis un deuxième projet, Coopères, dans le cadre de l'appel « Publication et édition scientifiques ouvertes » du Comité pour la science ouverte (COSO) avec le soutien d'Héritages.

Ainsi, Héritages s'engage résolument dans des projets transversaux et des réseaux professionnels contribuant au développement des publications scientifiques ouvertes. Par ailleurs, grâce à Solène Hazouard, chargée de communication éditoriale (CY Cergy Paris Université), Héritages s'est emparé des outils de diffusion scientifique sur internet valorisant la prolifique activité de ses membres. Directrice de rédaction, Solène Hazouard anime deux *Carnets* sur la plateforme Hypotheses.org : le *Carnet des chercheurs d'Héritages*⁷ et le *Carnet des doctorants d'Héritages*⁸, qui valorisent leurs recherches, activités et publications, les doctorants étant étroitement associés au processus éditorial, afin de les professionnaliser et de les familiariser avec les pratiques scientifiques. ■

5. <https://callisto-formation.fr>

6. Voir dans ce numéro l'article de Véronique Cohoner et Juliette Hueber, p. 101.

7. <https://umrheritages.hypotheses.org/>

8. <https://docheritages.hypotheses.org/>

BÉROSE, UNE EXPÉRIENCE ENCYCLOPÉDIQUE UNIQUE **EN HUMANITÉS NUMÉRIQUES**

Bérose est une encyclopédie internationale des histoires de l'anthropologie. Pionnière dès 2004, nativement en science ouverte, multilingue, unique en son genre dans le paysage international des humanités numériques, *Bérose* occupe une niche thématique riche, avec un fort potentiel de développement du savoir cumulatif et de publication de recherches inédites. Reconnue au niveau international, *Bérose* a connu un développement scientifique accéléré ces cinq dernières années et une production éditoriale intense. Depuis ses débuts, *Bérose* bénéficie du soutien du ministère de la Culture.

CHRISTINE LAURIÈRE

Chargée de recherche au CNRS, ex-codirectrice de l'UMR 9022 Héritages, codirectrice de *Bérose* – *Encyclopédie internationale des histoires de l'anthropologie* et de la collection des *Carnets de Bérose*

Une place exceptionnelle dans l'histoire de l'anthropologie
Bérose a conquis une place d'excellence dans le domaine de l'histoire de l'anthropologie, à l'échelle française et internationale. C'est le résultat de projets scientifiques réussis (« Action concertée incitative », Agence nationale de la recherche [ANR], programme de numérisation de revues piloté par la Bibliothèque nationale de France [BnF et Gallica, site internet]), dont Daniel Fabre, Claudie Voisenat et Jean-Christophe Monferran ont été les initiateurs. Depuis 2017, avec une nouvelle codirection franco-portugaise

(Christine Laurière et Frederico Delgado Rosa), le projet a changé de dimension : capitalisant sur son développement passé, *Bérose* est devenue une encyclopédie en ligne d'envergure mondiale revendiquant une pratique et une écriture renouvelées de l'histoire des anthropologies, plurielle, qui prennent acte du tournant numérique dans les sciences humaines et sociales. *Bérose* est tout à la fois une structure éditoriale publiant des articles scientifiques dans l'encyclopédie et des ouvrages dans la collection des *Carnets de Bérose*, un outil collaboratif de recherche, une biblio-

thèque de ressources numérisées et une plateforme de dossiers documentaires numériques faisant l'objet d'une construction scientifique mûrement réfléchie. Soutenue par des partenaires et des institutions scientifiques français (ministère de la Culture, CNRS, UMR 9022 Héritages depuis 2021), *Bérose* est hébergée par IR* Huma-Num¹, très grande infrastructure de recherche dans le domaine des humanités numériques, qui assure la pérennité de son fonctionnement et de ses contenus. Doté d'un conseil scientifique international², d'une vingtaine d'équipes de recherche³

1. Voir dans ce numéro l'article d'Olivier Baude et Nicolas Larrousse, p. 11.

2. <https://www.berose.fr/article1065.html>

3. <https://www.berose.fr/rubrique964.html>



Günter Tessmann et son phonographe sur la véranda à Nkolentangan, Guinée équatoriale, Noël 1907.

© Völkerkundesammlung der Hansesstadt Lübeck

et d'un réseau de contributeurs⁴ en expansion constante provenant de tous les continents, *Bérose* est un éditeur scientifique (ISSN 2648-2770) publiant tout au long de l'année des articles en plusieurs langues (français, anglais, espagnol, portugais et italien), qui ont fait l'objet d'une évaluation non anonyme par les directeurs et le comité éditorial. Entre 2017 et 2023, près de 280 articles ont été publiés (environ 8 500 pages), signés par plus de 240 auteurs. *Bérose* a créé plus de 250 dossiers documentaires. La fréquentation mensuelle croît régulièrement et avoisine les 15 000 visites. La collection des *Carnets de Bérose* (ISSN 2266-1964, 13 volumes parus⁵), référencée dans le SUDOC⁶, totalise plus de 45 000 consultations.

Des projets scientifiques et techniques pour rayonner

Bérose a signé un partenariat scientifique avec la revue étatsunienne *History of Anthropology Review*, qui relaie l'annonce des publications de *Bérose* à son lectorat anglophone, ces publications étant également diffusées *via* une newsletter trimestrielle.

En octobre 2020, *Bérose* a signé une convention avec le ministère de la Culture et la Huit production, mettant ainsi progressivement en ligne la collection de 25 films d'entretiens réalisés avec des anthropologues français, « L'ethnologie en héritage⁷ ». Grâce à une subvention du Fonds national pour la science ouverte (FNSO) obtenue pour un projet intitulé FORBEROSE, débuté en mars 2022 pour une durée de trois ans, un important saut qualitatif va permettre une meilleure mise en œuvre des principes FAIR.

Il s'agit d'améliorer et de consolider le référencement, la visibilité et l'indexation des contenus ; de renforcer la qualité des métadonnées et leur interopérabilité ; de mettre en place des pratiques numériques vertueuses ; et de standardiser nos capacités en matière de publication numérique ouverte. Côté scientifique, *Bérose* est aussi lauréat

Affiche d'une rencontre au Musée du quai Branly pour la présentation d'un *Carnet de Bérose*.

pour 2022-2026 d'un *International Research Network* (IRN) du CNRS, « Histoire transatlantique des anthropologies d'Amérique latine⁸ », consortium rassemblant neuf institutions latino-américaines et européennes.

La science ouverte, terreau idéal pour une encyclopédie innovante

L'écosystème de la science ouverte constitue le terreau idéal pour des initiatives scientifiques de type encyclopédique, grâce à ses qualités inhérentes : souplesse, évolutivité, interactivité et intégration de contenus documentaires multimédia complémentaires, lectorat accru. Par définition interminable, constamment enrichie, la publi-

cation numérique d'une encyclopédie comme *Bérose* montre des atouts indéniables par rapport à une édition imprimée.

Initiative française de portée internationale, *Bérose* contribue au rayonnement de la science ouverte francophone et démontre sa capacité à s'adresser à des publics divers (étudiants, professeurs, chercheurs, grand public curieux) de nombreuses nationalités. Peu visible dans le secteur de l'édition française imprimée, l'histoire de l'anthropologie gagne en rayonnement scientifique grâce à son inscription dans un réseau de science ouverte internationale qui en démultiplie les auteurs, les chercheurs et les lecteurs, leur donnant ainsi conscience de ce qu'ils représentent. ■

4. <https://www.berose.fr/article1075.html>

5. <https://www.berose.fr/rubrique240.html>

6. Le catalogue du Système universitaire de documentation (SUDOC) est le catalogue collectif français réalisé par les bibliothèques et centres de documentation de l'enseignement supérieur et de la recherche.

7. <https://www.berose.fr/article2530.html>

8. <https://www.berose.fr/article2525.html>

Apprendre à penser science ouverte dans un laboratoire de recherche en architecture : l'UMR CNRS Ambiances Architectures Urbanités (AAU)¹

La science ouverte infuse aujourd'hui toute la vie scientifique d'un laboratoire. Si elle a d'abord concerné la diffusion des résultats de la recherche, avec le libre accès aux publications, elle déborde désormais largement ce cadre pour se disséminer dans toutes les activités de recherche.

FRANÇOISE ACQUIER

Ingénieure d'études au CNRS, chargée de ressources documentaires-appui à la recherche, UMR AAU Centre de recherche sur l'espace sonore & l'environnement urbain (CRESSON)

LAURENCE BIZIEN

Chargée d'études documentaires au ministère de la Culture, responsable du centre documentaire recherche-appui à la recherche, UMR AAU Centre de recherche nantais architectures urbanités (CRENAU)

LOUISE-ANNE CHARLES

Bibliothécaire, chargée de l'organisation des services aux chercheurs du BnF DataLab, alternante au CRESSON en 2021-2022

Ouvrir la science : de l'ouverture des publications à la science ouverte

Comment peut se positionner au sein de ce mouvement un laboratoire de recherche en architecture et urbanisme ? Quelle politique mener dans un laboratoire commun à deux Écoles nationales supérieures d'architecture (ENSA) pour la science ouverte ? Retour d'expérience, notre article analyse la mise en place progressive de cette démarche.

Comme dans toutes les bibliothèques universitaires devant composer avec la croissance du coût de la documentation scientifique, les catalogues de nos centres documentaires ont rapidement signalé des ressources « gratuites », dont des revues en ligne, en participant notamment à la veille collaborative des réseaux ArchiRèS² puis Mir@bel³.

Le développement des revues en accès libre a été concomitant de celui des archives ouvertes, avec les premiers dépôts de l'unité mixte de recherche (UMR) Ambiances Architectures Urbanités (AAU) en 2006 dans la plateforme d'archivage HAL-SHS (sciences de l'homme et de la société). Correspondantes pour l'information scientifique et technique (IST) de l'Institut des sciences humaines et sociales (INSHS) du CNRS, les documentalistes ont dès lors accompagné les chercheurs dans leurs pratiques d'auto-archivage, tout en alimentant les collections par des dépôts massifs. Aux campagnes de numérisation des thèses et rapports de recherche du Centre de recherche sur l'espace sonore & l'environnement urbain (CRESSON)⁴ et de la revue nantaise *Cahiers du LAUA/Lieux communs*⁵ ont succédé une deuxième collection « Revue »⁶ et la page unifiée⁷ des actes des congrès du réseau international Ambiances. S'ajoutent à cet « écosystème » les mémoires de Master encadrés par le Centre de recherche nantais architecture urbanités (CRENAU) versés dans la collection DUMAS (Dépôt universitaire de mémoires après soutenance) de l'École nationale supérieure d'architecture de Nantes⁸.

Au moment de créer la revue *Ambiances*⁹ en 2013, le laboratoire a privilégié le portail *Revue.org* du Centre pour l'édition électronique ouverte (CLÉO) (devenu OpenEdition Journals) et le modèle de publication « diamant » (accès ouvert et sans frais). La plateforme de blogging scientifique *Hypothèses* a été largement investie par les équipes depuis un premier *Carnet* en 2009 jusqu'aux dix *Carnets* actuels¹⁰.

Cette dynamique installée demande néanmoins d'être soutenue pour la prise en compte par les

1. AAU comprend le CRESSON à Grenoble et le CRENAU à Nantes.

2. ArchiRèS est un réseau de bibliothèques (dont des centres de documentation de laboratoire), d'écoles d'architecture et de paysage, et de partenaires : les vingt Écoles nationales supérieures d'architecture (ENSA) du ministère de la Culture ; l'École spéciale d'architecture ; l'École Camondo ; l'Université catholique de Louvain – Faculté d'architecture, d'ingénierie architecturale, d'urbanisme LOCI, site de Bruxelles ; l'Université libre de Bruxelles – Faculté d'architecture ; l'Académie libanaise des Beaux-Arts (ALBA) ; l'École nationale d'architecture de Rabat ; la Cité de l'architecture et du patrimoine : <https://www.archires.archi.fr/labos>

3. Pour chaque revue, Mir@bel propose des liens vers le texte intégral en ligne, les sommaires, les résumés ou l'indexation des articles et permet de rebondir sur de nombreux sites complémentaires : <https://reseau-mirabel.info/site/page/presentation> (voir dans ce numéro l'article de Cyrielle Avisse et Yannick Bernardie, p. 124).

4. Les rapports figurent dans la collection HAL du Bureau de la recherche architecturale, urbaine et paysagère (BRAUP). <https://lcv.hypotheses.org/13731>

5. <https://hal.science/REVUE-LIEUX-COMMUNS/>

6. <https://shs.hal.science/REVUE-TROUBLE-DANS-LES-COLLECTIONS/>

7. <https://hal.science/RESEAUINTERNATIONALAMBIANCES/>

8. Voir dans ce numéro l'article de Frédéric Gendre, Christophe Hugot, Joachim Schopfel et Agnès Souchon, p. 49, et le focus de Marjorie Borios, p. 52 : <https://dumas.ccsd.cnrs.fr/MEM-ENSA-44/>

9. <https://journals.openedition.org/ambiances/>

10. <https://hal.science/hal-04004646v1>

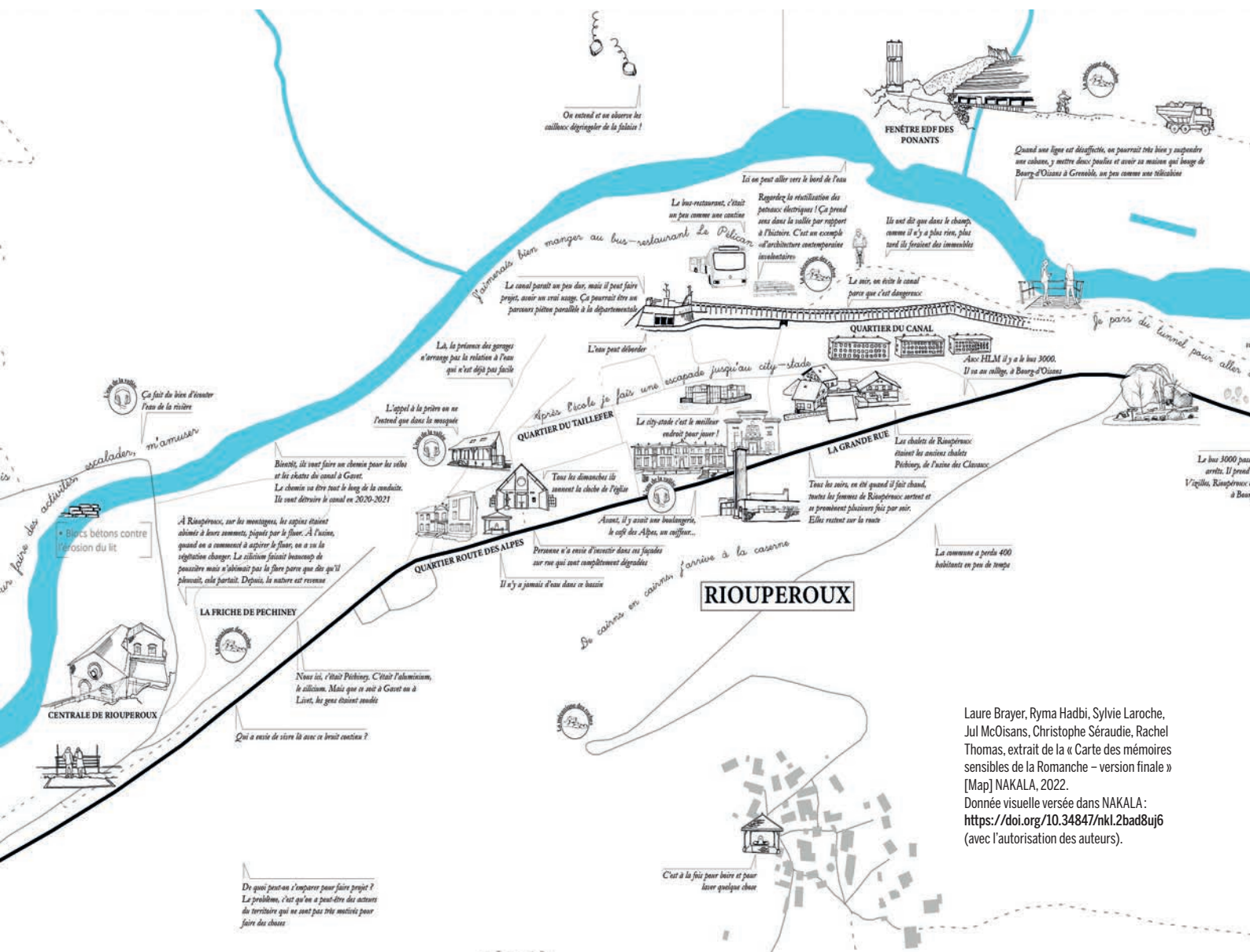
Dans le discours des chercheurs, la science ouverte reste fortement reliée au mouvement du libre accès des publications, la gestion des données de la recherche apparaissant comme un chantier naissant très différent.

chercheurs des modalités concrètes d'une science ouverte, comme l'ont démontré les questionnements sur le libre accès lors des 9^e Journées d'études du réseau Médecin¹¹, accueillies à Nantes en 2019, ou l'étude réalisée en 2022 au CRESSON¹². Dans le discours des chercheurs, la science ouverte reste fortement liée au mouvement du libre accès des publications, la gestion des données de la recherche apparaissant comme un chantier naissant très différent. Le périmètre de la science ouverte, s'appliquant à toutes les étapes de la recherche, est un défi pour les accompagnants et les chercheurs.

L'engagement du laboratoire

En mentionnant en 2020 dans son règlement intérieur l'obligation de déposer dans HAL les publications, le laboratoire ouvrait la voie à la formalisation de sa politique en faveur de la science ouverte. Un petit groupe se constitue autour des résultats d'enquêtes¹³ sur les usages et besoins en matière de gestion des données, afin d'organiser des actions de sensibilisation (présentations de l'infrastructure de recherche

11. <https://www.canal-u.tv/chaines/medici/journees-medicin/9es-journees-medicin-nantes-24-26-septembre-2019>
12. Louise-Anne Charles, *Apprendre à penser en données de la recherche : le cas de l'équipe de recherche en SHS, AAU-Cresson*, Mémoire de Master 2 SIBIST, ENSSIB, 2022.
13. <https://labeledoc.hypotheses.org/5990>



Laure Brayer, Ryma Hadbi, Sylvie Laroche, Jul McOisans, Christophe Séraudie, Rachel Thomas, extrait de la « Carte des mémoires sensibles de la Romanche – version finale » [Map] NAKALA, 2022. Donnée visuelle versée dans NAKALA: <https://doi.org/10.34847/nkl.2bad8uj6> (avec l'autorisation des auteurs).

14. OPIDoR : Outil d'aide à la création en ligne de plans de gestion de l'Institut de l'information scientifique et technique (INIST) du CNRS. <https://tour.opidor.fr/nantes/>
15. <https://aau.archi.fr/laboratoire-aau/science-ouverte/>
16. https://aau.archi.fr/uploads/2022/09/Guide_donnees_FA_LC_01_09_2022.pdf
17. <https://scienceouverte.univ-grenoble-alpes.fr/donnees/cellule-data/> (voir dans ce numéro l'article de Lucie Albaret et Violaine Louvet, p. 66).
18. <https://isidore.science/vocabularies>
19. <https://labedoc.hypotheses.org/?p=10424>
20. Voir dans ce numéro l'article de Roselyne Aliacar, p. 73.
21. <https://lcv.hypotheses.org/17028>
22. <https://lcv.hypotheses.org/17446> (voir dans ce numéro l'article de Laetitia Bracco, p. 26).
23. <https://www.canal-u.tv/chaines/aau>

IR* Huma-Num, séances du jeu coopératif nommé GOpenDoRe [Go open données de la recherche], accueil de l'OPIDoR Tour¹⁴) et de rédiger les 3 axes de son engagement¹⁵, publié en 2021. Pour accompagner ses membres dans la rédaction de plans de gestion des données (PGD), deux projets grenoblois servent de terrain à la rédaction d'un *Guide pratique de la gestion des données*¹⁶ et d'un PGD commenté, avec l'aide de la Cellule Data Grenoble Alpes¹⁷. Ces documents offrent une vision d'ensemble des solutions et outils existants, comme le thésaurus ArchiRès, vocabulaire du ministère de la Culture et référentiel du moteur de recherche ISIDORE donnant accès aux données numériques des SHS¹⁸, permettant l'indexation des données versées dans l'entrepôt de données pour les sciences humaines et sociales NAKALA¹⁹ et un possible moissonnage dans l'entrepôt national Recherche Data Gouv²⁰.

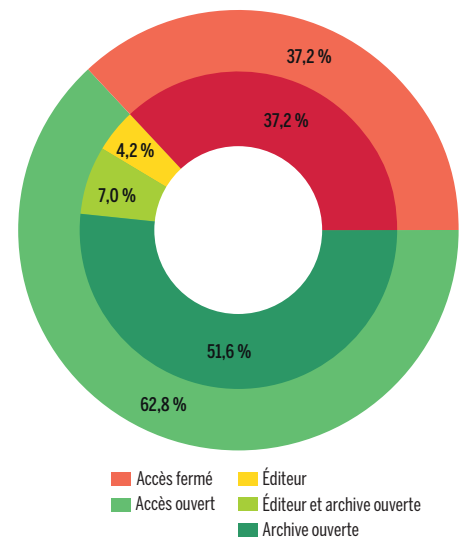
Les moteurs : entre injonctions et émulation

Différents leviers, incitatifs ou contraignants, servent de moteurs à l'accélération de la démarche. Bénéficiaires d'un financement de l'Agence nationale de la recherche (ANR), les porteurs de projet s'engagent à rendre disponibles leurs publications en libre accès et à rédiger un PGD. Le CNRS vise 100 % des publications de ses laboratoires en accès ouvert et seules les publications déposées dans HAL sont retenues pour le bilan annuel d'activité des chercheurs. Le dernier accord-cadre ministère de la Culture-CNRS demande la mutualisation et la mise à disposition des données utiles à la recherche, cette démarche intégrant par ailleurs le processus d'évaluation du Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur (Hcéres).

Des événements pour motiver les équipes sont mis en place, comme la participation en 2021 au « HALathon »²¹ national de l'association CasuHAL regroupant les utilisateurs professionnels de HAL.

Baromètre de la science ouverte du laboratoire AAU

Proportion des publications en accès ouvert entre 2015 et 2020 mesurée en janvier 2022



Licence CC-BY

Mesurer l'efficacité de l'action collective et en afficher les bons résultats, tel est aussi l'objectif de notre déclinaison du Baromètre de la science ouverte²². Dernière initiative, la chaîne²³ AAU de Canal-U diffuse les podcasts des séminaires.

Les perspectives

Grâce à l'accompagnement de différents acteurs universitaires ou du CNRS, en particulier la Maison des sciences de l'homme (MSH) Ange-Guépin et la Cellule Data Grenoble Alpes, il s'agit d'ancrer la conservation et la valorisation de nos données, afin de répondre aux critères d'intégrité et de reproductibilité, avec par exemple la rédaction de *data papers* (« articles de données »). S'investir dans un projet relatif à l'éducation ouverte est aussi envisagé... ■



Rencontre AAU 3, podcast de la chaîne Canal-U. <https://rencontreaau.hypotheses.org/?p=2200>

De la pratique quotidienne à une charte institutionnelle : la démarche vers la science ouverte de l'INHA

L'Institut national d'histoire de l'art (INHA) est un établissement public à caractère scientifique, culturel et professionnel destiné à promouvoir la recherche scientifique en histoire de l'art. Depuis plusieurs années, l'INHA s'est engagé dans une démarche de mise à disposition et d'ouverture des données (codes sources, métadonnées, publications, images) issues des programmes de recherche de l'Institut et de ses partenaires, auprès de la communauté scientifique et, plus largement, de toute la société.

Une bonne pratique de terrain (*Open Data, Open Access, Open Content, Open Source*¹)

Dès sa naissance, en 2001, l'INHA a mis en place une Cellule d'ingénierie documentaire (CID) au sein du Département des études et de la recherche (DER), et c'est à partir de 2004 qu'une réflexion a été menée autour de la gestion des données numériques de la recherche. En 2006, L'Institut s'est doté d'un outil de gestion documentaire numérique pour les données issues des programmes de recherche de l'établissement : AGORHA², accessible en ligne au public dès 2011.

Face à l'essor de ces actions, la Cellule d'ingénierie documentaire a été transformée en un Service numérique de la recherche (SNR), reprenant mais élargissant également les missions de l'ancienne Cellule.

Une nouvelle version d'AGORHA, conçue comme plateforme des données de la recherche, a été mise en ligne en 2021, avec une attention particulière portée aux principes FAIR (facile à trouver, accessible, interopérable, réutilisable), ainsi qu'à l'ouverture et au partage des données de la recherche.

La plateforme compte à ce jour près de 250 000 notices documentaires publiées, dont environ 180 000 sont illustrées. Toutes les métadonnées sont mises à disposition, depuis 2018, sous une licence ouverte Creative Commons Attribution 4.0 International (CC-BY 4.0)³; les informations concernant les droits de chaque image sont précisées dans la page où l'image apparaît, avec la précision de la source et de la licence.

Le Département de la bibliothèque et de la documentation (DBD) a mis en ligne, en 2006, une bibliothèque numérique⁴, qui met à disposition de toutes et de tous des documents numérisés de la bibliothèque de l'INHA entrés dans le domaine public (à la date du 10 mars 2023, cela représente 31 261 documents, soit 1 039 085 images). Les documents et les métadonnées associées sont placés depuis 2016 sous « Licence ouverte / *Open License* » Etalab⁵.

La bibliothèque de l'INHA signale en outre son catalogue et les métadonnées associées, en format ouvert et interopérable, dans le SUDOC⁶, le catalogue collectif du système universitaire de documentation. Elle alimente aussi le signalement des archives et des manuscrits *via* Calames⁷, catalogue en ligne des archives et des manuscrits de l'enseignement supérieur.

Une attention particulière est également portée par l'INHA sur la question des droits des images et sur l'ouverture des contenus visuels, en particulier à travers le travail mené au sein de l'Institut depuis 2017 par la mission Images/Usages sur les enjeux liés à la diffusion en ligne des images patrimoniales (ouverture, partage, accessibilité, inclusivité...). La mission a publié un *Guide pratique pour la recherche et la réutilisation des images d'œuvres d'art* en 2021⁸ et poursuit son activité à travers différentes actions de formation destinées aux professionnels.

L'INHA adhère ainsi à une politique d'ouverture des contenus visuels numériques, en promouvant le partage et la réutilisation de ses ressources et en favorisant des partenariats avec des fournisseurs de

FEDERICO NURRA

Chef du service numérique de la recherche à l'Institut national d'histoire de l'art (INHA)

1. Pour données ouvertes, accès ouvert, contenus ouverts, sources ouvertes (logiciels-outils libres).
2. <https://agorha.inha.fr/>
3. <https://creativecommons.org/licenses/by/4.0/deed.fr>
4. <https://bibliotheque-numerique.inha.fr/>
5. Dans le cadre de la politique du gouvernement en faveur de l'ouverture des données publiques (*Open Data*), Etalab a conçu la « Licence ouverte / *Open License* ».
6. <http://www.sudoc.abes.fr>
7. <http://www.calames.abes.fr>
8. Librement téléchargeable à l'adresse : https://www.inha.fr/_resources/PDF/2021/2021_LIVRET_GUIDE-PRATIQUE_web_15062021.pdf

Filtrer les résultats

Filtres généraux ▼ Appliquer

Filtres additionnels ▼ Appliquer

Date d'exécution ▼

Lieu de création ▼

Lieu de découverte ▼

Lieu de conservation ▼





Type d'œuvre ▼

Sujet ▼

Matériaux ▼

Techniques ▼

Degrés d'attribution ▼

	<p>Œuvre / Peinture LA VIERGE ET L'ENFANT AVEC LE PETIT SAINT JEAN-BAP [L-]</p> <p>Auteur : Reni Guido, copié d'après</p> <p>Date de création : 19e siècle</p> <p>Lieu : Eglise paroissiale Saint-Saturnin (Pont-Saint-Esprit)</p>	<p>Modifiée le : 02/12/2021</p> <p>Notice associée à la base Répertoire des tableaux italiens dans les collections publiques françaises (XIIIe-XIXe siècles), RETIF</p>
	<p>Œuvre / Peinture SAINTE MARIE MADELEINE PÉNITENTE</p> <p>Auteur : Cavedone Giacomo, entourage de</p> <p>Date de création : 18e siècle</p> <p>Lieu : Musée des Beaux-Arts et d'Archéologie (Besançon)</p> <p>Cote / Numéro : INV. 984.1.26 (numéro d'inventaire)</p>	<p>Modifiée le : 31/10/2022</p> <p>Notice associée à la base Répertoire des tableaux italiens dans les collections publiques françaises (XIIIe-XIXe siècles), RETIF</p>
	<p>Œuvre / Peinture LE SUICIDE DE CLÉOPÂTRE</p> <p>Auteur : Dandini Cesare, attribué à</p> <p>Date de création : 17e siècle</p> <p>Lieu : Musée des Beaux-Arts (Dunkerque)</p> <p>Cote / Numéro : P.27 (numéro d'inventaire)</p>	<p>Modifiée le : 25/10/2022</p> <p>Notice associée à la base Répertoire des tableaux italiens dans les collections publiques françaises (XIIIe-XIXe siècles), RETIF</p>
	<p>Œuvre / Peinture PORCIA</p> <p>Auteur : Anonyme florentin (peintre) - 17e siècle, attribué à</p> <p>Date de création : 17e siècle</p> <p>Lieu : Musée d'Art et d'Histoire (Chaumont)</p> <p>Cote / Numéro : 1859-31 (numéro d'inventaire)</p>	<p>Modifiée le : 25/10/2022</p> <p>Notice associée à la base Répertoire des tableaux italiens dans les collections publiques françaises (XIIIe-XIXe siècles), RETIF</p>

Liste de résultats d'une recherche dans la plateforme AGORHA en mode d'affichage « cartouche ».

Collections numérisées de la bibliothèque de l'INHA
Institut national d'histoire de l'art
INHA


EXPOSITIONS VIRTUELLES COLLECTIONS SERVICES
Rechercher...

Documents

258 résultat(s)

Filtré par : Estampes > Estampes > Fonds d'estampes d'artistes > Estampes d'Henri de Toulouse-Lautrec (1864-1901) > Estampe

Résultats par page : 10 Trié par : Pertinence Notices courtes

☆  **Yvette Guilbert : à Menilmontant de Bruant**

Auteur : Toulouse-Lautrec, Henri de (1864-1901)
Bruant, Aristide (1851-1925)

Période : 19e siècle


Date : 1898

Sujet : Guilbert, Yvette (1865-1944) -- Portraits

Identifiant numérique : NUM Mfilm 347 (15, 320)

Bibliothèque d'origine : Bibliothèque de l'Institut National d'Histoire de l'Art, collections Jacques Doucet

Corpus : Estampes > Fonds d'estampes d'artistes > Estampes d'Henri de Toulouse-Lautrec (1864-1901)

☆  **L'Anglais au Moulin Rouge**

Auteur : Toulouse-Lautrec, Henri de (1864-1901)

Période : 19e siècle

Date : 1892

Sujet : Warrener, William (1861-1934) -- Portraits

Identifiant numérique : NUM Mfilm 347 (15, 166)

Bibliothèque d'origine : Bibliothèque de l'Institut National d'Histoire de l'Art, collections Jacques Doucet

Corpus : Estampes > Fonds d'estampes d'artistes > Estampes d'Henri de Toulouse-Lautrec (1864-1901)

Tranche chronologique

[Progression bar]

Date de publication

[Dropdown menu]

Corpus

Estampes

Fonds d'estampes d'artistes 258

Affiches du Salon des Cent 4

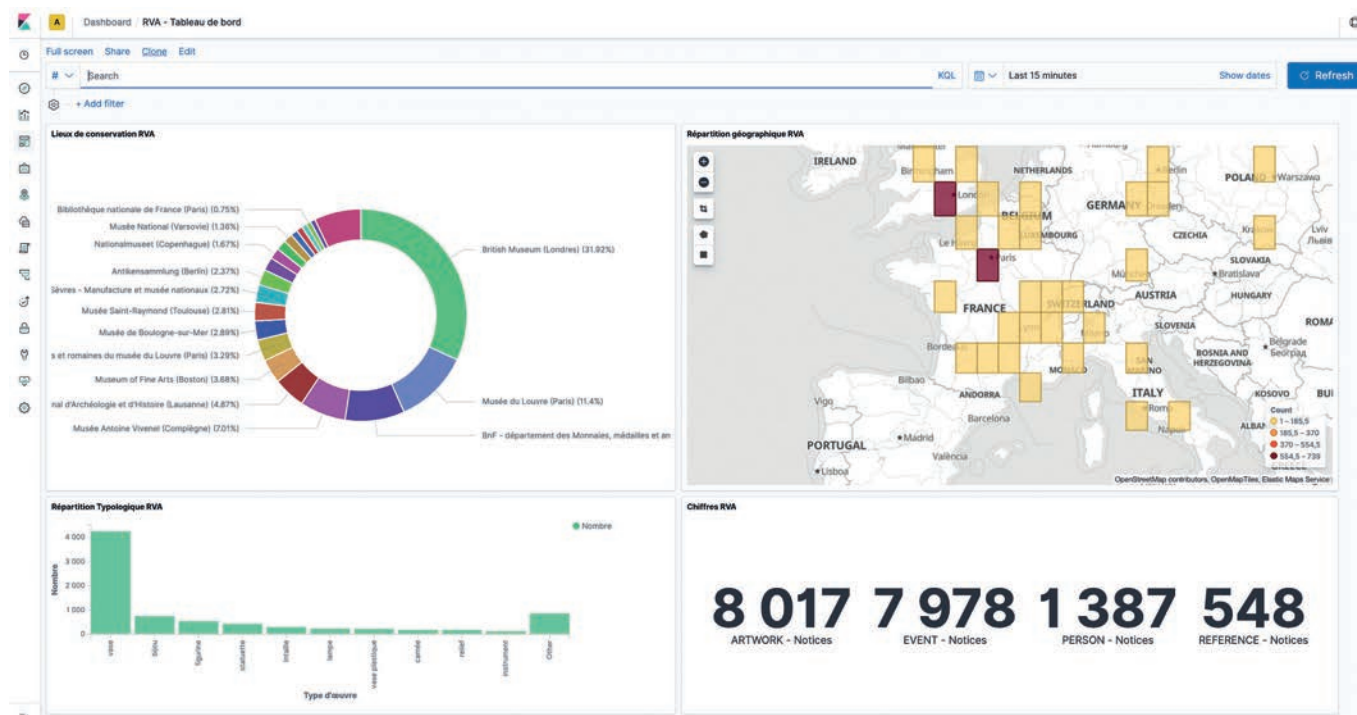
Type de document

Estampe

Auteur

Toulouse-Lautrec, Henri de (1864-1901)	258
Pellet, Gustave (1859-1919)	5
Goudezki, Jean (1866-1977)	4
Donnay, Maurice (1859-1945)	3
Théâtre de l'Oeuvre (Paris)	3
Barrucand, Victor (1864-1934)	2
Bruant, Aristide (1851-1925)	2
Byl, Arthur (1877-1908)	2
Théâtre Antoine-Simone Berriau (Paris)	2
Adhémar, Jean (1908-1987)	1

Liste de résultats d'une recherche dans les collections numérisées de la bibliothèque de l'INHA.



documentation visuelle qui mettent à disposition les images et les contenus visuels en licence ouverte (notamment, l'INHA demande l'attribution de licences ouvertes Creative Commons sur les images mises à disposition par les musées partenaires).

Dans le cadre de la production d'outils numériques pour la gestion, de la mise en ligne et de la valorisation des données de la recherche, et dans un souci de transparence et de garantie envers les développements numériques, l'INHA s'est engagé à publier le code source utilisé dans le cadre des recherches de l'Institut, par le biais du dépôt GIT institutionnel⁹. L'INHA encourage la mise à disposition du code source de toute production numérique issue des programmes de recherche de l'Institut et soutient la diffusion et la réutilisation du code source pour accroître l'impact scientifique de ses travaux.

Une charte pour promouvoir les valeurs de la science ouverte

Cette expérience « sur le terrain » et cette pratique quotidienne nous ont conduits à nous interroger sur l'importance d'institutionnaliser ces démarches d'ouverture et de mise à disposition des données de la recherche ainsi que des publications scientifiques, que l'Institut a pratiquées, d'une façon « informelle », tout au long de sa vie.

L'INHA adhère aux principes promus par la « science ouverte » (*Open Science*), en accord avec le Plan national pour la science ouverte (PNSO) : rendre la recherche scientifique et les données produites accessibles à toutes et à tous, qu'il s'agisse des publications scientifiques, des données de la recherche ou des méthodes et des outils de travail.

C'est pour cette raison que la Direction générale de l'INHA a initié, au printemps 2022, la rédaction d'une *Charte d'établissement pour la science ouverte*.

Cette charte, validée par un comité de relecture interservices et adoptée à l'unanimité par le conseil scientifique de l'Institut le 10 mars 2023, sera révisée et mise à jour périodiquement, en adéquation avec les transformations du monde numérique et les évolutions des plans nationaux et internationaux pour la science ouverte.

En adéquation avec le deuxième Plan national pour la science ouverte du ministère de l'Enseignement supérieur et de la Recherche (2021-2024) et avec le soutien du ministère de la Culture, cette charte formalise une stratégie de mise à disposition et d'ouverture des données, des publications scientifiques et des codes sources des outils développés par l'Institut.

Afin de consolider cette démarche, un plan de formations aux pratiques et aux outils numériques en histoire de l'art et en archéologie a été mis en œuvre en 2023. Ces formations, destinées à la communauté des historiennes et des historiens de l'art ainsi qu'aux archéologues, mobilisent toutes les composantes de l'Institut. Ce plan a pour but de former et de sensibiliser cette communauté aux enjeux de la science ouverte, de promouvoir une gestion responsable de l'identité numérique, d'inciter à l'utilisation de modèles, formats et standards qui permettent l'interopérabilité et la pérennité des données, d'accompagner la diffusion ouverte des procédés et des résultats de la recherche publique menée au sein de l'établissement.

Avec cette charte, l'INHA souhaite donc adhérer pleinement aux valeurs promues par la science ouverte, afin de rendre la recherche scientifique en histoire de l'art plus accessible, transparente, collaborative et efficace, et de favoriser une participation large de la communauté scientifique et du public aux résultats de nos recherches. ■

Visualisation, dans un tableau de bord, des données produites dans le cadre d'un programme de recherche de l'INHA.

9. Un dépôt GIT est un entrepôt virtuel de projets : <https://gitlab.inha.fr/>

Comment mesurer l'efficacité d'une politique dans le domaine de la recherche ?

Un exemple avec le Baromètre lorrain de la science ouverte

La science ouverte est un domaine qui a été largement investi par les bibliothèques universitaires depuis de nombreuses années. D'abord mouvement militant, elle s'est institutionnalisée jusqu'à faire l'objet d'un Plan national pour la science ouverte (PNSO) lancé par le ministère de l'Enseignement supérieur et de la Recherche. L'impulsion politique ayant été donnée, il était nécessaire de mettre en œuvre un outil de suivi de l'efficacité de ces actions. C'est le point de départ du Baromètre français de la science ouverte (BSO), dont la première version portant sur l'ouverture des publications scientifiques a été publiée dans la foulée.

LAETITIA BRACCO

Conservatrice des bibliothèques, responsable de l'atelier de la donnée ADOC Lorraine et du pôle bibliométrie, cheffe de projet Baromètre français de la science ouverte pour les données de la recherche et codes logiciels, mission appui recherche - Direction de la documentation de l'Université de Lorraine.

Cet outil a rapidement intéressé les établissements et notamment les bibliothèques universitaires ; l'Université de Lorraine y a vu une occasion de mesurer sa propre action en matière de science ouverte, en développant la première déclinaison locale de cet outil. Repris désormais par de nombreux établissements, ce code a permis de faire fleurir de multiples Baromètres locaux. En 2022, le Baromètre a vu son périmètre s'étendre : ce sont à présent les données et codes logiciels issus du travail de recherche scientifique qui font l'objet d'une étude approfondie pour en observer l'ouverture, et une communauté se fédère au sein du club utilisateurs du Baromètre.

C'est en 2018 qu'a été publiée en France la première version du Plan national pour la science ouverte (PNSO)¹. La science ouverte y était alors désignée comme « la diffusion sans entrave des publications et des données de la recherche » par opposition, d'une part, aux publications scientifiques accessibles uniquement sur abonnement et, d'autre part, aux données de la recherche qui restent sur les disques durs de leurs producteurs. Dès lors, afin d'évaluer l'efficacité de cette politique publique, il était nécessaire de développer un outil de suivi pour mesurer les effets de cette stratégie. Or rapidement, ce besoin s'est également fait ressentir

au sein des établissements de l'enseignement supérieur et de la recherche, notamment pour ajuster les actions de sensibilisation des services d'appui à la recherche auprès de leur communauté scientifique. C'est tout l'enjeu du développement du Baromètre lorrain de la science ouverte.

Décliner à l'échelle locale des indicateurs nationaux

En 2018, le Département des outils d'aide à la décision du ministère de l'Enseignement supérieur et de la Recherche publiait la première édition du BSO. Il proposait alors des graphiques permettant d'apprécier l'ouverture des publications scientifiques (articles, chapitres d'ouvrages, actes de conférences...) par année, par éditeur ou encore par discipline. À la suite de la publication de cet outil de mesure ouvert (*Open Source*), ne s'appuyant pas sur des outils commerciaux, l'Université de Lorraine a rapidement manifesté son fort intérêt, en tant qu'établissement résolument engagé en faveur de la science ouverte².

Cependant, la transposition de cet outil national à l'échelle d'un établissement a nécessité un important travail d'adaptation mené au sein des bibliothèques universitaires. La méthodologie pour réaliser et mettre

1. <https://www.enseignementsup-recherche.gouv.fr/fr/le-plan-national-pour-la-science-ouverte-les-resultats-de-la-recherche-scientifique-ouverts-tous-49241>

2. <https://scienceouverte.univ-lorraine.fr/>

https://www.metmuseum.org/art/collection/search/436106?searchField=All&sortBy=relevance&what=Oil+paint&showOnly=openAccess&ip=science&ipfise=0&ipp=40&pos=1



Jacques-Louis David, *Antoine Laurent Lavoisier (1743-1794) et Marie Anne Lavoisier (Marie Anne Pierrette Paulze, 1758-1836), 1788*, Metropolitan Museum of Art, New-York.

en ligne cet outil ayant fait l'objet d'un article détaillé³, il n'est pas nécessaire de s'y attarder. Il faut néanmoins souligner que le défi le plus important à relever pour un établissement ou une unité de recherche qui souhaite suivre la progression de la science ouverte réside dans la constitution de son corpus, afin d'obtenir la liste la plus complète possible de ses publications scientifiques.

En effet, la méthodologie employée en France pour constituer le corpus national, basée sur le *scraping* de pages Web⁴ pour rassembler un maximum de publications avec au moins un auteur français, n'est pas adaptée pour obtenir le corpus fin d'une université ou d'un laboratoire. Le dépôt systématique de l'ensemble des publications dans l'archive ouverte HAL⁵ reste le meilleur moyen pour obtenir facilement un corpus à une échelle restreinte.

La méthodologie employée par le BSO lorrain étant publique et ouverte, ce sont rapidement des dizaines d'autres établissements d'enseignement supérieur⁶ qui ont réutilisé l'outil pour faire leur propre BSO.

Portée et usages de l'outil

Que faire des résultats obtenus ? Sont-ils des indicateurs qui pourraient servir à évaluer une unité, voire un chercheur-auteur ? Ce serait prendre à rebours la logique de l'outil. Car le BSO, qu'il soit national ou *a fortiori* local, n'a pas vocation à attribuer des notes, mais bien à identifier la progression et les besoins des communautés scientifiques en matière d'ouverture de leur recherche.

De plus, le BSO permet également de mieux connaître les pratiques de publications des chercheurs ;

par exemple, d'établir auprès de quels éditeurs ou revues telle ou telle discipline a l'habitude de publier. Le BSO peut aussi être un outil de pilotage en matière de politique documentaire, notamment pour le choix des abonnements (ou désabonnements) à des ressources électroniques.

Élargir les indicateurs aux données et aux logiciels

On l'a vu, le BSO a d'abord porté uniquement sur les publications scientifiques. Or, dès la première version du PNSO, les données de la recherche puis les logiciels (deuxième PNSO⁷) sont considérés comme des productions scientifiques à part entière, dont l'ouverture doit également être encouragée.

En 2021, le MESR, l'Université de Lorraine et l'INRIA (Institut national de recherche en informatique et en automatique) ont ainsi obtenu un financement du Plan de relance⁸, afin de porter le projet d'extension du BSO aux données de la recherche et aux codes logiciels. Ce projet, toujours en cours, est jalonné par deux grandes étapes :

- constituer la base la plus complète possible de publications françaises en texte intégral depuis 2013 et y appliquer des méthodes de *machine learning*⁹ pour y détecter des mentions de données et de logiciels ;
- établir le corpus des jeux de données publiés par les chercheurs français et déposés dans des entrepôts de données tels que Recherche Data Govu¹⁰.

Dans les deux cas, il s'agira de mesurer le taux d'utilisation, de partage et d'ouverture des données et logiciels, dans la même logique que celle qui sous-tend les indicateurs relatifs aux publications.

Cette extension va de pair avec l'évolution de l'évaluation de la recherche, telle qu'exprimée au sein de la déclaration de San Francisco¹¹, qui appelle à « tenir compte de la valeur et de l'impact de tous les résultats de travaux de recherche (y compris les jeux de données et les logiciels) en plus des publications scientifiques ». La reconnaissance de la valeur scientifique de ces productions est en effet un levier fort pour leur ouverture et leur partage.

Une communauté d'utilisateurs qui se structure

La mise en œuvre des BSO français, puis lorrain, a entraîné le lancement d'une communauté d'utilisateurs toujours plus grande¹², qui se structure aujourd'hui au sein du Club utilisateurs du BSO.

Les établissements représentés viennent de tous les horizons : universités, organismes de recherche, écoles, voire unités de recherche, comme l'UMR Ambiances Architectures Urbanités (AAU), au sein des Écoles nationales supérieures d'architecture (ENSA) de Nantes et Grenoble¹³.

Autour d'une liste de diffusion, cette communauté échange sur ses pratiques en matière de constitution de corpus, de présentation des résultats ou encore d'usage des indicateurs. Des webinaires sont régulièrement organisés pour présenter les avancées du BSO national et pour susciter le débat. C'est ainsi un cercle vertueux qui se construit. ■

3. Laetitia Bracco, « Mesurer l'ouverture de la science : le cas de l'Université de Lorraine », *Revue française des sciences de l'information et de la communication* [en ligne], n° 24, 2022, mis en ligne le 24 mars 2022, consulté le 1^{er} février 2023. <http://journals.openedition.org/rfsic/12474> ; DOI : <https://doi.org/10.4000/rfsic.12474>

4. Technique permettant d'extraire automatiquement des informations présentes sur une page Web (par exemple, le titre et l'auteur d'un article) pour les placer dans une base de données.

5. <https://hal.science/>

6. Par exemple : les universités d'Évry, Lyon 1, Grenoble Alpes, Rennes 1, Paris-Saclay... Mais aussi des organismes de recherche comme l'INRAE ou le CEA.

7. <https://www.ouvrirlascience.fr/deuxieme-plan-national-pour-la-science-ouverte-pns0/>

8. Fonds « Transformation numérique de l'État et des territoires », volet « Développer l'utilisation de la donnée dans votre administration ».

9. Ou « apprentissage automatique » : champ de l'intelligence artificielle dédié à l'apprentissage automatique, par une machine, de certaines opérations. Par exemple, la rendre capable de déterminer la discipline d'un article scientifique en l'analysant.

10. Voir dans ce numéro l'article de Roselyne Aliacar, p. 73. <https://recherche.data.gouv.fr/fr>

11. <https://sfdroa.org/read/read-the-declaration-french/>

12. <https://barometredelascienceouverte.esr.gouv.fr/declinaisons/bs0-locaux>

13. Voir dans ce numéro l'article de Françoise Acquier, Laurence Bizien et Louise-Anne Charles, p. 20.

Un projet européen de *Cloud* collaboratif pour le patrimoine culturel

Sous l'impulsion du Parlement européen et sur la base d'un rapport d'experts consultable en ligne¹, la Commission européenne a lancé un projet de *Cloud* collaboratif pour le patrimoine culturel (*European Collaborative Cloud for Cultural Heritage – ECCCH*). Inscrite dans le paysage de la science et des données ouvertes, la conception de ce nouvel instrument à destination des professionnels du patrimoine et des chercheurs résultera d'une série d'appels à projets lancés sur la période 2023-2027 dans le cadre du programme Horizon Europe et bénéficiera d'un important financement à hauteur de plus de 110 millions d'euros.

LIVIO DE LUCA

Directeur de recherche au CNRS et directeur de l'UMR CNRS/ministère de la Culture : MAP (Modèles et simulations pour l'architecture et le patrimoine)

PASCAL LIÉVAUX

Adjoint au chef de la délégation/Direction générale des patrimoines et de l'architecture/Délégation à l'inspection, à la recherche et à l'innovation/ministère de la Culture

Pourquoi un *Cloud* sur le patrimoine culturel ?

Comme l'a bien mis en lumière la manifestation européenne organisée par la Fondation des sciences du patrimoine en collaboration avec le ministère de la Culture et le CNRS dans le cadre de la présidence française du Conseil de l'Union européenne les 15 et 16 mars 2022², le patrimoine est au cœur des grandes problématiques contemporaines, qu'elles soient sociales, économiques ou environnementales. La recherche pour la connaissance, la conservation, la restauration et la transmission du patrimoine constitue désormais un domaine scientifique à part entière désigné sous le nom de « sciences du patrimoine » ou *heritage science*. Ces travaux, menés par des chercheurs de champs disciplinaires très divers appartenant aux sciences humaines et sociales, aux sciences expérimentales ou aux sciences du numérique, produisent une quantité considérable de données, tant aux niveaux nationaux qu'euro péen.

On sait par ailleurs la grande richesse et diversité du patrimoine en Europe, qu'il soit matériel, immatériel ou numérique, ainsi que le grand nombre d'institutions qui en ont la responsabilité (musées, centres d'archives, bibliothèques, sites archéologiques, Monuments historiques, centres de recherche, etc.). Ces dernières sont de grandes productrices de connaissance qu'elles restituent sous des formes très variées (inventaires, catalogues raisonnés, catalogues d'expositions, articles scientifiques, etc.). Ces connaissances sont pour une grande part organisées sous la forme de bases de données conservées dans les établissements, sans garantie de conservation à très long terme ni grandes possibilités de partage. Au-delà

de cette production interne, un nombre croissant de professionnels du patrimoine (conservateurs, restaurateurs, régisseurs, architectes, archéologues, etc.) est impliqué dans des projets de recherche, en association avec des scientifiques extérieurs à leur institution, eux-mêmes producteurs de données.

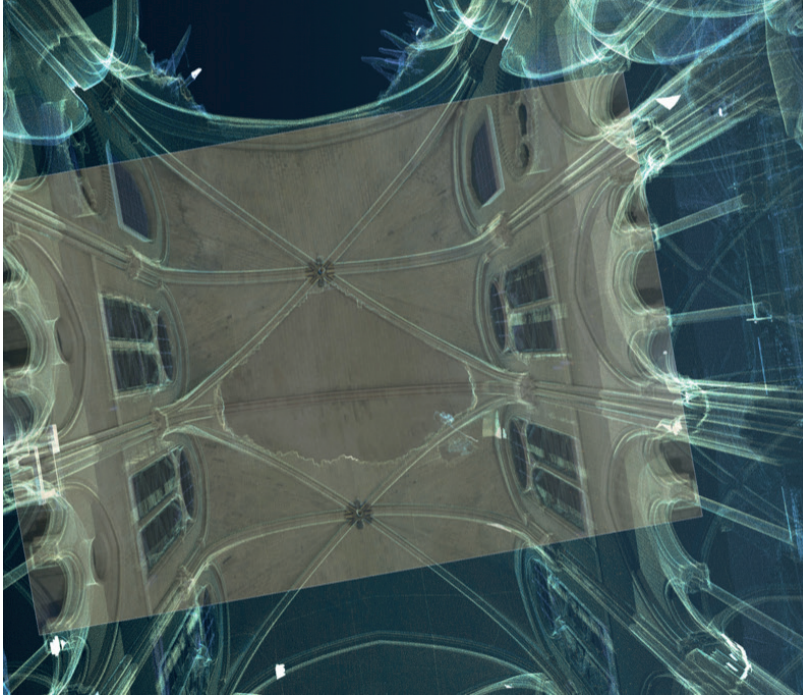
Dans le même temps, les institutions patrimoniales ont connu des évolutions majeures, à la fois par le recours croissant aux technologies numériques (plans de numérisation, doubles numériques, publication de ressources dans le Web sémantique, etc.) et par de nouveaux modes d'action et de gouvernance appuyés sur la recherche et associant une plus grande diversité d'acteurs sur un mode collaboratif. Ces évolutions sont productrices de grandes masses de données hétérogènes et aujourd'hui extrêmement dispersées pour lesquelles se pose d'une manière criante la question de leur gestion, de leur conservation à long terme et de leur partage avec une grande diversité d'acteurs. Si ces impératifs sont un défi pour de grandes institutions telles que le Louvre ou le château de Versailles dotées de certains moyens humains et budgétaires, on imagine aisément que des entités moins importantes, qui constituent la grande majorité des établissements patrimoniaux, ne sont pas en mesure de le relever. D'où la nécessité d'une approche conjointe et collaborative de niveau européen dont le futur *Cloud* serait la concrétisation.

La gestion des données du patrimoine

Les données numériques du patrimoine, produites pour la plupart dans un cadre public, se caractérisent par leur hétérogénéité et par la grande diversité de leurs producteurs. Elles ont une forte valeur culturelle,

1. Rapport des experts (*Report on a European collaborative Cloud for cultural heritage. Ex-ante impact assessment*). <https://op.europa.eu/en/publication-detail/-/publication/90f1ee85-ca88-11ec-b6f4-01aa75ed71a1/language-en>

2. http://www.sciences-patrimoine.org/2022/10/pfue-publication_actes/



Spatialisation 3D d'une photographie de l'arc-doubleau de la nef de Notre-Dame de Paris avant incendie, superposée au nuage de points de l'état de la cathédrale après incendie.

scientifique et, potentiellement, économique qu'il convient de maîtriser et de conserver dans la sphère publique, au bénéfice de tous³. Il en va de la souveraineté culturelle de l'Europe, des pays qui la constituent et de la capacité européenne à présenter une alternative sérieuse aux GAFAM.

En outre, ces données numériques jouent un rôle fondamental dans la structuration de la recherche sur le patrimoine, notamment par leur nature transversale par rapport aux champs disciplinaires et aux modes opératoires. Elles relèvent des multiples regards mobilisés pour l'étude des œuvres et des objets et leur accessibilité est une condition essentielle à la bonne compréhension, conservation-restauration et transmission de ces derniers.

Les appels à projets porteront sur la conception de la plate-forme et sa gouvernance, la mise en place de l'entité de pilotage et de coordination des actions de développement de nouveaux instruments collaboratifs, et l'expérimentation de nouveaux usages en grandeur réelle. Sur chacun de ces points, le ministère de la Culture (MC), le ministère de l'Enseignement supérieur et de la Recherche (MESR) et le CNRS peuvent identifier des acteurs du secteur (établissements culturels et laboratoires de recherche) susceptibles de répondre, en s'associant avec des partenaires européens, à ces appels de préfiguration qui dessineront les contours du futur *Cloud*. Ils pourront notamment s'appuyer sur les expertises développées au sein des projets tels qu'ESPADON⁴, E-RIHS⁵ et IPERION HS⁶ dont certains ont été évoqués dans de précédents numéros de cette revue.

Une avancée majeure au croisement de trois défis scientifiques et technologiques

La conception d'un *Cloud* collaboratif européen pour le patrimoine culturel se situe à l'intersection de trois grands défis qui visent à répondre aux exigences spécifiques des communautés du patrimoine culturel et à favoriser une coopération active des professionnels pour la coproduction d'une nouvelle génération de ressources numériques.

Le premier défi consiste à passer de la numérisation fragmentée et épisodique à des contreparties numériques « permanentes » d'objets patrimoniaux. Ce qui permettrait de disposer de données numériques

plus complètes et de mieux structurer les informations, notamment en intégrant progressivement les nouvelles acquisitions, analyses et enrichissements liés aux activités de recherche, de conservation et de diffusion.

Le deuxième défi est de passer d'une quantité de données produites au quotidien par les divers acteurs, aujourd'hui non ou peu mémorisées, à un *continuum* numérique capable d'assurer leur traçabilité et leur interopérabilité avec le double numérique de l'objet patrimonial. Cette approche permettrait une meilleure coordination des opérations menées par les différents producteurs de données.

Enfin, le troisième défi consiste à passer d'une production et gestion de données fragmentées et individuelles à un système sociotechnique sophistiqué prenant la forme d'une plateforme distribuée, adaptable et ouverte.

Relever ces trois défis serait une véritable révolution dans l'acquisition et la gestion des données du patrimoine culturel européen. Cela conduirait à la production de biens communs numériques, à la généralisation de bonnes pratiques et à la mise en place d'un outil mutualisé de partage de données qui offrirait aux grandes institutions comme aux plus modestes un accès privilégié à des services régulièrement optimisés et pérennes, facilitant les collaborations entre professionnels et chercheurs à un niveau européen. On imagine l'effet démultiplicateur d'un tel dispositif sur la recherche et l'innovation dans le domaine du patrimoine, notamment en lien avec les industries culturelles créatives. Il permettrait en outre la mise à disposition du public le plus large, par exemple *via* Europeana⁷, de certaines données ainsi produites et mises en forme.

S'il répond finalement aux objectifs fixés par le rapport des experts, on voit que le potentiel d'un tel projet peut être considérable, à commencer par la mise en cohérence d'initiatives européennes et nationales pour le moment assez éparpillées et leur articulation avec diverses initiatives nationales, en particulier, pour la France, le projet ESPADON, visant à la création d'un « objet patrimonial augmenté », ou le système d'information mis en place par le groupe de travail « données numériques » du chantier scientifique de Notre-Dame⁸ porté par le CNRS et le MC. ■

3. Gestion des données basée sur les principes FAIR (*Findable, Accessible, Interoperable, Reusable* pour « Facile à trouver, accessible, interopérable et réutilisable »).

4. Voir dans ce numéro l'article de Vincent Detalle et Romain Thomas, p. 68 : ESPADON (en sciences du patrimoine, l'analyse dynamique des objets anciens et numériques). <http://www.sciences-patrimoine.org/espaddon/>

5. E-RIHS (*European Research Infrastructure for Heritage Science* pour désigner l'infrastructure de recherche dédiée à l'étude des matériaux du patrimoine culturel et naturel). <https://www.e-rihs.eu/>

6. IPERION HS organise, notamment, des formations et offre un accès à un très large ensemble d'instruments scientifiques de haut niveau, de méthodes, de données, d'outils et d'expertise dans le but de faire avancer la connaissance et l'innovation dans le domaine des sciences du patrimoine.

7. Marie-Véronique Leroi, « Europeana, un espace de données du patrimoine culturel et une interface pour la recherche », *Culture et Recherche*, n°142, 2022, p. 51-52. Europeana est une plateforme numérique européenne lancée en novembre 2008 par la Commission européenne. Elle donne accès à des ressources (livres, matériel audiovisuel, photographies, documents d'archives, etc.) numériques des institutions culturelles de l'Union européenne. Europeana n'archive pas les œuvres, mais sert uniquement d'interface de recherche.

8. Livio De Luca et Violette Abergel, « Vers des biens communs numériques pour construire des cathédrales de données et de connaissances pluridisciplinaires sur le patrimoine européen », *Culture et Recherche*, n° 142, 2022, p. 104-106.

Ouvrir les données de la Bibliothèque nationale de France à la recherche

Métadonnées, documents numérisés, archives du Web ou *logs*¹ : appréhender ces données comme source et terrain pour la recherche pose des problématiques d'ordre à la fois méthodologique, technique et juridique. Quelle est la singularité des données de la Bibliothèque nationale de France (BnF)? Quelles en sont les modalités d'accès et de réutilisation? Quelles sont les limites à leur ouverture et à leur exploitation?

ARNAUD LABORDERIE

Chef de projet Gallica, chargé de l'exploitation des données pour la recherche, Bibliothèque nationale de France (BnF)

FLORENCE TFIBEL

Adjointe à la cheffe du service Diffusion des métadonnées, et responsable de l'équipe Services aux professionnels, BnF

Des données riches et très diverses

Les données sont au cœur des activités de la BnF. Ses agents contribuent à en produire quotidiennement : métadonnées bibliographiques décrivant les collections, documents numérisés ou collectés nativement, faisant l'objet de traitements permettant d'exploiter leur contenu (OCR², tables des matières...), données d'usage de la bibliothèque et de Gallica par le public, etc. Ces données, nécessaires au bon fonctionnement de l'établissement, sont également une source très riche pour les chercheurs, qui doivent néanmoins en comprendre toute la diversité pour pouvoir les exploiter.

Métadonnées des catalogues et données numérisées de Gallica

Les métadonnées bibliographiques de la BnF sont hétérogènes car elles proviennent de sources diverses, depuis les métadonnées fournies dans le cadre du Dépôt légal éditeurs jusqu'aux notices rédigées manuellement et enrichies par des experts, en passant par des données obtenues par calcul ou par chargement³, ce qui implique des niveaux de qualité différents. Elles sont également diverses dans leur structuration et leur format⁴, dont le choix est lié à la nature des collections décrites⁵ ainsi qu'à leurs usages. Les chercheurs peuvent ainsi accéder à différentes qualités et structurations de données selon leurs besoins.

Avec plus de 10 millions de documents, Gallica est l'une des plus grandes bibliothèques numériques au monde dotée d'une large diversité documentaire : livres, périodiques, images, manuscrits, cartes, objets, partitions, sons, vidéos... Les documents numérisés sont issus des collections patrimoniales de la BnF mais également de bibliothèques partenaires. Le processus de numérisation produit différents types de données : des données brutes (images au format TIFF ou JPG, sons et vidéos au format MPEG), des données dérivées (OCR⁶, tables des matières) et des métadonnées (METS⁷, Manifest⁸).

La collection numérique s'est constituée sur une trentaine d'années au cours desquelles les besoins ont évolué tandis que les technologies se sont considérablement améliorées (niveau de résolution, OCR, normalisation), d'où une certaine hétérogénéité de qualité des données. Si la quantité prévaut dans la logique de la numérisation de masse et reste nécessaire pour interroger largement la collection par mots-clés, la qualité des données est primordiale pour les chercheurs en

1. Dans un contexte informatique, un *log* est un « journal » qui désigne la documentation automatiquement générée et horodatée des événements concernant un système particulier. Pratiquement tous les systèmes et logiciels produisent des « fichiers journaux ».

2. OCR, pour *Optical Character Recognition*, est la reconnaissance optique de caractères, processus permettant de convertir une image de texte en format de texte lisible par une machine.

3. Opération permettant d'intégrer des données en masse dans les catalogues.

4. Interمارc pour le Catalogue général, XML-EAD pour le catalogue BnF Archives et manuscrits, Dublin Core pour la bibliothèque numérique Gallica, RDF pour le portail data.bnf.fr, WARC pour les archives du Web...

5. XML-EAD est un format particulièrement adapté à la description des fonds d'archives, là où le format Interمارc permet de décrire une gamme très large de types de documents dans une logique bibliographique.

6. L'OCR (*Optical Character Recognition*) est une technologie de reconnaissances de textes imprimés à partir d'images numérisées. Disponible dans Gallica depuis 2005, l'OCR constitue aujourd'hui une chaîne d'entrée interne à la BnF, qui est engagée notamment dans des programmes de rétroconversion.

7. METS (*Metadata Encoding and Transmission Standard*) est un standard permettant d'encoder les métadonnées descriptives, administratives et de structure, spécifiques aux objets numériques.

8. Manifest est un fichier XML décrivant la structure d'un document numérique ou le contenu d'un paquet, d'une application logicielle, etc.



Le voyageur contemplant la mer des données.

humanités numériques. Prise entre ces deux besoins, la BnF tâche de trouver un point d'équilibre pour fournir les données les plus qualitatives aux usagers.

Les archives du Web et les données d'usage

Les archives de l'Internet français sont un autre réservoir de données majeur. Ces collections, que la BnF collecte puis 2006 et dont les fonds remontent à 1996, représentent aujourd'hui 48 milliards d'URL⁹. Le Web présente un cas singulier de données pour la plupart librement accessibles en ligne mais qui, lorsqu'elles sont collectées par la BnF dans le cadre du dépôt légal, deviennent des données patrimoniales avec des restrictions d'accès.

Un autre type de données susceptibles d'intéresser les chercheurs, mais qui restent d'accès réservé, sont les logs de connexion aux serveurs de Gallica. Initialement conservées par la BnF uniquement à des fins de sécurité et d'évaluation de la qualité de service, ces données, qui enregistrent l'activité des utilisateurs, ont été anonymisées pour des raisons juridiques et éthiques afin d'être mises à la disposition des chercheurs.

Ouvrir les données : une question juridique et technique

Dès 2014, la BnF s'est engagée dans une démarche d'ouverture de ses données dont elle encourage les appropriations les plus variées¹⁰. Cette ouverture pose des questions à la fois juridiques et techniques autour de l'accès et des usages de recherche.

Une offre globale de diffusion des données : services et outils

La BnF a mis en place de la documentation et des outils permettant de rechercher, d'extraire et de réutiliser ses données afin de constituer des corpus numériques. Le Catalogue général permet un export des notices au format CSV¹¹, mais aussi dans des formats techniques plus utilisés dans le monde des bibliothèques (ISO27.09). Les usagers de Gallica peuvent récupérer les données au fil de leur consultation grâce aux fonctionnalités du visualiseur (téléchargement, partage, etc.) et des plugins associés (Zotero, IIIF)¹². Data.bnf.fr expose sur le Web, selon un modèle entité-relations, les données produites par l'établissement et propose un accès centré sur les œuvres, auteurs, thèmes, dates ou lieux ; un SPARQL endpoint¹³ permet d'interroger la base et de récupérer les données.

Ces données peuvent être exposées sous d'autres formes afin d'être récupérées en nombre et exploitées par des machines grâce à des API¹⁴. Il est ainsi possible de requêter et de récupérer les métadonnées du Catalogue général ou de Gallica par flux (SRU – Search and Retrieve via URL), ponctuellement (Z39.50¹⁵) ou sous forme de lots de données (OAI-PMH¹⁶). Le portail <https://api.bnf.fr>¹⁷ documente l'offre d'API et de jeux de données. Des extractions complètes des bases sont également disponibles depuis les plateformes de données ouvertes <https://www.data.gouv.fr/fr/> et <https://data.culture.gouv.fr>¹⁸.

9. Voir Vladimir Tybin, « Les collections du dépôt légal du Web de la BnF au cœur des réseaux de coopération internationale pour la recherche », *Culture et Recherche*, n° 143, 2022, p. 126-127.

10. <https://www.bnf.fr/fr/reutiliser-les-donnees-de-la-bnf>

11. CSV (Comma-separated values) désigne un format de fichiers dont le rôle est de présenter des données séparées par des virgules. Il s'agit d'une manière simplifiée d'afficher des données afin de les rendre transmissibles d'un programme à un autre.

12. Les plugins sont de petits programmes complémentaires qui ajoutent des fonctions aux applications Web et programmes de bureau : voir Sophie Bertrand, Gaël Chenard, Stéphane Pillorget, Carine Prunet (coordinatrice), Régis Robineau, « IIIF : découverte et interopérabilité sans frontières des images patrimoniales », *Culture et Recherche*, n° 143, 2022, p. 111-117.

13. Format de requête permettant d'interroger le langage de base du Web sémantique Resource Description Framework (RDF) : <https://api.bnf.fr/sparql-endpoint-de-databnfr>

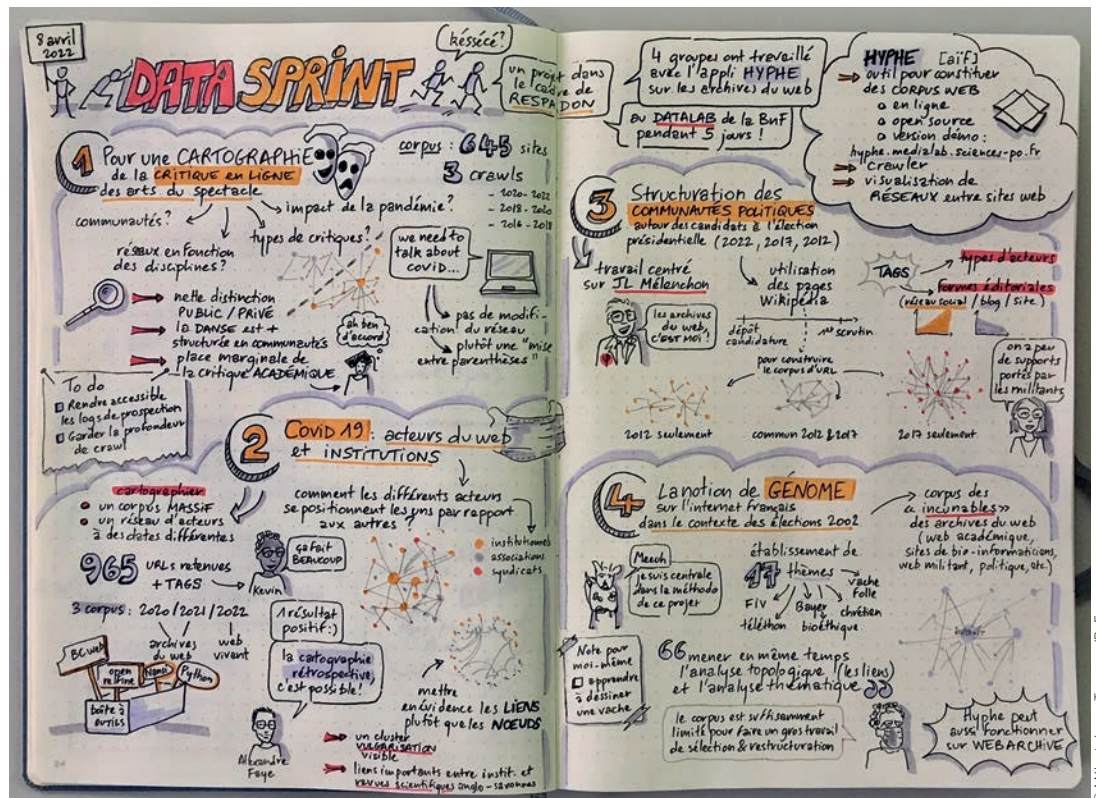
14. Les API (application programming interface) sont des interfaces de programmation applicatives permettant à des systèmes informatiques de communiquer entre eux et d'échanger des données de manière standardisée.

15. Protocole informatique utilisé par les bibliothèques pour interroger simultanément plusieurs catalogues.

16. Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) est un protocole informatique permettant d'échanger des métadonnées en interrogeant des entrepôts OAI, qui est largement répandu dans les institutions patrimoniales et notamment les bibliothèques.

17. <https://api.bnf.fr>

18. Voir dans ce numéro l'article de Roselyne Aliacar, p. 73.



Sketchnote du *Datasprint* sur les archives du Web, organisé au BnF DataLab du 4 au 8 avril 2022, en partenariat avec Sciences Po Paris.

Des métadonnées accessibles librement ou à certaines conditions

La grande majorité des données de la BnF sont accessibles et exploitables librement. Les quelque 15 millions de notices bibliographiques et 5 millions de notices d'autorité du Catalogue général sont sous licence libre Etalab, comme celles du Catalogue collectif de France (CCFr), donc réutilisables librement et gratuitement sous réserve d'en mentionner la source. La collection numérique de Gallica comprend des documents de toutes les époques, de l'Antiquité à nos jours, librement et gratuitement accessibles en ligne et exploitables pour ceux entrés dans le domaine public, seule leur utilisation commerciale étant soumise à redevance.

Des contraintes juridiques peuvent néanmoins peser sur les usages de certains ensembles de données. Rappelons que la BnF est soumise à deux cadres législatifs : le Code du patrimoine (CP) et le Code de la propriété intellectuelle (CPI), qui encadrent à la fois la collecte, l'accès et la réutilisation des données avec un certain nombre d'exceptions. Les documents sous droits de Gallica sont ainsi soumis au CPI et ne sont donc accessibles que dans l'enceinte de la BnF, comme les archives du Web¹⁹, constituées par dépôt légal (CP article L.131-132).

Enfin, pour accompagner les usages et encadrer les pratiques en matière de fouille de textes et de données (TDM²⁰), deux nouvelles exceptions ont été introduites dans le CPI²¹, dont une concerne les fouilles réalisées à des fins de recherche (CPI article L.122-5-3-II). Celles-ci sont désormais possibles sans accord préalable des ayants droit et sans que ceux-ci puissent s'y opposer, permettant également aux institutions

de recherche de conserver les corpus, sans pouvoir les diffuser, mais rendant possible la reproductibilité des traitements et la vérification des résultats. Cette grande avancée pose néanmoins des difficultés pour les chercheurs en termes de citabilité, notamment pour les archives du Web : si les permaliens permettent de citer les URL des archives, les chercheurs ne peuvent pas publier librement leurs sources au-delà du droit de courte citation (CPI article L.122-5).

Développer les usages et collaborer avec les chercheurs sur les données

Afin de faciliter et d'encourager l'usage de ses données pour la recherche, la BnF s'investit dans divers projets dont le but est de composer avec les contraintes techniques et juridiques existantes, et s'attache à accompagner au mieux les chercheurs. Accessibles sur demande sous forme de jeux de données anonymisées, les logs de connexion aux serveurs de Gallica, par exemple, ont pu faire l'objet de plusieurs travaux²², dont l'objectif n'était pas de connaître les usagers ni leurs profils mais, en partant des traces d'usages que sont les logs, de mieux connaître les sessions de consultation et d'identifier des parcours de recherche dans la bibliothèque numérique.

Depuis une dizaine d'années, Gallica collabore à des projets de recherche dans le domaine des humanités numériques. La coopération avec les chercheurs permet de co-construire des outils qui s'inscrivent dans l'écosystème de Gallica, comme Gallicagramm, outils et services qui, potentiellement, peuvent devenir des fonctionnalités de Gallica au bénéfice de tous les usagers, comme Gallica Images²³.

Dans le cas des archives du Web, le projet ResPaDon (2020-2023)²⁴ expérimente des accès distants sécurisés, permettant la consultation et l'exploitation des collectes appréhendées comme source primaire et données de recherche, dans un réseau de bibliothèques partenaires. Un prototype est actuellement testé à l'Université de Lille.

Ouvert en 2021, le BnF DataLab (voir le focus de Marie Carlin, p. 34) s'inscrit dans ce *continuum* d'outils, de services et de projets. Ceux issus de l'appel à projets 2021-2022²⁵ sont exemplaires des problématiques qui se posent aux chercheurs travaillant sur les données de la BnF : correction de l'OCR, extraction des entités nommées et modélisation des sujets ; indexation et création automatisées de corpus numériques sur des notions relativement récentes, comme l'environnement, grâce à des techniques de traitement automatique du langage (TAL) ; transcription automatique de l'écriture manuscrite (HTR²⁶). À l'issue de ces projets, les équipes de recherche produisent des données enrichies (corrections, annotations, modèles, etc.). La science ouverte prescrit de partager

ces données et de les rendre accessibles selon les principes FAIR²⁷. Pour la BnF, réintroduire ces enrichissements dans ses propres données est un défi.

Conclusion

Qu'elles soient issues de la numérisation ou nativement numériques, toutes les données proposées par la BnF s'inscrivent dans le *continuum* de ses collections. La BnF mobilise différents niveaux d'expertise pour répondre aux besoins des chercheurs et résoudre aux mieux les difficultés scientifiques et techniques qu'ils rencontrent. Concernant le cadre juridique, l'exception en matière de fouille de textes et de données (*Text Data Mining-TDM*) est une avancée majeure dans la prise en compte des usages de recherche. Dans le cadre de la rédaction du décret d'application de la loi Darcos du 30 décembre 2021 qui instaure le dépôt légal des documents numériques, le ministère de la Culture porte une attention particulière aux modalités d'accès aux données, avec le souci de l'équilibre des territoires, ouvrant sans doute de nouvelles perspectives aux chercheurs. ■

24. ResPaDon (Réseau de partenaires pour l'analyse et l'exploration de données numériques) est financé par le GIS Collex-Persée, porté par l'Université de Lille et la BnF, en partenariat avec Sciences Po Paris et le Campus Condorcet.
<https://respadon.hypotheses.org/>

25. En savoir plus : <https://www.bnf.fr/fr/les-projets-de-recherche-bnf-datalab>

26. L'HTR (*Handwritten Text Recognition*) est une technologie de reconnaissance de l'écriture manuscrite à partir d'images numérisées. Voir dans ce numéro un exemple avec l'article d'Edouard Bouy, p. 110.

27. « Facile à trouver, accessible, interopérable et réutilisable ».

Schéma numérique de la Bibliothèque nationale de France

« C'est ce bel écho entre savoir et légèreté qui fait la force de la BnF et qui permet de rendre le savoir accessible à tous ! »
Laurence Rogot, Présidente de la BnF, 27 avril 2020.

Cette cartographie originale est le fruit d'une réflexion collective sur l'appropriation de la Bibliothèque nationale de France. Elle s'inscrit dans la continuité des précédents schémas numériques de l'établissement : comme eux, elle permet d'observer les axes stratégiques, les reliefs, les distances et les parcours au sein de l'univers numérique de la Bibliothèque et de son écosystème. Pour mieux favoriser l'appropriation d'un discours commun sur ce sujet complexe, pour mieux transmettre sa vision d'avenir, elle fait aujourd'hui le pari de la métaphore comme véhicule d'humanisme et de poésie : le présent support, qui s'est nourri des riches collections de cartes et plans de la Bibliothèque et viendra s'enrichir à son tour, est en effet une occasion qui vous est donnée, à vous agents de la BnF, à vous professionnels extérieurs, partenaires, prestataires, à vous, curieux de tous horizons, de vivre une expérience et d'explorer véritablement ces territoires.

Le « monde numérique » qui se dévoile en cet automne 2020 paraît durant une période – la crise sanitaire – qui, à bien des égards, a bouleversé notre rapport au numérique, rendu plus évidents les enjeux de la médiation virtuelle, de la personnalisation du rapport aux publics, ou du travail à distance, et donné un visage plus précis que jamais à la fracture numérique.

Le schéma numérique prend acte de ces défis nouvellement formulés, mais suggère également à quel point un recul critique permanent, transversal, endurant, le plus à même de nourrir une stratégie et une vision, confère à une institution la capacité à faire face aux situations les plus inhabituelles ou inattendues. Ainsi, l'environnement numérique de travail du personnel comme des publics apparaît bien avant mars 2020 comme le terrain privilégié des mutations à engager. Cette carte s'assoit sur des fondations solides – ses régions, ses réalités géographiques, qui sont autant d'échos aux missions de la Bibliothèque – mais, parce que la réflexion sur le numérique est amenée à se poursuivre, parce que les conclusions de la crise sanitaire continuent de se préciser, elle a aussi été conçue pour évoluer au fil d'éditions à venir : si vous regardez bien, vous y verrez d'ailleurs ses géographes à l'œuvre.

(BnF) Bibliothèque nationale de France
NOVEMBRE 2020
<https://www.bnf.fr/fr/la-bnf/schemas-numeriques>



Légende

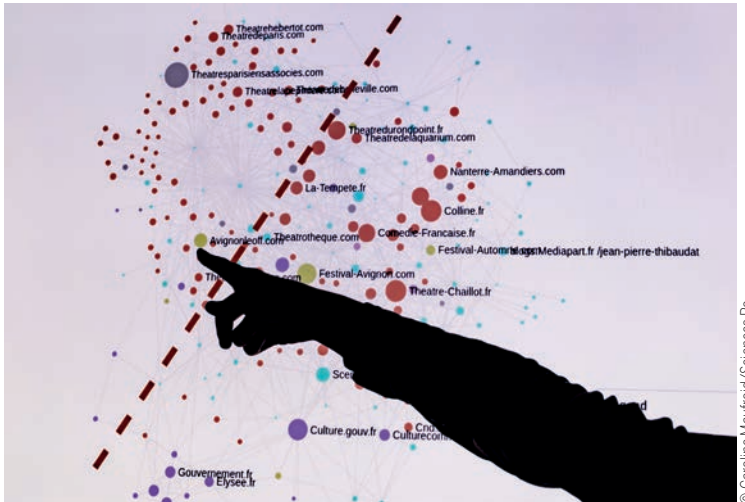
Données	Cartes documentaires	Objets de la connaissance	Lib des données	Manus	Stratégie	Voies	Champs	Outils	Événements
• Collections numériques • Collections numériques à vocation de conservation • Collections numériques à vocation de médiation • Collections numériques à vocation de recherche • Collections numériques à vocation de diffusion	• Objets de la connaissance • Objets de la connaissance à vocation de conservation • Objets de la connaissance à vocation de médiation • Objets de la connaissance à vocation de recherche • Objets de la connaissance à vocation de diffusion	• Objets de la connaissance • Objets de la connaissance à vocation de conservation • Objets de la connaissance à vocation de médiation • Objets de la connaissance à vocation de recherche • Objets de la connaissance à vocation de diffusion	• Objets de la connaissance • Objets de la connaissance à vocation de conservation • Objets de la connaissance à vocation de médiation • Objets de la connaissance à vocation de recherche • Objets de la connaissance à vocation de diffusion	• Objets de la connaissance • Objets de la connaissance à vocation de conservation • Objets de la connaissance à vocation de médiation • Objets de la connaissance à vocation de recherche • Objets de la connaissance à vocation de diffusion	• Objets de la connaissance • Objets de la connaissance à vocation de conservation • Objets de la connaissance à vocation de médiation • Objets de la connaissance à vocation de recherche • Objets de la connaissance à vocation de diffusion	• Objets de la connaissance • Objets de la connaissance à vocation de conservation • Objets de la connaissance à vocation de médiation • Objets de la connaissance à vocation de recherche • Objets de la connaissance à vocation de diffusion	• Objets de la connaissance • Objets de la connaissance à vocation de conservation • Objets de la connaissance à vocation de médiation • Objets de la connaissance à vocation de recherche • Objets de la connaissance à vocation de diffusion	• Objets de la connaissance • Objets de la connaissance à vocation de conservation • Objets de la connaissance à vocation de médiation • Objets de la connaissance à vocation de recherche • Objets de la connaissance à vocation de diffusion	• Objets de la connaissance • Objets de la connaissance à vocation de conservation • Objets de la connaissance à vocation de médiation • Objets de la connaissance à vocation de recherche • Objets de la connaissance à vocation de diffusion

LE BNF DATALAB :

LE LABORATOIRE DES HUMANITÉS NUMÉRIQUES DE LA BIBLIOTHÈQUE NATIONALE DE FRANCE



Datasprint ResPaDon dans les espaces du BnF DataLab, du 4 au 8 avril 2022.



MARIE CARLIN

Conservatrice des bibliothèques,
coordinatrice du DataLab de la BnF

1. Voir dans ce numéro l'article
d'Olivier Baude et Nicolas Larrousse,
p. 11.

Ouvert en octobre 2021 pour accueillir les chercheurs travaillant sur les collections numériques de la BnF, le BnF DataLab a été conçu pour favoriser l'accès et l'exploitation des données de la BnF. Plus qu'un service de fourniture de données brut, le BnF DataLab est une coordination de

services et d'expertises qui accompagnent les projets de recherche depuis la constitution des corpus jusqu'à la valorisation des résultats de recherche.

Pour rester au plus près de la recherche en train de se faire, le BnF DataLab s'est inscrit dès son origine au sein d'un réseau de partenaires de grands acteurs du monde de la recherche, tels que le CNRS, via l'infrastructure¹ IR* Huma-Num, ou l'équipe-projet dédiée aux humanités numériques, l'Observatoire des textes, des idées et des corpus » (ObTIC) de Sorbonne-Université.

Lieu de conversation scientifique et de co-construction des savoirs entre chercheurs et professionnels des bibliothèques, le BnF DataLab poursuit également l'ambition d'être

le laboratoire des outils de la bibliothèque de demain. En proposant un lieu d'expérimentation de nouvelles technologies et d'amélioration des outils, en valorisant la diffusion des jeux de données de façon transversale et transnationale, le BnF DataLab inscrit encore plus durablement la BnF dans la dynamique des *Collections as Data*. ■



Signalétique du BnF DataLab en salle X des espaces recherche de la BnF (site de Tolbiac).

Pour tout savoir :

<https://www.bnf.fr/fr/bnf-datalab>

Pour soumettre un projet,
avoir plus d'informations :

<https://www.bnf.fr/fr/postulez-au-bnf-datalab>

Reconstituer l'histoire de la France « par le bas » : un siècle de recensements de population décryptés par le projet Socface

Socface est un projet de recherche, soutenu par l'Agence nationale de la recherche (ANR), sur les recensements de la population française de 1836 à 1936. Il mobilise des chercheurs en sciences humaines et sociales, des ingénieurs et des archivistes, et illustre de nombreux aspects de la science ouverte ainsi que les apports et défis de la reconnaissance automatique d'écriture manuscrite¹.

Le projet Socface² a pour ambition de transcrire automatiquement l'ensemble des listes nominatives des recensements de 1836 à 1936 (soit vingt recensements) pour produire, étudier et diffuser une base de données des individus ayant vécu en France durant cette période. Soutenu par l'Agence nationale de la recherche (ANR)³, ce projet illustre de nombreux aspects de la science ouverte ainsi que les apports et défis de la reconnaissance automatique d'écriture manuscrite.

Il met aussi en évidence l'appétit sans cesse croissant des différents utilisateurs des archives pour les données nominatives : aujourd'hui, l'écrasante majorité des recherches faites dans les services d'archives porte sur ce type de sources. Chaque personne a vocation à être représentée aux archives, dès lors que sa vie a connu quelques événements, heureux ou, le plus souvent, malheureux.

Socface mérite une place à part en raison de son ampleur : il porte en effet sur un corpus très vaste – une même typologie, traitée sur 100 ans, conservée dans près de 100 structures de métropole et d'outre-mer.

Aux origines du projet

L'intérêt croissant pour les données individuelles, en particulier nominatives, est alimenté par les développements techniques (facilité de numérisation, diffusion des images sur le Web, améliorations des techniques de reconnaissance automatique d'écritures, etc.) tout autant qu'il les nourrit : la demande des usagers (chercheurs, généalogistes ou amateurs éclairés) motive les campagnes de numérisation tout comme l'appétence de la recherche quantitative en

sciences sociales pour des données « micro » stimule le développement de la reconnaissance automatique d'écriture manuscrite.

Socface illustre parfaitement ce cercle vertueux autour d'une source unique (les recensements) qui fait partie des rares typologies de documents à avoir été presque intégralement numérisées par les services d'archives, créant un corpus qui devrait dépasser à terme les 10 millions d'images malgré les destructions, volontaires ou accidentelles. Cette numérisation quasi exhaustive était une condition préalable pour qu'un tel projet de recherche puisse être réalisé.

Cette condition remplie, la gourmandise des historiens pour cette masse de données ne suffisait pas ; encore fallait-il imaginer un système efficace pour extraire le texte contenu dans ces millions d'images. Les progrès considérables de la reconnaissance automatique de l'écriture manuscrite ces dernières années, grâce aux avancées des technologies de l'intelligence artificielle, permettent d'envisager cette extraction. Les documents historiques manuscrits, du Moyen Âge à nos jours, sont désormais à la portée d'une transcription automatique permettant une exploitation directe. Cette reconnaissance automatique prend tout son sens pour des traitements à très large échelle pour lesquels une transcription manuelle, même collaborative, n'est pas envisageable.

Le rôle du collaboratif dans la reconnaissance d'écriture

Pour autant, la reconnaissance d'écriture n'est pas un vase clos, entièrement autonome. En effet, le développement d'un système de reconnaissance d'écriture

CHRISTOPHER KERMORVANT

Président de la société Teklia

LIONEL KESZTENBAUM

Directeur de recherche à l'Institut national d'études démographiques (INED)

MANONMANI RESTIF

Cheffe de projet du portail FranceArchives, Service interministériel des Archives de France (SIAF)

1. En anglais : *Handwritten Text Recognition* (HTR).

2. <http://www.socface.org/>

3. <https://anr.fr/Projet-ANR-21-CE38-0013>

© Archives municipales de Rennes, 1F 90J

Recensement du canton sud-est de Rennes : ville et banlieue, 1896.

performant nécessite une phase d'entraînement des modèles sur des données annotées, par des techniques d'apprentissage automatique supervisé. Les modèles les plus récents, fondés sur des technologies de *Deep Learning*⁴, peuvent être entraînés avec un protocole beaucoup plus simple que leurs prédécesseurs. Aujourd'hui, il n'est plus nécessaire de transcrire précisément les documents, en indiquant la position et le contenu des lignes de texte. Il est possible d'entraîner les modèles à partir de données saisies dans un formulaire, comme on le ferait pour un dépouillement d'archives. Ce protocole, beaucoup plus rapide et naturel, permet de faire appel à des volontaires pour réaliser les annotations.

Le projet Socface a ainsi ouvert une dizaine de campagnes d'annotations collaboratives pour créer des données d'entraînement en utilisant la plateforme Callico de Teklia⁵.

Par ailleurs, les annotations déjà existantes, réalisées par les cercles généalogiques ou dans les services d'archives départementales, peuvent aussi être utilisées pour entraîner la machine. De fait, la qualité de la reconnaissance est améliorée par tout un ensemble d'informations extérieures : de la liste des noms de famille (et de leur fréquence) jusqu'au nom des lieux-dits de chaque commune, en passant par une estimation grossière des distributions par âge au

cours du temps, tout ce qui peut donner à la machine une idée, même vague, de « l'univers des possibles » est précieux.

En ce sens, Socface est très directement un produit de la science ouverte.

Traiter, analyser et diffuser des millions d'images

Le siècle d'histoire française auquel s'intéresse Socface est marqué par des changements spectaculaires souvent résumés par quelques concepts généraux esquissés à grands traits : urbanisation, industrialisation, transition démographique. Pourtant, on connaît encore relativement mal la variation spatiale de ces phénomènes sur le territoire métropolitain, leurs mécanismes et leurs conséquences. L'apport de Socface, en particulier en appariant les individus entre les recensements pour reconstituer leurs trajectoires (migratoires, professionnelles, familiales), est de permettre d'étudier cette hétérogénéité, de saisir comment ces trajectoires rencontrent, ou pas, la « Grande Histoire », comment elles sont influencées par elle et l'influencent en retour.

Un second produit direct du projet sera de diffuser librement ces données pour permettre à tout un chacun d'y accéder. Pour les archives, cette mise à disposition d'un grand volume de données, tant dans

4. Connue en français sous le nom « apprentissage profond », le *Deep Learning* est une branche de l'intelligence artificielle. L'apprentissage profond est un procédé d'apprentissage automatique utilisant de très grands réseaux de neurones, possédant des centaines de couches et des milliards de paramètres (neurones). Cette technique a permis des progrès importants et rapides dans les domaines de l'analyse du signal sonore ou visuel et notamment de la reconnaissance faciale, de la reconnaissance vocale, de la vision par ordinateur et du traitement automatisé du langage.

5. <https://callico.teklia.com>

nom	Chévenet	prénom	Charles	date_naissance	1873	lieu_naissance	Montrigny Corrèze	nationalite	F	profession	chef	statut	Néant
nom	Théveut	prénom	Marguerite	date_naissance	1875	lieu_naissance	Noulener P. Loire	nationalite	F	profession	épouse	statut	Néant
nom	Legouet	prénom	Marie Antré	date_naissance	1866	lieu_naissance	Laoure de G	nationalite	F	profession	chef	statut	Néant
nom	Jolivet	prénom	Marie	date_naissance	1910	lieu_naissance	Venas	nationalite	F	profession	chef	statut	couturière
nom	Jolivet	prénom	Simon	date_naissance	1908	lieu_naissance	Venas	nationalite	F	profession	frère	statut	Mécanicien M Ravet
nom	V. Jaulines	prénom	Marie	date_naissance	1852	lieu_naissance	Marigoy	nationalite	F	profession	grand mère	statut	Néant
nom	Lamarque	prénom	Virginie	date_naissance	1870	lieu_naissance	Douvz ord	nationalite	F	profession	chef	statut	Néant
nom	Sabathie	prénom	Joseph E desnoud	date_naissance	1863	lieu_naissance	Toulonne	nationalite	F	profession	chef	statut	Receveur Enregistrement
nom	Sabalhier	prénom	Marguerite	date_naissance	1863	lieu_naissance	Gremoble	nationalite	F	profession	épouse	statut	Néant
nom	Ballaire	prénom	Albert	date_naissance	1884	lieu_naissance	Vemennes	nationalite	F	profession	chef	statut	profice de Pait
nom	Ballaire	prénom	Adélaïde	date_naissance	1876	lieu_naissance	Maillet	nationalite	F	profession	épouse	statut	Néant
nom	Rouy	prénom	Louis	date_naissance	1895	lieu_naissance	Bionde	nationalite	F	profession	chef	statut	Platrice peintre patron
nom	Rouy	prénom	Emilienne	date_naissance	1896	lieu_naissance	Hérison	nationalite	F	profession	épouse	statut	Néant
nom	Roux	prénom	Jean Robert	date_naissance	1920	lieu_naissance	Herisses	nationalite	F	profession	frères	statut	apprenti plâtre employeur M Noot
nom	Lepée	prénom	Louis	date_naissance	1893	lieu_naissance	Bizenemille	nationalite	F	profession	chef	statut	Agriculteur patron
nom	Pepie	prénom	Marie	date_naissance	1903	lieu_naissance	Gaumond	nationalite	F	profession	épouse	statut	Néant
nom	Pepce	prénom	Jean	date_naissance	1925	lieu_naissance	Heinson	nationalite	F	profession	frères	statut	Néant
nom	Lepée	prénom	Edmond	date_naissance	1928	lieu_naissance	Heinson	nationalite	F	profession	frères	statut	Néant
nom	Guillomet	prénom	Henri	date_naissance	1910	lieu_naissance	Herison	nationalite	F	profession	chef	statut	Camier patron
nom	Guillomet	prénom	Marguerite	date_naissance	1913	lieu_naissance	Chaubeux	nationalite	F	profession	épouse	statut	Néant
nom	Guillomet	prénom	Philippe	date_naissance	1882	lieu_naissance	Estesancies	nationalite	F	profession	père	statut	Entrepreneur frances froblies
nom	V. Veunat	prénom	Adélaïde	date_naissance	1882	lieu_naissance	Venisa P	nationalite	F	profession	chef	statut	Néant
nom	Riotte	prénom	Jean	date_naissance	1911	lieu_naissance	Heinson	nationalite	F	profession	frères	statut	Neveu Courreur employeur Min Naves
nom	Ve Martinet	prénom	Marie	date_naissance	1860	lieu_naissance	Venise	nationalite	F	profession	chef	statut	ménagère Diven
nom	V. Lachasagne	prénom	Clarice	date_naissance	1870	lieu_naissance	Heinson	nationalite	F	profession	chef	statut	Néant
nom	Lachanagne	prénom	Georges	date_naissance	1892	lieu_naissance	Veus	nationalite	F	profession	frères	statut	Electricien employeur S T M C
nom	V. Périnet	prénom	Hélène	date_naissance	1862	lieu_naissance	Charbens	nationalite	F	profession	chef	statut	Néant
nom	Bergerat	prénom	Jean	date_naissance	1864	lieu_naissance	Heinson	nationalite	F	profession	chef	statut	Jardinier employeur Dr Benon
nom	Bergerat	prénom	Jean	date_naissance	1902	lieu_naissance	Heinson	nationalite	F	profession	frères	statut	Carnis employeur Guillounet
nom	V. Malochet	prénom	Amilie	date_naissance	1865	lieu_naissance	Maillet	nationalite	F	profession	chef	statut	Néant

Transcription automatique d'une page de recensement.

la base de noms de FranceArchives⁶ que sur les sites Web des services d'archives, représente une formidable opportunité de développer de nouveaux services pour leurs publics attachés à la micro-histoire individuelle. Elle ouvre aussi des perspectives de mutualisation du réseau des archives pour augmenter le stock des métadonnées archivistiques interoperables.

À terme, Socface représentera un prodigieux effet levier. D'un côté, il poussera, inévitablement, à la numérisation des recensements manquants, voire à

leur identification. De l'autre, il pourra constituer un socle sur lequel mettre en œuvre d'autres dépouillements de sources à grande échelle. Plus largement, il devrait favoriser la concertation entre les archivistes et le monde de la recherche, les premiers pouvant réinterroger leurs politiques de numérisation, par exemple en développant une dimension nationale autour de typologies d'envergure, tandis que le second devra être plus attentif à reverser aux services d'archives les données qu'il produit. ■

6. <https://francearchives.gouv.fr/fr/basedenoms>

Annotation d'une page de recensement à l'aide de Callico.

Le cercle vertueux des ressources documentaires du Centre de recherche du château de Versailles : l'exemple du projet VERSPERA

Dès sa création en 2006, sous la houlette du ministère de la Culture, le Groupement d'intérêt public (GIP) – Centre de recherche du château de Versailles (CRCV) a eu la volonté de mettre à la disposition sans entraves le fruit de ses dépouillements et de ses diverses recherches, se plaçant ainsi dans la lignée du mouvement de la science ouverte. D'abord sous format expérimental, le CRCV a mis en ligne, par différents moyens (site Internet, revue en ligne, bases de données, etc.), des ressources numériques en tendant à suivre les principes qui régissent les données FAIR (« Facile à trouver, accessible, interopérable et réutilisable »)¹.

MATHIEU DA VINHA

Directeur scientifique du Centre de recherche du château de Versailles

BENJAMIN RINGOT

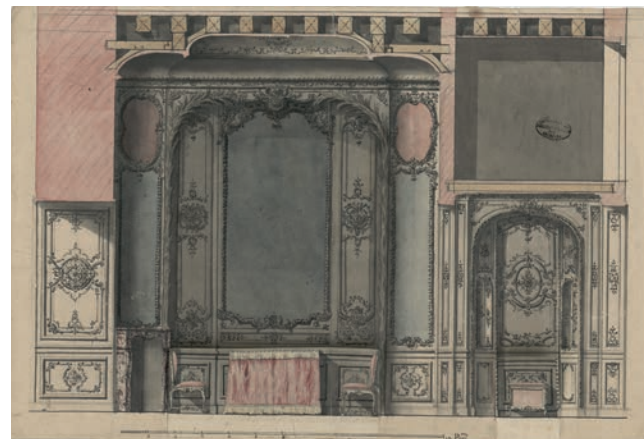
Adjoint au directeur scientifique du Centre de recherche du château de Versailles

VERSPERA : une libre circulation des sources

Le projet de recherche VERSPERA², lancé en 2013 et piloté par le CRCV, avec le soutien financier de la Fondation des sciences du patrimoine et du ministère de la Culture, vise à préserver et à rendre publics, *via* Internet, les plans du domaine de Versailles sous l'Ancien Régime, conservés aux Archives nationales et à la Bibliothèque nationale de France, puis à en restituer certains espaces grâce à la modélisation 3D élaborée par le laboratoire ETIS (Équipes traitement de l'information et systèmes, UMR 8051, Cergy Paris Université/ENSEA Cergy/CNRS)³.

La numérisation de près de 9000 documents graphiques a créé un corpus digital unique mis en ligne sur la banque d'images du CRCV, le tout en haute définition et augmenté de nombreuses métadonnées. Il est dorénavant largement diffusé aussi bien au sein de la communauté scientifique qu'auprès du grand public et permet à tous de pouvoir le consulter librement et de le réutiliser pour appréhender, voire s'approprier, l'histoire du château de Versailles.

Créant un véritable écosystème, les données de VERSPERA alimentent aujourd'hui les cinq pôles du CRCV. En effet, l'existence de ces numérisations a grandement facilité leur utilisation comme support, notamment aux programmes de recherche, aux ressources documentaires, mais aussi aux enseignements et formations, dans lesquels le CRCV est



Coupe et élévation du cabinet de chaise du roi et du cabinet voisin, au premier étage [1732].

Archives nationales (Paris), O/1/1770, dossier 1, n° 29. Cliché Archives nationales (France) – Avec le soutien de la Fondation des sciences du patrimoine/LabEx Patrima.

particulièrement engagé avec ses différents partenaires, aussi bien à des fins de recherche que de médiation.

Un apport indéniable pour la formation

Au sein du CRCV, plusieurs types de formations sont dispensés à l'attention des élèves de ses partenaires de l'enseignement supérieur. Les numérisations des plans sont dorénavant mises à leur disposition

1. Site internet institutionnel du CRCV : <https://chateauversailles-recherche.fr/> ; portail de ressources du CRCV : <http://www.chateauversailles-recherche-ressources.fr/jlbweb/jlbWeb?html=accueil>

2. <https://verspera.hypotheses.org/>

3. Sur le détail des différentes facettes de ce partenariat, voir : Michel Jordan et Benjamin Ringot, « Le projet VERSPERA. Numérisation, recherche et modélisation 3D des plans d'Ancien Régime de Versailles » (<https://eac.ac/articles/3432>), dans Stéphane Patin (dir.), *Les enjeux du numérique en sciences sociales et humaines. Vers un homo numericus ?*, Éditions des archives contemporaines, 2020.



Vue de la restitution en 3D d'un projet pour la petite galerie de Mignard. CC-BY Licence professionnelle Métiers du numérique, Cergy Paris Université.

leur facilitant grandement la réalisation d'études spécifiques en confrontant simultanément plusieurs sources iconographiques, chose difficilement réalisable avec les documents originaux qui vont de quelques centimètres à plusieurs mètres. Plusieurs mémoires d'étude de première année de second cycle de l'École du Louvre ont ainsi pu être rédigés à partir de VERSPERA, tel celui sur *Les salles des bains du château de Versailles de Louis XIV à Louis XVI*⁴.

Par ailleurs, chaque année, le projet VERSPERA, dans son volet modélisation, est couplé à la licence professionnelle « Métiers du numérique » de CY Cergy Paris Université. Les plans, coupes et élévations permettent la réalisation du modèle 3D puis l'ajout de texture. Cette confrontation de la 2D des plans à la 3D du modèle informatique génère des discussions, des interrogations et parfois des remises en question qui nourrissent la formation des étudiants mais font aussi évoluer les connaissances sur Versailles. Les échanges entre les étudiants et les chercheurs du CRCV permettent ainsi d'arriver à une proposition d'« habillage » destiné à la médiation du projet auprès du grand public.

Une médiation numérique facilitée

Dans le cadre des travaux menés par le CRCV, le corpus VERSPERA a rendu possible un projet de recherche appliquée, pensé comme un outil de médiation culturelle et scientifique : Immersailles⁵.

Ce projet de science ouverte a réuni le CRCV, l'Université Gustave Eiffel (Département informatique de l'institut universitaire de technologie de Marne-la-Vallée) et l'Université Rennes 2 dans l'idée de rendre librement accessible sur Internet la localisation – sur des plans d'époque – des appartements des occupants du château de Versailles sous l'Ancien Régime.

Dès son lancement, Immersailles a été conçu et développé par son créateur, Thomas Fressin, dans

une logique évolutive de science ouverte et avec l'intégration de corpus issus de données ouvertes. En effet, chaque individu y fait l'objet d'une fiche biographique dont les informations proviennent de plusieurs bases de données interopérables : la base biographique du CRCV, le référentiel Wikidata⁶ ou encore le fichier Bossu de la Bibliothèque nationale de France (BnF)⁷, créant ainsi une nouvelle fiche composite de référence librement accessible à toutes et tous.

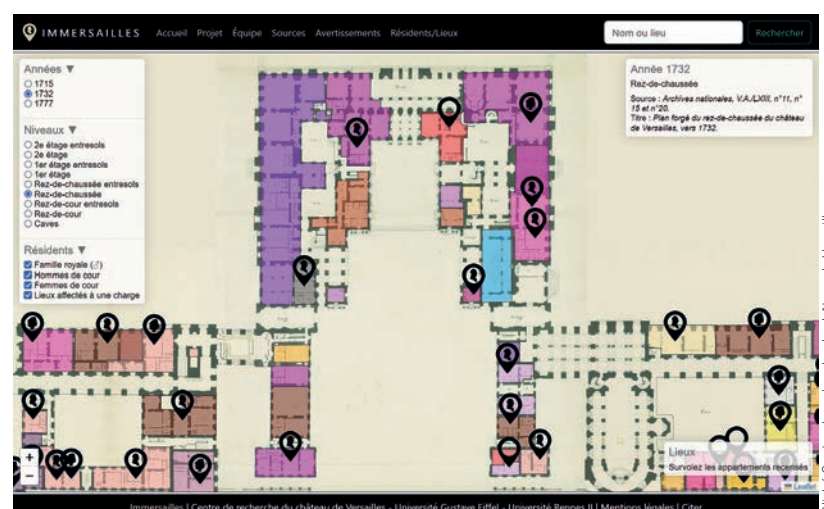
À travers ces quelques exemples, on s'aperçoit que la multiplication des outils développés et l'ensemble des données mises à disposition par le CRCV, tel un véritable écosystème, ont créé un cercle vertueux qui, finalement, dépasse la simple volonté de donner à voir ces différents matériaux et dans lequel outils et données se répondent et alimentent ses différentes activités. ■

4. Claire Boulestin, *Les salles des bains du château de Versailles de Louis XIV à Louis XVI*, Paris, mémoire de l'École du Louvre, 2018.

5. <https://www.chateauversailles-recherche.fr/immersailles/>

6. Wikidata est une base de connaissances libre et gratuite qui peut être lue et modifiée tant par des personnes que par des dispositifs informatisés.

7. Le fichier Bossu, entré par legs en 1987 et composé de 130 000 fiches biographiques rédigées par l'historien Jean Bossu, est un fichier nominatif des maçons français réputé exhaustif, en l'état des sources existantes, jusqu'en 1850.



Cliché Centre de recherche du château de Versailles

Site Immersailles.

HALUR: des outils connectés à HAL au service des chercheurs

HALUR¹ est une suite d'outils créée par deux ingénieurs, Laurent Jonchère (Université de Rennes) et Olivier Troccaz (UMR CNRS 6566), centrée sur l'accès ouvert aux publications des chercheurs. Ces six applications Web exploitent les possibilités de l'interface de programmation (API) de l'archive ouverte nationale HAL², où sont référencées les publications scientifiques de l'Université de Rennes, dont une majorité sont disponibles en texte intégral.

LAURENT JONCHÈRE

Ingénieur d'études au Service commun de la documentation – Université de Rennes et administrateur du portail HAL-Rennes

OLIVIER TROCCAZ

Ingénieur de recherche UMR CNRS 6566 Centre de recherche en archéologie, archéosciences, histoire (CReAAH)

1. <https://halur1.univ-rennes1.fr/index.php>

2. <https://hal.science/>

3. <https://appui-evaluation.hceres.fr/hal> : outil dérivé d'une application similaire de l'Institut national de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE), elle-même directement inspirée d'ExtrHAL.

La première brique HALUR a été créée en 2013 sous le nom d'ExtrHAL. Il s'agissait de répondre à un besoin urgent des unités de recherche d'exporter la bibliographie complète de leurs productions pour l'évaluation du Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur (Hcéres). Nous avons adapté et enrichi un script du chercheur Philippe Gambette (Université Gustave Eiffel), croisé à Lyon lors d'une formation HAL.

Le principe sera repris sur tous les outils ultérieurs de HALUR : l'utilisateur saisit le code collection HAL de son laboratoire, correspondant à l'ensemble des publications de cette équipe, puis renseigne quelques paramètres (dates, type de document, etc.) avant de valider sa requête. L'outil, que nous avons souhaité partager avec la communauté universitaire, a rencontré un vif succès et a été adopté par un nombre important d'établissements et de laboratoires.



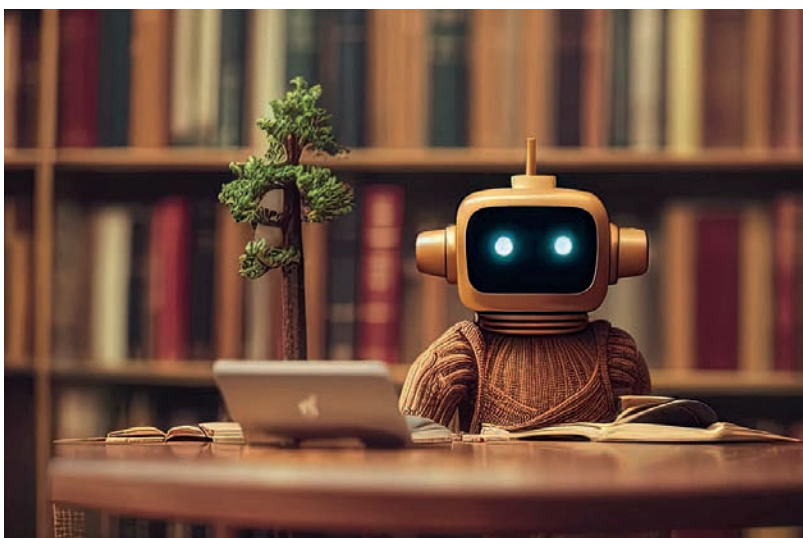
<https://unsplash.com> – Photo de Scott Graham

Nous avons été depuis lors très sollicités pour effectuer des mises à jour et apporter un support technique, mais aussi pour animer des formations. Dix ans plus tard, le Hcéres s'est inspiré d'ExtrHAL pour créer sa propre application d'export³, en s'appuyant sur notre expertise.

Référencer la bibliographie des laboratoires dans HAL

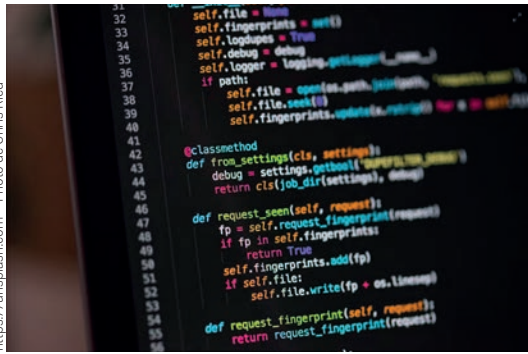
Encore faut-il, pour exporter la bibliographie d'un laboratoire ou d'une institution, que ses publications aient bien été toutes référencées dans HAL. C'était une gageure, car il n'était évidemment pas question de le faire article par article, et nous ne pouvions pas non plus compter sur une participation active de tous les chercheurs. Nous avons ainsi créé les deux briques d'une chaîne de dépôt massif, **OverHAL** et **Zip2HAL**, la première pour convertir en un format lisible par les machines (XML) des listes de publications issues de bases de données bibliographiques internationales (Web of Science, Pubmed), la seconde pour déposer en masse ces publications dans HAL *via* des webservice, après vérification des métadonnées.

Ces deux outils sont principalement destinés à un administrateur HAL de l'université qui se charge



<https://pixeid.com> – (licence CC-BY)

Un robot d'intelligence artificielle apprend dans une bibliothèque.



https://unsplash.com - Photo de Chris Ried

de déposer dans HAL la production scientifique pour le compte des unités, à la place des auteurs. L'administrateur HAL extrait ainsi chaque mois les nouvelles publications indexées dans les bases Web of Science et Pubmed, afin de les référencer dans HAL en utilisant la chaîne OverHAL-Zip2HAL. Un algorithme établit la correspondance entre des chaînes de caractères (noms d'auteurs, affiliations) et les personnes et laboratoires enregistrés dans les référentiels HAL. Grâce à ce dispositif, il nous a été possible d'atteindre assez vite l'exhaustivité du référencement bibliographique des unités de recherche, principalement sous forme de métadonnées (titre, auteur, résumé, etc.).

Ajouter et diffuser le texte intégral des articles

Cependant, le but d'une archive ouverte est avant tout d'offrir un accès ouvert au texte intégral de ces publications, et pas de se limiter à un simple catalogue. Pour cela, nous avons fait évoluer **OverHAL** de manière à pouvoir collecter les manuscrits des articles auprès des auteurs. En effet, la loi pour une République numérique⁴ autorise la diffusion dans HAL du « Manuscrit auteur accepté » (MAA), mais pas du PDF éditeur publié, sauf si celui-ci a été publié sous une licence ouverte Creative Commons (CC-BY)⁵. Nous avons donc adopté une nouvelle procédure, qui complète celle du référencement de la chaîne OverHAL-Zip2HAL : l'administrateur HAL fait chaque semaine une veille sur les plateformes éditeur afin de repérer les derniers articles publiés, qu'il collecte dans l'outil libre Zotero⁶, en y ajoutant les adresses mail des auteurs à contacter. Il importe ensuite ces données dans OverHAL pour envoyer des messages personnalisés aux auteurs de correspondance, les invitant à lui transmettre le « Manuscrit auteur accepté » par retour de mail. En cas de non-réponse, un mail de relance est envoyé. Cette fonctionnalité, depuis sa mise en place en 2013, a permis de récolter un très grand nombre de « Manuscrits auteurs acceptés », qui ont pu être ajoutés aux références bibliographiques déjà enregistrées dans HAL. Elle a en outre offert l'avantage de mieux nous faire connaître auprès des auteurs, d'engager avec eux des échanges fructueux et de les informer sur les différentes possibilités de HAL.

Assurer la qualité des métadonnées

En parallèle, une quatrième brique, **CrossHAL**, a été développée pour ajouter de façon systématique le PDF éditeur sous licence CC-BY, mais également pour enrichir les métadonnées des dépôts

existants d'une collection HAL, à partir de données de bases de données externes interrogeables avec des interfaces de programmation d'application (ou *Application Programming Interface* – API). Manière à la fois d'augmenter le taux d'accès ouvert des publications HAL et de veiller à la qualité et à la richesse descriptive du référencement.

Évaluer le taux d'accès ouvert aux publications

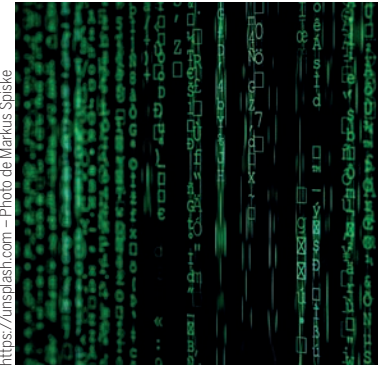
Toutes ces procédures ont été formalisées en 2017 dans une politique d'accès ouvert de l'Université de Rennes, invitant les chercheurs à ajouter ou faire ajouter le MAA de leurs articles dans HAL. Un nouvel outil, **VizuHAL**, a été développé à cette occasion⁷ pour mesurer chaque année la progression de l'accès ouvert, avant que le Baromètre de la science ouverte (BSO)⁸ du ministère de l'Enseignement supérieur et de la Recherche (MESR) n'en prenne le relais à partir de 2021 : dans la dernière édition du BSO, 80 % des publications 2021 de l'Université de Rennes étaient en accès ouvert, soit 13 points au-dessus de la moyenne nationale⁹. Succès qui témoigne de l'efficacité des procédures de référencement-dépôt mises en place avec les outils HALUR.

Vérifier sa liste de publications HAL

La dernière brique, **MyHAL**, a vu le jour en 2020, suite à l'obligation faite aux chercheurs du CNRS de déposer le texte intégral de leurs articles dans HAL, afin de valider le bilan annuel de leur production dans l'application CRAC du CNRS. Il fallait leur offrir un outil simple d'utilisation, leur permettant de consulter d'un clic la totalité de leurs publications HAL et de repérer d'un coup d'œil les PDF déjà déposés. Comme la plupart des articles étaient déjà référencés dans HAL par nos soins, il ne leur restait plus qu'à ajouter le « Manuscrit auteur accepté » aux notices sans texte intégral, au moyen d'un petit bouton créé à cet effet. Plus largement, MyHAL est utilisé par tous les chercheurs de l'Université, affiliés ou non au CNRS.

Pérenniser ces outils au service de la communauté

C'est ainsi qu'au fil des années, la suite HALUR s'est développée et enrichie, avec la volonté d'ouvrir ces outils à tous, autant que possible. À la base, nul cahier des charges, nul comité de pilotage ou autre *task force*. Juste la collaboration libre et informelle entre deux collègues passionnés, donnant forme à leur imagination à partir de besoins concrets. Revers de la médaille, la fragilité d'un dispositif qui repose essentiellement sur deux personnes, s'agissant de la maintenance et du support technique (mais l'aventure dure avec succès depuis 10 ans !), et sur une mémoire serveur limitée qui ne permet pas de traiter de gros volumes de requêtes ou de fichiers. Fort heureusement, des solutions émergent, ce qui devrait permettre à terme d'assurer leur pérennité et de mieux les partager avec la communauté des établissements représentés dans HAL¹⁰. ■



https://unsplash.com - Photo de Markus Spiske

4. <https://scienceouverte.univ-rennes.fr/que-dit-la-loi>
5. https://fr.wikipedia.org/wiki/Creative_Commons
6. Zotero est un logiciel de gestion bibliographique complet, libre et gratuit.
7. Avec la contribution de Thierry Fournier, Directeur adjoint du service commun de documentation de l'Université de Rennes.
8. Voir dans ce numéro l'article de Laetitia Bracco, p. 26.
9. <https://scienceouverte.univ-rennes.fr/barometre-science-ouverte-edition-2022>
10. Une enquête Flash d'octobre 2021 de l'association CasuHAL regroupant les utilisateurs HAL plaçait ExtrHAL et OverHAL dans le top 7 des chantiers prioritaires pour HAL : <https://www.casuhal.org/category/enquetes/>

OCdHAL, un outil de contrôle des publications déposées dans HAL¹

Évaluation, enquête, rapport d'activité, liste des publications scientifiques, production d'indicateurs bibliométriques... les activités conduites par les laboratoires de recherche sont régulièrement évaluées, notamment par le Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur (Hcéres)² selon un rythme quinquennal.

PATRICIA REYNIER

Ingénieure d'études au CNRS, chargée de ressources documentaires, laboratoire GIPSA-lab de l'INP (Images Parole Signal automatique), UMR CNRS-Université Grenoble Alpes

PASCAL BELLEMAIN

Ingénieur d'études au CNRS, informaticien et développeur

Un outil de valorisation

Un laboratoire de recherche doit ainsi pouvoir produire la liste des publications scientifiques de ses chercheurs et disposer rapidement et facilement de quelques statistiques ou indicateurs bibliométriques, afin de répondre aux demandes de ses tutelles, de valoriser sa production scientifique et de contribuer à la politique de science ouverte. La plateforme de l'archive ouverte nationale HAL est devenue un outil commun à l'ensemble des chercheurs, utilisée par les unités de recherche à la fois comme archive institutionnelle et comme vitrine pour leur production scientifique, avec une ou plusieurs collections à gérer. L'application OCdHAL³, « Outil de contrôle des données issues de HAL », propose un véritable tableau de bord permettant de voir et de contrôler en un seul écran toutes les métadonnées des documents déposés dans HAL et de gérer les collections d'une structure.

Un dépôt dans HAL peut être pauvre et peu homogène. Afin d'obtenir une liste fiable et complète, il est donc souvent nécessaire pour le gestionnaire de la collection d'intervenir dans HAL, directement ou indirectement. Indirectement ? C'est là qu'entre en action le logiciel OCdHAL ! Pour l'utiliser, il suffit d'avoir un compte sur HAL : l'application se présente sous la forme d'un tableur dynamique, les métadonnées rapatriées de HAL étant en colonnes et les « articles » en lignes. L'opérateur visualise les données qu'il peut modifier, ajoutant l'information absente ou corrigeant une erreur (les modifications sont synchronisées entre HAL et OCdHAL). OCdHAL permet aussi de poser des alertes, de définir des analyses, de préparer des listes selon ses propres nomenclatures (ou celles attendues par le Hcéres) et d'éditer une liste de publications avec la fonctionnalité ExtrHAL⁴ réalisée par l'Université de Rennes 1, intégrée dans OCdHAL...



La fleur emblématique d'OCdHAL et son abeille, inlassable travailleuse.

Avec des finalités multiples...

OCdHAL permet d'« enrichir » un dépôt et de « nettoyer » la collection, selon divers degrés de contrôles :

- contrôler le périmètre. Il s'agit de s'assurer par des alertes que les documents présents dans la collection relèvent bien du laboratoire, des années concernées et de son/ses affiliation(s) ;
- tendre vers l'harmonisation des métadonnées. Les métadonnées descriptives de l'article sont repérées comme erronées ou absentes ? Elles peuvent être corrigées, complétées selon les normes et formats en vigueur et des règles d'écriture. OCdHAL reprend le code couleur des référentiels de HAL (AURÉHAL)⁵. Le « nettoyage » des formes des auteurs, des affiliations ou encore des titres de revues devient aisé. Le fait de poser des alertes permet de repérer les anomalies (doublons, champs absents, auteurs sans identifiant IdHAL⁶, etc.) des données de HAL ;
- enrichir les métadonnées et appliquer les principes FAIR⁷ (« Facile à trouver, accessible, interopérable et réutilisable », par l'homme et la machine). Visualiser les données permet de compléter les

1. Hyper Articles en ligne (HAL), est une plateforme nationale d'archives ouvertes <https://hal.science/>

2. <https://www.hceres.fr/fr/le-hceres>

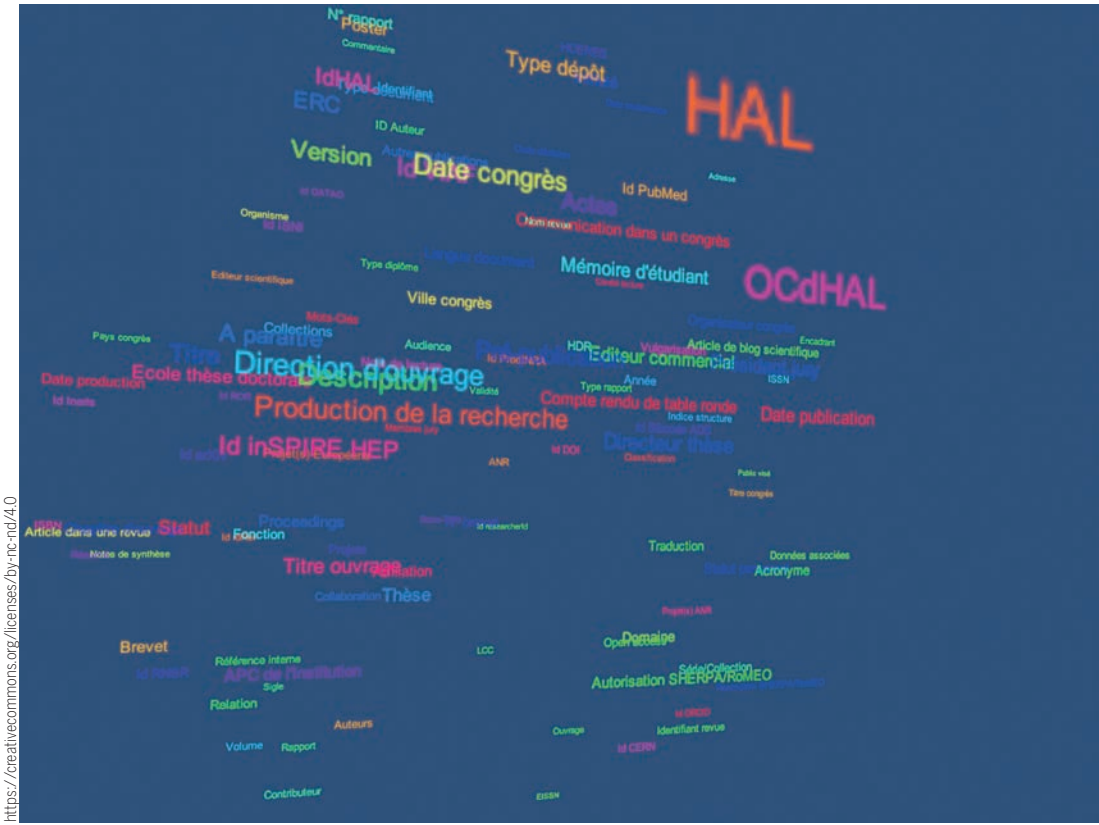
3. <https://ocdhal.univ-grenoble-alpes.fr/>

4. Voir dans ce numéro l'article de Laurent Jonchère et Olivier Troccaz, p. 40. <https://halur1.univ-rennes1.fr/ExtrHAL.php>

5. <https://aurehal.archives-ouvertes.fr/>

6. <https://doc.archives-ouvertes.fr/identifiant-auteur-idhal-cv/>

7. <https://www.ccsd.cnrs.fr/principes-fair/>



Les mots d'OCdHAL en 3D.

éléments nécessaires à l'édition d'une liste des publications, mais aussi de repérer les publications absentes ou qui pourraient faire l'objet d'un dépôt en texte intégral, grâce aux champs liés à SHERPA/RoMEO⁸ (base de données qui recense les politiques des revues scientifiques en matière de droit d'auteur et d'auto-archivage). Il ne reste plus qu'à contacter les auteurs des articles, afin qu'ils rendent accessible leur article, **dans le respect du droit** ;

- récupérer l'information hors de HAL et d'OCdHAL. Par exemple, un clic sur le titre d'une revue ou son ISSN renvoie vers le catalogue du SUDOC⁹ (système universitaire de documentation) et permet de vérifier l'exactitude du titre, des années de parution, de l'ISSN, de l'éditeur, d'attribuer la forme correcte d'un titre de revue dans les articles de sa collection et de mettre à jour la fiche du référentiel. Un clic sur un titre de pré-publication/document de travail ouvre un moteur de recherche, permettant de vérifier s'il y a une publication ou parution et de saisir alors l'information.

Pour servir l'évaluation de la recherche...

Afin d'adapter OCdHAL aux demandes des évaluateurs, de nouvelles colonnes ont été introduites, comme la qualification d'une collaboration (internationale, nationale, académique) ou encore le statut du personnel.

Ce travail effectué, une liste des publications pourra être éditée, ainsi que des indicateurs choisis :

- la production de « Tableaux de données » permet de regrouper les différentes informations demandées par le Hcéres (identifiant ORCID¹⁰ d'un auteur, publications d'un doctorant, etc.). À l'opérateur de sélectionner les données, de poser des filtres et d'exporter les métadonnées dans un tableau qui alimentera le rapport ;

- la production de la « Liste des publications » permet de paramétrer la liste des données demandées (les 20 % les plus significatifs de la production pour le Hcéres par exemple) et le formulaire ExtrHAL permet de faciliter l'édition d'une liste dans le style de son choix ;
- le module « Analyses » permet de générer des indicateurs prédéfinis ou de les personnaliser. Chaque opérateur peut conserver ses propres indicateurs d'une session à l'autre, en créer de nouveaux et les partager avec un autre usager. Les indicateurs permettent de répondre non seulement à des demandes internes comme externes, mais aussi à des enquêtes.

Quelles perspectives ?

Jusqu'où irons-nous ? Actuellement, l'application est réalisée par une chargée de documentation, administratrice d'un portail HAL et référente HAL d'une UMR, et par un développeur, personnel du CNRS. D'abord outil interne, OCdHAL est devenu un logiciel, avec un manuel d'utilisation. Depuis 2016, il est disponible en version Web. Nous ne cessons de l'améliorer, en répondant aux attentes des utilisateurs et en étant attentifs aux évolutions de HAL. La communauté de OCdHAL compte 600 utilisateurs et 110 utilisateurs d'OCidHAL¹¹ (logiciel dérivé dédié à la gestion d'un auteur/d'un identifiant d'auteur idHAL).

Aujourd'hui, plus de 1 000 collections, provenant de 600 structures de l'enseignement supérieur et de la recherche, mais aussi de la santé et de la culture, comme les laboratoires de recherche architecturale (Ambiances Urbanités Architectures [AAU], Architecture Urbanisme Société : savoir enseignement recherche [AUSser], Architecture, environnement et cultures constructives [AE&CC]) ou le Musée du quai Branly... sont chargées, complétées, enrichies. À vous d'essayer ! ■

8. <https://v2.sherpa.ac.uk/romeo/>
 9. <http://www.sudoc.abes.fr/cbs/>
 10. ORCID attribue depuis 2012 des identifiants ORCID (ou ORCID IDs), chaque identifiant étant composé d'un code alphanumérique qui permet d'identifier les chercheurs et auteurs de contributions académiques et scientifiques, et est progressivement imposé par les grands éditeurs : <https://orcid.org/>
 11. <https://ocdhal.univ-grenoble-alpes.fr/OCdHAL.OCidHAL>

Données et rapports d'opérations archéologiques à l'heure de la science ouverte

Aujourd'hui, l'enjeu prioritaire de la diffusion de la recherche archéologique est la mise à disposition de tous, chercheurs, amateurs éclairés et grand public, des rapports d'opérations archéologiques programmées et préventives, dont le nombre s'est considérablement accru avec le développement de l'archéologie préventive depuis la fin des années 1990 et la loi de 2001 qui l'a institutionnalisée¹.

JEAN-MANUEL CONILLEAU

Chargé d'études documentaires,
Direction régionale des affaires
culturelles (DRAC) Bretagne-Service
régional de l'archéologie (SRA)

CHRISTIAN CRIBELLIER

Adjoint au sous-directeur
de l'archéologie, Direction générale
des patrimoines et de l'architecture,
Sous-direction de l'archéologie,
ministère de la Culture

STÉPHANE DESCHAMPS

Conservateur régional de l'archéologie,
DRAC Île-de-France-SRA

CYNTHIA DOMENECH-JAULNEAU

Chargée de la documentation
scientifique, DRAC Île-de-France-SRA

THIERRY LORHO

Conservateur régional adjoint
de l'archéologie,
DRAC Centre-Val de Loire-SRA

YVES MENEZ

Conservateur régional de l'archéologie,
DRAC Bretagne-SRA

MARTIAL MONTEIL

Professeur d'archéologie romaine,
directeur de la revue *Gallia* et
d'*Archéologie de la France - Informations*
(AdIFI), Nantes Université - Laboratoire
de recherche Archéologie et
architectures (LARA-UMR CNRS 6566)
du Centre de recherche en archéologie,
archéosciences, histoire (CReAAH)

MARIE-ALICE VIRLOUVET

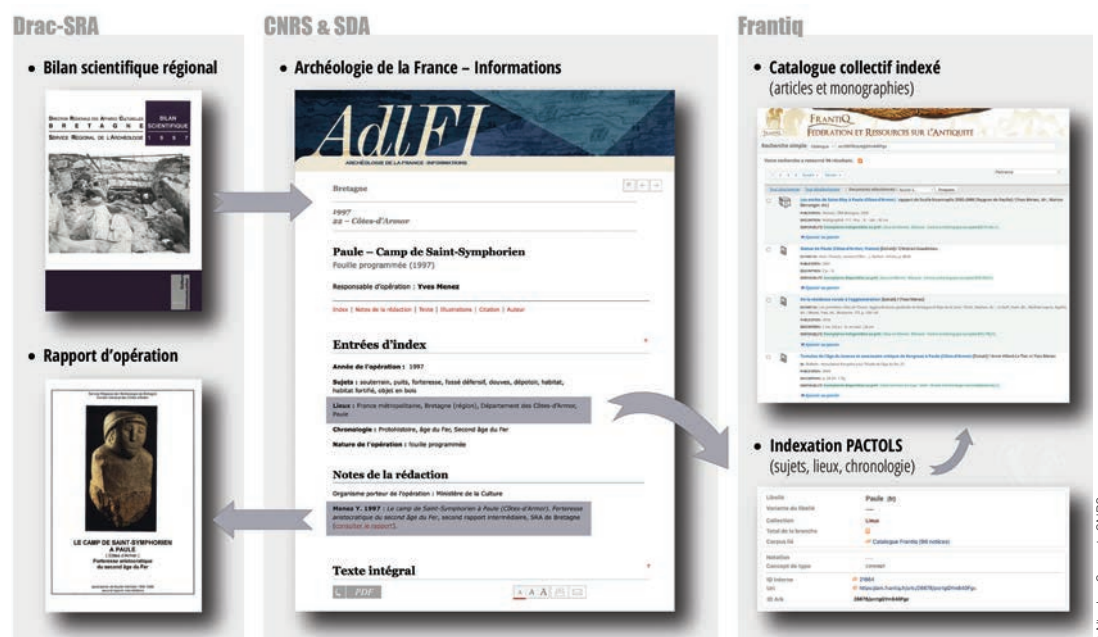
Chargée d'études documentaires,
DRAC Île-de-France-SRA

Les acteurs

Plusieurs Directions régionales des affaires culturelles-Services régionaux de l'archéologie (DRAC-SRA) ont ainsi développé des programmes de numérisation des rapports produits depuis les années 1940 et mis à disposition ces ressources sur des portails numériques régionaux dédiés, en complément d'un soutien apporté aussi aux publications. L'Institut national de recherches archéologiques préventives (INRAP) et quelques autres opérateurs d'archéologie

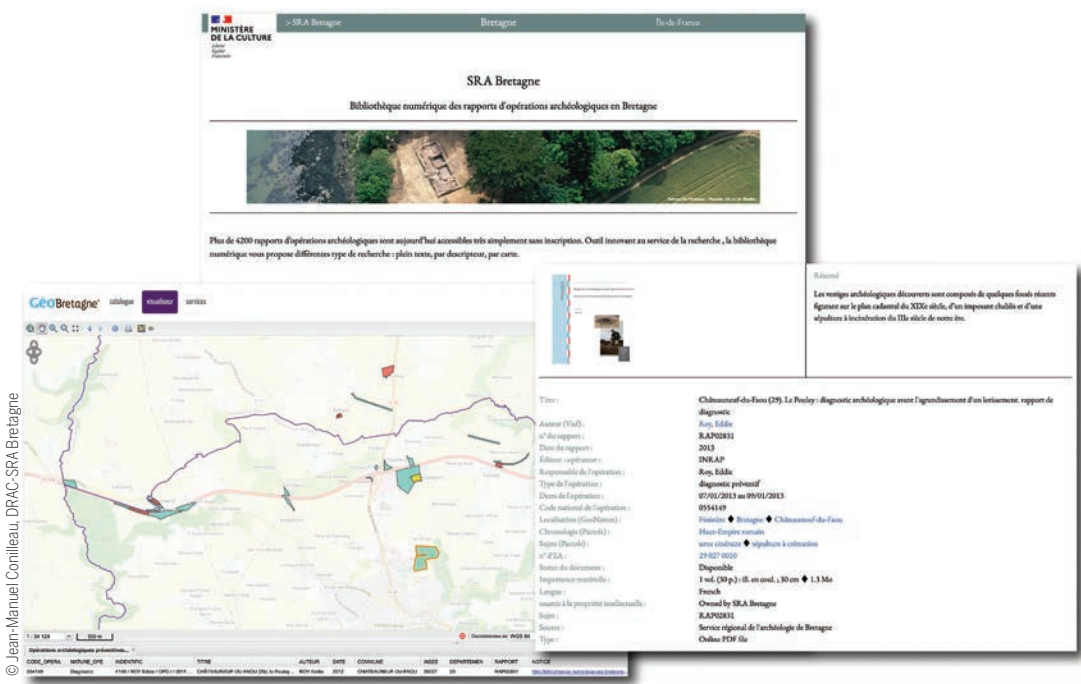
préventive se sont également engagés dans une telle diffusion.

Afin que les travaux réalisés atteignent le niveau de visibilité attendu au plan national et international, reflet du dynamisme de la recherche archéologique française, un portail national dédié agrégeant les différentes bases institutionnelles existantes ou en cours de développement devrait à terme faciliter l'accès à cette information scientifique de premier ordre. Il devrait aussi assurer leur mise en lien avec les notices



1. Loi n° 2001-44 du 17 janvier 2001 relative à l'archéologie préventive.

Des rapports d'opérations archéologiques aux monographies et périodiques indexés par le catalogue Frantiq : une circulation de l'information via les notices d'*Archéologie de la France - Informations* grâce à des référentiels partagés.



Interface de consultation de la bibliothèque numérique des rapports d'opérations archéologiques en Bretagne, accessible également par une interrogation cartographique des opérations réalisées sur le site GéoBretagne.

de synthèse publiées par *Archéologie de la France – Informations (AdIFI)*, revue coéditée par la sous-direction de l'archéologie et l'Institut des sciences humaines et sociales du CNRS (InSHS). Héritière des *Informations archéologiques* publiées par la revue *Gallia*, elle a pour mission de mettre en ligne celles-ci ainsi que les résumés des opérations archéologiques publiés à partir de 1992 par les SRA dans des *Bilans scientifiques régionaux* annuels. Cette riche information est harmonisée, structurée et indexée pour une diffusion numérique gratuite et en flux continu sur le portail OpenEdition Journals².

Ouverture sous vigilance

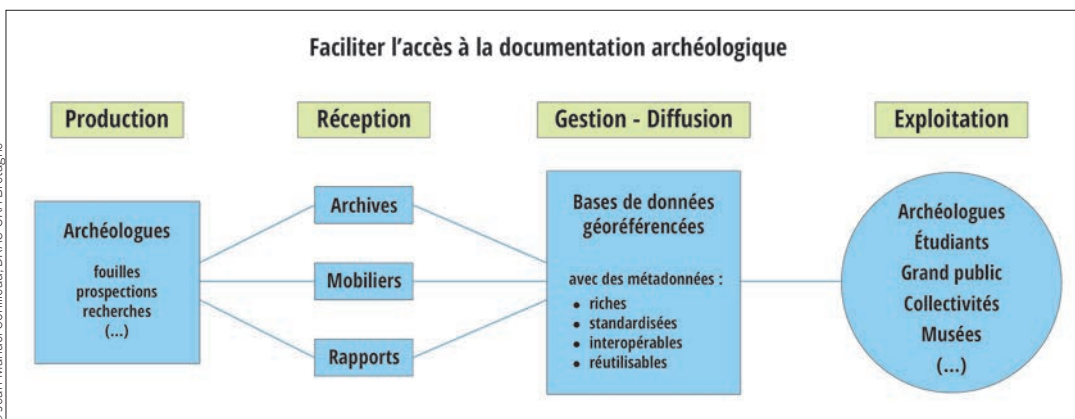
Parmi les DRAC-SRA bien avancées dans le processus, celle de Bretagne a, depuis 2010, mis en ligne 3 000 rapports d'opérations archéologiques qui ont été numérisés au format PDF-A (normes CINES³), ocrés⁴ et catalogués, puis déposés dans une bibliothèque numérique développée sous le logiciel libre Omeka. Une seconde étape a conduit à verser sur l'infrastructure de données spatialisées GéoBretagne⁵, développée sous GeOrchestra, l'information spatiale produite dans l'application nationale de gestion de la carte archéologique Patriarche. Il s'agit des sites archéologiques, de leurs zones de protection, des zones

de présomption de prescription archéologique (ZPPA) et des fouilles réalisées depuis 1950. Des liens entre ces deux outils ont été mis en place par Marie-Dominique Pinel et Thierry Lorho afin de permettre aux chercheurs et au grand public un accès par la cartographie ou la base de données⁶.

Une précaution a été prise lors de la mise en ligne : les lieux les plus susceptibles en Bretagne de livrer des objets métalliques et donc d'intéresser les usagers illégaux de détecteurs de métaux, ne sont pas identifiables comme tels parmi les 23 000 sites actuellement connus. Un texte d'introduction précise par ailleurs les protections réglementaires, notamment l'interdiction de fouilles ou de prospections non autorisées par le préfet de région. De même, les dispositions du Code de la propriété intellectuelle sont rappelées.

La réflexion se poursuit actuellement sur la mise en ligne d'autres données, dans le cadre de bases interopérables ayant des référentiels communs avec celles déjà mises en œuvre. Un objectif est de permettre aux chercheurs d'accéder en ligne aux objets issus des opérations, conservés dans cinq dépôts, et ce sous des formats compatibles avec les normes définies pour les musées de France, avec les inventaires complets et les clichés les plus significatifs. Un autre projet est la mise en ligne de la documentation issue

- <https://journals.openedition.org/adifi/>
- Le CINES est le Centre informatique national de l'enseignement supérieur.
- L'action d'ocriser un document consiste à en extraire le contenu via la reconnaissance des caractères, pour pouvoir ensuite le modifier.
- <https://cms.geobretagne.fr/>
- <https://www.inrap.fr/une-experience-d-open-accessopen-data-la-drac-bretagne-met-en-ligne-ses-donnees-13179>



L'objectif du système en cours de constitution dans les SRA de Bretagne et d'Île-de-France est de faciliter la consultation de l'essentiel de la documentation archivée par l'intermédiaire de bases de données géo-référencées et interopérables.

des opérations (relevés en plan et en coupe, photographies, dessins d'objets, restitutions, etc.), avec des inventaires conformes aux normes des archives, ainsi que les documents les plus remarquables (vues de chantier, de vestiges significatifs, etc.).

Au-delà de la numérisation...

La DRAC-SRA d'Île-de-France s'est également engagée en 2018 dans une politique de diffusion de la documentation scientifique. En 2020-2021, 4 000 rapports, des plus anciens jusqu'en 2004, ont ainsi été numérisés. L'enjeu était de répondre aux demandes croissantes d'accès à distance aux rapports, tout en se conformant à la loi pour une République numérique de 2016 et à la politique de données ouvertes (*Open Data*) du ministère de la Culture.

La diffusion des données se veut également protectrice des sites contre le pillage, des dispositions du Code de la propriété intellectuelle et des données personnelles. Ainsi, afin d'éviter les fouilles illégales, la priorité a été donnée à la mise en ligne des rapports de fouilles préventives et, pour les diagnostics, de ceux menés en contexte urbain. Par la suite, un examen des rapports liés à des sites présentant *a priori* plus de risques de pillage sera fait en amont de leur mise en ligne. S'agissant des données personnelles, une anonymisation partielle du PDF est réalisée, si nécessaire.

Comme en Bretagne, et pour chaque rapport, une notice de la bibliothèque numérique synthétise des métadonnées bibliographiques en lien avec l'opération, issues de la carte archéologique, auxquelles s'ajoutent une indexation (sujet, lieu, chronologie), un résumé et la localisation sur la carte issue d'OpenStreetMap. Du côté des utilisateurs, l'accès à l'interface est libre et la recherche se fait en plein texte, mais aussi par les termes d'indexation ou, visuellement, sur la carte. La mise en ligne est prévue au 1^{er} semestre 2023.

Parallèlement, le SRA a engagé la numérisation d'une partie de ses archives, dont celles de la circonscription des Antiquités préhistoriques d'Île-de-France. Dès le début des années 1970, les informations concernant les découvertes préhistoriques et

Afin d'éviter les fouilles illégales, la priorité a été donnée à la mise en ligne des rapports de fouilles préventives et, pour les diagnostics, de ceux menés en contexte urbain.

Par la suite, un examen des rapports liés à des sites présentant *a priori* plus de risques de pillage sera fait en amont de leur mise en ligne.

protohistoriques de la région ont été consignées de façon systématique et normalisées sur des fiches, rangées et triées par départements puis par communes. En tout, 10 980 fiches de cet inventaire ont été numérisées, océrisées et indexées. Un long travail de recollection est en cours pour faire le lien entre les sites enregistrés dans Patriarche et ceux de ces fiches. L'objectif est une mise en ligne en projet avec le consortium Paris Time Machine (IR* Huma-Num⁷), interrogeable *via* une entrée géographique et thématique, afin d'ouvrir de nouvelles possibilités de recherches.

L'importance des réseaux

La clé de voûte commune de l'évolution récente de ces bibliothèques numériques, gérées en Bretagne par Laure Cadars, puis Jean-Manuel Conilleau, et en Île-de-France par Marie-Alice Virlovet, est la participation au réseau Frantiqu⁸ et la collaboration avec le consortium MASA⁹ (Mémoires des archéologues et des sites archéologiques). Un outil plus performant a ainsi été développé dans une optique d'ouverture des données (FAIR¹⁰), tandis que l'indexation des rapports a été améliorée, en enrichissant et en standardisant les métadonnées et en les rendant interopérables. Le vocabulaire a été aligné sur différents référentiels normés et reconnus internationalement : les thésaurus PACTOLS¹¹ pour les sujets et GeoNames¹² pour les localisations, le VIAF (*Virtual International Authority File*)¹³ pour les auteurs. Le jeu de données est également aligné sur l'ontologie du CIDOC-CRM¹⁴. Dans un avenir proche, la recherche simultanée dans les bibliothèques numériques des deux SRA pourrait être rendue possible depuis l'interface publique du catalogue Frantiqu. ■

7. Voir dans ce numéro l'article d'Olivier Baude et Nicolas Larrousse, p. 11.

8. La Fédération Frantiqu est constituée en groupement de service* (GDS 3378) du CNRS ; elle rassemble une quarantaine d'unités mixtes du CNRS, des services du ministère de la Culture, des musées d'archéologie, des collectivités territoriales et d'autres établissements publics. Bibliothécaires et documentalistes du réseau partagent des outils communs et offrent des services associés aux archéologues, en formation ou professionnels.

9. Le consortium MASA a pour objectif la diffusion et la mise en œuvre des principes FAIR (« facile à trouver, accessible, interopérable et réutilisable ») dans la communauté archéologique. Il comprend dix partenaires : Maison des sciences de l'homme du Val de Loire, Maison de la recherche en sciences humaines de Caen, Maison de l'Orient et de la Méditerranée de Lyon, Maison méditerranéenne des sciences de l'homme, Maison des sciences de l'homme mondes de Nanterre, Musée d'archéologie nationale de Saint-Germain-en-Laye, groupement de service Frantiqu, INRAP, Maison interuniversitaire des sciences de l'homme - Alsace, Réseau des écoles françaises à l'étranger (ResEFE).

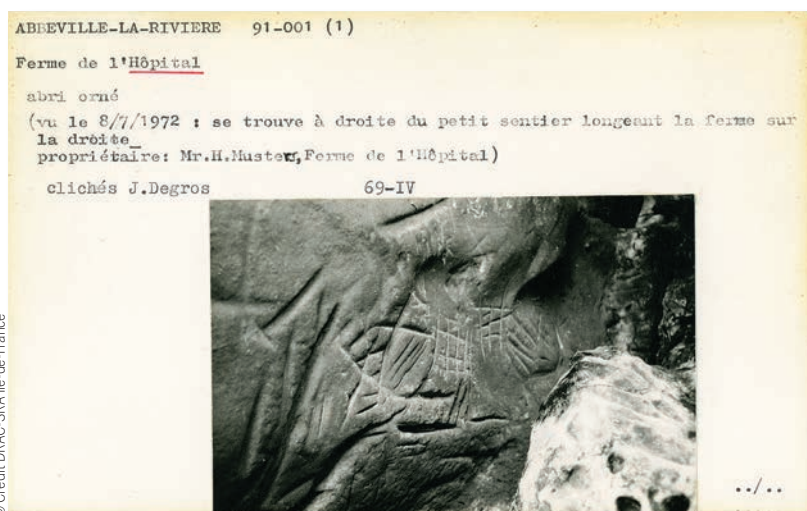
10. <https://www.ccsd.cnrs.fr/principes-fair/>

11. <https://pactols.frantiqu.fr/>

12. <https://www.geonames.org/>

13. <https://viaf.org/>

14. <https://www.cidoc-crm.org/>



Exemple d'une fiche de site (gravures pariétales) établie en 1972, conservée au sein du SRA d'Île-de-France et numérisée pour être mise en ligne avec d'autres archives semblables.

La science ouverte à l'INRAP

Avec l'apparition de méthodes et d'outils de production de données archéologiques nativement numériques, les archéologues se sont trouvés confrontés à une évolution profonde et durable des modalités d'enregistrement et de diffusion de ces données, dont la plus grande partie est produite aujourd'hui par l'archéologie préventive. Avec plus de 2 000 opérations archéologiques par an, l'Institut national de recherches archéologiques préventives (INRAP) est l'un des principaux producteurs et gestionnaires de ces données. Il se trouve, de ce fait, concerné en premier chef par ces problèmes, raisons pour lesquelles il a déjà élaboré un certain nombre de protocoles et d'outils, tels Dolia, Caviar, HAL, Archipel, publications scientifiques en accès ouvert. Cet engagement implique une réflexion sur la structuration des activités de gestion, d'archivage et de partage des données, préalable à la diffusion des résultats de la recherche dans le respect des principes de l'accès ouvert (*Open Access*) et des données ouvertes (*Open Data*).

Historique et politique de l'établissement

Depuis presque une décennie, la notion d'ouverture des données a constitué un objectif fort des acteurs de la politique d'établissement de l'INRAP. Ces efforts ont conduit à l'organisation, en 2017, d'un séminaire pour répondre à l'appel à projet Open Access Week 2016-2017 proposé par le consortium Couperin¹. À cette occasion, des ateliers ont permis à une cinquantaine d'agents de l'institut de réfléchir aux enjeux et aux conditions de la mise en œuvre d'une démarche d'accès ouvert à l'INRAP.

Un double mouvement de fond intervient dans l'écosystème de la recherche archéologique : les données ouvertes et la science ouverte. Pour les données ouvertes, la loi pour une République numérique met à la charge de l'Institut l'obligation par principe de publier en ligne les documents administratifs². La science ouverte s'appuie quant à elle sur l'opportunité que représente la mutation numérique pour développer l'accès ouvert aux publications et – autant que possible – aux données de la recherche.

Fort des tutelles des ministères de la Culture, de l'Enseignement supérieur et de la Recherche et des missions que la loi lui confère, l'INRAP, principal institut en Europe dédié à la recherche archéologique en contexte préventif, s'est engagé dans une démarche de partage des données, en développant les actions suivantes :

- harmoniser les pratiques relatives à la production des données en définissant un socle commun minimal pour permettre l'interopérabilité de multiples systèmes d'enregistrement et en accompagnant les personnels dans cette démarche, notamment en développant des actions d'information, de formation et en proposant des guides de bonnes pratiques ;
- renforcer la démarche de partage des données issues des opérations archéologiques dès l'émission d'un

arrêté de prescription en mettant en place un plan de gestion de données, en développant un système raisonné de gestion et stockage des données produites et en accompagnant les personnels dans une démarche de description normalisée pour la structuration et le signalement des données (métadonnées) ;

- améliorer la visibilité de l'activité de recherche en archéologie préventive (mise en avant des projets collectifs et des équipes) ;
- valoriser la production scientifique des archéologues (mise en valeur des personnes) ;
- développer un environnement virtuel de recherche commun dédié à l'archéologie préventive pour faciliter la recherche des sources et de la documentation inédite ;
- mettre à disposition des outils en ligne permettant aux archéologues de contribuer à l'enrichissement de cet environnement de recherche ;
- développer de nouveaux supports de diffusion de la production scientifique (bases de données, rapports spécialisés, articles, etc.) qui permettront des liens entre des contenus éditorialisés et les données qui leur sont associées.

Aujourd'hui, la mise en œuvre d'un programme sur les données scientifiques numériques de l'institut, Archipel, doté d'un financement du fonds pour la transformation de l'action publique, s'inscrit pleinement dans les ambitions de la science ouverte.

Bonnes pratiques et outils

Le volume, le temps contraint, la diversité des producteurs rendent difficile la bonne gestion des données produites au cours des opérations en vue de leur conservation et de leur partage.

Seuls les rapports d'opération font l'objet d'un signalement systématique et d'une conservation pérenne assurée par le Centre informatique national

MARC BOUIRON

Directeur scientifique et technique de l'Institut national de recherches archéologiques préventives (INRAP)

EMMANUELLE BRYAS

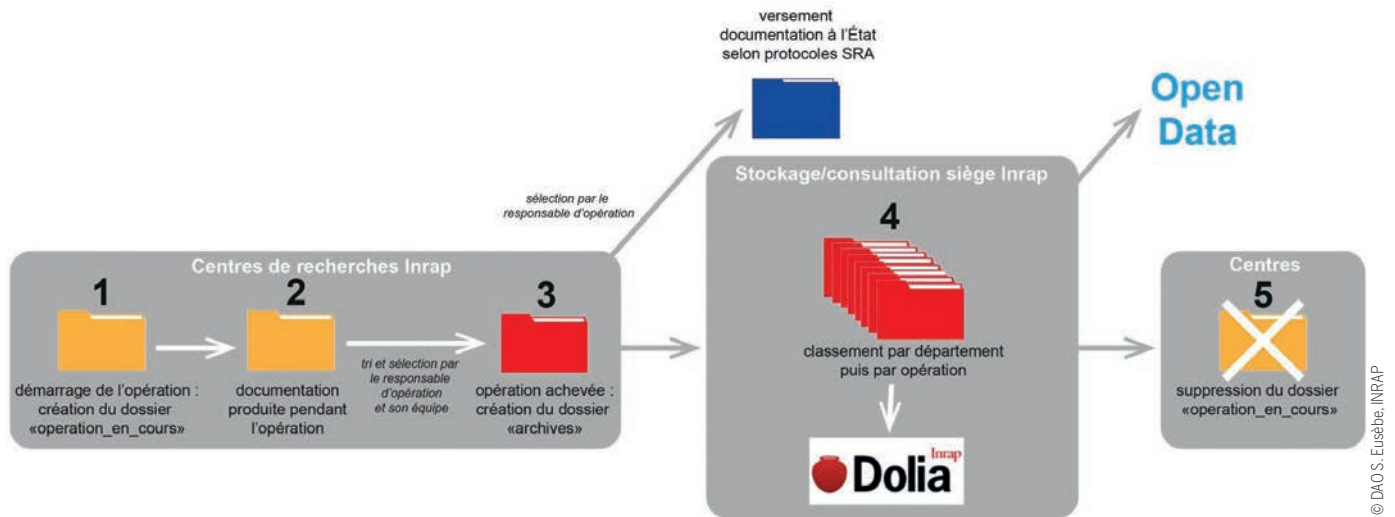
Chargée de ressources et d'ingénierie documentaire à l'INRAP

KAI SALAS ROSSENBACH

Chargé des programmes de recherches stratégiques à l'INRAP

1. Séminaire des 23 et 24 mars 2017, par l'INRAP avec le soutien du consortium Couperin, du projet européen NEARCH et de l'Université René-Descartes, coordonné par Emmanuelle Bryas, Gilles Bellan et Federico Nurra (INRAP). Voir dans ce numéro l'article de Federico Nurra, p. 23.

2. Loi n° 2016-1321 du 7 octobre 2016 pour une République numérique : <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000033202746>



Processus de diffusion et d'archivage des opérations archéologiques.

3. L'*Universal Machine Readable Cataloguing* (UNIMARC) est un format permettant le catalogage, le traitement et l'échange de données bibliographiques sous forme informatisée.

4. Le thésaurus PACTOLS est un vocabulaire contrôlé, normalisé et multilingue pour l'archéologie, depuis la préhistoire jusqu'à la période contemporaine et pour les sciences de l'Antiquité sous tous leurs aspects. Il a été créé par Frantiq (Fédération et ressources sur l'Antiquité constituée en groupement de service du CNRS n° 3378), qui rassemble une quarantaine d'unités mixtes du CNRS, des services du ministère de la Culture, des musées d'archéologie, des collectivités territoriales et d'autres établissements publics.

5. Ark (*Archival Resource Key*) est un format d'identifiant créé en 2001 par la *California Digital Library* (CDL) qui a vocation à identifier des ressources de tous types – physiques (échantillons destinés à une expérience scientifique, produits éditoriaux, etc.), numériques (livres numérisés, notices de catalogue, etc.) ou même immatériels (concepts). Son but est de fournir des identifiants adaptés aux besoins des producteurs et diffuseurs de données sur le Web, mais également capables de durer sur le long terme.

6. Le Dublin Core est un format descriptif simple et générique créé en 1995.

7. FAIR : *Findable, Accessible, Interoperable, Re-usable* pour « facile à trouver, accessible, interopérable et réutilisable ».

de l'enseignement supérieur (CINES). Ils sont référencés dans le catalogue documentaire en ligne Dolia, construit dans le respect des règles qui permettent l'interopérabilité des données : la modélisation des données selon le format UNIMARC³, l'utilisation de référentiels métier et des thésaurus PACTOLS du réseau Frantiq⁴, et l'attribution à chaque référence d'un identifiant numérique pérenne, un lien Ark⁵. À partir de cette description, il est possible de générer les métadonnées du rapport, qui sont régulièrement exportées dans le catalogue collectif indexé de Frantiq et la plateforme européenne des données archéologiques Ariadne+, et formatées en Dublin Core⁶ pour l'archivage au CINES des rapports dans un format compatible avec la conservation et le partage.

Dolia est également une bibliothèque numérique dans laquelle est mis en ligne l'ensemble des rapports de diagnostic et de fouilles pour les personnels de l'INRAP. Les utilisateurs externes à l'Institut n'ont accès qu'à une sélection de rapports de fouilles. L'ouverture plus large des rapports d'opération pose un certain nombre de questions auxquelles l'INRAP répond petit à petit afin de remplir à terme son obligation d'ouverture des données.

Le reste des données produites au cours des opérations fait l'objet d'une politique de gestion raisonnée. Les archéologues sont invités à nettoyer et à classer les données qu'ils produisent dans une arborescence partagée afin de créer un dossier rassemblant l'ensemble de la documentation produite par opération. Celui-ci est déposé sur un serveur central sécurisé accessible aux agents de l'Institut et référencé dans Dolia au même titre que le rapport. L'objectif est de faciliter le versement à l'État, mais également de constituer à l'INRAP un fonds documentaire de référence permettant la réutilisation, la diffusion et le partage de tout ou partie de la documentation produite. Le travail se poursuit afin de permettre l'ouverture de ces données à tous.

L'INRAP mène une politique en faveur d'une adoption large des systèmes d'informations géographiques qui permettent d'appréhender l'information archéologique dans sa dimension spatiale. Les données produites sont versées dans le catalogue de visualisation de l'information archéologique (Caviar).

Ce catalogue n'est pour le moment accessible qu'aux personnels de l'Institut.

La mise en œuvre de l'ouverture

Le programme Archipel en cours de construction a deux objectifs principaux : harmoniser, clarifier et stabiliser les pratiques et les systèmes de gestion des données numériques scientifiques et construire une plateforme d'accès aux données archéologiques numériques issues des recherches de l'INRAP.

Le premier objectif doit permettre de mettre en cohérence les initiatives internes avec l'état de l'art : produire des données plus FAIR⁷, en faisant évoluer les systèmes existants, en améliorant la gestion à l'INRAP et en clarifiant la stratégie de partage ; organiser la gouvernance des données et les processus humains ; promouvoir de nouvelles méthodes d'analyse de données.

Le second objectif est donc de construire une plateforme d'accès aux données archéologiques numériques issues des recherches de l'INRAP, par le biais de la donnée elle-même ou de sa métadonnée. Celle-ci devra en priorité répondre aux besoins de la recherche scientifique et permettre un accès aux données sur le Web, universel, ouvert à tous.

Quelques chantiers à venir

Après analyse des besoins et clarification des aspects juridiques, deux chantiers importants nous attendent : la construction d'un modèle de données reposant sur les ontologies et le développement d'un moteur de recherche intégré dans une interface utilisateur ergonomique.

Parallèlement, l'INRAP poursuit sa politique de libre accès : la revue à comité de lecture de l'Institut, *Archéopages*, est maintenant en ligne sur OpenEdition.org, de même que la collection *Recherches archéologiques*, coéditée avec CNRS Éditions et accessible sur le portail HAL de l'INRAP.

Par ailleurs, a été développée une nouvelle collection entièrement en ligne, les *Documents d'archéologie préventive* (DAP), qui accueille une sélection de rapports d'opération de fouille de l'Institut, enrichis d'une introduction qui permet aux auteurs de contextualiser et d'actualiser si nécessaire les résultats du rapport initial. ■

Valoriser en diffusant les travaux d'étudiants : le portail d'archives ouvertes DUMAS

Comment valoriser les travaux d'étudiants? Inscrit dans le mouvement de la science ouverte, le portail DUMAS (Dépôt universitaire de mémoires après soutenance) donne accès à plus de 47 000 mémoires issus des établissements de l'enseignement supérieur. Il s'agit d'une infrastructure unique dont le but est d'accroître la visibilité et l'accessibilité des travaux des étudiants, de valoriser les activités pédagogiques et de formation à la recherche des universités, écoles et laboratoires et de préserver ce riche patrimoine institutionnel.

Enjeux

La France s'inscrit dans le mouvement de la science ouverte pour une « diffusion sans entrave des résultats, des méthodes et des produits de la recherche scientifique¹ ». Cette nouvelle politique s'adresse avant tout aux scientifiques – chercheurs, enseignants-chercheurs, ingénieurs, doctorants, etc. – dans les établissements et organismes de recherche mais engage également l'enseignement supérieur dans un travail commun sur les ressources éducatives libres et

Rendre visibles et valoriser les meilleurs de ces travaux, créer des vitrines pour les parcours, diplômes et établissements, et contribuer par là à l'insertion professionnelle et scientifique des étudiants, dans le respect de leurs droits d'auteur.

dans le développement d'unités d'enseignement ou de parcours de formation en science ouverte aux niveaux licence et master.

Se pose aussi, dans ce contexte, la question de la valorisation des travaux des étudiants du deuxième cycle de l'enseignement supérieur, notamment des mémoires de fin d'études scientifiques et des thèses d'exercice dans les domaines de la santé. Il s'agit d'un patrimoine institutionnel d'une grande richesse et diversité qui couvre l'ensemble des disciplines scientifiques mais qui reste trop souvent invisible et mal conservé. Or les principes, outils et procédures de la science ouverte proposent aujourd'hui des solutions pour relever les enjeux de ces travaux : rendre visibles et valoriser les meilleurs de ces travaux, créer des vitrines pour les parcours, diplômes et établissements, et contribuer par là à l'insertion professionnelle et scientifique des étudiants, dans le respect de leurs droits d'auteur.

DUMAS

La plateforme d'archives ouvertes HAL, créée en 2001 par le Centre pour la communication scientifique directe du CNRS (CCSD) devient un support intéressant pour valoriser de la littérature grise, donc non publiée, qui abonde dans les établissements supérieurs d'enseignement et de recherche. Il devenait difficile de contenir et de conserver cette production de travaux dits « non aboutis » ou « inachevés », bien que véritable richesse et matériau pour la recherche elle-même. Valoriser ces travaux en archives ouvertes, c'est à la fois donner une forme de reconnaissance au travail

FRÉDÉRIC GENDRE

Ingénieur d'études, responsable de la bibliothèque Humanités (Université de Lille), documentaliste de la *Revue des sciences humaines*

CHRISTOPHE HUGOT

Ingénieur d'études, responsable de la bibliothèque des sciences de l'Antiquité (Université de Lille)

JOACHIM SCHÖPFEL

Maître de conférences à l'Université de Lille, membre du Groupe d'études et de recherche interdisciplinaire en information et communication (GERiiCO), Unité de recherche CNRS 4073)

AGNÈS SOUCHON

Ingénieure d'études, responsable de la bibliothèque de l'unité de formation et de recherche (UFR) Arts et Sciences humaines (Université Grenoble Alpes), administratrice du portail DUMAS

1. Deuxième Plan national pour la science ouverte 2021-2024 : <https://www.ouvrirlascience.fr/deuxieme-plan-national-pour-la-science-ouverte/>

de recherche effectué pendant au moins un an, et permettre à un large public d'en prendre connaissance, peut-être pour mieux en explorer certaines pistes ou établir des collaborations. C'est aussi et surtout une manière de participer au mouvement national de la science ouverte².

À partir de 2006, une réflexion sur la conservation et la valorisation des travaux académiques est menée au sein de l'UFR Arts et Sciences humaines de l'ex-Université Pierre-Mendès-France, (Université Grenoble Alpes aujourd'hui), aboutissant en 2008 au lancement de DUMAS³, portail pluridisciplinaire de HAL adapté aux différents types de mémoires.

Le nombre annuel de dépôts dans DUMAS est passé d'une centaine en 2008 à près de 7 000, 13 ans après (figure 1). L'évolution des consultations et des téléchargements des fichiers PDF au fil des ans montre l'intérêt que la base suscite (figure 2).

À ce jour, DUMAS propose plus de 47 000 mémoires en texte intégral. 63 % relèvent des formations en médecine et santé, 33 % sont issus des formations en SHS. On y trouve 1 503 mémoires en architecture et aménagement de l'espace, 733 mémoires en art et histoire de l'art et 68 mémoires en héritage culturel et muséologie.

Parmi ses 120 collections, DUMAS compte 7 collections d'établissements sous la tutelle du ministère de la Culture, dont 5 Écoles nationales supérieures d'architecture (ENSA), l'École du Louvre et le Muséum national d'Histoire naturelle, ainsi que 2 laboratoires de recherche (CRATERre, Centre international de la construction en terre⁴, et PACEA, De la préhistoire à l'actuel : culture, environnement et anthropologie – UMR 5199) totalisant 1 176 mémoires en libre accès. En 2023, le réseau des bibliothèques des écoles d'art et de design (BEAR⁵), qui rassemble une soixantaine de structures documentaires, choisit de donner plus de visibilité aux « mémoires-objets » des étudiants en les diffusant sur le portail DUMAS (voir focus p. 52).

Le rôle prégnant des enseignants

Le dépôt des mémoires n'est pas en auto-archivage, il se fait après soutenance et sous la responsabilité des enseignants, bibliothécaires ou documentalistes.

Le rôle des enseignants et, en particulier, des responsables de formation est essentiel. Ils doivent sensibiliser les étudiants, valider la sélection des travaux à déposer et obtenir les autorisations nécessaires. Parfois ils déposent les mémoires, créent les métadonnées et s'occupent de la curation de la

2. Ouvrir la science : <https://www.ouvrirlascience.fr/accueil/>
3. <https://dumas.ccsd.cnrs.fr/>
4. Depuis 1979, CRATERre œuvre à la reconnaissance du matériau terre afin de répondre aux défis liés à l'environnement, à la diversité culturelle et à la lutte contre la pauvreté. Dans cette perspective, CRATERre poursuit trois objectifs : mieux utiliser les ressources locales, humaines et naturelles ; améliorer l'habitat et les conditions de vie ; valoriser la diversité culturelle.
5. https://www.bsad.eu/index.php?lvl=cmepage&pageid=6&id_rubrique=23

Figure 1 – Évolution des dépôts dans DUMAS (2008-2023)

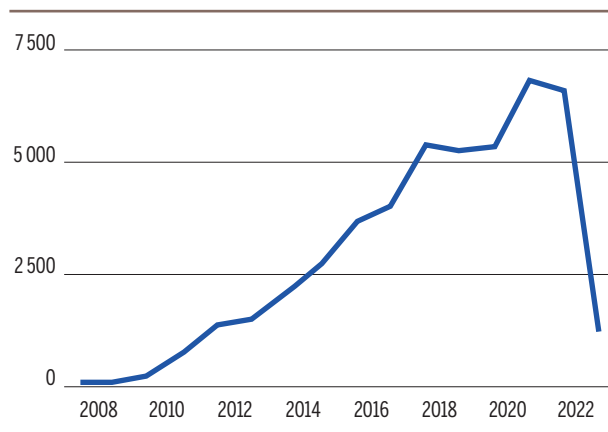
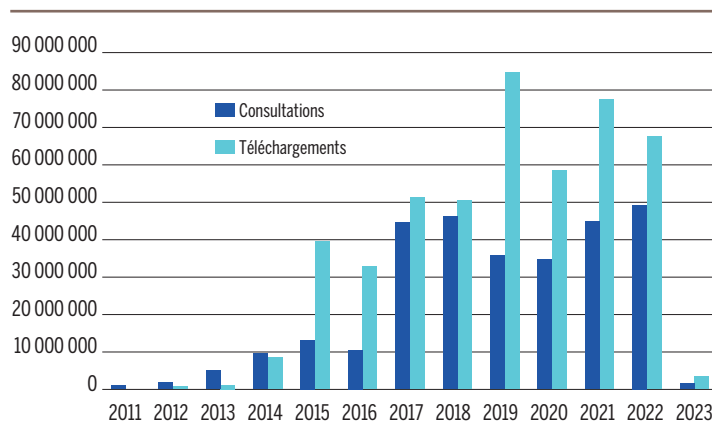


Figure 2 – Évolution des consultations et téléchargements de DUMAS (2011-2023)



Le campus Pont de Bois.



Étudiants en cours.

collection. Ils encouragent aussi la valorisation des mémoires sur les pages Web de la formation, dans les guides des études, dans l'intranet des établissements ou *via* les réseaux sociaux.

À titre d'exemple, la collection des mémoires du département Sciences de l'information et du document de l'Université de Lille, mise en place en 2010 à l'initiative d'une équipe d'enseignants-chercheurs, contient aujourd'hui plus de 600 mémoires de master, DESS, DEA⁶ et maîtrise de 1996 à 2022 sur les bibliothèques, la gestion documentaire, les sites Web, la littérature de jeunesse, etc. Certains ont été téléchargés plusieurs milliers de fois, partout dans le monde.

Et... le chaînon indispensable : les bibliothécaires

Les bibliothécaires ont été déterminants pour convaincre le conseil de la faculté des Humanités de l'Université de Lille d'adopter DUMAS comme archive officielle et unique pour les mémoires de master 1 et 2, remplaçant leur archivage papier. Il a également été proposé et obtenu qu'un niveau d'entrée soit exigé, la mention « très bien », afin que les mémoires aient valeur d'exemple méthodologique et rédactionnel.

Le rôle des bibliothécaires consiste par ailleurs, et peut-être essentiellement, une fois le principe de l'archivage mis en place, à accompagner les étudiants dans la préparation de leurs travaux pour la publication (respect des normes bibliographiques, présentation) ainsi qu'à rappeler aux différents acteurs universitaires la nécessité de faire vivre ces archives. Ce sont enfin les bibliothécaires qui renseignent la grille de saisie et intègrent le fichier au sein du portail, veillant à la qualité et à la communicabilité de la notice.

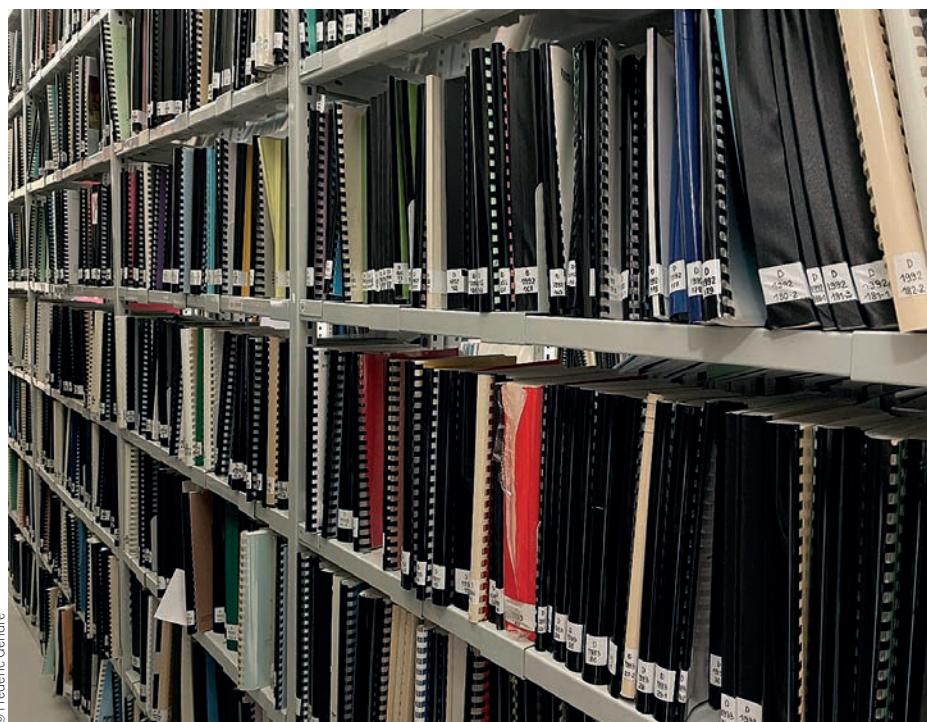
Points de vigilance !

Le dépôt dans DUMAS nécessite une double voire triple autorisation : par l'auteur du mémoire (propriété intellectuelle), par le jury de soutenance (qualité, absence de données personnelles) et, le cas échéant, par l'organisme d'accueil du stage (absence d'information confidentielle). Parfois, la diffusion en libre accès fait craindre le risque de plagiat ; au contraire, ils en sont protégés par leur caractère public, ce que ne permettent pas les archives papier.

Visibles à l'échelle internationale, les mémoires contribuent à valoriser les activités pédagogiques et d'initiation à la recherche des établissements et participent de leur mémoire institutionnelle. Avec seulement 3 % des mémoires sur DUMAS en anglais ou d'autres langues, la diversité linguistique y est certes peu présente, mais beaucoup plus de mémoires sont diffusés avec un titre, un résumé et des mots-clés en anglais.

DUMAS peut jouer le rôle d'archive pérenne des formations, les fichiers déposés bénéficiant de l'archivage à long terme du CINES (Centre informatique national de l'enseignement supérieur), à charge des établissements de faire le lien avec leur système de gestion locale.

Sur DUMAS, le mémoire le plus ancien date de 1942⁷. La connaissance des mémoires anciens est un



© Frédéric Gendrie

Collection de mémoires, bibliothèque Georges Lefebvre.

enjeu particulier, car peu catalogués et souvent dispersés : trop peu de mémoires ont fait l'objet d'une publication, d'une recension, voire d'une simple mention bibliographique, alors que l'intérêt historiographique est pourtant important, tant pour l'information donnée sur le sujet traité que pour la connaissance de l'évolution des attendus de la recherche et de la pédagogie dans un champ disciplinaire. L'enregistrement, la numérisation et la diffusion de mémoires anciens, parfois centenaires, permettraient de répondre à ce double objectif.

Questions (très) ouvertes

De plus en plus d'établissements ont fait le choix de DUMAS. D'autres ont mis en place leur propre plateforme de mémoires. Difficile de dire quelle est la meilleure solution.

Mais il est certain qu'une plateforme nationale mutualisée est moins onéreuse à mettre en place et à maintenir, que sa masse critique augmente la visibilité et l'impact des dépôts, avec l'avantage incontournable d'offrir aux étudiants un lieu mieux identifié où ils peuvent trouver aussi bien des sources documentaires que des exemples à suivre.

Au fur et à mesure des besoins des contributeurs, le portail évolue. Aujourd'hui, en participant à DUMAS, trois écoles d'art (École européenne supérieure d'art de Bretagne [EESAB], École supérieure des beaux-arts de Nîmes [ESBAN], Institut national supérieur d'enseignement artistique Marseille-Méditerranée [INSEAMM]) apportent un nouveau défi, celui de la représentation virtuelle de l'œuvre artistique que peut constituer un mémoire⁸. Comment représenter un objet d'art en format numérique sur une plateforme de mémoires habituellement textuels ? Comment valoriser une nouvelle forme de mémoire sans trahir la pensée de l'auteur, la création elle-même ? Un enjeu peut-être pour les humanités numériques ? ■

6. Le DESS (Diplôme d'études supérieures spécialisées) est un ancien diplôme de l'enseignement supérieur de niveau bac + 5 préparant à la vie professionnelle ; le DEA (Diplôme d'études approfondies) est également un ancien diplôme de l'enseignement supérieur de niveau bac + 5 qui avait pour objet de sanctionner la première année de troisième cycle en vue de la préparation d'un doctorat. L'inscription à cette année de formation était subordonnée à l'obtention d'une maîtrise ou au bénéfice d'un niveau jugé équivalent.

7. Mémoire en paléontologie du diplôme d'études supérieures de la faculté des sciences de l'Université de Grenoble : <https://dumas.ccsd.cnrs.fr/dumas-00988256v1>

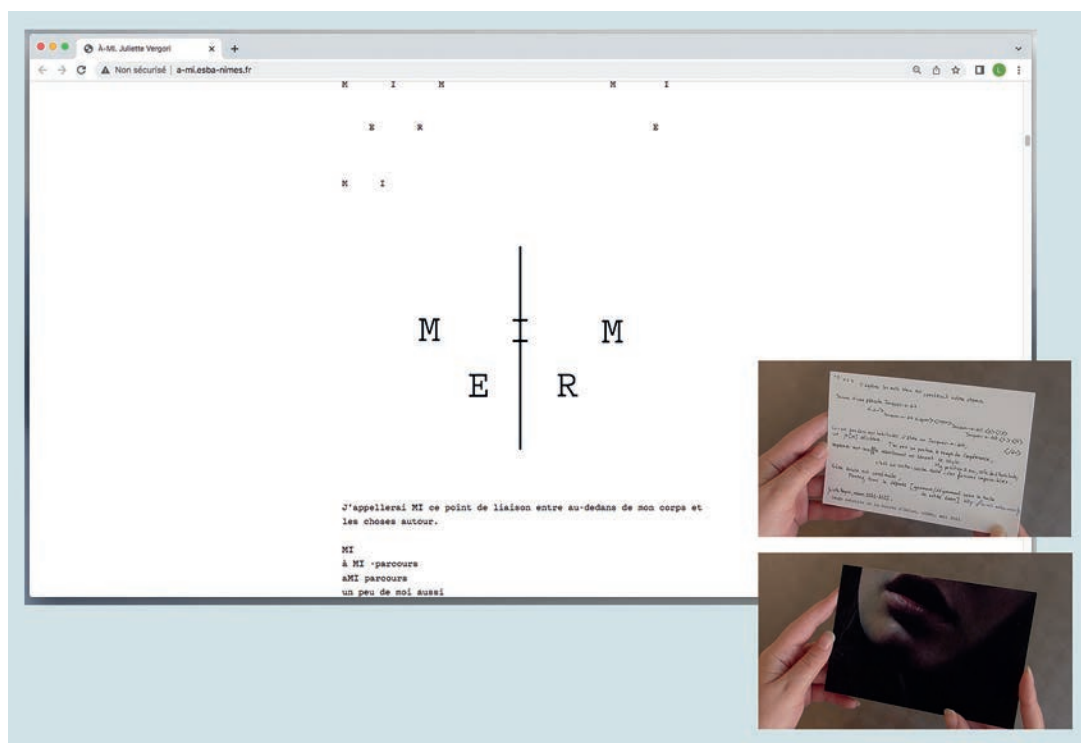
8. Voir dans ce numéro le focus de Marjorie Borios, p. 52.

Bibliographie

Sophie Chauvin, Gabriel Gallezot et Joachim Schöpfel, Eric Kergosien et Bernard Jacquemin, « Master's theses and open scholarship: a case study », *Digital Library Perspectives*, vol. 34, n° 4, 2018, p. 276-287. [10.1108/DLP-07-2018-0021](https://doi.org/10.1108/DLP-07-2018-0021)

Joachim Schöpfel, Sylvain Vanacker, Eric Kergosien et Bernard Jacquemin, « Master's theses and open scholarship: a case study », *Digital Library Perspectives*, vol. 34, n° 4, 2018, p. 276-287. [10.1108/DLP-07-2018-0021](https://doi.org/10.1108/DLP-07-2018-0021)

LA MISE EN LIGNE DES MÉMOIRES DU DIPLÔME NATIONAL SUPÉRIEUR D'EXPRESSION PLASTIQUE (DNSEP) DANS LE PORTAIL DUMAS, UN ATOUT POUR LA VALORISATION DE LA RECHERCHE EN ÉCOLE D'ART ET DE DESIGN



Juliette Vergori, À-Mi, 2022.

© Lexane Laplace et Juliette Vergori

MARJORIE BORIOS

Responsable de la bibliothèque et des ressources numériques de l'École supérieure des beaux-arts de Nîmes (ESBAN)

Le mémoire du DNSEP, un document hybride

Une des étapes dans l'obtention du DNSEP, conférant le grade de master, est la remise par les étudiants et les étudiantes d'écoles d'art et de design d'un mémoire de recherche en vue d'une soutenance. Ce document témoigne d'une pensée critique et problématise un questionnement lié au travail artistique. « La forme du mémoire constitue en elle-même un objet de recherche [...] : les formes expérimentales [...] sont fortement encouragées. L'écrit peut donc s'articuler avec des éléments sonores et visuels [...], des liens Internet, etc. L'écriture elle-même peut [...] mêler différents genres et registres, tels que l'enquête, l'expression poétique et fictionnelle, etc. L'articulation de ces différentes formes d'expression peut aboutir à une édition, une installation, une vidéo, un site Internet, etc.¹ »

« La forme du mémoire constitue en elle-même un objet de recherche [...] : les formes expérimentales [...] sont fortement encouragées. L'écrit peut donc s'articuler avec des éléments sonores et visuels [...], des liens Internet, etc. »

On retrouve donc, dans les bibliothèques de ces écoles, des mémoires aux formes et aux supports variés : un artefact en forme de pierre à briser pour accéder à la clé USB cachée à l'intérieur, une carte postale nous invitant à accéder à un site internet, un film, des podcasts, des pochettes sérigraphiées, des dépliants imprimés en riso², des objets réalisés avec une imprimante 3D, un plateau

de jeu, etc. Une version numérique est également stockée sur support informatique. Chaque école définit ses propres critères de conservation et de mise à disposition de ce fonds. Cependant, les bibliothécaires et documentalistes ont constaté, d'une part, des problèmes dans la conservation pérenne des versions numériques des mémoires ainsi que des fichiers vidéo, audio, etc. associés ;

1. ESBAN, *Livret de l'étudiant·e 2022-2023*, p. 42-43

2. Technique d'impression de copies multiples sur des papiers non couchés (c'est-à-dire brut).

et, d'autre part, un accès plutôt confidentiel à ces documents essentiellement consultés sur place par les étudiantes et étudiants.

DUMAS, un partenaire de choix pour valoriser la recherche en école d'art

Depuis une dizaine d'années, l'association des Bibliothèques d'écoles d'art en réseau (BEAR) propose aux écoles partenaires de signaler ces mémoires dans son catalogue documentaire, la Base spécialisée en art et design (BSAD)³. La BSAD recense, à ce jour, plus de 2 000 mémoires. En 2022, BEAR a mis en place un groupe de travail dédié exclusivement à l'évolution de la gestion et de la valorisation des mémoires du DNSEP. Ce groupe s'est vite intéressé aux possibilités d'accessibilité, de stockage pérenne et de valorisation qu'offre le portail DUMAS.

La collaboration avec Agnès Souchon⁴, administratrice de DUMAS, a entraîné une réflexion quant à nos besoins et aux solutions d'hébergement des mémoires aux formats hybrides sur cette plateforme institutionnelle. D'un côté, nous avons dans nos fonds des documents en version PDF, mais aussi des objets qui n'ont pas encore d'existence numérique; de l'autre, il

s'agit pour DUMAS de s'interroger sur le stockage et la mise en ligne de mémoires aux formats inhabituels (films, podcast, sites internet, etc.). Une éventuelle collaboration avec MédiHAL⁵ est aussi envisagée. Trois écoles se sont portées volontaires pour tester le dépôt des mémoires du DNSEP dans DUMAS. Une collection BEAR a été créée afin de centraliser tous les dépôts des écoles d'art et de design sur une page dédiée⁶: les écoles de Marseille, Nîmes et Rennes disposent également d'une collection propre.

L'école supérieure des beaux-arts de Nîmes (ESBAN), précurseur de la mise en ligne

Le projet, porté par la responsable de la bibliothèque de l'ESBAN, s'inscrit dans une dynamique de valorisation de la recherche par l'école et fait écho à la réalisation d'une plateforme numérique dédiée au second cycle⁷. Les équipes enseignantes ainsi que les étudiants et étudiantes sont sensibles aux questions d'archives ouvertes et se sont montrés très favorables à la mise en ligne des mémoires. Une fois l'accord de la direction du mémoire obtenu, chaque étudiant et étudiante devra dorénavant remplir une fiche descriptive (résumé, mots-clés etc.) et signer une autorisation de diffusion



© Lexane Laplace

Else Bedoux, *Six lignes de chant d'amour et une prière*, mémoire, ESBAN, 2022.

du mémoire. Depuis mars 2023, l'ESBAN dispose d'un espace dans DUMAS et a commencé à déposer des mémoires des diplômées 2022, sous licence libre CC-BY-NC-ND 4.0⁸.

Cet espace est appelé à s'enrichir au cours des prochains mois et les écoles suivront avec attention les évolutions à venir de DUMAS. De leur côté, elles devront imaginer des solutions pour rendre accessibles en ligne certains supports échappant au cadre traditionnel, par exemple en créant des vidéos ou en photographiant la manipulation de certains mémoires objets. ■

3. <https://www.bsad.eu/index.php>
4. Voir dans ce numéro l'article de Frédéric Gendre, Christophe Hugot, Joachim Schöpffel et Agnès Souchon, p. 49.
5. MédiHAL est une archive ouverte qui permet de déposer des données visuelles et sonores (images fixes, vidéos et sons), produites dans le cadre de la recherche scientifique : <https://media.hal.science/>
6. <https://dumas.ccsd.cnrs.fr/MEM-BEAR>
7. <https://esba-nimes.fr/cycle/home/>
8. <https://dumas.ccsd.cnrs.fr/MEM-ESBAN/>



© Lexane Laplace

Anaïs Roubertou, *Les langages des cabanes*, 2022.

Digital Library :

le Plan national de numérisation du patrimoine culturel du ministère de la Culture italien

L'Institut central pour la numérisation du patrimoine culturel, Digital Library, du ministère de la Culture italien réalise, coordonne et favorise le Plan national de numérisation du patrimoine culturel (PNN), qui incarne et diffuse la vision stratégique par laquelle le ministère veut promouvoir et organiser le processus de transformation numérique des institutions patrimoniales pour la période 2022-2026.

ANTONELLA NEGRI

Architecte, Institut central pour la numérisation du patrimoine culturel, ministère de la Culture italien (article traduit de l'italien)

1. <https://digitallibrary.cultura.gov.it/>
2. Intitulé MIC3 1.1.
3. <https://op.europa.eu/en/publication-detail/-/publication/90f1ee85-ca88-11ec-b6f4-01aa75ed71a1/language-en>
4. Voir dans ce numéro l'article de Pascal Liévaux et Livio De Luca, p. 28.

Archives photographiques de la Surintendance de l'architecture, des beaux-arts et du paysage de Gênes. Négatifs photographiques sur plaque de verre.

La Digital Library du ministère de la Culture italien

Pendant la crise de la pandémie de Covid-19, le numérique s'est affirmé comme un instrument essentiel de l'évolution des institutions et des lieux culturels italiens. La pandémie a surtout démontré que le numérique était une chance extraordinaire pour augmenter l'accessibilité aux biens culturels à davantage de publics, d'atteindre des cibles générationnelles et géographiques éloignées, et de tisser de nouvelles relations entre les biens culturels, les individus et les communautés. La réalisation de cet objectif a toutefois nécessité la mise en œuvre d'une stratégie susceptible d'harmoniser la dimension culturelle avec les aspects managériaux et techniques, mais surtout capable de déterminer et d'engager l'évolution des systèmes nécessaires dans l'environnement numérique.

C'est dans cette perspective que l'Institut central pour la numérisation du patrimoine culturel, nommé également Digital Library, a entamé la rédaction du Plan national de numérisation du patrimoine culturel.

La Digital Library¹ a été fondée en 2020 par le ministère de la Culture italien pour coordonner et promouvoir les programmes de numérisation du patrimoine culturel conservé par les institutions patrimoniales italiennes. À ce titre, et en s'appuyant sur des dispositifs techniques préalablement établis, elle a élaboré le Plan national de numérisation du patrimoine culturel et en supervise la mise en œuvre, tout en accompagnant les institutions et les lieux culturels dans leur transformation numérique. La Digital Library est chargée de définir, dans une logique d'écosystème, le cadre dans lequel chaque institution va concevoir sa propre trajectoire de changement, que celui-ci vise à l'amélioration de la gestion des biens conservés ou bien à la refonte des interactions avec le patrimoine culturel et sa valorisation.

Elle a, pour ce faire, bénéficié de moyens attribués par le Plan national de relance et de résilience italien (PNRR) qui soutient, notamment, la transition numérique des États et a été identifiée par le ministère de la Culture comme la structure responsable de la mise en œuvre de l'investissement « Stratégies et plateformes numériques pour le patrimoine culturel² », qui poursuit (conformément aux dernières initiatives de la Commission européenne, y compris la mesure spécifique pour la période 2023-2025³, concernant la création d'un *Cloud* européen collaboratif pour le patrimoine culturel⁴) l'objectif de créer un écosystème numérique de la culture reposant sur une infrastructure qui intègre et fédère la connaissance du patrimoine culturel et offre aux utilisateurs, des services d'accès innovants.



Le Plan national de numérisation du patrimoine culturel⁵

Le Plan national de numérisation du patrimoine culturel (PNN) incarne la vision stratégique avec laquelle le ministère entend promouvoir et organiser le processus de transformation numérique pendant les cinq prochaines années. Ce plan s'adresse en premier lieu aux musées, aux archives, aux bibliothèques, aux instituts et aux lieux culturels de l'État qui possèdent, gèrent et valorisent les biens culturels. Il crée un cadre méthodologique, intellectuel et professionnel de référence pour la réalisation des objectifs du PNRR.

Le Plan national de numérisation est également le résultat d'un processus de partage et d'échanges avec différentes institutions culturelles : il ne s'agit pas d'un document normatif, mais d'une référence méthodologique utile aux institutions et professionnels qui, aussi bien dans le domaine public que privé, se reconnaissent dans les valeurs qu'il énonce.

En raison de ses caractéristiques, il peut être défini comme un document « ouvert », parce qu'il est librement accessible⁶, « dynamique », parce qu'il est régulièrement mis à jour, « partagé », parce qu'il est le résultat d'un processus de participation à plusieurs niveaux et parce qu'il peut avoir plusieurs degrés d'application et de scalabilité⁷.

La structuration et le périmètre du PNN

Le PNN se divise en trois sections interconnectées : la vision qui préfigure la transformation et les opportunités du changement, en indiquant les objectifs à long terme ; la stratégie, qui définit la voie à suivre pour mettre en œuvre et atteindre les objectifs ; les lignes directrices, outils opérationnels qui soutiennent la planification et l'exécution des activités liées à la numérisation du patrimoine culturel.

Les lignes d'action stratégique représentent les intentions du ministère de la Culture pour initier, faciliter et accélérer les changements structurels que l'innovation technologique suscite dans le secteur culturel. Elles concernent trois domaines : les technologies permettant la mise en œuvre, les

Les lignes d'action stratégique représentent les intentions du ministère de la Culture pour initier, faciliter et accélérer les changements structurels que l'innovation technologique suscite dans le secteur culturel.

processus de travail et les personnes. Les technologies soutiennent les processus, eux-mêmes pilotés par les acteurs dans une perspective d'évolution consciente.

Le contexte technologique favorable

La première ligne d'action prévoit donc la création d'un contexte technologique favorable qui repose sur la conception d'une stratégie de gestion des données et des ressources numériques du patrimoine culturel, en conformité avec les directives nationales et européennes : les données sont le cœur de la transformation numérique. Pour cette raison, des actions ont été menées pour créer une infrastructure et des logiciels (respectivement *Hardware* et *Software Cloud*) à disposition de l'écosystème numérique de la culture. Il est nécessaire de réaliser cette infrastructure par la création de plateformes de service à l'échelle nationale, en incitant et en soutenant les développeurs et les entreprises pour la production de services numériques à diffuser sur le marché, capables d'associer le patrimoine culturel aux acteurs et aux industries créatives.

Les processus clés de la transformation numérique, conformément aux principes FAIR (« facile à trouver,

5. Pour une bibliographie exhaustive : <https://digitallibrary.cultura.gov.it/il-piano/>

6. <https://docs.italia.it/italia/icdp/>

7. La scalabilité est un terme employé dans le domaine de l'informatique matérielle et logicielle pour définir la faculté d'un produit informatique à s'adapter aux fluctuations de la demande en conservant ses différentes fonctionnalités : <https://partecipa.gov.it/processes/piano-nazionale-digitalizzazione-patrimonio-culturale/f/144/>

Archives d'État de Palerme. Registres cadastraux anciens.



© Ministère de la Culture italien



© Ministère de la Culture italien

Musée archéologique national d'Adria
(Direction régionale des musées
de la Vénétie).
Découvertes archéologiques
en céramique.

accessible, interopérable, réutilisable »), peuvent être identifiés, non seulement dans la numérisation et la gestion du cycle de vie de la ressource numérique, mais aussi dans les politiques actualisées d'accès, de valorisation et de réutilisation des reproductions numériques des biens culturels et dans la conception des services et des modèles pour la création de valeur culturelle, sociale et économique.

Mais pour que la transition numérique du patrimoine produise de la valeur culturelle et sociale, il est nécessaire que les personnes soient au centre du processus de changement : cela peut se faire par la formation et la mise à jour des compétences grâce à un programme d'apprentissage tout au long de la vie (*Lifelong Learning*), mais aussi et surtout par la construction d'un réseau d'échanges autour du patrimoine culturel, capable de stimuler la réutilisation des ressources numériques et les mécanismes de co-création et de production de valeurs.

Les outils et la méthodologie du Plan national de numérisation (PNN)

Les instruments associés au PNN indiquent et suggèrent des méthodologies et procédures pour la mise en œuvre des processus identifiés dans la stratégie que l'on peut facilement représenter avec le cycle de vie des données : « produire les données, gérer les données, faire circuler les données, créer des services à partir des données, évaluer la maturité numérique ».

Les lignes directrices se composent par conséquent de cinq documents :

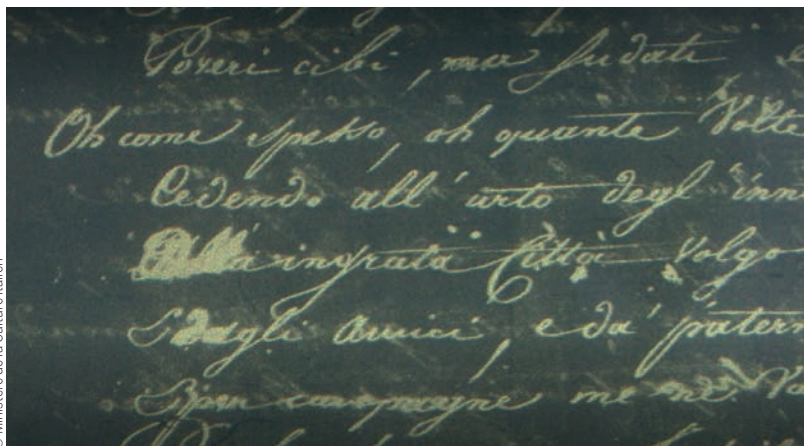
1. Les « lignes directrices pour la numérisation du patrimoine culturel » définissent les approches et les procédures pour la création, la réalisation de métadonnées et le stockage numérique des objets du patrimoine culturel analogique.
2. Les « lignes directrices pour la rédaction du plan de gestion des données (*Data Management Plan*) » décrivent et analysent de quelle manière stocker, partager et conserver les données issues des projets de numérisation du patrimoine culturel et des bases de données existantes. Elles sont accompagnées des « questions fréquemment posées⁸ » sur les données ouvertes.
3. Les « lignes directrices pour l'acquisition, la circulation et la réutilisation des reproductions des biens culturels dans le domaine numérique » encadrent l'environnement réglementaire et fournissent les outils opérationnels nécessaires pour identifier les cadres législatifs appropriés aux typologies de biens culturels et aux modalités de reproduction, d'acquisition d'images et de matériels audiovisuels, en lien avec les différents usages et les droits d'auteur (propriété intellectuelle et droit à la reproduction). L'objectif est de permettre aux institutions et aux utilisateurs de distinguer de manière explicite les limites et les possibilités de réutilisation des reproductions mises en ligne par les institutions.

8. Pour FAQ : *Frequently Asked Questions*.



© Ministère de la Culture italien

Archives d'État de Florence. Cartes anciennes.



© Ministère de la Culture italien

Bibliothèque centrale nationale de Rome, Centre national d'étude des manuscrits. Microfilm d'un manuscrit sur film négatif, visualisé au moyen d'une visionneuse rétroéclairée.

4. Les « lignes directrices pour la classification des produits et services numériques, des processus et modèles de gestion » identifient et décrivent les différentes taxonomies et cartographies des processus et des services précisés dans le PNN.
5. L'« introduction à la méthodologie pour l'évaluation de la maturité numérique des institutions culturelles » présente les modèles (*Digital Maturity Model*) qui peuvent être adoptés par les institutions culturelles en soulignant les possibilités d'application au patrimoine public et en proposant l'utilisation, à l'échelle nationale, de la méthodologie d'évaluation la plus appropriée pour permettre aux institutions de comprendre clairement leur niveau initial et de gérer plus efficacement leurs processus de transition numérique.

Une nouvelle notion de patrimoine culturel européen

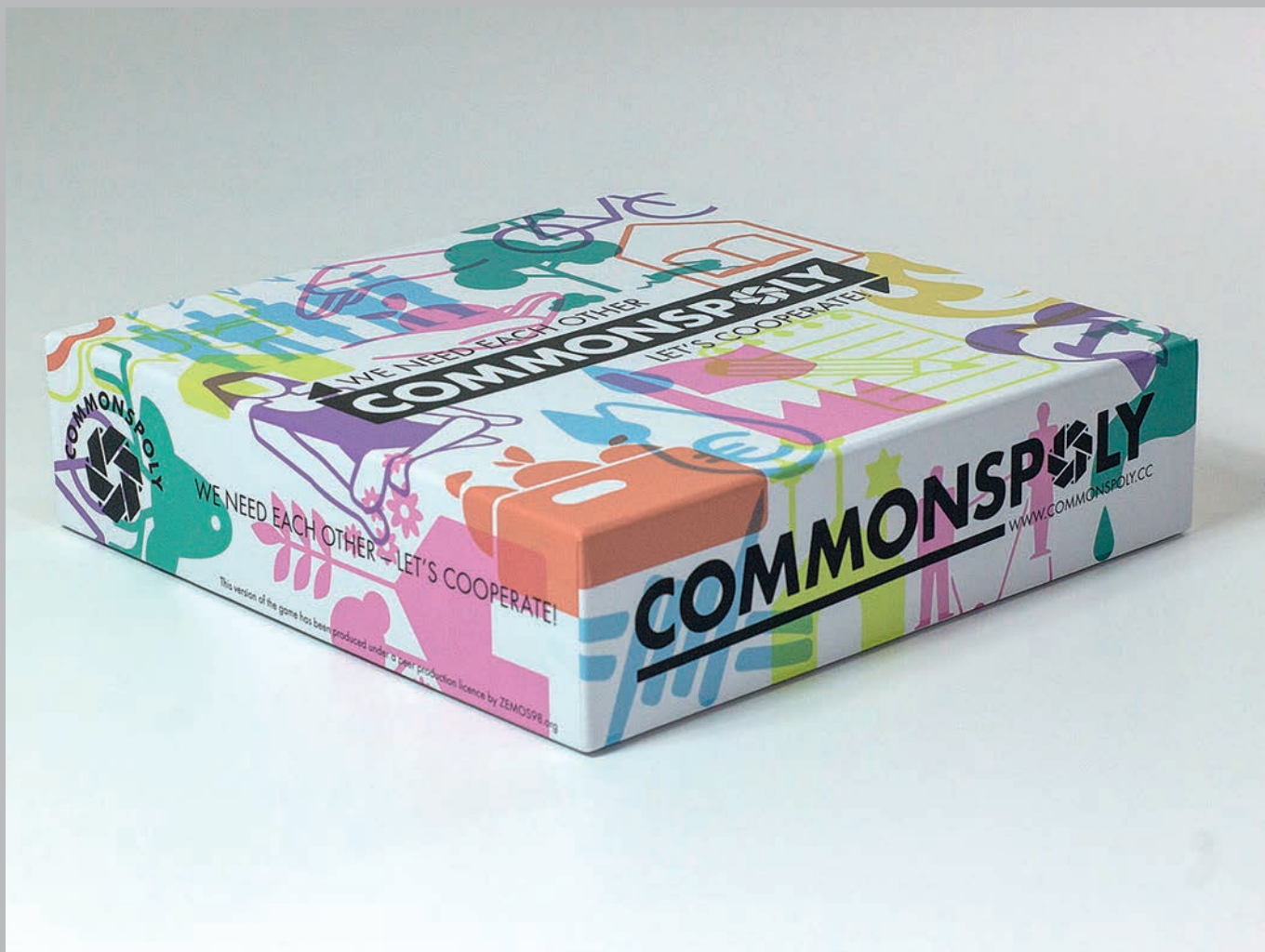
Ainsi, en replaçant la stratégie et les mesures du PNN dans le contexte européen et en se référant aux constatations de la Commission européenne sur le changement en cours dans les institutions culturelles, marqué notamment par l'accélération de la numérisation, par de nouveaux modèles de gouvernance stimulés par la science co-créative et par la participation d'un large éventail de parties prenantes, une nouvelle notion de patrimoine culturel européen « vert (durable), innovant (qui soutient l'emploi dans les Industries culturelles et créatives – ICC) et numérique émerge. Le patrimoine culturel numérique contemporain n'est pas seulement technologiquement

La nécessité de se doter d'un Plan national de numérisation du patrimoine culturel répond donc à l'exigence de pleine adhésion de l'Italie, en tant qu'État membre de l'Union européenne, au système de coopération culturelle européen.

prégnant, mais il ouvre aussi la voie à un nouveau monde numérique centré sur l'homme, où le *Cloud* collaboratif européen pour le patrimoine culturel jouera un rôle déterminant⁹ ».

La nécessité de se doter d'un Plan national de numérisation du patrimoine culturel répond donc à l'exigence de pleine adhésion de l'Italie, en tant qu'État membre de l'Union européenne, au système de coopération culturelle européen. La construction d'un patrimoine culturel numérique national, composé d'objets pouvant être localisés dans l'espace d'information de l'infosphère (entendu comme la totalité de l'espace d'information et de tout système capable d'interagir avec lui), peut représenter une opportunité pour l'Italie de retrouver et de jouer un rôle pertinent sur la scène culturelle mondiale, sans pour autant renoncer à sa tradition historique de protection des biens culturels. ■

9. Pere Brunet, Livio De Luca, Eero Hyvönen et al., *Report on a European collaborative cloud for cultural heritage. Ex-ante impact assessment*, European Commission, Directorate-General for Research and Innovation, 2022. <https://data.europa.eu/doi/10.2777/64014>



Jeu Commonspoly est un jeu de société sous licence libre probablement inspiré du Monopoly qui apprend aux joueurs à se familiariser avec le concept des communs

© Felipe González Gil, Creative Commons CC0 1.0 don universel au domaine public. https://commons.wikimedia.org/wiki/File:Commonspoly_Game_Box.jpg

**Un bien commun :
stocker
pour capitaliser**

Données culturelles et science ouverte : pratiques et enjeux du ministère de la Culture

Dans ses *Stratégies de recherche 2017-2020* et *2022-2026*, le ministère de la Culture affirme la diversité de ses pratiques visant à accroître les connaissances dans les domaines d'action scientifique du ministère. La transformation numérique est ainsi identifiée comme source de changements significatifs dans les pratiques de la recherche, du recueil des données à leur conservation et réutilisation, de l'établissement des sources aux outils utilisés pour les analyser et jusqu'au partage et à la diffusion de ses résultats.

ROMAIN DELASSUS

Chef du Service du numérique,
Secrétariat général, Service du
numérique, ministère de la Culture

AURÉLIEN CONRAUX

Administrateur ministériel des données
délégué, Secrétariat général, Service du
numérique, Département du numérique
pour la transformation des politiques
culturelles et de l'administration
des données, ministère de la Culture

Les données semblent toutefois offrir une promesse plus large encore que la recherche, au point que les poncifs « d'or noir » et de « pépites » reviennent régulièrement dans les rencontres de dirigeants d'établissements – souvent associés d'ailleurs à l'expression de difficultés pour extraire la valeur de ces ressources très diverses ou à la remise à plus tard de leur exploitation.

Nous souhaitons ici peindre la *veduta* des données culturelles en 2023, une vue topographique d'après nature et depuis le chevalet privilégié du Service du numérique.

Si l'on définit la « donnée » comme une information numérique pouvant faire l'objet d'un traitement informatique aisé, ou dans les termes officiels une « représentation d'une information sous une forme conventionnelle destinée à faciliter son traitement¹ », nous savons déjà que l'ensemble comprend dans les institutions culturelles à la fois des informations dans des applications (chiffres, calculs, textes, symboles) et des métadonnées (des informations descriptives d'autres informations numériques : l'auteur d'un fichier, la date de mise à jour...).

Dans les institutions culturelles, administration centrale, services déconcentrés ou opérateurs se trouvent en fait au moins cinq ensembles distincts :

- des données de gestion (données RH, financières, statistiques et indicateurs sur les processus administratifs...);
- des données d'usages (offres culturelles et données de fréquentation);
- des référentiels permettant de partager des éléments aussi bien techniques que scientifiques (les

thésaurus, ontologies et autres systèmes normalisés descriptifs des œuvres, des sites et des événements);

- des contenus culturels numérisés ou nativement numériques (reproductions d'archives, d'ouvrages conservés en bibliothèques, de tableaux, d'objets et de statues dans les musées...);
- des métadonnées, notamment les catalogues et inventaires, associés aux biens culturels.

Depuis plus d'une décennie, le ministère et ses opérateurs étendent progressivement le périmètre des données ouvertes, notamment à destination des chercheurs : publications par le ministère de guides sur les données ouvertes (*Open Data*), placement sous licence Etalab² des données bibliographiques de la Bibliothèque nationale de France (BnF)³, y compris celles de Gallica, ouverture en décembre 2016 de la plateforme de données ouvertes data.culture.gouv.fr⁴ permettant la mise en valeur des données du ministère (la liste des Monuments historiques, les aides et subventions...).

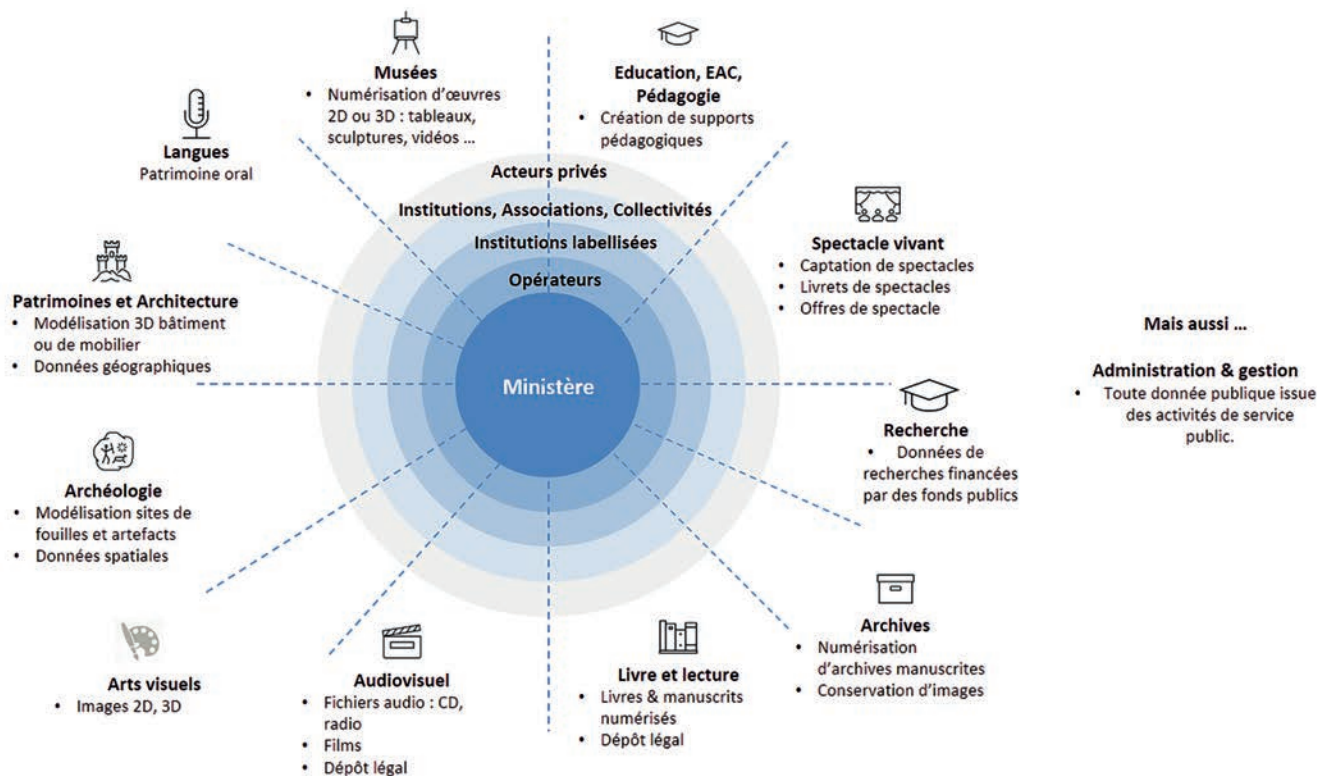
Fin 2022, l'ouverture d'un catalogue des données du ministère de la Culture (MICCADO) recensant plus largement les jeux de données existant dans les services a été suivie d'une consultation (close en février 2023) sur les attentes et priorités d'ouverture et de montée en qualité pour les données culturelles. La publication des jeux de données fondant la recherche interne (les données des enquêtes auprès des publics d'archives, sur site et en ligne, les sources non protégées des études du DEPS sur la statistique culturelle...) enrichit ainsi progressivement la plateforme.

1. Arrêté du 22 décembre 1981 sur l'enrichissement de la langue française.

2. Etalab est une licence libre française créée par la mission Etalab afin d'encadrer l'ouverture des données de l'État français.

3. Voir l'article d'Arnaud Laborderie et de Florence Tîbel dans ce numéro, p. 30.

4. Voir l'article de Roselyne Aliacar dans ce numéro, p. 73.



© Ministère de la Culture

La diversité des données culturelles.

Cette mission de partage et de diffusion, qui est le pendant pour les données de la mission ministérielle envers les œuvres⁵, devient pourtant une modalité de la politique générale des données du ministère, plutôt qu'une finalité en soi.

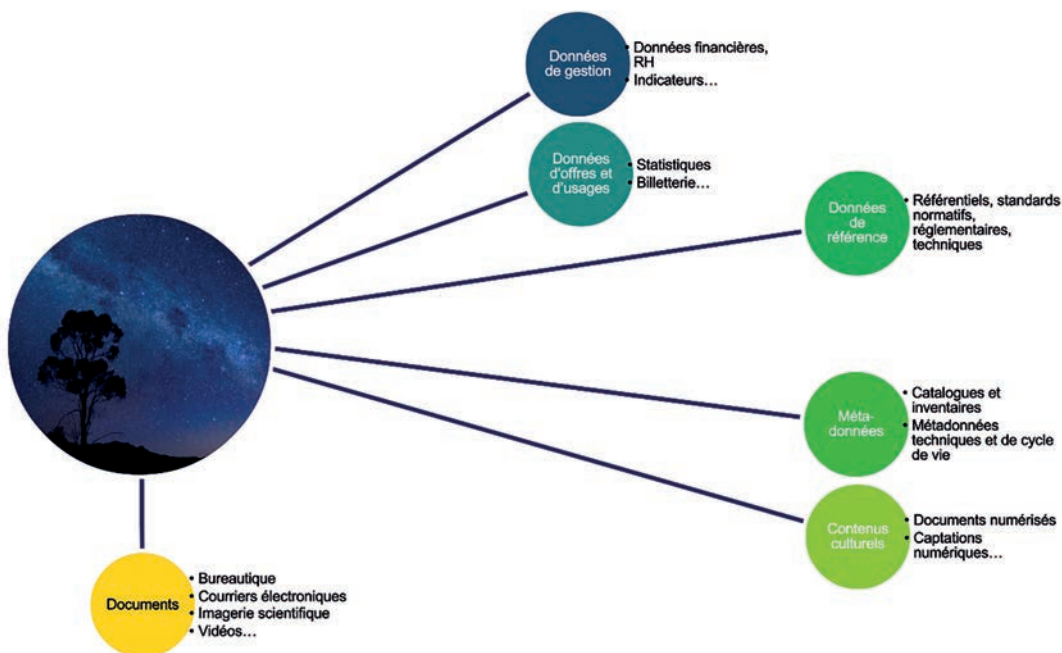
Avec la création du Service du numérique (SNUM) début 2021 et la publication en septembre de la même année d'une *Feuille de route des données et contenus culturels*, puis d'une *Stratégie numérique* en 2022, l'objectif principal associé aux données s'est en effet déplacé, pour s'élargir et placer les *Data* au cœur du système d'information.

L'explosion des pratiques numériques dans la société et dans l'administration, avec ses impacts

structuraux sur les processus de création, de diffusion et de relation avec les publics, ne permet plus de confiner des données dans des silos applicatifs, comme une collecte finale d'informations documentant les politiques publiques. Au contraire, pour que les institutions culturelles se saisissent, évaluent et maîtrisent ces nouveaux usages, les données doivent être rassemblées pour pouvoir être comparées, mises en relation et transcrites en informations porteuses de sens.

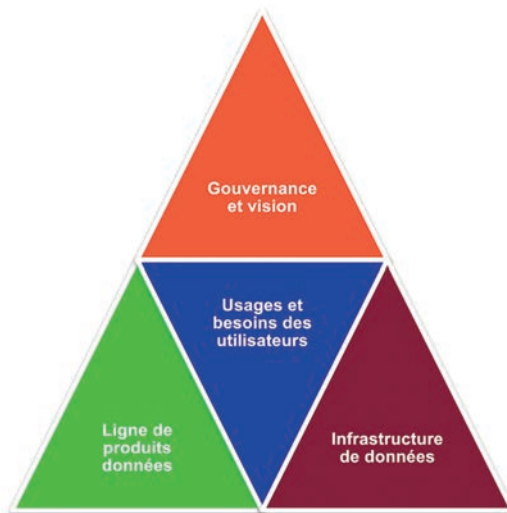
Là où la recherche vise souvent à présenter des jeux de données d'intérêt particulier, l'administration des données portée par le SNUM vise d'abord à établir un socle collectif, aussi large que possible, alimenté automatiquement avec des données nouvelles et doté

5. C'est même la première des missions confiées aux ministres de la Culture que de « rendre accessibles au plus grand nombre les œuvres capitales de l'humanité et d'abord de la France » (décret 2022-844 du 1^{er} juin 2022 relatif aux attributions du ministre de la Culture, reprenant à l'identique les précédents).



© Ministère de la Culture

Les catégories de données et de contenus numériques culturels.



La pyramide d'un écosystème « données » bâti autour des usages.

© Ministère de la Culture

des résultats peut révéler des pistes d'améliorations des politiques publiques.

L'activation des données et contenus culturels à des fins opérationnelles aussi variées que possible est donc le principe fondateur de la démarche d'administration ministérielle des données. Elle permet surtout de démultiplier les résultats puisque cette masse critique de données exploitables peut évidemment être utilisée pour la recherche, soit directement par des chercheurs extérieurs s'il s'agit de données ouvertes (*Open Data*), soit selon les procédures d'accès *ad hoc* si les jeux de données ne sont pas publics. L'enjeu, déjà mentionné dans le rapport Bothorel de 2020⁸, est alors de trouver le bon équilibre entre recherche et données parfois confidentielles, plus souvent loin d'être au niveau de qualité requis pour une exploitation sans explicitation du contexte de production.

Au bout du chemin, reste aussi pour le SNUM et les responsables informatiques des établissements culturels à réintégrer dans les descriptions d'œuvres la richesse des métadonnées développées par les chercheurs pour tel ou tel corpus. Pour ne prendre qu'un exemple, comment ajouter dans les catalogues (donc dans des métadonnées descriptives généralement sous licence Etalab) les milliers de localisations d'estampes réalisées par Johanna Daniel, chercheuse en train d'achever une thèse sur *Les vues d'optique, une production européenne d'estampes semi-fines (1740-1830)*⁹, avec un corpus dispersé entre des dizaines d'institutions européennes ? Et demain, comment associer responsables de collections, chercheurs et *Data scientists* pour décupler les informations grâce à l'analyse méthodologique par intelligence artificielle des collections nationales : indexer le contenu des livres et de la presse ancienne, pouvoir explorer les moindres détails des tableaux, se voir proposer des associations entre objets 3D, textes et documents multimédias ou audiovisuels ? ■

6. Une interface de programmation d'application (*Application Programming Interface*) est une grammaire technique par laquelle un logiciel peut être interrogé par d'autres, permettant une industrialisation des échanges fiables, à côté des requêtes humaines. Il s'agit ici de renvoyer soit des données de description soit des images à partir d'une requête envoyée par une machine.

7. Pour identifier ces jeux de données « à forte valeur ajoutée », le SNUM échange beaucoup avec les opérateurs, en bilatéral et au sein de communautés, mais a également organisé une consultation ouverte, du 6 décembre 2022 au 6 février 2023.

8. <https://www.gouvernement.fr/rapport/11979-rapport-sur-la-politique-publique-de-la-donnee-des-algorithmes-et-des-codes-sources>

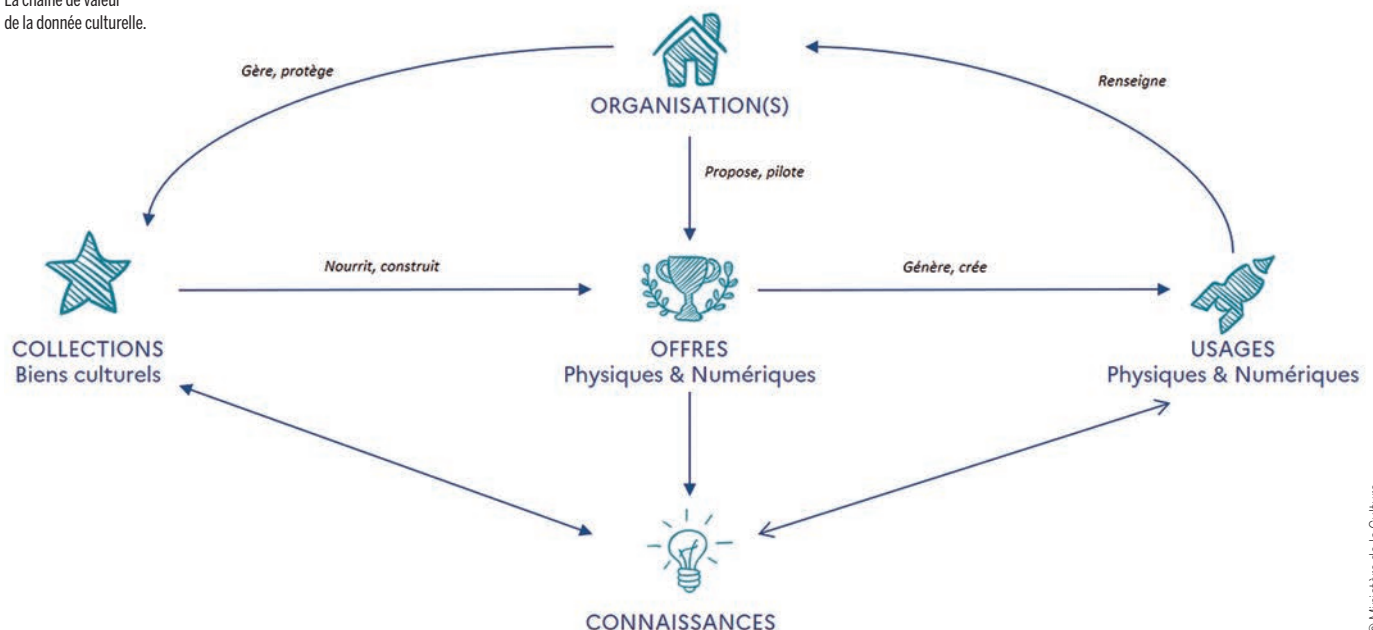
9. <https://ig.hypotheses.org/avis-propos-2/mon-projet-de-these>

de connecteurs (API⁶ notamment pour des interrogations de masse par des machines).

Depuis 2021, le SNUM construit ainsi avec ses partenaires à la fois des circuits de gouvernance stratégique et des processus permettant le traitement des données les plus utiles⁷ et leur montée en qualité, le tout appuyé sur une révision de l'infrastructure informatique du ministère.

La crise Covid a également fait apparaître avec clarté qu'une action publique appuyée sur des tableaux de bord calculés à partir de données de qualité est à la fois plus fiable et précise ; mieux adaptée aux réalités du terrain car la visualisation peut devenir dynamique, permettant par exemple le passage d'une carte de l'échelle nationale à la communauté de communes ; plus efficace car la fréquence de mise à jour de la donnée est connue et adaptée à l'enjeu ; productrice d'un meilleur service public. L'usage responsable et documenté des données devient la règle et le dialogue entre services producteurs des données et analystes

La chaîne de valeur de la donnée culturelle.



© Ministère de la Culture

Heurs et malheurs d'un programme européen :

le cas du projet Rurland (*Rural Landscape in North Eastern Roman Gaul*)

Lancé en 2014, le projet européen Rurland (*Rural Landscapes in North-Eastern Roman Gaul*) avait pour objectif d'étudier les campagnes de l'Antiquité sur un vaste espace géographique entre la Seine et le Rhin, de la fin du I^{er} siècle av. J.-C. au début du V^e siècle de notre ère. Il s'étendait donc sur différents pays modernes (France, Belgique, Pays-Bas, Luxembourg, Allemagne, Suisse) aux langues, aux traditions de recherche, aux pratiques administratives et patrimoniales différentes.

Son intention était de répondre à une question fondamentale de la recherche archéologique actuelle : comment réaliser une synthèse scientifique raisonnée de la masse d'informations qu'apportent aujourd'hui les nombreuses fouilles préventives nécessitées par le développement économique moderne ? Dans le même temps se déroulait Outre-Manche une expérience similaire, avec des objectifs identiques. Les deux projets ayant été publiés de manière quasi simultanée, il vaut la peine de les comparer, d'évaluer leurs résultats et d'en tirer, si faire se peut, quelques leçons pour l'avenir.

Un projet pionnier

Les programmes financés par l'ERC (*European Research Council*) sont rares car fortement sélectionnés. Mais ils présentent d'incontestables avantages, en particulier des financements importants qui favorisent un changement d'échelle, offrant aux chercheurs la possibilité d'aller très au-delà de leurs ambitions ordinaires. En l'espèce, le budget Rurland approchait du million d'euros pour quatre ans. Mais cette manne financière n'est pas accordée, évidemment, sans contrôle ni contrepartie. En l'occurrence, elle n'est pas distribuée à une ou plusieurs équipes mais à un chef de projet (PI ou *Principal Investigator*), seul responsable scientifique sous la garantie financière d'une institution de recherche qui dispose d'un compte public (CNRS/universités). L'autre obligation,

tout aussi normale, réside dans une large diffusion des acquis à travers les médias scientifiques classiques ; progressivement, la mise en ligne avec accès libre des bases de données créées dans ce cadre scientifique et des résultats de la recherche est ainsi devenue une obligation contractuelle.

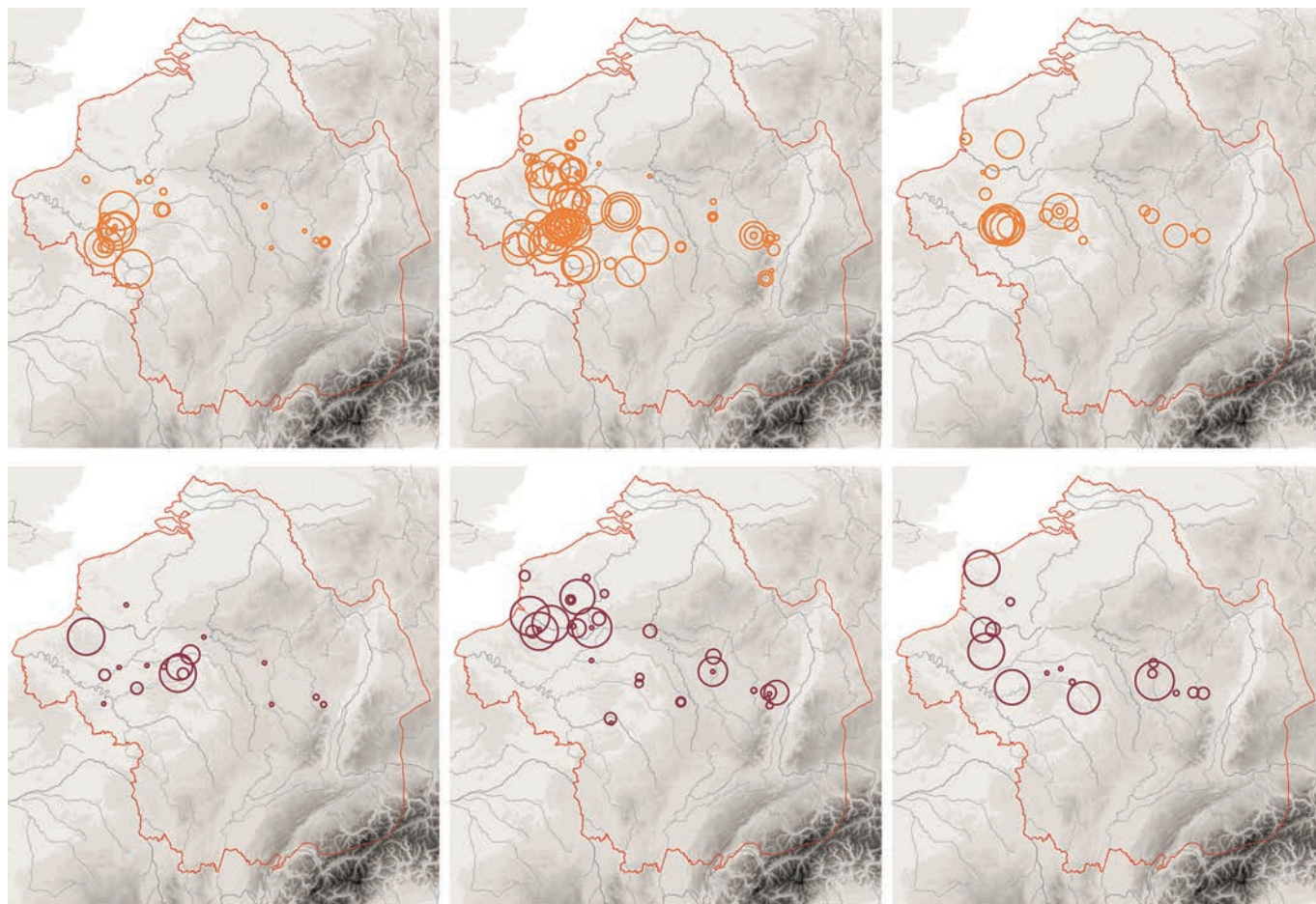
Le projet Rurland avait initialement prévu d'asseoir sa recherche sur l'exploitation des cartes archéologiques des différents pays concernés, en finançant, à ce titre, un ingénieur de recherche chargé de créer et de gérer une base homogène centralisée à partir d'un Système d'information géographique (SIG). Mais c'est là que les difficultés ont commencé. À l'échelle européenne, l'incompatibilité des différents systèmes entre eux – quand ils étaient informatisés et quand leur accès était possible et immédiat – a constitué un obstacle infranchissable. En France même, l'état réel de la carte archéologique et, tout simplement, ses possibilités réduites d'accès, notamment au niveau central, n'ont pas permis une compilation rapide et efficace des données.

Dans un petit nombre de régions, toutefois, l'investissement personnel du conservateur régional a permis de lever les principaux obstacles. Enfin, à l'époque où débutait ce projet, la base Dolia de l'Institut national de recherches archéologiques préventives (INRAP)¹ était encore balbutiante et n'a pu apporter qu'une contribution limitée. À ce propos, il est très intéressant de comparer cette situation à celle de nos voisins

MICHEL REDDÉ

Directeur d'études émérite à l'École pratique des hautes études (EPHE)-Université Paris Sciences & Lettres (PSL), UMR 8210 Anthropologie et histoire des mondes antiques (ANHIMA)

1. Voir dans ce numéro l'article de Marc Bouiron, Emmanuelle Bryas, Kai Salas-Rossenbach, p. 47.



© V. Zech-Matterne pour Ruriland

blés nus épeautre

Évolution des proportions relatives de blés nus (en haut) et de blé épeautre (en bas). Nombres minimum d'individus (NMI) moyens calculés à l'échelle de chacun des sites pour l'ensemble des structures ayant livré des restes carpologiques. Sites de plus de 100 restes d'espèces domestiques, à l'exception des céréales et légumineuses indéterminées. Phase 1 (à gauche) : début I^{er} siècle-30 av. J.-C. ; phase 2 (au centre) : 30 av. J.-C.-275 ap. J.-C. ; phase 3 (à droite) : dernier quart du III^e siècle-I^{er} siècle ap. J.-C.

britanniques investis dans une recherche similaire sur les campagnes de la Bretagne romaine. Ces derniers ont pu en effet exploiter, dès la fin des années 2000, une masse sélectionnée de 2 575 dossiers de fouilles préventives, numérisés par les services patrimoniaux compétents ou déjà en ligne, offrant ainsi aux chercheurs la possibilité de classer et de cartographier aisément les données, après normalisation. Cet effort collectif, à l'échelle d'un programme national, a été évidemment particulièrement efficace et productif, la base de données ainsi créée reste disponible et utilisable.

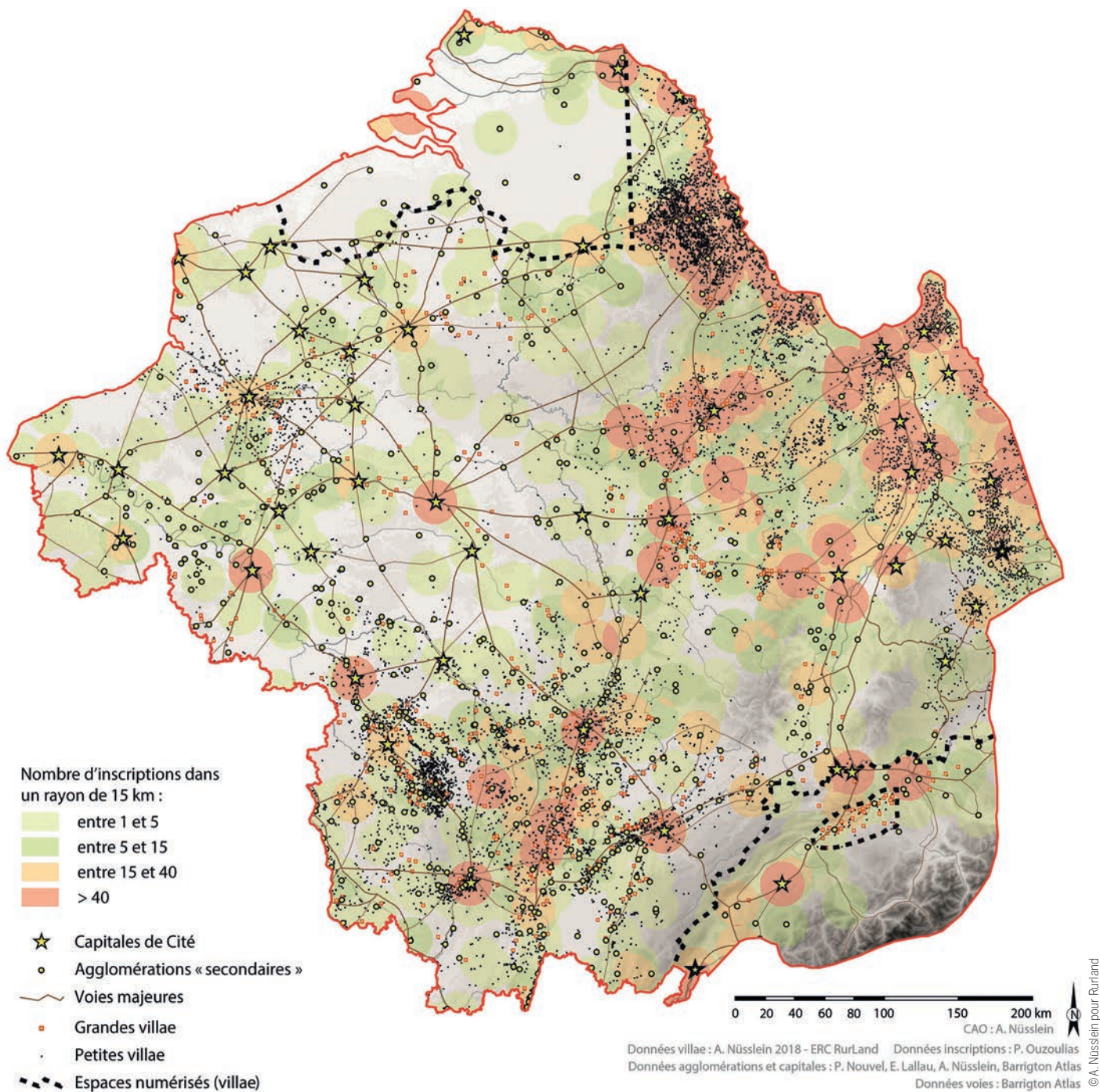
Des acquis essentiels

Le programme Ruriland a donc fonctionné « à l'ancienne », en collationnant les données disponibles selon deux méthodes : 1) la consultation sur place des rapports de fouilles (version « papier »), ce qui a mobilisé un post-doctorat pendant deux ans ; il s'agit là d'une tâche ingrate, longue, onéreuse ; 2) l'organisation de séminaires mensuels d'une ou deux journées réunissant, à l'INHA, les chercheurs intéressés et disponibles. Il faut saluer ici l'enthousiasme manifesté par ces derniers, leur apport scientifique et le soutien très réel de l'INRAP qui a permis à ses collaborateurs de participer à ce projet. Une partie de ces séminaires a été mise en ligne et reste disponible sur un blog dédié²,

mais l'essentiel des recherches a été publié de manière « classique » dans deux monographies successives, depuis lors disponibles librement sur le portail HAL de l'éditeur. Ces dernières intègrent naturellement les contributions des différents chercheurs européens qui ont participé au projet.

Si l'on tente de dresser un bilan de ces quatre années de recherche, on peut sans doute souligner plusieurs aspects non négligeables du point de vue scientifique : on citera ici, en particulier, la collaboration étroite entre spécialistes de l'âge du fer, de la période romaine et de l'Antiquité tardive, que la tradition académique a trop tendance à séparer, mais aussi la synergie apportée par l'apport des différentes disciplines investies dans l'archéologie de terrain (sciences de la terre, carpologie, zoologie...), indispensables pour mettre aujourd'hui en évidence le fonctionnement des systèmes agro-pastoraux anciens. De ce point de vue, Ruriland a apporté un saut qualitatif significatif dans notre compréhension du monde rural de la Gaule et a permis, au moins partiellement, de dépasser les frontières administratives de l'hexagone. Il est hautement souhaitable que la recherche soit désormais étendue au-delà de la zone initialement dévolue à ce projet. On notera au passage avec satisfaction que, malgré les obstacles rencontrés, les résultats scientifiques sont cohérents avec ceux que le programme britannique

2. <https://ruriland.hypotheses.org/>



Cartographie des zones de développement dans le nord-est de la Gaule.

a mis en évidence. Faute de place, on n'en donnera pas ici de résumé détaillé, mais le lecteur intéressé en trouvera une synthèse commode dans un récent article des *Annales. Histoire, sciences sociales* (voir Bibliographie). On y verra par exemple une illustration du dynamisme économique très différencié au sein de la zone étudiée et l'attraction exercée à la fois par le réseau urbain et la frontière romaine de Germanie.

Les mécomptes rencontrés au démarrage du projet devraient en revanche inviter à une réflexion collective sur l'accès effectif à la carte archéologique, son utilisation, ses fonctionnalités, les éléments qu'elle contient³. ■

Bibliographie

Michel Reddé, « Le projet "Ruriland" », dans Michel Reddé (dir.), *Gallia Rustica 1. Les campagnes du nord-est de la Gaule, de la fin de l'âge du fer à l'Antiquité tardive*, Ausonius, coll. « Mémoires », 2017. hal-03029620

Michel Reddé (dir.), *Gallia Rustica 2. Les campagnes du nord-est de la Gaule, de la fin de l'âge du fer à l'Antiquité tardive*, Ausonius, coll. « Mémoires », 2018. hal-03225681

Michel Reddé, « Le développement économique des campagnes romaines dans le nord de la Gaule et l'île de Bretagne. Des approches renouvelées », *Annales. Histoire, sciences sociales*, 2022-1, p. 105-145. <https://www.cairn.info/revue-Annales-2022-1-page-105.htm>

3. Voir dans ce numéro l'article de Jean-Manuel Conilleau, Christian Cribellier, Stéphane Deschamps, Cynthia Domenech-Jaulneau, Thierry Lorho, Yves Menez, Martial Monteil et Marie-Alice Virlovet, p. 44.

La Cellule Data Grenoble Alpes au service des communautés scientifiques de Grenoble pour la gestion de leurs données de recherche

Face à l'augmentation notable des demandes autour de la gestion des données de recherche depuis quelques années, une cellule d'accompagnement, la Cellule Data Grenoble Alpes (CDGA)¹, a été mise en place, structure transverse pilotée par l'Unité d'appui et de recherche (UAR) GRICAD – Grenoble Alpes recherche infrastructure de calcul intensif et de données, centre de calcul et de données, et la BAPSO (Bibliothèque universitaire, service de soutien à la recherche), dans un contexte institutionnel local favorable à la science ouverte.

LUCIE ALBARET

Chargée de mission science ouverte de l'Université Grenoble Alpes

VIOLAINE LOUVET

Ingénieure de recherche au CNRS, administratrice des données de recherche de l'Université Grenoble Alpes, Laboratoire Jean Kunztmann (UMR 5224 CNRS)

Cette cellule regroupe des personnes issues de ces deux organisations, des laboratoires et établissements du site (UGA, CNRS, Commissariat à l'énergie atomique et aux énergies alternatives [CEA], Institut national de recherche en informatique et en automatique [INRIA]) et de leurs services d'appui, avec des compétences, des métiers et des implications différents (informaticiens, chercheurs, bibliothécaires, documentalistes, juristes, délégués à la protection des données).

Les laboratoires de l'École nationale supérieure d'architecture (ENSA) Grenoble participent au réseau de référents « données », qui permet des échanges sur le thème de la gestion des données de la recherche, entre la Cellule Data et les chercheurs et ingénieurs du site, mais aussi entre laboratoires du site.

Cette complémentarité est à la fois un atout (un enrichissement) et un défi : se comprendre, fédérer les différentes approches et points de vue, organiser l'accompagnement sont des enjeux importants.

L'Université Grenoble Alpes (UGA) est une université pluridisciplinaire comptant près de 60 000 étudiants. Plus de 3 000 enseignants-chercheurs et presque 3 000 doctorants travaillent au sein d'une centaine de laboratoires. Dans un contexte de recherche extrêmement dynamique (l'UGA fait partie de l'association UDICE² qui réunit 10 universités françaises), l'impact de la science ouverte, l'explosion des volumes de données collectées ou disponibles, mais aussi les transformations liées au numérique et à l'intelligence artificielle bousculent les méthodes de travail tout en apportant de nouvelles opportunités à l'ensemble des communautés de recherche.

Accompagner les changements de pratiques

Afin de faire face aux sollicitations croissantes des scientifiques dans tous ces domaines, a naturellement émergé une structure opérationnelle organisée en réseau et rassemblant des personnes d'horizons différents apportant des compétences complémentaires : la Cellule Data Grenoble Alpes (CDGA).

Les laboratoires de l'ENSA Grenoble participent au réseau de référents « données », qui permet des échanges sur le thème de la gestion des données de la recherche, entre la Cellule Data et les chercheurs et ingénieurs du site, mais aussi entre laboratoires du site.

1. <https://scienceouverte.univ-grenoble-alpes.fr/donnees/cellule-data/>

2. <https://www.udice.org/>



La CDGA repose sur un socle cœur constitué d'informaticiens issus de l'unité GRICAD³ et de bibliothécaires issus des Bibliothèques et appui à la science ouverte (BAPSO⁴). Ce premier noyau, qui inclut des personnes à temps plein sur la cellule, est renforcé par des membres de laboratoires et de services d'appui (délégue à la protection des données, Service d'expérimentation et de développement, Institut national de recherche en sciences et technologies du numérique-INRIA-Grenoble). La CDGA compte à ce jour 24 personnes, issues de beaucoup des organismes et établissements de recherche présents sur le site : Université Grenoble Alpes, CNRS, Commissariat à l'énergie atomique et aux énergies alternatives (CEA), INRIA.

Cette approche par la complémentarité des expertises est tout à la fois essentielle pour répondre de façon pertinente aux questionnements très variés des membres des laboratoires, et particulièrement enrichissante pour les membres de la cellule.

La CDGA a également construit des liens étroits avec différents services de l'université (autour de la valorisation ou du juridique par exemple).

Au-delà du périmètre local, ses activités s'inscrivent dans l'écosystème national qui se construit autour des données et s'appuient sur les réseaux de professionnels existants.

Assurer la proximité avec les laboratoires, notamment ceux des Écoles nationales supérieures d'architecture (ENSA)

Ce groupe opérationnel au service des communautés de recherche s'appuie sur un réseau de référents « données » présents dans la plupart des laboratoires du site et assurant un relais efficace et de proximité auprès de leurs collègues. Leur rôle est essentiel pour garantir la justesse de l'accompagnement proposé et son évolution en cohérence avec les besoins des utilisateurs.

Depuis la création de la CDGA en 2019, les services à la donnée ont évolué et se sont largement étoffés⁵ sur un périmètre plus large intégrant la réalité du terrain : outre la question des données, celle des logiciels est également au cœur des préoccupations, et le *continuum* entre publications, données et logiciels doit être couvert. La CDGA travaille ainsi de façon très étroite avec la cellule HAL UGA sur les publications ouvertes.

Les sciences humaines et sociales (SHS) au cœur du processus

Toutes les communautés scientifiques sont accompagnées par la CDGA, en particulier les laboratoires en SHS, comprenant les laboratoires de l'ENSA Grenoble.

La thématique des données en SHS devrait connaître dans les prochaines années une forte accélération grâce au projet Excellence GATES (*Grenoble Attractiveness and Excellences*)⁶ présenté par l'université et soutenu par le ministère de l'Enseignement supérieur et de la Recherche.

En parallèle, un travail de recensement des développements logiciels de ces laboratoires aura lieu cette année afin de valoriser cette production scientifique bien souvent ignorée pour ces thématiques scientifiques.

Dessiner des perspectives en s'appuyant sur un socle solide

La Cellule Data Grenoble Alpes existe depuis maintenant 4 ans et le recul sur ces années de fonctionnement permet de bien identifier ses atouts, que sont la complémentarité des expertises et l'envie de partager ses compétences et de travailler ensemble, qui ont permis de développer une offre de service large et qui commence à être connue par les chercheurs. Elle a gagné également une certaine visibilité à l'échelle nationale, grâce à la labellisation « atelier de la donnée⁷ » de l'initiative Recherche Data Govv⁸.

La CDGA est à un tournant de sa structuration, et doit relever des défis majeurs comme sa pérennisation dans la durée, l'intégration des ingénieurs du projet GATES, le développement du lien avec l'ensemble des disciplines et la veille pour faire évoluer les services et les compétences de ces membres.

Et le véritable défi reste la construction d'une culture commune autour des données et des codes en lien avec les publications.

L'expansion de la CDGA s'est faite en parallèle du développement d'une politique volontariste⁹ de l'université autour de la science ouverte, à travers l'adoption de textes importants comme la charte science ouverte et le schéma directeur de la science ouverte qui explicitent l'ambition de l'UGA sur ces sujets. Les actions de l'opérationnel et du politique sont particulièrement cohérentes et s'enrichissent mutuellement, permettant de lancer la CDGA sur des rails durables. ■

3. <https://gricad.univ-grenoble-alpes.fr/>

4. <https://bibliotheques.univ-grenoble-alpes.fr/appui-aux-chercheurs/guichet-de-la-science-ouverte/>

5. <https://scienceouverte.univ-grenoble-alpes.fr/donnees/cellule-data/offre-de-services/>

6. <https://www.univ-grenoble-alpes.fr/actualites/a-la-une/actualites-universite/le-projet-gates-attractivite-et-excellences-a-l-universite-grenoble-alpes-labellise-dans-le-cadre-de-l-appel-a-projets-excellences-de-france-2030-1102335.kjsp>

7. <https://recherche.data.gouv.fr/fr/page/ateliers-de-la-donnee-des-services-generalistes-sur-tout-le-territoire>

8. <https://recherche.data.gouv.fr/fr/>; voir dans ce numéro l'article de Roselyne Aliacar, p. 73.

9. <https://scienceouverte.univ-grenoble-alpes.fr/science-ouverte-grenoble-alpes/les-engagements-de-luga/>

ESPADON :

renforcer l'écosystème instrumental et numérique des sciences du patrimoine

Financé de 2021 à 2029 en tant qu'équipement d'excellence (EquipEx) Patrimex dans le cadre du Plan d'investissement d'avenir (PIA 3) de l'Agence nationale de la recherche, le projet ESPADON (en sciences du patrimoine, l'analyse dynamique des objets anciens et numériques) vise à enrichir le domaine interdisciplinaire des sciences du patrimoine sur le plan instrumental et numérique.

VINCENT DETALLE

Professeur, CY Cergy Paris Université,
coordinateur consortium ESPADON

ROMAIN THOMAS

Maître de conférences en histoire de
l'art moderne, Université Paris Nanterre,
coordinateur adjoint consortium
ESPADON

Il est porté par la Fondation des sciences du patrimoine (FSP), qui réunit depuis une décennie un grand nombre d'institutions culturelles françaises et de laboratoires de recherche dédiés à l'étude et à la conservation du patrimoine matériel.

Au cours de la même période, les États européens et les institutions patrimoniales se sont engagés dans une transformation numérique majeure. Si l'objectif de faciliter l'utilisation des ressources à l'échelle européenne dans un contexte multiculturel et multilingue est largement atteint, l'archivage numérique des données de la recherche et la dématérialisation du patrimoine culturel doivent être constamment revisités en raison des nouvelles évolutions technologiques, des nouveaux paradigmes et des nouveaux modèles économiques¹.

Un projet pour les sciences du patrimoine

Afin d'être aussi inclusif que possible, ESPADON a été construit en rassemblant un certain nombre d'institutions éducatives et culturelles et d'unités de recherche touchant au champ des sciences du patrimoine. Ce dernier peut être décrit à l'aide d'un schéma conceptuel centré autour des objets culturels.

Les principales approches s'articulent autour de trois axes :

- les questions relatives à l'étude synchronique de l'objet, c'est-à-dire l'objet considéré dans son état actuel : caractérisation matérielle actuelle de l'objet et questions de conservation ;
- la recherche diachronique sur l'objet, c'est-à-dire l'histoire de l'objet de sa création à aujourd'hui, à différentes échelles (microscopique à macroscopique) ;
- la gestion des données relatives à l'objet (de la production jusqu'à leur dissémination envers les publics les plus larges).

Une diversité d'acteurs, qu'il s'agisse de chercheurs ou de professionnels, d'individus ou d'institutions, contribuent à ces questions en interdisciplinarité (par exemple anthropologues, archéologues, historiens, historiens de l'art, *Data Scientists*, juristes [du patrimoine culturel], physico-chimistes, physiciens, etc. ; mais aussi des professionnels comme les archivistes, conservateurs, bibliothécaires, médiateurs, etc.).

Les objectifs stratégiques d'ESPADON : instrumentaux et informationnels

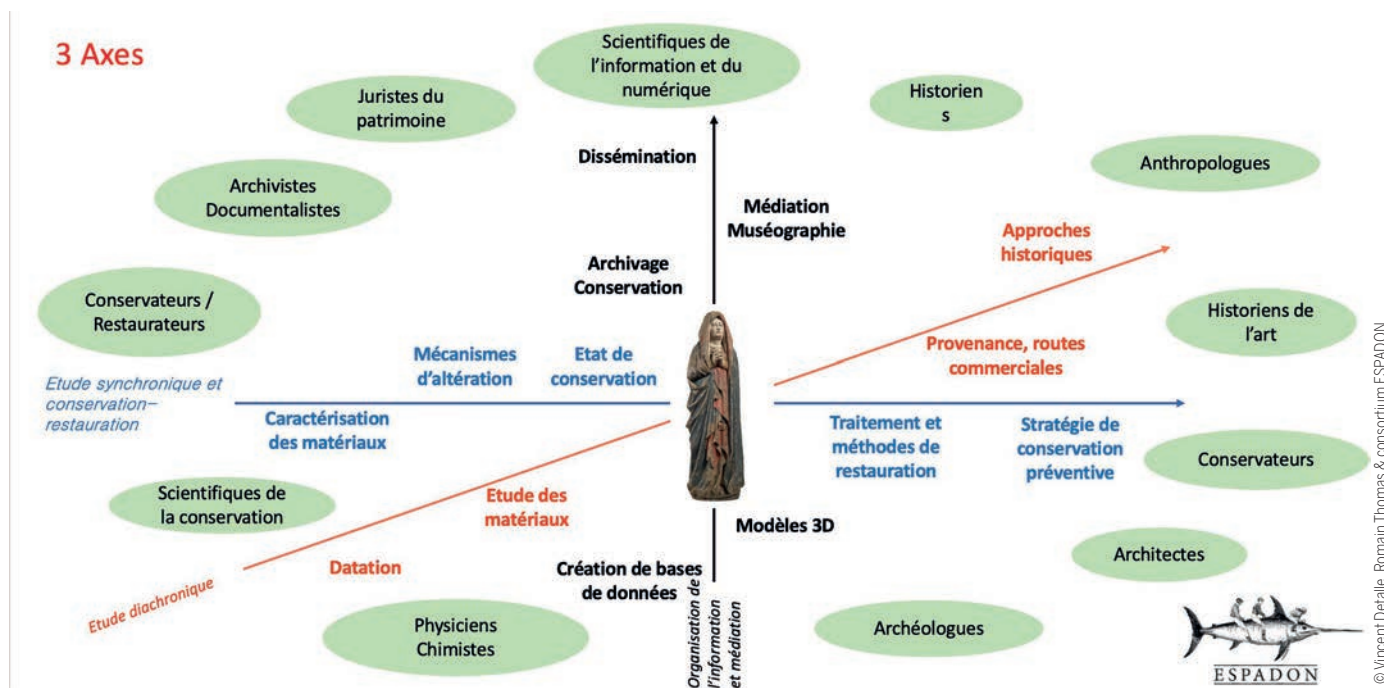
Le premier objectif d'ESPADON est de relever un défi instrumental. S'il s'inscrit dans la continuité de l'EquipEx Patrimex, les évolutions importantes apparues au cours de la dernière décennie permettent aujourd'hui d'améliorer la caractérisation d'un objet en passant de techniques d'analyse ponctuelle à une approche analytique en 2D, 2,5D (surface + épaisseur) ou 3D. Le temps est venu d'assurer une telle mise en œuvre pour les objets du patrimoine, en mettant à la disposition de la communauté de nouveaux moyens instrumentaux de tomographie 3D multi-échelles et d'imagerie physico-chimique 2D. En couplant l'instrumentation analytique et structurale, on peut assurer des capacités d'analyse multi-dimensionnelle, multi-échelles et multi-temporelle afin d'obtenir une approche multimodale.

Du point de vue de l'information numérique, l'ambition d'ESPADON est de relever le défi du traitement, du stockage et de l'échange de données massives en faisant émerger un écosystème numérique interopérable. Il s'inscrit ici dans la continuité des actions Parcours et SoCoRe! (Science ouverte pour la conservation/restauration du patrimoine culturel)².

Les actions seront naturellement articulées avec la transition numérique qui s'opère dans les différentes institutions à l'échelle nationale. Elles fédéreront les

1. Voir dans ce numéro l'article de Livio De Luca et Pascal Liévaux, p. 28.

2. <http://www.sciences-patrimoine.org/projet/parcours-patrimoine-culturel-et-restauration-conservation-ontologie-pour-lusage-dun-referentiel-commun-aux-differentes-sources-de-donnees/> ; <http://www.sciences-patrimoine.org/projet/socore/>



communautés académiques et professionnelles en France et seront l'occasion de mettre en place des actions de formation. Au niveau européen, le projet s'inscrit dans la perspective de l'initiative E-RIHS DIGILAB³. ESPADON permettra donc à la communauté des sciences du patrimoine d'améliorer considérablement l'accès à l'information.

Comparé à d'autres domaines de recherche, le patrimoine culturel est spécifique en raison de la variété (en termes de matériaux, de formes, d'échelle, etc.) des objets étudiés et de leur unicité. En outre, les pratiques institutionnelles sont également très diverses, ce qui rend plus difficile (mais nécessaire) l'objectif de collecter toutes sortes d'informations ou de connaissances générées autour d'un même objet patrimonial (objet matériel). Le futur écosystème numérique permettra la représentation de ces informations et l'accès à ces informations dans un espace de données unique, basé sur une architecture de stockage distribuée et interopérable.

Vers un modèle d'« objet patrimonial augmenté »

À cette fin, ESPADON vise à développer un modèle d'« objet patrimonial augmenté ». L'écosystème ne sera pas un simple agrégateur, mais bien un nouveau média pour les sciences du patrimoine, partagé par la communauté de recherche des sciences du patrimoine ainsi que par les professionnels. Les mondes éducatif et socio-économique auront accès à une partie de ces informations selon les règles juridiques idoines.

Les « objets du patrimoine » considérés sont très divers, allant des artefacts de musée aux Monuments historiques en passant par les sites archéologiques. Leur patrimonialisation implique l'existence d'une documentation forgée par des professionnels de la conservation, des archives, mais aussi les chercheurs des divers domaines disciplinaires. Afin d'approfondir la connaissance de ces objets, ils peuvent faire l'objet de nouvelles enquêtes à propos, par exemple,

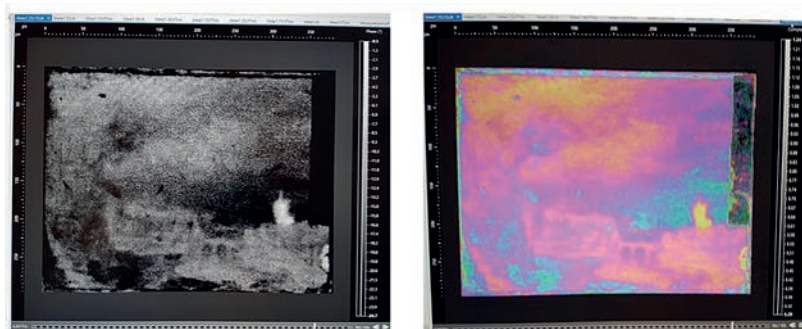
des phénomènes élémentaires, moléculaires ou structurels, des données climatiques dans un contexte global en évolution; ou encore relevant des sciences humaines et sociales. Les méthodologies appliquées livreront un ensemble de données très hétérogènes relatives à l'objet, qu'il s'agira alors de lier, d'organiser et de visualiser.

ESPADON vise ainsi à offrir une plateforme de services numériques assurant la numérisation, la sémantisation et l'interopérabilité des données. Une telle plateforme permettra, par exemple, la comparaison, la classification et l'apprentissage automatique.

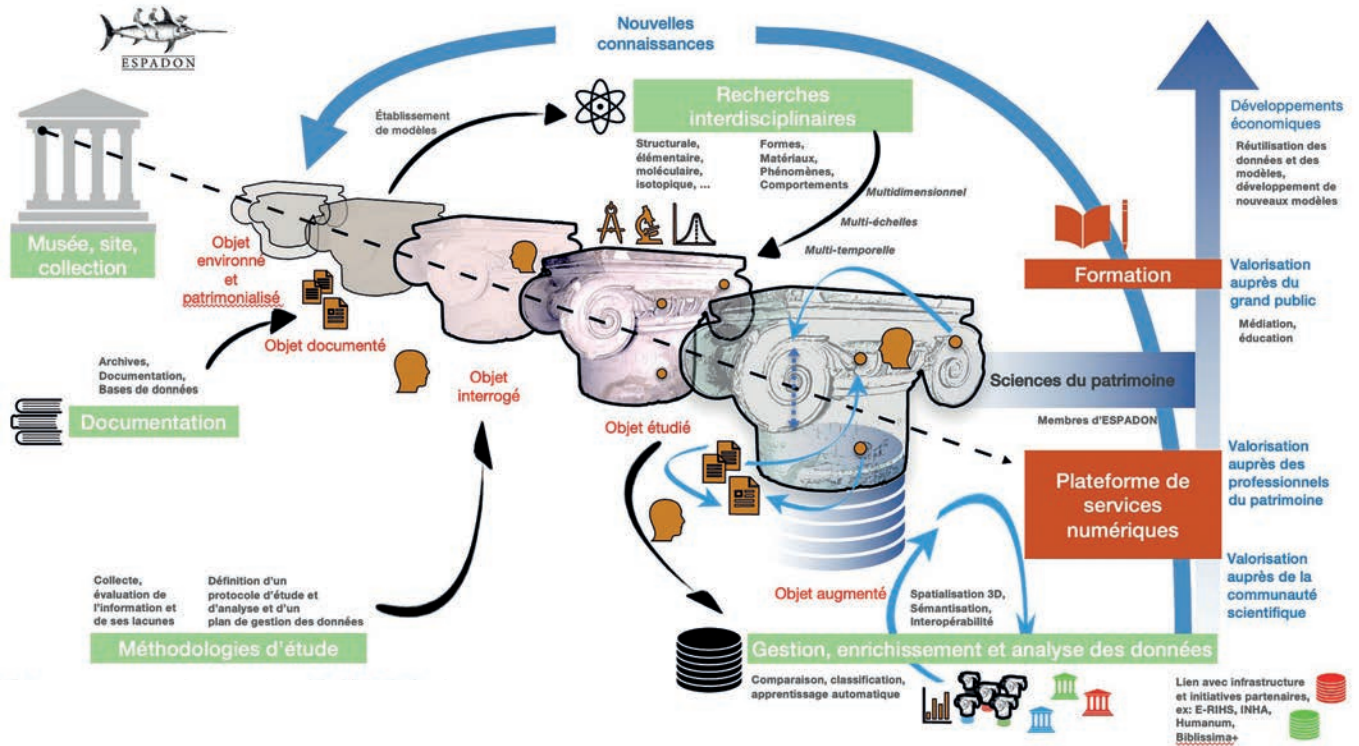
Proposition pour un schéma conceptuel des sciences du patrimoine fondé sur l'objet matériel.

3. DIGILAB : accès à une infrastructure numérique pour le traitement des données quantitatives, mettant en œuvre une politique sur la (ré-)utilisation des données, sur le choix des formats de données, etc.

Instrumentation analytique et structurelle pour des capacités d'analyse multi-dimensionnelle, multi-échelles et multi-temporelle.



IR Camera 7.5-11.5 μm, 50 mm with ring 0.25",
 $L_c = 0.55 \text{ m}$, $L_r = 0.8 \text{ m}$, $E(f) = 3 \text{ kJ (9)}$,
 $\nu(\text{Hz})$ – the Fourier frequency, Image in Phase Shifts or Amplitude.



© Livio De Luca, Elsa Marguin-Hamon, Vincent Detalle, Romain Thomas & consortium ESPADON

Le concept de l'objet patrimonial augmenté.

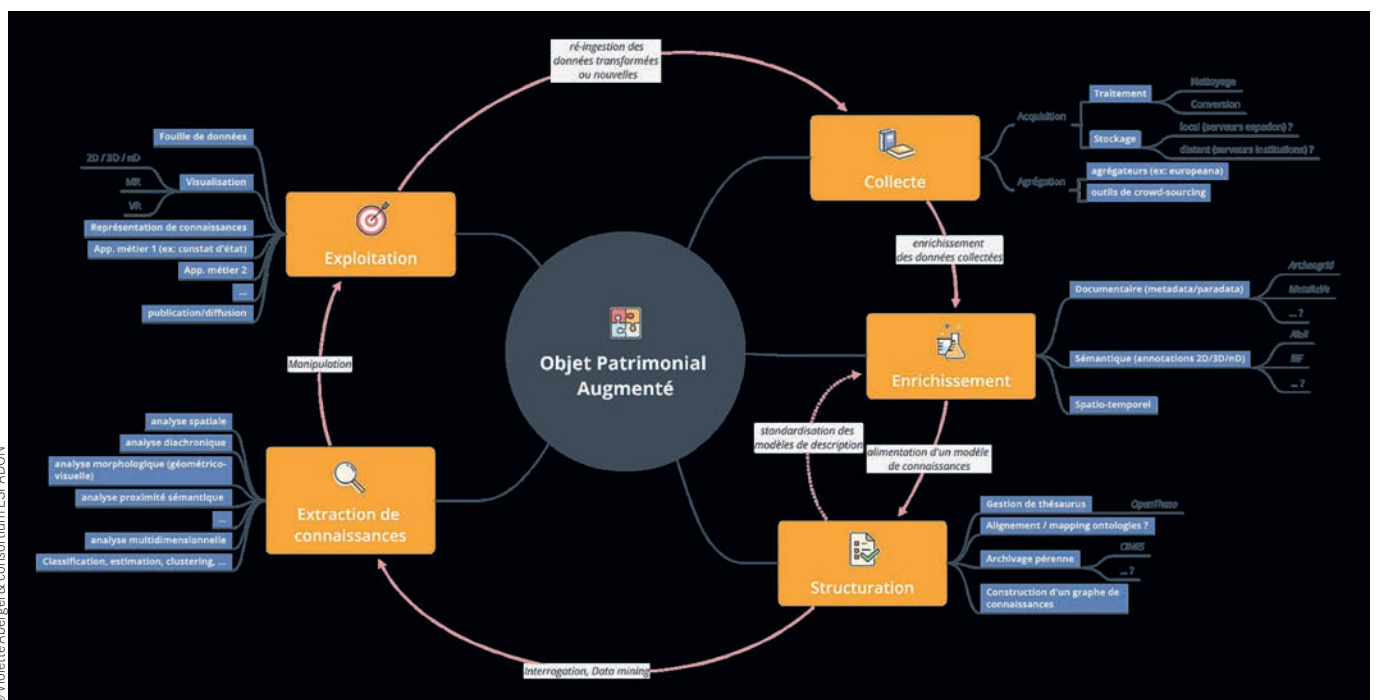
Les états généraux des données du patrimoine

Au cours de l'année 2022 et au début de l'année 2023, à travers la diffusion d'un questionnaire en ligne et l'organisation d'une série de consultations mobilisant les divers domaines de recherche et les professions variées s'intéressant aux « objets » du patrimoine culturel, ESPADON a mené les « états généraux des données du patrimoine ». Il s'agit d'une grande enquête visant à faire le bilan des usages et des pratiques actuelles concernant la production de connaissances et les données associées, à l'échelle nationale. Cette enquête avait également pour but de recueillir les souhaits de tous en vue de l'élaboration du futur écosystème numérique ESPADON, quant à

l'accessibilité des données des objets du patrimoine, à leur création, à leur gestion et à leur manipulation (visualisation, analyse, traitement, interrogation...).

Le questionnaire mis en ligne a été élaboré de manière interdisciplinaire et interprofessionnelle et est destiné à l'ensemble des acteurs (anthropologues, archéologues, conservateurs, documentalistes, historiens, historiens de l'art, professionnels de la médiation, régisseurs, restaurateurs, scientifiques de la conservation, spécialistes des bibliothèques, etc.). Le pari du consortium ESPADON est d'impliquer en amont du projet l'ensemble des acteurs de la communauté. De l'implication de tous dépendra la capacité à élaborer un écosystème numérique répondant le plus largement possible aux besoins. ■

Schéma d'une plateforme de services numériques.



© Violette Abergel & consortium ESPADON

Interopérabilité des données culturelles et patrimoniales : le FAIR sans en avoir l'air

Dans un monde numérique saturé par la circulation et la création de données, les principes FAIR (acronyme pour « facile à trouver, accessible, interopérable, réutilisable ») guident les bonnes pratiques en matière de gestion des données pour la science ouverte. Mis en regard de l'expertise des institutions patrimoniales, ces principes invitent à l'exploration de nouveaux savoir-faire communs, entre la recherche et le domaine culturel.

Dans un monde numérique, les données sont partout : chaque activité, qu'elle soit de nature professionnelle, administrative ou personnelle, fait appel à des outils informatiques et génère des données. Nous produisons des données quand nous naviguons sur Internet ou sur les réseaux sociaux, quand nous communiquons avec nos proches, quand nous effectuons des démarches et, bien sûr, quel que soit notre métier, quand nous travaillons. La recherche n'y fait pas exception : des simples métadonnées des articles publiés dans HAL aux corpus les plus complexes étudiés dans le contexte des humanités numériques, l'activité scientifique génère des volumes sans cesse croissants de données. Plus ou moins structurées, plus ou moins spécifiques à chaque projet, elles impliquent une gestion de leur cycle de vie : combien de temps faut-il les conserver ? Qui pourrait avoir besoin de les réutiliser ? Quel formalisme adopter pour anticiper au mieux ces différents besoins ? Depuis 2016, les principes FAIR¹ préconisent la marche à suivre dans le contexte de la science ouverte.

Les professionnels de l'information face aux données

Dans le domaine des institutions culturelles, l'élaboration de données accessibles et interopérables n'est cependant pas une nouveauté. Le numérique et le Web ont contribué à généraliser des concepts et des méthodes de gestion de l'information qui avaient auparavant fait l'objet de décennies de maturation dans les bibliothèques et les archives.

Élaborer un catalogue ou un inventaire suivant des règles précises, s'assurer qu'une information reste accessible au sein d'une masse de milliers d'objets, analyser la nature, la forme et le contenu d'un



Photo Emmanuelle BERMÈS, CC-BY

EMMANUELLE BERMÈS

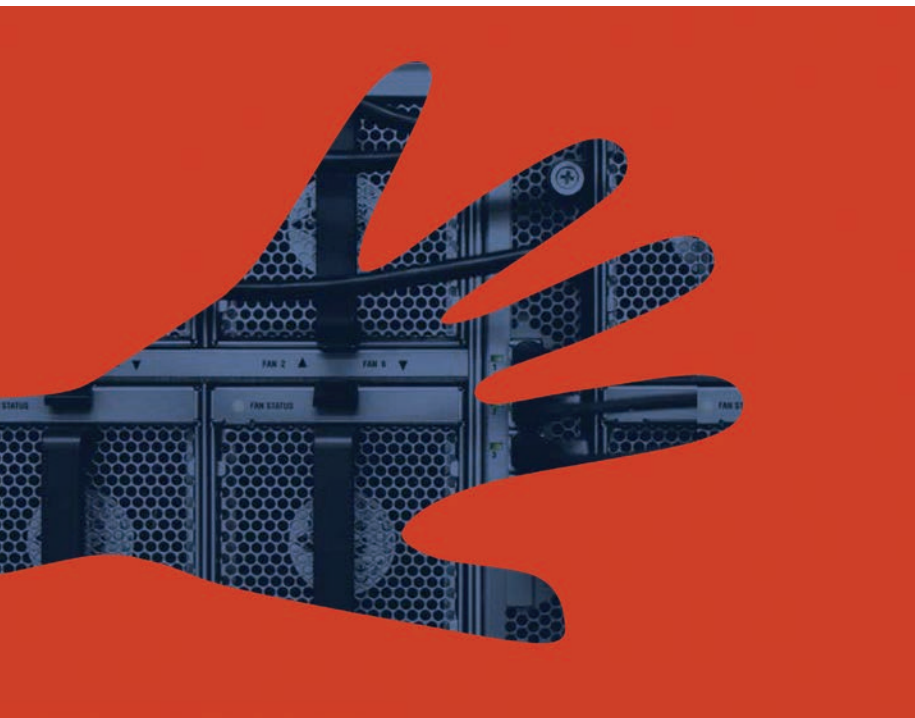
Maîtresse de conférences à l'École nationale des Chartes

Prise de vue d'un catalogue sur fiches conservé sur le campus d'OCLC à Dublin, Ohio.

document ou encore mettre en œuvre les moyens de sa conservation sur le long terme : ces savoir-faire traditionnels des professionnels de l'information se sont aujourd'hui adaptés à la gestion des données.

Au tournant des années 2010, la rencontre entre ces traditions documentaires et le projet du Web de données, tel qu'il avait été formulé par Tim Berners-Lee en 2006², a débouché sur de nouvelles perspectives d'interopérabilité et de découvrabilité. L'enjeu était de créer un espace global de données liées entre elles, mais aussi d'ancrer sur le Web de bonnes pratiques de gestion des données : utilisation d'identifiants et de standards, création de liens facilitant la réutilisation. Si cette vision n'a pas eu l'impact espéré sur l'évolution globale du Web et les applications industrielles du numérique³, elle a contribué à l'émergence de modèles conceptuels comme le CRM (*Conceptual Reference Model*) dans les musées, LRM (*Library Reference Model*) dans les bibliothèques et RIC (*Records In Context*) dans les archives, qui fournissent un cadre pour penser la structure des données de manière globale.

1. Publiés pour la première fois dans la revue *Scientific Data* : <https://www.nature.com/articles/sdata201618>
2. <https://www.w3.org/DesignIssues/LinkedData.html>
3. Gautier Poupeau, « Les technos du Web sémantique ont-elles tenu leurs promesses ? », *Les petites cases*, 6 octobre 2018. <http://www.lespetitescases.net/les-technos-du-web-semantique-ont-elles-tenu-leurs-promesses>



La machine et l'humain.

Illustration par Céline Leclaire (accord pour publication)

« FAIRiser » les données : de nouveaux défis

Le contexte actuel implique cependant un changement d'échelle : la volumétrie des données concernées, leur portée globale, le nombre des acteurs concernés ou encore la granularité de plus en plus fine de l'unité minimale d'information (la numérisation et la fouille de textes et de données⁴ nous donnent désormais accès à des unités de sens bien plus fines que le document ou même le chapitre ou l'article) rendent encore plus vital le besoin de partager ces méthodes, bien au-delà des seuls professionnels de l'information.

Tout projet de recherche doit désormais définir son « plan de gestion de données » et ainsi apporter des garanties sur ses bonnes pratiques dans ce domaine : ne pas créer inutilement des données qui existent déjà, favoriser l'utilisation de standards, disposer d'une stratégie pour que les données restent utilisables au-delà de la durée du projet... Autant d'activités qui bénéficient de l'expérience des bibliothécaires, archivistes ou documentalistes qui interviennent en appui à la recherche dans les universités.

Les bibliothèques numériques, au premier chef Gallica⁵, apportent à la fois une inspiration pour définir les conditions de la « FAIRisation » des données de la recherche et un réservoir contenant des masses de données répondant à ces mêmes exigences de normalisation et d'ouverture.

Les apports théoriques du Web sémantique restent en outre précieux et peuvent s'appliquer aux données produites par la recherche telles que référentiels de concepts, de personnes ou de lieux, corpus de textes ou d'images, etc. Toutefois, l'application des principes FAIR à certains corpus protégés notamment par le droit d'auteur ou la réglementation sur les données personnelles peut soulever des problèmes : c'est le cas pour les corpus Web, constitués pour étudier à des fins scientifiques des contenus libres d'accès sur Internet. Leur stabilisation sous la forme d'une archive s'appuie, à la Bibliothèque nationale de France (BnF) et

à l'Institut national de l'audiovisuel (INA), sur le cadre juridique du dépôt légal⁶, qui impose des restrictions d'accès nécessaires pour préserver les équilibres économiques, souvent fragiles, des acteurs du Web, ainsi que le droit à l'oubli.

En outre, il est légitime de s'interroger sur l'investissement consenti pour rendre les données de la recherche interoperables et les conserver à long terme, alors même que, par nature, elles sont souvent spécifiques à une discipline, une équipe de recherche, voire un projet. Une approche archivistique intégrant les notions de durée d'utilité administrative et d'élimination pourrait être éclairante pour éviter une logique d'accumulation difficilement soutenable, aussi bien d'un point de vue budgétaire qu'écologique.

Face à ces tensions propres à la période de transition numérique que nous traversons, la seule référence aux principes FAIR pourrait présenter des risques. Par chance, la place grandissante du numérique dans les activités de recherche rend de plus en plus nécessaire l'émergence d'une logique de collaboration, faisant dialoguer chercheurs, ingénieurs et professionnels de l'information, notamment dans le contexte des laboratoires (Labs) qui émergent dans les institutions patrimoniales (tels que le BnF DataLab⁷, opéré en partenariat avec IR⁸ Huma-Num⁸, le Lab de l'INA ou encore celui des Archives nationales).

Ces nouveaux lieux de savoirs, où les pratiques de gestion des collections rencontrent celles de la recherche, voient émerger des dynamiques nouvelles, circulaires, où les données culturelles nourrissent la démarche scientifique qui vient elle-même les enrichir de nouvelles données. Et où se créent, en commun, de nouveaux « savoir-FAIR ». ■

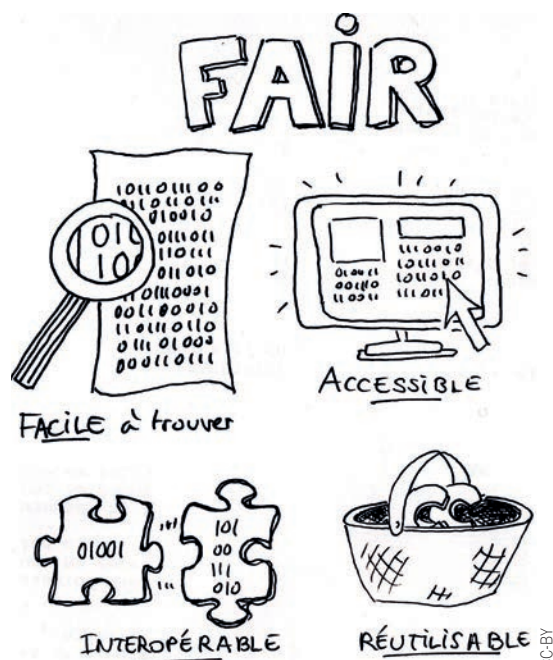
4. Voir dans ce numéro l'article de Didier Thebault, p. 85.

5. Voir dans ce numéro l'article d'Arnaud Laborderie et Florence Tribel, p. 30. La bibliothèque numérique de la BnF et de ses partenaires contient aujourd'hui plus de 10 millions de documents numérisés : <https://gallica.bnf.fr>

6. Vladimir Tybin, « Les collections du dépôt légal du Web de la BnF au cœur des réseaux de coopération internationale pour la recherche », *Culture et Recherche* n° 143, 2022, p. 126-127.

7. Voir dans ce numéro le focus de Marie Carlin, p. 34.

8. Voir dans ce numéro l'article d'Olivier Baude et Nicolas Larrousse, p. 11.



Les principes FAIR : sketchnote par Emmanuelle Bermès.

Les données ouvertes au ministère de la Culture : mises à disposition et usages

En décembre 2016, avec la mise en place de la plateforme de données ouvertes data.culture.gouv.fr (moissonnée quotidiennement par la plateforme interministérielle data.gouv.fr)¹, le ministère de la Culture a souhaité offrir un espace de valorisation de ses ressources culturelles numériques à l'attention des chercheurs, étudiants, journalistes, et aussi des développeurs et des entreprises.

Les objectifs de data.culture.gouv.fr sont de promouvoir l'ouverture des données publiques au sein du ministère, de fédérer ses usagers et ses producteurs autour d'un écosystème culturel des données ouvertes et liées, et d'associer les citoyens à l'amélioration de la qualité des données.

Pour quels publics ?

La plateforme s'adresse aux usagers qui recherchent des informations en leur proposant une visualisation des données attractive et lisible, et s'adresse également aux réutilisateurs des données, auxquels elle propose des exports sous différents formats, et des API².

Les outils de visualisation de data.culture.gouv.fr permettent une meilleure valorisation des ressources et, en conséquence, une meilleure visibilité du travail des services producteurs. L'acculturation et l'implication des producteurs constituent une étape fondamentale dans la démarche d'ouverture.

À cette fin, il est nécessaire de les accompagner dans les différentes phases du processus en tant que de besoin.

Quelles données ?

Les données ouvertes (*Open Data*) sont des données librement accessibles et gratuitement réutilisables par tous. Il est donc nécessaire, pour chaque jeu de données, de s'assurer dans un premier temps que les données sont dans le périmètre de l'ouverture et de veiller à leur conformité juridique et réglementaire.

Il s'agit de vérifier que des tiers ne détiennent pas des droits de propriété intellectuelle sur les données, qu'il n'y a pas de données à caractère personnel (sauf exception) et que ces données n'affectent pas la vie privée des citoyens (secret statistique). Les données

couvertes par un secret protégé par la loi sont également exclues.

Une fois délimité le périmètre des données à ouvrir, il faut considérer la question de la qualité des données, garante d'usages réels des données : des données fiables, non agrégées, facilement réutilisables par des machines, régulièrement mises à jour et documentées pour permettre aux usagers d'en comprendre la structure et le sens.

Les données publiées sur data.culture.gouv.fr concernent aussi bien l'activité des institutions culturelles que les collections, les vocabulaires scientifiques, ou encore les événements culturels.

Pour illustrer toute la diversité de l'intérêt que les utilisateurs pourront trouver à la plateforme, nous prendrons quelques exemples dans le domaine des archives :

- des données de type liste : l'annuaire des services d'archives publiques en France qui va renseigner son public sur les horaires d'ouverture des services, leur géolocalisation et l'adresse électronique du contact, etc. Ce jeu de données est actualisé au fil de l'eau par le producteur, et mis à jour automatiquement sur data.culture.gouv.fr ;
- des données sur l'activité des services publics d'archives, collectées annuellement par le biais d'une enquête menée par le Service interministériel des Archives de France (SIAF) ; plusieurs années sont disponibles et permettent ainsi des comparaisons dans le temps ;
- des données d'enquête des pratiques auprès des usagers en ligne : deux enquêtes menées en 2014 et en 2021 vont permettre une mise en perspective de ces usages ;

ROSELYNE ALIACAR

Chargée de mission Circulation et ouverture des données/Département du numérique pour la transformation des politiques culturelles et de l'administration des données/Secrétariat général/Service du numérique/ministère de la Culture

1. <https://data.culture.gouv.fr/pages/home>

2. API pour *Application Programming Interface* (ou « interface de programmation d'application ») est une interface qui permet de « connecter » un logiciel ou un service à un autre logiciel ou service afin d'échanger des données.

- des corpus de données historiques issues de ces inventaires : par exemple, « les députés bannis de la Convention nationale pour avoir voté la mort du roi Louis XVI » ; les « Pièces de théâtre écrites par des femmes, et représentées à Paris entre 1809 et 1906 » ; « Les enquêtes statistiques sur les moulins à farine diligentées par l'État en 1794, 1801 et 1809 afin d'évaluer les moyens de production de farines, d'un point de vue quantitatif, qualitatif et aussi prospectif » ; « Les dénaturisés de Vichy » ; etc.

Quelles finalités ?

Ces jeux de données sont très utiles en particulier aux chercheurs, historiens, sociologues, généalogistes, etc. Mais aussi, grâce à leur accessibilité facilitée par leur exposition sur data.culture.gouv.fr, ces informations connues essentiellement des spécialistes atteindront un public plus large. Elles permettent de découvrir des aspects peu connus de notre histoire et favorisent un éveil de la curiosité pour l'histoire de France chez le plus grand nombre.

La finalité de la plateforme n'est pas seulement de mettre à disposition des données, mais aussi de veiller à ce que ces données (contenu et présentation) correspondent aux besoins des chercheurs et des communautés réutilisatrices. C'est pourquoi le ministère a lancé une consultation publique en décembre 2022. L'objectif était de recueillir les demandes d'un maximum d'utilisateurs afin de permettre d'identifier les données à améliorer et les nouvelles données à ouvrir. Le calendrier de mise en œuvre des actions correspondantes sera publié avant l'été 2023.

En répondant ainsi au plus près des attentes, nous espérons renforcer le cercle vertueux de la science ouverte (*Open Science*) et des données ouvertes (*Open Data*) : les réutilisateurs alimentés en données pourront à leur tour les analyser, les associer et les transformer pour produire de nouveaux services accessibles et utiles à d'autres publics. ■



Crédits photo : Rythmes par Robert Delaunay, 1885-1941 – 1934 – Centre Pompidou – National Museum for Modern Art – Industrial Design Centre, France – CC-BY-NC-SA. https://www.europeana.eu/item/2063621/FRA_280_010

Bandeau de la page d'accueil
de la plateforme de données ouvertes
du ministère de la Culture :
<https://data.culture.gouv.fr/pages/home/>

La plateforme ouverte du patrimoine (POP) : un outil de mutualisation et de diffusion des données scientifiques des services culturels¹

Mise en ligne en septembre 2018, la Plateforme ouverte du patrimoine (POP)² s'inscrit dans la riche tradition de description et de diffusion de données scientifiques voulue par le ministère de la Culture depuis plus de cinquante ans. Réunissant des producteurs de données aux statuts hétérogènes, POP permet la production et l'exposition de plus de quatre millions de notices qui décrivent des œuvres conservées ou étudiées en France. Quatre ans après son lancement porté par beta.gouv³, la plateforme se lance dans une refonte, visant à améliorer autant la qualité des données que leur réutilisation.

Des bases historiques

POP présente huit bases de données, désormais enrichies par près de six cents institutions différentes⁴. Cinq de ces bases ont une portée réglementaire, indiquée dans le Code du patrimoine dans ses livres « Monuments historiques » et « Musées », ainsi que dans la loi du 13 août 2004 organisant la décentralisation de l'inventaire général⁵.

Actuel catalogue collectif des collections des 1 216 musées de France, la base Joconde est l'héritière de bases de données spécialisées (peintures, dessins, estampes, sculptures, objets d'art, archéologiques, ethnographiques), progressivement fusionnées de 1975 à 2004 sous l'égide du service des Musées de France (SMF). Joconde diffuse actuellement 658 752 notices illustrées descriptives d'objets, émanant de plus de 570 musées bénéficiant de l'appellation « Musée de France ». L'enrichissement de Joconde figure dans les missions des Musées de France précisées par la loi du 4 janvier 2002 : rendre accessibles au public les collections des musées et contribuer au progrès de la connaissance et de la recherche. La création et la modification des notices dans Joconde sont à la charge de l'équipe du musée, qui procède aux imports de notices avec l'accompagnement du SMF. Outre la publication sur Joconde, le SMF facilite la mise à disposition de 16 vocabulaires permettant de décrire les collections. Ces vocabulaires concernent les auteurs, la représentation, la dénomination, les lieux de création, la découverte-collecte, etc. Joconde est connectée à Muséofile, répertoire des Musées de France, accessible sur POP.

La base Rose-Valland (MNR-Jeu de Paume) reprend deux anciennes bases fusionnées en 2022 : MNR et Jeu de Paume. Conçue en 1997, MNR contient 2 143 notices (dont 2 210 illustrées) décrivant chacune un bien « MNR ». Créé en 1949, le statut Musées nationaux récupération (MNR) désigne, parmi les 60 000 œuvres récupérées en Allemagne après le conflit et renvoyées en France, les 2 200 qui n'ont pu être restituées. Confiés à la garde des Musées nationaux, ces biens, qui n'ont pas tous été spoliés⁶, continuent d'être étudiés par la Mission de recherche et de restitution des biens culturels spoliés entre 1933 et 1945 (M2RS) et les musées dépositaires afin de déterminer leur historique et leur provenance. MNR propose ainsi le dernier état connu de la recherche. Jeu de Paume contient, pour sa part, 246 notices décrivant les œuvres exposées sous l'Occupation par l'*Einsatzstab Reichsleiter Rosenberg* (ERR).

Disponibles dans la base Enluminures, près de 110 000 notices illustrées décrivent des enluminures et éléments de décor, provenant de plus de 8 000 manuscrits médiévaux conservés dans les bibliothèques publiques françaises – hors la Bibliothèque nationale de France. La base procède d'Initiale, administrée depuis les années 1980 par l'Institut de recherche sur l'histoire des textes (IRHT). Elle est régulièrement enrichie *via* des campagnes de numérisations subventionnées par le service du Livre et de la Lecture.

Forte de 13 899 notices partiellement illustrées (50 %), la base Autor a repris le contenu de Poinçons, répertoire des orfèvres enregistrés à la Garantie de

CAMILLE DUCLERT

Directrice adjointe de la Médiathèque du patrimoine et de la photographie (MPP), Conservatrice du patrimoine, ministère de la Culture

1. Cet article a été relu par le comité métiers de POP.

2. <https://www.pop.culture.gouv.fr/>

3. beta.gouv.fr est l'incubateur de services numériques de l'État : <https://beta.gouv.fr/>

4. Les chiffres cités dans cet article sont à la date du 28 mars 2023.

5. Loi n° 2004-809 du 13 août 2004 relative aux libertés et responsabilités locales : <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT00000804607/>

6. La proportion d'œuvres spoliées parmi les MNR reste incertaine. Toute œuvre spoliée est susceptible d'être restituée à son propriétaire.



La plateforme POP regroupe les contenus numériques du patrimoine français afin de les rendre accessibles et consultables au plus grand nombre

Recherchez parmi plus de 3 millions de documents

Rechercher

Les bases actuellement dans POP

- Collections des Musées de France Base Joconde
- Patrimoine mobilier Base Palissy
- Patrimoine architectural Base Mérimée
- Photographies Base Mémoire
- Rose Valland Base MNR-jeu de Paume

Page d'accueil de POP, www.pop.culture.gouv.fr, visitée le 28 mars 2023.

Paris entre 1875 et 1989, supervisé par la Mission de l'inventaire général du patrimoine culturel (MIGPC). Deux autres ressources sont appelées à rejoindre Autor : un répertoire biographique des peintres-verriers des XIX^e et XX^e siècles, élaboré par le centre André-Chastel, et celui des producteurs de fonds de la Médiathèque du patrimoine et de la photographie (MPP).

Comptant 2,5 millions de notices décrivant des images fixes, la base Mémoire est alimentée depuis 1995 par deux producteurs différents : d'une part, les services régionaux de l'inventaire (SRI) qui étaient positionnés en DRAC avant leur transfert et, de l'autre, la MPP (service du ministère de la Culture). Les photographies versées par les SRI proviennent des dossiers d'inventaire tandis que les images mises en ligne par la MPP illustrent autant ses fonds

archivistiques (dossiers de protection ou de restauration des Monuments historiques) que les dons et dépôts consentis par des photographes tels Jacques Henri Lartigue, Willy Ronis, André Kertész, Gilles Caron, Christine Spengler, etc. Les règles de saisie des notices ne sont pas normatives, mais relèvent d'un système de description commun aux services producteurs.

Les bases Mérimée et Palissy connaissent la même bipartition. Pour le compte du service des Monuments historiques, la MPP crée dans POP des notices pour tout édifice ou objet/ensemble mobilier protégé au titre des Monuments historiques (46 148 notices dans Mérimée MH et 287 963 dans Palissy MH). Côté inventaire, en application de la loi de décentralisation, chaque SRI veille au versement dans les bases nationales de fiches synthétisant les données



OAR00240, Harpe, Sébastien Erard (1751-1831), Paris, musée de la Musique (affectataire musée du Louvre, département des Objets d'art).
© Réunion des musées nationaux Grand Palais
© Musée du Louvre, département des Objets d'art
© Musée de la Musique, Germain Claude



AP2122T001053, Sculpture : dieu d'Euffigneix, musée d'Archéologie nationale, Jean Roubier (1896-1981).

© Donation Jean Roubier, Ministère de la Culture (France), Médiathèque du patrimoine et de la photographie, diffusion RMN-GP

© Ministère de la Culture (France), Médiathèque du patrimoine et de la photographie (objets mobiliers)



AP40W00370, Tricycle ferroviaire dit kalamazoo, Landes (Sabres), classé au titre des Monuments historiques le 27 mars 1990.



APMH00013766, Abside de la cathédrale Saint-Bénigne, Médéric Mieusement (1840-1905).

© Ministère de la Culture (France), Médiathèque du patrimoine et de la photographie, diffusion RMN-GP

d'études et d'inventaires produites et diffusées dans un outil régional (Gertrude). MPP et SRI suivent le même système de description pour chaque base, et les vocabulaires et référentiels sont identiques.

POP, un accélérateur de la diffusion des connaissances

En hébergeant ces huit bases sur une seule plateforme, POP fait donc cohabiter les données des Musées de France, des services régionaux de l'Inventaire, de l'administration des Monuments historiques, de la M2RS, de l'IRHT, de la MIGPC et de la MPP. À ce regroupement de données métiers répondent des choix architecturaux et techniques (sélection de MongoDB⁷ et exposition des modèles de données des bases) témoignant d'une volonté forte de diffusion des connaissances.

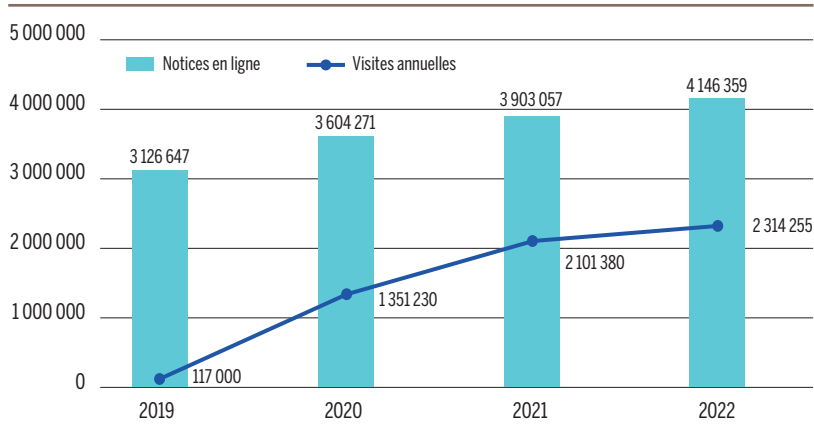
Une interrogation unique sur six champs communs à toutes les bases (titre, dénomination, localisation et auteur) a ainsi été mise en place, permettant une recherche plus fluide et plus rapide sur un nombre croissant de données.

Des champs ont été confortés, voire ajoutés dans les huit bases afin de permettre des rebonds vers d'autres bases de données : bases internes de POP, catalogues en ligne des musées, bases de données spécialisées sur les recherches de provenance des œuvres, base Agorha de l'INHA depuis Mérimée, base Archiv'MH de la MPP depuis Mérimée et Palissy, lien vers les dossiers Inventaires des bases régionales.

Par ailleurs, une part significative des données de POP sont transmises au site www.data.culture.gouv.fr, la plateforme de données ouvertes du ministère⁸. Y sont disponibles les données de Mérimée MH (parmi les plus populaires), de Joconde, d'Autor, de Muséofile et de Palissy MH (sauf pour les objets à la propriété privée). Seules les données produites par les SRI ne s'y trouvent pas, leur diffusion relevant des régions.

Indépendamment de la croissance du nombre de notices et de visiteurs, les services métiers, producteurs des données de POP, ont souhaité dès 2020 de nouvelles fonctionnalités. Cette nouvelle phase est conjointement portée par l'Atelier du numérique, incubateur des projets numériques du ministère, et

POP, 2019-2022



Statistiques 2019-2022 illustrant le rapport entre l'augmentation du nombre des notices mises à disposition et l'augmentation du nombre de visiteurs par an.

© Ministère de la Culture (France), Médiathèque du patrimoine et de la photographie

l'équipe POP. Plusieurs chantiers ont été identifiés afin d'augmenter significativement le nombre de notices en ligne et la qualité des données.

Le premier s'attellera en 2023 à automatiser et à normaliser les versements des données, notamment depuis les bases des SRI et celles des Musées de France par le recours au protocole OAI-PMH et aux API. L'objectif est autant de fluidifier ces opérations afin d'inciter les services à verser que de soumettre les données elles-mêmes à des contrôles de termes, grâce à la connexion de POP aux outils gérant les référentiels.

Un second enjeu touche au perfectionnement de la recherche. Il s'agit d'y intégrer les termes et la hiérarchisation des vocabulaires *via* l'auto-complétion, entre autres, et de la rendre plus intuitive pour l'internaute.

Enfin, l'équipe POP entend faciliter la réutilisation des données exposées, actuellement limitée à une impression .pdf de la notice. Un export des résultats de recherche en format CSV⁹ est une étape largement attendue.

L'ensemble de ces évolutions confortera la place stratégique de POP dans la diffusion de données culturelles du ministère et participera à la mission de transmission incarnée par le ministère depuis plus de cinquante ans. ■

7. MongoDB (de l'anglais *humongous* qui peut être traduit par « énorme ») est un système de gestion de base de données orienté « documents », répartisable sur un nombre quelconque d'ordinateurs et ne nécessitant pas de schéma prédéfini des données.

8. Voir dans ce numéro l'article de Roselyne Aliacar, p. 73.

9. Un fichier CSV (valeurs séparées par des virgules) est un type de fichier spécial qu'il est possible de créer ou de modifier dans Excel. Plutôt que de stocker les informations en colonnes, les fichiers CSV les stockent en les séparant par des points-virgules.

La Research Data Alliance (RDA):

pour des passerelles sociales et techniques afin de permettre le libre partage des données, leur réutilisation et leur interopérabilité scientifiques

La Research Data Alliance (RDA)¹ est une organisation internationale créée en mars 2013 par la Commission européenne, le gouvernement australien, la National Science Foundation et le National Institute of Standards and Technologies des États-Unis, avec pour mission de faciliter le partage des données de la recherche. En mars 2023, elle rassemblait plus de 13 000 membres venant de 150 pays, constituant un forum international neutre unique de chercheurs et professionnels des données scientifiques travaillant dans de multiples disciplines, domaines et thématiques et appartenant à différents types d'organisations à travers le monde.

FRANÇOISE GENOVA

Animatrice du Nœud national RDA France, directrice de recherche émérite, Centre de données astronomiques de Strasbourg, Infrastructure de recherche CNRS (IR) 7550, Observatoire astronomique de Strasbourg, Université de Strasbourg

Toute personne intéressée par le sujet peut en effet devenir membre de la RDA : l'adhésion individuelle est gratuite, mais suppose l'engagement d'adhésion aux valeurs de celle-ci : ouverture, consensus, inclusivité, harmonisation, pilotage par la communauté, but non lucratif et neutralité technologique. Les organisations qui partagent ces objectifs sont aussi invitées à rejoindre la RDA moyennant une cotisation annuelle.

Les activités de la Research Data Alliance

Ce sont les membres de la RDA qui proposent et mènent ses activités, pour construire des passerelles sociales et techniques qui facilitent le partage ouvert et la réutilisation des données. Celles-ci sont portées par des groupes d'intérêt, qui prennent en charge un sujet sur la durée, et de groupes de travail, qui sont mis en place pour 18 mois, avec pour objectif de produire une ou des recommandations qui seront endossées par la RDA. Ces recommandations et les productions des groupes d'intérêt sont publiées sur le site de la RDA² et peuvent être utilisées librement.

Le domaine couvert par la RDA est très vaste, et cette diversité qui la caractérise se retrouve dans les sujets abordés par ses membres au sein de plus de



90 groupes d'intérêt et groupes de travail. La RDA abrite par exemple de très nombreuses activités autour des principes FAIR (« facile à trouver, accessible, interopérable, réutilisable ») qui caractérisent les propriétés des données dans le cadre de la science ouverte. Elle étudie aussi des sujets plus « sociaux », tels que la professionnalisation de la gestion des données ou à la manière d'amener les chercheurs à s'intéresser aux données.

Parmi ses contributions importantes à la mise en œuvre de la science ouverte, on peut citer CoreTrustSeal³, qui résulte de la fusion de certifications antérieures, réalisé par un groupe de travail de la RDA et qui est devenu le cadre de certification de

1. <https://www.rd-alliance.org/>

2. <https://www.rd-alliance.org/recommendations-outputs>

3. <https://www.coretrustseal.org/>



© Paula Afrianto

base des entrepôts de données préconisé dans le Plan national pour la science ouverte (PNSO).

La RDA a également mis en place récemment la possibilité de proposer une Communauté de pratique (CoP) pour les domaines déjà bien présents en son sein : une Communauté de pratique est le point focal de la RDA pour un domaine spécifique.

La Research Data Alliance, RDA France et la stratégie française pour la science ouverte

La RDA et son chapitre français RDA France ont le soutien du Fonds national pour la science ouverte. RDA France⁴, hébergé par le CNRS, fait circuler les informations sur les activités de la RDA, encourage la communauté française à participer à celle-ci et travaille à l'acculturation de la communauté française

à la science ouverte pour ce qui concerne le partage des données.

Elle gère des listes d'information, organise des ateliers et a des activités plus spécifiques sur la certification des entrepôts et services de données⁵ au sein d'un groupe de travail commun avec le Collège données de la recherche du Comité pour la science ouverte.

RDA France soutient aussi la participation française aux réunions plénières bisannuelles⁶ de la RDA. Comme celles de la RDA, ses activités sont ouvertes à toute personne intéressée.

La RDA fait progresser de nombreux sujets qui s'inscrivent pleinement dans la stratégie nationale pour la science ouverte. RDA France contribue à créer une communauté nationale autour des données de la recherche, et permet à celle-ci de s'inscrire dans un paysage international en évolution rapide. ■

La 20^e plénière de la RDA (Göteborg, Suède, 21-23 mars 2023) a marqué son dixième anniversaire. Elle avait été créée dans la même ville en mars 2013.

4. <https://www.rd-alliance.org/groups/rda-france>

5. <https://www.ouvrirlascience.fr/certification-des-entrepots-et-services-de-donnees/>

6. Trois établissements français de l'enseignement supérieur et de la recherche ont rejoint la RDA : l'Université Grenoble Alpes, l'Université de Lorraine et l'Université de Strasbourg.



© RDA

Logo des 10 ans de la RDA.

Les archivistes, des professionnels au service de la diminution des impacts environnementaux du numérique : l'apport de l'outil Archifiltre

À l'ère de la dématérialisation effrénée et à l'heure d'une prise de conscience croissante de l'impact environnemental que représente le numérique, il peut paraître antinomique d'associer professionnels de la conservation et diminution de l'impact environnemental lié au numérique. Produire et conserver des documents et données numériques est, par essence, énergivore. Pourtant, l'archiviste peut jouer un rôle positif dans ce domaine en s'appuyant sur son savoir-faire de sélectionneur et sur certains outils pour lui permettre d'exercer ses talents.

CHLOÉ MOSER

Adjointe à la cheffe de la mission des Archives de France auprès des ministères sociaux, Service interministériel des Archives de France, ministère de la Culture, cheffe de produit Archifiltre

L'explosion de la production numérique et son impact environnemental

Il est courant d'associer la notion de virtuel au numérique. Le volume occupé par toutes les données produites et conservées par l'ensemble des utilisateurs du numérique est, cependant, bien réel même si invisible aux yeux des utilisateurs... Son impact est d'autant moins virtuel que nos données et fichiers numériques font l'objet d'une multiplication des copies de sauvegarde pour leur sécurité, ce qui rend leur volume encore plus significatif.

Plusieurs institutions (l'ADEME, Agence de la transition écologique, ou l'Institut numérique responsable) et événements ont rendu visible cette problématique de la « face cachée du numérique ». Ils ont permis de communiquer sur les problématiques environnementales posées par le numérique, à savoir les ressources en énergie nécessaires pour la fabrication des matériels (78 % de l'impact environnemental du numérique) ou pour le fonctionnement des réseaux pour le stockage des données (4 à 22 %)¹.

Face à la réalité de cette croissance numérique, en tant qu'il accompagne et collecte les documents et données produits, l'archiviste n'est pas responsable des volumes générés ni des supports de

production choisis : ce qu'il conserve après sélection ne sera que le reflet de la façon dont les services travaillent et produisent leurs données. Par ailleurs, pour traiter ces gisements de données numériques, l'archiviste pouvait manquer de repères et avait du mal à mettre en application son pouvoir d'intervention face à des masses abstraites et difficiles à appréhender.

Un outil dans la mallette des archivistes pour le numérique : Archifiltre

Dans ce contexte, les outils Archifiltre ont été conçus à partir d'une réflexion sur la visualisation. L'archiviste ayant perdu la capacité d'appréhender l'information dans sa globalité dans l'environnement numérique, l'automatisation de la visualisation lui permet de retrouver la capacité de mettre en œuvre ses compétences pour plusieurs activités de son cœur de métier : le tri des documents ou mails *via* l'élimination des éléments redondants ou dont la durée de conservation est échu, la description des documents à conserver, la préparation des paquets de documents ainsi sélectionnés, l'accompagnement de la production en amont de la collecte directement auprès des producteurs et, bientôt, la préparation de l'archivage de documents et données issus de systèmes d'information.

Archifiltre-Docs, pour les documents, et Archifiltre-Mails, pour les messageries, ont ainsi été développés comme des outils d'aide à la décision, à la sélection et au traitement. Ce sont des outils libres et gratuits, dont le code source est ouvert².

Ces outils ont été conçus à partir de 2018, initialement dans le cadre d'un défi du dispositif « Entrepreneurs d'intérêt général³ », dispositif de développement soutenu par Etalab⁴ à la Direction interministérielle du numérique, pour accompagner l'archiviste ou tout producteur de documents dans l'analyse des ensembles volumineux de fichiers, une identification plus intuitive des éliminables et la

Plusieurs institutions (l'ADEME, Agence de la transition écologique, ou l'Institut numérique responsable) et événements ont rendu visible cette problématique de la « face cachée du numérique ».

1. De nombreuses infographies de l'ADEME présentent une illustration de cette réalité : <https://infos.ademe.fr/magazine-avril-2022/faits-et-chiffres/numerique-quel-impact-environnemental/>

2. Les deux outils de la suite Archifiltre sont téléchargeables sur le site <https://archifiltre.fr/>

3. <https://eig.etalab.gouv.fr/>

4. Etalab est une administration publique française qui vise à améliorer le service public et l'action publique grâce aux données. Etalab contribue ainsi à la transformation de l'action publique et à l'innovation publique grâce au numérique.

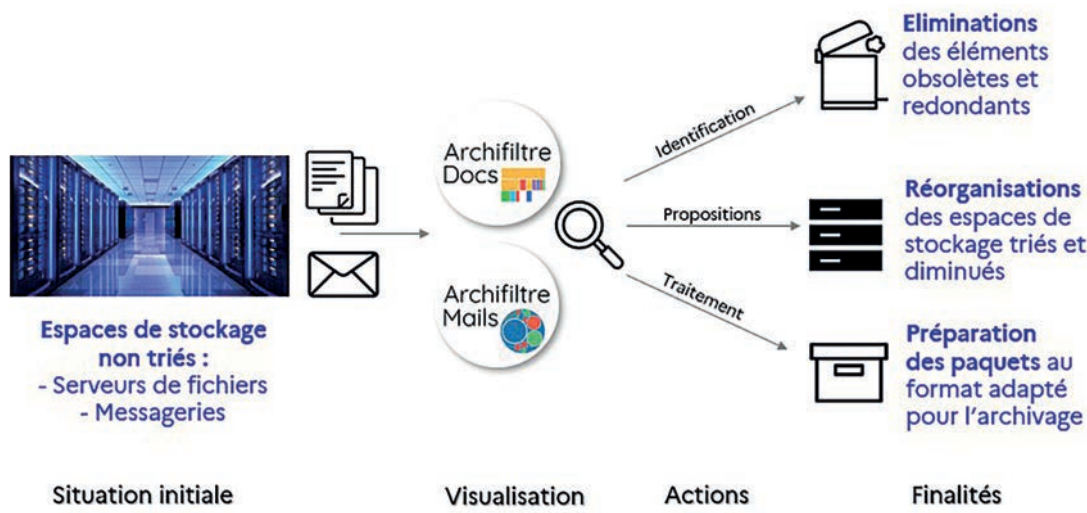


Schéma des fonctionnalités principales des outils Archifiltre.
© C. Moser

facilitation des réorganisations d'espaces pour une plus grande efficacité.

Ils ont été, enfin, développés comme une brique dans l'écosystème des outils de préparation pour l'archivage électronique (notamment dans des systèmes tels que VITAM⁵). Depuis 2019, les développements se poursuivent dans le cadre de la Fabrique numérique des ministères sociaux avec le soutien du Service interministériel des archives de France (SIAF).

Comment Archifiltre peut-il contribuer à diminuer l'impact du stockage numérique ?

L'archiviste est le premier promoteur au sein de son institution de la réduction des espaces de stockage de documents. Ce rôle n'est évidemment pas né avec la production numérique, c'est déjà celui qu'il exerce dans l'environnement papier. Le volume occupé par les documents papier ne pouvant augmenter sans fin, principalement pour des raisons de coûts et de place dans les locaux, l'archiviste est un expert de l'information reconnu qui sait accompagner les producteurs de documents pour le tri et la conservation de l'essentiel : ce qui est encore à conserver pour des raisons juridiques ou d'utilité administrative, et ce qui mérite d'être conservé à titre historique.

Comme corollaire de cette compétence de sélection, l'archiviste porte également depuis longtemps la sensibilisation aux conséquences d'une conservation trop large et longue de documents inutiles. À côté des habituels arguments de manque d'efficacité au quotidien dans des arborescences complexes, de risques juridiques et de coût de l'augmentation continue des espaces de stockage, la sensibilisation à l'impact environnemental du numérique est une nouvelle façon de prôner la vertu d'un tri régulier et collectif. Archifiltre est ainsi utilisé auprès des services producteurs, dans les opérations de nettoyage et de tri

des espaces de stockage, que ce soit au quotidien dans l'accompagnement des producteurs ou lors d'événements particuliers comme les *Cleaning Days* internes à l'institution ou les initiatives internationales comme le *Digital Cleanup Day*⁶ qui remportent du succès dans les équipes qui les proposent.

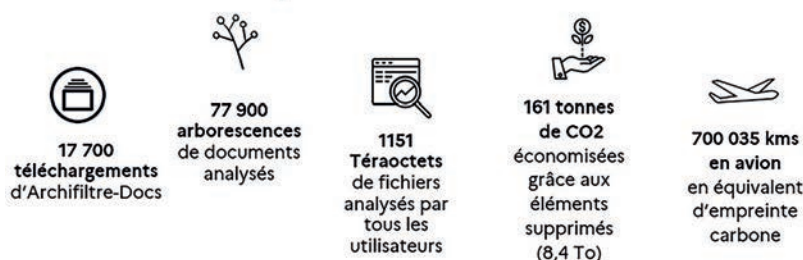
Depuis 2021, la start-up d'État Archifiltre calcule ses impacts et les illustre en affichant les volumes cumulés de documents analysés dans ses statistiques, les volumes supprimés par l'ensemble des utilisateurs et l'économie que ces volumes représentent en équivalent CO₂, valeur ensuite convertie dans une unité de mesure appréhendable plus facilement par les utilisateurs. L'interface du deuxième outil de la suite, Archifiltre-Mails, s'inspire de cette représentation pour placer le gain environnemental dans l'espace de travail des utilisateurs.

Si l'archiviste n'est qu'un des acteurs dans la chaîne de production des données et documents et n'a pas la maîtrise de la transformation numérique ni du volume exponentiel de données et documents produits, il peut cependant réellement apporter sa contribution à la sensibilisation et à la réduction de l'impact du stockage de données obsolètes ou redondantes. Dans ce contexte, les outils Archifiltre ont été conçus comme des aides pour relever ces défis. ■

5. Pour en savoir plus sur ce programme interministériel initié en 2015 : <https://www.programmevitam.fr/>
6. <https://digital-cleanup-day.fr/>

L'impact des outils Archifiltre en lien avec les statistiques d'utilisation.
© C. Moser

L'impact d'Archifiltre



Compter les mots pour remonter le temps : Gallicagram et Gallicagrapher, deux outils d'exploration des archives numérisées de la BnF

Les archives numérisées de la Bibliothèque nationale de France, Gallica, constituent un trésor de données ouvertes. Les auteurs ont développé deux logiciels de lexicométrie, qui mesurent et permettent de visualiser l'évolution de l'usage des mots au cours du temps, et facilitent l'accès au contexte des occurrences.

BENJAMIN AZOULAY

Administrateur-élève des affaires maritimes, ministère de la Mer et École normale supérieure Paris-Saclay

BENOÎT DE COURSON

Doctorant au Max Planck Institute de Freiburg (Allemagne)

WILL GLEASON

Développeur indépendant

La Bibliothèque nationale de France (BnF) dispose d'un des plus riches fonds imprimés au monde. Elle est en effet dépositaire du dépôt légal : chaque livre ou numéro de presse publié depuis 1537 doit y être déposé pour archivage et désormais pour consultation par les chercheurs.

Depuis 1997, la BnF entreprend une numérisation massive de ses fonds et les verse en libre accès sur la plateforme Gallica. Cette « bibliothèque virtuelle de l'honnête homme¹ » est aussitôt devenue un outil de travail incontournable pour les chercheurs en humanités.

De par son volume (400 000 livres et 4,5 millions de numéros de presse océrisés² en français), Gallica se prête à merveille à la lexicométrie, c'est-à-dire au traitement quantitatif des textes. L'accès transparent aux données est ici un atout précieux. Développés à cette fin, les sites Gallicagram et Gallicagrapher visent à offrir aux chercheurs en humanités une interface Web pour exploiter et visualiser ces données.

Les outils lexicométriques existants n'appliquent guère les principes de la « science ouverte ». L'outil de référence en la matière, Google Books Ngram Viewer³, est peu utilisé par les chercheurs, embarrassés par la constitution opaque de son corpus et déçus de ne pouvoir accéder au contexte des occurrences⁴. Plus souple et transparent, Frantext⁵ permet des relevés syntaxiques dans près de 6 000 œuvres (soit 266 millions de mots), mais son volume est insuffisant pour des analyses quantitatives diachroniques⁶, et Frantext (par ailleurs payant) a récemment cessé d'assortir les données récoltées de graphiques.

Gallicagram et Gallicagrapher : deux logiciels pour démocratiser le traitement automatique de Gallica

Gallicagram⁷ est un logiciel permettant de visualiser l'évolution de l'usage des mots au cours du temps, en fouillant, parmi d'autres corpus, la presse et les livres numérisés de Gallica. La croissance des données interrogées à partir de la Révolution rend le corpus de presse (certainement le plus intéressant pour l'historien) particulièrement fiable entre 1789 et 1950.

Disposant d'un corpus ouvert, l'historien est à même de savoir dans quoi il cherche, et d'accéder au contexte des occurrences. Pour ce faire, Gallicagrapher⁸ exploite les API⁹ de Gallica et présente le contexte immédiat de chaque occurrence, directement dans le logiciel, sur le modèle de Frantext. Cela facilite l'analyse des résultats correspondant aux courbes affichées, ce qui permet, par exemple, de lever les ambiguïtés sur les homonymes et les erreurs de reconnaissance optique des caractères.

Les deux logiciels cherchent à appliquer les principes de la science ouverte et collaborative. Ils sont

1. <https://gallica.bnf.fr/edit/und/a-propos>

2. Le terme « océrisation » dérive de l'abréviation OCR : *Optical Character Recognition*, c'est-à-dire en français : « Reconnaissance optique des caractères » (ROC, peu utilisé). Techniquement, il s'agit du traitement d'une image (le texte est scanné, comme par une photocopieuse) sur laquelle on fait intervenir un logiciel de reconnaissance de caractères : le logiciel déchiffre les formes et les traduit en lettres.

3. <https://books.google.com/ngrams>

4. François Héran, « Les mots de la démographie des origines à nos jours : une exploration numérique », *Population*, vol. 70, 2015, p. 525-566.

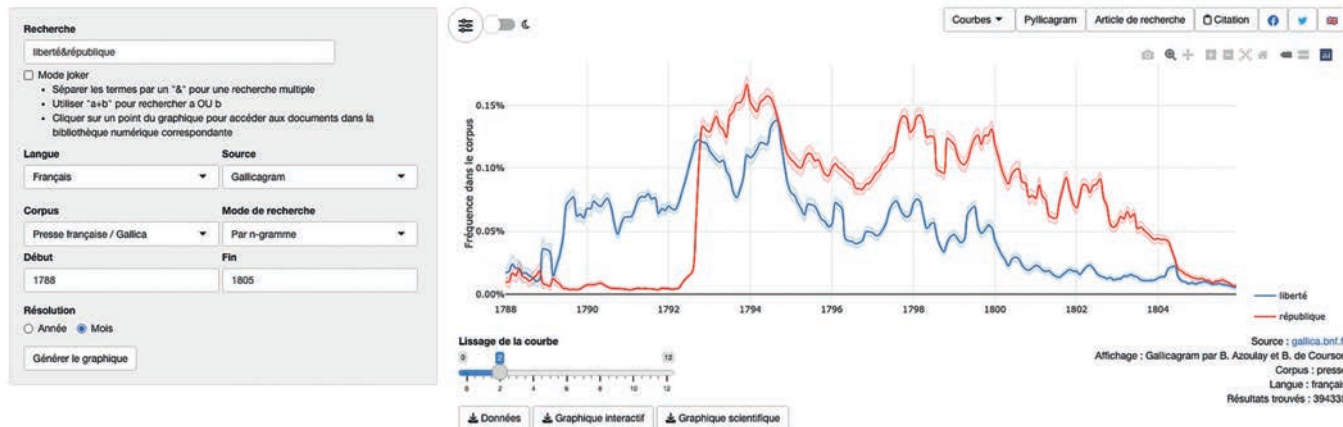
5. <https://www.frantext.fr>

6. Nous développons ce point dans l'article de Benoît de Courson, Benjamin Azoulay, Clara de Courson, Laurent Vanni et Étienne Brunet, « Gallicagram : les archives de presse sous les rotatives de la statistique textuelle », *Corpus*, n° 24, 2023. <https://doi.org/10.4000/corpus.7944>

7. <https://shiny.ens-paris-saclay.fr/app/gallicagram>

8. <https://www.gallicagrapher.com/>

9. Une API (*Application Programming Interface* ou « interface de programmation d'application ») est une interface logicielle qui permet de « connecter » un logiciel ou un service à un autre logiciel ou service afin d'échanger des données et des fonctionnalités.



Interface de Gallicagram : <https://shiny.ens-paris-saclay.fr/app/gallicagram>

Page 1 de 645 > >>

Document	Gallica	Contexte gauche	Pivot	Contexte droit
> (5) Le Mémorial des Pyrénées : polittk 1888-08-07	Image		GRÈVES	A PARIS Les terrassiers La réunion quotidienne des ouvriers grévistes terrassier
> (7) Le Petit républicain : journal quoti 1888-08-03	Image	donner leur adhésion à la	grève	des terrassiers. les cochers de fiacre Après deux discours des citoyens Winant e
> (4) Les Chantiers de l'Exposition univ 1888-08-01	Image	La	Grève	et l'Exposition de 1889. —Les Chantiers de Paris. — Les Travaux de Paris. — Une
> (12) Le Cri du peuple : journal politiq. 1888-08-01	Image	LA	GRÈVE	DES MINEURS (De notre correspondant) Saint-Etienne, 30
> (1) La Lanterne de Boquillon / par A. I 1888-08-19	Image	mirent en	grève	et les garçons de café «c extras », battus par Lozé II à Aboukir, se jurèrent de
> (7) Le Clairon du Lot : journal monarc 1888-08-07	Image	que les progrès faits par la	grève	dans d'autres corporations ne pouvaient qu'encourager les terrassiers à persévér
> (8) Le Cri du peuple : journal politiq. 1888-08-21	Image	La	grève	doit se transformer, comme s'est transformée l'industrie et la guerre elle-même
> (9) Le Cri du peuple : journal politiq. 1888-08-02	Image		GRÈVE	ET SYNDICAT Huit mille ouvriers se sont mis en
> (10) Le Radical algérien : paraît tous l 1888-08-05	Image	E» c est comme ce a partout 1 V La	grève	des mineurs. — On écrit du Saint Etienne
> (7) Le Cri du peuple : journal politiq. 1888-08-03	Image	UNE NOUVELLE	GRÈVE	Les verriers à vitres de Saint-Etienne Le verre à vitre. — L'usine Velin. — Un r
> (8) La Presse 1888-08-02	Image	Une	grève	se résume toujours pour les travailleurs par ces trois phases: Plus de salaires,
> (9) L'intransigeant 1888-08-06	Image	Le jour où toutes les corporations du bâtiment se mettraient simultanément ^n	grève	, on verrait ce que pèse le capital devant le travail, et il faudrait bien que l
> (8) Le Radical 1888-08-09	Image	LE DEVOIR DES PATRONS La	grève	des terrassiers semble entrer enfin dans la voie de l'arrangement
> (10) Le Cri du peuple : journal politiq. 1888-08-11	Image	La	grève	des ouvriers verriers. — Im- portante réunion ouvrière. — Pas de désordres
> (7) Le Cri du peuple : journal politiq. 1888-08-14	Image	Les terrassiers en	grève	ne font que réclamer do leurs exploiters les salaires et les conditions que le

Open Source¹⁰, ce qui permet le réemploi du code dans le cadre de projets tiers. Gallicagram est aussi *Open Data*: la base de données constituée par le décompte des milliards de mots des corpus numérisés de Gallica (presse et livres français) est accessible par API¹¹. Les deux logiciels collaborent puissamment à travers leurs API respectives: Gallicagram fournit ses données au graphique affiché par Gallicagrapher, qui lui renvoie le contexte des occurrences (figure ci-dessus).

Ceux-ci étant destinés à une population de chercheurs inégalement à l'aise avec l'informatique, l'ergonomie est un enjeu central. Gallicagrapher ne fait au fond qu'enrober les API de Gallica pour présenter de façon plus intuitive et plus maniable des données que le chercheur aurait pu trouver manuellement.

Mais il y a fort à parier que sans ce tour de force ergonomique, il n'en aurait pas eu le courage. Notons aussi que le développement de telles applications Web interactives a été facilité par les avancées récentes des langages de programmation (respectivement Shiny et React pour les deux logiciels).

Des obstacles persistants à l'ouverture des données

Ces logiciels se sont développés en tirant profit des données ouvertes de Gallica – mais aussi, avouons-le, en contournant les barrières à l'entrée de sources qui le sont moins.

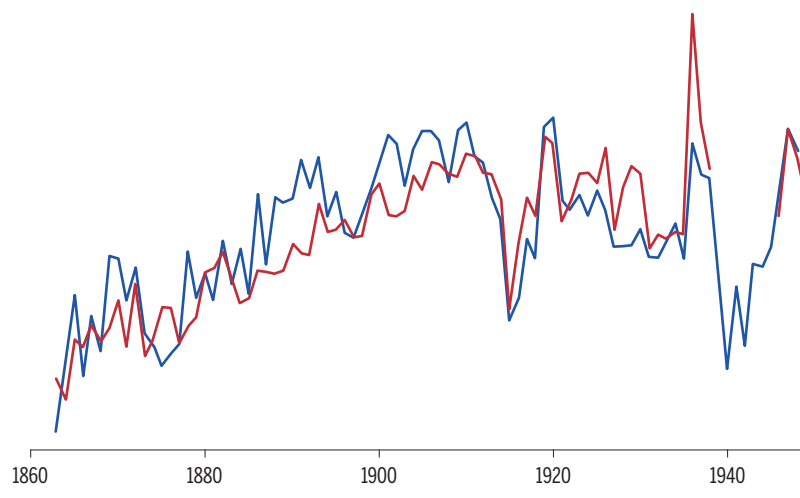
Nous nous sommes surtout heurtés à Retronews, un service développé par BnF-Partenariats pour

Fonction « contexte » de Gallicagrapher : <https://www.gallicagrapher.com/> Gallicagram utilise l'API de Gallicagrapher pour proposer une présentation analogue sous ses graphiques.

10. https://github.com/regicid/docker_gallicagram, <https://github.com/gleasonw/gallica-grapher>

11. <https://github.com/regicid/pyllicagram>

Superposition de la fréquence occurrence de « grève(s) » dans la presse de Gallica avec le nombre de grèves effectivement recensées



monétiser certaines archives de presse. Outre l'aspect étonnant du rétablissement des droits d'auteur s'agissant de documents tombés dans le domaine public, la valeur ajoutée du site est faible (« ré-reconnaissance » optique des caractères des fichiers numérisés et interface graphique plus moderne). Si Retronews n'avait été qu'un doublon de Gallica, le mal serait bénin ; mais la BnF a engagé un authentique versement de ses archives numériques vers l'opérateur, et ce pour une bonne moitié des journaux de la collection. L'exploitation d'une option cachée du site Retronews nous a permis d'extraire massivement le texte de ces journaux, sans lesquels notre base serait largement incomplète. Lorsque l'on effectue une recherche avancée sur Gallica, le contexte des occurrences issues de documents de Retronews est masqué¹². Pourtant, les API de Gallica pouvaient y accéder – ce blocage semble donc délibéré. La fonction « contexte » de Gallicagrapher permet d'en rétablir l'accès en exploitant cette bizarrerie.

Par ailleurs, l'exploitation statistique de corpus soumis aux droits d'auteur se situe dans une zone grise juridique. Gallicagram ne diffuse que des nombres (les fréquences d'occurrence des mots au cours du temps), et non les textes eux-mêmes. Les ayants droit ne sont donc pas lésés, et l'on peut penser que notre usage tomberait dans l'exception au droit d'auteur « fouille de textes »¹³, récemment introduite en droit français¹⁴. Mais l'incertitude demeure, et ce flou pousse aujourd'hui de nombreux chercheurs à

L'exploitation statistique de corpus soumis aux droits d'auteur se situe dans une zone grise juridique. Gallicagram ne diffuse que des nombres (les fréquences d'occurrence des mots au cours du temps), et non les textes eux-mêmes. Les ayants droit ne sont donc pas lésés.

l'autocensure. Il est heureux que Retronews, contacté après l'extraction des données (*scraping*)¹⁵, ait décidé de ne pas chercher noise à deux jeunes développeurs bénévoles. Mais il serait préférable qu'une doctrine plus claire apaise la conscience et les nuits des chercheurs. ■

12. Par exemple : <https://t.ly/ghX1>

13. Voir dans ce numéro l'article de Didier Thebault, p. 85.

14. Articles L.122-5-3 et R.122-23 du Code de la propriété intellectuelle. Décret n° 2022-928 du 23 juin 2022 portant modification du Code de la propriété intellectuelle. Ordonnance n° 2021-1518 du 24 novembre 2021.

15. Le *Web scraping* est une technique permettant l'extraction des données d'un site via un programme, un logiciel automatique ou un autre site.

L'objectif est donc d'extraire le contenu d'une page d'un site de façon structurée.

Le *scraping* permet ainsi de pouvoir réutiliser ces données.

Faites parler vos données!

Le *Text Mining* (la fouille de textes) en documentation

La fouille de textes, également connue sous le nom de *Text Mining*, est devenue l'un des outils les plus importants pour extraire des informations de grandes quantités de données textuelles. Cette technique permet aux chercheurs, entreprises et gouvernements de découvrir des informations, des tendances, des expressions d'opinions et autres ressentis, souvent dispersés dans des données textuelles plurielles, telles que les articles de presse, les messages sur les réseaux sociaux et les documents gouvernementaux.

La fouille de textes est principalement une technique d'analyse de données qui permet de traiter et d'extraire des informations à partir de textes non structurés. Elle est de plus en plus utilisée dans le domaine de la documentation pour améliorer la recherche d'informations et optimiser la gestion de la connaissance, mais avec quels usages, quelle méthodologie et quelles compétences mobilisées ?

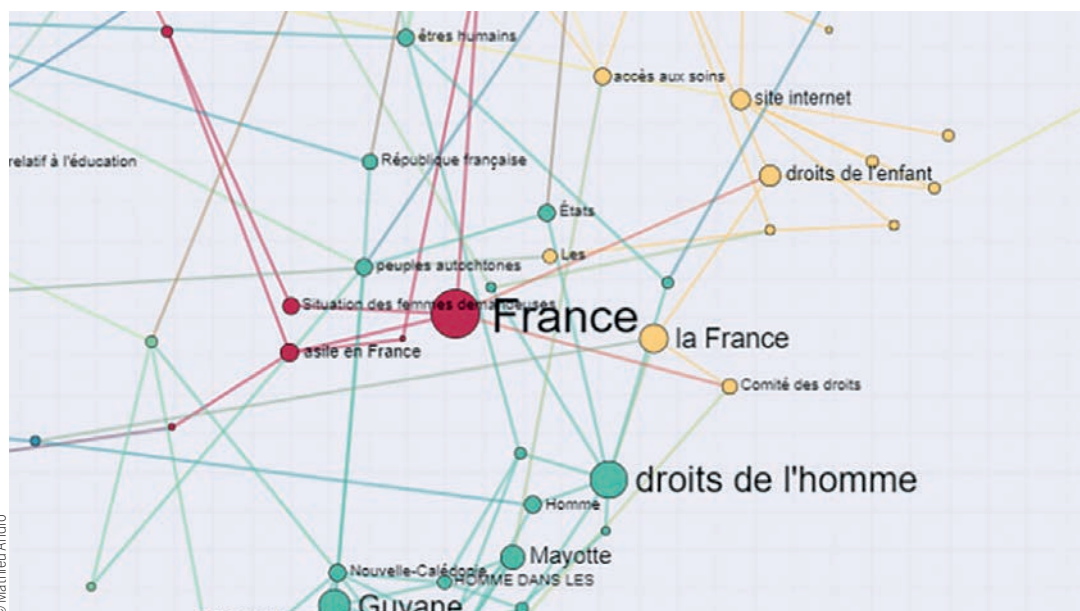
Quels usages ?

Les applications de la fouille de textes en documentation sont nombreuses.

Elles permettent, notamment, d'extraire des informations clés d'un document pour en faciliter la compréhension et la mémorisation ; d'identifier les termes les plus importants d'un document pour les utiliser comme mots-clés pour les indexer ; de détecter les thèmes principaux d'un corpus de documents pour faciliter la navigation ; d'identifier puis d'extraire des

DIDIER THEBAULT

Responsable de projet, Secrétariat Général, Département des études, de la prospective, des statistiques et de la documentation, Mission de la politique documentaire, ministère de la Culture



informations contextuelles pour établir des liens entre différents documents.

Elle améliore la qualité des informations disponibles, facilite l'accès à l'information pertinente et favorise la collaboration entre les utilisateurs. Cependant, il est important de noter qu'elle nécessite une certaine expertise pour être correctement utilisée et que les résultats obtenus peuvent varier en fonction de la qualité des données utilisées.

Quelles étapes clés de la fouille de textes? Mode d'emploi

Les étapes clés de la fouille de textes comprennent généralement cinq phases:

- la collecte de données: recueillir des données textuelles à partir de différentes sources, telles que des articles de presse, des blogs, des forums en ligne, etc.;
- le prétraitement: nettoyer et préparer les données pour l'analyse en utilisant des techniques telles que la segmentation de phrases, la *tokenisation*¹, la suppression des mots vides et la normalisation des données;
- l'analyse des données: appliquer des techniques statistiques et d'apprentissage automatique pour extraire des informations utiles à partir des données textuelles, telles que la classification de documents, l'extraction de thèmes et l'analyse de l'expression de sentiments et ressentis;
- la visualisation: présenter les résultats de l'analyse sous forme de graphiques et de tableaux pour

faciliter la compréhension et la communication des résultats;

- l'interprétation des résultats: interpréter les résultats de l'analyse et en tirer des conclusions utiles à destination des utilisateurs. La fouille de textes demeure un procédé itératif.

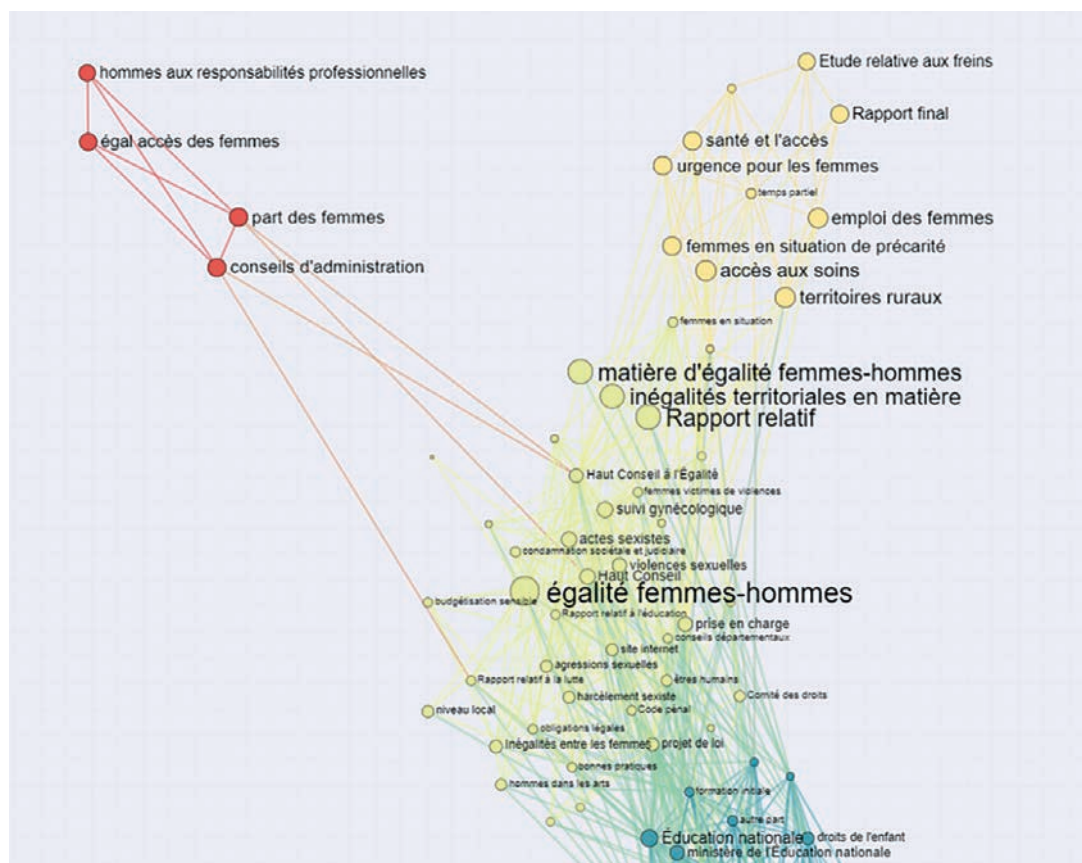
La part du documentaliste! Évolution du métier et nouvelles compétences

Le métier de documentaliste est en constante évolution. Les avancées technologiques, notamment dans le domaine de l'intelligence artificielle et du traitement automatique du langage naturel, ont un impact important sur la façon dont les documentalistes collectent, organisent et diffusent l'information.

Il est probable que l'utilisation de l'intelligence artificielle et de la fouille de textes continuera à croître, permettant aux documentalistes de traiter de plus grandes quantités de données, plus rapidement et de manière plus efficace. Ces outils pourraient également permettre une meilleure analyse des données pour identifier les tendances qui se dégagent plus fortement et les thèmes pertinents pour les utilisateurs.

Les documentalistes continueront ainsi à jouer un rôle clé dans la gestion de la connaissance au sein des structures de recherche et des entreprises, en aidant les acteurs à trouver des informations répondant au plus près à leurs besoins et susceptibles d'accompagner des décisions éclairées.

1. La *tokenisation* permet de réduire l'exposition des données sensibles à des applications, utilisateurs ou processeurs et ainsi limiter les risques. Une fois *tokenisées*, les données critiques sont réputées désensibilisées puis peuvent circuler et être traitées ou stockées dans les systèmes.



Retour d'atelier de fouilles de textes : méta-analyse (combinaison de résultats statistiques d'une série d'études indépendantes sur un problème donné) du corpus de documentation administrative.

Ils pourraient également endosser un rôle plus incitatif et actif dans la création de contenus, en utilisant les données collectées pour créer des guides, des tutoriels et d'autres types de contenus, destinés à aider les utilisateurs à tirer le meilleur parti des informations disponibles.

Nul doute qu'en étant de plus en plus impliqués dans la gestion des données de manière responsable, ils pourront contribuer à l'amélioration de la qualité des opérations relatives à la fouille de textes.

Ces perspectives prometteuses offrent de nombreuses opportunités pour les professionnels qui sont prêts à s'emparer des nouvelles technologies et à explorer ces nouveaux rôles. Dans ce cadre, la formation des documentalistes à ces évolutions s'avère essentielle; s'en saisir est une ouverture pleine de potentialités.

Le statut d'exception de la fouille de textes dans le droit

Un décret précise les exceptions pour la fouille de textes et de données, suscitant l'inquiétude chez les éditeurs en France. L'article d'Antoine Oury dans *Actualité*² le présente ainsi (extraits) :

« Par ordonnance, et avec plusieurs mois de retard, la France avait finalement transposé les dispositions de la directive européenne sur le droit d'auteur de 2019, notamment celles instituant des exceptions au droit d'auteur en vue de la fouille de textes et de données.

Fin 2021, cette ordonnance indiquait que la fouille de textes et données *ne pouvait faire l'opposition des titulaires de droit d'auteur*, dès lors qu'elle est effectuée «aux seules fins de la recherche scientifique par les organismes de recherche, les bibliothèques accessibles au public, les musées, les services d'archives ou les institutions dépositaires du patrimoine cinématographique, audiovisuel ou sonore, ou pour leur compte et à leur demande par d'autres personnes, y compris dans le cadre d'un partenariat sans but lucratif avec des acteurs privés». »

Quid de l'utilisation de la fouille de textes en science ouverte ?

La fouille de textes est un domaine en pleine expansion qui utilise des techniques de traitement automatique du langage naturel pour extraire des informations à partir de textes bruts.

La science ouverte est un levier qui vise à rendre les données scientifiques et les résultats de recherche accessibles à tous, et l'utilisation de la fouille de textes y joue, par conséquent et nécessairement, un rôle important.

La science ouverte englobe, on le sait, des domaines aussi variés de la connaissance que les sciences

L'un des principaux avantages de l'utilisation de la fouille de textes en science ouverte est la possibilité de capter des « signaux faibles », de découvrir des tendances (évolutions des enjeux sociétaux) et des relations intrinsèques entre les données scientifiques, qui passeraient autrement inaperçues.

humaines et sociales, les sciences économiques, les sciences de la santé, les sciences de l'environnement et les sciences de la vie. Les données scientifiques sont souvent disponibles sous forme de textes tels que des articles de revues, des rapports de recherche, des thèses et des brevets. La fouille de textes permet de traiter ces textes et d'extraire des informations ciblées.

L'un des principaux avantages de l'utilisation de la fouille de textes en science ouverte est la possibilité de capter des « signaux faibles », de découvrir des tendances (évolutions des enjeux sociétaux) et des relations intrinsèques entre les données scientifiques, qui passeraient autrement inaperçues.

Les chercheurs peuvent ainsi identifier les thèmes les plus couramment mentionnés dans un corpus de documents scientifiques, les auteurs les plus prolifiques dans un domaine donné, ou encore les publications les plus citées.

Enfin, la fouille de textes peut également être utilisée pour automatiser les tâches fastidieuses et chronophages telles que la classification de documents, la segmentation de phrases, la reconnaissance de la langue, la génération de textes et la détection de modes de pensée. Cela permet aux chercheurs de consacrer plus de temps à l'analyse qualitative des résultats plutôt qu'à la collecte quantitative de données.

La fouille de textes est devenue une pratique qualitative venant enrichir les outils qui sous-tendent et alimentent les manières de produire de la recherche en science ouverte.

Elle semble ainsi promise à un bel avenir... ■

2. Antoine Oury, *Actualité* du 24 juin 2022 : <https://actualitte.com/article/106646/droit-justice/un-decret-precise-les-exceptions-en-vue-de-la-fouille-de-textes-et-de-donnees>

BigScience, une collaboration internationale à l'origine du plus gros modèle de langue multilingue ouvert au monde

Fin 2022, la présentation du système conversationnel ChatGPT par le consortium américain OpenAI a montré l'avancement des technologies de simulation d'expression humaine. ChatGPT, avec sa capacité à produire du texte en réponse à une sollicitation (*prompt*) dans plus d'une dizaine de langues dont le français, a impressionné, effrayé, interrogé bien au-delà du secteur de l'innovation.

LUCIE GIANOLA

Chargée de mission pour les technologies, la recherche et l'innovation, Délégation générale à la langue française et aux langues de France, Mission des langues et du numérique, ministère de la Culture

Objet de nombreuses spéculations, ChatGPT n'est pourtant qu'un exemple des avancées de la recherche en traitement automatique des langues (TAL). Son retentissement tient en grande partie à sa mise à disposition du public sous la forme d'une interface facile d'utilisation, gratuite, qui permet à tout un chacun de laisser libre cours à ses expérimentations du modèle : on a vu circuler sur les réseaux toutes sortes de textes et d'échanges, poèmes ou encore « pièges » demandant au système comment ramasser des œufs de vache destinés à montrer ses limites de raisonnement.

En effet, ChatGPT – et tous les systèmes de traitement de la langue – n'est rien d'autre qu'une énorme architecture statistique qui, après avoir été « entraînée » sur un volume très important de données textuelles annotées, calcule tout simplement, compte tenu du contexte, quel est le mot le plus probable à venir à la suite. Ainsi, si la syntaxe et le style du système sont impressionnants, ses capacités sémantiques et de raisonnement sont encore en dessous de celles d'un humain.

BigScience et le modèle BLOOM

L'entraînement et la gestion des données utilisées pour le développement sont fondamentaux pour la performance des modèles de langue. On fournit à l'algorithme de très grandes quantités de données annotées par des humains, qui servent d'exemples de ce qu'on en attend ou non.

Par exemple, pour le développement de ChatGPT, des annotateurs humains d'une société kenyane ont été recrutés pour annoter et ainsi filtrer les contenus toxiques (violents, sexuels, racistes, etc.), non sans susciter des questions éthiques importantes sur les conditions de travail de ces annotateurs peu rémunérés et soutenus vis-à-vis de la nature des contenus auxquels ils étaient confrontés.

L'origine des données d'entraînement de ChatGPT reste relativement floue. Lorsqu'on interroge le *chatbot*, il répond avoir été « entraîné sur un large corpus de données textuelles provenant d'Internet, comprenant des pages Web, des livres électroniques, des articles de presse, des discussions en ligne, des publications sur les réseaux sociaux et d'autres sources en langue anglaise », et que le corpus n'est « pas publiquement disponible dans son intégralité en raison de restrictions de droits d'auteur et de confidentialité ».

Des initiatives un peu moins médiatiques mais tout aussi performantes existent aussi, dont certaines s'inscrivent dans une démarche d'ouverture, de transparence et de contrôle de la qualité des données d'entraînement utilisées. C'est le cas du modèle BLOOM¹, développé par le consortium BigScience et piloté par la jeune pousse (*start-up*) franco-américaine Hugging Face. Plus de 1 000 chercheurs internationaux ont contribué au développement de BLOOM pendant un an. L'entraînement du modèle s'est fait sur

1. <https://bigscience.huggingface.co/blog/bloom>



© Cyril Frésillon CNRS Images

le supercalculateur du CNRS Jean Zay, la plus puissante infrastructure de ce type en France².

BLOOM est un modèle doté de 176 milliards de paramètres capable de générer du texte en 46 langues dites « naturelles » (langues utilisées par les humains pour communiquer entre eux) et 13 langages de programmation. Parmi les langues prises en charge, en plus du français, de l'anglais, du chinois ou encore de l'espagnol, on trouve des langues régionales, comme le catalan ou le basque, et des langues rares et peu dotées en technologies comme le kinyarwanda (14 millions de locuteurs), le népalais (16 millions de locuteurs), le xhosa (langue à clics bantoue parlée en Afrique du Sud par environ 8 millions de locuteurs) ou le fon (langue parlée au Bénin, au Nigeria et au Togo par 4 millions de locuteurs).

Parmi les corpus utilisés pour l'entraînement de BLOOM, on trouve un corpus comprenant plus de 360 langues issues de BloomLibrary.org³, un site compilant des livres et histoires en accès ouvert dans plus de 500 langues, le corpus OSCAR⁴, un corpus ouvert comprenant 166 langues destiné à l'apprentissage automatique, ou encore The Pile⁵, un corpus agrégeant 22 sources comme le corpus multilingue Europarl (actes du Parlement européen) ou le projet Gutenberg (littérature du domaine public).

L'intérêt des données ouvertes pour les technologies

L'exemple de BLOOM illustre bien l'intérêt de poursuivre le développement de ressources linguistiques de qualité, de volumes importants et surtout multilingues. Le projet a pu capitaliser sur un certain nombre de ressources préexistantes, construites avec rigueur et comprenant des langues qui ne bénéficient pas des avancées technologiques produites par les

acteurs commerciaux. Or, on sait aujourd'hui qu'une langue ne disposant pas de technologies numériques est une langue en danger : ne pouvant être utilisée dans la sphère numérique, sur Internet, à l'aide d'un smartphone ou d'une tablette, son espace d'utilisation est réduit.

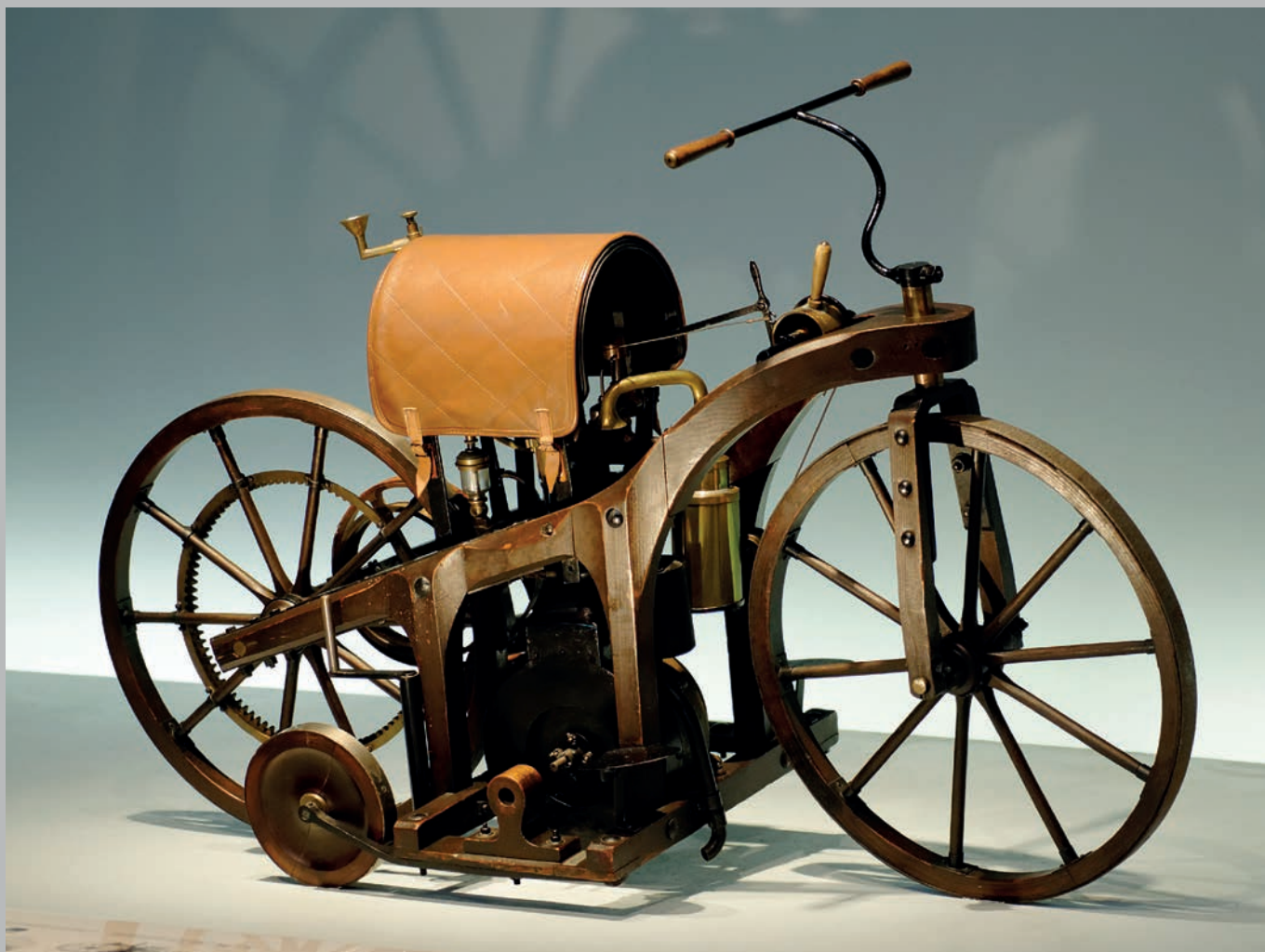
BLOOM représente donc une vision très différente des technologies de la langue et de l'intelligence artificielle de celle portée par les géants américains tels qu'OpenAI, Facebook (qui vient de présenter son modèle « façon » ChatGPT, LLaMA), Google ou Microsoft, qui produisent des modèles dont l'architecture et les données d'entraînement ne sont pas transparentes.

Outre les enjeux de diversité linguistique que nous avons évoqués se trouvent aussi des enjeux de souveraineté numérique. Il faut donc souhaiter que BLOOM puisse trouver son marché et ses applications dans le contexte européen. ■

Le « supercalculateur Jean Zay » du CNRS, dans les locaux de l'Institut du développement et des ressources en informatique scientifique (IDRIS) sur le campus de l'Université Paris-Saclay.

- <https://www.cnrs.fr/fr/jean-zay-le-supercalculateur-le-plus-puissant-de-france-pour-la-recherche>
- <https://huggingface.co/datasets/sil-ai/bloom-lm>
- <https://oscar-project.org/>
- <https://pile.eleuther.ai/>

On sait aujourd'hui qu'une langue ne disposant pas de technologies numériques est une langue en danger : ne pouvant être utilisée dans la sphère numérique, sur Internet, à l'aide d'un smartphone ou d'une tablette, son espace d'utilisation est réduit.



Prototype de moto à moteur à gaz de 1885,
des inventeurs allemands Gottlieb Daimler et
Wilhelm Maybach, Musée Mercedes-Benz à Stuttgart

© RaBoe, Creative Commons BY-SA 3.0 DE

https://commons.wikimedia.org/wiki/File:Mercedes-Benz-Museum_2015-01_by-RaBoe_027.jpg

De nouveaux modes de transmission et de diffusion

Pourquoi le développement de la science ouverte a-t-il impérativement besoin d'éditeurs?

Les éditeurs jouent un rôle essentiel pour assurer la qualité des publications scientifiques et la plus large diffusion des connaissances. Et la remise en cause par la « stratégie de non-cession des droits » du droit d'auteur au fondement de leur travail risque de les fragiliser durablement. C'est pourquoi elle n'est pas la meilleure voie pour assurer le développement de l'accès ouvert (*Open Access*), ni pour garantir la liberté d'expression des chercheurs.

FRANÇOIS GÈZE

Président du groupe des éditeurs universitaires du Syndicat national de l'édition

Lancée en 2002 pour réagir à la hausse du prix des abonnements aux revues scientifiques STM (sciences, technique et médecine)¹ publiées par les grands éditeurs anglophones, l'*Initiative de Budapest pour l'accès ouvert*² s'est depuis élargie à la promotion de la « science ouverte », devenue désormais largement consensuelle, dans le monde de la recherche comme dans celui de l'édition scientifique.

En France, sa mise en œuvre par les acteurs publics de la recherche a toutefois donné lieu à de nombreux débats, toujours vifs à ce jour. On l'a vu notamment lors de l'élaboration de la loi pour une République numérique d'octobre 2016³, dont un article a stipulé que l'auteur d'un écrit scientifique financé sur fonds publics a le « droit de mettre à disposition gratuitement dans un format ouvert, par voie numérique, [...] la version finale de son manuscrit acceptée pour publication », au plus tôt six mois après celle-ci pour les STM et douze mois pour les sciences humaines et sociales (SHS).

Un an plus tôt, une vingtaine d'éditeurs généralistes (dont Gallimard, Le Seuil, La Découverte ou Minuit) ou spécialisés en SHS avaient pourtant alerté dans une tribune :

« En invitant le législateur à déterminer une période d'embargo la plus courte possible, il n'est pas certain que ce projet s'engage dans la bonne direction. Le remède proposé, la gratuité de l'accès à ces écrits, risque d'être pire que le mal, leur « capture abusive ». [...] L'espace du numérique ne doit pas relever d'un système unique: il



À livre ouvert.

© Pexels-Pixabay (<https://www.pexels.com/fr-fr/photo/illustration-de-livres-ouverts-rouges-261806/>)

doit, au contraire, être un lieu d'expérimentation, d'écosystèmes hybrides où coexistent des modèles économiques et techniques mixtes, associant le payant et le gratuit, modèles qui doivent être explorés et expérimentés sans *a priori* doctrinaires grâce à un patient et rigoureux travail commun entre acteurs privés et publics⁴. »

Force est de constater que ce propos modéré n'a guère été entendu, même si le Plan pour la science ouverte (PNSO) adopté par le ministère de l'Enseignement supérieur et de la Recherche (MESR) en 2018 (et prolongé en 2021) affirme vouloir « préserver un objectif de bibliodiversité ».

1. Voir dans ce numéro l'article d'Odile Contat et Anne-Solweig Gremillet, p. 8.

2. <https://www.ouvrirlascience.fr/initiative-de-budapest-pour-laccès-ouvert/>

3. <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000033202746>

4. Olivier Bétourné *et al.*, « L'édition de savoir ne doit pas être soumise au seul modèle de la gratuité », *Le Monde*, 7 octobre 2015.



Indispensable échelle.

À l'inverse, on a vu depuis le CNRS et plusieurs universités obliger leurs chercheurs à déposer dès leur parution leurs publications sur une archive ouverte, en parfaite contradiction avec la loi. Suivant en cela les préconisations du MESR, ils ont promu, comme l'Agence nationale de la recherche (ANR), la « stratégie de non-cession des droits » des chercheurs consistant à publier leurs écrits en accès gratuit, sous licence Creative Commons⁵. Cela afin de contourner la disposition de la loi de 2016 (inscrite dans le Code de la recherche, article L. 533-4).

Ces décisions apparaissent comme le fruit de pré-supposés loin d'être démontrés. La première justification avancée est souvent d'élargir, par la gratuité, l'accès aux connaissances scientifiques et de favoriser ainsi la science collaborative et citoyenne. Une justification pourtant assez illusoire. Car, d'une part, la plupart des écrits scientifiques publiés dans les revues académiques sont incompréhensibles par le commun des mortels, dès lors qu'ils recourent inévitablement à des concepts et des formulations réservés aux spécialistes. D'autre part, la gratuité des écrits scientifiques ne garantit en rien leur large diffusion dans le monde académique : plusieurs études, certes à compléter, montrent que la notoriété de la revue compte bien plus que son caractère payant ou non⁶, et que la simplicité de l'accès en ligne est un facteur d'audience déterminant – comme en atteste le fait que les chercheurs français préfèrent majoritairement recourir au moteur de recherche Google Scholar (ou aux réseaux sociaux académiques privés comme ResearchGate

et Academia) et au site pirate SciHub pour accéder aux travaux qui les intéressent, plutôt qu'aux outils des organismes publics ou à des « archives ouvertes » comme HAL⁷.

La seconde justification avancée par certains acteurs publics français est que l'accès ouvert permettrait de réduire le coût excessif des abonnements aux publications scientifiques numériques. Mais une étude publiée par le MESR montre que cet espoir est infondé : elle indique que, en 2020, les dépenses françaises d'abonnements aux périodiques électroniques étaient de 87,50 M€ et les dépenses de financement des coûts (*Articles Processing Charges*, « frais de publication ») pour les articles publiés en libre accès de 30,10 M€ ; et que ces deux postes, loin que le premier diminue à mesure que le second augmente, passeraient en 2030 (à évolution identique) respectivement à 97,50 M€ et 50,60 M€ (soit une augmentation du total de 26 %) ⁸. Ce qui n'est pas surprenant, puisque la France contribue environ à hauteur de 2,5 % des articles publiés dans le monde et à seulement 2,2 % des dépenses d'abonnements aux publications scientifiques mondiales : le budget de financement en amont de la totalité des coûts d'édition des articles de chercheurs français pour les rendre gratuits serait donc nécessairement supérieur à celui actuellement consacré aux abonnements.

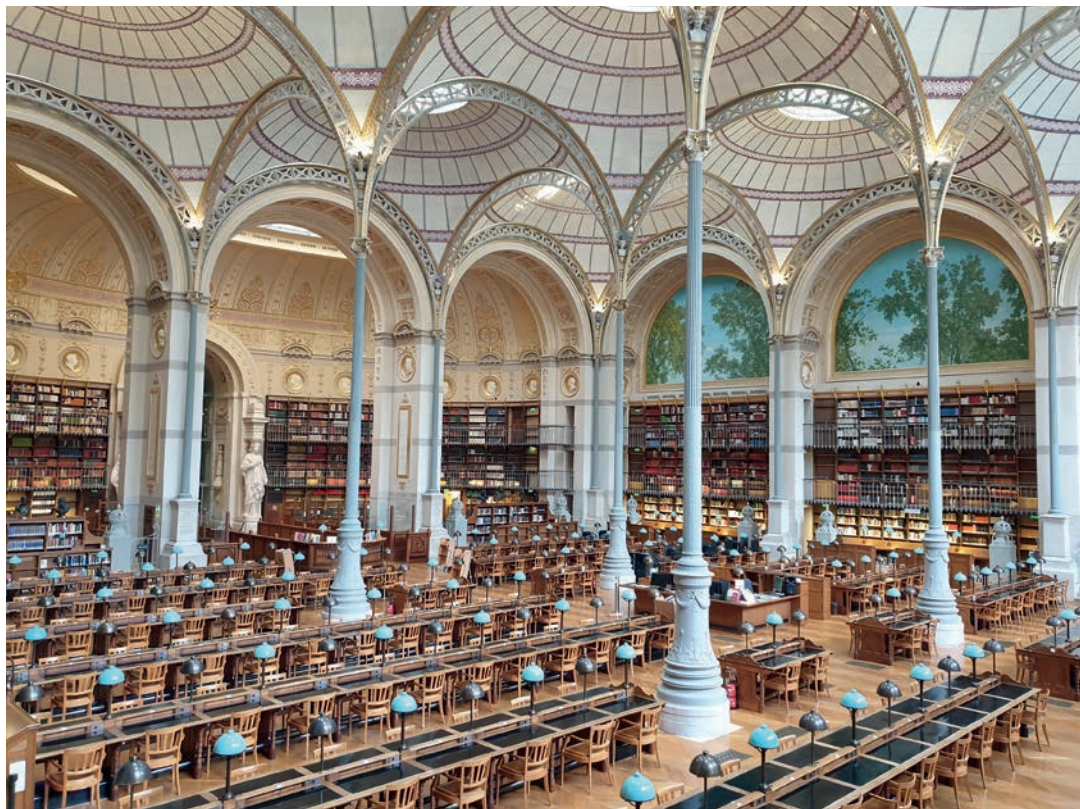
C'est notamment ce qui rend hautement aléatoire le succès du modèle de publication scientifique en accès ouvert dit « diamant » (sans paiement par le lecteur ni par l'auteur), aujourd'hui promu par le MESR. Dans ce modèle, l'indispensable travail d'édition serait financé

5. Alain Schuhl, « Il n'y a pas de raison que les scientifiques fassent une cession exclusive gratuite de leurs œuvres aux éditeurs », *CNRS Info*, 1^{er} décembre 2022.

6. Voir notamment Abdelghani Maddi et David Sapinho, « Does open access really increase impact? A large-scale randomized analysis », 26th International conference on science and technology indicators, septembre 2022. <https://www.hal.inserm.fr/CAMPUS-CONDORCET/hal-03694666v1>

7. Françoise Rousseau-Hans *et al.*, *Les pratiques de publications et d'accès ouvert des chercheurs français en 2019*, Couperin, juin 2020. <https://hal.science/cea-02450324>

8. Antoine Blanchard *et al.*, *Retrospective and Prospective Study of the Evolution of APC Costs and Electronic Subscriptions for French Institutions*, Comité pour la science ouverte, décembre 2022. <https://hal-lara.archives-ouvertes.fr/OUVRIR-LA-SCIENCE/hal-03909068>



Salle Labrouste, Bibliothèque nationale de France.

© Pecardille/Creative Commons BY-4.0 (https://commons.wikimedia.org/wiki/File:Salle_Labrouste_BNF_INHA.jpg)

9. Pierre Henriët, Laure Darcos et Pierre Ouzoulias, *Pour une science ouverte réaliste, équilibrée et respectueuse de la liberté académique*, rapport de l'Office parlementaire d'évaluation des choix scientifiques et technologiques (OPECST), mars 2022.
10. Jean-Philippe Mochon, *Avis du médiateur du livre sur l'édition scientifique dans le contexte des politiques en faveur de la science ouverte*, 13 avril 2023.
11. Voir dans ce numéro, l'article de Christine Cherbut, p. 95.
12. Voir dans ce numéro l'article de David Pouchard, p. 97.

par des « sponsors » ou par l'État, sur un hypothétique budget très supérieur à ses moyens. De plus, comme l'a souligné en 2022 un rapport parlementaire, « en reposant exclusivement sur des subventions publiques, ce modèle économique s'apparente à une étatisation pure et simple de l'édition du savoir⁹ ». Et c'est bien là que le bât blesse. Car de fait, les promoteurs de l'accès ouvert « à la française » semblent considérer que le travail de l'éditeur scientifique est inutile, voire parasite, comme si le numérique permettait miraculeusement de supprimer les tâches qu'il assure classiquement : validation éditoriale des textes proposés, vérification des informations, corrections stylistiques, mise aux normes typographiques, diffusion large, etc. Tout cela a un coût. *A contrario*, le mythe du « tout

gratuit » cache souvent des dépenses en salaires de fonctionnaires et mène de plus à une récupération de ces contenus gratuits par les GAFAM pour leur propre profit.

En France, les éditeurs scientifiques qui assurent ces tâches publient surtout des SHS et sont très majoritairement privés. Leur diversité est une garantie de liberté d'expression pour les auteurs, de même qu'elle assure efficacement la diffusion des connaissances au-delà de cercles académiques (par des revues et livres de vulgarisation), ce qui n'a jamais été du ressort des éditeurs publics, comme l'a souligné le médiateur du livre dans son avis d'avril 2023¹⁰.

C'est pourquoi les éditeurs du Syndicat national de l'édition (SNE) et de la Fédération nationale de la presse d'information spécialisée (FNPS) sont convaincus que l'avancée de la science ouverte ne sera possible qu'à deux conditions : rechercher les voies de la transition en ce sens par des travaux « communs entre acteurs privés et publics » comme ceux conduits depuis 2022 par l'Observatoire de l'édition scientifique¹¹ ; préserver à tout prix le droit d'auteur, condition d'une science ouverte de qualité à la différence de la « stratégie de rétention des droits », comme le montre l'exemple d'un compromis législatif intelligent trouvé pour les images illustrant les écrits scientifiques¹². ■



Les multiples voies de l'édition scientifique.

© Eric Carza/Pexels-Pixabay (https://www.pexels.com/fr-fr/photo/circulation-autoroute-longue-exposition-rapide-5395585/)

L'Observatoire de l'édition scientifique (OES)

Créé en décembre 2021 pour cinq ans, à l'initiative du ministère de l'Enseignement supérieur et de la Recherche et du ministère de la Culture, l'Observatoire de l'édition scientifique (OES) est un lieu de concertation et de dialogue entre les acteurs de l'édition scientifique, privés et publics, le monde de la recherche et de celui de l'information scientifique. Il fait suite au comité de suivi de l'édition scientifique¹ (2017-2019), chargé d'observer les effets de la mise en place de la loi de 2016 dite « pour une République numérique »².

Une instance d'observation et de dialogue

La mission de l'OES est d'enrichir la connaissance sur l'édition scientifique et d'éclairer les orientations en matière de soutien et de développement de l'édition scientifique française. Dans ce but, l'OES réalise des études et collecte des données sur les revues et ouvrages scientifiques et sur leur évolution dans un contexte d'essor accéléré des principes de la science ouverte, en créant les conditions d'un dialogue constructif entre les différents acteurs. La composition de l'OES reflète la diversité et le pluralisme de ces acteurs, apportant expertise complémentaire et richesse des points de vue. Y sont représentés de manière paritaire les éditeurs privés et publics, d'une part, et les usagers de l'édition scientifique que sont les chercheurs et les professionnels de l'information scientifique, d'autre part.

Tous les pans de la recherche et de la diffusion des savoirs ont connu, au cours des dernières décennies, des bouleversements importants. La numérisation de l'information, l'ouverture des données de la recherche, l'essor de la science ouverte et citoyenne forment un environnement nouveau pour l'ensemble des acteurs et organisations. Dans ce contexte en évolution rapide, l'édition scientifique est directement et particulièrement concernée par la transformation des modèles de production et de diffusion des résultats de la recherche, comme cela a été souligné dans deux récents rapports de l'OPECST³ et du médiateur du livre⁴.

Une première feuille de route ambitieuse

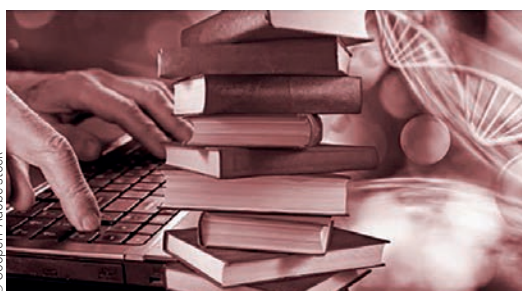
Pour accompagner ces changements, il est nécessaire d'objectiver les analyses, en les appuyant sur des données précises, produites dans un cadre méthodologique solide. Dans la même logique, l'élaboration de définitions et de références communes est un passage obligé pour faciliter le dialogue et les échanges entre



Une instance d'observation et de dialogue.

© Puckillustrations-Adobe stock

© Cooperr-Adobe stock



CHRISTINE CHERBUT

Présidente de l'OES, directrice de recherche à l'Institut national de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE)

BENOÎT EPRON

Vice-président de l'OES, professeur associé à la Haute École de Gestion de Genève (HES-SO/HEG)

ANNE-SOLWEIG GREMILLET

Chargée d'études Édition scientifique et services à la recherche, MESR, secrétaire de l'Observatoire de l'édition scientifique (OES)

l'ensemble des parties prenantes. L'OES s'est donc doté d'une feuille de route ambitieuse et a constitué en son sein des groupes de travail chargés d'approfondir et de documenter des sujets pour lesquels le besoin de concertation et de connaissance est jugé prioritaire.

Un premier besoin est celui de définitions et d'indicateurs communs à l'ensemble des acteurs. Le groupe de travail se focalise actuellement sur la mise en œuvre d'indicateurs de références sur les revues scientifiques françaises, en sciences humaines et sociales et en sciences, technologies et mathématiques.

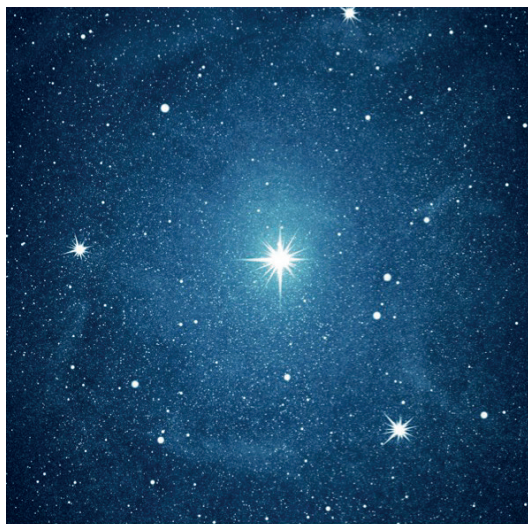
1. Daniel Renoult, *L'édition scientifique de revues : plan de soutien et évaluation des effets de la loi du 7 octobre 2016*, rapport du Comité de suivi de l'édition scientifique (CSES), 2019.

2. <https://www.legifrance.gouv.fr/dossierlegislatif/JORFDOLE00031589829/>

3. Pierre Henriot, Laure Darcos et Pierre Ouzoulias, *Pour une science ouverte réaliste, équilibrée et respectueuse de la liberté académique*, rapport au nom de l'Office parlementaire d'évaluation des choix scientifiques et technologiques, 2022. Voir dans ce numéro l'article de François Gèze, p. 92.

4. Jean-Philippe Mochon, *Avis du médiateur du livre sur l'édition scientifique dans le contexte des politiques en faveur de la science ouverte*, 13 avril 2023.

Connaissance de l'édition scientifique.



© Robert-Adobe stock

Pour une édition scientifique plurielle et rayonnante.

Le projet s'appuie sur une collecte large de données afin de qualifier et de suivre les évolutions du paysage des revues selon plusieurs variables : type de structure porteuse, modèles de diffusion et de financement, politiques de licences et de dépôts en archives ouvertes, audience de ces revues. À l'occasion de ce travail, l'OES s'est accordé sur un lexique commun et a précisé les définitions de différents termes du domaine de l'édition scientifique.

La vitalité de l'édition scientifique française, dans toute sa diversité, est une des conditions du dynamisme et du rayonnement de la recherche et de la culture francophone, ainsi que du pluralisme dans le débat démocratique.

Un autre travail de nature plus prospective vise à cerner l'évolution des pratiques des scientifiques en lien avec leur environnement documentaire, dans un contexte de science ouverte. Il s'agit d'évaluer les changements de comportements des scientifiques en tant que lectrices et lecteurs d'articles scientifiques aujourd'hui largement proposés en accès libre, d'une part; d'identifier les conditions nécessaires et les contraintes à lever pour que l'accès ouvert facilite l'utilisation des contenus, d'autre part.

Un troisième besoin concerne une meilleure connaissance de l'édition des livres scientifiques, notamment son économie. Afin d'évaluer l'impact de l'accès ouvert sur le marché des ouvrages scientifiques, une étude rassemble et intègre plusieurs types de données (ventes, accès, consultations, dissémination, disponibilité sur les plateformes, citations...) portant sur un vaste échantillon d'ouvrages disponibles en accès ouvert. L'analyse de ces données sera croisée avec une analyse plus précise des pratiques des différents acteurs concernés par les livres (bibliothécaires, éditeurs et chercheurs). Une autre étude cherche à mieux comprendre les postes de coûts des ouvrages de recherche, dans l'objectif d'aider à élaborer des modèles économiques efficaces et vertueux pour l'édition de ce type de livres, en particulier avec une visée de publication en accès ouvert pour toutes les disciplines.

Pour une édition scientifique francophone, plurielle et rayonnante

La vitalité de l'édition scientifique française, dans toute sa diversité, est une des conditions du dynamisme et du rayonnement de la recherche et de la culture francophone, ainsi que du pluralisme dans le débat démocratique. À travers ses travaux, l'ambition de l'OES est de participer activement aux réflexions sur les trajectoires que peuvent suivre les différentes parties prenantes dans un contexte d'accès ouvert, afin d'éclairer les pouvoirs publics pour l'accompagnement des éditeurs français, dans la concertation, l'anticipation et le respect des fondements économiques de leur activité. Dans ce cadre, l'OES porte une attention soutenue au respect de la liberté académique, à l'intégrité scientifique et aux droits des auteurs. ■

La diffusion ouverte des travaux de recherche intégrant des images

La récente consécration dans le Code de la propriété intellectuelle d'un nouveau dispositif visant à faciliter l'usage des œuvres relevant des arts visuels à des fins d'illustration de publications scientifiques ouvertes témoigne de ce que le droit d'auteur peut être un instrument efficace en faveur de la promotion de l'accès ouvert (*Open Access*).

Aby Warburg, *Atlas Mnémosyne*.



© Warburg Institute Archive

Les enjeux liés à l'utilisation des images dans les travaux de recherche

L'usage des œuvres de l'esprit dans un but pédagogique ou scientifique est permis lorsque la durée de protection de ces œuvres a expiré ou, s'agissant de celles qui sont encore protégées, dans le cadre de certaines exceptions au droit d'auteur. Le champ restreint de ces exceptions ne permet toutefois pas d'exploiter largement les œuvres des arts visuels (œuvres des arts graphiques, plastiques, photographiques, architecturaux, etc.) à des fins exclusives d'illustration de publications diffusées en ligne sans restriction d'accès.

Aussi des voix se sont-elles faites entendre ces dernières années pour affirmer que les difficultés d'obtention des droits d'auteur font barrage à la diffusion sous forme numérique des travaux scientifiques

qui s'appuient sur des corpus importants d'images, qu'elles soient patrimoniales ou d'actualité, ce qui empêche leur mise à disposition en accès ouvert et réduit leur visibilité et leur impact¹. Or, dans notre société de l'image, un usage des images dans un but pédagogique ou scientifique est nécessaire pour faire face au déferlement d'images, notamment détournées ou truquées.

Les coûts de transaction générés – à la fois le temps de recherche des ayants droit, qui ne sont pas toujours membres d'un organisme de gestion collective, et le paiement de droits – sont parfois jugés excessifs au regard de l'économie de la publication dans les disciplines, relevant principalement des sciences humaines et sociales (SHS), qui étudient les œuvres (histoire et théorie de l'art) ou utilisent des images

DAVID POUCHARD

Adjoint à la cheffe du bureau de la propriété intellectuelle, Secrétariat général, Service des affaires juridiques et internationales, Sous-direction des affaires juridiques, ministère de la Culture

1. Martine Denoyelle, Katie Durand, Elli Doukariidou-Ramantani et Johanna Daniel, *Droits des images, histoire de l'art et société*, rapport sur les régimes de diffusion des images patrimoniales et leur impact sur la recherche, l'enseignement et la mise en valeur des collections publiques, INHA, octobre 2018. <https://shs.hal.science/halshs-02066987>

protégées comme sources (sciences du patrimoine, archéologie, sociologie, *Visual Studies*, *Media Studies*, sémiologie, sciences de l'information et de la communication, etc.).

Les réflexions engagées en vue de répondre à la difficulté ainsi identifiée ont toutefois rapidement écarté toute solution de nature à affaiblir la protection des auteurs des arts graphiques et plastiques, dans un contexte de paupérisation marqué. Divers éléments de diagnostic permettent en effet de caractériser de façon objective une dégradation de la situation économique et sociale d'une grande partie des artistes-auteurs qui se traduit par une érosion de leurs revenus.

Le rapport du gouvernement au Parlement sur la situation des arts visuels de 2017 l'a rappelé : sur les 65 000 auteurs des arts visuels rattachés au régime social des artistes-auteurs, seuls 10 000 perçoivent un revenu supérieur à 1 430 euros par mois en moyenne sur cinq ans².

Une solution équilibrée : la licence collective étendue

En réponse à ces enjeux croisés, l'article 16 de la loi n° 2020-1674 du 24 décembre 2020 de programmation de la recherche (LPR) pour les années 2021 à 2030³ a privilégié la voie de la licence collective étendue qui paraît la mieux à même de concilier les intérêts des auteurs et ceux des chercheurs.

Cette nouvelle forme d'exercice des droits, consacrée par la directive sur le droit d'auteur et les droits voisins dans le marché unique numérique du 17 avril 2019⁴, permet à un organisme de gestion collective de conclure des autorisations d'exploitation non seulement au nom de ses membres mais aussi des titulaires de droits non adhérents, dès lors que cet organisme est représentatif des œuvres concernées.

L'article 6 de l'ordonnance n° 2021-1518 du 24 novembre 2021⁵ complétant la transposition de la directive précitée du 17 avril 2019 consacre ainsi la possibilité de recourir à une licence collective étendue autorisant l'utilisation d'œuvres des arts visuels dans des publications et travaux issus d'une activité de recherche et d'enseignement supérieur publics, cela à la double condition que ces publications et travaux soient diffusés en accès ouvert sur Internet et qu'ils soient produits et diffusés dans un cadre non lucratif.

Les travaux et publications concernés sont principalement publiés sur des plateformes de publication de revues, d'ouvrages scientifiques et de carnets de recherche en sciences humaines et sociales (OpenEdition, Cairn.info, Persée, etc.), sur un site Internet propre à une revue, un éditeur ou un chercheur, dans une archive ouverte (HAL) ou sur le site d'un établissement d'enseignement supérieur et de recherche ou d'un organisme de recherche.

En réponse à ces enjeux croisés, l'article 16 de la loi n° 2020-1674 du 24 décembre 2020 de programmation de la recherche (LPR) pour les années 2021 à 2030 a privilégié la voie de la licence collective étendue qui paraît la mieux à même de concilier les intérêts des auteurs et ceux des chercheurs.

Le système repose expressément sur la conclusion d'un accord – en cours de négociation – entre le ministère en charge de l'Enseignement supérieur et de la Recherche et les organismes de gestion collective des droits des artistes des arts visuels concernés⁶. Les auteurs pourront obtenir une rémunération en contrepartie de l'utilisation de leurs œuvres, mais ceux qui ne souhaitent pas profiter de ce système pourront exercer un droit de retrait (*opt-out*) et négocier sur une base individuelle l'utilisation de leurs œuvres.

Pour préparer la conclusion de cet accord, une étude commandée par le ministère en charge de la Recherche a permis d'estimer le nombre de reproductions d'œuvres des arts visuels dans les publications scientifiques qui entrent dans le champ de la mesure⁷. Le nombre de nouvelles images protégées publiées chaque année est estimé, sur la base du corpus des publications scientifiques parues en 2019, entre 55 840 et 66 500.

Face à l'ampleur des œuvres concernées, il y a tout lieu de penser que l'accord à venir, et qui sera étendu par arrêté de la ministre de la Culture, fournira une contribution sensible à la politique de la science ouverte. ■

2. Le rapport de Bruno Racine sur le statut des artistes-auteurs, *L'auteur et l'acte de création*, de janvier 2020, confirme ce diagnostic – 46 % des artistes plasticiens perçoivent moins de 5 000 euros par an – et la nécessité de conforter les dispositifs de rémunération des auteurs. <https://www.culture.gouv.fr/Espace-documentation/Rapports/L-auteur-et-l-acte-de-creation>

3. <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000042738027>

4. Directive (UE) 2019/790 du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique et modifiant les directives 96/9/CE et 2001/29/CE. <https://eur-lex.europa.eu/legal-content/FR/ALL/?uri=celex:32019L0790>

5. Ordonnance n° 2021-1518 du 24 novembre 2021 complétant la transposition de la directive 2019/790 du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique et modifiant les directives 96/9/CE et 2001/29/CE. <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000044362034>

6. Deux organismes de gestion collective ont, à ce jour, été agréés par le ministère de la Culture pour mener ces négociations : la Société des auteurs dans les arts graphiques et plastiques (ADAGP) et la Société des auteurs des arts visuels et de l'image fixe (SAIF).

7. Pierre-Carl Langlais et Antoine Blanchard, *Étude sur l'utilisation d'œuvres relevant des arts visuels dans les publications scientifiques*, rapport de recherche du ministère de l'Enseignement supérieur et de la Recherche, août 2021. <https://hal.science/hal-03682113>

La science ouverte et l'évaluation de la recherche

La science ouverte : une utopie en marche, un mouvement en profondeur. Réservés à quelques adeptes au début des années 2010, les mots « science ouverte » sont maintenant connus de tous dans la communauté scientifique française. La science ouverte transforme les pratiques scientifiques, avec toutefois des différences sensibles entre pays et entre communautés scientifiques.

Il s'agit d'une transformation en profondeur, dont les racines sont anciennes. Comme toutes les grandes transformations culturelles, elle donne lieu à plusieurs « grands récits ». Certains évoquent l'essor des publications scientifiques en accès libre et les premières plateformes d'archives ouvertes dans les années 1990. D'autres rappellent les succès des logiciels *Open Source* et racontent l'histoire de communautés d'utilisateurs-développeurs contribuant à la mise au point de logiciels dont l'ouverture était le gage essentiel de la qualité. D'autres soulignent que la science ouverte trouve ses origines au sein des communautés de la « grande physique ». Parce que l'exploration des « deux infinis » s'appuie sur des très grands instruments (télescopes et collisionneurs) dont la mise au point et le coût considérable nécessitent une coopération étroite au niveau mondial, les données issues des expériences ont été *nécessairement* partagées au sein des communautés scientifiques. Une révolution était en marche : le Cern, qui s'est attaché au début des années 1990 les services d'un informaticien visionnaire, Tim Berners-Lee, pour travailler sur les questions liées au partage des données et des documents scientifiques, a été le berceau du *World Wide Web*.

Toutes ces facettes – les publications en accès ouvert, les logiciels « libres », le partage des données – participent toutes d'une « utopie » nourrie par la conviction que les produits de la recherche financée sur crédits publics sont des « biens communs » qui ont vocation à être accessibles à tous. Le partage au sein des communautés scientifiques fait avancer la science ; il contribue à la qualité de la recherche (*via* par exemple la mise en qualité des données). Plus largement, le partage avec l'ensemble de la société est une condition pour que la science tienne toute sa promesse d'apport au progrès intellectuel, culturel, économique et social.

Bien sûr, comme pour toute « belle utopie », la confrontation avec la réalité soulève des difficultés.

Certains prônent une science « ouverte et gratuite », oubliant parfois qu'un bien, même public, a un coût et que la question de sa gratuité est une question de choix d'un modèle économique. Les débats entre promoteurs et détracteurs de la science ouverte vont bon train, et recèlent parfois des vraies oppositions, comme cela est normal lorsque des modèles économiques établis sont remis en cause.

Cette utopie est clairement en marche, et la France fait partie des pays où le dynamisme de la



THIERRY COULHON

Président du Haut Conseil d'évaluation de la recherche et de l'enseignement supérieur (Hcéres)

BERNARD LARROUTUROU

Directeur du Département d'évaluation des organismes de recherche, Hcéres



Perspective d'un matériel de base de données.

science ouverte est très visible. Les Plans nationaux pour la science ouverte (PNSO) portés depuis 2018 par le ministère de l'Enseignement supérieur et de la Recherche ont eu un impact indéniable. Des modes de publications nouveaux font l'objet d'initiatives variées, la « bibliodiversité » prend son essor, le CNRS vise à atteindre en 2023 l'objectif de 100 % des publications en accès ouvert. La mise en qualité des données et leur partage se heurtent à des contraintes plus dures : d'une part, l'objectif n'est pas de partager 100 % des données (penser à la protection des données personnelles, ou aux différents secrets protégés par la loi) ; d'autre part, la mise en qualité des données et la quête de reproductibilité de la recherche qui s'y attache posent des questions très difficiles dans certains domaines comme la biologie ou les sciences sociales, et touchent à des sujets essentiels, notamment à l'intégrité scientifique. Les pratiques évoluent en profondeur, les chercheurs sont formés à la science ouverte dès leur thèse de doctorat.

Évaluer la science ouverte

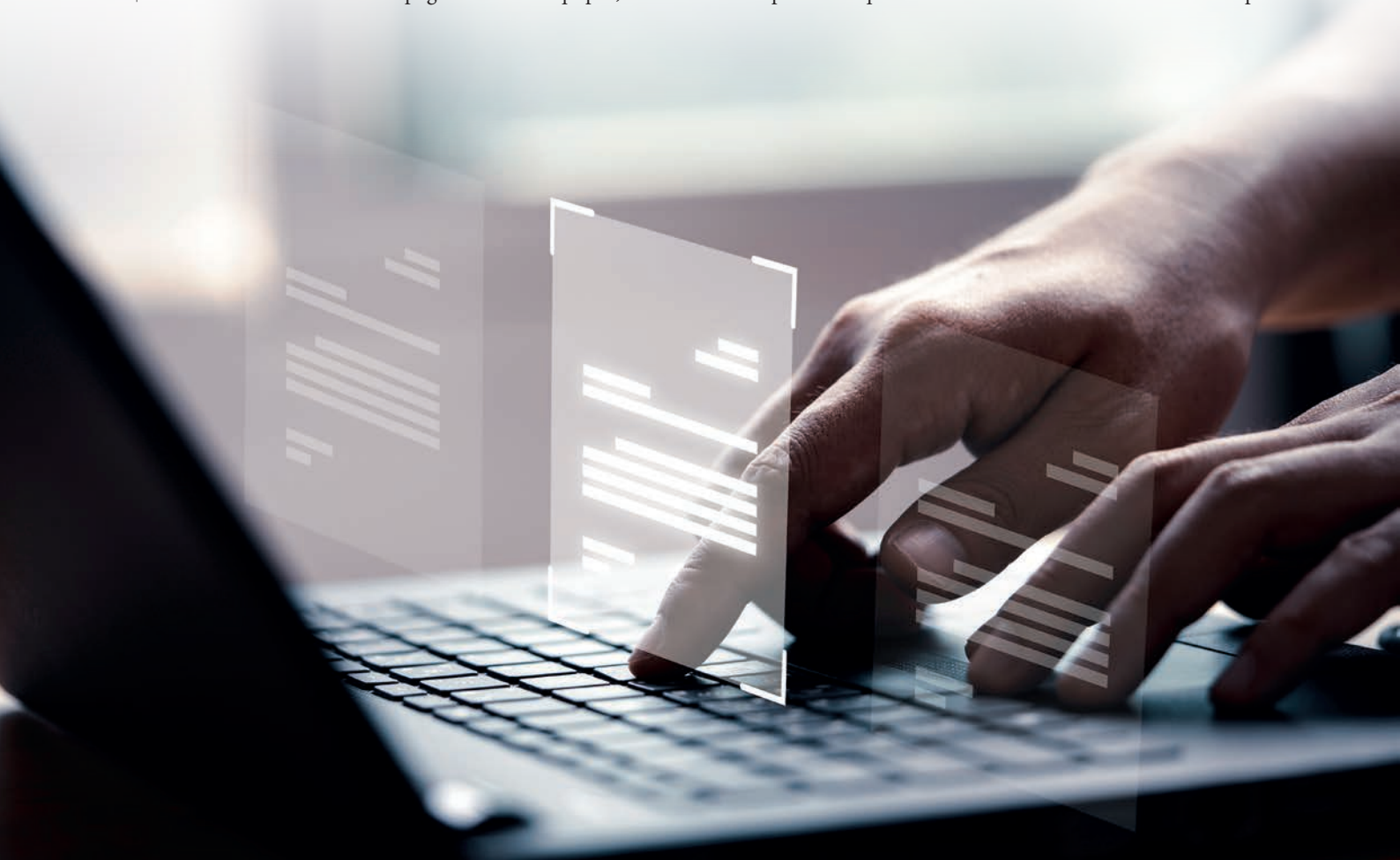
Le Hcéres, observateur de ces évolutions, adapte ses méthodes et ses critères pour les mesurer et les analyser dans ses évaluations. Les analyses bibliométriques incluent désormais des « indices d'accès ouvert ». Les évaluations des laboratoires de recherche et de formations doctorales examinent comment les différentes dimensions de la science ouverte sont prises en compte dans la vie des laboratoires de recherche et les formations doctorales. Les évaluations des universités et des organismes de recherche conduisent à examiner comment ces établissements mettent en place les dispositifs (outils, méthodes, formations, accompagnement des équipes, mais aussi entrepôts

de données, etc.) nécessaires pour progresser en matière de gestion, de mise en qualité et de partage des données.

Des évolutions plus profondes de l'évaluation de la recherche

Plus profondément, parce que la science ouverte porte une vision de la science comme bien commun à partager au sein de nos sociétés, elle ébranle le système de reconnaissance en vigueur dans le monde scientifique. Les promoteurs de la science ouverte critiquent le système commercial international des publications scientifiques, contestent le bien-fondé de certains de ses indicateurs de succès et appellent à prendre des distances par rapport à la « course à la quantité » dans laquelle s'inscrit trop souvent la compétition scientifique. Visant à privilégier le partage et la qualité des résultats de la recherche, ces réflexions appellent à changer les méthodes d'évaluation et soulignent la nécessité de prendre en compte l'ensemble des productions de la recherche, pas seulement les publications, et de chercher à évaluer leur impact.

On voit ainsi, dans plusieurs pays, se mettre en place des changements dans les méthodes d'évaluation de la recherche. Par exemple, le CNRS demande désormais qu'un chercheur évalué présente un petit nombre de ses productions en expliquant pourquoi il ou elle considère que ce sont les productions les plus pertinentes, et sans présenter sa liste de publications. Le Hcéres est partie prenante de ces évolutions, qui n'en sont qu'à leur début. Si elles ne portent pas sur la science ouverte en elle-même, elles participent, au fond, de la même vision que l'on s'est permis plus haut de qualifier comme une – belle et salutaire – utopie. ■



L'utilisation des technologies IIF dans les publications scientifiques à forts contenus visuels : les expérimentations de la pépinière de revues DeVisu, membre du réseau Repères

La volonté de valoriser les fonds conservés par les institutions patrimoniales est à l'origine de nombreux programmes de numérisation et de mise à disposition des collections publiques. Les chercheuses et chercheurs dont les travaux se basent sur l'étude de sources visuelles y trouvent ainsi des matériaux pour mener à bien leur recherche. Afin de faciliter le partage et la réutilisation de ces corpus d'images, des solutions techniques ont été développées au sein des institutions mais elles restent encore trop peu investies au regard des possibilités qu'elles offrent à l'édition scientifique. C'est pourquoi des expérimentations sont menées dans le cadre de la pépinière de revues DeVisu, avec pour objectif de mettre en œuvre la démarche et les outils nécessaires à l'exploitation des possibilités en matière d'édition scientifique.

InVisu et DeVisu : présentation

Le laboratoire InVisu¹, Unité mixte d'appui à la recherche (UAR) de l'Institut national d'histoire de l'art (INHA)² et du Centre national de la recherche scientifique (CNRS), accompagne les renouvellements méthodologiques en histoire de l'art, en s'appuyant sur l'expérimentation d'outils numériques en adéquation avec les bonnes pratiques de la science ouverte³.

Depuis sa création en 2008, InVisu propose des services d'accompagnement aux équipes de recherches afin qu'elles s'approprient les pratiques et outils numériques nécessaires à la valorisation et à la diffusion de leur production⁴ selon les principes dits FAIR⁵ : « facile à trouver, accessible, interopérable et réutilisable ».

C'est dans ce cadre qu'en 2020, le laboratoire a lancé DeVisu⁶, pépinière de revues à destination de la communauté des historiennes et historiens de l'art, en collaboration avec l'INHA. DeVisu est à la fois une plateforme de diffusion de revues en accès ouvert, en modèle diamant⁷, et un service d'aide à la montée en qualité éditoriale des revues. Nous travaillons pour cela avec les critères QUERO⁸, qui permettent de vérifier tous les points nécessaires au bon fonctionnement d'une revue : composition des comités, régularité des parutions, spécificités techniques.

La pépinière accompagne à ce jour trois revues : *Photographica*⁹, *Modes Pratiques*¹⁰ et *ABE Journal*¹¹ et s'inscrit dans l'offre de services proposée par



1. Page d'accueil d'InVisu (« L'information visuelle et textuelle en histoire de l'art : nouveaux terrains, corpus, outils »). <https://invisu.cnrs.fr>
2. Voir dans ce numéro l'article de Federico Nurra, p. 23.
3. Les principes de la science ouverte sont de mettre à disposition de façon pérenne, libre et gratuite, les données et résultats de la recherche afin qu'ils puissent être largement réutilisés. Consortium Couperin, *Guide des bonnes pratiques* : <https://scienceouverte.couperin.org/guide-des-bonnes-pratiques/> ; Christine Hadrossek, Joanna Janik, Maurice Libes, Violaine Louvet, Marie-Claude Quidoz, Alain Rivet et Geneviève Romier, *Guide de bonnes pratiques sur la gestion des données de la Recherche*, 2023. <https://hal.science/hal-03152732v2/document>

4. Parmi les services mis en place au sein du laboratoire InVisu, on peut notamment citer les résidences au sein desquelles des postdoctorantes et postdoctorants sont accueillis afin de les accompagner dans la publication d'un catalogue numérique rassemblant leur corpus de thèse. <https://www.inshs.cnrs.fr/fr/cnrsinfo/retours-dexperience-sur-les-residences-dinvisu>.

5. Acronyme pour Findable, Accessible, Interoperable et Reusable.
6. <https://devisu.inha.fr/>

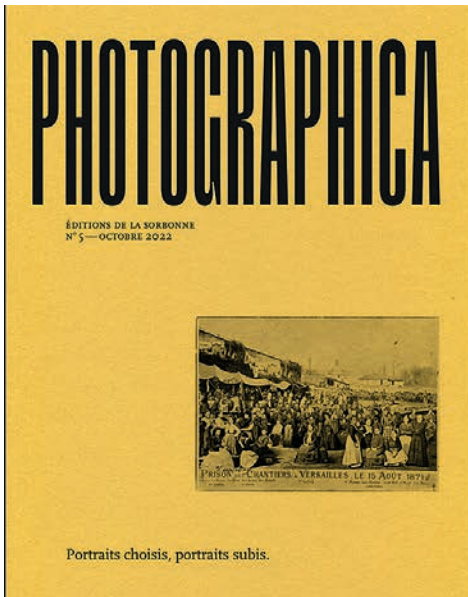
VÉRONIQUE COHONER

Ingénieure, Pôle numérique-soutien aux revues et publications en libre accès, UAR CNRS 3491 – Maison des sciences de l'homme Ange Guépin réseau Repères-Nantes Université

JULIETTE HUEBER

Responsable éditoriale et de l'ingénierie documentaire, Laboratoire InVisu, Unité d'appui et de recherche (UAR) CNRS-INHA 3103

7. Modèle « diamant » : modèle selon lequel la publication et la diffusion sont totalement gratuites pour le lecteur et pour l'auteur : <https://skosmos.lotterre.fr/TSO/fr/page/-HS79BV00-V> Et voir dans ce numéro l'article de François Gêze, p. 92.
8. Françoise Catherine Gouzi, Jean-Luc de Ochandiano et Armelle Thomas, *Critères QUERO – Préconisations pour la qualité des revues scientifiques : science ouverte, bonnes pratiques éditoriales et référencement*, 2020. <https://hal.science/hal-03060601/>
9. <https://devisu.inha.fr/photographica/>
10. <https://devisu.inha.fr/modespratiques/>
11. <https://journals.openedition.org/abe/>



Couvertures des revues *Photographica*, *Modes Pratiques* et *ABE Journal*.

12. Voir dans ce numéro l'article de David Pouchard, p. 97.

13. Acronyme pour *International Image Interoperability Framework*: <https://iif.io/>; voir : Sophie Bertrand, Gaël Chenard, Stéphane Pillorget, Carine Prunet et Régis Robineau, « IIF : découverte et interopérabilité sans frontières des images patrimoniales », *Culture et Recherche*, n° 143, 2022, p. 111-117.

14. On peut notamment citer pour la France les travaux de l'infrastructure numérique de recherche et de service *Biblissima+* consacrée à l'histoire de la transmission des textes anciens, de l'Antiquité à la Renaissance en Orient comme en Occident. <https://biblissima.fr/>

15. <https://jhna.org/issues/>

l'ensemble des pépinières de revues membres du réseau Repères.

Mais, s'agissant d'une pépinière thématique destinée aux revues à forts contenus visuels, son objectif est également de mettre en œuvre un ensemble de services basés sur la mise à disposition de fonds iconographiques libres et en très bonne définition, par les institutions patrimoniales. Ceci dans le but de proposer à la communauté des historiennes et historiens de l'art les possibilités offertes par des technologies innovantes en matière de présentation, de navigation et d'annotation des images¹². Il s'agit donc de (re)mettre l'image au cœur des publications scientifiques tout en maintenant son inscription dans les chaînes éditoriales existantes et efficaces.

Des expérimentations pour l'utilisation des technologies IIF¹³ en contexte éditorial

Cette spécialisation de la pépinière DeVisu a conduit ses membres à s'intéresser aux technologies IIF, ensemble de spécifications dont l'objectif est de doter les bibliothèques numériques d'un cadre technique commun pour la diffusion et la manipulation des images. Ces technologies sont largement utilisées par les institutions culturelles et dans le cadre de projets scientifiques¹⁴ mais ne sont que très rarement exploitées par les revues qui continuent, à quelques exceptions¹⁵ près, d'insérer dans les articles des images statiques sans annotation ni capacité de se focaliser sur des détails.

© Jean-Christophe Carius, Service numérique de la recherche du Département des études et de la recherche de l'INHA

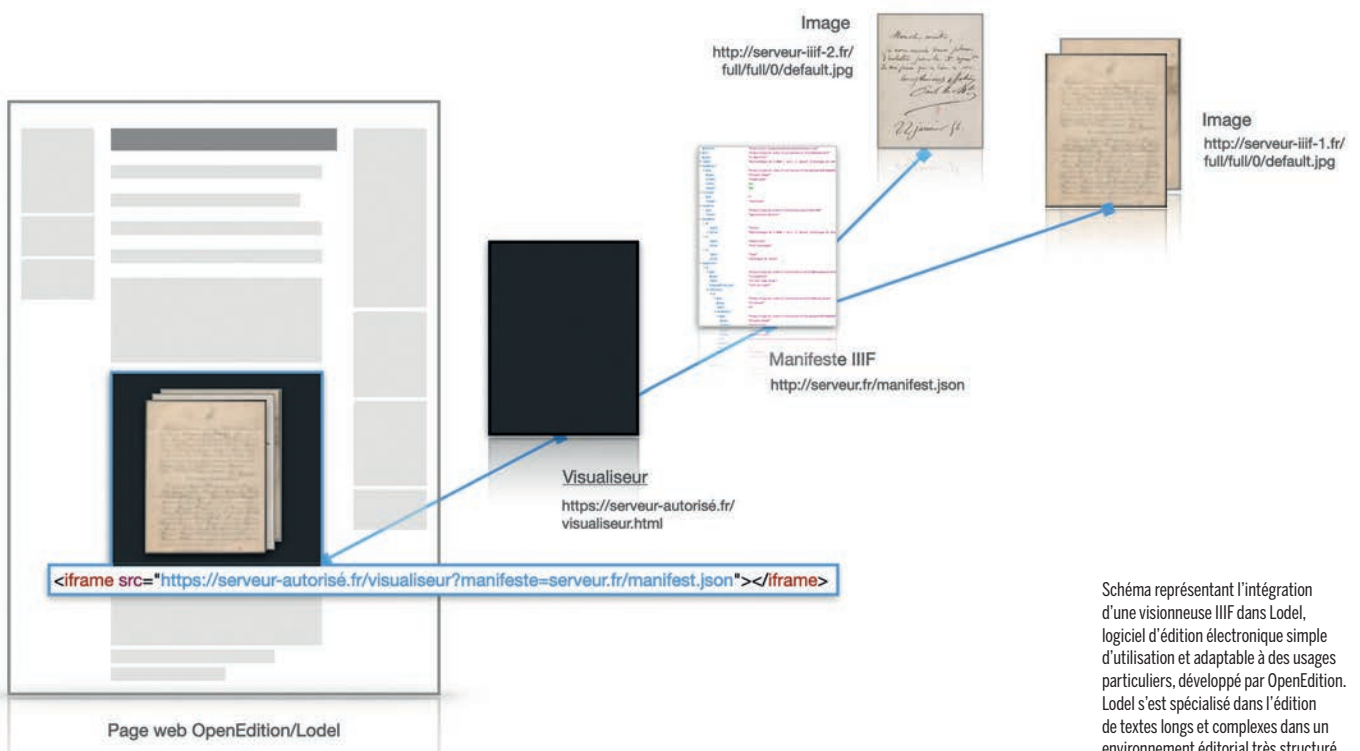
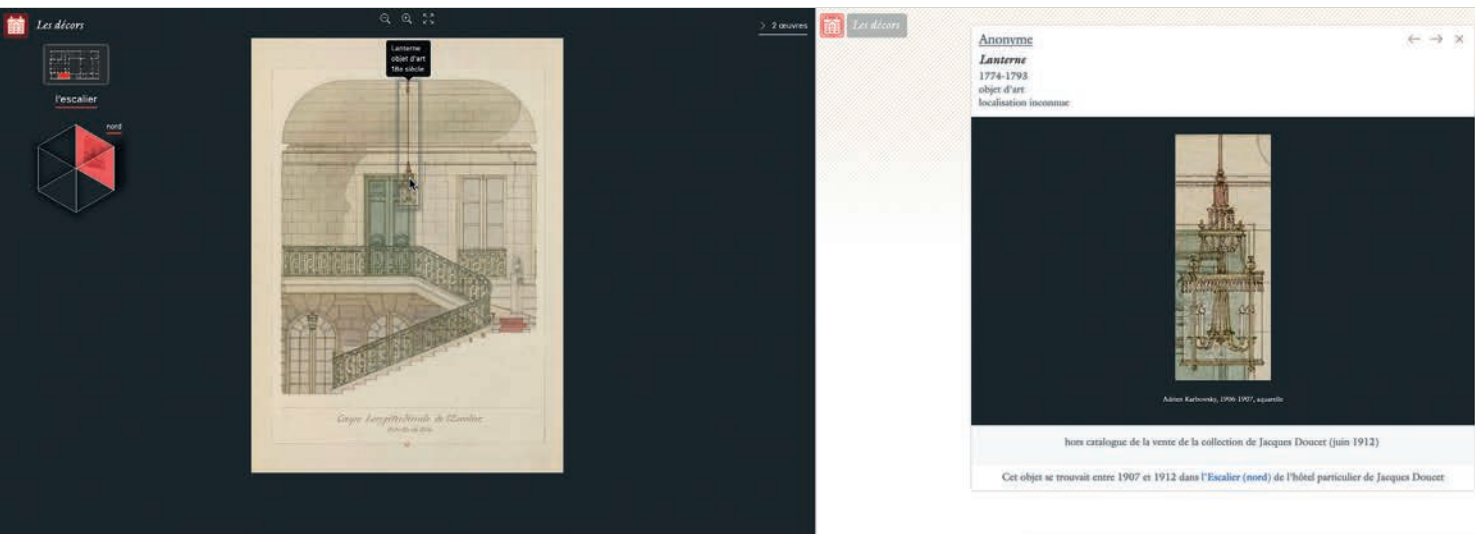


Schéma représentant l'intégration d'une visionneuse IIF dans Lodel, logiciel d'édition électronique simple d'utilisation et adaptable à des usages particuliers, développé par OpenEdition. Lodel s'est spécialisé dans l'édition de textes longs et complexes dans un environnement éditorial très structuré.



Des expérimentations ont été effectuées en collaboration avec le service numérique de la recherche de l'INHA qui travaille sur l'utilisation des technologies IIF pour la publication de sources enrichies. Ce service a ainsi développé la plateforme PENSE¹⁶ qui vise à mettre en ligne des sources enrichies en histoire de l'art. Ces expérimentations nous ont ainsi amenés à valider l'intégration de visionneuses IIF sur la plateforme OpenEdition Journals et à enclencher un travail avec les acteurs de la chaîne Métopes¹⁷ autour de l'intégration du traitement IIF dans leur modèle éditorial. Ces expérimentations ont été menées en priorité, OpenEdition Journals et Métopes étant des outils largement utilisés en sciences humaines et sociales.

La prochaine étape sera de tester l'apport scientifique de l'usage de l'IIF en édition avec un panel de revues en histoire de l'art et en sciences humaines

Cette façon de réutiliser et valoriser les fonds conservés démontre que leur mise à disposition en très haute définition-résolution, sous des licences ouvertes permettant leur libre réutilisation, et avec des technologies de type IIF n'est pas du tout anecdotique et accessoire.

et, au-delà et parallèlement, de développer des outils permettant une mise en œuvre effective.

Outre l'apport scientifique, ces travaux visent à la réutilisation du patrimoine mis à disposition par les institutions culturelles et patrimoniales. Cette façon de réutiliser et de valoriser les fonds conservés démontre que leur mise à disposition en très haute définition-résolution, sous des licences ouvertes permettant leur libre réutilisation, et avec des technologies de type IIF n'est pas du tout anecdotique et accessoire. Au contraire, elle offre aux chercheuses et chercheurs de nouvelles possibilités tant pour leurs travaux de recherche que pour leur mise à disposition et leur diffusion.

Participation à l'offre de service commune de Repères

Au terme de cette expérimentation, et afin d'en permettre une large appropriation, il s'agira également pour DeVisu de partager son travail au sein des membres du réseau Repères et plus largement au sein de la communauté de l'édition scientifique publique.

Le réseau Repères¹⁸ fédère seize plateformes de revues scientifiques en accès ouvert; il organise le partage de compétences complémentaires entre métiers, et la production de ressources aux niveaux français et belge. Ainsi, les travaux effectués au sein de chaque plateforme ont vocation à être partagés et diffusés, par exemple: évolutions techniques des logiciels d'édition électronique, adaptation des modèles éditoriaux, webinaires et formations. Chaque plateforme, pépinière ou pôle éditorial du réseau, peut ainsi travailler pour elle-même (du fait de ses spécificités propres), avec et pour les autres. L'effet de mutualisation du réseau joue à plein, au bénéfice de l'édition scientifique en accès ouvert, et en lien bien sûr avec ses partenaires.

Ainsi, les actions menées au sein de DeVisu ne se limitent pas à sa propre communauté, mais permettent de faire profiter la communauté scientifique de son expertise sur la question du traitement des images pour la recherche. ■

Un exemple d'utilisation des technologies IIF au sein de l'INHA. Dans la plateforme d'exposition numérique de sources enrichies (PENSE), rubrique « Les décors Karbowsky » (<https://karbowsky.inha.fr>), puis « Les décors d'Adrien Karbowsky pour Jacques Doucet », lorsque l'utilisateur clique dans l'image de l'escalier sur la zone active (« lanterne objet d'art XVIII^e siècle »), cela donne accès à une visualisation du détail.

16. <https://pense.inha.fr/>

17. Pour « Méthodes et outils pour l'édition structurée », <http://www.metopes.fr/>

18. Basile Bayoux, Véronique Cohoner et Armelle Thomas, « Repères, un réseau pour les plateformes de revues », *Ar(ables)ques*, n° 108, mars 2023, p. 8-9, DOI : 10.35562/arabesques.3135, <https://publications-prairial.fr/arabesques/index.php?id=3135>; Armelle Thomas, « Le réseau de pépinières de revues scientifiques Repères. Un révélateur de la complémentarité des métiers de l'infodoc et de l'édition », *i2D - Information, données & documents*, vol. 2, n° 2, 2020, p. 144-148, DOI : 10.3917/i2d.202.0144, <https://www.cairn.info/revue-i2d-information-donnees-et-documents-2020-2-page-144.htm>

Les enjeux du libre accès pour les pratiques et expertises de la documentation : quels impacts de la science ouverte sur les activités documentaires en administration centrale ?

Le libre accès¹ aux données de la recherche est un des volets de la science ouverte entendue comme « la diffusion sans entrave des publications et des données de la recherche [qui] induit une démocratisation de l'accès aux savoirs, utile à la recherche, à la formation, à l'économie, à la société » (premier Plan national pour la science ouverte, 2018). Dans ce contexte, le libre accès est considéré comme « la mise à disposition gratuite sur l'Internet public des productions de la recherche avec pour seule limite de garantir aux auteurs un contrôle sur l'intégrité de leurs travaux et le droit à être correctement reconnus et cités » (Initiative de Budapest en 2002, complétée par les Déclarations de Bethesda en juin 2003 et de Berlin en octobre 2003)².

STÉPHANE MARGUERIN

Adjoint au chef de la Mission de la politique documentaire, ministère de la Culture

MIGUEL SAYOUS

Chargé de mission, Secrétariat Général, Département des études, de la prospective, des statistiques et de la documentation, Mission de la politique documentaire, ministère de la Culture

Une constellation d'initiatives privées et publiques a ainsi vu le jour et accélère ce mouvement visant à rendre accessibles les travaux de recherche, en utilisant des outils variés : archives ouvertes et plateformes de diffusion, notamment.

Enjeux du libre accès pour la politique documentaire : une continuité des missions, des pratiques plus sophistiquées

Les Plans nationaux pour la science ouverte (PNSO 2018 et 2021)³ ont permis aux documentalistes de se saisir de ce sujet, de développer des partenariats avec des chercheurs et de prendre des initiatives nouvelles. Le libre accès des données et des contenus de la recherche questionne une compétence traditionnelle des professionnels de la documentation et induit une nouvelle façon de diffuser les savoirs. Par leur mission « régaliennne » d'acquisition, de référencement, catalogage et indexation des documents, mais aussi de valorisation et de diffusion des connaissances, les

centres de documentation sont nécessairement au cœur de cette évolution.

La gestion du libre accès implique un enrichissement plus sélectif des fonds documentaires. En effet, la production scientifique s'accélérate et s'intensifie au fil des années, il devient délicat de suivre ce flux de publications en constante évolution : 3 millions d'articles publiés dans 33 100 journaux d'éditeurs scientifiques, selon le rapport STM (2018)⁴. Aujourd'hui, les deux tiers des publications scientifiques françaises sont en accès ouvert, selon les données du Baromètre français de la science ouverte (décembre 2022)⁵. Face à ce foisonnement, il ne s'agit pas de marginaliser les documents imprimés mais d'organiser la complémentarité des ressources, en veillant à préserver la cohérence des fonds documentaires.

L'apport d'une expertise en matière de visibilité des ressources et de médiation prend une importance accrue lorsqu'il s'agit de valoriser les données ouvertes de la recherche.



Jean-Honoré Fragonard, *Le philosophe (Saint-Jérôme)*, 1764, Hamburger Kunsthalle : <https://online-sammlung.hamburger-kunsthalle.de/en/objekt/HK-777>

Une organisation des ressources documentaires intégrant des flux immatériels complète une gestion en termes de stocks (collections papiers) et nécessite une véritable synergie entre les chercheurs et le responsable de la politique documentaire dans la sélection des corpus pertinents.

Au ministère de la Culture, la prise en compte du libre accès dans la politique documentaire s'organise selon trois volets :

- collecter et valoriser la production scientifique du ministère pour la rendre accessible aux communautés de chercheurs, mais également au grand public. C'est tout particulièrement le cas pour les études et statistiques ;
- identifier et référencer la production administrative interne (« littérature grise ») en favorisant sa réutilisation et la libre diffusion des contenus existants, dès lors qu'ils sont communicables ;

L'apport d'une expertise en matière de visibilité des ressources et de médiation prend une importance accrue lorsqu'il s'agit de valoriser les données ouvertes de la recherche.

- identifier et valoriser, notamment auprès des chercheurs, les publications en libre accès relevant du champ d'études du ministère de la Culture.

Deux leviers de valorisation des documents, données ouvertes incluses, l'un physique (le futur centre de ressources, non seulement lieu d'échanges entre chercheurs et acteurs de la Culture, mais aussi d'accès à des ressources libres sélectionnées par les documentalistes, ouvrira en 2024), l'autre numérique (un portail documentaire) soutiennent cette ambition.

Si l'impact du libre accès n'a que des effets positifs sur l'enrichissement des collections, il rend nécessaire une montée en compétences des professionnels et des modifications éventuelles dans l'organisation du travail.

L'incidence du libre accès sur l'expertise et les savoir-faire : du traitement documentaire au management de la connaissance

Les ressources en libre accès viennent modifier en profondeur l'expertise des documentalistes sur deux aspects : la constitution d'un corpus de référence pour les chercheurs ; un accompagnement à la publication de leurs travaux.

Organiser et classer les documents en libre accès : constituer des corpus de référence pour les chercheurs

Les ressources ouvertes sélectionnées par les documentalistes ne passent pas systématiquement par le circuit d'acquisition traditionnel et ne sont pas nécessairement identifiées dans le catalogue de la bibliothèque au même titre que les documents imprimés. Dans cette hypothèse, il appartient aux

1. Le libre accès (*Open Access*) est une combinaison de l'accès ouvert (les textes sont accessibles sans identification et sans paiement) et d'une licence libre, permettant la réutilisation dans certaines conditions.

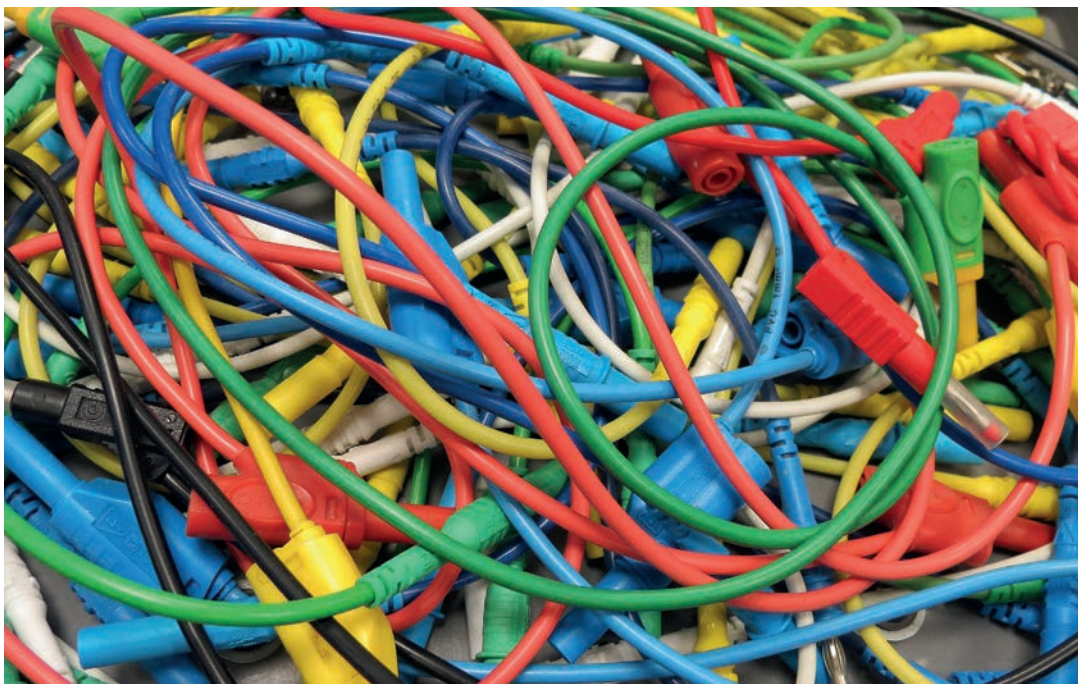
2. Initiative de Budapest pour l'accès ouvert (2002) : <https://www.budapestopenaccessinitiative.org/read/french-translation/>
Déclaration de Bethesda (2003) : https://www.ouvrirelscience.fr/wp-content/uploads/2018/11/D%C3%A9claration-Bethesda_fr.pdf
Déclaration de Berlin (2003) : https://www.ouvrirelscience.fr/wp-content/uploads/2018/11/D%C3%A9claration-Berlin_fr.pdf

3. Plan national pour la science ouverte (PNSO 2018) : https://www.enseignementsup-recherche.gouv.fr/sites/default/files/content_migration/document/PLAN_NATIONAL_SCIENCE_OUVERTE_978672.pdf ;

Deuxième Plan national pour la science ouverte (PNSO 2021-2024) : https://www.enseignementsup-recherche.gouv.fr/sites/default/files/content_migration/document/2e-plan-national-pour-la-science-ouverte-2021-2024-7794.pdf

4. <https://www.issn.org/fr/le-rapport-stm-2018-annonce-le-developpement-des-services-autour-de-l-issn/>

5. Baromètre français de la science ouverte : <https://barometredelascienceouverte.esr.gouv.fr/> ; <https://www.enseignementsup-recherche.gouv.fr/fr/barometre-francais-science-ouverte-2022-89981> ; voir dans ce numéro l'article de Laetitia Bracco, p. 26.



European Science Photo Competition,
Electric cables Tangled cables, 2015
(Licence CC-BY-SA-4).

6. Voir dans ce numéro les articles de Frédéric Gendre, Christophe Hugot, Joachim Schöpfel et Agnès Souchon, p. 49, et le focus de Marjorie Borios, p. 52.

7. <https://creativecommons.org/>

8. Le grand principe du Plan S : les publications financées par des fonds publics doivent être librement et immédiatement accessibles dans des revues ou des plateformes en accès ouvert ou dans des archives ouvertes à compter de janvier 2021 :

<https://www.ouvrirlascience.fr/principes-et-recommandations-pour-la-mise-en-oeuvre-du-plan-s/>

documentalistes de proposer des dispositifs d'accès à ces ressources tels les outils de découverte pour faciliter leur repérage par les lecteurs. À défaut, ils peuvent également signaler ces documents dans le catalogue de la structure documentaire par une notice enrichie de métadonnées pour rendre les documents plus visibles, dans le respect des règles de réutilisation spécifiées dans la licence d'utilisation et en cohérence avec les fonds documentaires.

Garantir un accès pérenne aux résultats de la recherche et accroître leur visibilité : un enjeu de publication

Pour les documentalistes, les principaux axes de développement consisteraient à :

- garantir l'existence d'identifiants pérennes ou de liens statiques pour la publication sur un site dédié ;
- accompagner le chercheur dans son parcours de publication et inciter au dépôt de sa production sur une archive ouverte, telle que HAL⁶ ;

- réaliser des actions de médiation valorisant les ressources numériques et les documents en accès libre, et plus particulièrement les publications du ministère ;
- acquérir des compétences juridiques pour accompagner les chercheurs dans la gestion de leurs droits numériques. Des revues en libre accès proposent des articles sous une licence de diffusion, souvent de type Creative Commons (CC)⁷. Le ministère de la Culture s'inscrit dans cette pratique. Les études du Département des études, de la prospective, des statistiques et de la documentation sont publiées sous une licence régissant les conditions de réutilisation, de distribution et de partage.

Depuis 2021, le Plan S⁸ de la cOAlition S (consortium de bailleurs) préconise la licence Creative Commons (CC-BY), l'auteur conservant ses droits (BY, ou « attribution ») : toute exploitation de l'œuvre est permise pourvu qu'elle et ses auteurs soient correctement cités).

Les professionnels de la documentation peuvent, au prix d'une redéfinition de leurs métiers, s'emparer de toutes les opportunités de la science ouverte et jouer un rôle de premier plan dans la diffusion de la connaissance scientifique comme « experts de la donnée » (*Data Librarian*), à la croisée de la documentation, de l'informatique et de l'accompagnement à la valorisation de la recherche. ■



Concept d'hébergement de site Web avec circuits :
https://fr.freepik.com/photos-gratuite/concept-hebergement-site-web-circuits_26412535.htm#query=open%20data&position=29&from_view=search&track=ais

La base EROS du C2RMF : mise en perspective de cette nouvelle ressource publique avec les programmes de recherche ESPADON, E-RHIS et IIIF

Dès sa genèse, la base de données EROS¹ du C2RMF² a été conçue selon les standards du W3C³ et des normes préconisées de formats et de structuration de données pour garantir la présence dans l'Internet de données liées. Son architecture permet donc intrinsèquement la mise en œuvre d'une version publique d'EROS.

Une histoire d'ouverture

Élaborée en 1999 à partir de logiciels libres (*Open Source*), la base de données EROS a suivi l'évolution du Web en intégrant de nouvelles fonctionnalités, comme notamment l'implantation d'un nouveau moteur de

recherche permettant de faire des requêtes selon plusieurs modes de recherches avancées, ou le développement d'un visualiseur permettant d'afficher et de superposer des images multispectrales pour examiner leur différence par variation d'opacité. Avec la crise de

LUC BOUILLER

Chef de la mission Nouvelles technologies de l'information, Centre de recherche et de restauration des musées de France (C2RMF)



1. *European Research Open System.*
2. Centre de recherche et de restauration des musées de France.
3. Le *World Wide Web Consortium*, abrégé par le sigle W3C, est un organisme de standardisation à but non lucratif, fondé en octobre 1994 et chargé de promouvoir la compatibilité des technologies du *World Wide Web* telles que HTML, par exemple. Fonctionnant comme un consortium international, il comptait 452 membres au 1^{er} octobre 2021.

Bronzino, *La déploración sur le corps du Christ*, 1543-1545, Besançon, musée des Beaux-Arts et d'Archéologie.

La photographie en infrarouge en fausses couleurs est un procédé de traitement numérique de l'image, dont le rendu coloré est décalé dans le spectre. Cette manipulation permet de repérer des différences d'homogénéité dans certaines plages colorées, de repérer les zones de réserve en liaison avec le dessin préparatoire, ou d'identifier certains pigments grâce à la couleur qu'ils prennent par cette manipulation.

4. *International Image Interoperability Framework* est une communauté internationale qui a mis au point un protocole d'échange pour rendre les documents numériques interoperables, ainsi qu'un ensemble de standards à la fois client et serveur pour le développement de plusieurs API (interface de programmation d'application) : <https://iif.io/> ; voir : Sophie Bertrand, Gaël Chenard, Stéphane Pillorget, Carine Prunet et Régis Robineau, « IIF : découverte et interoperabilité sans frontières des images patrimoniales », *Culture et Recherche*, n° 143, 2022, p. 111-117 ; et dans ce numéro l'article de Véronique Cohoner, Juliette Hueber, p. 101.

5. Quelques potentiels risques juridiques sont à l'étude, principalement au regard de l'exploitation commerciale des créations intellectuelles des agents dans le cadre de missions de service et de la cession des droits d'auteur pour les chercheurs-enseignants et les restaurateurs extérieurs au Centre (voir dans ce numéro l'article de David Pouchard, p. 97).

6. Il s'agit de plusieurs téraoctets de données et d'imageries.

7. Sophie Bertrand, Gaël Chenard, Stéphane Pillorget, Carine Prunet et Régis Robineau, « IIF : découverte et interoperabilité sans frontières des images patrimoniales », *Culture et Recherche*, n° 143, 2022, p. 111-117.

Covid-19 et la généralisation du télétravail, des moyens spécifiques ont été mis en œuvre afin de maintenir un accès à distance ainsi que les performances et la fiabilité de l'infrastructure informatique de la base de données. Et paradoxalement, cette situation inédite nous a poussés à accélérer notre planification des tâches à accomplir pour l'ouverture d'EROS, et plus spécifiquement la virtualisation de certaines parties de notre infrastructure.

Cette technique de virtualisation nous a permis de créer des processus et des services qui ne sont pas intrinsèquement liés à du matériel informatique et qui peuvent être déployés, dupliqués ou migrés à la demande, de manière flexible sur un ou plusieurs serveurs. Ce travail de virtualisation a facilité l'introduction de nouvelles fonctionnalités, de services et plus particulièrement des améliorations concernant le visualiseur IIF⁴ qui seront intégrés dans la version publique.

Ce projet de mise à disposition publique d'une grande partie des connaissances produites relatives à l'étude et à la restauration des œuvres d'art se fera de manière progressive⁵. Le projet n'a pas vocation, pour le moment, à produire un contenu éditorial supplémentaire, car les ressources sont suffisamment riches et nombreuses⁶ pour envisager des collaborations et des initiatives innovantes avec d'autres institutions publiques, des musées, des écoles, des universités ou des industries créatives françaises, et même étrangères puisqu'EROS est nativement multilingue.

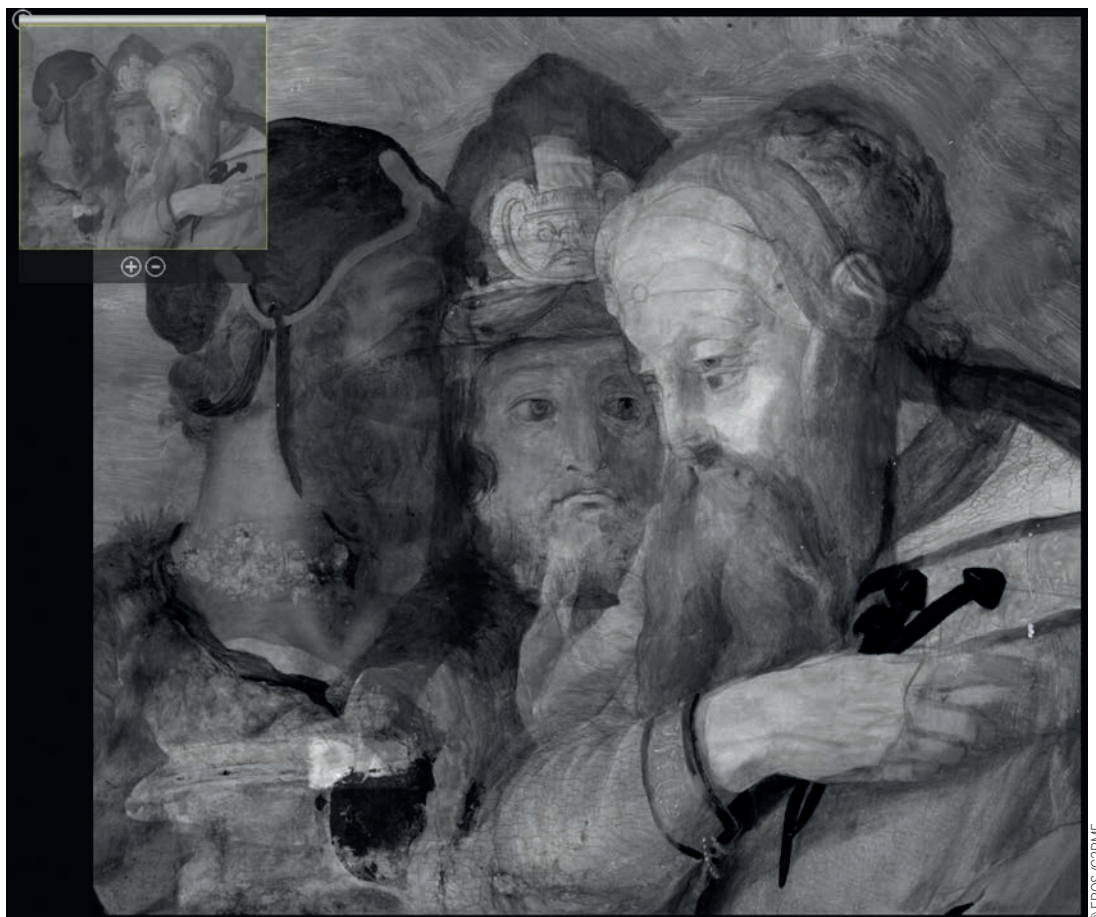
En effet, les glossaires d'indexation des rapports d'étude et de restauration des œuvres existent en allemand, anglais, catalan, chinois, danois, espagnol, français, italien, japonais, portugais et russe. Les termes permettent de sélectionner des rapports par leur contenu ; ils ont également été définis dans toutes ces langues afin d'éviter toute ambiguïté d'interprétation et leurs définitions sont également consultables en ligne.

Les interfaces d'interrogation et de saisie ont également été traduites du français vers les autres langues. À terme, cette version publique pourrait être un point d'ancrage pour le développement de missions d'accompagnement et de médiation avant, pendant et après une visite dans un musée. Elle pourrait également permettre aux mobinautes des niveaux ou des parcours différents de leurs visites selon leurs intérêts, leur disponibilité, ou encore faciliter leur compréhension de la recherche scientifique en conservation-restauration.

En effet, le visualiseur IIF développé pour EROS est le garant de la traçabilité comme de l'origine des images ; il offre des solutions de partage et d'interopérabilité, d'annotation des images stockées sur des serveurs distants, et empêche la duplication des fichiers sources des images pour des exploitations externes. Cette mise à disposition de contenu scientifique ouvre le champ des possibles à de multiples réutilisations⁷.


Plus fondamentalement, l'ouverture d'EROS s'inscrit dans l'essor de la science ouverte et c'est dans la communauté des sciences du patrimoine qu'il y a

La technique de la réflectographie infrarouge exploite le fait que la plupart des pigments utilisés en peinture deviennent plus ou moins transparents lorsqu'ils sont examinés dans une partie du spectre électromagnétique normalement invisible à l'œil nu, situé entre 780 et 1800 nanomètres. Ce sont principalement les pigments noirs, à base de carbone, qui restent visibles sous ce rayonnement, permettant de mettre au jour l'exécution d'un dessin préparatoire.




Déposition de Croix

artiste: **BRONZINO** Agnolo (1503-1572)
 lieu de conservation: France, Besançon, musée des beaux-arts et d'archéologie
 institution propriétaire: France, Besançon, musée des beaux-arts et d'archéologie
 numéro d'inventaire: D 799_1.29
 numéro de référence C2RMF: F10121
 collection patrimoniale: peinture
 école: Florence
 date de l'oeuvre: 1543 - 1545
 typologie: retable
 matériaux: bois
 techniques: peinture à l'huile
 dimensions: 2680 x 1730 mm
 signature, date, inscription, cachet: signature
 localisation de l'inscription: en bas à droite
 transcription de l'inscription: OPERA DEL **BRONZINO** FIORENTINO
 mode d'acquisition: saisie révolutionnaire
 informations diverses: retable exécuté pour l'oratoire d'Éléonore de Tolède au Palazzo Vecchio à Florence
 sources documentaires: Le musée des beaux-arts et d'archéologie de Besançon, Paris, musée et monuments de France-Fondation Paribas-RMN, 1994, p.30-31
 dossier de restauration n°: P7753
 secteur: R-art pictural / CR-peinture
 numéro de régie: MPR 1990
 domaine: peinture
 saisie de la notice: [Christine Desoprez](#)
 date de saisie de la notice: 09/03/2009



92 image(s) 12 document(s)

EROS :: Lightbox - Mozilla Firefox
 www.c2rmf.cnrs.fr/raais/oeuvres/F10121/images?page=8



EROS :: Mozilla Firefox
 www.c2rmf.cnrs.fr/raais/oeuvres/F10121/docs

document 1:
 numéro du document: **7426**
 type de document: rapport de laboratoire
 date: 15/09/2005
 auteur: Laval Eric
 problématique, résumé: identification des repeints anciens, approfondissement de la technique originale
 lieu d'archivage: C2RMF, Paris, Carrousel
 institution propriétaire du document: C2RMF (Centre de Recherche et de Restauration des Musées de France)
 numéro de référence C2RMF: E10121

document 2:
 numéro du document: **5186a**
 type de document: rapport de laboratoire
 date: 06/02/2005

© EROS/C2RMF

Exemple d'une fiche détaillée.

un intérêt croissant pour une convergence de travaux menés depuis plusieurs années. Des travaux qui partagent une même philosophie: la mutualisation de la connaissance produite par diverses institutions de la recherche scientifique à l'aide des technologies du Web sémantique et des données liées.

Pour la recherche scientifique en conservation-restauration du patrimoine en Europe

Mais pour que cette mutualisation de la connaissance puisse être tangible, il faut d'abord s'accorder sur un ensemble de classifications méthodiques, de nomenclatures, qui intègrent des vocabulaires contrôlés, des thésaurus et des ontologies spécifiques. Rappelons que les enjeux sont ceux définis par les principes du FAIR qui attestent que les données sont « cherchables » (faciles à trouver et accessibles) et modifiables (interopérables et réutilisables). Ces travaux ont d'ores et déjà été menés en partie dans le cadre des projets PARCOURS et SoCoRe!, et ils alimentent aujourd'hui les travaux d'identification de briques logicielles dans le projet d'EquipEx+ESPADON⁸.

Ces aspects conceptuels et techniques sont également réintroduits aujourd'hui dans le cadre de nouvelles réflexions sur les principes d'alignement des ontologies et du développement des thésaurus dont l'objectif est de créer un graphe de connaissances dans le domaine de la conservation-restauration du patrimoine culturel. Un graphe de connaissances est une représentation de l'ensemble des connaissances relatives à un domaine précis structurées par une

ontologie (modèle de données), des thésaurus (vocabulaires contrôlés) et des ressources (auteurs, chercheurs) et constitue un élément clé pour effectuer des recherches performantes avec des applications de nouvelles générations couplées à des réseaux de neurones convolutifs⁹.

Les données sont aujourd'hui bien évidemment au cœur des préoccupations des établissements de recherche et chaque établissement met en place des équipes dédiées à la transformation numérique de leurs activités à haut niveau afin d'être capable, face aux enjeux d'avenir, de répondre efficacement et avec la plus grande visibilité sur leurs expertises et de participer à des projets nationaux ou européens tels que ESPADON ou E-RHIS¹⁰.

EROS est une base de données conséquente de la recherche scientifique en conservation-restauration du patrimoine en Europe, mais son contenu est également au croisement de plusieurs domaines scientifiques. Ainsi, sa mise à disposition se révèle d'une grande importance pour ces deux futures plateformes qui promeuvent l'émergence de nouvelles méthodes de travail collaboratives, d'interdisciplinarité et de sciences participatives. Dès lors, nous constatons que l'ouverture des données engendre, dans une dimension prospective qui s'enrichit alors, à partir de l'identification formalisée des besoins actuels insatisfaits et des besoins émergents, celui d'une nécessité d'outils et de compétences, pour l'aide à l'analyse et aux traitements de données brutes dans une logique proche du *Big Data* et des impératifs que posent les technologies aujourd'hui présentes intégrant l'intelligence artificielle. ■

8. <http://socore.etic-lab.fr/>; voir dans ce numéro l'article de Vincent Detalle, Romain Thomas, p. 68.
9. Un réseau de neurones convolutif permet de détecter la présence de formes en appliquant un calcul local, appelé convolution, à un groupe de pixels. Ce calcul de convolution s'applique à différentes échelles d'une image et permet d'identifier progressivement le contenu d'une image par association et recoupement. Les réseaux convolutifs ont beaucoup d'applications en vision par ordinateur, pour la reconnaissance d'images et de vidéos, mais aussi en reconnaissance sonore, dans les systèmes de recommandation et dans le traitement de la langue naturelle.
10. E-RHIS (*European Research Infrastructure for Heritage Science*) est une infrastructure de recherche dédiée à l'étude des matériaux du patrimoine culturel et naturel.

« Lettres en Lumières » : l'intelligence artificielle appliquée aux décrets des États de Bourgogne

Les Archives de France ont entrepris il y a 25 ans une numérisation massive du patrimoine écrit. Ce qui se joue aujourd'hui est l'océrisation de ces images numériques, c'est-à-dire leur transcription automatique en textes, leur accessibilité dans leur intégralité, et leur « traduction » dans le français d'aujourd'hui. Plus encore, l'indexation de ces textes est un enjeu pour la recherche en science ouverte¹. Plusieurs millions de textes du XVIII^e siècle sont conservés en France. À ce titre, les registres des États de Bourgogne (1636-1787) sont signifiants (pour l'écriture) et porteurs de signifié (richesse des sujets abordés).

EDOUARD BOUYÉ

Directeur des Archives départementales
de la Côte-d'Or

1. Voir dans ce numéro l'article
de Christopher Kermorvant,
Lionel Keszenbaum, Manonmani Restif,
p. 35.

La lumière du paradis

Les archives ont leur enfer. Mais il est bien différent de celui des bibliothèques. Il s'agit des documents anciens, accumulés parfois depuis des décennies ou des siècles, sans jamais avoir été pourvus d'inventaire. C'est le dernier cercle de l'enfer. Faisons le chemin de Dante vers le purgatoire, et considérons ces fonds dotés d'un inventaire-sommaire conçu et imprimé au XIX^e siècle, que l'archiviste du XXI^e siècle rétro-convertit et met en ligne.

Pour la série C des Archives départementales de la Côte-d'Or, cette entrée en purgatoire s'est produite en 2015. Elle s'est accompagnée, dans les services départementaux et municipaux d'archives, de la numérisation et de la mise en ligne des archives décrites. Les archives sont trouvables, consultables et lisibles de loin : en vingt ans, le progrès est considérable.

Il reste cependant une marche à gravir pour arriver au paradis : c'est d'offrir aux chercheurs du texte lisible par machine, c'est de transformer les archives en données, c'est de mettre en lumière les lettres conservées dans les archives.

Mettre en lumière les lettres du siècle des Lumières

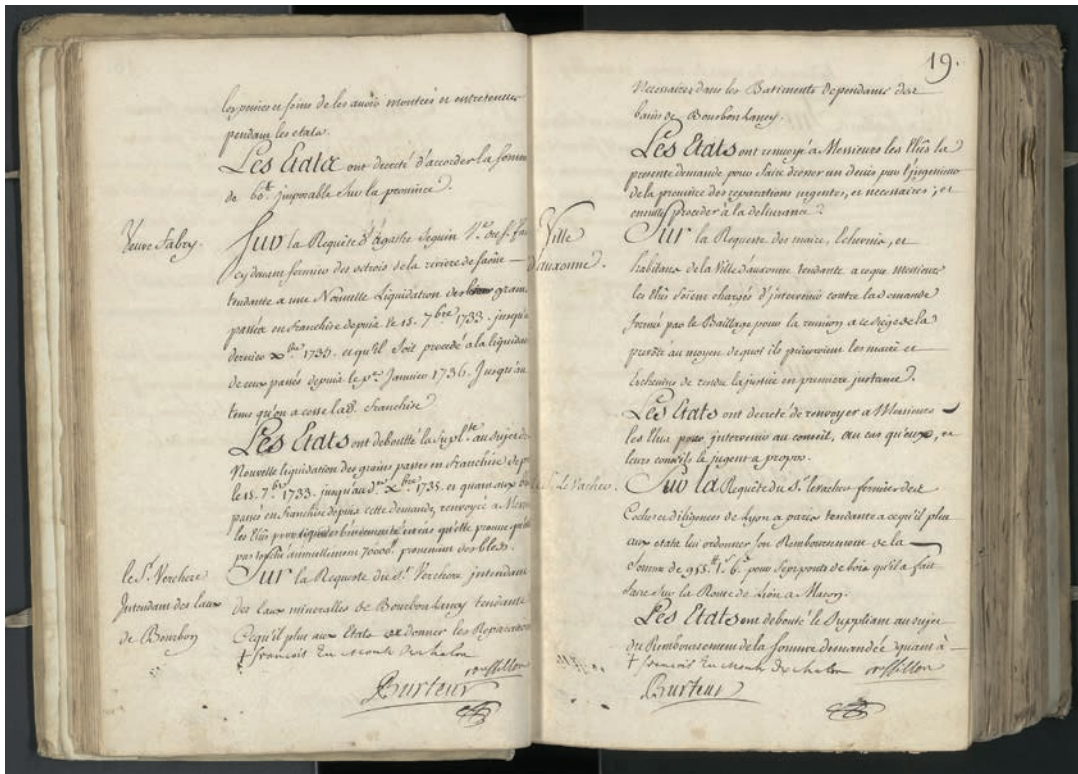
Le projet « Lettres en Lumières » naît en juin 2020 de la rencontre d'un archiviste avec un professeur d'université en intelligence artificielle, qui met sur pied un projet de recherche. Florian Fizaine commence

par consacrer son stage de fin d'année de l'École supérieure d'ingénieurs Numérique et matériaux (ESIREM) de Dijon à ce projet. Le département de la Côte-d'Or finance, depuis novembre 2021, ses trois ans de thèse. La thèse est dirigée par Annie Vinter, professeur de psychologie, ce qui donne à « Lettres en Lumières » une approche neuro-inspirée de l'écriture. La codirection en est assurée par Michel Paindavoine, spécialiste d'intelligence artificielle.

Le choix se porte sur l'écriture administrative du XVIII^e siècle pour plusieurs raisons : les projets en cours en France et en Europe ne portent pas sur ce type d'écriture ; cette écriture est répandue dans les archives françaises du XVIII^e siècle, ce qui garantit la portabilité de l'application comme l'intérêt de nombreux « clients » potentiels ; l'écriture est lisible par un jeune ingénieur du XXI^e siècle comme par des paléographes d'un niveau moyen. Le corpus choisi est la série de registres de décrets des États de Bourgogne au XVIII^e siècle étudiés par Jérôme Loiseau, professeur d'histoire moderne à l'Université de Dijon, qui fait lui aussi partie de l'équipe. Les délibérations de cette assemblée couvrent les sujets les plus divers et racontent à chaque fois une histoire, ce qui constitue une « vérité de terrain » attrayante pour les collaborateurs bénévoles.

Segmenter et transcrire

La lecture se décompose en deux étapes : segmentation des lignes de textes puis transcription ligne



Pages des décrets des États de Bourgogne, 1742, Archives départementales de la Côte-d'Or, C3004, f. 64v-65.

par ligne. Pour la première opération est utilisé un algorithme de *Deep Learning* (apprentissage profond) appelé « Mask-RCNN » ; pour la seconde est appliqué et optimisé un deuxième algorithme d'apprentissage profond intitulé « Transformer ».

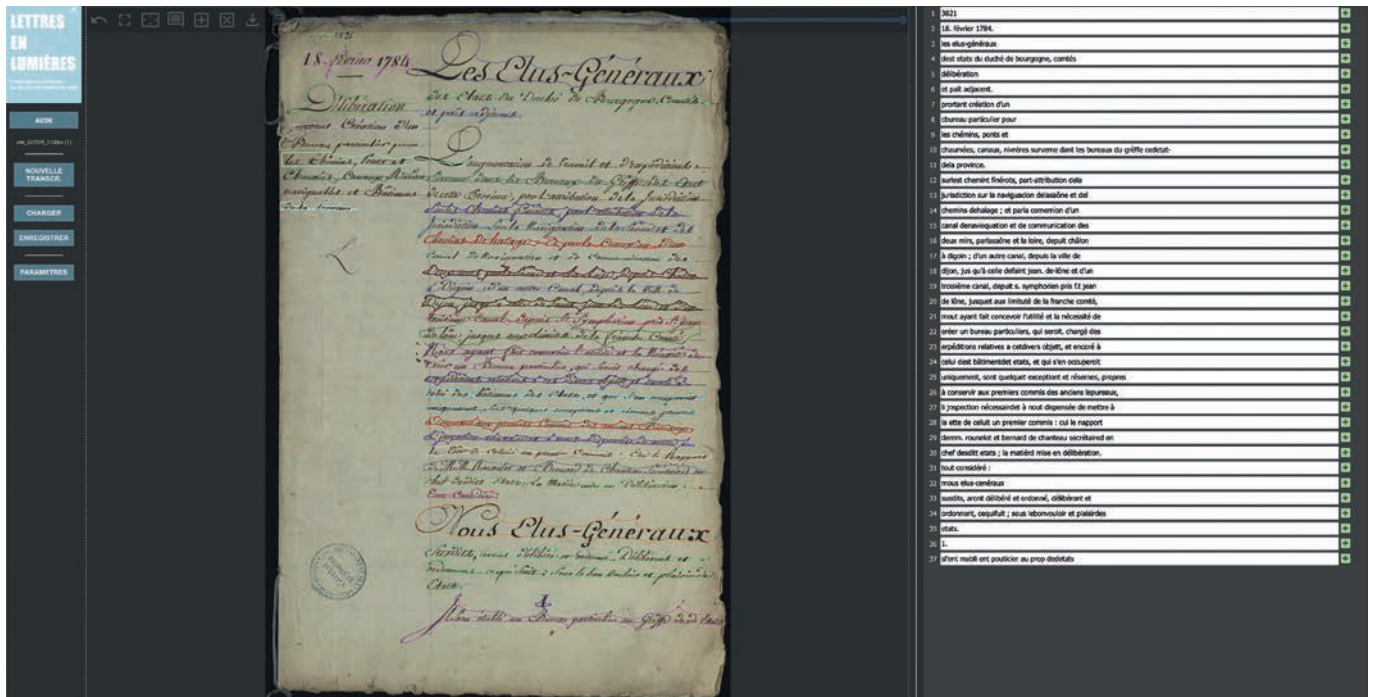
Les premiers travaux ont porté sur la segmentation ; pour la seconde année de thèse, les efforts se concentrent sur la transcription. La machine fut d'abord entraînée en « tache d'huile », c'est-à-dire en remontant le temps des registres des décrets ; elle l'est désormais en « tache de léopard », en testant

des écritures d'une autre époque du XVIII^e siècle, ou d'autres lieux, en partenariat avec d'autres services d'archives et avec l'Institut national du patrimoine (INP). On pourrait imaginer à terme que le système soit embarqué sur les scanners, afin que soient HTRisées (transcrites par reconnaissance automatique de l'écriture manuscrite)², en temps réel, les pages d'écritures aussitôt après leur numérisation.

Le texte produit est brut, l'HTRisation produit une édition de texte au kilomètre et juxtalinéaire. Les données ne sont pas enrichies, car « Lettres en

2. Proviend de l'anglais *Handwritten Text Recognition* (HTR).

Le processus de transcription automatique se découpe en deux parties : la segmentation des lignes, puis la transcription de ces lignes. Exemple de segmentation et de transcription d'une page des décrets des États de Bourgogne par « Lettres en Lumières ».



Lumières » se borne à offrir des milliers de pages de traitement de texte à la curiosité des chercheurs, à la voracité des robots et à l'ingéniosité des analystes de données (*Data Analysts*).

Interface

La particularité de « Lettres en Lumières » est d'embarquer dans le projet à la fois des paléographes volontaires et un jeune ingénieur doctorant. Les Archives départementales fournissent la matière première et organisent la transcription des lots d'images par les élèves des quatre cours de paléographie qu'elles dispensent. Puis elles valident les transcriptions

fournies au doctorant qui va en alimenter la machine, *via* une interface conviviale qui a été créée par Raphaël Lefevre, ingénieur frotté de lettres, lui aussi collaborateur bénévole du projet.

Le projet rassemble sciences dures et sciences molles ; personnels de conservation et personnels de recherche ; professionnels et bénévoles. Chacun apporte une contribution indispensable à l'objectif commun ; mais le dialogue entre les différentes parties nécessite parfois des « interfaces » humaines pour éviter les malentendus et les incompréhensions. C'est toute la richesse et la difficulté d'une véritable interdisciplinarité, moins entendue comme la juxtaposition que comme le croisement des disciplines et des profils. Le dialogue est passionnant mais exigeant. Ou, pour mieux dire : passionnant car exigeant.

Un grand nombre d'utilisateurs potentiels

L'objectif est de produire un système portable adapté aux écritures du XVIII^e siècle, que pourraient utiliser les services en charge du patrimoine écrit, sociétés savantes et associations de généalogie, laboratoires de recherche et chercheurs isolés, historiens des familles et des territoires, etc. *Open Data* ou licence ? Privé ou public ? *Start-up* ou Société coopérative de production (SCOP)³ ? Ce qui est certain, c'est que le produit final ne sera pas un produit fini, puisqu'il faudra sans cesse continuer de l'entraîner à lire de nouvelles écritures.

Mais Perrette doit pour l'heure se concentrer sur son pot au lait, c'est-à-dire sur le palais des États de Bourgogne, avant de songer à des châteaux en Espagne. ■

Le projet rassemble sciences dures et sciences molles ; personnels de conservation et personnels de recherche ; professionnels et bénévoles. Chacun apporte une contribution indispensable à l'objectif commun ; mais le dialogue entre les différentes parties nécessite parfois des « interfaces » humaines pour éviter les malentendus et les incompréhensions.



Charles-Joseph Le Jolivet, *Élévation perspective du palais des États de la province de Bourgogne, érigé sur la place royale de la ville de Dijon, 1784*. Archives départementales de la Côte-d'Or, 1.Fi.9.

3. Une SCOP est une Société coopérative ouvrière de production qui peut être sous la forme d'une Société anonyme (SA), d'une Société à responsabilité limitée (SARL) ou d'une Société par actions simplifiée (SAS). Les salariés sont en même temps les associés majoritaires de la SCOP qui peut être créée dans tous les secteurs d'activité.

Des modèles tridimensionnels pour les grottes ornées : quels usages ?

La fragilité des grottes ornées paléolithiques n'est plus à démontrer. Monuments naturels, ces sites subissent les effets anthropiques comme naturels qui mettent en danger leur authenticité et leur pérennité. À cela s'ajoutent leur accès souvent difficile et les risques encourus pour les étudier tout en les préservant. Dès leur découverte, des mesures sont désormais prises pour éviter leur dégradation. Après une présentation de la politique nationale relative aux grottes ornées quant à leur enregistrement tridimensionnel fortement préconisé par le ministère de la Culture, le projet d'un jumeau numérique de la grotte Cosquer, associé à son exploitation de façon interdisciplinaire dans le cadre de la nouvelle recherche mise en place récemment, démontre tout l'intérêt du développement de l'acquisition des données tridimensionnelles de ce patrimoine pour des usages multiples, partagés et en devenir.

Des acquisitions tridimensionnelles des grottes ornées, mais à quels desseins ?

Les acquisitions 3D des grottes ornées interviennent fréquemment dans le cadre de la conservation, comme l'analyse du microclimat souterrain de la grotte de Lascaux avec le simulateur « Lascaux » (Lacanette et Malaurent, 2010), ou encore comme support d'étude des sols, plafonds et parois ornées permettant l'analyse et la cartographie de toutes les traces repérées en préservant les sites originaux. Le Centre national de

préhistoire (CNP) expérimente actuellement, avec l'accord du Centre des monuments nationaux (CMN), l'étude de la grotte de Combarelles III, pour les relevés 3D des parois avec le logiciel libre Blender et la possibilité d'annoter en réalité virtuelle avec la technologie Unreal. Cette expérience a pour but de mesurer la distance entre l'étude en virtuel et en réel d'un site. Il est également possible de simuler des hypothèses archéologiques comme la comparaison entre des entités graphiques de deux sites contemporains (Pinçon *et al.*,

GENEVIÈVE PINÇON

Directrice du Centre national de préhistoire (CNP), Direction générale des patrimoines et de l'architecture, Sous-direction de l'archéologie, ministère de la Culture

CYRIL MONTOYA

Conservateur régional de l'archéologie adjoint, Direction régionale des affaires culturelles (DRAC) Occitanie, Service régional de l'archéologie et Laboratoire méditerranéen de préhistoire Europe Afrique (LAMPEA), UMR 7269 CNRS, Aix-Marseille Université et ministère de la Culture

SYLVAIN RASSAT

Ingénieur d'études en archéologie, LAMPEA, UMR 7269 CNRS, Aix-Marseille Université et ministère de la Culture



Habiter au Magdalénien : simulation des hypothèses archéologiques appliquée à l'abri sculpté du Roc-aux-Sorciers.



Pingouins de la grotte Cosquer.

© Equipe Cosquer - ministère de la Culture

2008) ou de rendre compte de la topographie ancienne des lieux au moment de leur fréquentation par les paléolithiques. Mais ces travaux et ces rendus innovants qui peuvent être partagés avec la communauté scientifique et les publics ne sont possibles que si la ressource 3D est de qualité et accessible.

C'est pourquoi le ministère de la Culture réalise, soutient et incite l'acquisition des nouvelles ressources 3D sur les grottes ornées. Ces relevés permettent de disposer d'un état du site (constat d'état) et en cela d'une archive patrimoniale qui peut être mise à la disposition des chercheurs et des projets de médiation comme ce fut le cas pour les centres d'interprétation Lascaux 4 (2016), Chauvet 2 (2015) et Cosquer Méditerranée (2022 avec 600 000 visiteurs depuis son ouverture). La re-matérialisation de grottes ornées demande une intervention humaine pour restituer véritablement l'épiderme des parois au plus près de la réalité. Les acquisitions numériques se font selon un cahier des charges de plus en plus précis et exigeant afin d'éviter d'avoir à revenir trop souvent dans le site original, même si ces enregistrements peuvent être refaits régulièrement pour un bon suivi de l'état sanitaire et complétés au fil des recherches.

Lors d'une découverte, les spéléologues veillent aux impacts de leur passage dans ces lieux vierges et, après l'expertise patrimoniale et archéologique, un relevé 3D est préconisé pour en avoir une archive et disposer d'une ressource sur laquelle pouvoir inventorier les vestiges anthropiques et réfléchir à leur conservation préventive.

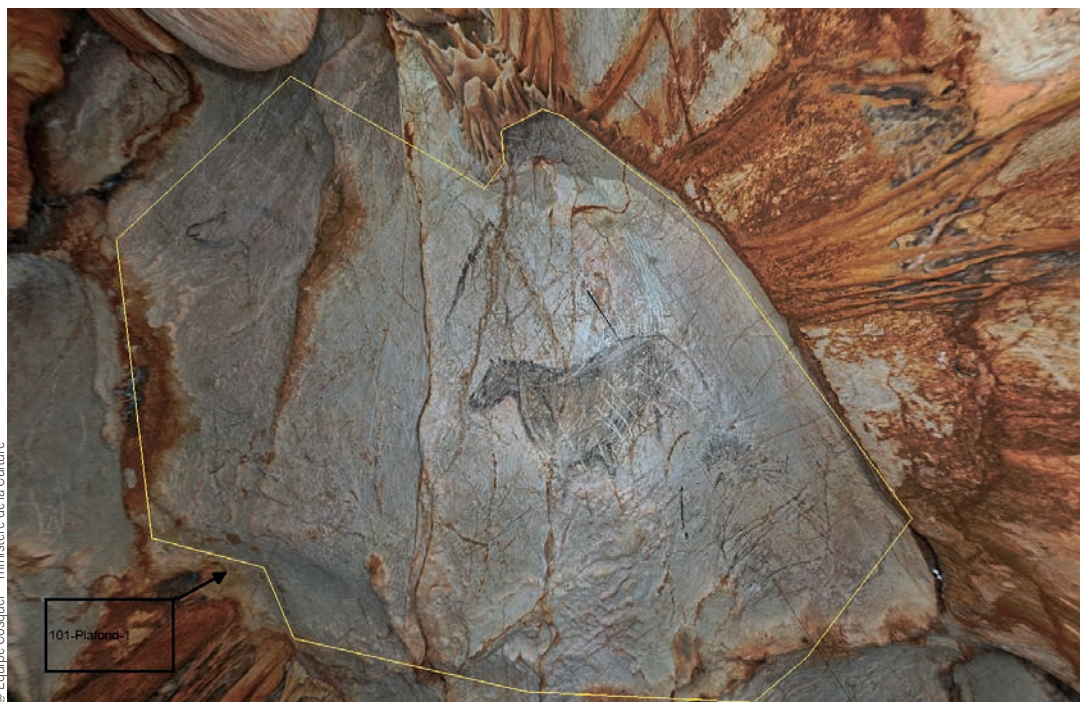
Ces données tridimensionnelles doivent être documentées précisément, techniquement et contextuellement. Leur statut de données publiques s'impose pour une utilisation la plus large possible dont beaucoup d'usages sont encore à venir. Dans le cadre du projet THEIA porté par la société Mercurio avec le CNP (appel à projet Programme d'investissements

d'avenir - PIA 4), des expériences participatives pour lire les gravures sont développées. Chacun pourra découvrir en apprenant et en faisant sien ce patrimoine exceptionnel qui d'inaccessible redevient accessible. Ces usages entraînent un nouveau positionnement des chercheurs alors amenés à évoluer vis-à-vis d'une société civile de plus en plus impliquée pour appréhender son patrimoine, ce que l'équipe de recherche de la grotte Cosquer met également en œuvre.

Grotte Cosquer et modèle numérique 3D : un outil scientifique collaboratif

Découverte probablement au moins depuis le milieu des années 1980, la grotte Cosquer (Marseille, Bouches-du-Rhône) a été déclarée par son inventeur, Henri Cosquer, en septembre 1991. Son accès (- 37 m) est aujourd'hui submergé en raison de la remontée du niveau marin depuis le début de l'Holocène (autour de 9 000-8 000 ans BP non cal.; Sartoreto *et al.*, 1995). Le site est classé au titre des Monuments historiques depuis le 2 septembre 1992.

Patrimoine d'exception en Méditerranée, cette cavité est un des sanctuaires à dispositifs pariétaux majeurs du Paléolithique supérieur européen. Les premières études (Clottes et Courtin, 1994; Clottes, Courtin et Vanrell, 2005) ont montré la richesse, la diversité et l'originalité des manifestations graphiques pariétales. Si la représentation d'espèces marines (pingouins, phoques...) reste minime sur les parois de la cavité (environ 4 % du total des entités), elle constitue néanmoins une originalité de ce site dont le littoral se situait, au plus près, à environ 6-8 km de distance lors du dernier maximum glaciaire (il y a 21 000 à 18 000 ans). La grotte conserve également de nombreuses traces d'activités au sol (foyers, systèmes d'éclairage, outils en silex...) parfaitement préservées. Les mesures radiocarbones fournissent un premier cadre chronologique de la grotte Cosquer



Extraction à partir du semis de points haute définition et d'un lecteur 3D (plafond secteur 101).

qui s'échelonne sur au moins 11 millénaires (entre 27 000 et 16 000 BP non cal. ; Valladas *et al.*, 2017).

En raison de l'accélération du réchauffement climatique, la remontée brutale du niveau marin continue de détruire les parois ornées et les sols archéologiques. Face à ce constat, la DRAC PACA a décidé, en 2017, la mise en place d'une campagne de numérisation 3D (laser et photogrammétrie) de la cavité à des fins d'études scientifiques et comme archive patrimoniale du monument (acquisition : sociétés Fugro/Immadras). L'acquisition de ces données 3D, et leur mise à disposition par l'État, a permis la réalisation de la réplique Cosquer Méditerranée. En outre, consciente de l'enjeu conservatoire et scientifique, la Direction générale des patrimoines et de l'architecture (DGPA) a décidé, en août 2020, la constitution d'une nouvelle mission scientifique, appuyée par le Département des recherches archéologiques subaquatiques et sous-marines (DRASSM), afin de développer une recherche pluridisciplinaire destinée à la sauvegarde par l'étude de ce patrimoine exceptionnel menacé.

Un modèle numérique 3D et un projet de jumeau numérique

Grâce à la convention entre la DRAC PACA et le laboratoire de Préhistoire pilotant le projet de recherche (LAMPEA, UMR 7269), l'équipe a pu entamer l'exploration et l'exploitation scientifique des données 3D (microtopographie, inventaire des entités graphiques, constats d'état, etc.). En raison de la complexité d'accès au site archéologique, le modèle 3D a été conçu, dès les premiers travaux euclidiens de l'équipe, comme un outil indispensable pour l'analyse de la topographie des lieux et l'étude immersive de la grotte dans sa totalité (volumétries, éclairage, flux d'air, etc.). Cette jumélisation est, dès lors, devenue un référentiel commun à tous les acteurs du programme de recherche qu'ils soient capables de plonger ou non.

Ce clone de la grotte Cosquer sera donc un laboratoire de recherche pluridisciplinaire et collaboratif ainsi qu'un véritable écosystème numérique en devenir.

Un accord de recherche entre le centre scientifique et technique du bâtiment (CSTB), l'unité de recherche LAMPEA et le ministère de la Culture (Sous-direction de l'archéologie-CNP) est en cours de finalisation. Toujours dans le cadre de la mission confiée par la Direction générale des patrimoines et de l'architecture, cette ouverture institutionnelle permettra au futur jumeau numérique « Cosquer » de répondre aux exigences de la science ouverte.



Vue immersive d'un assemblage des extractions des sols et plafonds du secteur 101. Vue arrière avec spéléothèmes.

Ce futur outil en quatre dimensions (référentiel euclidien et variable temporelle), regroupant les problématiques de conservation et de recherche, est, de par la complexité d'accès et d'étude, indispensable pour la gestion conservatoire et scientifique de la grotte Cosquer.

Issue de la gestion de l'aménagement du territoire et du parcours de vie des bâtiments, la maquette numérique 3D ou jumeau numérique est un outil bien défini avec des formats interoperables, standardisés (City Gml et IFC) et des chaînes opératoires/processus éprouvés (collaboratives, traçabilité des fichiers et de leurs modifications, etc.).

L'objectif est d'appliquer le maquettage 3D pour fournir, non pas une cinématique esthétisante de plus, mais un nouvel outil patrimonial aux potentialités d'études immersives importantes. En combinant les leviers 3D à l'annotation scientifique en ligne, le futur jumeau numérique « Cosquer » permettra d'analyser, de conserver et de partager cet ensemble archéologique majeur. Un accent prioritaire sera mis sur les zones en cours de destruction et celles risquant d'être détruites, à court et moyen terme, par la remontée du niveau marin. Il s'agit en outre d'optimiser l'approche et la collaboration scientifique entre les différents spécialistes au chevet de ce monument, et de permettre, grâce aux déploiements d'outils novateurs, la mise en place d'un dialogue interdisciplinaire.

Ainsi, ce jumeau numérique doit aider à la décision et à la visualisation des données complexes relevant des sciences de la terre, de l'environnement et de la société. Notre démarche se veut exemplaire quant à la protection et à la gestion conservatoire du patrimoine archéologique, archivistique et numérique et s'inscrit dans le mouvement des sciences citoyennes et d'apprentissage par la science.

Une série de hackathons dédiés à la grotte Cosquer et à son jumeau numérique est envisagée pour faire émerger un écosystème numérique « Cosquer » structuré. Des procédures de transfert de technologies (dépôts de logiciels et stages par exemple) et de valorisations scientifiques (publications, communications, etc.) pourront également être mises en place comme facteurs d'appropriation par et pour d'autres publics qu'académiques. ■

LA GROTTTE COSQUER

NUMÉRISÉE :

UN OUTIL POUR
LA CONNAISSANCE,
LA CONSERVATION
ET LA MÉDIATION

En 2016 et en 2018, les services patrimoniaux (Conservation régionale des Monuments historiques – Service régional de l'archéologie, CRMH – SRA) de la Direction régionale des affaires culturelles de Provence-Alpes-Côte d'Azur (DRAC PACA) ont élaboré un cahier des charges en vue de réaliser un relevé 3D de l'intérieur de la cavité et de ses parties immergées. Cette entreprise est confiée, après appels d'offres, aux sociétés Fugro/Immadras. Elle est menée sous le contrôle scientifique de la DRAC PACA. L'acquisition des données, en cours de réalisation pour la partie sous-marine et finalisée pour la partie terrestre, concerne 2 500 m² avec une définition inframillimétrique pour les parties ornées de la cavité, ce qui constitue pour une grotte ornée préhistorique française une première à ce niveau de précision.

Les objectifs de cette réalisation sont triples. En premier lieu, doter les services de l'État d'un outil de référence pour le suivi sanitaire de la cavité afin de prévoir si nécessaire, en concertation avec les organismes de conservation et de recherche dédiés, les mesures d'urgence utiles pour la conservation du site. Cet objectif est une priorité absolue dans la mesure où l'on sait que la grotte peut disparaître à une échéance plus ou moins lointaine suite à une conjonction de plusieurs phénomènes tels que la remontée du niveau de la mer, la sismicité de la zone et possiblement les conséquences de diverses pollutions. La numérisation permet de disposer d'un référentiel à un instant « T » pour la gestion patrimoniale et le suivi sanitaire au quotidien. Ce relevé de base sera complété par la suite, à des fréquences régulières, par d'autres relevés s'appuyant sur le relevé initial, afin de mesurer les évolutions structurelles et surfaciques des zones les plus fragilisées. La nécessité de mener des campagnes de relevés complémentaires est envisagée pour tenir compte dans le temps des progrès techniques des matériels et de leur maniabilité dans un contexte très complexe. Le deuxième objectif est de servir à la connaissance scientifique de la cavité, avant tout pour enrichir sur le plan documentaire le dossier des œuvres pariétales, peintes et gravées. Si le corpus

BIBLIOGRAPHIE

Jean Clottes et Jean Courtin, *La grotte Cosquer. Peintures et gravures de la caverne engloutie*, Seuil, 1994, 199 p., 191 fig.

Jean Clottes, Jean Courtin et Luc Vanrell, *Cosquer redécouvert*, Seuil, 2005, 256 p., 209 fig.

Delphine Lacanette et Philippe Malaurent, « La 3D au service de la conservation des grottes ornées, l'exemple de Lascaux et du simulateur Lascaux », *In Situ*, n°13, 2012. <https://journals.openedition.org/insitu/6793> ; DOI : 10.4000/insitu.6793

Geneviève Pinçon, Camille Bourdier et Oscar Fuentes, « Les sculptures pariétales magdaléniennes du Roc-aux-Sorciers (Vienne) et de la Chaire-à-Calvin (Charente) :

œuvre d'un groupe culturel magdalénien ou d'un seul et même artiste », dans Robert Vergnien (dir.), *Virtual Retrospect 2007*, Ausonius, 2008, p. 13-20. hal-01764531

Stéphane Sartoretto, Jacques Collina-Girard, Jacques Laborel et Christophe Morhange, « Quand la Grotte Cosquer a-t-elle été fermée par la montée des eaux ? », *Méditerranée*, tome 82, 1995, n° 3-4, p. 21-24.

Hélène Valladas, Anita Quiles, Emmanuelle Delqué-Kolic, Évelyne Kaltnecker, Christophe Moreau, Edwige Pons-Branchu, Luc Vanrell, Michel Olive et Xavier Delestre, « Radiocarbon dating of the decorated Cosquer Cave (France) », *Radiocarbon*, vol. 59, n°2, 2017, p. 621-633.



© Michel Olive (DRAC PACA-SRA)

Opération de relevés 3D dans la grotte Cosquer par les sociétés Fugro/Immadras.

des œuvres peintes est aujourd'hui parfaitement stabilisé, il n'en est rien pour les gravures.

L'expertise des relevés numériques réalisée doit permettre de compléter les observations d'ores et déjà enregistrées, autoriser des compléments, ouvrir à de nouvelles interprétations et entrer avec une précision jusqu'ici jamais égalée dans une analyse du geste des hommes préhistoriques. Le relevé numérique inframillimétrique permet de reconstruire le geste à l'origine du tracé, de consolider une chronologie relative des œuvres réalisées en identifiant, par des séries de coupes des traits gravés, les recoupements et la direction des épaufrures dans la paroi.

Les relevés numériques sont également une base de référence pour les travaux archéologiques programmés dans le site au cours des années à venir, compte tenu des difficultés d'accès et de travail *in situ*. Ce référentiel numérique le plus complet possible mis à disposition des équipes de chercheurs permet ainsi en amont de mieux préparer les missions dans la cavité, sachant que le temps de présence est limité et que les conditions d'accès sont très fortement contingentes par les impératifs climatiques.

Le troisième objectif est de permettre de disposer d'un document non seulement le plus précis et le plus développé possible, mais aussi utile pour la réalisation de supports de médiation afin de donner la possibilité au plus grand nombre de pouvoir visiter cette cavité, par ailleurs l'une des plus importantes actuellement connues pour la richesse de son corpus d'œuvres inventoriées (517 représentations), son iconographie unique comprenant des représentations de faune marine et sa situation géographique.

Les services de la DRAC PACA qui sont responsables de la gestion patrimoniale et scientifique de ce site archéologique, référence internationale, se doivent de disposer d'un outil permettant de répondre à ces trois objectifs et de conserver la mémoire de toutes les actions

qui sont menées dans la cavité. Pour ce faire, un système d'information géographique (SIG) est en cours de construction. Son point d'entrée est le relevé numérique de référence. Le SIG est constitué de deux parties.

La première est une partie administrative dans laquelle est intégrée toute la documentation (correspondances, rapports techniques, etc.) relative au site, depuis sa déclaration officielle auprès des services de l'État. L'accès à cette partie du SIG est réservé à l'administration.

La seconde partie intéresse directement la connaissance scientifique. Elle regroupe l'intégralité de la documentation archéologique (photos, relevés, rapports, publications). En posant le curseur sur une œuvre ou une zone, le chercheur a la possibilité de connaître alors l'intégralité de la documentation disponible constituant ce socle de connaissances. Ce second volet du SIG doit être à terme accessible à toute la communauté scientifique, après autorisation de l'administrateur du site et validation préalable du projet scientifique par la commission territoriale de la recherche archéologique sud-est. En permettant la consultation de cette base de données dont l'accès sera défini au regard du projet scientifique, l'équipe scientifique sera informée, après l'évaluation de son rapport remis à l'issue de l'opération, que la totalité de la documentation constituée sera versée dans le SIG ainsi que toutes les publications s'y rapportant. Grâce à la technologie numérique la plus performante disponible, actuellement mise en œuvre par des opérateurs de haut niveau, et au suivi par l'administration des évolutions en la matière, toutes les garanties sont réunies pour que l'enregistrement et la conservation de la mémoire de ce site majeur de la préhistoire occidentale puissent être réalisés en sachant que les générations futures n'auront peut-être pas l'immense chance d'avoir la possibilité de disposer de ce témoignage de notre passé. ■

XAVIER DELESTRE

Conservateur général du patrimoine,
Service régional de l'archéologie,
ministère de la Culture

DELPHINE LECOUVREUR

Ingénieure du patrimoine,
chargée de conservation et de
restauration des patrimoines des
Bouches-du-Rhône,
Conservation régionale des Monuments
historiques, hors bâtiments publics
d'Aix-en-Provence, ministère de la
Culture

MICHEL OLIVE

Ingénieur d'Études, Service régional
de l'archéologie, DRAC PACA,
ministère de la Culture

De la bibliothèque numérique à un outil de gestion documentaire : l'Institut national du patrimoine (INP)

L'Institut national du patrimoine a lancé en avril dernier sa nouvelle médiathèque numérique qui intègre progressivement la documentation des œuvres traitées dans les ateliers du Département des restaurateurs. Cette médiathèque est à présent un outil documentaire normé qui permet l'accès en libre consultation d'un plus grand nombre de documents¹.

ODILE BLANC

Responsable de la bibliothèque et de la documentation des œuvres, Institut national du patrimoine (INP)

CLAUDE KERMEL

Responsable du développement multimédia, INP

SANDIE LE CONTE

Adjointe au directeur des études, Département des restaurateurs, responsable du Laboratoire de recherche, INP

AMÉLIE MÉTHIVIER

Adjointe au directeur des études, Département des restaurateurs, chargée de la formation initiale, INP

OLIVIER ZEDER

Directeur des études, Département des restaurateurs, INP

De la bibliothèque numérique à un outil de gestion documentaire

La restauration des biens culturels, telle qu'elle est enseignée au Département des restaurateurs de l'INP, réserve une large place à la documentation produite lors des interventions réalisées sur les objets qui lui sont confiés : photographies sous différents rayonnements, radiographies, numérisations 3D et analyses conduites par le laboratoire, rapports de restauration des œuvres et des objets, mémoires de fin d'études rédigés par les élèves et rapports de chantier-école. Ces travaux constituent une documentation inédite qui enrichit les dossiers d'œuvres des établissements prêteurs auxquels elle doit être accessible, ainsi qu'aux professionnels de la conservation-restauration. À ce titre, elle est conservée et consultable à la bibliothèque de l'établissement.

Aujourd'hui, ce sont plus de 5 000 documents imprimés qui intègrent rétrospectivement le catalogue de la bibliothèque. Parmi eux, les 574 mémoires de fin d'étude des élèves restaurateurs, partie la plus visible, la plus consultée et accessible en ligne, figureront au catalogue du système universitaire de documentation (SUDOC) que l'INP a rejoint en 2022. Une volonté de diffusion étendue des données de la documentation des biens culturels et l'exigence actuelle d'un accès numérique ont donné un nouveau souffle à la médiathèque numérique de l'INP qui aura ainsi connu trois phases de développement.

En 2006, la bibliothèque numérique est mise en place : on y trouve, sous forme d'une liste, des bibliographies, des dossiers de formations et, après identification, les mémoires des élèves restaurateurs.

En 2010, la bibliothèque numérique devient la médiathèque numérique. Les documents sont désormais indexés par auteurs, thématiques,

mots-clés et un moteur de recherche permet d'explorer l'ensemble du fonds qui s'enrichit des rapports de jury des concours et des retransmissions audio/vidéo des manifestations scientifiques.

Enfin, en 2023, un outil de gestion des actifs numériques (DAM²) et une interface Web sont déployés. Les ressources documentaires sont désormais indexées selon les recommandations du *Dublin Core*³ et des thésaurus sont mis en place. Le fonds s'enrichit des données de la documentation des œuvres.

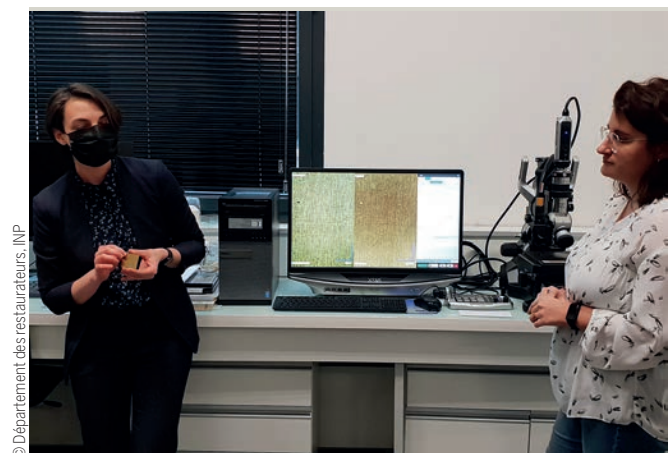
La documentation du Département des restaurateurs

Qu'il ait la forme d'une fiche ou d'un dossier, un rapport de restauration comprend un constat d'état, un diagnostic, une proposition de traitement ainsi que le traitement réalisé et, si nécessaire, un rapport d'étude et d'analyses. Ce dernier peut être de

1. <https://mediatheque-numerique.inp.fr/>

2. Pour *Digital Asset Management*.

3. Le *Dublin Core* est un format descriptif simple et générique créé en 1995 à Dublin.



© Département des restaurateurs - INP

Étude de matériaux à l'INP.



Restauration de tableau.



Restauration de tissus.

différents formats et formes qui s'étendent de l'imagerie scientifique (techniques d'image sous différents rayonnements, à différentes échelles) à l'analyse du micro-échantillon pour caractérisation quantitative chimique, ce qui rend complexe la mise en forme homogène des rapports.

Le moment du traitement permet de comprendre finement la matérialité et de synthétiser toutes les informations acquises par la recherche bibliographique, l'observation et les analyses sur les modes de fabrication (traces d'outils, aspect des revêtements, modes de fixation, assemblage, description des matériaux), sur l'histoire matérielle de l'objet, en particulier ses anciennes restaurations (matériaux et technique) et sur ses altérations.

À l'INP, la documentation constituée par les rapports de restauration, les mémoires de cinquième année et les rapports d'analyse du laboratoire est particulièrement intéressante du fait de la grande variété de typologies, de techniques, de matériaux (naturels ou synthétiques), de contextes de conservation des objets qui sont traités dans les différents ateliers (peinture, sculpture, arts du feu, mobilier, arts graphiques et livre, photographie et image numérique, arts textiles). Les mémoires constituent notamment une source d'étude technique et historique particulièrement approfondie sur des objets aussi divers qu'une armure indienne du XVI^e siècle en métal ou qu'un lé de papier peint du XIX^e siècle. Cette variété rend l'indexation des documents particulièrement capitale pour que la ressource soit exploitable.

Les rapports de chantier-école⁴ apportent la dimension du travail *in situ* dans les institutions patrimoniales, plus particulièrement dans les Monuments historiques. Ils témoignent de l'approche globale inhérente à ces projets collectifs et détaillent la démarche mise en œuvre pour mener à bien le traitement de conservation curative et/ou de restauration.

La vocation de la documentation est d'alimenter la recherche, d'aider les professionnels dans leur travail (*via* les évaluations de traitement des projets scientifiques et techniques et la bibliographie des mémoires) et d'améliorer la compréhension des techniques de

création et de conservation-restauration. Au-delà de l'utilisation des données de caractérisation par les restaurateurs pour adapter leur protocole, l'accès aux rapports d'étude est une manne pour les historiens de l'art et les scientifiques de la conservation au sens large. En effet, les approches comparatives de corpus sont indispensables à une meilleure compréhension des objets et des techniques. Cette documentation témoigne également de l'évolution des pratiques en conservation-restauration.

La médiathèque et son intégration dans le paysage des sciences du patrimoine

Les réseaux des sciences du patrimoine aux échelles régionale (Domaine de recherche et d'innovation majeurs « Patrimoines matériels – innovation, expérimentation, résilience » [DIM PAMIR]), nationale (Fondation des sciences du patrimoine, ministère de la Culture) et européenne (Cloud européen, *Joint Programming Initiative*)⁵ stimulent par leurs appels à projet et initiatives des réflexions sur l'accessibilité et le partage des données, toutes disciplines confondues, concernant les objets des collections publiques. C'est cette réflexion qui est aussi menée à l'INP pour travailler à rendre interrogeable le fruit des travaux de notre Institut. Citons en exemple le projet d'EquipEx+ ESPADON⁶ qui envisage de proposer « l'objet du patrimoine augmenté ». L'objectif sera de moissonner dans les différentes bases de données publiques celles qui sont relatives aux objets des collections. Cette opération nécessite ainsi un travail de mise à niveau de l'indexation des informations afin que ces données puissent être retrouvées.

Les rapports de restauration, d'études et d'analyses et les mémoires de fin d'études sont appelés à être moissonnés par ces outils. Notre réflexion se porte donc sur la mise en place de thésaurus construits en partenariat avec les acteurs des sciences du patrimoine. Elle porte également sur la possibilité, à partir du numéro d'inventaire de l'objet, de pouvoir accéder aussi bien à des rapports de restauration qu'à des résultats d'analyses ou des photographies. ■

4. Pour un exemple de chantier-école de l'INP, voir l'article de Pauline Chassaing et Emmanuel Poirault, « Le chantier-école du temple de Gongshu », *Culture et Recherche*, n° 143, 2022, p. 87-88.

5. Voir dans ce numéro l'article de Livio De Luca et Pascal Liévaux, p. 28.

6. Voir dans ce numéro l'article de Vincent Detalle et Romain Thomas, p. 68.

Archives, humanités numériques et études critiques en design : Problemata.org, une plateforme digitale *Open Source* pour le design...

Par quoi passe l'histoire d'une discipline comme le design ? Le musée propose des objets et des *Period Rooms*, qui, ressortant des arts appliqués, frayent avec les arts majeurs d'un temps révolu et les arts contemporains. Mais les archives, des textes et des images entre autres, sont tout autant les matériaux des chercheurs que des praticiens. En France où l'histoire du design n'est pas constituée comme champ académique et où les études critiques accusent un retard certain, Problemata.org est la tentative collective en école d'art et à l'université de permettre l'éclosion des études historiques et critiques en design.

CATHERINE GEEL

Professeure titulaire des Écoles nationales supérieures d'art (histoire et théorie du design), École nationale supérieure d'art et de design de Nancy (Ensad Nancy), École normale supérieure (ENS) Paris-Saclay, Sciences Po Paris, et Chercheure au Centre de recherche en design ENS Paris-Saclay/École nationale supérieure de création industrielle (Ensci-Les Ateliers)

1. <http://www.problemata.org/fr/>

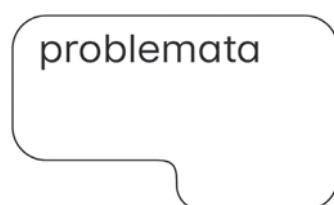
Une plateforme de recherche dédiée¹

Problemata est donc un site de ressources et de publications sur le design. Conçue comme un projet de recherche en humanités numériques grâce au CNRS, la plateforme est consultable gratuitement, s'inscrivant dans les principes de la science ouverte. Comme une bibliothèque de recherche « ouverte », elle permet autant à des chercheurs qu'à des étudiants, des curieux ou des amateurs de lire les recherches actuelles

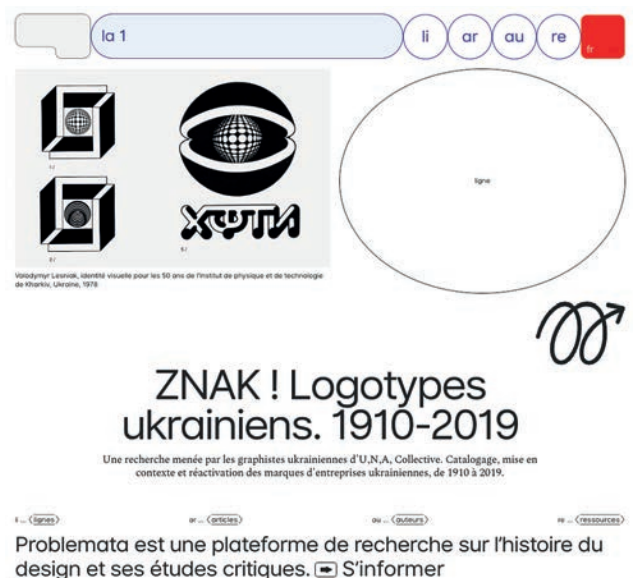
en histoire du design, de regarder et d'arpenter des collections visuelles et de retrouver ou de découvrir des épisodes, des analyses ou des hypothèses historiques nouvelles.

Manières de faire histoire autrement...

Pour le design, il n'est pas possible de rester seulement à l'échelle du poster qui récapitule l'évolution des modèles historiques de chaises ou de l'écriture



Logo Problemata.
© officeabc



Problemata,
page d'accueil.

pour décrypter les histoires des environnements et des artefacts construits dans lequel nous évoluons. Du smartphone au paysage, de la petite cuillère plaquée argent à l'affiche publicitaire, du spécimen de caractère typographique aux arts ménagers... la plateforme, structurée en « collections de ressources » et en « lignes de recherches » thématiques et problématisées, permet de publier et de diffuser des travaux où les croisements et les écarts vont des cas d'études aux théorisations.

En s'éloignant d'une histoire monographique ou des seuls objets iconiques, les chercheurs et les étudiants qui se tournent vers la recherche critique ou historique convoquent d'autres méthodes, d'autres disciplines ou des sous-champs (histoire matérielle, culture visuelle, histoire culturelle, sociologie, histoire et philosophie des techniques, culture populaire, histoire des expositions) et peuvent utiliser le genre varié des *Studies* (études).

Attraction des formes françaises

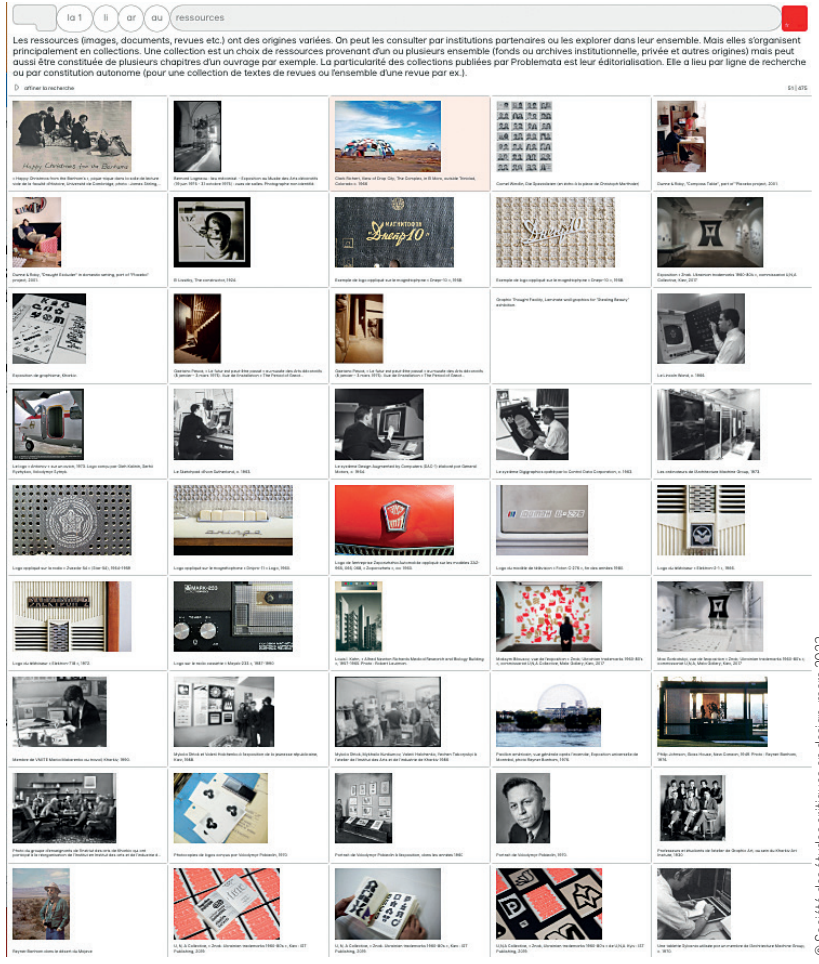
Les partenariats avec les bibliothèques des musées et des fonds permettent aux chercheurs de reconfigurer la manière d'éditer, de monter et d'organiser selon leur sujet les corpus des fonds d'archives, et aux visiteurs de découvrir des textes, des dessins et des images d'une autre façon que celle du nécessaire classement archivistique.

Problemata, plateforme *Open Source* (outil-logiciel libre), a donc aussi pour objectif de valoriser le patrimoine et la recherche francophone documentaire à un niveau international. Elle contribue à stimuler l'exploration et les recherches sur des fonds déposés dans des institutions francophones; la traduction vers l'anglais d'un certain nombre de textes de chercheurs contemporains y participe.

C'est pourquoi la plateforme a fait l'objet d'une recherche graphique et typographique, en partenariat avec l'Atelier national de recherche typographique, et développe certains outils et briques logiciels en prenant appui sur l'infrastructure de recherche² IR* Huma-Num.



Visuel Problemata. © officeabc



Problemata, page des ressources.

© Société des études critiques en design, mars 2023

Des Problemata pour interroger nos sociétés

Initiées par Aristote, les *Problemata*, sont au départ une suite de textes sous forme de questions-réponses. Devenues des compilations en volumes, elles sont élaborées jusqu'au Moyen Âge pour des disciplines scientifiques. Cette manière de formuler des problématiques et des réponses, la nature concrète des problèmes posés et les angles théoriques ou réflexifs abordés sont des modalités que la plateforme, au-delà de son nom, emprunte. L'ambition de Problemata est donc de conforter, d'augmenter et de donner plus largement accès à la façon dont les connaissances historiques des pratiques du design s'articulent légitimement aux problématiques actuelles des sociétés. ■

2. Voir dans ce numéro l'article d'Olivier Baude et Nicolas Larrousse, p. 11.

Les institutions partenaires de Problemata

- | | |
|--|--|
| Bibliothèque Kandinsky – Centre Pompidou | École supérieure d'art et de design de Reims (ESAD Reims) |
| Bibliothèque du musée des Arts décoratifs de Paris | École supérieure d'art et de design Valence Grenoble (ESAD Valence Grenoble) |
| Centre national des arts plastiques (CNAP) | Design Spot – Centre de design du cluster Paris-Saclay |
| Cité du Design | Institut supérieur des arts de Toulouse |
| Centre de recherche en design – ENSCI-Les Ateliers – ENS Paris-Saclay | Laboratoire ACCRA – EA 3402 – Université de Strasbourg |
| École nationale supérieure d'art et de design de Nancy (ENSAD-Nancy) et Atelier national de recherche typographique (ANRT) | Laboratoire ECCLA – EA 3069 – Université Jean Monnet Saint-Étienne |
| École nationale supérieure des arts décoratifs (ENSAD Paris) | Maison des sciences de l'homme (MSH) Paris-Saclay |

La Plateforme éditoriale numérique Plara : recherches et pratiques artistiques en partage

Plara, plateforme éditoriale numérique de recherche en arts, a été inaugurée en janvier 2023, en phase pilote. Elle s'articule autour de trois visées : la circulation des productions de la recherche en art au sein de la communauté des pairs ; la diffusion des processus des investigations en cours ; le partage critique collectif autour des travaux des chercheuses, chercheurs et artistes¹.

CORINNE DISERENS

Directrice de l'École nationale supérieure de Paris-Cergy (ENSAPC), août 2019-février 2023

1. <https://plara.fr>

Plara a été créée à l'initiative de trois écoles supérieures d'art : l'École nationale supérieure d'arts de Paris-Cergy (ENSAPC), l'École supérieure d'arts et médias de Caen-Cherbourg (ESAM Caen-Cherbourg) et la Villa Arson de Nice, en partant du constat qu'il n'existait pas en France d'espace de publication trans-écoles d'art pour valoriser la recherche-crédation et publier les travaux des artistes-chercheuses et chercheurs. Ces trois établissements ont développé chacun des doctorats de création, par le projet, en partenariat étroit avec des écoles doctorales. Il s'agit là de formats de recherche innovants et nourris par l'entrelacement des approches sensibles et scientifiques, des pratiques artistiques et de la pensée théorique.

L'émergence de cette plateforme vise ainsi à répondre à l'absence d'outils permettant aux artistes-chercheuses et chercheurs d'effectuer une médiation de leur travail, de partager le développement de leurs recherches et de présenter les résultats de celles-ci. À la différence d'une revue qui publie selon une périodicité fixe des contenus dont la forme est stabilisée, Plara encourage les transformations et accompagne le développement de chaque projet, depuis ses prémices jusqu'à sa finalisation, selon les temporalités et dans les formes qui lui sont propres : traductions, associations de textes, d'images fixes ou en mouvement (films et vidéos), de sons, d'applications numériques, de dispositifs 3D, de réalité virtuelle/augmentée, etc.

Plateforme évolutive et modulaire, extrêmement souple dans ses formes et rigoureuse dans ses contenus, Plara accueille la multiplicité et la complexité des dispositifs de la recherche en arts et soutient l'engagement des artistes-chercheuses et chercheurs dans un dialogue transdisciplinaire ancré tout autant dans des contextes situés que dans des altérités géographiques.



Joan Ayrton, *Tokyo, woman walking*, 2021.



© Nicola Lo Calzo

Nicola Lo Calzo, *Géraldo De Souza and the Mask of Death, Beach of Ouidah, Benin, 2011.*

À partir de l'inscription de l'expérimentation et de l'invention en art dans leurs nécessaires conditions de renouvellement et leurs méthodologies, il s'agit d'explorer archives, résonances artistiques, ancrages historiques et migratoires, configurations visuelles et discursives qui engagent des gestes artistiques et renvoient à un espace à partir duquel il est possible d'aborder les questions liées à la mémoire, à la décolonisation institutionnelle et à la construction de récits alternatifs. En d'autres termes, il est question de se situer au cœur de ce conflit de temps dont est constitué tout présent et qu'il appartient aux pratiques artistiques contemporaines d'activer.

Aujourd'hui plus que jamais, il est crucial d'aborder collectivement les modalités selon lesquelles les dispositifs de partage se doivent d'être hospitaliers à la vitalité de la recherche et à la sédimentation d'idées, et de générer des potentiels qui activent les dialogues et le pouvoir de l'art dans une multiplicité de modes

d'expérience avec des publics renouvelés, afin de favoriser l'invention d'un commun imprévisible à une époque de privatisation et privation des « communs » du monde, et de sonder leur rôle catalyseur dans la création de systèmes de connaissance au sein de configurations sociétales qui prennent en compte les changements de paradigmes culturels et politiques.

Ce sont les enjeux indissociables de la recherche, la production, la diffusion, la transmission de la création artistique et de modes de pensée et d'agir hétérogènes qui ont motivé l'émergence de Plara.

Plara a reçu le soutien du Plan de relance – Innovation et transformation numérique de l'État (ITN5). Les personnes en charge de Plara sont Brice Giacalone, responsable de la recherche de l'ESAM Caen-Cherbourg, Martina Olivero, responsable des études et de la recherche de l'ENSAPC, et Vittorio Parisi, directeur des études et de la recherche de la Villa Arson de Nice. ■

ArchiRès/Mir@bel : un partenariat qui valorise la science ouverte en architecture

Le réseau Mir@bel (Mutualisation d'informations sur les revues et leurs accès dans les bases en ligne)¹ s'inscrit dans le long mouvement d'ouverture des revues scientifiques en répertoriant de nombreux périodiques dont ceux aux contenus scientifiques librement accessibles. Le réseau ArchiRès², regroupant les bibliothèques d'écoles nationales supérieures d'architecture et de paysage (ENSA-P) du ministère de la Culture, collabore au sein de Mir@bel afin de valoriser les périodiques traitant d'architecture et de domaines associés.

CYRIELLE AVISSE

Documentaliste, École nationale supérieure d'architecture de Lyon (ENSAL), référente Mir@bel pour le réseau ArchiRès

YANNICK BERNARDIE

Aide-documentaliste, École nationale supérieure d'architecture de Marseille (ENSAM), référent Ressources numériques au sein de la commission « Périodiques » ArchiRès

Mir@bel, acteur essentiel du référencement de la science ouverte

Depuis 2009, le réseau Mir@bel rassemble des professionnels de la documentation autour de sa plateforme Web de signalement de revues. Piloté par Sciences Po Lyon, Sciences Po Grenoble, l'École nationale des travaux publics de l'État et la Maison des sciences de l'homme de Dijon, il compte plus de 130 partenaires et met à disposition des internautes les données de 15 000 revues dont 10 000 proposent du texte intégral en accès libre.

La plateforme donne à voir l'écosystème complet de chaque revue signalée :

- informations bibliographiques ;
- liens d'accès aux contenus (texte intégral en ligne, sommaires, résumés et indexation des articles) ;
- liens vers les réseaux sociaux ;
- liens vers les informations de bibliométrie et de politique d'accès (base de référencement des revues en accès ouvert *Directory of Open Access Journals* (DOAJ)³, répertoire des ressources scientifiques et universitaires en libre accès ROAD⁴, base de données Sherpa Romeo⁵ regroupant les politiques de libre accès des éditeurs) ;
- affichage des labels (sceaux DOAJ et Latindex pour les revues académiques éditées en Amérique ibérique et lusophone).

Le référencement des revues, la mise à jour des données ainsi que le contrôle qualité de ces dernières sont réalisés par les partenaires *via* une veille manuelle. Elle est complétée par des traitements automatisés permettant : l'import des accès en ligne pour certaines ressources (Cairn.info, Érudit, OpenEdition Journals, Persée...) ; l'enrichissement des données par la mise en place d'un cercle vertueux de curation des données avec l'Agence bibliographique de l'enseignement supérieur (Abes) et les centres ISSN⁶ (centre ISSN France sous l'égide de la Bibliothèque nationale

de France [BnF]) attribuant le numéro international qui permet d'identifier une revue ; l'ajout de liens ou de rebonds vers d'autres plateformes en ligne (DOAJ, ROAD, Sherpa Romeo, Latindex, archives ouvertes HAL...).

Mir@bel propose des données ouvertes et réutilisables grâce à son API (interface qui permet à deux applications de communiquer entre elles). Elles peuvent enrichir les catalogues de bibliothèques (dont les logiciels sont compatibles) ou être affichées dans les outils de découverte (systèmes de recherche destinés aux bibliothèques). Elles sont également utilisées par l'Abes (catalogue du Système universitaire de documentation SUDOC, Base de connaissance nationale BACON et application Web IdRef [Identifiants et Référentiels]) et lors des campagnes de subventions aux revues de l'Institut des sciences humaines et sociales (InSHS) du CNRS (obligation pour les revues d'avoir une entrée sur le portail).

Depuis février 2022, les éditeurs de revues académiques françaises peuvent déclarer et rendre publiques leurs politiques éditoriales sur Mir@bel. Après un processus de modération, elles y sont mises en ligne puis transmises à Sherpa Romeo afin d'accroître la visibilité de ces revues à l'international.

Lauréat de l'appel à projets du Fonds national pour la science ouverte avec Mir@bel2022, le réseau vise à ancrer les revues scientifiques françaises à l'international, en les accompagnant vers la base de référencement des revues en accès ouvert *Directory of Open Access Journals* (DOAJ) et en signalant toutes les revues et éditeurs français.

La valorisation des périodiques numériques dans les bibliothèques des ENSA-P

Le réseau ArchiRès regroupe les bibliothèques des 20 ENSA-P, placées sous la double tutelle du ministère de la Culture et du ministère de l'Enseignement



Logo du réseau Mir@bel (logo sous licence ouverte Etalab).

1. <https://reseau-mirabel.info/>

2. <https://www.archires.archi.fr/accueil>

3. <https://doaj.org/>

4. <https://road.issn.org/>

5. <https://v2.sherpa.ac.uk/romeo/>

6. L'*International Standard Serial Number* ou « Numéro international normalisé des publications en série » est un numéro international qui permet d'identifier de manière unique une publication en série. Il concerne donc les journaux, les revues et les collections de monographies, quel que soit le support.



Logo du réseau ArchiRès réalisé à l'occasion des 50 ans du réseau.

supérieur et de la Recherche. La commission « Périodiques » du réseau assure la gestion des fonds de périodiques et engage des actions communes pour assurer des missions de conservation, valorisation et diffusion de ces collections. Dans cette commission, la référente Mir@bel, sous l'égide du référent pour les périodiques numériques, est l'interlocutrice principale pour les questions concernant Mir@bel. Depuis 2015, 16 ENSA-P (sur 20!) participent en effet à l'enrichissement et à la mise à jour de la plateforme pour les revues liées aux thématiques de l'architecture, de l'urbanisme, de l'aménagement du territoire et du paysage (sans statuts juridiques, ArchiRès ne peut conventionner en son nom. Le partenariat est donc entendu entre chaque école et Mir@bel).

Cet engagement répond à la volonté des bibliothèques des ENSA-P de donner à leurs usagers accès à une information de qualité pour les périodiques qu'elles mettent à disposition sous format papier et en ligne, *via* un abonnement ou en accès libre. Cette collaboration permet également de valoriser les travaux de numérisation rétrospective réalisés par certains éditeurs et institutions, dont la Cité de l'architecture et du patrimoine (partenaire d'ArchiRès depuis 2008) et la BnF.

Seules 134 revues concernant l'architecture, 81 l'aménagement du territoire et 135 les études urbaines et rurales sont actuellement répertoriées, mais ArchiRès entend améliorer et intensifier le signalement des titres sur ces thématiques en augmentant le nombre de revues suivies par les ENSA-P (actuellement 103) et en créant de nouvelles notices dans Mir@bel.

Possédant des périodiques du monde entier et seul réseau documentaire français spécialisé en architecture, ArchiRès participe régulièrement à l'attribution d'ISSN pour les revues sur support papier ou électronique. Son portail met à disposition des internautes une liste de revues disponibles en ligne et en libre accès, *via* la rubrique « Revues en accès ouvert » de l'onglet « Ressources en ligne ». Dans le catalogue commun, les notices de périodiques référencés dans Mir@bel comportent la mention « @ Accès en ligne disponible ». En cliquant sur cette dernière, l'utilisateur est redirigé vers la notice du périodique sur Mir@bel. Cette interface, personnalisable en fonction de son établissement, permet de lister les revues présentes dans Mir@bel et disponibles dans sa bibliothèque, d'accéder directement à une revue en ligne lorsque sa bibliothèque dispose d'un abonnement et que la connexion est possible *via* un serveur proxy. En facilitant l'accès aux périodiques numériques à distance, Mir@bel est un outil précieux.

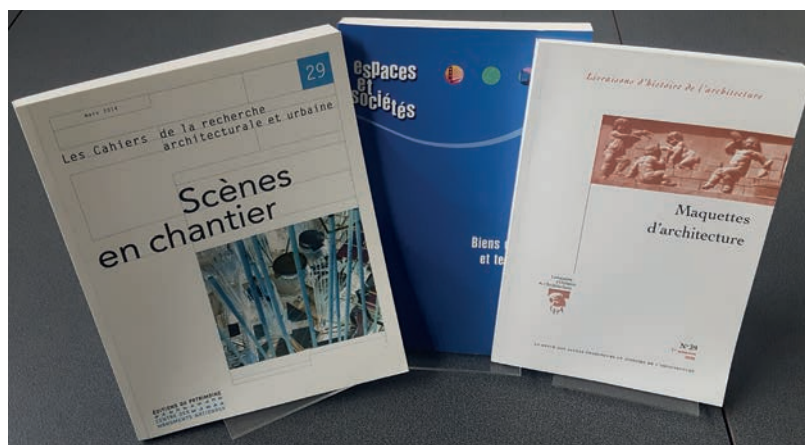
ArchiRès poursuit ses efforts pour la science ouverte

La mise en ligne du nouveau portail ArchiRès en 2022 a permis l'émergence d'une offre numérique commune pour les usagers des ENSA-P. Pour l'instant, seuls les livres numériques et vidéos à la demande sont concernés, mais une étude est en cours pour analyser la mise en place d'une telle offre pour les périodiques numériques. Les éditeurs commerciaux de revues d'architecture proposent rarement une offre numérique adaptée aux usages des bibliothèques (connexion avec adresse électronique nominative et mot de passe ou par adresse IP), ce qui en rend l'accès à distance souvent impossible pour nos lecteurs. Ces revues étant référencées dans Mir@bel, leur visibilité et celle de leurs éditeurs sont accrues. Ainsi, le réseau ArchiRès espère que ces derniers seront plus disposés à revoir les modalités de consultation en ligne de leurs publications périodiques (*via* une passerelle proxy par exemple).

Les veilleurs Mir@bel des ENSA-P, au sein de la commission des « Périodiques » ArchiRès ou du groupe Lab&doc de ses bibliothèques recherche, sont des interlocuteurs privilégiés auprès des enseignants-chercheurs exerçant dans les laboratoires de recherche des ENSA-P pour favoriser la sensibilisation au libre accès et la transition numérique des revues qu'ils éditent.

Le mouvement d'ouverture des revues scientifiques connaît un timide essor en France dans le domaine de l'architecture et le projet Mir@bel2022 peut devenir un véritable levier pour convaincre les éditeurs concernés de participer significativement au mouvement d'ouverture de la science. En étant mieux référencées et plus visibles à l'international, ces revues verront leur notoriété progresser. ■

Sélection de revues référencées dans Mir@bel proposant du texte intégral librement accessible en ligne.



Les recherches culturelles participatives dans la science ouverte

L'implication des citoyens, aux côtés des chercheurs, dans la construction commune de connaissances scientifiques caractérise les démarches de recherches participatives. Si ces pratiques de participation des non-scientifiques professionnels dans les recherches sont depuis longtemps présentes et visibles dans le champ des sciences de l'environnement (observatoires de la biodiversité, mesure collaborative de la qualité de l'air, détections de cratères...), elles se développent également dans les domaines de la culture.

ALEXANDRA VILLARROEL

Coordinatrice Vigie-Muséum et Particip-Arc, Muséum national d'Histoire naturelle (MNHN)

FRANÇOIS MILLET

Directeur de projets Sciences & Société, Le Dôme – Relais d'osciences

Que ce soit pour étudier de nouvelles topologies de mises en scène, caractériser l'environnement sonore d'un territoire, documenter, transcrire et enrichir des archives en ligne, ou encore modéliser collectivement des propositions d'aménagement urbain, toutes ces initiatives mettent en œuvre une articulation entre différentes natures de savoirs, dans un objectif de recherche. Les recherches culturelles participatives embarquent ainsi une grande diversité d'acteurs : universités, centres de culture scientifique et technique, archives, bibliothèques, associations, sociétés savantes, tiers lieux...

Les sciences participatives au cœur des enjeux de la science ouverte

Si la définition usuelle de la science ouverte reste trop souvent cantonnée aux principes d'accès ouvert aux données et aux publications, il est intéressant de souligner la conception promue par l'UNESCO. Elle définit ainsi la science ouverte comme

« un concept inclusif qui englobe différents mouvements et pratiques visant à rendre les connaissances scientifiques multilingues, librement accessibles à tous et réutilisables par tous, à renforcer la collaboration scientifique et le partage des informations au profit de la science et de la société, ainsi qu'à ouvrir les processus de création, d'évaluation et de diffusion des connaissances scientifiques aux acteurs de la société au-delà de la communauté scientifique traditionnelle¹ ».

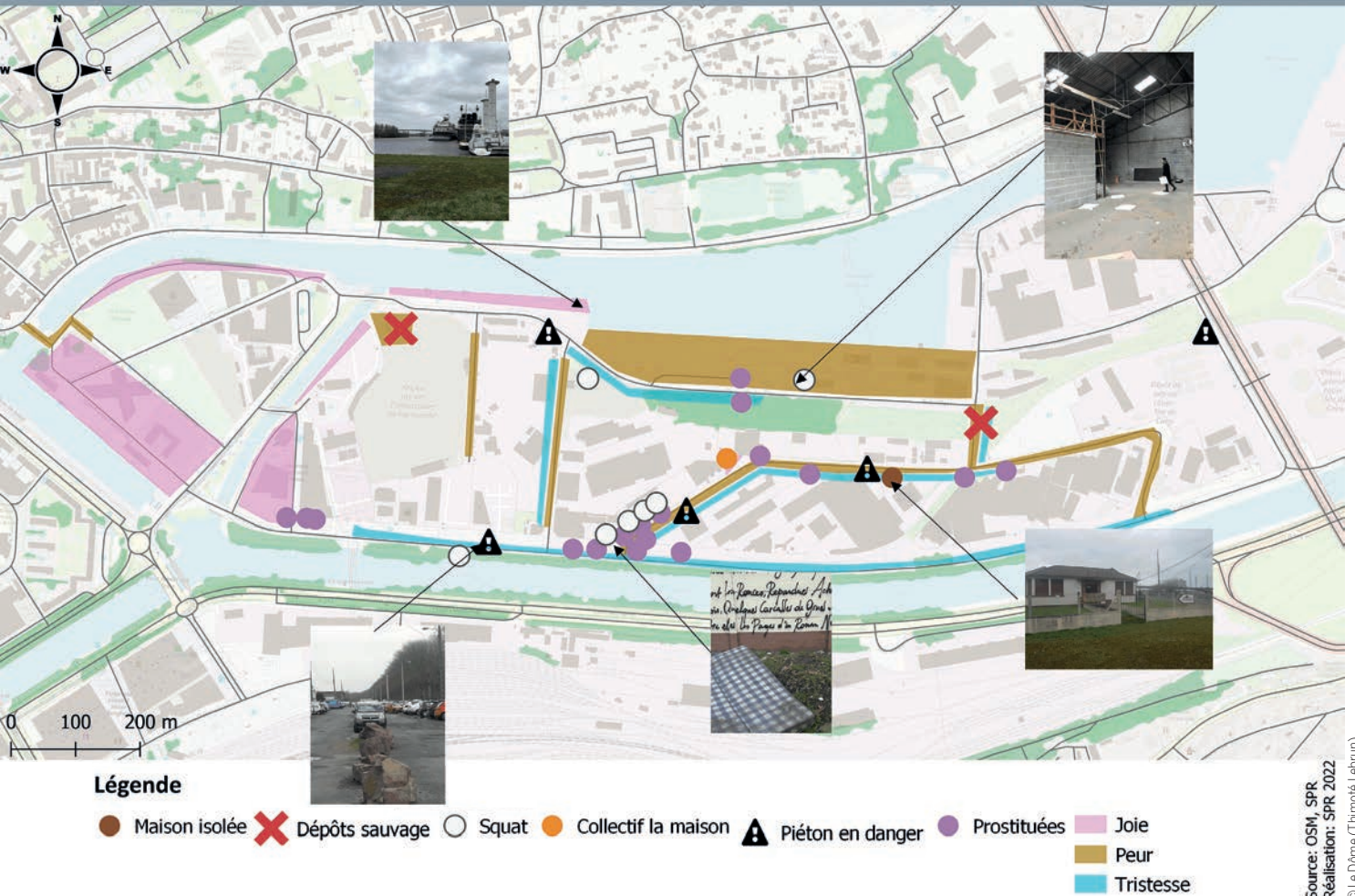
La notion de connaissance, plus que les données, est ici mise en avant, et l'ouverture dans l'élaboration même de ces connaissances, y compris la co-construction des problématiques de recherche et des protocoles de collectes de données, apparaît comme centrale.

Une telle définition renverse ainsi le paradigme et montre de quelle façon les recherches culturelles participatives, de par leur articulation, à tous les niveaux de la construction de la donnée et de la connaissance, entre différents acteurs académiques et non académiques, sont au cœur des enjeux de la science ouverte. Celles-ci intègrent, en effet, les quatre piliers énoncés par l'UNESCO : connaissances scientifiques ouvertes, infrastructures de la science ouverte, participation ouverte des acteurs de la société et dialogue avec d'autres systèmes de connaissance. Les dispositifs de participation et d'animation des communautés de participants favorisent les interactions entre les savoirs portés par les chercheurs et par les contributeurs. C'est par le dialogue et la compréhension de enjeux et attentes mutuels que la dynamique collective peut s'enclencher. Le cas échéant, les plateformes numériques adossées aux projets permettent aux participants d'accéder à l'ensemble des contenus produits par la communauté, et ouvrent parfois même la possibilité de commenter (ajout de métadonnées) ou de valider collectivement les données produites. Au-delà de la collecte, l'analyse des données, la diffusion et la publication des résultats peuvent également faire l'objet d'une co-construction. L'enjeu des dispositifs à mettre en place ne se situe donc plus tant dans l'ouverture et l'accessibilité à tous des données produites, mais dans la façon d'embarquer les citoyens à tous les niveaux de la construction des connaissances.

En effet, si l'ouverture des données s'adresse avant tout aux communautés professionnelles, et éventuellement amatrices, elle n'est pas à destination du public. Les recherches culturelles participatives portent donc un double enjeu d'une mise en culture de la donnée.

1. Projet de recommandation de l'UNESCO sur une science ouverte : https://unesdoc.unesco.org/ark:/48223/pf0000378841_fre

Emotions ressenties lors des balades urbaines sur la presqu'île de Caen



Retranscription et mise en forme de productions de participants à un atelier de co-urbanisme.

Prendre part, apporter sa part, retirer une part...

Le premier enjeu est celui de la participation de publics dans ces recherches participatives, et celle-ci ne se décrète pas. Selon Joëlle Zask², participer nécessite de prendre part, d'apporter sa part et de retirer une part. *Prendre part* signifie être partie prenante d'un groupe, d'un collectif ou d'une communauté, c'est être avec autrui. *Apporter sa part* correspond ici à la production de données pour les recherches culturelles participatives, qu'elles soient brutes ou analysées. Mais c'est en grande partie sur le fait de *retirer une part* que reposent l'adhésion et la contribution à la production de données pour ces programmes. Cette part qui est retirée constitue un levier de la participation qui doit être reconnu. Il peut s'agir d'une reconnaissance en rémunération, qu'elle soit financière ou en nature. Des leviers existent aussi en termes de valorisation individuelle des personnes participantes ou bien d'une affirmation militante dans le thème proposé. La personne participante peut également en retirer une part en termes de savoir ou de savoir-faire acquis. Assurer un accès aux données constitue alors la part la plus simple et directe que peuvent retirer les participants. C'est un moyen pour les équipes de recherche, l'institution culturelle ou

de médiation, de redonner à chacun le produit de sa contribution.

À l'heure où la collecte de données et sa capitalisation constituent un nouveau modèle économique basé sur l'instrumentalisation et la privatisation de la créativité collective, redonner à chacun le fruit de sa contribution constitue une marque déontologique qui peut également assurer l'adhésion nécessaire aux projets. Documenter et restituer une recherche culturelle participative relève d'une démarche de transparence qui contribue à la confiance et à la reconnaissance de la structure de médiation partie prenante de sa mise en œuvre. Cette documentation et ces données permettent également la reconnaissance de la contribution des publics à la démocratie en tant que participation des individus à l'action collective.

Le deuxième enjeu est donc d'être en mesure de restituer cette part produite par les personnes participantes *via* une documentation adaptée. Qu'il s'agisse de data visualisation ou de compte rendu d'atelier, qu'elle soit produite par les équipes de recherche elles-mêmes ou par une structure de médiation, cette documentation permet également de maintenir la participation et de recruter de nouvelles personnes. Elle contextualise le projet, indique aux potentielles recrues son état d'avancement et témoigne de son

2. Joëlle Zask, *Participer. Essai sur les formes démocratiques de la participation*, Le bord de l'eau, 2011, 200 p.



Atelier sur le site archéologique de Bibracte.

© Projet Bulliot, Bibracte et Moi (transcription collaborative des archives du site de Bibracte et plus particulièrement des carnets de fouille de Jacques-Gabriel Bulliot, inventeur de Bibracte au 19^e siècle)

4. Nicolas Belett Vigneron, Sébastien Beyou, Xavier Coadic et Émilie Picton, « De l'hypothèse de la documentation comme technique de résistance et du wiki comme objet de ces résistances », *Sens Public*, 2019, 36 p.

caractère participatif en incarnant ou en rendant visibles la diversité des participations et leur accessibilité. À la différence des données utilisées dans les sciences participatives, davantage tournées vers le quantitatif, les documents et productions des recherches et innovations participatives sont aussi de nouvelles formes de données qualitatives.

Une culture commune

Reste que cette action de documentation confère un pouvoir à la structure qui la produit. C'est elle qui détient la capacité de produire du commun avec la parole ou la donnée brute des personnes participantes. Un pouvoir qui est parfois délégué à une personne ou structure tierce à qui l'on confie le travail de restitution. Cette posture et ce pouvoir qui pourraient

être perçus comme contradictoires avec les aspirations de la science ouverte sont surtout un palliatif à une absence, une invitation à la documentation, qui semble en réalité n'être le fait que d'un petit nombre d'individus au sein de tels groupes sociaux⁴.

C'est en ce sens que les recherches culturelles participatives ont un rôle moteur à jouer au sein de la science ouverte. De par cet effort de co-construction de connaissances, de documentation, d'écriture collective, de systématisation des données produites avec les participants, elles dépassent la question de l'ouverture et de la mise à disposition des données. Elles mettent en récit ces productions collectives dans l'optique de les rendre réellement appropriables par le plus grand monde et de créer une culture commune. ■

À la différence des données utilisées dans les sciences participatives, davantage tournées vers le quantitatif, les documents et productions des recherches et innovations participatives sont aussi de nouvelles formes de données qualitatives.

Plateformes culturelles contributives face au défi de l'ouverture de données

Avec le terme « plateformes culturelles contributives », nous faisons référence à une large variété de dispositifs numériques qui donnent au citoyen la possibilité de partager en ligne ses savoirs, ses productions, ses récits ou ses mémoires. Le spectre est très étendu et peut aller des plateformes où l'artiste amateur dépose ses œuvres d'art numériques et l'internaute ses critiques, comme DeviantArt, Babelio ou SensCritique, à des projets institutionnels ou scientifiques de transcription ou d'annotation de manuscrits tels que Transcribe Bentham ou Les herbonautes.

Ce type de dispositif est devenu très populaire auprès des institutions culturelles (archives, musées, bibliothèques, etc.) qui ont trouvé dans les plateformes numériques un outil indispensable pour faciliter leurs démarches de recherche participative.

Nous pouvons citer à titre d'exemple le blog Archives nationales participatives¹ qui donne accès aux différentes campagnes de transcription collaborative de l'institution, le projet Passeports intérieurs² des archives de Saint-Brieuc qui s'appuie sur la plateforme Zooniverse ou encore la plateforme Histoire de Nature du Muséum national d'Histoire naturelle, qui permet aux citoyens de déposer des témoignages ou récits autour des changements environnementaux³.

À ce propos, il est intéressant de mentionner le réseau Particip-Arc⁴ qui a été lancé en 2019 sous l'égide du ministère de la Culture et avec le pilotage du Muséum national d'Histoire naturelle. L'objectif de ce réseau est de faciliter les échanges entre acteurs portant des projets de sciences participatives et de recherches culturelles ou intéressés par ce type de démarche. Dans ce contexte, un groupe de réflexion a travaillé autour du rôle des plateformes contributives⁵ pour construire des corpus de données standardisées selon une méthode collaborative et scientifique. La question du statut de données collectées s'est révélée particulièrement controversée. Si les institutions universitaires sont désormais sensibles aux obligations de

MARTA SEVERO

Professeure des universités, directrice adjointe du Laboratoire Dispositifs d'information et de communication à l'ère numérique (DICEN), Équipe d'accueil CNRS 7339: Conservatoire national des arts et métiers, Université Gustave Eiffel, Université Paris-Nanterre

1. <https://archivnat.hypotheses.org/>
2. <https://www.zooniverse.org/projects/archivesdesaint-brieuc/passports>
3. <https://www.changing-natures.org/fr/>
4. <https://www.participarc.net/>
5. Voir dans ce numéro l'article de François Millet et Alexandra Villarroel, p. 126.

particip ARC

QUI SOMMES-NOUS ? PARTICIPEZ ! ANNUAIRE DES PROJETS CENTRE DE RESSOURCES ACTUALITÉS & ÉVÉNEMENTS

particip ARC

Le réseau des acteurs engagés pour les sciences participatives dans les domaines de la culture.

Conservation du patrimoine, archives, archéologie, architecture et aménagement urbain, informatique documentaire, musique, arts du spectacle, patrimoine écrit, ethnologie, sociolinguistique, arts plastiques... Dans tous ces secteurs, des démarches participatives liées aux projets de recherche émergent, accompagnées de leur lot de questionnements spécifiques.

Quelles sont les spécificités des recherches participatives dans le champ culturel ? Comment mieux associer les différents publics à tous les stades de la recherche ? Comment accueillir les volontés citoyennes de mener des recherches ? Comment restituer les résultats aux partenaires, aux participants et au public en général ?

Le réseau Particip-Arc, soutenu par le Ministère de la Culture, ambitionne la mise en réseau des acteurs des recherches culturelles participatives, pour favoriser le partage d'expériences et l'analyse réflexive de pratiques mais aussi pour promouvoir et développer les sciences participatives dans les domaines de la culture.

EN SAVOIR PLUS

Le portail de Particip-Arc
(<https://www.participarc.net/>).



ANR COLLABORA
(Couverture-logo de *Culture en partage*:
<https://anr-collabora.parisnanterre.fr/index.php/2022/05/19/parution-culture-en-partage/>).

6. Cet article a été réalisé dans le cadre du projet ANR COLLABORA (ANR-18-CE38-0005). Pour plus d'informations, voir : <https://anr-collabora.parisnanterre.fr/>

7. <https://anr-collabora.parisnanterre.fr/observatoire/>

8. Ces plateformes sont présentées de manière détaillée dans l'ouvrage : Marta Severo, Sébastien Schulz, et Olivier Thuillas, *Culture en partage. Guide des plateformes culturelles contributives*, FYP Éditions, 2022.

9. Cette licence concerne les bases de données et oblige ceux qui utilisent les données d'une base à les maintenir ouvertes et disponibles sous la même licence.

10. Les licences Creative Commons (CC) sont dédiées aux contenus et autorisent un droit d'accès sur un contenu, tout en le combinant avec différents types d'obligations, comme la nécessité d'en mentionner la paternité (BY), l'impossibilité d'en faire un usage commercial (non commercial, NC) ou encore le devoir de proposer sous la même licence (*Share-alike*, SA) le contenu utilisé.

l'ouverture de données, cela reste plus complexe dans d'autres milieux où des questions de droits d'auteur peuvent entraver l'ouverture. Parfois, pour une institution culturelle, il n'est pas évident de mettre à disposition sous licence libre des corpus de données liés à ses collections parce que l'ouverture doit se confronter à des contraintes d'ordre juridique, éthique et parfois politique.

Dans le cadre du projet ANR COLLABORA⁶, nous avons construit un observatoire des plateformes contributives culturelles pour rendre compte de la diversité de ce phénomène⁷. En nous appuyant sur cet observatoire, nous pouvons approfondir la question de l'ouverture de données sur un échantillon de 54 plateformes⁸. Cet observatoire n'a pas seulement concerné des plateformes institutionnelles et publiques, mais il a aussi cherché à couvrir la variété du phénomène en considérant des réalités associatives et commerciales.

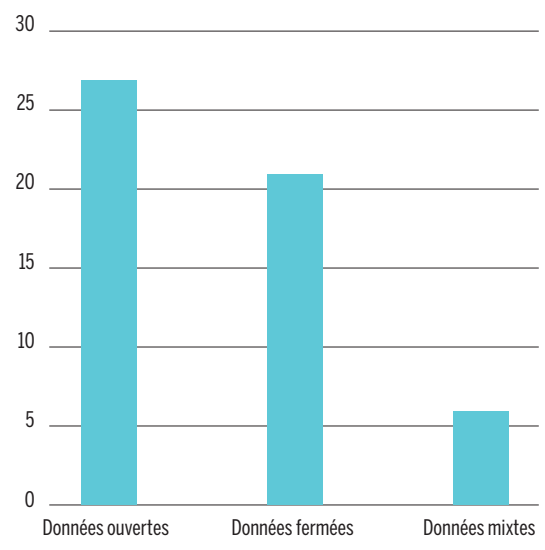
Ces plateformes hébergent une grande diversité de données : vidéos, images, textes, etc. La moitié d'entre elles mettent à disposition ces données dans un format que nous pouvons qualifier d'« ouvert ». Ce terme regroupe les licences qui autorisent au minimum le libre accès.

Par exemple, les organismes publics, comme le ministère de la Culture avec sa plateforme #culturechezvous, sont légalement tenus de mettre à disposition leurs contenus en données ouvertes. Certaines plateformes n'en ont pas l'obligation légale, mais défendent le principe de la libre circulation de l'information. Elles attribuent à leurs données les licences Open Database Licence⁹ (par exemple Open Food Facts) et Creative Commons¹⁰ (comme Nantes Patrimonia).

Pour les projets qui privilégient le droit d'auteur et la propriété intellectuelle, nous avons qualifié leurs données de « fermées ». Ces plateformes ne permettent que la consultation du site pour un usage personnel. Ainsi, sur la plateforme Arte Audio Blog, toutes les composantes du site, comme sa structure générale, les sons et les logos, sont protégées et soumises au droit de propriété intellectuelle.

Enfin, une dernière catégorie relève de plateformes dont les données sont « mixtes ». Certaines plateformes combinent des données ouvertes et des données fermées (voir figure ci-dessous). Ainsi, la plateforme Wiki-Brest recueille des contenus issus d'habitants de la région. Si la licence Creative Commons leur est recommandée, chaque contributeur est libre de choisir la licence qui lui convient le mieux, selon qu'il souhaite autoriser la réutilisation de son contenu, son usage commercial ou sa modification.

Distribution des plateformes par statut des données



Il faut noter que le statut juridique des données n'est souvent pas clair. Certaines plateformes n'indiquent pas explicitement quels sont les usages autorisés de leur contenu. Cela génère des ambiguïtés qui ont conduit à des situations de controverses importantes mettant en péril la survie même des plateformes. Par exemple, fin 2013, la National Music Publishers Association (NMPA) a demandé à cinquante sites Web, dont la plateforme Rap Genius (devenue Genius), de retirer les paroles et les chansons qu'ils utilisaient sans en avoir reçu l'autorisation légale. Quelques mois plus tard, Genius a annoncé avoir passé des accords commerciaux avec des majors de l'industrie musicale, et a pu ainsi poursuivre son activité d'explication des paroles de chansons de rap.

En conclusion, ces exemples nous montrent la complexité de la question de l'ouverture de données dans le domaine culturel entre dynamiques institutionnelles et commerciales. Cependant, nous pouvons observer un mouvement général vers l'ouverture qui permettra de réfléchir à de nouvelles initiatives de circulation et de réutilisation des données autour de nos objets patrimoniaux et culturels. ■

Les Cahiers du Studio – Théâtre de Vitry

Les *Cahiers du Studio*, c'est l'histoire d'une application numérique qui collecte, archive, édite et publie le récit d'un processus de recherche ou de création théâtrale. L'objet, lui-même sujet de recherche, est aussi un outil ouvert et collaboratif de documentation pour une prise de notes multimédias.

Les *Cahiers du Studio* ont été conçus dans le cadre du projet artistique de la metteuse en scène Bérangère Vantusso pour le studio-théâtre de Vitry qu'elle dirige depuis 2017. Fondé par le metteur en scène Jacques Lassalle, il est un lieu de créations et de recherches scéniques qui accueille des équipes artistiques en résidences.

Par son expérience de metteuse en scène, Bérangère Vantusso constate l'isolement des artistes qui racontent peu leur manière de faire face à la création. Elle croit au bienfait des échanges entre pairs et propose la conception d'une plateforme numérique qui serait un endroit de ressources et de partage des questionnements liés à la fabrication théâtrale. Elle porterait le récit des processus qui mènent à la création et s'adresserait à tous.

Le temps de la construction

En tant que responsable du pôle recherche du studio-théâtre, je coordonne la conception et le développement de cet outil. Je pars donc en quête d'un développeur-designer. Je découvre le travail de Louis Éveillard de l'Atelier des chercheurs et son application *do.doc*¹ destinée aux élèves des écoles primaires, conçue pour documenter et créer des récits à partir d'activités pratiques.

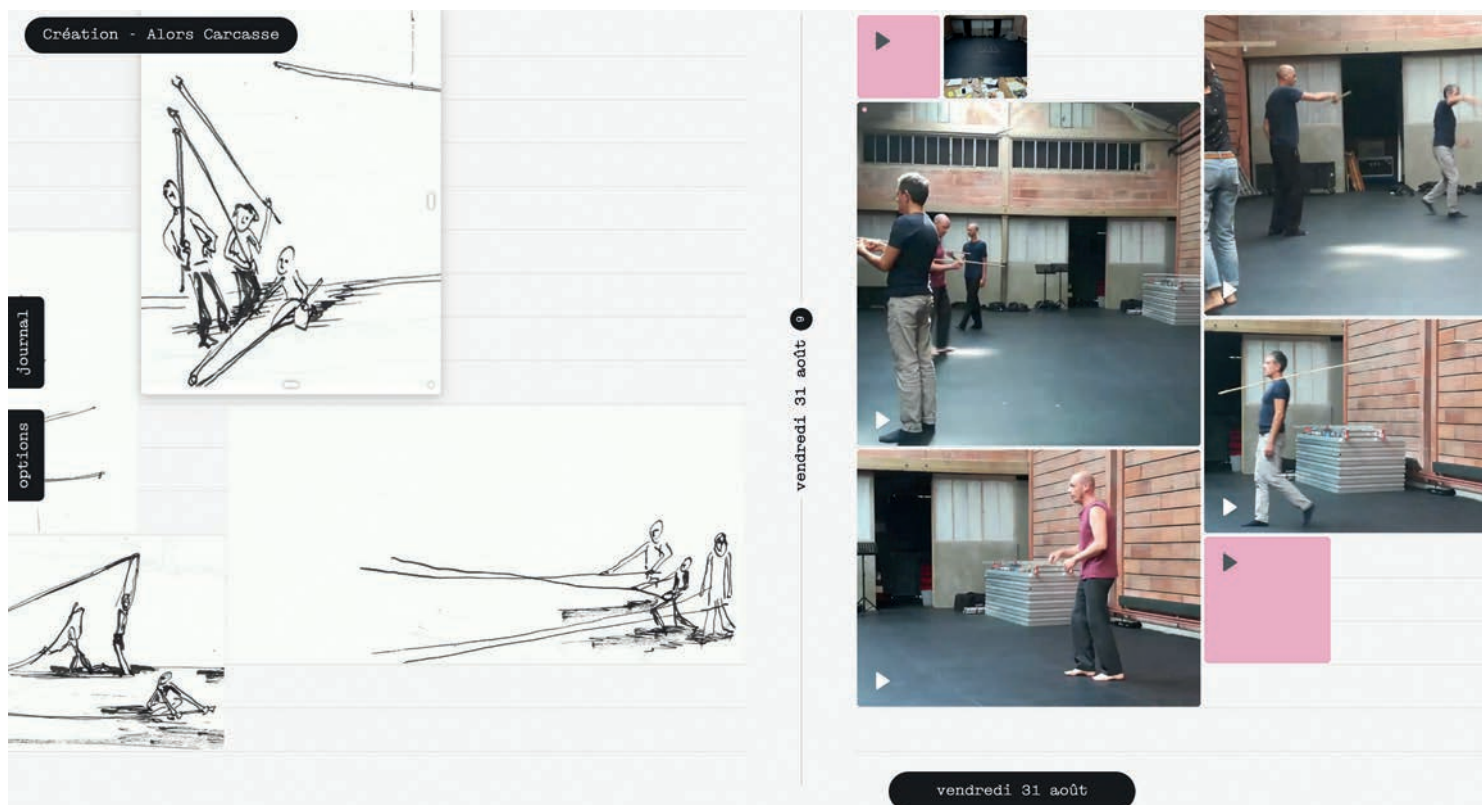
Louis Éveillard a le désir de développer des applications accessibles au plus grand nombre, toujours en logiciel libre et gratuit (*Open Source*), dans un souci de recherche constante et de partage des connaissances. Son approche, qu'il qualifie de « située », suppose des temps d'observation, de nombreux échanges, ainsi que l'expérimentation de prototypes fonctionnels.

FLORENCE KREMPER

Responsable du pôle recherche et de la communication du studio-théâtre de Vitry, conseillère artistique

1. <https://dodoc.fr/>

Capture d'écran d'un *Cahier du Studio* qui raconte le processus de création de *Alors Carcasse* de Mariette Navarro, mis en scène par Bérangère Vantusso.



Ces derniers évoluent de manière itérative et donnent lieu à des versions successives qui se précisent au fur et à mesure des temps d'observation, des tests et des retours collectés. Diffusés sous licence libre et documentés, ils sont ensuite facilement réutilisables et appropriables par d'autres personnes, pour d'autres usages et d'autres situations.

Ensemble, nous décidons de concevoir une application numérique *Open Source*, à caractère expérimental, qui se co-construit avec les équipes artistiques et évolue en fonction des usages.

Nos influences

Nous nous inspirons d'expériences faites par d'autres, je pense par exemple à *Rekall* conçu par Clarisse Bardiot². Nous nous nourrissons des échanges précieux avec des chercheuses en études théâtrales telles que Julie Valéro et Gay McAuley (fondatrice de la discipline *The Performance Studies*, un programme de recherches sur les processus de création des arts de la scène qui utilise des méthodes issues de l'ethnographie pour observer les temps de répétitions) et des recherches universitaires menées dans le champ de la

génétique théâtrale. Toutes ces ressources dessinent *Les Cahiers du Studio*.

La documentation

S'intéresser au processus de création, c'est entrer dans le monde de la fabrication, un temps intime, protégé, parfois âpre, de labeur, avec ses essais, ses tentatives, ses erreurs.

Pour ces raisons, nous proposons que le processus soit mis en récit par ceux qui font, à partir d'une documentation faite de l'intérieur, matériau de travail qui circule au sein de l'équipe artistique, une documentation réalisée par tous les protagonistes de l'aventure : metteurs en scène, interprètes, éclairagistes, techniciens, costumiers, etc.

Pour donner envie aux artistes de rendre visible leur travail sans paraître intrusif, Louis Éveillard conçoit un outil de travail qui sert aussi bien le processus que la collecte des documents, un espace collaboratif de partage des ressources au service de l'équipe : brouillons de création, cahiers de mise en scène, cahiers des comédiens, plans et maquettes scénographiques, mais aussi documentation directe

2. <https://dorradh.hypotheses.org/category/logiciel-rekall>





© Ivan Boccara

L'intérieur du studio-théâtre.

des temps à la table, des répétitions grâce à la capture en temps réel *via* divers outils : caméra HD, appareil photo, zoom enregistreur, téléphone portable pour une capture de médias directement reliée à l'application.

Vigilants à l'égard de la dimension collective de l'art théâtral, nous développons en écho une application collaborative : plusieurs auteurs peuvent éditer du contenu en même temps, voir ce que chacun écrit et collaborer directement.

La chronologie à l'œuvre dans les processus de création et de recherche apparaît essentielle pour comprendre les chemins pris par les différents acteurs de la création. Il est donc décidé que *Les Cahiers du Studio* prennent la forme d'une ligne de temps sur laquelle se déposent des traces multimédias.

Très vite nous dissociions le moment de la collecte de celui de la mise en récit de cette documentation. Le collectage de traces n'est possible et visible que par l'équipe artistique, une collecte rapide, faite sur le moment mais qui peut aussi se faire après-coup. Le dispositif accepte l'ajout de médias, qui peuvent être modifiés ou supprimés.

Le récit

Le récit par l'équipe artistique se fait à partir de la documentation, une fois la création terminée. Construire ce récit, c'est raconter de manière subjective et sensible la fabrication d'une création. L'équipe artistique, en situation d'observatrice de sa propre pratique, dans une posture réflexive, se pose alors la question de la transmission d'un savoir-faire. C'est le temps du partage. L'étape de la formulation est délicate pour les artistes qui délivrent une part intime de leur façon de créer, un contenu fragile et partiel. Ce récit est publié sur le site du studio-théâtre³.

Pour donner envie aux artistes de rendre visible leur travail sans paraître intrusif, Louis Éveillard conçoit un outil de travail qui sert aussi bien le processus que la collecte des documents, un espace collaboratif de partage des ressources au service de l'équipe...

Les Cahiers du Studio sont en *Open Source* pour permettre à tous et toutes, équipes artistiques, lieux d'expérimentation et de création, chercheurs... de l'utiliser à leur tour. Nous pouvons alors échanger sur ses modalités, mutualiser son développement et poursuivre l'évolution de ses potentialités.

Les Cahiers du Studio font partie intégrante du projet du studio-théâtre. Les récits des artistes invités en résidences forment peu à peu une mémoire vive et inédite des différentes façons de faire l'art théâtral.

Nous aimerions aussi qu'ils soulèvent l'intérêt du public pour le spectacle vivant par une meilleure connaissance des mécaniques de création, et qu'ils puissent être une ressource pour tout chercheur intéressé par l'analyse des processus de création ou de recherche théâtrale. ■

3. <https://www.studiotheatre.fr/les-cahiers>

Ossia Score,

un logiciel de création interdisciplinaire au cœur de la plateforme logicielle libre du Studio de création et de recherche en informatique et musiques expérimentales (SCRIME)

Le SCRIME est une plateforme de recherche et de création expérimentale de l'Université de Bordeaux qui est destinée à fournir des équipements et des outils logiciels pour répondre aux problématiques scientifiques et artistiques dans le domaine du son et de la musique. La plateforme logicielle est composée exclusivement d'outils libres. Le cœur de la plateforme est le logiciel Ossia score issu du projet i-score et dont le développement est soutenu par l'association ossia.io. Le projet i-score est issu d'une initiative collaborative de recherche et de développement visant à proposer et à faire évoluer des modèles et outils pour l'écriture intermedia interactive.

PIA BALTAZAR

Développeuse multimédias et design interactif, collectif ossia.io, SCRIME

JEAN-MICHAËL CELERIER

Ingénieur-chercheur, collectif ossia.io, SCRIME

MYRIAM DESAINTE-CATHERINE

Directrice du SCRIME, professeure des universités, Laboratoire bordelais de recherche en informatique (LaBRI UMR 5800), Université de Bordeaux

Ce projet s'articule autour du séquenceur Ossia score, et est le résultat d'une collaboration soutenue sur une quinzaine d'années par une communauté d'artistes, de scientifiques et d'ingénieurs. Une présentation plus exhaustive du projet ainsi que le téléchargement de ses résultats, sous licence *Open Source*, sont disponibles sur le site Web ossia.io¹. Ce projet est né, d'une part, d'un modèle théorique informatique permettant de spécifier des relations temporelles entre des événements interactifs, ainsi que d'une proposition d'implémentation logicielle de ce modèle, conçue avec la société Blue Yeti et, d'autre part, de la volonté de praticiennes et de praticiens du spectacle vivant, de la composition musicale et des arts numériques de mutualiser leurs efforts de développement logiciel pour disposer d'outils de création partagés. Ce projet a été largement soutenu par l'Agence nationale de la recherche (ANR), le SCRIME et le Groupe de musique électro-acoustique (GMEA)².

La quasi-totalité des participantes et des participants à l'initiative d'ossia.io, depuis son origine, ont des profils hybrides se situant pour la plupart à différentes intersections ou combinaisons de pratiques artistiques et d'ingénierie, et de démarches scientifiques. Cependant, chacune et chacun d'entre elles et eux s'est engagé dans l'initiative à partir d'une position donnée, qui impliquait de définir les priorités au service de cette position. Pour les scientifiques, l'intérêt principal

était de produire des modèles génériques permettant de produire ou d'analyser des scénarios interactifs ou d'en étudier les usages. Pour les actrices et les acteurs artistiques, la priorité était de pouvoir favoriser une démarche intuitive, voire sensorielle, pour intégrer l'usage du logiciel dans une pratique créative. Enfin, du point de vue de l'ingénierie, il s'agissait en quelque sorte de combiner ces deux aspects, en s'appuyant sur les modèles théoriques, pour en rendre l'usage ergonomique et « intuitif ». Cette diversité d'intérêts se reflète également dans une diversité de temporalités et de modalités de valorisation du travail : les praticiennes et les praticiens ont en effet l'habitude de développer des dispositifs « dans l'urgence », au fil des productions. La valorisation de leur travail se fait sur le résultat artistique, les détails de la réalisation important peu. De leur côté, les chercheuses et les chercheurs se situent davantage sur le temps long de la production du savoir transférable. L'évaluation de leur travail se fait sur la publication écrite de ces savoirs.

Par ailleurs, la création d'un logiciel libre pose la question du modèle économique : en l'état actuel, il consiste en la réalisation de formations au logiciel en école et de soutien aux artistes dans la démarche de production. Le développement, effectué *via* la plateforme de développement collaborative Github, permet à une équipe de contributeurs principaux de gérer et d'intégrer les évolutions proposées par la

1. <https://ossia.io/>

2. Implanté à Albi depuis 1977, le Groupe de musique électro-acoustique (GMEA) fait partie des six Centres nationaux de création musicale (CMCN) labellisés par le ministère de la Culture.

ossia score, an interactive sequencer for the intermedia arts

ossia.io

art open-source qt osc creative-coding sequencer interactive midi digital-art osc-messages open-sound-control hacktoberfest intermedia media-art

View license

1.2k stars

56 watching

95 forks

10 years old

v2.1.8 (Latest) 2 weeks ago

Sponsor this project

celier Jean-Michel Celier

opencollective.com/ossia

Contributors 27

Environments 1

github-pages (Active)

Releases

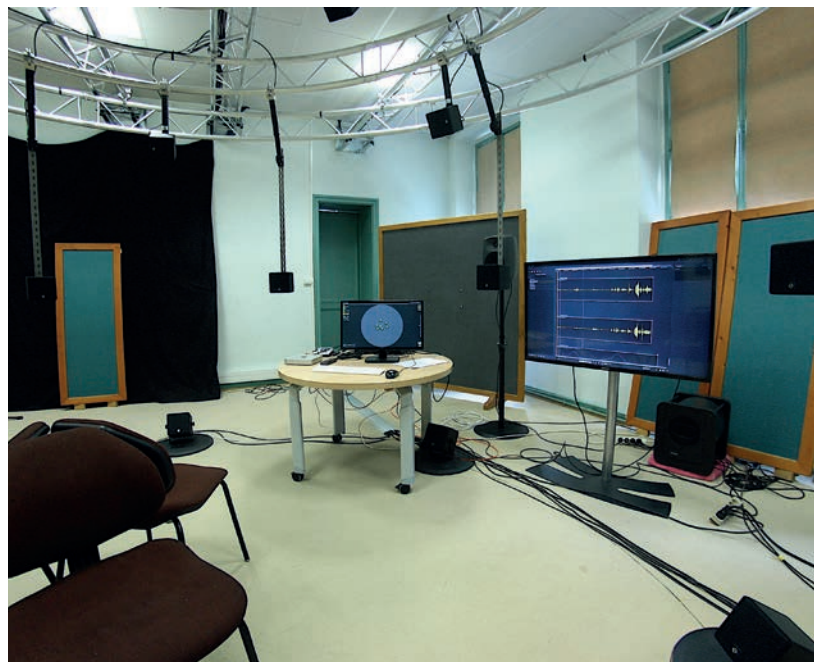
Releases are available for Windows, Linux (via AppImage) and Mac OS X.

Présentation d'Ossia score (<https://ossia.io/score/about.html>).

communauté, ainsi que de suivre les divers rapports de bugs et d'assigner leurs priorités en fonction des objectifs de développements de l'équipe ossia, qui cherche un équilibre entre les évolutions issues de la recherche scientifique et les besoins de production artistiques.

Enfin, en quelques chiffres : la base logicielle (score et libossia) comporte une quarantaine de contributeurs directs de code, notamment avec des intégrations dans les environnements créatifs numériques les plus populaires.

Sur un an, entre mars 2022 et mars 2023, environ un millier de contributions ont été réalisées par une dizaine de personnes distinctes. Et le logiciel Ossia score a été téléchargé 88 000 fois au 16 mars 2023... ■



Espace de travail du Studio de création et de recherche en informatique et musiques expérimentales (SCRIME).

Le laboratoire Perception, représentations, image, son, musique (PRISM) ou la vie des archives entre recherche et création

À l'heure de la généralisation des réseaux sociaux ou des humanités numériques, les unités de recherche deviennent des acteurs importants de la conservation et de la valorisation des données artistiques et scientifiques. L'enjeu est de concevoir des projets de recherche interdisciplinaires menés en partenariat avec des acteurs de la société (artistes, citoyens...) comme le montrent les différentes expérimentations conduites au sein du laboratoire PRISM¹.

RÉMI ADJIMAN

Professeur des universités en sciences de l'information et de la communication, Aix-Marseille Université

CAROLINE BOË

Doctorante en théorie et pratique de la création artistique et littéraire, compositrice, Aix-Marseille Université

PASCAL CESARO

Maître de conférences en études cinématographiques, Aix-Marseille Université

CHRISTINE ESCLAPEZ

Professeure des universités en musique et musicologie, Aix-Marseille Université

JULIEN FERRANDO

Maître de conférences en musique et musicologie, Aix-Marseille Université

MURIEL PIQUÉ

Doctorante en théorie et pratique de la création artistique et littéraire, chorégraphe, Aix-Marseille Université

JOSÉPHINE SIMONNOT

Ingénieure de recherche, responsable de la valorisation des données audiovisuelles, Aix-Marseille Université, UMR CNRS 7061-PRISM

Des bases de données ouvertes

« Sons de France » (Rémi Adjiman)²

Imaginé par des chercheurs au sein du département Sciences, arts et techniques de l'image et du son (SATIS) de l'Université d'Aix-Marseille, cette sonothèque en ligne permet au public d'accéder à des sons d'ambiance ainsi qu'à des réponses impulsives en haute qualité. Elle s'adresse non seulement à tous les professionnels, créateurs de contenu audiovisuel ou radiophonique ou de jeux vidéo, mais aussi aux amateurs de milieux sonores (*soundscape*). Son interface permet d'effectuer des recherches selon différents critères, suggérés parmi un thésaurus de plus de 2 500 termes. Son design original, conçu par des chercheurs et des ingénieurs du son, offre une description adaptée aux besoins des différents publics³.

« Anthropophony » (Caroline Boë)⁴

Anthropophony est une sonothèque collaborative conçue comme un petit musée des sons du quotidien qui nous envahissent, des sons *inframances* que nos oreilles filtrent⁵. L'écologie sonore⁶, le mouvement

environnemental, l'anthropocène ou l'écologie y sont questionnés grâce aux commentaires des utilisateurs. Les sons sont accessibles soit dans une liste avec des critères de recherche, soit à partir d'une cartographie. Chacune et chacun peut contribuer à ces archives sonores mais aussi les télécharger pour les utiliser (licence Creative Commons BY).

« Compose & Danse » (Muriel Piqué)⁷

Compose & Danse est une application Web (lauréate de l'appel à projets « Services numériques innovants » du ministère de la Culture), librement accessible, participative et collaborative, dédiée à l'art chorégraphique. Celle-ci expérimente une médiation numérique entre habitants et artistes et privilégie l'accès à la fabrique du mouvement dansé. Cet outil propose à tout internaute de faire connaissance avec les processus de création d'artistes contemporains. Des artistes, pédagogues et thérapeutes ont ainsi imaginé, en reliance avec l'histoire de la danse, des scénarios à danser à la portée de tous⁸.

Des fonds d'archives valorisés

Deux projets sont en cours d'exploitation au sein de PRISM (Julien Ferrando) :

- le fonds Misraki qui comporte de multiples documents (films, partitions) conservés sur des supports anciens inédits (séances d'enregistrements par exemple) ;
- le fonds Marcel Pagnol⁹. Le projet Les sons de Marcel Pagnol (soutenu par la Société d'accélération du transfert de technologies [SATT] Sud-Est) préfigure une collaboration avec le futur

1. <https://www.prism.cnrs.fr>

2. <https://www.sonsdefrance.fr/#/home>.

3. Rémi Adjiman, « Sémiotique des sons et cognition située », dans François Bobrie, Jean-François Bordron et Gérard Chandès (dir.), *Les sens du son*, Solilang, 2015. hal-01626799

4. [Anthropophony.org](http://anthropophony.org)

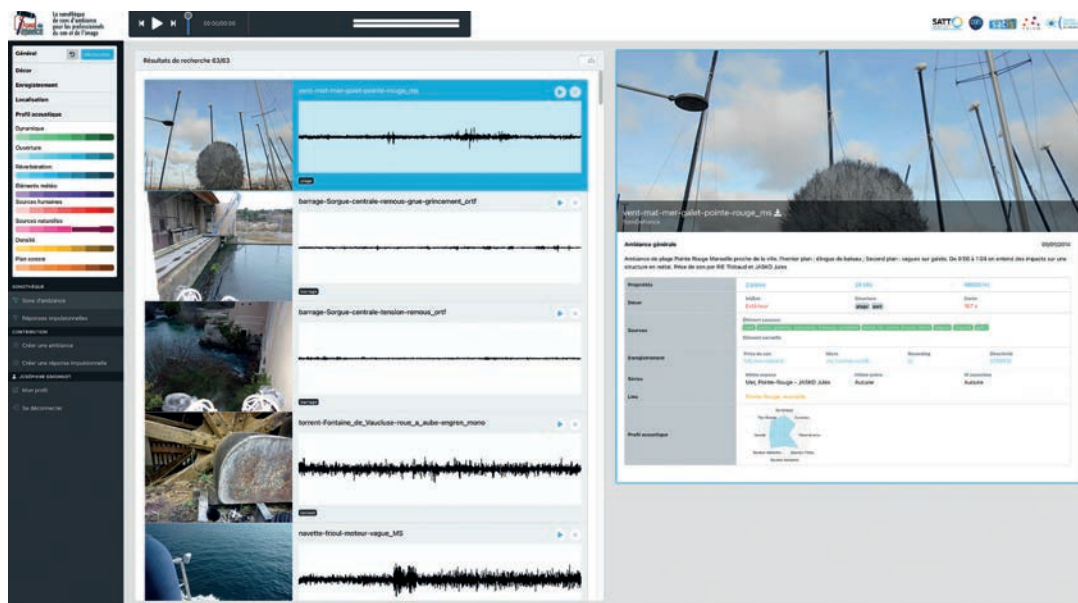
5. Caroline Boë, « Ces sons qui nous envahissent : Field Recordings, Soundwalks et Soundscapes éco-artistes », *Filigrane*, n° 26 [en ligne], 2022.

6. Raymond Murray Schafer, *Le paysage sonore. Le monde comme musique*, Wild project 2010.

7. <https://compose-danse.art/>

8. Muriel Piqué, « Compose & Danse », *Culture et Recherche*, n° 140, 2019-2020, p. 36-37.

9. Julien Ferrando, « Le son et la musique chez Marcel Pagnol : une dialectique singulière dans l'écriture cinématographique », dans Philippe Gonin et Jérôme Rossi (dir.), *Le cinéma populaire français et ses musiciens*, Éditions universitaires de Dijon, 2020, p. 243-266.



Capture d'écran de la plateforme Web audio Sons de France (<https://sonsdefrance.com/#/home>).

musée Marcel Pagnol qui sera implanté dans la commune d'Allauch en 2025-2026. Ces fonds constituent des ressources précieuses pour le musicologue, notamment pour l'analyse des processus de création image/son. Dans le cas du fonds Pagnol déposé au Centre national du cinéma et de l'image animée (CNC, Bois-d'Arcy) a été retrouvé un documentaire sonorisé inédit sur la ville de Marseille, *Les chansons de Marseille* (réalisé par Toé en 1950). Ce film a été numérisé en janvier 2023 et sera accessible sur une plateforme spécifique, en cours de conception.

Des réusages des archives audiovisuelles

« Di-Archiv » (Pascal Cesaro)¹⁰

Dialoguer avec les archives télévisuelles est un programme de recherche-création qui a été commandité, en 2021, par le Musée des civilisations de l'Europe et de la Méditerranée (MUCEM) dans le cadre de l'exposition « VIH/sida, l'épidémie n'est pas finie ! » L'objectif était de documenter le travail de remémoration de cette exposition qui retrace l'histoire sociale des luttes contre le sida. Le travail a été développé par le laboratoire PRISM en partenariat avec le centre Norbert Elias et a donné lieu à la création d'une œuvre audiovisuelle collective, conçue à partir d'un dispositif audiovisuel utilisant le réemploi d'archives sur le sida. Les chercheurs ont fait l'hypothèse qu'éprouver l'archive permet à la personne qui témoigne de retrouver un contexte particulier, une mémoire difficile et favorise l'émergence d'une parole non conventionnelle qui permet de dépasser les discours convenus et militants¹¹.

« [CO S-A] – Corsica Sud-Archives » (Christine Esclapez et Joséphine Simonnot)¹²

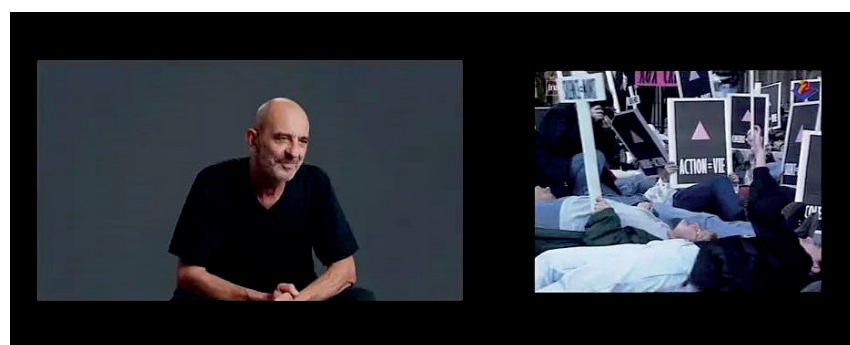
Le projet [CO S-A] est porté par PRISM en collaboration avec le Centre national de création musicale (CNCM) VOCE (Pigna, Corse) et la Maison méditerranéenne des sciences de l'homme d'Aix-en-Provence (MMSH-Aix). Son objectif est de réfléchir de façon collaborative à la *vie* de l'archive après sa mise en archive.

Le corpus de référence est le Repertorium VOCE. Numérisé à partir de 2010 grâce à l'aide du ministère de la Culture, de la MMSH-Aix et de la collectivité territoriale de Corse, ce fonds d'archives résulte de plusieurs collectes initiés dès les années 1970. Le projet [CO S-A] repose sur trois axes de recherche : recréation à partir d'archives par les artistes ; actualisation et mise en conformité selon les principes FAIR¹³ ; valorisation des archives pour le grand public (ballade sonore composée à partir de certaines archives du Repertorium mixées avec des captations de sons de gestes des artisans de Pigna au travail)¹⁴.

Conclusion

Préludes à des recherches scientifiques en émergence, ces projets montrent combien la science ouverte offre un cadre pertinent de développement de recherches fondamentales, mais aussi appliquées dans le domaine des arts, sciences et sociétés¹⁵. La participation du laboratoire PRISM aux consortiums Musica 2 et Sound of Life de l'Institut des sciences humaines et sociales du CNRS (InSHS) consolide cette démarche d'ouverture. Si la pérennité des données est un des enjeux de l'équipe, leurs réutilisations interrogent ce qui fonde une part de la démarche scientifique d'un laboratoire interdisciplinaire comme PRISM : comment continuer à explorer et à préserver l'histoire de nos mémoires artistiques ? ■

10. <https://www.prism.cnrs.fr/projet/dialoguer-avec-les-archives-la-television-face-au-choc-du-sida/>
11. Pascal Cesaro, Sandrine Musso, Mario Fanfani et Emmanuel Vigier, *Dialoguer avec les archives télévisuelles*, co-édition Anamosa-Mucem, 2021 ; Pascal Cesaro, Sandrine Musso, Mario Fanfani et Emmanuel Vigier, *Ce n'est pas à proprement parler une épidémie*, film d'installation, production : MUCEM-Promenades film-PRISM, 2021.
12. <https://cosa.hypotheses.org>
13. « Facile à trouver, accessible, interopérable et réutilisable ».
14. Caroline Boë et Christine Esclapez, « Écoute(s) active(s) en "territoire-laboratoire" infra-insulaire. Pigna, une maison de l'écoute », dans Brigitte Marin, Anthony Santilli et Arturo Gallia (dir.), *Dos ou face au monde. Habiter les îles*, Imbernon, à paraître.
15. Joséphine Simonnot, « Musiques enregistrées : nouveaux défis, nouvelles pratiques », dans Achille Davy-Rigaux, Catherine Deutsch, Hamdi Makhoul et Anas Ghrab (dir.), *La musicologie francophone et la circulation des savoirs dans un contexte multiculturel : interpénétration et résilience*, Sotumédias Éditions, 2022, p. 129-149.



Photogramme extrait du film *Ce n'est pas à proprement parler une épidémie*, 2 h 11 min.

CULTURE ET RECHERCHE

Tous les numéros de *Culture & Recherche* sont disponibles au format pdf sur le site internet du ministère de la Culture.
www.culture.gouv.fr/Sites-thematiques/Enseignement-superieur-et-Recherche/La-revue-Culture-et-Recherche

Pour s'inscrire sur la liste de diffusion, ou pour tout renseignement :
culture-et-recherche@culture.gouv.fr



Le dernier numéro paru en 2022

n° 143 – Automne-hiver 2022, **La recherche culturelle à l'international**

Ce numéro interroge la place de la recherche culturelle soutenue par le ministère de la Culture dans l'internationalisation des pratiques et des productions des communautés scientifiques. Il présente la recherche culturelle au sein de la mondialisation, aborde l'extension fulgurante des relations entre les sociétés et leurs valeurs. Cette production scientifique sans frontières repose sur une nouvelle stratégie de coopération internationale dans le cadre de partenariats plus équilibrés, qui facilitent la circulation des connaissances, l'amplification des talents et suscitent de nouvelles opportunités. Le renforcement mutuel des capacités à produire de la recherche et de l'innovation est aussi un marqueur de la diplomatie scientifique, en adoptant un autre regard sur l'altérité et l'externalité.

Trois grandes thématiques sont abordées : le partage et la valorisation de l'expertise culturelle française qui s'exerce dans un univers désormais fortement déterritorialisé ; le développement et l'expérimentation de l'interculturalité qui reposent, dans ce voisinage mondial, sur une diversité d'acteurs qui « font corps » ; les transitions innovantes d'un continent numérique, infrastructure propre de recherche tentant de répondre aux grands défis sociétaux et portée par une interopérabilité des outils.

Numéros récents



n° 142 – Printemps-été 2022



n° 141 – Printemps-été 2020



n° 140 – Hiver 2019-2020



n° 139 – Printemps-été 2019



n° 138 – Automne-hiver 2018



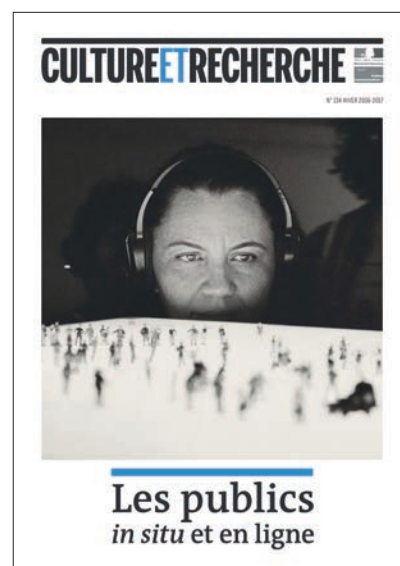
n° 137 – Printemps-été 2018



n° 136 – Automne-hiver 2017



n° 135 – Printemps-été 2017



n° 134 – Hiver 2016-2017



n° 133 – Été 2016



n° 132 – Automne-hiver 2015-2016



n° 131 – Printemps-été 2015

CULTURE ET RECHERCHE

COMITÉ ÉDITORIAL ET PROGRAMMATION

Le comité éditorial est piloté par la Sous-direction Formation-Recherche et son bureau de la recherche au sein de la Délégation générale à la transmission, aux territoires et à la démocratie culturelle. Ses membres représentent :

- La délégation générale à la langue française et aux langues de France.
- Le département des études, de la prospective, des statistiques et de la documentation, Secrétariat général.
- La direction générale de la création artistique.
- La direction générale des médias et des industries culturelles.
- La direction générale des patrimoines et de l'architecture.

Directeur de la publication : **Emmanuel MARCOVITCH** / Directeur de cabinet de la ministre de la Culture

Rédactrice en chef : **Catherine GRAINDORGE** / Délégation générale à la transmission, aux territoires et à la démocratie culturelle / Sous-direction des formations et de la recherche

COMITÉ ÉDITORIAL

Solène BELLANGER

Cheffe de la Mission Recherche / Direction générale de la création artistique / Sous-direction des enseignements spécialisés et supérieur et de la recherche

Laurence BIZIEN

Invitée expertise Science ouverte, Chargée d'études documentaires, École nationale supérieure d'architecture de Nantes / Centre de recherche nantais Architectures Urbanités (CRENAU)

Jean-Christophe BONNISSENT

Chargé de mission / Délégation générale à la langue française et aux langues de France / Mission Emploi et diffusion de la langue française

Bastien CHASTAGNER

Chef du bureau Accès aux archives et de l'animation du réseau / Service interministériel des Archives de France / Sous-direction du pilotage, de la communication et de la valorisation des archives

Claire CHASTANIER

Adjointe à la Sous-directrice des collections / Direction générale des patrimoines et de l'architecture / Service des Musées de France / Sous-direction des collections

Christian CRIBELLIER

Adjoint au Sous-directeur / Direction générale des patrimoines et de l'architecture / Service du patrimoine / Sous-direction de l'archéologie

Aude CROZET

Chargée de mission / Direction générale des patrimoines et de l'architecture / Service du patrimoine / Sous-direction de l'archéologie / Bureau du patrimoine archéologique

Brigitte GUIGUENO

Adjointe au Sous-directeur / Direction générale des patrimoines et de l'architecture / Service interministériel des archives de France / Sous-direction du pilotage, de la communication et de la valorisation des archives

Priscilla GUSTAVE-PERRON

Cheffe du bureau Recherche / Délégation générale à la transmission, aux territoires et à la démocratie culturelle / Sous-direction des formations et de la recherche

Laurence ISNARD

Cheffe du bureau des acquisitions, de la restauration, de la conservation préventive et de la recherche / Direction générale des patrimoines et de l'architecture / Service des Musées de France / Sous-direction des collections (jusqu'en mars 2023)

Judith KAGAN

Cheffe du bureau de l'expertise et des métiers / Direction générale des patrimoines et de l'architecture / Service du patrimoine / Sous-direction des Monuments historiques et des sites patrimoniaux

Isabelle-Cécile LE MÉE

Chargée de mission / Direction générale des patrimoines et de l'architecture / Délégation à l'inspection, à la recherche et à l'innovation

Pascal LIÉVAUX

Adjoint au Chef de la Délégation / Direction générale des patrimoines et de l'architecture / Délégation à l'inspection, à la recherche et à l'innovation

Wilfried MULLER

Chargé de mission / Direction générale des médias et des industries culturelles / Service du Livre et de la Lecture / Département des bibliothèques / Bureau du patrimoine

Carine PRUNET

Adjointe au chef du bureau de la diffusion numérique des collections / Direction générale des patrimoines et de l'architecture / Service des Musées de France / Sous-direction des collections

Pierre-Jean RIAMOND

Chef du bureau du patrimoine / Direction générale des médias et des industries culturelles / Service du Livre et de la Lecture / Département des bibliothèques

Mickaël ROBERT-GONÇALVES

Chargé de mission / Direction générale de la création artistique / Sous-direction des enseignements spécialisés et supérieur et de la recherche / Mission recherche (jusqu'en mars 2023)

Éric ROUARD

Chef de la Mission de la politique documentaire / Secrétariat Général / Département des études, de la prospective, des statistiques et de la documentation / Mission de la politique documentaire

Miguel SAYOUS

Chargé de mission / Secrétariat Général / Département des études, de la prospective, des statistiques et de la documentation / Mission de la politique documentaire

Valérie WATHIER

Adjointe à la Cheffe de bureau / Direction générale des patrimoines et de l'architecture / Service de l'architecture / Sous-direction de l'enseignement supérieur et de la recherche en architecture / Bureau des enseignements

CULTURE ET RECHERCHE



Créée en 1985, la revue *Culture & Recherche* présente et explicite la recherche culturelle menée au sein du ministère de la Culture dans toutes ses composantes : patrimoines, création, médias, industries culturelles, développements technologiques appliqués au secteur culturel.

Chaque numéro est consacré à un axe prioritaire de l'action du ministère dont les travaux sont menés par les acteurs et partenaires de la recherche culturelle financée et soutenue par le ministère de la Culture. La revue a pour objectif de diffuser et rendre visible la recherche culturelle auprès d'un public élargi de professionnels du secteur culturel et de la recherche. Elle offre un regard sur l'actualité de la recherche culturelle développée au sein des différentes structures du ministère (établissements, structures de recherche, laboratoires, réseau des écoles de l'ESC) et qui s'inscrivent, notamment, dans l'accord-cadre avec le CNRS ou les appels à projets.

Ce numéro 144 de la revue *Culture & Recherche* est dédié à la science ouverte. Il présente les productions de la recherche culturelle des acteurs et partenaires institutionnels du ministère de la Culture dans un contexte national, européen et international.

Parmi les modes de « faire de la recherche » aujourd’hui, « les pratiques scientifiques plus ouvertes, transparentes, collaboratives et inclusives, associées à des connaissances scientifiques plus accessibles et vérifiables et soumises à l’examen et à la critique, améliorent l’efficacité, la qualité, la reproductibilité et l’impact de l’entreprise scientifique » (Recommandation de l’UNESCO sur une science ouverte, 2021).

Trois écosystèmes sont sollicités : les données ouvertes (*Open Data*), le libre accès (*Open Access*) et des outils-logiciels libres et gratuits (*Open Source*).

Chercheurs, ingénieurs, informaticiens, bibliothécaires et documentalistes, usagers et société civile se mobilisent pour expérimenter et mettre en œuvre ce régime de production des savoirs partagés, afin de répondre, notamment, à des enjeux à fort impact sociétal.

Les modèles existants sont bousculés : l’économie de l’édition, la protection du droit d’auteur, l’intégrité scientifique, l’évaluation des chercheurs, le type de technologies utilisées, la transformation des (très) grandes infrastructures et le financement de la recherche, la formation et la médiation scientifiques...

Ce changement de paradigme « faire science par, avec et pour la société » s’accompagne aujourd’hui d’une démarche qualitative en quête de plus grande accessibilité et de découvrabilité.

Ce récit est porté par trois axes : un environnement multidisciplinaire, fédéré et ouvert ; un bien commun : stocker pour capitaliser ; de nouveaux modes de transmission et de diffusion.


Directeur de la publication : **Emmanuel MARCOVITCH**
Directeur de cabinet de la ministre de la Culture

Rédactrice en chef : **Catherine GRAINDORGE**
Délégation générale à la transmission, aux territoires
et à la démocratie culturelle / Sous-direction
des formations et de la recherche

Impression : **DILA**/Département de l’imprimerie,
75015 Paris

ISSN papier : 0765-5991 – ISSN en ligne : 1950-6295

Réalisation : **Transfaire**
contact@transfaire.com

 **PEFC** 10-31-2190 / Certifié PEFC

 **IMPRIM'VERT**
