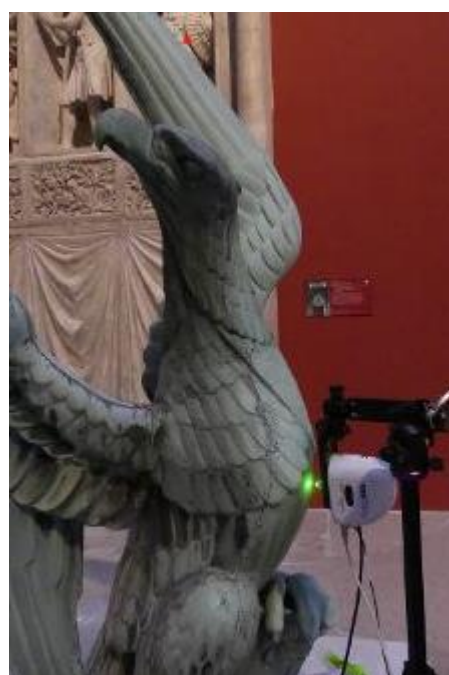


## SCIENCE OUVERTE DONNEES OUVERTES

### PATRIMOINE ET DONNEES

*Bien gérer ses données, dès le début des projets de recherche doit devenir la règle en sciences du patrimoine.*

Début 2019, le Laboratoire de Recherche des Monuments Historiques (LRMH), en collaboration avec le Centre de Recherche sur la Conservation des Collections, fait l'acquisition d'un spectromètre Raman mobile pour l'étude sur site des peintures murales et des œuvres en métal classées monument historique, et des collections naturalistes du Museum National d'Histoire Naturelle. L'objectif est de mieux caractériser les matériaux constitutifs de ces objets, matériaux de la couleur et liquides de conservation de spécimens, pour améliorer la compréhension de leur production et de leur vieillissement. Cet appareil va produire au fil des années de nombreux jeux de données qui, au-delà de l'observation initiale, sont susceptibles d'être réutilisés dans le temps pour évaluer l'impact des conditions de conservation ou encore pour les comparer avec des objets de facture similaire mais analysés à distance de plusieurs années selon la même méthode analytique. Au-delà de l'appareillage technique, c'est toute une problématique de gestion des données qui s'ouvre dans ce cadre.



De fait, la gestion et la mise à la disposition de tous, des données produites par la recherche scientifique sont devenus des sujets particulièrement importants ces dernières années dans le contexte du développement de la science ouverte.

L'Europe a ainsi mis en place l'obligation, pour tout projet de recherche financé par ses soins, de produire un *plan de gestion de données*, rejointe en cela par l'Agence Nationale de la Recherche en France dans le cadre des projets qu'elle finance.

Dans une démarche d'avant-garde, la France s'est par ailleurs dotée d'un ambitieux *plan national pour la science ouverte*<sup>1</sup> lancé en juillet 2018 par le ministère de l'enseignement supérieur, de la recherche et de l'innovation, relayé en juillet 2021 par une deuxième édition couvrant largement les sujets des publications, des données et des logiciels de recherche.

---

<sup>1</sup> <https://www.enseignementsup-recherche.gouv.fr/cid132529/le-plan-national-pour-la-science-ouverte-les-resultats-de-la-recherche-scientifique-ouverts-a-tous-sans-entrave-sans-delai-sans-paiement.html>

Il est cependant difficile d'expliciter ce qu'une gestion raisonnée des données de la recherche peut être lorsque l'on reste à un niveau générique, avec le risque d'être écrasé par des abstractions telles que les principes FAIR<sup>2</sup>, qui occultent la complexité de ce que peut être une donnée de recherche dans des champs disciplinaires particulier.

Pour appréhender le sujet de façon plus concrète, nous reprenons ici quelques éléments issus de l'expérience acquise dans le cadre du projet Dopamine<sup>3</sup>, soutenu par le DIM (Domaine d'intérêt majeur) « Matériaux Anciens et Patrimoniaux », qui est le réseau francilien de recherche consacré à l'étude des matériaux anciens (dont le coordinateur est Loïc Bertrand, président du comité Patrimoine d'IESF). À ce titre, les auteurs du présent article ont eu l'occasion d'interagir avec des projets de recherche et des institutions patrimoniales impliqués dans des études sur les matériaux anciens tels qu'on les rencontre dans les œuvres d'arts, les collections archéologiques ou paléontologiques. Pour la plupart, ces projets impliquent des objets issus de collections institutionnelles ou privées. Ils sont soumis à des analyses physico-chimiques permettant d'en analyser la composition ou la structure pour par exemple comprendre les techniques ou processus naturels qui les ont formés, ou encore en étudier le vieillissement. Le fruit de ces études se présente avant tout sous la forme des données numériques d'observation qui vont servir de base à toutes les analyses et les conclusions qui seront tirées de ces recherches.

Ainsi, l'ensemble des recherches concernées a basculé presque malgré elles dans la numérique, sans que les communautés n'aient été préparées à cela. En particulier, des questions aussi essentielles que le devenir de ces données, la possibilité de les réutiliser pour d'autres études, la capacité de comparer les expériences entre elles, restent souvent en suspens, et au-delà de la maîtrise des techniques numériques correspondantes, le chercheur doit apprendre à maîtriser les méthodes nécessaires à une gestion pérenne de ses propres données.

Car, en dernier lieu, ces données scientifiques sont potentiellement un réel trésor : bien les gérer permet de créer un véritable patrimoine numérique permettant de mieux comprendre, dans un ensemble cohérent, les différentes composantes de notre patrimoine humain ou environnemental. Mais ce trésor est fragile. Il est facile de perdre des données, ou simplement de ne juste pas les retrouver dans un océan numérique mal catalogué, de ne pas savoir en garantir leurs conditions de production ou l'authenticité.

Afin d'éviter l'obsolescence naturelle des jeux de données et donc toute possibilité d'une quelconque réutilisation, il est indispensable d'intégrer au processus de recherche une documentation systématique de toute information pertinente relative à celles-ci. La base est bien sûr de décrire le ou les formats<sup>3</sup> dans lesquels les données sont représentées, ainsi que la nature de l'observation qu'elles traduisent. Mais cette documentation doit aller beaucoup plus loin si on veut assurer que ces données s'intègrent correctement dans un processus scientifique reproductible et durable. C'est en effet toute la démarche en amont qu'il faut pouvoir tracer, un peu comme on trace l'origine d'un produit dans un supermarché.

---

<sup>2</sup> <https://www.nature.com/articles/sdata201618>

<sup>3</sup> <http://www.dim-map.fr/projets-soutenus/dopamine/>

<sup>4</sup> Les formats de données sont encore trop dépendants des équipements d'analyse utilisés et ne font encore que très peu l'objet de travaux internationaux de normalisation.

Il faut bien sûr identifier l'objet qui a été observé et potentiellement touché par la méthode physique ou chimique d'analyse, ainsi que bien sûr l'institution patrimoniale qui a fourni cet objet et en a autorisé l'étude. C'est un élément indispensable pour reconnaître à celle-ci la contribution initiale, mais aussi pour éviter de refaire une analyse existante, effectuer des comparaisons avec des observations passées ou futures ou encore comparer des observations sur des cohortes d'objets de même nature.

La documentation des caractéristiques propres à l'expérience est aussi importante pour bien être en mesure d'en interpréter les résultats, qu'il s'agisse de la nature de l'appareil ou du dispositif expérimental ou des paramètres utilisés pour réaliser l'expérience (par exemple les fréquences et la puissance d'un rayon de lumière). Si les données subissent des traitements numériques ultérieurs (filtrage, lissage, extraction de paramètres), il faut savoir décrire les différentes étapes correspondantes et idéalement pouvoir faire référence au code logiciel utilisé. On touche là un aspect plus général de toute réflexion liée à la gestion des données de recherche : leur production, et donc leur réutilisation, est étroitement liée à différents logiciels dont il faut documenter les caractéristiques.

Au-delà des aspects techniques il faut savoir reconnaître le travail effectué par tous les acteurs qui ont contribué à l'expérience du chercheur. Qu'ils s'agissent des acteurs patrimoniaux, des techniciens de la plateforme technique utilisée ou des collègues qui ont fourni un soutien conceptuel ou matériel à telle ou telle étape de la recherche. Là encore, la documentation doit être précise et recenser les noms, institutions et rôles de ces différents acteurs, de façon à pouvoir les citer et leur attribuer la reconnaissance nécessaire quand il est fait référence au jeu de données lui-même.

Ce travail peut paraître lourd et fastidieux, mais il fait intrinsèquement partie de toute démarche numérique. Bien plus, si on n'intègre pas la démarche de documentation au déroulement de la recherche elle-même, il est quasiment impossible de reprendre l'ensemble des éléments nécessaires, alors que le chercheur est déjà passé à l'expérience suivante. Pour ne prendre qu'un aspect parmi tant d'autres, la sélection des données à conserver dans le cadre d'un tri qui va exclure les tests de mise en route ou les observations ratées ou superfétatoires, ne peut être faite qu'en continu. Ne pas y réfléchir conduit à un engorgement numérique où les « bonnes » données sont encore moins identifiées.

Sous l'angle du numérique, on voit qu'on se rapproche des pratiques de disciplines expérimentales où la gestion d'un carnet de laboratoire accompagnant toute expérience est une obligation dans bien des laboratoires. On pourrait souhaiter étendre ce concept à la systématisation d'un carnet de laboratoire numérique qui permette de référencer toutes les étapes de production, de traitement et de sauvegarde des données.

Enfin, et paradoxalement, l'un des aspects les plus importants pour la réutilisation des données n'est pas technique, ni même véritablement légal, mais bien social : la complexité des acteurs (chercheurs, institutions patrimoniales, équipements) impliqués dans un travail de recherche en sciences du patrimoine fait que la diffusion d'un résultat quelconque ne peut s'imaginer sans un consensus sur le devenir des données. Pour des raisons scientifiques de primeur, ou encore afin de préserver la communication autour d'objets sensibles (œuvres d'art majeures, travaux de restauration d'un monument particulièrement exposé médiatiquement), on peut vouloir limiter ou retarder la diffusion des résultats. Au contraire, dans certains domaines comme l'étude du comportement de certains matériaux en situation de restauration où seule la mise en commun de grandes quantités de données va permettre de dégager des tendances, on peut vouloir rendre obligatoire la diffusion précoce des observations disponibles.

Alors que certaines institutions se positionnent clairement en faveur d'une ouverture massive, comme on l'a vu récemment avec la mise en ligne de la totalité de la production des rapports du LRMH via son portail SYNAPSE<sup>4</sup> (Système Numérique Appliqué au Patrimoine et à ses SciencEs) ou l'ouverture de millions d'images par le Smithsonian<sup>5</sup>, cet accord entre acteurs doit passer par des mécanismes de concertation telles que ceux mis en œuvre dans le cadre de l'initiative de charte de réutilisation des données<sup>6</sup>.

En adoptant une démarche vertueuse de gestion des données en sciences du patrimoine, qui intègre tout à la fois une maîtrise fine des aspects numériques et une intégration des obligations légales d'une démarche éthique, on se donne les moyens de créer un véritable patrimoine numérique qui accompagne le patrimoine tangible.

Stéphanie DUCHENE

Ingénieure d'études au Laboratoire de Recherche des Monuments Historiques

Laurent ROMARY

Directeur de recherche à l'INRIA, membre du Comité Patrimoine d'IESF

*Illustration : Aigle de Saint Jean l'Évangéliste, flèche de la cathédrale Notre-Dame de Paris – sculpture en fonte de fer et cuivre en cours d'analyse - crédits photographiques LRMH Stéphanie Duchêne*

---

<sup>4</sup> <https://www.lrmh.fr/centre-de-ressources-synapse.aspx>

<sup>5</sup> <https://www.si.edu/openaccess>

<sup>6</sup> <https://datacharter.hypotheses.org/77>