

CULTURE ET RECHERCHE

MINISTÈRE DE LA CULTURE
Liberté
Égalité
Fraternité

N° 148 PRINTEMPS-ÉTÉ 2025



Recherche Culture et politiques publiques

Circulation(s), usage(s) et ouverture(s) des données culturelles

À la croisée de l'action publique, du numérique et des droits culturels, la circulation des données culturelles progresse, tandis que le développement des réutilisations s'affirme comme un axe stratégique des années à venir pour les institutions culturelles et la société.

AURÉLIEN CONRAUX

Administrateur ministériel des données, algorithmes et codes sources délégué, Service du numérique, ministère de la Culture

EUDES PEYRE

Chargé de mission circulation et ouverture des données, Service du numérique, ministère de la Culture

L'historique : les données culturelles parmi les données publiques

Revenir sur l'histoire des données culturelles et leurs usages, c'est déjà poser la question d'une particularité de la Culture, qui se reflèterait dans les informations représentées « sous une forme conventionnelle destinée à faciliter son traitement¹ », notamment par des machines.

Tant pour le ministère que pour ses opérateurs, les données de gestion (données RH, financières, données statistiques et indicateurs) comme les données d'usages utilisées ne diffèrent bien sûr pas essentiellement des données équivalentes d'autres secteurs de l'État.

À ce titre, les grands principes structurants de la réglementation sur les données publiques s'appliquent évidemment, tels qu'ils se sont sédimentés et précisés depuis la Révolution : la transparence de l'administration inscrite dans la Déclaration des droits de l'homme et du citoyen² ; l'accès des citoyens aux archives et aux documents administratifs (dont les données font partie) formalisé par les lois de 1978 et 1979 qui établissent également les contreparties de protection de certaines informations comme les données à caractère personnel ; l'encouragement progressif à la réutilisation des données publiques, jusqu'au principe d'ouverture par défaut des données publiques et d'intérêt général, qui est le cadre par défaut depuis la Loi pour une République numérique de 2016.

Des points d'attention – on a envie de parler d'expertise – existent cependant : les données relatives aux offres culturelles, les reproductions numériques d'œuvres en bibliothèques, archives ou musées, comme les captations de spectacles ou la documentation 2D ou 3D des sites et monuments portent en elles toutes les dimensions culturelles. De cette richesse découlent concrètement plusieurs spécificités de collecte, traitement et diffusion des données : le partage des données avec les citoyens, les chercheurs et les industries culturelles participe

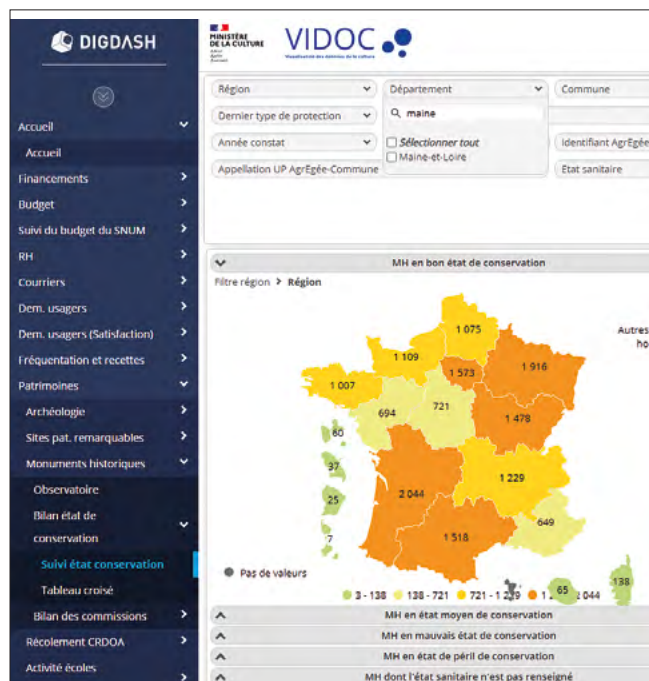
d'un projet de société ; la diffusion sur le très long terme de ressources numériques amène des besoins d'identifiants et d'accès persistants ; les modèles de données peuvent être très sophistiqués pour décrire des contenus culturels d'époques et de styles artistiques variés ; des référentiels et des notices d'autorité très structurés sont indispensables pour la recherche et pour distinguer auteurs, interprètes, traducteurs, éditeurs, titulaires de droits, etc.

L'ouverture des données culturelles s'insère donc dans une dynamique générale forte (la France est depuis 2021 au premier rang européen de l'ouverture des données selon l'*Open Data Maturity Report* de l'Union européenne³) mais a justifié une politique propre du ministère et de ses

1. Définition de la donnée, arrêté du 22 décembre 1981 sur l'enrichissement de la langue française. France Terme : <https://www.culture.fr/franceterme/terme/INFO118?domaine=0&q=donn%C3%A9e> (site consulté le 27 mars 2025).

2. Article 15 : « La société a le droit de demander compte à tout agent public de son administration ». <https://www.conseil-constitutionnel.fr/le-bloc-de-constitutionnalite/declaration-des-droits-de-l-homme-et-du-citoyen-de-1789> (page consultée le 17 mars 2025).

3. <https://data.europa.eu/en/publications/open-data-maturity/2024> (page consultée le 27 mars 2025).



établissements, souvent d'ailleurs progressive et pragmatique.

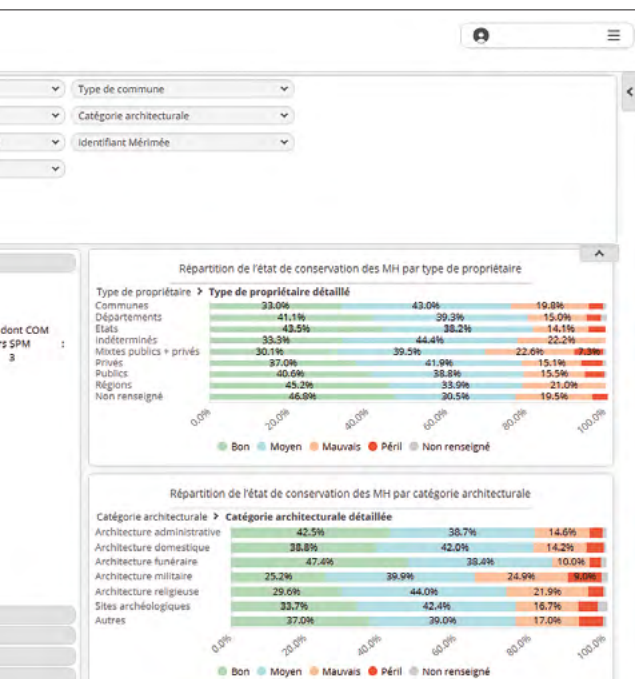
Les plans de numérisation, qui existaient de longue date dans les institutions culturelles pour faciliter la communication de documents fragiles, ont subi à partir de 2005 un accroissement majeur⁴. Cela a amené progressivement les établissements publics à ouvrir les métadonnées associées aux biens culturels numérisés, puis l'ensemble des données d'autorité et de description de leurs catalogues⁵, ouvrant des domaines entiers à la recherche. Au niveau ministériel, le développement des bases de données patrimoniales documentant les fonds (Joconde, FranceArchives, etc.) illustre également cet effort de représentation de la diversité culturelle sur le Web.

Depuis 2009, le ministère a produit des rapports sur l'ouverture et la réutilisation des données culturelles⁶, puis des guides pratiques⁷ et un *Vade-Mecum pour les producteurs de données*. Plus récemment, un accompagnement important s'est porté sur l'agrégation des données culturelles et leur découvrabilité dans les portails locaux, nationaux ou européens, à la fois sous forme de notices individuelles et dans un graphe de données liées permettant des interrogations sémantiques.

La France est depuis 2021 au premier rang européen de l'ouverture des données...



La maturité des pratiques d'ouverture des données dans l'Union européenne en 2024 : <https://www.data.gouv.fr/fr/posts/la-france-classee-premiere-europeenne-en-matiere-dopen-data-pour-la-4e-annee-consecutive/>



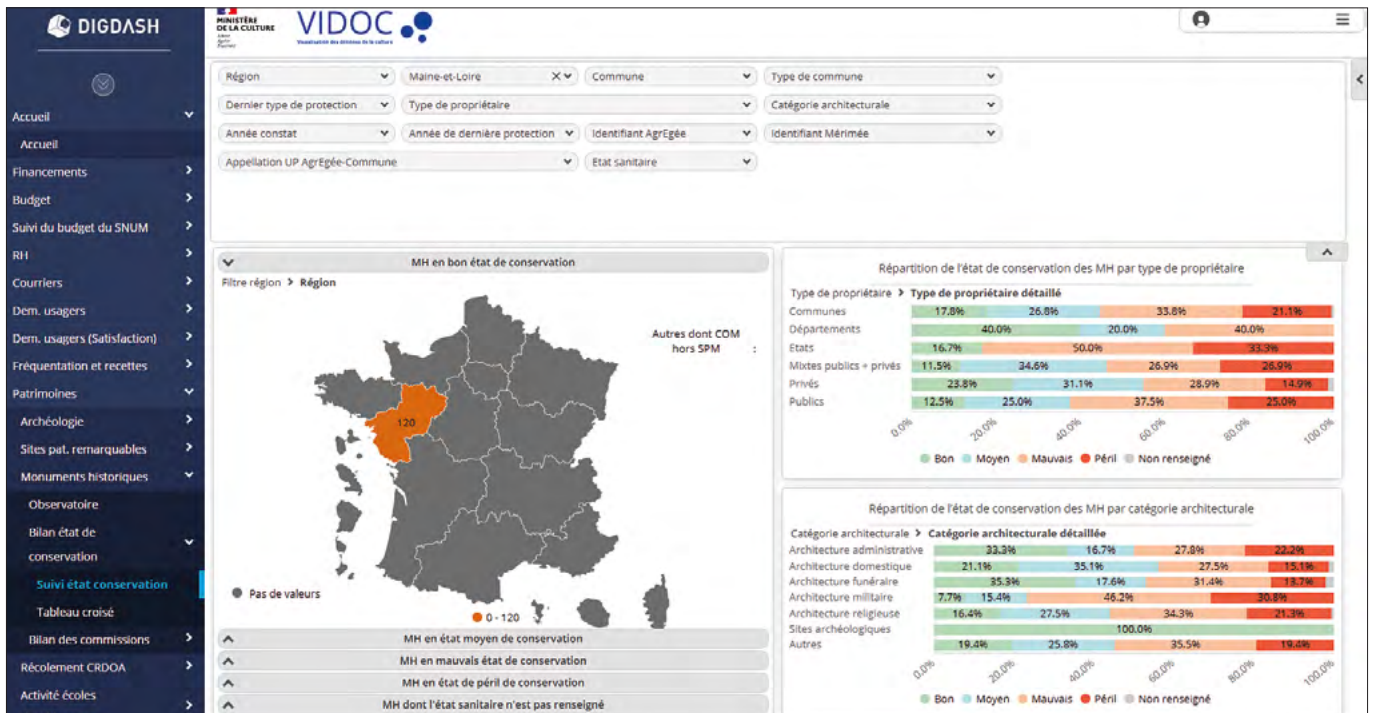
Le tableau de bord de l'état de conservation des monuments historiques (vision nationale dans VIDOC). Source : ministère de la Culture.

4. Notamment sous l'effet de considérations politiques, exposées par exemple par le président de la BnF Jean-Noël Jeanneney dans ses publications, par exemple « Quand Google défie l'Europe » dans *Le Monde* du 22 janvier 2005 : https://www.lemonde.fr/archives/article/2005/01/22/quand-google-defie-l-europe-par-jean-noel-jeanneney_395266_1819218.html (page consultée le 18 mars 2025).

5. Hormis quelques informations à caractère personnel traitées dans le cadre de processus RGPD ou, pour les archives, les inventaires se rapportant à des documents classifiés.

6. Voir notamment *Partager notre Patrimoine culturel. Propositions pour une charte de la diffusion et de la réutilisation des données publiques culturelles numériques*, Secrétariat général du ministère de la Culture et de la Communication, Mission de la communication interne/HFSI, mai 2010, ou la synthèse récente de Samuel Goëta, *Les données de la démocratie. Open data, pouvoirs et contre-pouvoirs*, C&F éditions, 2024.

7. Par exemple *Ouverture des données publiques culturelles. Guide pratique*, 2019 : <https://www.culture.gouv.fr/Thematiques/innovation-numerique/Faciliter-l-acces-aux-donnees-et-aux-contenus-culturels/Ouvrir-et-diffuser-les-donnees-et-contenus-culturels-en-pratique> (page consultée le 18 mars 2025).



Le tableau de bord de l'état de conservation des monuments historiques (vision nationale dans VIDOC).
Source : ministère de la Culture.

Reste l'équation à résoudre entre le souhait d'ouvrir le plus largement possible les reproductions numériques des œuvres libres de droit et le besoin pour les établissements de trouver des ressources financières pour des numérisations ambitieuses – et pour la préservation numérique qui suit. Le problème a trouvé dans le cadre légal en vigueur des réponses distinctes, de l'*Open Content*⁸ des Archives nationales depuis 2017 à des politiques de redevances assumées.

2020-2024 : la mise en avant des cas d'usages

Après ce premier acte de l'ouverture des données culturelles, la création d'un Service du numérique au ministère de la Culture a permis d'engager le chantier pour une meilleure circulation des données à l'intérieur même du ministère. Avec la publication en septembre 2021 d'une *Feuille de route des données et contenus culturels*⁹, puis d'une *Stratégie numérique* en 2022, les *data* se sont positionnées au cœur du système d'information. À partir de cas d'usages concrets discutés avec les services métiers du ministère, ce sont donc les jeux de

La publication des données reste un axe majeur, comme en témoignent les articles du numéro 144 de *Culture et Recherche* dédié à la science ouverte.

données ouvertes demandés par les utilisateurs¹⁰, mais aussi les données internes qui sont mobilisés.

Ce changement de perspective permet une acculturation pragmatique des services, avec un meilleur accompagnement humain et technologique. Des ateliers d'échanges, une gouvernance régulière dans un comité stratégique des données pour le suivi des projets et un comité ministériel du numérique dirigé par la ministre mettent ainsi des besoins en face de l'infrastructure de données en construction, des API et d'une équipe « ligne de produits données ».

La publication des données reste un axe majeur, comme en témoignent les articles du numéro 144 de *Culture et Recherche* dédié à la science ouverte¹¹. Les nouveaux usages des données transforment cependant la manière dont l'information est utilisée au ministère, avec une objectivation quantifiée et rapide de phénomènes auparavant appréhendés plutôt par des remontées qualitatives, comme les enquêtes et le suivi d'activités :

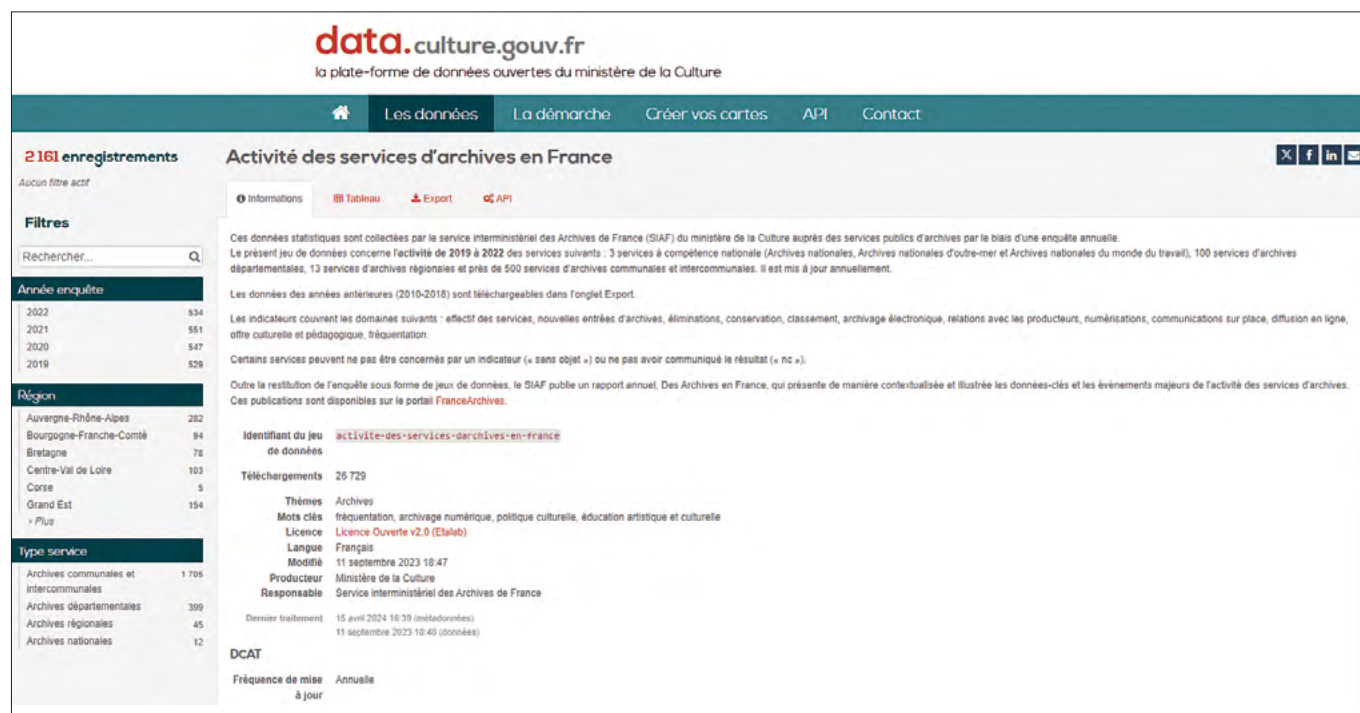
- les enquêtes sectorielles annuelles, qui aboutissent à des rapports très fouillés mais après un

8. C'est-à-dire la libre possibilité de réutilisation des reproductions numériques d'œuvres libres de droit de propriété littéraire ou artistique. Par exemple un registre des mariages du XVIII^e siècle, un plan du Génie militaire du XIX^e siècle, etc.

9. <https://www.culture.gouv.fr/espace-documentation/Documentation-administrative/Feuille-de-route-Donnees-et-contenus-culturels> (page consultée le 18 mars 2025).

10. De décembre 2022 à février 2023, le ministère a par exemple ouvert une consultation en ligne pour voter sur les priorités d'ouverture entre des jeux de données proposés par les services et identifier d'autres besoins. <https://www.culture.gouv.fr/actualites/Open-data-de-nouveaux-jeux-de-donnees-adaptés-aux-attentes-des-utilisateurs> (page consultée le 8 mars 2025).

11. Voir par exemple Roselyne Aliacar, « Les données ouvertes au ministère de la Culture : mises à disposition et usages », *Culture et Recherche*, n° 144, printemps-été 2023, p. 73-74 ; Aurélien Conraux et Romain Delassus, « Données culturelles et science ouverte : pratiques et enjeux du ministère de la Culture », *Culture et Recherche*, n° 144, printemps-été 2023, p. 60-62.



Les données de l'enquête annuelle sur les services d'archives en France (<https://data.culture.gouv.fr>).

important délai d'analyse de réponses textuelles, sont progressivement data-ifiées : remplacement des tableaux Excel par des logiciels d'enquêtes en ligne, suppression de champs de commentaires libres, alignement d'une partie des réponses sur des référentiels contrôlés (les nomenclatures de l'Institut national de la statistique et des études économiques – INSEE, les noms géographiques, les catégories des labels...), etc. Cela permet une utilisation immédiate de la donnée et des traitements informatiques plus développés, ainsi qu'une comparaison facilitée avec les analyses effectuées par le service statistique ministériel¹² ;

- la disponibilité des données dans l'infrastructure centralisée du système d'information (et non plus l'éparpillement dans la base attachée à chaque application mise en production) autorise le croisement des données et la construction de tableaux de bord partagés et interactifs pour suivre l'activité du ministère et évaluer l'impact des politiques publiques.

Le processus d'alimentation automatisée reste encore limité : pour la plateforme de données ouvertes du ministère, il concerne 4 jeux de données seulement (essentiels) sur les 170 mis à disposition en 2025¹³. La remontée automatique des données depuis les systèmes d'information sources devient ici essentielle pour économiser le temps des agents et utiliser des informations aussi récentes que possible. Si pour certains indicateurs l'année écoulée reste encore la référence temporelle pertinente, le mois est désormais une échelle appropriée pour les visualisations de données RH ou de fréquentation des lieux culturels, et la semaine ou le jour pour les relations avec les citoyens.

Et d'ici à 2035 ?

L'ouverture, l'interopérabilité, les réutilisations et la circulation des données culturelles sont au cœur des réflexions engagées depuis 2009. Si des avancées fondamentales ont été accomplies, le contexte a profondément évolué avec l'accélération de la demande de données, la popularisation de l'intelligence artificielle et la montée en puissance des droits culturels. Si l'on tente de se projeter vers le 50^e anniversaire de *Culture et Recherche*, c'est-à-dire à un horizon où les difficultés techniques actuelles, mais aussi les soucis de qualité et de complétude des données comme les sujets d'organisation auront été résolus, que voit-on se dessiner ?

Des utilisations stratégiques

Il semble évident que la donnée culturelle, devenue encore plus accessible, compréhensible, interopérable et mobilisable, ne sera plus uniquement convoquée *a posteriori* mais *ex ante*, pour anticiper l'impact territorial, économique et social des politiques culturelles, et probablement associer davantage les citoyens à l'élaboration des actions locales.

D'une part, l'usage anticipé des données pour la définition des politiques publiques semble avoir bientôt la maturité nécessaire pour produire des résultats pertinents et complémentaires de l'analyse subjective des experts métiers. Mieux formalisées *en matière de données* et plus tôt les visions stratégiques des cadres dirigeants et des décideurs politiques s'appuieront sur la science des données. À un terme plus ou moins distant existeront ainsi des jumeaux numériques où tester l'impact des propositions envisagées en regard de myriades d'informations et de variables. En

12. Le Département des études, de la prospective, des statistiques et de la documentation (DEPS) du ministère de la Culture, département responsable de 40 % des jeux de données publiés sur la plate-forme de données ouvertes du ministère.

13. <https://data.culture.gouv.fr/pages/home/> (page consultée le 18 mars 2025).

Un appel à contributions : le catalogue des données pour la transition écologique (<https://ecologie.data.gouv.fr/>).

14. Voir les propositions de Christine Hennion, Magali Altounian et Bertrand Monthubert, *Rapport de la mission data et territoires*, 2023 : <https://cnig.gouv.fr/rapport-de-la-mission-data-et-territoires-a26165.html> (page consultée le 18 mars 2025).

15. Le Conseil supérieur de la propriété littéraire et artistique consacre depuis plusieurs années une part importante de son activité aux sujets amenés par l'IA, notamment la rémunération des ayants droit dans les corpus d'entraînement des modèles et la définition des conditions d'un usage équitable (c'est-à-dire consenti et rémunéré) des œuvres sous droits.

16. Gallica Images est un projet d'intelligence artificielle de la BnF soutenu dans le cadre du dispositif « France 2030 ». Voir par exemple : <https://www.bnf.fr/fr/les-projets-en-intelligence-artificielle-la-bnf#:text=Le%20projet%20industriel%20Gallica%20Images.journaux%20et%20revues%20et%20de> (page consultée le 18 mars 2025).

facilitant le croisement entre jeux de données issus de la culture, de l'éducation ou du tourisme, ces technologies permettront d'adapter les recommandations aux besoins réels, à l'échelle du pays comme à celle du quartier ou de l'environnement quotidien. Cela nécessitera une interopérabilité numérique accrue entre les informations de l'État et celles venant des collectivités territoriales¹⁴.

L'intégration des citoyens dans la gestion et l'exploitation des données culturelles ouvre également la voie à une participation accrue et à un écosystème plus inclusif et interactif. La rencontre et l'articulation, bien que prospective à ce stade, de l'ouverture des données et des initiatives liées aux droits culturels pourraient se révéler extrêmement enrichissantes pour la conception des politiques culturelles.

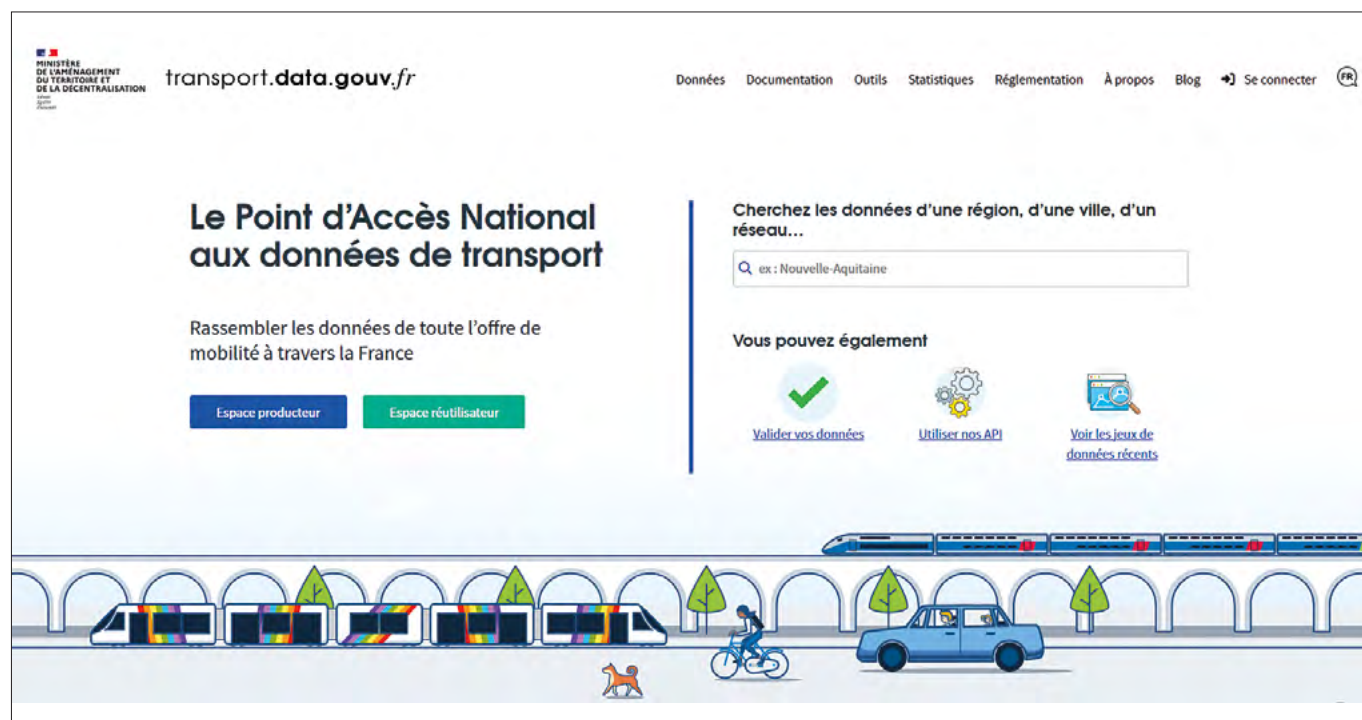
Enfin, le lien entre données culturelles et droit restera étroit, tant pour définir dans leurs métadonnées quels biens culturels et reproductions sont libres de droit, et ainsi concourir à la mission ministérielle de diffusion large des œuvres, que pour faire respecter par les mêmes moyens la propriété littéraire et artistique lorsqu'elle existe¹⁵, permettre l'attribution de droits d'auteur fiables et la rémunération équitable de la création.

Des implémentations techniques très diverses

Au niveau technique, les enjeux majeurs de 2025 vont probablement continuer à structurer la décennie à venir : outre la qualité des données déjà citée et les sujets informatiques, le traitement des données culturelles par algorithmes, notamment par des intelligences artificielles spécialisées, jouera un rôle essentiel dans la description et la découvrabilité des ressources culturelles. À la Bibliothèque nationale de France, Gallica Images par exemple donnera accès à l'ensemble des représentations graphiques, estampes, dessins et photographies existant dans les collections d'imprimés¹⁶, par une segmentation des images et une caractérisation par IA, avec une découvrabilité à terme incomparable avec les possibilités des plus minutieux dépouillements humains.

Pour aboutir à ce résultat, le travail fondamental sur la qualité des données et la technique, déjà engagé depuis plusieurs années, ne change pas. Le premier défi consiste à assurer l'interopérabilité et la normalisation des données culturelles, ouvertes comme internes : standardisation des formats, harmonisation des identifiants culturels (œuvres, artistes, institutions) et structuration des modèles de métadonnées adaptés aux usages des acteurs publics comme privés. Les traitements algorithmiques de masse peuvent ensuite montrer leur pleine puissance.

En matière d'ouverture et de circulation des données, l'enjeu est d'adopter une approche connectée aux usages, en ajustant les stratégies aux pratiques des réutilisateurs – qu'ils soient acteurs culturels, collectivités ou citoyens. Il s'agit



de relier les actions aux attentes, d'encourager des écosystèmes ouverts, mutualisés et évolutifs, et de doter les données culturelles des outils nécessaires à leur exploitation afin de renforcer leur ancrage territorial. Poursuivre la dynamique d'industrialisation des traitements semble d'ores et déjà indispensable pour répondre à la demande croissante et garantir une mise à jour efficace des jeux de données. L'automatisation des processus et l'interopérabilité des infrastructures allègent en effet la charge des producteurs tout en assurant une diffusion plus fiable, qui elle-même incite davantage des tiers à réutiliser les informations.

La mise en place d'une page dédiée aux données culturelles sur data.gouv.fr, sur le modèle des verticales développées autour des données écologiques, météorologiques et de transport¹⁷, facilitera les échanges entre producteurs et réutilisateurs en rassemblant plus simplement les contributions du ministère, des collectivités et des établissements recevant les publics. Au-delà des pratiques actuelles, on peut envisager que citoyens, associations et institutions seront intéressés par des jeux de données « à la demande », générés sur mesure pour répondre à des besoins spécifiques et à des usages personnalisés des données culturelles. ■

Un point central : le « Point d'accès national aux données de transport » (<https://transport.data.gouv.fr>).

17. Le Point d'accès national aux données de transport rassemble ainsi les données de toute l'offre de mobilité à travers la France, et la réglementation impose à la fois l'alimentation et l'usage des données aux institutions concernées. <https://transport.data.gouv.fr/> (page consultée le 18 mars 2025).

Poursuivre la dynamique d'industrialisation des traitements semble d'ores et déjà indispensable pour répondre à la demande croissante et garantir une mise à jour efficace des jeux de données.