

Nathalie Casemajor (2023) « Cultures de données : regard transatlantique sur la mise en données des contenus et des publics », communication au symposium Culture en régime numérique. Questionner les pratiques, catégories et méthodes, DEPS / Ministère de la Culture / et Université Sorbonne Paris Nord, EXPERICE, Paris.

Cultures de données : regard transatlantique sur la mise en données des contenus et des publics

La « culture de données » apparaît comme nouvelle catégorie d'action publique aujourd'hui mobilisée tant au niveau des politiques gouvernementales de transition numérique que des politiques culturelles. Au Québec, le *Guide des bonnes pratiques : découvrabilité et données en culture* publié en 2021 par le Ministère de la culture et des communications se donne pour objectif « d'améliorer la découvrabilité des contenus culturels [et d']aider le milieu culturel à évoluer vers une “culture des données” »¹. Celle-ci est présentée comme « une clé importante pour une transformation numérique réussie »².

En France, la terminologie de la culture des données est également mobilisée. C'est le cas dans la *Feuille de route sur la politique du ministère de la Culture concernant les données, les algorithmes et les codes sources* (2021). On peut y lire que « le ministère souhaite développer une culture commune de la donnée » pour contribuer à atteindre ses objectifs de « diffuser et rendre visible les contenus culturels », d'améliorer « l'efficacité des actions du ministère » et de contribuer au développement économique du secteur³. Pour développer cette culture commune, la feuille de route préconise d'« acculturer » à l'utilisation de la donnée « à tous les niveaux de l'organisation » du « ministère et [de] son réseau d'opérateurs »⁴.

¹ MINISTÈRE DE LA CULTURE ET DES COMMUNICATIONS DU QUÉBEC, *Guide des bonnes pratiques : découvrabilité et données en culture*, Paris, MCCQ, 2021, p. 5.

² *Ibid.*

³ MINISTÈRE DE LA CULTURE (FRANCE), *Feuille de route sur la politique du ministère de la Culture concernant les données, les algorithmes et les codes sources*, Paris, ministère de la Culture, 2021, p. 25, 4.

⁴ *Ibid.*, p. 24.

Ce chapitre propose d'inscrire la découvrabilité dans une problématique de recherche plus large : celle de la (ou plutôt des) cultures de données. La prescription faite aux professionnel.les et aux organismes du secteur culturel de développer une culture des données est sous-tendue par un présupposé : soit il existerait chez eux un manque de culture des données, soit ils possèderaient déjà une culture des données mais celle-ci serait insuffisante ou inadaptée. Or ce présupposé doit être interrogé : il existe en effet dans le secteur culturel une multiplicité de pratiques relatives aux données, inscrites dans des modes sociaux hétérogènes. Et par ailleurs il existe une multiplicité de types de données en jeu, autant qu'une multiplicité d'infrastructures de données peu interopérables.

J'aborderai les cultures de données sous l'angle de leur « dynamique de catégorisation » selon les mots de Bénédicte Zimmerman⁵, c'est-à-dire (en ce qui concerne ce chapitre) la production de leurs significations dans le champ des politiques culturelles et des pratiques professionnelles du secteur de la culture. Quelles sont les formes actuelles de cadrage de la culture de données ? Comment ces cadrages façonnent-ils les significations et les pratiques liées aux cultures de données ? La notion de cadrage est mobilisée ici au sens d'un processus de sélection de certains aspects d'une réalité perçue⁶, produisant un schéma d'interprétation qui cadre la compréhension de l'objet « culture de données ». Ces questions nous conduisent à resituer les cultures de données dans le contexte des transformations plus larges qui touchent l'accès à la culture dans les environnements numériques.

Mon analyse s'appuie sur l'étude d'un contexte organisationnel particulier : celui de Bibliothèque et Archives nationales du Québec (BAnQ), une société d'État opérant sous la tutelle du gouvernement provincial du Québec. BAnQ est un cas d'étude intéressant parce qu'il s'agit d'une organisation hybride. Elle rassemble différents univers de pratique en

⁵ Bénédicte ZIMMERMANN, *Éléments pour une socio-histoire des catégories de l'action publique. Historicités de l'action publique*, Paris, PUF, 2003, p. 241-258.

⁶ Robert N. ENTMAN, *Projections of Power: Framing News, Public Opinion, and US Foreign Policy*, Chicago, University of Chicago Press, 2004. Contrairement aux travaux d'Entman, le rôle des médias dans la production des cadrages n'est pas abordé ici. L'analyse porte plutôt sur les documents de politique publique et de gouvernance institutionnelle, ainsi que sur les dispositifs sociotechniques de gestion de données.

lien avec les cultures de données : celui des archives et celui des bibliothèques, qui eux-mêmes associent plusieurs logiques d'action (valorisation du patrimoine et services d'accès libre pour le public).

Se pencher sur le domaine des bibliothèques et archives permet aussi d'examiner les nouveaux paradigmes des données à l'aune de référentiels de pratique établis de longue date. En effet, les milieux documentaires sont parmi les premiers dans le secteur culturel à avoir développé une expertise liée à la numérisation des systèmes d'information. En portant un regard transatlantique – depuis le Québec – je souhaite aussi donner un éclairage sur certaines tendances qui traversent les frontières nationales.

Pour étudier le cas de BAnQ, je m'appuie sur un matériau d'observation participante recueilli depuis 2020 dans le cadre d'un partenariat de recherche avec BAnQ autour de la gouvernance des données d'usage⁷. Je m'appuie également sur une série de groupes de discussion menés en 2021 auprès de professionnels de la culture au Québec dans le contexte d'un projet de recherche⁸. J'ai également mené diverses observations dans le cadre d'ateliers et de communautés de pratique rassemblant des professionnels du secteur culturel au Québec et en France.

1. Conceptualiser la notion de culture de données

L'appellation « *data culture* » a fait son apparition dans les années 1960, au moment de l'essor de l'industrie informatique. Elle a connu un regain d'intérêt dans les années 2010, parallèlement à l'expansion du phénomène de « mise en données » (*datafication*) de la

⁷ Nathalie CASEMAJOR, Guillaume SIROIS, Guy BELLAVANCE, Nathalie VACHON, François DERBAS THIBODEAU, Alexandra BEAUSÉJOUR-VILLEMAIRE et Alexis MORIN, *Gouvernance des données d'usage dans les bibliothèques et archives : sondage sur les perceptions des usagers et usagères de Bibliothèque et Archives nationales du Québec*, Rapport de recherche, Montréal, Institut national de la recherche scientifique, 2023. Nathalie CASEMAJOR, Guillaume SIROIS, Lily-Cannelle MATHIEU, Alexis MORIN et Louis MELANÇON, *La gouvernance des données d'usage : enjeux éthiques et perceptions des publics dans les bibliothèques et archives, Synthèse de la littérature*, Institut national de la recherche scientifique, Montréal, 2021.

⁸ *La mise en données des publics de la culture. Portrait des pratiques actuelles (2020-2021)*, projet dirigé par Guillaume Sirois, avec Nathalie Casemajor et Guy Bellavance, financé par InventT (Université de Montréal).

société. Pour situer plus précisément mon angle d'analyse, je propose de distinguer trois plans des cultures de données⁹. Premièrement celui d'une *culture de données de type holistique*, au sens de cadres d'expérience d'un monde mis en données. Deuxièmement, celui des *cultures d'usage*, qui se situe du côté des publics, ceux-ci développant une série de pratiques liées à la réception de l'offre et aux usages quotidiens (ordinaires) des dispositifs connectés. Et troisièmement, le plan des *cultures de production*, qui est resserré sur l'activité professionnelle ou amateur engagée dans la création ou la transformation de données. C'est sur ce dernier plan des cultures de production que portera mon analyse, et en particulier sur la sphère des pratiques professionnelles.

Dans les travaux académiques, on peut identifier plusieurs approches épistémologiques des cultures de données (je me limiterai ici aux champs francophone et anglophone). Les approches francophones se situent principalement dans le domaine de l'information documentation, mais aussi de la communication organisationnelle et de l'économie de la culture. Tandis que les travaux anglophones se situent plutôt dans le champ des études critiques de données et des études culturelles.

Dans le cadre de mes propres travaux, je m'appuie à la fois sur les études critiques de données¹⁰ et sur les études des infrastructures technologiques¹¹. Les travaux en sociologie des bases de données¹² sont également éclairants puisqu'ils mettent l'accent sur les modes de coordination du travail des données au sein de mondes sociaux hétérogènes.

⁹ Nathalie CASEMAJOR, « Les cultures de données : genèse d'une notion, enjeux épistémologiques et implications pratiques », in Florence MILLERAND, Alexandre COUTANT, Guillaume LATZKO-TOTH, Mélanie MILLETTE (sous la dir. de) *Datafication et publics de données. Penser la mise en donnée de la société*, Montréal, Presses de l'Université de Montréal, à paraître.

¹⁰ Rob KITCHIN, « Big Data, new epistemologies and paradigm shifts », *Big Data & Society*, vol. 1, n° 1, 2014. Andrew ILIADIS et Federica RUSSO, « Critical data studies: An introduction », *Big Data & Society*, vol. 3, n° 2, 2016.

¹¹ Paul N. EDWARDS, Geoffrey C. BOWKER, Steven J. JACKSON et Robin WILLIAMS, « Introduction: an agenda for infrastructure studies », *Journal of the Association for Information Systems*, vol. 10, n° 5, 2009, p. 364-374. Brian LARKIN, « The politics and poetics of infrastructure », *Annual Review of Anthropology*, n° 42, 2013, p. 327-343.

¹² Patrice FLICHY et Sylvain PARASIE, « Présentation. Sociologie des bases de données », *Réseaux*, vol. 2, n° 178-179, p. 9-19.

Je retiens les éléments de compréhension suivants de la notion de culture de données, non pour en figer une définition, mais pour guider l'analyse :

La notion de culture de données réfère à un assemblage de répertoires collectifs de représentation, d'action et de justification situé dans un contexte spécifique et caractérisé par une sensibilité, un univers de sens et une rationalité construits en relation avec des ensembles de données.

Ces répertoires contribuent à façonner les données, leur statut et leurs effets sur le monde. Ils reposent sur des valeurs, des normes (explicites ou implicites), des littératies, des pratiques, des affects et des dispositifs techniques.¹³

À partir de la configuration particulière des pratiques à BAnQ, je souhaite donc examiner comment les différentes significations que prend cette catégorie de culture de données articulent des répertoires propres aux domaines des bibliothèques et des archives. Je souhaite en particulier montrer comment ces répertoires évoluent au regard des grandes dynamiques qui façonnent l'accès à la culture dans les environnements numériques ; c'est-à-dire la découvrabilité, la dynamique de *datafication* (ou « mise en données ») et la dynamique de plateformisation.

Par découvrabilité, dans le contexte de BAnQ, j'entends à la fois la *visibilité* (*ranking*), autrement dit la capacité d'un contenu à apparaître dans un flux de données où on ne cherchait rien en particulier ; la « *trouvabilité* », soit la capacité à trouver ce que l'on cherche ; et la *recommandation*, qui consiste à suggérer des contenus spécifiques à un utilisateur, soit en fonction de la similarité entre des contenus, ou en fonction de son profil ou de son comportement d'utilisateur.

¹³ Nathalie CASEMAJOR, « Les cultures de données », art. cité.

Quant aux travaux sur la plateformes¹⁴, au-delà des études de plateformes spécifiques et de leurs algorithmes de recommandation, ils se penchent sur les processus de développement *d'infrastructures de données reprogrammables, à haut degré de connectivité au sein d'un écosystème web*. Je m'intéresse en particulier aux effets idéologiques de la plateformes au sens où ils valorisent un modèle organisationnel qui instaure une économie et une politique techniciste des flux de données.

Enfin les travaux sur la mise en données éclairent les processus de découvrabilité en étudiant la transformation en ressources de données de pan de plus en plus larges de la vie. Cette mise en données du monde se déploie dans le contexte d'une intensification des modes d'exploitation des données par de nouveaux outils algorithmiques d'analyse automatisée. Elle a de puissants effets idéologiques que José Van Dijck¹⁵ a nommé « dataïsme » pour qualifier la croyance en l'objectivité des données et ses associations avec des logiques qualifiées par d'autres travaux « d'extractivisme de données » et d'accumulation monopolistique de ressources de données au profit d'un capitalisme de plateforme¹⁶.

C'est à travers une perspective critique des dynamiques de la découvrabilité, de la plateformes et de la mise en données que j'aborderai donc l'analyse des cultures de données comme catégorie d'action dans le contexte de Bibliothèque et Archives nationales du Québec (BAnQ).

¹⁴ Anne HELMOND, « The Platformization of the Web: Making Web Data Platform Ready », *Social Media+ Society*, vol. 1, n° 2, 2015, p. 1-11.

¹⁵ José VAN DIJCK, « Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology », *Surveillance & Society*, vol. 12, n° 2, 2014, p. 197-208.

¹⁶ Nick SRNICEK, *Capitalisme de plateforme. L'hégémonie de l'économie numérique*, Montréal, Lux, 2018. Tiziana TERRANOVA, *After the Internet: Digital Networks Between Capital and the Common*, MIT Press, Cambridge (MA), 2022.

2. Développer la culture des données ?

2.1 Un chantier de gouvernance

Dans le cadre d'un large « Plan de transformation numérique », BAnQ a publié en 2021 une *Stratégie numérique*, dont l'une des orientations est formulée ainsi : « l'innovation est accélérée par l'exploitation des données et par l'intelligence artificielle »¹⁷. Dans cette perspective, l'exploitation des données vise à « améliorer[r] l'expérience des usagers » en exploitant l'historique d'emprunt des documents pour cibler les « besoins » et les « intérêts » des usagers¹⁸. Elle vise aussi à améliorer la « performance » de l'organisation via « l'intelligence d'affaire », tout en devenant un modèle de « pratiques exemplaires » sur le plan éthique¹⁹.

Cette stratégie de valorisation s'applique aux « données institutionnelles » de BAnQ, définie comme : « l'ensemble des données produites ou reçues par BAnQ [...] qui sont nécessaires pour [...] documenter la réalisation de ses missions, pour effectuer des opérations, pour prendre des décisions ou pour rendre compte de ses activités »²⁰. Ces données institutionnelles « englobent les données descriptives, transactionnelles et d'usage »²¹.

Au cours de l'élaboration de sa nouvelle stratégie numérique, l'institution a lancé en 2020 un « chantier » sur les données. Il s'attèle à des enjeux d'ordre juridique, technologique et organisationnel liés à la gestion des données. Parmi les actions entreprises par ce chantier, on trouve la création d'un laboratoire d'innovation et la définition d'un mode de gouvernance des données. Pour cela, une dizaine de cadres et d'agent.es ont été rassemblé.es dans un Comité de gouvernance des données.

¹⁷ BAnQ, *Stratégie numérique 2021-2028*, Montréal, BAnQ, 2021, p. 14.

¹⁸ *Ibid.*

¹⁹ *Ibid.*

²⁰ BAnQ, *Directive relative à l'utilisation ou à la communication sans consentement de renseignements personnels à des fins d'étude, de recherche ou de production de statistiques*, Montréal, BAnQ, 2022, p.4.

²¹ *Ibid.*

Ce chantier a aussi mené à l'élaboration d'un corpus réglementaire visant à clarifier les règles et processus entourant la gestion des données, dont *une Politique sur le partage et la valorisation des données* (2022) et une *Feuille de route sur l'intelligence artificielle* (2023). Examinons maintenant plus en détail les enjeux de coordination de l'activité soulevés lors de ce chantier sur les données. Il s'agit premièrement, selon les sources internes avec qui j'ai pu échanger, de forger une vision commune de la culture des données à BAnQ ; et deuxièmement, de faire adhérer à cette vision les membres du personnel de BAnQ (pas une mince affaire puisqu'ils s'élèvent à plus de 600).

2.2 Vision commune et diversité de cultures de données

Un premier défi rencontré lors de ce chantier a été de se mettre d'accord, au sein de l'organisation, sur des principes communs qui permettraient de guider l'exploitation des données. Or les activités liées aux données varient au sein de l'institution, à la fois en raison de cultures de métier différentes, mais aussi en raison des fonctions différentes occupées par les membres du personnel.

Les travaux de recherche sur les cultures de données réfutent à juste titre les approches réductrices selon lesquelles il n'existerait qu'*une* culture de données homogène et partagée au sein d'une même organisation, voire au sein du secteur culturel. On observe plutôt, à l'intérieur d'une organisation une *diversité* de cultures de données. À BAnQ, cette diversité est liée au fait que l'institution rassemble trois entités distinctes, réunies assez récemment (en 2006) : les archives nationales (fondées en 1920 avec une mission de préservation de l'histoire du Québec) ; une bibliothèque nationale (créée en 1968 et dédiée à l'édition québécoise et à la conservation et à la mise en valeur de collections patrimoniales) ; et une bibliothèque publique située à Montréal (la Grande bibliothèque, créée en 2005 pour offrir un service de consultation et d'emprunt de documents à la fois sur place et en ligne).

Certes, ces trois entités partagent des valeurs communes au sein d'une même culture organisationnelle – il s'agit notamment de l'accès au savoir et de la production de données de référence et de qualité. Mais en raison de leurs cultures de métier, ces entités

entretiennent des relations différentes aux données : les normes professionnelles et les référentiels de pratique divergent entre le domaine des bibliothèques et celui des archives par exemple la normalisation bibliographique via le format MARC (*Machine-Readable Cataloging*) n'est pas la même que les standards de description des fonds d'archives (tel le standard ISAD(G) - *General International Standard Archival Description*). Par ailleurs, les infrastructures de données utilisées par les bibliothécaires et les archivistes à BAnQ différent— ainsi le système de gestion des requêtes de consultation des documents d'archive est distinct du système de gestion des emprunts des documents de la Grande bibliothèque.

Outre ces trois entités (archives, bibliothèque patrimoniale et bibliothèque publique), le comité sur la gouvernance des données rassemble des membres d'autres services, qui ont encore d'autres rapports aux données, liées à leurs fonctions au sein de l'organisation : le secrétariat général s'intéresse aux données sous l'angle de la gestion opérationnelle ; le service juridique se préoccupe notamment du respect des lois relatives à la protection des renseignements personnels ; et la direction numérique joue un rôle de premier plan dans la stratégie de « transformation numérique » de l'organisation. Cette dernière, aux côtés du service des communications, adopte la posture plus favorable à l'exploitation des données.

Mais à l'intérieur même de ces services, on peut encore distinguer ce que Sarah Labelle²² nomme : des « profils data » (ceux qui possèdent un « usage et une connaissance avancée des données », tels les analystes de données, les chargés de marketing et les gestionnaires de systèmes d'information) ; et des profils « non data », des professionnels qui font parfois de la saisie de données mais pour lesquels l'exploitation des données n'est pas au cœur du métier en termes de prise de décision ou de développement de nouveaux services. On peut donc mesurer le potentiel de faire face à des visions divergentes de l'exploitation des données au sein de l'institution – des divergences qui sont qualifiées de « freins » par une source interne de BAnQ. La définition d'une culture de données « relève donc d'un

²² Sarah LABELLE (sous la dir. de), *Vers une organisation data driven : mettre la culture des données au cœur de l'organisation*, Paris, Infolab - FING, 2017.

processus de négociation » entre les différentes perspectives, comme le mentionne à juste titre Anne Lehmans²³.

Face à cette multiplicité d'univers de pratique et de points de vue, une consultation interne a été menée à BAnQ pour tenter de dégager une vision collective. Cette consultation a réuni lors d'ateliers une trentaine de professionnels clés (issus entre autres des services juridiques, des technologies de l'information, des finances et des ressources humaines). L'objectif était d'identifier des principes autour desquels se dégageait une « forte adhésion des participants » (selon une source interne), par exemple l'intérêt supérieur du service rendu aux usagers.

Sans présager des divergences précises qui ont animé la consultation à BAnQ, on peut identifier parmi les discussions en cours dans le milieu culturel certains points de tension concernant l'exploitation des données. Il s'agit notamment de l'éthique de la transformation du travail en lien avec l'automatisation de certaines activités. Il s'agit aussi du profilage des usagers et usagères, qui soulève des enjeux juridiques de protection des renseignements personnels, mais aussi des enjeux de valeur liés à une conception des publics comme audience à cibler grâce aux données. Il peut aussi s'agir de luttes de position entre métiers au sein de l'organisation. Ou encore de tensions hiérarchiques entre une vision essentiellement portée par des cadres, et le reste des employés.

2.3 Le prisme du manque et de l'acculturation

En 2021, un cadre de BAnQ estimait qu'il n'y avait « pas de culture des données » au sein de l'organisme. Les actions entreprises depuis lors ont surtout porté sur la définition d'une vision stratégique suivie d'une campagne interne pour « faire adhérer [plus largement les membres du personnel] à cette culture de la donnée », et sur la mise en place d'un dispositif de gouvernance, suivi d'une campagne pour « faire sortir les informations du comité »

²³ Anne LEHMANS, « Données ouvertes et redéfinition de la culture de l'information dans les organisations. Vers une culture de la donnée », *Communication et organisation*, n° 51, 2017, p. 20.

(propos recueillis auprès de cadres de BANQ), c'est-à-dire sensibiliser les membres du personnel aux enjeux et les informer sur leurs responsabilités et leurs obligations. Deux ans plus tard, en 2023, l'instauration d'une culture de données est toujours un « défi », selon une source interne.

Il pourrait être tentant d'expliquer ce défi par le manque : manque d'adhésion à la vision stratégique ; manque de littératie des données de la part des employé.es ; manque d'interopérabilité entre les infrastructures de données. C'est la lecture à laquelle invite l'optique des stades de développement de la « maturité numérique » qui est prônée par certaines approches gestionnaires de la « transformation numérique » des organisations²⁴. Dans cette perspective, le manque de culture des données serait un facteur d'immaturité numérique, et « l'acculturation » à la culture de données serait au contraire un facteur de développement organisationnel.

Le vocable de l'acculturation pour référer au développement d'une culture de données est problématique. Il induit une conception linéaire, mécaniste et unilatérale du changement culturel²⁵ : il suffirait que les professionnel.les assimilent la vision adéquate et les compétences adéquates pour effacer tous les freins empêchant d'accéder à un stade de développement présenté comme supérieur. Outre sa propension idéologique, cette approche ne permet pas de comprendre les tensions qui se nouent dans les débats portant sur les principes supérieurs qui devraient guider le travail avec les données.

En contrepoint de l'optique du manque, je propose plutôt d'envisager les défis rencontrés à BANQ comme la conséquence d'un surplus : je veux dire par là que la multiplicité des significations et des pratiques qui se télescopent sous le vocable de culture de données viennent brouiller le sens de ce répertoire d'action en amalgamant des logiques disparates, voire contradictoires.

²⁴ Kim Viborg ANDERSEN et Helle Zinner HENRIKSEN, « E-government maturity models: Extension of the Layne and Lee model », *Government Information Quarterly*, vol. 23, n° 2, 2006, p. 236-248.

²⁵ Zohra GUERRAOUI, « De l'acculturation à l'interculturalisation : réflexions épistémologiques », *L'Autre*, vol. 10, n° 2, 2009, p. 195-200.

Dans la suite de ma présentation, je propose de réexaminer l'évolution des cultures de données à BAnQ selon trois cadrages coexistant qui en façonnent les significations et les pratiques : 1) un cadrage par la *documentation des objets culturels* ; 2) un cadrage par *l'accès public aux objets* ; 3) un cadrage par la *gestion organisationnelle*. L'intérêt de cette approche par cadrages coexistants est d'envisager les cultures de données comme des systèmes complexes traversés par des ambivalences et des contradictions internes, certaines étant provoquées ou exacerbées par les logiques de la mise en données et de plateforme.

Cette approche permet aussi de s'éloigner d'une analyse de type diffusionniste qui formulerait le problème sous l'angle de la rencontre entre des cultures de données différentes. Le défaut de celle-ci étant de tendre vers une description essentialisée, de figer certains traits de ces cultures pour montrer ensuite comment ils se transforment au contact les unes des autres.

3. Trois cadrages des cultures de données

3.1 Documentation des objets culturels

Le premier cadrage de la culture des données que l'on peut identifier à BAnQ est celui de la gestion et de la production de données dans des systèmes documentaires. Dans cette perspective, la culture de données réfère notamment à la production de données descriptives, c'est-à-dire les données de catalogage qui fournissent de l'information sur les objets culturels (leurs caractéristiques, leur provenance, leurs droits d'utilisation)²⁶. Ce cadrage a connu plusieurs inflexions majeures dans les trente dernières années en lien avec de nouveaux paradigmes de valorisation des données : il s'agit, des données ouvertes, de la production d'outils automatisés « compagnons » de métier et des intelligences artificielles (IA) génératives. Je m'attarderai ce sur dernier paradigme.

²⁶ BAnQ, *Politique sur le partage et la valorisation des données*, Montréal, BAnQ, 2022.

Le développement d'outils tels que ChatGPT, fondés sur de grands modèles de langue, suppose d'entraîner des systèmes informatiques sur de larges corpus de documents. Dans ce contexte, les établissements patrimoniaux sont sollicités comme fournisseurs de données d'entraînement. C'est le cas en France avec l'appel à projet « Communs numériques pour l'intelligence artificielle générative » lancé en 2023 par la Banque Publique d'Investissement. Ce nouveau mode de valorisation des corpus documentaires relève d'une mise en données du patrimoine dans le cadre d'un développement industriel et commercial.

Il suscite des tensions entre certains référentiels de pratique chez les professionnel.les du domaine des bibliothèques et archives. Ces tensions sont liées, d'une part, à la conciliation des valeurs d'innovation et de respect du droit de la propriété intellectuelle. Et d'autre part, elles découlent d'une forme de monétisation du patrimoine documentaire : en effet, pour s'engager dans un partenariat commercial de ce type, il faudrait savoir chiffrer la valeur monétaire de rente des collections. Cette monétisation du patrimoine (dans sa version mise en données) peut aussi entrer en contradiction avec la logique civique de l'ouverture des données, puisque l'accès privilégié à des données d'entraînement offre un avantage compétitif à un projet industriel de développement d'IA.

D'ailleurs, la nouvelle politique de valorisation des données à BANQ inclut une clause selon laquelle le principe d'ouverture « par défaut » des données institutionnelles peut être écarté au profit d'une entente commerciale avec un tiers si une « justification sérieuse et documentée » en est faite. En résumé, cette inflexion qualifie la culture de données comme un vecteur d'innovation technologique, voire de développement d'applications industrielles.

3.3 Accès public aux collections

Le deuxième cadrage de la culture de données que l'on peut identifier à BANQ est celui de l'accès aux collections sous l'angle de la découvrabilité. La culture de données réfère dans ce contexte à des modes d'exploitation des données en vue de favoriser l'accès aux collections. De multiples moyens sont utilisés ou expérimentés par les bibliothèques et

archives pour rendre plus découvrables les contenus de leurs collections. Ces nouveaux moyens sont liés au développement du web sémantique, des interfaces de recherche dites « intelligentes » et des robots conversationnels (un outil envisagé à BAnQ). Tous ces exemples sont essentiellement fondés sur les données descriptives des contenus.

D'autres dispositifs, liés à la recommandation algorithmique de contenus, misent plutôt sur le croisement entre données descriptives et données d'usage. Par données d'usage, dans le contexte de BAnQ, on entend des renseignements sur l'utilisateur qui sont issus de son activité et, plus précisément, de son interaction avec les systèmes d'information de BAnQ et leurs contenus. En 2023, BAnQ comptait près de 400 000 abonnés, environ 3 millions de documents empruntés ou consultés sur place et 46 millions de documents numériques empruntés ou consultés en ligne. Quelles sont les sources de données d'usage dans le contexte des bibliothèques et archives ? On a déjà évoqué tout à l'heure les requêtes de consultation des documents d'archives. On trouve aussi l'historique de navigation des usagers et usagères sur les sites web institutionnels (logs). Mais à BAnQ, la principale source de données d'usage que l'institution souhaite exploiter est l'historique d'emprunt de documents, fournie par les systèmes de gestion interne des abonnements à la Grande bibliothèque.

Les renseignements sur les abonnés collectés par BAnQ incluent quelques données socio-démographiques (notamment l'âge et le code postal) et surtout le comportement d'emprunt d'un.e abonné.e depuis son inscription à la bibliothèque. Il s'agit donc d'un réservoir très riche puisque ces données d'emprunt ne sont pas purgées régulièrement (comme l'y obligent certaines juridictions, notamment en lien avec le RGPD européen). Plus de 100 millions de transactions y sont amassées depuis la création de l'entrepôt de données.

Le cas de la recommandation de contenus fondée sur les données d'usage est particulier dans le contexte des bibliothèques et archives, et ce pour deux raisons. Contrairement à la visibilité des contenus sur les plateformes, sur laquelle les professionnels du domaine ont peu de prise, la recommandation algorithmique peut s'intégrer aux interfaces et moteurs de recherche institutionnels. Et deuxièmement, ce dispositif est aussi particulier parce qu'il

implique une mise en données des publics, ce qui produit certaines inflexions du cadrage de l'accès aux collections.

Dans cette inflexion, les géants du numérique (tels Amazon, Facebook ou YouTube) servent à la fois de modèle et de repoussoir pour les pratiques de découvrabilité. D'un côté, les services d'analyse de données sur les publics en ligne qui sont offerts par ces plateformes commerciales sont devenus des outils de référence incontournables pour le milieu culturel. Meta ou Google fournissent en effet des services clé en main sous la forme de tableaux de bord (Statistiques de *Meta Business Suite* pour Facebook et Instagram, ou *Google Analytics*). L'avantage pratique de ces services est qu'ils sont fournis par des tiers dans un format déjà interprétable. Mais évidemment, la propriété de ces données d'audience web reste dans les mains des plateformes et elles ne peuvent pas être librement exploitées par les organismes culturels.

Par ailleurs, le succès des systèmes de recommandation des plateformes commerciales est envié par de nombreux professionnels de la culture, au titre d'un principe d'efficacité. L'un des participants à nos entretiens collectifs évoque ces systèmes de recommandation comme « un beau rêve... » qui resterait hors d'atteinte dans le domaine des bibliothèques et archives au Québec. Il ne pourrait se réaliser, selon lui, « qu'aux États-Unis, jamais au Québec » en raison des limites technologiques des systèmes actuellement utilisés dans la province.

Mais c'est autour des enjeux de la personnalisation des services que l'on trouve des points d'inflexion particulièrement forts dans le cadrage de l'accès aux collections. Ainsi, la *Stratégie numérique* de BANQ énonce parmi ses priorités le fait de « pousser plus loin la personnalisation [des] services [...] en s'appuyant sur [...] des données et des statistiques d'utilisation reflétant [les] habitudes et [les] préférences [des usagers] »²⁷. Toutefois, on observe à BANQ – et plus largement au sein des institutions culturelles – une grande ambivalence concernant la place de la personnalisation dans les systèmes de recommandation algorithmique. L'objectif est d'optimiser l'accès grâce à la

²⁷ BANQ, *Stratégie numérique 2021-2028*, op. cit., p. 11.

personnalisation, mais pour reprendre les termes d'un cadre de BAnQ, « on ne veut pas faire de la recommandation trop personnalisée ».

Cette ambivalence est liée à une tension entre un imaginaire de l'efficacité algorithmique et une crainte de ses effets. On trouve donc d'un côté l'idée que l'on pourrait atteindre une adéquation optimale entre les contenus, les publics et les situations de découvrabilité : la stratégie numérique de BAnQ se donne ainsi pour ambition de « dispenser la bonne information, de la bonne façon, à la bonne personne et au bon moment »²⁸.

Cet imaginaire est aussi sous-tendu par une conception des données d'usage comme exhaustives, objectives et transparentes. Dans nos entretiens collectifs, certains professionnel.les évoquaient le fait que « les données comportementales servent à laisser parler les publics », elles donneraient une voix aux « usagers silencieux ». Mais cette optimisation algorithmique de la recommandation personnalisée est aussi perçue comme nuisible par divers.es professionnel.les – ce n'est pas tant l'illusion de l'optimisation parfaite qui est l'objet de leur critique, mais plutôt les effets de sa supposée grande puissance de profilage et prédiction des goûts.

En particulier, l'effet (réel ou supposé) de bulle de filtre fait office de repoussoir. Les professionnel.les avec lesquels nous avons échangé veulent éviter d'enfermer les publics dans des profils de recommandation trop ciblés. Il s'agit donc de recommander tout en ouvrant l'horizon de la découverte en dehors des sentiers battus. Amazon, qui a débuté dans le secteur de la vente de livres, est particulièrement désigné par les personnes que nous avons interviewées comme l'exemple à éviter, étant perçu comme nuisible à la diversité de l'offre culturelle.

Cette approche de la recommandation semi-personnalisée est aussi privilégiée dans d'autres systèmes de recommandation d'établissements publics, notamment en France. C'est le cas de l'offre du Pass Culture (destinée aux jeunes de 18 ans) où la curation des contenus « relève à la fois d'une sélection “manuelle” d'offres [et] d'une approche

²⁸ *Ibid.*

marketing »²⁹. C'est aussi le cas dans le système de recommandation de podcasts de Radio France qui se présente comme une « recommandation hybride »³⁰ incluant la curation manuelle, l'historique de lecture et le filtrage collaboratif en fonction de profils d'écoute.

En résumé, le cadrage de l'accès aux collections sous l'angle de la découvrabilité produit une inflexion de la relation aux publics. Leur mise en données soulève une tension entre d'un côté, un principe d'efficacité de l'accès, et de l'autre côté, le principe de la protection de la diversité de l'offre.

3.2 La gestion organisationnelle

Le troisième et dernier cadrage de la culture des données à BAnQ est celui de la gestion organisationnelle. Ici la culture des données prend le sens de stratégie de management pour optimiser la « performance organisationnelle » grâce à une prise de décision fondée sur l'analyse de données (ou « pilotage par les données »). Ce cadrage met – lui aussi – de l'avant un principe industriel d'efficacité. La mise en donnée y vise à rationaliser les activités de l'ensemble des services de l'organisation. Il suppose aussi une conception de ces données comme un actif stratégique de l'organisation.

À BAnQ, cette mise en données de la gestion se traduit par un fort accent mis par la direction actuelle sur l'optimisation des services et des ressources. Dans cette perspective, la découvrabilité est envisagée (selon la PDG) comme une « une question d'efficacité opérationnelle » qui est mesurable par des indicateurs (par exemple, nombre de documents patrimoniaux consultés en ligne). Ce cadrage de la culture de données s'appuie sur l'utilisation de systèmes d'information hérités de l'informatique décisionnelle (Progiciels de gestion intégrée ou ERP pour *Entreprise Ressource Planning*; Flichy, 2013). Ces

²⁹ Cyrille DELABRE et Tristan MARTIN-STOCKER, « Pass Culture », *Penser la découvrabilité des contenus culturels. Rencontres professionnelles*, Paris, BnF, 20 juin 2023.

³⁰ Bastien LUNETEAU et Valentin LECOMTE, « La mise en place d'une recommandation personnalisée de service public chez Radio France », *Penser la découvrabilité des contenus culturels. Rencontres professionnelles*, Paris, BnF, 20 juin 2023.

systemes ne sont pas nouveaux mais ils ont connu un développement accéléré depuis une dizaine d'année grâce à l'analyse de données en soutien à la prise de décision stratégique.

À BAnQ, cette orientation se traduit par une directive récente, de la part de la nouvelle direction, pour établir un « comptoir de données » : selon la PDG, « la réalisation des projets repose désormais sur l'utilisation de tableaux de bord et des cibles sont établies pour chacun d'entre eux »³¹. Mais reste la difficulté de définir les indicateurs d'atteinte de ces cibles, ce qui soulève, selon une source interne, « des enjeux de fiabilité des données ».

Pour produire ces rapports, BAnQ utilise le système PowerBI de Microsoft, une plateforme d'analyse de données massives qui permet (selon Microsoft) de prendre de meilleures décisions grâce aux données (« *make better decisions with data* ») – décisions prises du point de vue de l'exécutif, dans une démarche verticale (*top-down*). Dans la documentation et les tutoriels fournis aux professionnel.les pour se familiariser avec cet outil, Microsoft prescrit aux organisations de développer la « maturité » de leur propre culture de données³².

Un des indicateurs de cette maturité serait la capacité à prendre et justifier les décisions à partir de preuves fondées sur l'analyse de données et non en fonction d'un « savoir tribal non documenté », d'intuitions ou d'une expérience subjective, selon les termes de Microsoft³³. L'intérêt pour Microsoft de voir adopter une culture de données gestionnaire est bien évidemment de développer la commercialisation de sa plateforme.

D'autre part, le cadrage de la culture de données comme mode de gestion organisationnelle s'attache à un imaginaire de la plateformesation comme branchement de flux de données, dont il conviendrait de fluidifier la circulation.

³¹ Marie GRÉGOIRE, « Message de la présidente directrice générale », in BAnQ, *Rapport annuel d'activités 2022-2023*, Montréal, BAnQ, 2023, p. 14.

³² MICROSOFT, *Microsoft Fabric Adoption Roadmap: Data Culture*, learn.microsoft.com.

³³ « *Undocumented tribal knowledge* » ; « *on intuition and subjective experience* » ; *Ibid.*

À BANQ, l'une des ambitions du chantier sur les données était d'organiser les relations entre les « détenteurs principaux des données »³⁴ (selon les termes utilisés à BANQ) : c'est-à-dire les employé.es et cadres qui produisent au quotidien ces données dans la cadre de leur activité, et qui exercent un contrôle sur l'accès à ces données (ici le terme de détenteur qualifie le fait éventuellement, de pouvoir « retenir » des données). À l'échelle de l'institution, il s'agissait donc de faire tomber les barrières à la circulation des données ; ou pour reprendre les mots de la PDG d'« d'abattre les frontières entre les unités et les services de l'organisation »³⁵ pour permettre de brancher les services entre eux.

À l'échelle de ses relations avec le gouvernement, BANQ est aussi prise dans une injonction de se connecter, par la transmission de données, à un appareil gouvernemental qui s' imagine comme une « administration plateforme ». Un rapport de l'OCDE (2020) destiné aux gestionnaires d'administration publique préconise ainsi d'adopter « un modèle de secteur public axé sur les données » qui nécessite « la mise au point d'outils liés aux données communs mais flexibles (comme des plateformes de partage des données) »³⁶. Il faudrait pour cela dépasser « la fragmentation » liée à « l'identité de l'entité qui produit et administre les données » présentée comme une « difficulté héritées du passé » car elle produit des « cloisonnements », des « entrave[s] » à « l'instauration d'une administration intégrée et connectée »³⁷. Dans cette perspective, la culture des données est donc présentée comme un vecteur de modernisation de l'État.

Depuis 2021, BANQ est ainsi soumise à une nouvelle *Loi sur la gouvernance et la gestion des ressources informationnelles des organismes publics et des entreprises du gouvernement*. Cette loi établit que certaines données des établissements publics « constituent un actif informationnel stratégique du patrimoine numérique gouvernemental » dont il s'agit d'instaurer la « mobilité » et la « valorisation » (deux notions introduites par la loi) car il est « d'intérêt gouvernemental » pour l'administration

³⁴ « La direction responsable de l'activité ou du processus d'affaires auxquels se rattachent les données » ; BANQ, *Politique sur le partage et la valorisation des données institutionnelles*, Montréal, BANQ, 2022, p. 4.

³⁵ Marie GRÉGOIRE, « Message de la présidente directrice générale », *op. cit.*, p. 14.

³⁶ OCDE, *Axer le secteur public sur les données : marche à suivre*, Paris, Éditions de l'OCDE, 2020, p. 9, 26.

³⁷ *Ibid.*

et la prestation de service public³⁸. Il est donc exigé de BAnQ, en tant que « source officielle de données numériques gouvernementales » qu'elle transmette des rapports de données au gouvernement.

Ces directives de remontée des données vers un centre d'analyse gouvernemental – ou paragouvernemental – s'étendent, de plus en plus, en dehors du cadre des ministères et des établissements publics. Dans le secteur culturel, ces directives prennent la forme de plateformes de mutualisation de données. La Grande-Bretagne a été pionnière en la matière avec la création en 2012 de l'agence The Audience Agency, qui a créé l'outil Audience Finder. Il s'agit d'une plateforme de centralisation des données sur la fréquentation de plus de 800 organismes culturels. Ces données sont couplées avec d'autres données statistiques sur les populations de différents territoires (données sociodémographiques et des données sur le déplacement domicile-travail). En pratique, il s'agit d'un outil de marketing de ciblage de segments de publics qui fournit des recommandations aux organismes culturels pour mener des campagnes de promotion de leur offre, en particulier sur les médias sociaux.

En France, on peut citer le cas de SiBILL (pour Système d'Information BILletterie) un projet mis en place en 2018 par le ministère de la culture. Il s'agit d'un dispositif légal de remontée et de centralisation obligatoire (sous peine d'amende) des données de billetterie qui vise 15 000 structures de diffusion du spectacle vivant. Cette plateforme, qui était encore présentée en 2020 par le Ministère comme une « révolution statistique »³⁹ est à ce jour un échec, avec seulement un quart de structures ayant transmis leurs données en 2023. Des sources proches du dossier citent un défaut de conception de l'outil, d'autres une « tradition d'opacité » des structures de diffusion ou encore une crainte du « flicage » de la part du Ministère⁴⁰. Quoi qu'il en soit, cet exemple montre que dans les rouages d'un État conçu comme machine cybernétique, la mécanique des flux se grippe fréquemment à l'épreuve de la réalité.

³⁸ Conseil du Trésor du Québec, *Loi sur la gouvernance et la gestion des ressources informationnelles des organismes publics et des entreprises du gouvernement*, Québec, Conseil du Trésor, 2023, p. 11.

³⁹ Sandrine BLANCHARD, « Avec Sibil, le ministère de la culture compte obtenir une radiographie du spectacle vivant », *Le Monde*, 17 février 2020.

⁴⁰ Sandrine BLANCHARD, « Spectacle vivant : la fréquentation difficile à mesurer quatre ans après la création du système de billetterie Sibil », *Le Monde*, 28 octobre 2022.

Pour conclure, à travers l'exemple de BAnQ, j'ai voulu montrer comment les logiques de mise en données et de plateformes qui traversent le secteur culturel contribuent à infléchir les cadrages de la culture de données, mais que ces inflexions trouvent une résonance variable en fonction des milieux de pratique qui sont façonnés par une écologie très diverse et hétérogène de pratiques et d'infrastructures de données.