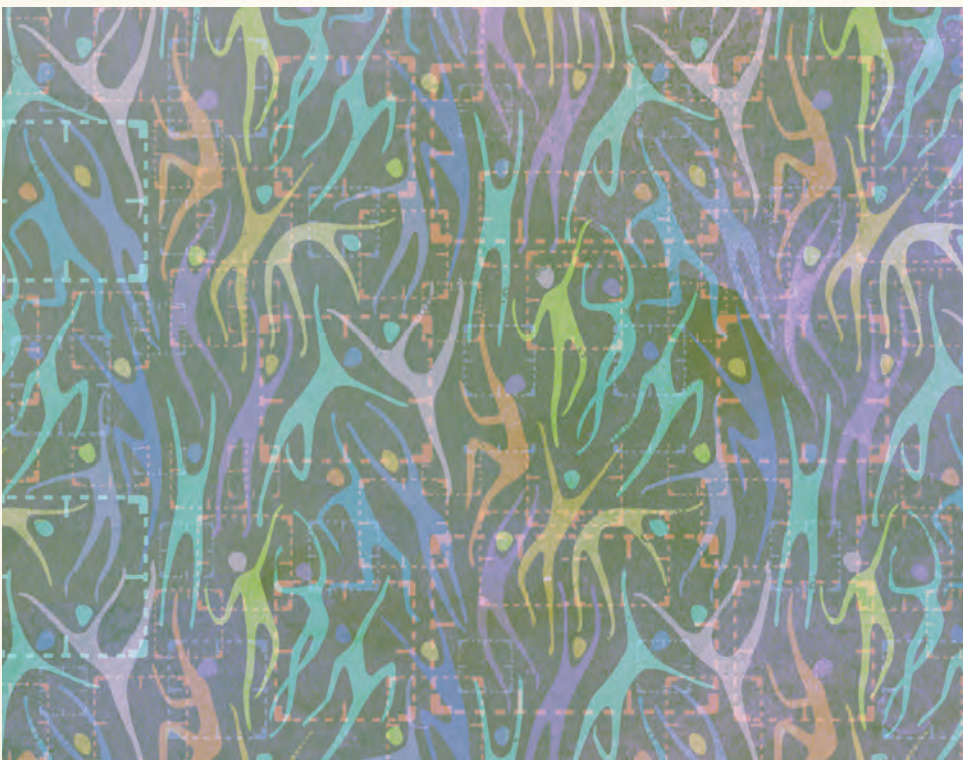




# **Le mouvement Wikimédia au Canada**

Communautés, institutions et culture libre

Sous la direction de Jean-Michel Lapointe et Marie D. Martel





La collection «**Parcours numériques**» est accessible gratuitement en édition augmentée sur [parcoursnumeriques-pum.ca](http://parcoursnumeriques-pum.ca).

Ce livre examine la participation des Canadiens et des Canadiennes aux plateformes Wikimédia en illustrant comment les idéaux de partage, d'ouverture et de collaboration se manifestent dans un pays marqué par sa diversité linguistique, culturelle et régionale. Il explore les enjeux à la fois éditoriaux et identitaires auxquels sont confrontés les internautes québécois, acadiens, atikamekw et anglo-canadiens qui cherchent à inscrire les spécificités de leur culture dans cet espace numérique mondialisé. L'ouvrage aborde également les motivations des institutions canadiennes qui collaborent avec ce mouvement international de bénévoles afin de rendre visibles et réutilisables les savoirs qui leur importent. Ce portrait inédit du Canada témoigne de la contribution inestimable d'un mouvement d'amateurs qui bouscule notre compréhension de la connaissance, démontrant que celle-ci ne relève pas exclusivement de l'expertise professionnelle.

**Jean-Michel Lapointe** est chargé de projets pédagonumériques au Carrefour d'innovation et de pédagogie universitaire de l'Université du Québec à Montréal.

**Marie D. Martel** est professeure agrégée à l'École de bibliothéconomie et des sciences de l'information de l'Université de Montréal.



24,95\$ • 20€

ISBN 978-2-7606-5109-8



## **LE MOUVEMENT WIKIMÉDIA AU CANADA**

La collection « **Parcours numériques** » est accessible gratuitement en édition augmentée sur **parcoursnumeriques-pum.ca**. La version enrichie comprend une bibliographie, des notes ainsi que de nombreux contenus complémentaires (vidéos, illustrations, etc.). La collection est dirigée par Michaël E. Sinatra et Marcello Vitali-Rosati.

# **Le mouvement Wikimédia au Canada**

Communautés,  
institutions  
et culture libre

SOUS LA DIRECTION  
DE JEAN-MICHEL LAPOINTE  
ET MARIE D. MARTEL



Les Presses de l'Université de Montréal

Mise en pages: Folio infographie

*Catalogage avant publication de Bibliothèque et Archives nationales  
du Québec et Bibliothèque et Archives Canada*

Titre: Le mouvement Wikimedia au Canada: communautés, institutions et culture libre / sous la direction de Jean-Michel Lapointe et Marie D. Martel.

Noms: Lapointe, Jean-Michel, éditeur intellectuel. | Martel, Marie D., éditeur intellectuel.

Collections: Parcours numériques.

Description: Mention de collection: Parcours numériques | Comprend des références bibliographiques.

Identifiants: Canadiana (livre imprimé) 20250023040 | Canadiana (livre numérique) 20250023059 | ISBN 9782760651098 | ISBN 9782760651104 (PDF) | ISBN 9782760651111 (EPUB)

Vedettes-matière: RVM: Wikipedia. | RVM: Wikidata. | RVM: Encyclopédies électroniques—Canada. | RVM: Bases de données en ligne—Canada.

Classification: LCC Z674.75.W55 M68 2025 | CDD 030.971—dc25

Dépôt légal: 1<sup>er</sup> trimestre 2025

Bibliothèque et Archives nationales du Québec

© Les Presses de l'Université de Montréal, 2025

Les Presses de l'Université de Montréal remercient de leur soutien financier le Fonds du livre du Canada, le Conseil des arts du Canada et la Société de développement des entreprises culturelles du Québec (SODEC).



Conseil des Arts  
du Canada

Financé par le gouvernement du Canada

Canada

SODEC

Québec

IMPRIMÉ AU CANADA

## **Introduction**

JEAN-MICHEL LAPOINTE ET MARIE D. MARTEL

Il existe une littérature abondante sur Wikipédia et la culture libre, mais rarement s'est-on attaché à décrire et à comprendre la participation spécifique de la population d'un pays aux plateformes Wikimedia. Ce livre a pour ambition d'examiner les initiatives wikimédiennes au Canada, en présentant des études qui éclairent la manière dont les idéaux de partage, d'accès ouvert et de collaboration, piliers de la culture libre, se manifestent au nord du 49<sup>e</sup> parallèle des Amériques.

Les questions pourraient s'énoncer en ces termes : comment les Canadiennes et les Canadiens participent-ils aux plateformes Wikimedia, et quels sont les enjeux et les dynamiques de cette participation ? Cette introduction souhaite déjà poser quelques repères et jalons concernant cette interrogation et pointer quelques-unes des limites qui la contraignent. D'emblée, il faut reconnaître que le Canada, traversé par des régionalismes complexes, des divisions linguistiques, ethnoculturelles et marqué par la présence historique des premiers peuples, rend périlleuse la production d'un portrait totalisant, quel que soit le thème. Pays

dont l'identité et l'histoire gagnent à être appréhendées au pluriel (Wright, 2020)<sup>1</sup>, le Canada est un miroir troublant du mouvement Wikimedia lui-même, où chaque point du réseau de personnes connectées à la plateforme peut ajouter sa touche aux projets, en fonction de ses intérêts de contribution, si bien que l'encyclopédie Wikipédia a déjà été envisagée comme un vaste territoire dynamique où différents projets et communautés se déploient sans qu'il soit possible de capturer une vue de l'ensemble des activités et des personnes qui y prennent part (McDowell et Vetter, 2021, p. 77). La multiplication des perspectives par les auteurs et autrices de ce collectif tend à refléter cette complexité croisée des territoires canadien et wikimédien.

Il semble, en outre, que les politiques de collecte de données sur les personnes qui contribuent aux différents projets ne permettent pas l'identification de leur pays de résidence. La culture de la communauté repose largement sur le respect de l'anonymat et ne requiert pas d'établir un lien avec l'identité civile (Konieczny, 2023). En conséquence, on ne peut guère compter sur cette avenue pour caractériser de manière précise la population canadienne contributrice. Cette absence de données particulières sur le profil des éditrices et éditeurs canadiens amène d'autres défis pour comprendre pleinement les dynamiques de participation qui nous intéressent. Cela dit, certains jeux de données anonymisés mis à disposition par l'équipe technologique de la Fondation Wikimedia permettent de géolocaliser les contri-

---

1. On trouvera toute la documentation sur les références ainsi qu'une bibliographie complète dans la version enrichie en ligne sur [parcoursnumeriques-pum.ca](https://parcoursnumeriques-pum.ca).

butions faites à Wikipédia, le projet le plus actif du mouvement. Depuis 2021, les versions linguistiques les plus actives au Canada sont, sans surprise, celles en anglais (entre 270 et 350 personnes actives mensuellement) et celles en français (entre 30 à 60 personnes actives mensuellement), mais l'on compte également une participation constante quoique de moindre envergure aux versions en farsi, en hébreu, en russe, en serbe, en ukrainien et en mandarin<sup>2</sup>, ce qui donne à penser que les personnes canadiennes issues de l'immigration maintiennent un lien avec leur culture et leur langue d'origine par la contribution à l'encyclopédie libre.

Il est aussi possible d'envisager ce questionnement à partir du récit de la naissance et de la croissance du « chapitre » canadien de Wikimedia. Cette piste permet de lever le voile sur les motivations et les défis rencontrés dès les premières initiatives et offre un point d'entrée crucial en vue de proposer quelques réponses; cette avenue souligne également l'importance des initiatives locales et des chapitres nationaux comme Wikimedia Canada, permettant d'accéder à un éclairage précieux sur les contributions régionales et les enjeux spécifiques.

L'organisme à but non lucratif Wikimedia Canada est officiellement reconnu depuis 2011 par la Fondation Wikimedia, l'organisation mère du mouvement. Créée en 2003 pour appuyer la contribution des bénévoles en administrant,

---

2. Données colligées des mois de janvier et juin pour les années 2021, 2022 et 2023 des *dumps* de géolocalisation : <https://dumps.wikimedia.org/other/geoeditors/>. Précisons que la participation constante des versions linguistiques autres que le français et l'anglais signifie plus de 100 modifications par mois faites par un nombre indéterminé de personnes allant de 1 à 10.

en développant et en maintenant l'infrastructure technologique de l'ensemble des projets gravitant autour de Wikipédia (fondé en 2001), la Fondation est dépositaire de la marque Wikimedia. Qui plus est, la Fondation soutient financièrement 37 chapitres à travers le monde, c'est-à-dire des organisations qui ont pour objectif de soutenir et stimuler le développement des projets Wikimedia sur leur territoire.

Du côté canadien, à partir de 2005, une poignée de bénévoles de l'Ontario, du Québec, de Colombie-Britannique et de l'Alberta entament les démarches en vue d'échafauder une structure qui permettrait de faire connaître et d'encourager la contribution à Wikipédia au pays. Ces personnes partagent également les idéaux du mouvement social de la culture libre à l'origine de Wikipédia qui les pousse à promouvoir la libre circulation des savoirs et de l'information afin qu'ils puissent être légalement copiés et redistribués. Ainsi, en vertu de sa licence libre, tout le contenu de Wikipédia peut être utilisé, modifié et partagé, pour autant que l'on cite sa provenance et que son caractère libre soit préservé.

Dès le début de leurs efforts collectifs, les militants qui poseront les bases organisationnelles de ce qui deviendra Wikimedia Canada demandent que soit aboli l'article 12 de la *Loi sur le droit d'auteur* qui traite de ce qu'on appelle le droit d'auteur de la Couronne. Cet article de loi, en vigueur au Canada depuis 1921 et commun à plusieurs pays du Commonwealth, stipule que les documents produits par l'appareil gouvernemental fédéral doivent attendre 50 ans avant d'entrer dans le domaine public, ce qui a pour effet d'empêcher des contenus financés par la population d'être réutilisés dans Wikipédia et les autres projets du mouve-

ment. À titre d'exemple, les photographies officielles des élus au Parlement canadien ne peuvent être utilisées pour illustrer les articles de l'encyclopédie, car elles appartiennent à Sa Majesté. Cette loi contraste avec celle en vigueur aux États-Unis, où tous les documents produits par le gouvernement central sont d'emblée placés dans le domaine public et peuvent être librement réutilisés. L'abolition ou la réforme de cet article de loi, qui est aussi demandée par les principales associations d'archivistes, de bibliothécaires, de chercheurs et d'enseignants du Canada, permettrait de préserver et de mettre en valeur les documents produits par les institutions publiques fédérales (Wakaruk, 2020).

Les actions des premiers bénévoles visant à créer une organisation canadienne qui serait reconnue par la fondation américaine, un labeur de longue haleine, se sont concrétisées à l'été 2011. Lorsque Wikimedia Canada (WMCA) devient officiellement affiliée à la Fondation Wikimedia, c'est le médecin James Heilman qui assure alors la première présidence de l'organisme. Depuis, WMCA est tenu à bout de bras par des bénévoles qui cherchent à stimuler le développement des contenus canadiens des projets en multipliant les occasions de rencontre et les événements de contribution.

Emblématique de cette approche, la première activité publique officielle du chapitre a lieu le 28 août 2011. Il s'agit d'une chasse photographique à travers la ville de Montréal dont l'objectif est de documenter plus de 375 lieux d'intérêts historiques afin de pouvoir illustrer avec des images sous licences libres les articles encyclopédiques qui leur sont consacrés. Nommée Wikipedia Takes Montréal, cette activité

est marquante à plus d'un titre. D'une part, elle vise à souligner le dixième anniversaire de Wikipédia et, d'autre part, à stimuler une participation inédite à la production de communs canadiens dans le paysage informationnel du pays. Cette activité parvient à attirer plus de 200 personnes grâce à une couverture médiatique pancanadienne, en français et en anglais. L'initiative locale comporte également une dimension internationale : préparée en amont avec l'administration municipale de la Ville de Montréal – une collaboration avec une institution publique qui deviendra un *modus operandi* pour WMCA –, elle bénéficie du soutien logistique d'un autre chapitre du mouvement, Wikimedia France. Cette activité connaît un aléa aussi imprévu que mémorable, alors que les restes de l'ouragan Irène s'invitent à Montréal le jour même, compromettant la qualité de la majorité des milliers de clichés. Cet échec relatif, loin de démotiver les organisateurs, jouit d'un certain retentissement au sein du mouvement.

S'il est une étape clé pour la reconnaissance du chapitre canadien du mouvement auprès des institutions du pays, en particulier celles relevant des GLAM (acronyme anglais pour désigner les « *Galleries, Libraries, Archives, Museums* », soit en français, les galeries, les bibliothèques, les archives et les musées), c'est l'établissement, à compter de 2013, d'une entente de partenariat avec Bibliothèque et Archives nationales du Québec (BAnQ). Véritable « partenariat public-commun », cette entente est toujours en vigueur une décennie plus tard et s'avère profitable aux deux parties, voire au-delà. Un partenariat public-commun (*public-commons partnership*), tel que décrit par David Bollier et Silke Helfrich (2022), est une forme de collaboration entre les entités publiques (gouvernements et institutions publiques)

et les communs (ressources gérées collectivement par une communauté suivant des règles établies par celle-ci); et, dans ce cas, les contenus wikimédiens créés à partir des ressources documentaires institutionnelles sont plus précisément conçus comme des communs de la connaissance. D'un côté, WMCA et le mouvement de bénévoles qu'il cherche à faire croître peuvent compter gracieusement sur un lieu physique de convergence et de socialisation hors ligne pour la communauté montréalaise par des soirées de formation intitulées – suivant une formule éprouvée dans la francophonie – Mardi, c'est Wiki. De l'autre, BANQ encourage les archivistes et les bibliothécaires de cette institution présente partout sur le territoire du Québec à prolonger leur expertise documentaire en valorisant les collections qu'ils et elles détiennent au sein des plateformes Wikimedia, et ce, en mettant à contribution leurs publics, à qui l'on propose de se former à la contribution aux projets Wikimedia (Boudreau *et al.*, 2016). Le bibliothécaire en chef du Canada, Guy Berthiaume, qui était à la tête de BANQ lorsque le partenariat a été établi, puis qui a répliqué celui-ci dans une certaine mesure lors de son passage à Bibliothèque et Archives Canada (BAC) à compter de 2017, avait ceci à dire dans son discours d'ouverture à la Conférence de l'Association des musées de l'Ontario cette même année:

Alors que les musées, les bibliothèques et les archives ont les connaissances et les documents, Wikipédia a une portée qu'aucune de nos institutions ne pourrait jamais atteindre à elle seule. [...]

Nous travaillons avec Wikimedia Canada, une société sans but lucratif qui vise à accroître le contenu canadien dans Wikipédia. L'idée était que le renforcement de notre présence

sur Wikipédia pourrait attirer de nouveaux utilisateurs vers notre site Web et nos collections. BAC a maintenant plus de 3 000 images intégrées dans des articles publiés sur Wikipédia. En retour, cela se traduit par environ 30 millions de consultations de documents de BAC, chaque mois (Berthiaume, 2017).

Les partenariats entre WMCA et les institutions de mémoire agissent comme catalyseur pour « wikifier » l'écosystème institutionnel non marchand du pays. Musées, sociétés savantes, universités, bibliothèques publiques et de recherche, organismes culturels et communautaires ont été nombreux à accueillir des activités de contribution et de formation, ou encore à intégrer une composante wikimédienne à leur stratégie organisationnelle. Cela passe parfois par l'embauche ponctuelle d'un ou une wikipédiste en résidence, bien que souvent l'expertise wikimédienne du personnel se déploie dans l'ombre, sans qu'il y ait une reconnaissance institutionnelle du travail contributif – une réalité fréquente au sein du mouvement Wikimedia (Stintson et Evans, 2018).

Mais avant même le partenariat avec BANQ, les premières interactions avec les bibliothèques se sont déroulées dans le cadre d'activités contributives publiques. La première de ces initiatives a eu lieu le 18 février 2012, à la bibliothèque de l'Université Laval, avec un atelier consacré à Jean Talon, premier intendant de la Nouvelle-France. Une année plus tard, le 6 avril 2013, la Bibliothèque Mile End (désormais appelée Bibliothèque Mordecai-Richler) accueille la première activité contributive dans une bibliothèque publique au Québec et, sous toutes réserves, au Canada. Cet événement faisait partie du Projet Mile End, lancé en février par Mémoire du Mile End, un organisme d'histoire locale, en

collaboration avec Wikimedia Canada. À Montréal, les bibliothèques publiques vont intégrer des ateliers de contribution à Wikipédia dans leur programmation régulière. Entre 2016 et 2018, plus d'une vingtaine d'ateliers se succèdent; la plupart d'entre elles sont organisées en partenariat avec le Café des savoirs libres, un collectif réunissant des bibliothécaires et des défenseurs de la culture libre, engagés dans la création de communs de la connaissance (Martel, 2021). La configuration des partenariats à partir du site des bibliothèques de proximité permet de réunir des acteurs issus des institutions publiques, mais aussi de la société civile (société d'histoire et groupes communautaires) en créant des réseaux wikimédiens situés localement. Dans la foulée, plusieurs ateliers ont également été mis en place dans le but de valoriser les communs du cinéma à la Cinémathèque québécoise, en présence même de créateurs et créatrices, figures clés du cinéma d'animation canadien.

Plusieurs de ces institutions ont orienté leurs activités wikimédiennes autour de la lutte contre les biais systémiques, la réduction du fossé des genres, la mise en visibilité de certaines communautés minorisées ou groupes marginalisés en réponse aux politiques publiques de réconciliation-équité-diversité-inclusion (RÉDI) qui sont désormais présentes dans la plupart d'entre elles au Canada.

En ce qui a trait à l'intégration du mouvement dans l'éducation supérieure, il importe de souligner les multiples rôles joués également par les bibliothécaires du pays pour ouvrir et baliser le chemin permettant d'établir des ponts avec un milieu universitaire d'abord méfiant vis-à-vis du potentiel jugé disruptif de Wikipédia auprès des usages entourant la documentation établis dans l'environnement numérique.

L'apport du milieu universitaire au mouvement est jugé à ce point central au sein de WMCA que les personnes qui gravitent autour du chapitre ont pris l'habitude de bonifier l'acronyme GLAM en y ajoutant un U à la fin. Benoit Rochon, président de WMCA de 2016 à 2019, raconte l'origine de l'expression GLAMU en ces termes : « Pour la petite histoire, j'avais écrit sur une diapositive "We will GLAM•U, universities" (à prononcer en anglais en référence à *We Will Rock You* de Queen) lors d'une conférence donnée en 2011 au Musée du champignon à Amsterdam, où il y avait dans l'assistance quelques professeurs d'université des Pays-Bas. » Ha-Loan Phan, bénévole depuis 2012 et présidente de WMCA de 2022 à 2024, explique les raisons pour lesquelles elle s'est attachée à populariser l'expression par la suite : « J'avais appris à contribuer toute seule pour publier des informations sur l'opossum après ma maîtrise consacrée à cet animal et je trouvais que la grande oubliée des GLAM était les U[niversités] alors que ce sont des lieux de production et de circulation du savoir. Les bibliothécaires universitaires ont toujours été des alliés privilégiés pour nos ateliers de formation : spécialistes des sources, ils donnaient l'assurance de leur expertise, permettant ainsi de rassurer les professeurs ! De plus, les bibliothèques universitaires sont des lieux parfaits pour héberger ce genre d'activité "parascolaire", avec les sources à portée de main » (Rochon et Phan, échange par courriels, 26 mai 2023).

Autre moment fort de l'histoire du mouvement canadien, la tenue en 2017 de la convention annuelle internationale Wikimania au centre-ville de Montréal a permis d'attirer de nouveaux adhérents et d'approfondir l'enracinement du chapitre canadien dans la métropole québécoise. Ce fut

l'occasion pour les wikimédiens et wikimédiennes des quatre coins du globe de découvrir la contribution du Canada au mouvement, mais aussi la dualité linguistique du pays et d'apprécier l'importance de ses cultures autochtones. À cet égard, Wikimania 2017 a été l'occasion de célébrer l'arrivée plus tôt cette année-là de la mise en service public de Wikipetcia, la version de l'encyclopédie en langue atikamekw nehiromowin.

Il faut aussi souligner la présence de l'autochtonie canadienne dans Wikipédia, qui permet de mettre en relief le rôle et les enjeux sous-jacents de la participation à l'encyclopédie pour œuvrer à la diversification et à la neutralité de l'information disponible en ligne. En effet, autant du point de vue des personnes qui peuvent participer à façonner les savoirs, mais aussi en raison du fait que ces derniers peuvent être exprimés dans une diversité de langues, le mouvement Wikimedia permet d'inscrire la pluralité des réalités canadiennes vécues au sein d'une des plateformes les plus visitées du Web. Cela dit, à ce jour, il n'existe que trois versions linguistiques autochtones: inuktitut, cri et atikamekw nehiromowin, cette dernière étant d'ailleurs fort dynamique. Les membres de la nation Atikamekw Nehirowisiwok, soit environ 8000 personnes réparties dans trois communautés dans le centre du Québec, sont toujours les locuteurs de leur langue ancestrale. De sa création en 2013 jusqu'à sa mise en service en 2017, cette version linguistique de Wikipédia a permis à une culture de tradition orale de s'appropriier et d'adapter à ses besoins une forme textuelle occidentale: l'encyclopédie. Cette présence numérique offre un moyen de stimuler la revitalisation de leur langue, en particulier par un dispositif de transmission intergénérationnelle entre les personnes

aînées qui détiennent les savoirs dont la langue est porteuse et les plus jeunes qui gravitent dans l'univers technologique contemporain. Ce projet de partage encyclopédique a également été l'occasion de redéfinir ce que signifie pour eux le principe d'ouverture cher au mouvement, ce qui les a amenés à exclure de leur Wikipédia les savoirs dits sacrés (Casemajor *et al.*, 2018). La longue histoire de dépossession qui structure les relations entre les colons d'origine européenne et les Autochtones du Canada explique le scepticisme légitime que ceux-ci peuvent avoir vis-à-vis de cet idéal d'ouverture. En tant que lieux de savoir foncièrement occidentaux, la science, les médias et Wikipédia ne sont pas reconnus par les Autochtones comme étant favorables à une véritable prise en compte de leurs perspectives sur le monde, mais Wikipédia se distingue néanmoins en ce qu'elle est de plus en plus considérée comme un espace de production de connaissances qui, en raison de son mode de fonctionnement décentralisé, peut être investi par les Autochtones du pays pour réécrire l'histoire d'une manière plus neutre (Lugosi *et al.*, 2022). Le chapitre canadien du mouvement Wikimedia se fait d'ailleurs le promoteur d'une telle approche et souhaite stimuler le développement de nouvelles versions linguistiques autochtones au cours de la prochaine décennie.

La tenue de Wikimania 2017 dans la métropole québécoise marque un tournant décisif pour WMCA, consolidant sa présence à Montréal et au sein de la culture francophone, malgré la prédominance de l'anglais au Canada. Cette particularité, reflétée dans le contenu de ce livre, rédigé majoritairement en français, s'explique non seulement par l'engagement vigoureux des institutions documentaires au Québec dans les projets wikimédiens, mais aussi par le

dévouement sans faille d'un bénévole. Travaillant dans l'ombre, ce dernier a joué un rôle crucial dans l'implantation et la croissance du chapitre canadien du mouvement. Benoit Rochon, contributeur à Wikipédia depuis 2003, membre fondateur de WMCA, mais aussi de WikiFranca – un organisme fondé en 2012 qui vise à faciliter la collaboration entre les bénévoles du mouvement de l'ensemble de la francophonie –, est l'instigateur des partenariats avec BANQ et BAC, tout comme d'un nombre incalculable d'activités de contribution un peu partout au pays et dont bon nombre dans la région montréalaise, incluant l'organisation de Wikimania à Montréal. Bien que la culture collaborative aime à faire valoir l'intelligence collective comme un processus où l'anonymat règne, rappelons que la contribution entre les personnes est inégalement répartie et obéit à une loi de puissance qui fait en sorte qu'une majorité de personnes participe peu, tandis qu'une infime minorité de personnes est responsable de la vaste partie des contributions (Shirky, 2008). Benoit Rochon appartient à cette minorité de personnes qui ont joué un rôle critique pour activer la collaboration et le développement de projets à portée locale, provinciale, pancanadienne et internationale au sein du mouvement Wikimédia.

En prenant la présidence de WMCA à partir de 2016, Benoit Rochon cherche à s'entourer de bénévoles au conseil d'administration qui, comme lui, ont une force de travail capable d'amener WMCA plus loin, en développant de nouveaux partenariats, mais aussi de nouvelles sources de financement. Leurs efforts sont couronnés de succès, mais la charge de travail repose exclusivement sur les épaules d'une équipe de bénévoles, ce qui exige de développer une

stratégie visant à professionnaliser l'organisation du chapitre en la dotant d'un personnel salarié capable de mener de front les dossiers requérant une attention soutenue, notamment lorsque l'on travaille avec des institutions. En comparaison avec les chapitres français ou allemand du mouvement, qui comptent respectivement plus de 12 et 90 salariés, la force de travail de Wikimedia Canada repose entièrement sur le temps libre que parviennent à dégager ses adhérents et les membres du conseil d'administration, accompagnés de mandats ponctuels accordés ici et là à des contractants. Cette manière de faire, qui a eu cours depuis la fondation de WMCA, commence à changer à partir de 2022. Le chapitre canadien se transforme profondément en mettant sur pied une première équipe opérationnelle, dont l'embauche du premier directeur général de son histoire, Louis Germain, qui était jusqu'alors directeur général de l'Association des archivistes du Québec. Signe d'une maturité croissante ou d'un désir d'exercer un dynamisme nouveau pour stimuler les activités du mouvement au Canada, cette nouvelle page est en train de s'écrire. La nouvelle équipe a relevé le défi de l'organisation de la Wikiconvention North America, en novembre 2023, à la Toronto Public Library, en réunissant les fervents et ferventes de tout le continent pour échanger des idées, renforcer les collaborations et célébrer le mouvement tout en s'interrogeant, entre autres, sur la question cruciale des impacts de l'intelligence artificielle sur les projets wikimédiens, la culture et les savoirs libres, mais aussi sur la démocratie considérant les risques de la manipulation de l'information, de la surveillance de masse, des biais algorithmiques, du microciblage politique et de la concentration du pouvoir.

Parmi les autres défis qui continueront d'aiguiller à l'avenir la culture libre canadienne, en particulier au sein du mouvement Wikimedia, relevons la nécessité d'élargir et de renouveler les communautés contributrices aux projets au-delà des cercles historiquement associés au mouvement (informaticiens libristes, bibliothécaires, archivistes, éducatrices et chercheurs). La pandémie a augmenté l'isolement des gens et a favorisé l'émergence de pratiques de socialisation à distance qui rendent plus difficile le développement de liens sociaux forts. Dans ce contexte, les activités de contribution et de réseautage en personne qui ont été importantes pour structurer et faire émerger des communautés wikimédiennes locales dans diverses villes canadiennes doivent reprendre vie. Qui plus est, l'évolution du rôle des institutions canadiennes partenaires du mouvement doit faire l'objet d'une attention particulière, car aucune pérennité n'est acquise en raison des changements au sein du personnel, en particulier des équipes de direction qui définissent les priorités organisationnelles.

Signalons aussi en terminant trois enjeux-phares qui sont susceptibles d'avoir un impact sur la vitalité du mouvement au Canada au cours des prochaines années. Dans un premier temps, le retour en force d'un souci pour l'éducation ouverte, les ressources et les pratiques éducatives libres, directement alignées avec les principes du mouvement Wikimedia, font l'objet d'un intérêt marqué au sein des ministères de l'Éducation des gouvernements provinciaux, mais aussi d'un plaidoyer renouvelé au sein des associations de bibliothécaires (ABRC, 2023). Ensuite, le fait que la science ouverte est désormais une politique publique au niveau fédéral et des organismes subventionnaires, ce qui a

le potentiel de transformer les pratiques des chercheurs et, plus globalement, de l'écosystème de la publication savante en le rapprochant d'un modèle de gouvernance auto-organisé similaire à celui du mouvement. Enfin, rappelons que le contexte légal en vigueur au Canada pose des limites importantes à la réutilisation des œuvres, qu'elles soient produites au sein du gouvernement fédéral tombant sous le droit d'auteur de la Couronne ou non, mais aussi en raison du recul du domaine public canadien depuis le 1<sup>er</sup> janvier 2023, qui est passé de 50 ans à 70 ans après la mort de l'auteur.

Ainsi, d'autres critiques internes qui s'inquiètent d'un possible épuisement du mouvement ouvert réclament non seulement l'apport d'autres voix, mais aussi d'autres narrations, comme celle des « communs » pour le revitaliser. Le mouvement Wikimedia et les concepts liés aux communs convergent en termes d'engagement participatif et collaboratif, de libre accès à la connaissance, de valorisation de la gestion collective des ressources.

\*\*\*

L'organisation du présent ouvrage suit trois grands axes thématiques d'approfondissement qui relient la majorité des textes et que nous avons regroupés en trois grandes sections en fonction de l'importance que leur accordent leurs auteurs et autrices: 1) les enjeux identitaires; 2) la collaboration avec les institutions; et 3) les formes de littérature que l'on développe en prenant part ou en interagissant avec les projets.

On constate que Wikipédia demeure l'objet d'investigation principale des analyses, suivi de près par Wikidata, le plus récent projet du mouvement, qui joue désormais un rôle central au sein de l'écosystème Wikimedia et au-delà; finalement, deux textes s'intéressent également à l'utilisation de la banque d'images Wikimedia Commons. Notons également que toutes les personnes qui signent un texte dans le présent ouvrage ont une pratique de contribution liée aux projets sur lesquels ils et elles élaborent un savoir.

Pour donner le coup d'envoi au premier axe thématique, qui interroge les enjeux identitaires au sein des plateformes de l'écosystème Wikimedia, Nathalie Casemajor analyse la cohabitation interculturelle qui a lieu au sein de trois versions linguistiques de Wikipédia: anglaise, française et atikamekw nehiromowin. Elle conceptualise les protocoles de pluralisation que ces trois encyclopédies ont élaborés au fil du temps pour faire exister les particularismes linguistiques et culturels canadiens dans ces espaces numériques mondialisés. Gabriel Arsenault et Mathieu Wade s'intéressent quant à eux au portrait qui se dégage de la nation acadienne dans la version française de Wikipédia. Leur propos est guidé par deux interrogations. D'une part, l'Acadie est-elle un territoire circonscrit aux provinces maritimes du Canada, ou plutôt une diaspora sans frontières? D'autre part, le peuple acadien est-il une réalité purement livresque appartenant au seul passé ou, au contraire, est-ce un sujet collectif vivant qui n'a pas dit son dernier mot? Cette exploration savante des espaces-temps acadiens débouche sur un appel à faire exister l'Acadie d'hier à aujourd'hui dans l'encyclopédie libre. Enfin, le texte de Marie D. Martel et Simon Villeneuve s'appuie sur les données massives tirées de Wikidata afin d'évaluer la

représentation des personnes s'identifiant comme femmes et hommes ainsi que le fossé des genres dans les articles biographiques de plus de 80 dictionnaires et encyclopédies de référence provenant d'une grande diversité de pays, du Canada à la Chine en passant par l'Allemagne et la Nouvelle-Zélande. Ce binôme démontre qu'il existe une sous-représentation majeure des personnes s'identifiant comme femmes dans tous ces ouvrages, y compris dans les publications canadiennes identifiées et – quoique dans une moindre mesure – dans les ressources biographiques du mouvement Wikimedia.

Le second axe thématique aborde l'intégration des projets Wikimedia au sein des institutions canadiennes, un sujet que nous avons commencé à décrire plus haut dans cette introduction. Ces collaborations donnent lieu à des enrichissements mutuels, mais exigent des institutions une volonté de dépasser certaines appréhensions, comme le fait d'admettre l'apport et l'expertise des amateurs, de comprendre l'intérêt de rendre accessibles les contenus qu'ils possèdent ou dont ils sont les dépositaires, et plus largement d'adopter des pratiques d'ouverture et de transparence, ce qui les force à développer des manières inédites de travailler et de prendre des décisions. Dans un premier temps, Stacy Allison-Cassin mène une réflexion sur la présence de la musique canadienne en ligne en proposant un retour critique sur une initiative qu'elle a menée qui visait à rendre les archives du Festival folk Mariposa accessibles au public via Wikidata. Ce faisant, elle brosse un portrait de l'appropriation de Wikidata au sein du milieu bibliothéconomique canadien et, plus largement, au sein des GLAM au pays. Cette contribution est suivie de deux autres études de cas.

Nathalie Thibault explique comment et pourquoi le Musée national des beaux-arts du Québec (MNBAQ) s'est engagé dans des initiatives collaboratives de type Open GLAM, mettant en place des chantiers wikimédiens pour rendre ses collections plus accessibles grâce à des licences libres et permettre la réutilisation, le partage et la modification de ses données et contenus culturels. Du côté des institutions gouvernementales, le texte de Miguel Tremblay présente un projet innovant de téléversement de métadonnées météorologiques produites par le gouvernement du Canada, un cas qui illustre une application du modèle de gouvernement ouvert grâce à Wikidata.

Le dernier axe thématique concerne le domaine des littératies. Un des leitmotivs partagés par les auteurs et autrices est de souligner les diverses compétences que l'on développe ou qui sont sollicitées dès lors que l'on prend part aux projets Wikimedia et à la vie communautaire du mouvement. La littératie se décline ici et là sous plusieurs angles, qui ont souvent tendance à s'agréger. Ainsi, la littératie informationnelle est aussi à bien des égards une littératie numérique, qui convoque également des compétences pouvant devenir plus spécifiquement juridiques, herméneutiques, mais aussi algorithmiques, relative aux données organisationnelles, relationnelles et technologiques. Au fil de leur parcours au sein du mouvement, les individus, les communautés et les institutions incorporent progressivement la plupart de ces formes de littératie requises pour mener à bien les initiatives dans lesquelles ils s'engagent. Ces multiples apprentissages attestent de la valeur indéniablement éducative des projets du mouvement Wikimedia. Cela dit, alors que la majorité des textes au sein du livre s'intéresse aux pratiques de production

amateurs ou expertes, le chapitre de Denise Smith se distingue par son ancrage dans le champ de connaissance traitant de la littératie en matière de santé, et ce, à la fois du point de vue de la communauté productrice de savoir que celui du grand public qui cherche à prendre des décisions éclairées en matière de santé. Situait son propos dans le contexte de l'*info-démie* de désinformation qui a déferlé durant la pandémie de COVID-19, Smith montre à quel point Wikipédia s'est révélé être une source d'information majeure pour les internautes du Canada.

En guise de postface, nous avons demandé au philosophe Pierre Lévy de prolonger et d'élargir la réflexion sur les processus d'intelligence collective étudiés tout au long de ce livre. Partant d'une prise en compte des pratiques de production et de partage des savoirs qui ont lieu au sein du mouvement Wikimedia, Lévy les réinscrit dans une problématisation plus large, ancrée à la fois dans l'histoire humaine et animale, mais aussi dans la plus récente actualité technologique de l'intelligence artificielle générative.

Outre la présentation des contenus de cet ouvrage, cette introduction visait à capter les principaux ressorts du mouvement Wikimedia au Canada à l'aide d'une brève chronologie et d'une cartographie des pôles institutionnels qui y sont liés. Rendez vous sur la plateforme en ligne associée au livre pour accéder à une autre cartographie, sonore, qui donne accès à un échantillon de la pluralité des voix constitutives du mouvement, et ce, grâce à une multitude de courtes capsules audio géolocalisées qui ont été enregistrées par des bénévoles de partout au Canada. Ils et elles nous expliquent dans leur langue les raisons et les motivations qui les incitent à s'engager et à participer aux projets Wikimedia.

AXE I

## **Identités**





## CHAPITRE 1

# Les protocoles de pluralisation

### *Négocier la cohabitation culturelle dans Wikipédia*

NATHALIE CASEMAJOR

Les principaux groupes linguistiques et culturels du Canada qui écrivent dans Wikipédia ont eu à faire reconnaître les particularités de leur société sur cette plateforme internationale depuis ses débuts. Fruit de plusieurs enquêtes ethnographiques, ce chapitre analyse et compare le fonctionnement de la cohabitation interculturelle qui a lieu dans les versions anglaises, françaises et atikamekw relativement aux particularismes du Canada. Les manières légitimes de faire exister le local dans le global dépendent de la culture éditoriale spécifique de chaque version linguistique de Wikipédia, qui est elle-même façonnée par les idées des groupes sociaux dominants qui habitent l'encyclopédie.

Wikipédia est un creuset où les Canadiens et Canadiennes cohabitent avec d'autres internautes situés aux quatre coins du globe. Organisée par langue et non par pays, l'encyclopédie

est investie par des internautes qui partagent une même langue, mais pas les mêmes usages ni les mêmes références : les Canadiens francophones jouent au hockey avec une *puck* tandis que les Français jouent avec un palet ; les Canadiens anglophones conduisent avec un *driver's licence* (permis de conduire), tandis que les Britanniques ont un *driving licence*. Ces différences culturelles font la richesse de l'encyclopédie, mais elles entraînent aussi de nombreux conflits d'édition. Comment les différents groupes linguistiques et culturels canadiens négocient-ils leur place dans Wikipédia ? L'objectif ici est d'analyser la cohabitation, au sein d'un même espace encyclopédique, entre membres de la communauté wikipédienne issus de différentes cultures. Je propose de comparer les cas des trois principales Wikipédias pour le Canada, soit les encyclopédies en anglais, en français et en atikamekw nehiromowin.

En croisant deux fils d'analyse, celui de la pluralisation culturelle et celui de la gouvernance de la plateforme Wikimedia, j'identifie deux types de protocoles d'édition qui influencent la pluralisation des savoirs dans ces espaces : l'établissement de dispositions communes à tous les groupes (*protocoles de communalité*), et l'introduction de clauses qui permettent l'expression particulière des différentes cultures (*protocoles de localité*). Ce travail s'appuie sur une série d'études ethnographiques, d'observations participantes et de recherches-actions menées depuis 2011 au sujet de la participation à Wikipédia (Casemajor *et al.*, 2019 ; Casemajor et Couture, 2020)<sup>1</sup>. Les observations présentées

---

1. Les observations faites dans ce chapitre documentent l'état des pages de Wikipédia en décembre 2021.

dans ce chapitre ont principalement été effectuées en 2021. Elles ont été enrichies par une analyse actualisée en 2024, qui a révélé que les ordres de grandeur dans les statistiques de contribution et de lectorat demeurent globalement inchangés.

## **Culture politique et politique des cultures**

### **Pluralisation culturelle**

La notion de pluralisation réfère aux processus de coexistence, d'inclusion et d'exclusion entre différents groupes sociaux au sein d'un même espace public. Les groupes qui cohabitent dans Wikipédia sont positionnés à différents niveaux de pouvoir et de légitimité dans le champ de la production de savoirs. Ces écarts entraînent une négociation politique mettant en jeu des rapports de pouvoir et de minorisation, qui s'expriment sous la forme de concertations et de conflits d'édition encyclopédique (Auray *et al.*, 2009). Dans cette perspective, l'analyse de la pluralisation met l'accent sur la nature évolutive d'un processus qui implique des ajustements constants entre différentes parties prenantes hétérogènes.

De nombreux travaux universitaires ont documenté les écarts entre groupes sociaux dans Wikipédia. Ces écarts sont abordés du point de vue des différences de genre (Hill et Shaw, 2013; Berson *et al.*, 2021), de littératie (Ford et Geiger, 2012) ou de niveau d'expérience dans l'encyclopédie (Hargittai et Walejko, 2008). Les enjeux de diversité culturelle dans Wikipédia ont aussi été étudiés sous l'angle de l'origine des contributeurs et contributrices (Massa et

Scrinzi, 2012) et des obstacles à l'inclusion des cultures minoritaires et des voix non occidentales (Reagle *et al.*, 2020). Toutefois, ces enjeux sont rarement examinés sous l'angle de la pluralisation culturelle à l'intérieur d'une même Wikipédia. C'est l'angle d'analyse auquel je vise à contribuer ici, en m'intéressant à l'expression des points de vue et des usages de la langue culturellement situés dans des territoires nationaux ou locaux.

Je parlerai de *protocoles* de pluralisation culturelle pour désigner un ensemble de procédures et de consensus formalisés au sujet de la cohabitation des différentes cultures. Ils s'incarnent à la fois dans les usages, les prises de décision et les systèmes techniques d'organisation de l'encyclopédie. Appliqué au domaine des plateformes numériques, le terme *protocole* suggère que les dimensions techniques et sociopolitiques de l'activité s'enchevêtrent dans un même dispositif informatisé. Dans quelle mesure les protocoles de gouvernance de Wikipédia sont-ils des opérateurs de pluralisation culturelle ? Je propose de distinguer ici deux types de protocoles parallèles et complémentaires qui prennent part à ce processus : des *protocoles de communalité* et des *protocoles de localité*. Tous deux jouent un rôle significatif dans la gouvernance de la pluralisation culturelle de Wikipédia.

## Gouvernance de Wikipédia

L'activité d'édition de Wikipédia est régie par une série de principes, règles et recommandations qui encadrent l'activité des contributeurs et contributrices. Ces cadres de gouvernance sont organisés selon une hiérarchie, où les directives les plus fixes et universelles à toutes les Wikipédias

priment sur celles qui sont plus spécifiques à chaque encyclopédie et plus malléables. Ils instaurent une gouvernance qualifiée de procédurale (Cardon et Levrel, 2009), au sens où les arguments qui visent la conformité aux conventions établies tendent à primer sur ceux qui prennent en compte le contexte socioculturel plus large d'un problème.

Au sommet de la hiérarchie des conventions de Wikipédia se trouvent premièrement les *principes fondateurs* (*pillars* en anglais), définis comme « le fondement intangible de toutes les règles et recommandations du projet ». On en compte cinq : encyclopédisme, neutralité de point de vue, licences libres, savoir-vivre communautaire et souplesse des règles. Ces principes sont des impératifs qui s'appliquent de façon non négociable à toutes les Wikipédias – contrairement aux autres consignes décrites plus bas, qui sont déterminées localement et adaptables à chaque version linguistique de l'encyclopédie. Deuxièmement, on trouve les *règles* (*policies*), soit des « normes largement acceptées » par les Wikipédiens et Wikipédiennes ; ces standards sont formellement adoptés par une prise de décision collective des membres. Troisièmement, en bas de la hiérarchie figurent les *recommandations* (*guidelines*), qui cristallisent des usages établis ; adoptées par consensus (discussion et sondage), elles peuvent être respectées moins strictement que les règles.

La Wikipédia en anglais est nettement plus formalisée que son homologue francophone, avec près de trois fois plus de règles et de recommandations établies. Cet écart s'explique principalement par le statut fondateur de l'encyclopédie anglophone, la première à avoir été créée en 2001, et encore de loin la plus développée. Les règles et recommandations de la Wikipédia anglophone servent souvent de

modèle aux autres Wikipédias, qui en ont repris et adapté plusieurs, sans toutefois les retenir toutes. En effet, les divergences entre ces encyclopédies découlent aussi de philosophies politiques, d'habitudes de gouvernance et de types de bassins de rédacteurs et rédactrices qui sont propres à chacune.

En 2021, la Wikipédia en anglais comptait 66 700 contributeurs et contributrices actifs, issus en majorité des États-Unis (35,4 %), du Royaume-Uni (11,3 %), de l'Inde (9,7 %), du Canada (4,6 %) et de l'Australie (4,2 %). Cette distribution est globalement la même pour les contributeurs et contributrices que pour les lecteurs et lectrices, avec des pourcentages légèrement inférieurs : 30,2 % de lecteurs et lectrices des États-Unis pour 3,6 % du Canada. On observe donc dans cette encyclopédie une position prépondérante des États-Unis, pays où le projet Wikipédia a vu le jour et qui abrite le siège social de la Fondation Wikimedia. Le Canada y occupe une place minoritaire, parmi une série de pays du Commonwealth. Néanmoins, tant les États-Unis que le Canada tendent à être surreprésentés par rapport à leur poids démographique dans le monde anglophone.

De son côté, la Wikipédia francophone affichait en 2021 environ 7 800 contributeurs et contributrices actifs, soit 8 fois moins que son homologue anglophone. Les principaux pays de contribution sont la France (71,4 %), la Belgique (5,3 %), le Canada (5 %), la Suisse (3,7 %) et l'Algérie (1,7 %). Là encore, la distribution des lecteurs et lectrices reproduisait globalement celle des contributeurs et contributrices à une moindre échelle (54,2 % pour la France, 3,5 % pour le Canada). Comparé à l'encyclopédie anglophone, on remarque que dans l'édition francophone, l'écart

entre le premier pays de contribution (la France) et les autres est considérablement plus élevé. La position de la France y est nettement plus dominante. En ce qui concerne la position du Canada, elle est globalement aussi minoritaire dans l'édition en français qu'en anglais, mais les Canadiens et Canadiennes y sont plutôt sous-représentés vu leur poids démographique dans le monde francophone.

Le cas de la Wikipédia en atikamekw nehiromowin est sensiblement différent des précédents. Créée plus récemment (en 2013), ses contenus sont encore très restreints (1 580 articles en 2021) et ses rédacteurs et rédactrices sont rares (4 actifs en moyenne sur les 12 derniers mois en 2021), mais elle suit une courbe de développement continu – une réussite étant donné le petit bassin de locuteurs de la langue (moins de 7 000). Cette encyclopédie ne contient que très peu de règles et recommandations formalisées. Ainsi, en 2021, la page décrivant les principes et les règles générales de la Wikipetcia Atikamekw Nehiromowin se limitait principalement à énoncer les cinq principes fondateurs. Cette moindre formalisation est surtout liée au stade encore émergent de cette encyclopédie. Mais elle tient aussi à la relation différente que les peuples autochtones entretiennent avec l'organisation politique et les savoirs : l'oralité y est traditionnellement le moyen de transmission des coutumes et des connaissances, tandis que la formalisation de règles écrites est plutôt un trait caractéristique de la modernité occidentale (Casemajor *et al.*, 2019).

Les prochaines sections nous permettront d'identifier les éléments de la liste des principes, règles et recommandations de chaque Wikipédia qui opèrent comme protocoles de communalité et de localité. Quels sont les points communs

et les différences à cet égard entre les diverses éditions linguistiques de l'encyclopédie ? Et comment expliquer ces différences ? S'il est possible de comparer en détail les protocoles de gouvernance de la Wikipédia en anglais et en français, la spécificité de la Wikipédia en atikamekw nehiromowin rend cette tâche hasardeuse. Elle requiert plutôt une analyse distincte, tout en la mettant en perspective avec les autres encyclopédies. Plus largement, ces questions soulèvent l'enjeu du partage des savoirs instauré par Wikipédia, fortement influencé par des dynamiques politiques (et géopolitiques), épistémologiques et technologiques.

## **Protocoles de pluralisation**

### **Protocoles de communalité**

Par *protocoles de communalité*, j'entends les dispositions qui s'appliquent communément à toutes les situations d'édition au sein d'une Wikipédia. Ces dispositions sont généralisées à l'ensemble des contenus, quels que soient les objets, les contextes, les contributeurs ou contributrices et les lecteurs ou lectrices concernés, ne tolérant a priori pas d'exception. Autrement dit, ces protocoles s'appliquent indépendamment des mondes culturels dans lesquels se situent les savoirs, leur production et leur interprétation. Ils visent à produire de la *communalité* au sens de « ce qui est commun » (en contrepoint de la *localité*, abordée plus loin).

### ***Neutralité de point de vue***

C'est dans les principes fondateurs de l'encyclopédie que l'on trouve le principal protocole de communalité, partagé par toutes les versions linguistiques. Il s'agit d'un des plus anciens principes de gouvernance de Wikipédia : le principe de *neutralité de point de vue* (*neutral point of view*). L'idée de neutralité y est décrite comme le fait de « présenter de manière impartiale les différents points de vue pertinents, en leur accordant une place proportionnelle à leur importance ». Il fonctionne comme un principe d'exclusion, puisqu'il a pour fonction d'écarter les contributions partisans, les opinions personnelles, les affirmations sérieusement contestées et les points de vue ultraminoritaires. Mais il est simultanément un principe d'inclusion, puisque c'est la présentation d'une multiplicité de points de vue (proportionnellement à leur poids relatif dans les sources fiables publiées sur le sujet) qui doit permettre l'atteinte de cet objectif. C'est ainsi que le principe de la neutralité de point de vue agit comme un protocole de pluralisation : il vise à créer un espace commun, qui exclut invariablement les points de vue non encyclopédiques, tout en incluant une multiplicité de points de vue jugés pertinents.

Au cœur du projet encyclopédique, ce principe a néanmoins fait l'objet de nombreuses critiques (Tkacz, 2014). Les politiques internes de la Wikipédia en anglais présentent plusieurs objections fréquemment opposées à ce principe de neutralité. L'une d'entre elles concerne « l'orientation anglo-américaine » de l'encyclopédie (*Anglo-American focus*). Expliquée par la prédominance des contributeurs et contributrices états-uniens et (dans une moindre mesure) britanniques, cette perspective culturelle dominante est citée

comme une disposition contraire au principe de neutralité, et destinée à être corrigée. Ces limites ne sont pas mentionnées dans la page décrivant le principe en français, qui reste descriptive et générale. En revanche, la Wikipédia en français contient une recommandation nommée *principe de moindre surprise* qui s'applique à un enjeu culturel similaire, mais de façon contraire.

### *Principe de moindre surprise*

Listée parmi les conventions de style, cette recommandation concerne principalement le style rédactionnel des articles et tout particulièrement le choix des titres. Elle « vise à ce que l'information soit aisément comprise par le lecteur » en s'appuyant sur des arguments de cohérence, de clarté et de compréhension par la frange la plus large du lectorat. Initialement mentionnée dans une résolution de la Fondation Wikimedia sur les sujets controversés et offensants, cette convention de moindre surprise a été intégrée par la Wikipédia dans un manuel d'aide à la rédaction des articles. Reprise par la Wikipédia francophone, elle a été formalisée en 2011 en tant que recommandation autonome, à la suite de discussions acrimonieuses entre Français et Québécois au sujet de l'utilisation d'expressions québécoises, peu courantes dans l'Hexagone. Par exemple, la tentative (infructueuse) de renommer l'article « Palet (sport) » en rondelle – terme utilisé au Québec dans le hockey sur glace – a donné lieu en 2008 à une guerre d'édition mémorable dans l'histoire des relations entre Québécois et Français au sein de Wikipédia.

Dans la Wikipédia anglophone, la convention de la moindre surprise n'a pas vocation à arbitrer des conflits liés

aux différences culturelles dans le rapport au savoir, telles que les diverses manières de nommer un article encyclopédique en fonction des usages de la langue dans chaque pays. En revanche, dans l'encyclopédie francophone, ce principe de moindre surprise a été formalisé en une convention qui sert en pratique à réguler l'usage des variations de la langue française. Plus précisément, elle induit l'usage du « français de France » comme standard commun par défaut, au détriment des autres variantes locales. Étant donné qu'une majorité de lecteurs et lectrices sont basés en France, « la terminologie la plus courante [est] généralement celle utilisée en France pour des raisons statistiques » afin de causer le moins de surprise au lectorat français. Cette convention de moindre surprise s'applique « pour tous les lecteurs, quels que soient leur pays et leur culture ». Elle établit ainsi un protocole de communalité linguistique fondé sur le particularisme de la France.

Largement appliquée en pratique, cette « moindre surprise » pour les Français et Françaises est aussi régulièrement contestée au motif qu'elle imposerait un « francocentrage » indu, autrement dit une violation du principe de neutralité, en imposant le point de vue culturel de la France sur les autres. En observant les pages de discussion où les contributeurs ou contributrices débattent à ce sujet, on trouve divers arguments qui nuancent le « francocentrisme » de la recommandation de moindre surprise, mettant en pratique des exceptions à la règle. Mais ces réflexions et ces exceptions ne sont pas clairement indiquées dans la page énonçant la recommandation. Elles restent donc relativement fragiles quand surviennent des conflits d'édition, où l'argument de la majorité française tend à l'emporter.

Dans la version en anglais, la recommandation de moindre surprise est tempérée par une autre convention nommée « *pas ce qui vous vient d'emblée à l'esprit* » (not “*what first comes to [your] mind*”). Elle stipule que les termes qui nous semblent les plus courants sont inévitablement interprétés en fonction de notre situation géographique, et que ces lectures partiales peuvent induire des biais systémiques nuisant au projet de l'encyclopédie.

Plus fondamentalement, une série de recommandations de la Wikipédia anglophone édictent qu'aucun pays ou groupe n'a priorité en termes de langue : « la Wikipédia anglaise ne préfère aucune variété nationale d'anglais à une autre ». Le groupe statistiquement majoritaire – les États-Uniens – n'a ainsi pas priorité sur les autres en théorie : il est précisé que « l'orthographe anglaise américaine ne doit pas être transformée en orthographe anglaise britannique, et vice versa » [traduction].

Une différence substantielle existe donc entre les Wikipédias en anglais et en français du point de vue de leurs protocoles de communalité. Du côté anglophone, l'orientation anglo-américaine de l'encyclopédie est d'emblée identifiée comme un biais systémique à corriger dans la mesure où la prévalence d'un point de vue culturel sur les autres est vue comme une entorse au principe cardinal de la neutralité. Tandis que du côté francophone, la préséance du point de vue français est vue comme un garant de neutralité au nom d'un principe de moindre surprise pour la majorité des lecteurs et lectrices, basés en France.

## Protocoles de localité

Contrairement aux protocoles de communalité, les protocoles de localité sont des clauses de gouvernance qui visent à permettre l'expression de la différence culturelle. Ils participent à la pluralisation culturelle dans la mesure où ils impliquent un processus politique de reconnaissance de la valeur de cultures distinctes (Honneth, 2004). Dans le cadre de Wikipédia, trois dynamiques de reconnaissance sont en jeu.

La première porte sur le caractère distinctif des cultures, c'est-à-dire sur l'existence de traits particuliers qui leur sont propres. Les groupes culturels se distinguent ainsi les uns des autres dans l'encyclopédie par un usage de la langue, une vision du monde, un panthéon de personnes notoires et des événements historiques spécifiques.

Une deuxième dynamique concerne la reconnaissance de la légitimité encyclopédique des particularismes culturels. Les protocoles mis en place dans Wikipédia à cet effet concernent les éléments suivants: le choix de sujets d'articles liés à une culture particulière (règle de notoriété; recommandation sur les sources fiables); l'expression de points de vue culturellement situés à l'intérieur des articles, des titres et des catégories d'indexation (conventions de titres); l'utilisation de variantes de la langue dans la rédaction des articles (conventions de style). Ces protocoles se situent au bas de la hiérarchie des normes de Wikipédia: il s'agit d'exception à des règles et surtout de recommandations, plus flexibles et définies au sein de chaque encyclopédie. Chaque Wikipédia possède donc ses propres protocoles de localité, adaptés aux différences culturelles observées dans son espace linguistique.

Troisièmement, la pluralisation culturelle dans Wikipédia passe par la reconnaissance qu'il existe certains obstacles à l'expression ou à l'inclusion de la diversité culturelle dans l'encyclopédie. Ces freins sont liés, d'une part, à des facteurs internes, en particulier à la démographie des contributeurs et contributrices, déséquilibrée en termes de géographie, de classe sociale et d'origine culturelle. D'autre part, les obstacles à la pluralisation sont causés par des facteurs externes. Parmi eux se trouve l'inégale distribution des ressources matérielles et symboliques entre pays ou groupes sociaux en ce qui concerne la production des savoirs (Fraser, 2004 ; Casemajor et Couture, 2020). Mais ces obstacles sont aussi dus à des biais cognitifs, institutionnels ou politiques qui favorisent certains des points de vue au détriment des autres. L'ensemble de ces facteurs peut freiner la capacité d'un groupe social à faire reconnaître sa singularité linguistique et culturelle dans l'encyclopédie ou à y rendre son point de vue visible et accessible. Ces obstacles entraînent une sous-représentation de certains groupes de population (et une surreprésentation d'autres), situation jugée nuisible à l'objectif de neutralité.

### « *Systemic bias* » vs « Internationalisation »

La reconnaissance des obstacles à la pluralisation culturelle se manifeste différemment dans les Wikipédias anglophone et francophone. En anglais, l'une des premières discussions à ce sujet a pris la forme d'une guerre d'édition d'anthologie, fondatrice et très violente, quelques mois seulement après la création de l'encyclopédie. Un contributeur nommé User 24 (banni depuis pour son attitude agressive) y contes-

tait la suppression d'un article intitulé «Licence virale» (*Viral licence*), dont l'ambition était de refléter une perspective internationale sur le droit d'auteur. L'article avait été supprimé au motif que l'article «Copyleft» couvrait déjà le sujet, certes d'un point de vue états-unien, mais que ce point de vue culturel était majoritaire par défaut étant donné le lectorat basé aux États-Unis. S'ensuivit une épique et très savante confrontation épistémologique sur la neutralité de point de vue et l'ethnocentrisme.

Par la suite, la réflexion s'est formalisée dans deux directions. D'une part, une série de wiki-projets a rassemblé des utilisateurs en groupes d'intérêt autour de sujets identitaires, culturels ou nationaux. Par exemple, les projets Groupes ethniques (2004), Canada (2004) ou Peuples indigènes de l'Amérique du Nord (2006) visent à développer la qualité des articles sur ces sujets particuliers. D'autre part, des wiki-projets ont été spécifiquement créés pour travailler à la cohabitation des différentes cultures dans Wikipédia, tel le projet Lutter contre les biais systémiques (*Countering systemic bias*) en 2004. Ce dernier joue un rôle structurant, puisqu'il a donné lieu à un essai du même nom (2009). Relativement stabilisé, mais toujours objet de débats plus ou moins houleux, cet essai sur les biais systémiques identifie des obstacles à la neutralité de point de vue dans Wikipédia : les points de vue qui y sont représentés sont jugés peu diversifiés en termes de classe sociale, de genre, d'origine culturelle et géographique.

Dans la Wikipédia en français, il existe aussi des groupes d'intérêt locaux, tel le projet Québec (2005). Mais le sujet général de la diversité culturelle et linguistique y est traité de manière relativement différente. Ainsi, la page en anglais

*Countering systemic bias* (Lutter contre les biais systémiques) est plutôt traduite en français par le terme *Internationalisation*, qui renvoie à l'injonction *globalize* de la Wikipédia anglophone : l'idée que Wikipédia est un projet mondial et qu'elle doit pouvoir être « lu[e] et compris[e] par n'importe quel lecteur francophone », tel qu'elle est formulée dans la page « Internationalisation ». Assez peu développée dans sa version en français, cette page a néanmoins le statut de recommandation, un état plus formalisé que l'essai en anglais. Son champ d'application est cependant débattu, entre une interprétation minimaliste qui la restreint à une convention de style (la forme) et une interprétation maximaliste qui en fait une convention sur le contenu encyclopédique (le fond).

Lors de sa création en 2005, la page « Internationalisation » s'appelait plutôt « Guide de défrancocentrage » : autrement dit, elle proposait des façons de mitiger le point de vue culturel dominant de la France (dit *francocentrisme*). Par la suite, le terme *francocentrisme* a été enlevé du corps du texte<sup>2</sup>. En page de discussion, un utilisateur défend même cette position : « Le fr [de l'adresse web fr.wikipedia] désigne principalement la France chez moi, et non la francophonie [...] personnellement, en tant que Français, je me fous éperdument de la "situation au Québec". » Chez ce contributeur, c'est tout bonnement le statut transnational de la Wikipédia francophone qui est contesté. Dans son état actuel, la recommandation d'internationalisation n'identifie pas explicitement l'existence de biais particulier lié au statut dominant de la France. Elle énonce plutôt de façon générale la nécessité d'« inclure différentes perspectives culturelles dans un article

---

2. Il n'apparaît que dans les liens connexes en bas de page.

afin de tenir compte de la diversité culturelle». Dans une bannière liée à la page, on trouve néanmoins la mention importante que «la Wikipédia en français n'est pas la Wikipédia de la France, du Québec ou du Gabon». Autrement dit, elle n'est pas l'apanage d'un pays en particulier. La recommandation d'internationalisation reste abondamment discutée : sa page de discussion est deux fois plus fournie que le contenu de la recommandation elle-même, ce qui témoigne d'un consensus peu stabilisé.

On peut de nouveau constater une différence notable entre les Wikipédias en français et en anglais. Du côté francophone, reconnaître l'existence de biais liés à la position dominante de la France est jugé non neutre par certains utilisateurs, qui limitent la recommandation sur l'internationalisation à des caractéristiques formelles, sans reconnaissance politique des enjeux de fond. Du côté anglais, le vocabulaire des biais systémiques pointe explicitement vers la reconnaissance de déséquilibres liés à des facteurs internes et externes, incluant des rapports de pouvoir géopolitiques qui nuisent à l'objectif de neutralité. L'essai «Systemic bias» va plus loin en proposant un rééquilibrage des sujets et des perspectives sous-représentées. L'étendue des correctifs à apporter reste toutefois débattue : «Wikipédia doit-elle refléter le monde tel qu'il se présente, ou tel que les Wikipédiens espèrent qu'il pourrait être?» [traduction]. Le consensus actuel veut que Wikipédia «ne répare pas les grands torts», car l'encyclopédie est vouée à synthétiser l'état actuel de l'information, et non à produire une pensée originale : «Wikipédia ne guide pas, nous suivons». Ainsi, les protocoles mis en place pour favoriser l'expression des particularismes culturels locaux dans Wikipédia portent essentiellement sur un rééquilibrage

interne à l'encyclopédie, et non sur une action politique visant à influencer l'état du monde.

### *Diversifier les sujets des articles (admissibilité)*

Outre la reconnaissance d'obstacles à la pluralisation culturelle, on trouve dans Wikipédia des séries de protocoles qui favorisent l'expression de particularismes culturels. Une première série concerne la création d'articles dans l'encyclopédie. Ces protocoles visent à diversifier les sujets des articles, en mitigeant certains effets de la recommandation de notoriété (*notability*) jugés défavorables à la diversité culturelle. Cette recommandation a d'abord été introduite en 2006 dans la Wikipédia anglophone pour en exclure les sujets non pertinents. Étroitement liée à la règle de vérifiabilité (*verifiability*), elle stipule que seuls sont admis les sujets « qui ont fait l'objet d'une attention suffisamment importante » en considérant « les preuves provenant de sources fiables et indépendantes ».

Cette recommandation, largement admise dans son principe général, a fait l'objet de nombreuses controverses sur ses modalités d'application. Les controverses opposent un parti « inclusionniste », qui prône une large rétention d'articles peu étayés, et un parti « suppressionniste<sup>3</sup> », qui prône une stricte sélectivité des articles et entraîne de ce fait la suppression de plusieurs articles traitant de cultures minoritaires. En effet, l'admissibilité d'articles qui traitent de ces cultures est plus difficile à établir en raison de la moindre

---

3. Voir les pages « Inclusionnisme » et « Suppressionnisme », principes politiques définis dans le site Méta de la plateforme Wikimedia : <https://meta.wikimedia.org/wiki/Inclusionnisme>, <https://meta.wikimedia.org/wiki/Suppressionnisme>.

disponibilité de sources de référence en anglais, ou du manque de sources tout court. En français, l'inclusion d'une diversité de perspectives internationales est perçue par certains contributeurs et contributrices comme un défi encore plus grand : « Dans beaucoup de domaines touchant les horizons culturels lointains, il y a peu de sources francophones comparées aux ressources anglophones », commente l'un d'entre eux. Les effets de la disponibilité et de la légitimité inégales des sources de référence ont été largement documentés par des travaux universitaires qui concluent à une exclusion de certains sujets et voix marginalisés (McDowell et Vetter, 2022 ; Lemieux *et al.*, 2023).

Les sujets qui traitent du Canada tendent plutôt à être surreprésentés dans les Wikipédias en anglais et en français au regard de son poids démographique global. Leur proportion reste toutefois très minoritaire comparativement à celle des États-Unis et de la France, qui dominent respectivement le nombre d'articles par pays dans les deux encyclopédies. Mais certains débats subsistent sur la couverture de sujets touchant la diversité culturelle interne au Canada. Pays multiculturel et multilingue<sup>4</sup>, le Canada est en effet lui-même caractérisé par une grande diversité culturelle à l'intérieur de ses frontières. Ainsi, plusieurs articles sur l'histoire de l'immigration chinoise dans différentes provinces ou municipalités du Canada ont fait l'objet d'une

---

4. L'État canadien a adopté une *Loi sur le multiculturalisme canadien* en 1971. Selon la *Loi sur les langues officielles* de 1969, le Canada est bilingue (anglais et français), mais il existe aussi plus de 70 langues autochtones parlées au pays. Bien qu'elles n'aient pas de statut officiel, elles sont constitutives de la mosaïque linguistique du pays qui peut ainsi être qualifié de multilingue.

demande de suppression ou de fusion, au motif que le sujet était déjà couvert par d'autres articles et que les sources citées étaient insuffisantes ou biaisées. L'article «Chinese Canadians in Greater Vancouver» a par exemple été maintenu à l'issue d'un long débat, en citant à l'appui l'existence de quatre ouvrages savants et d'une maîtrise universitaire sur le sujet.

Dans la Wikipédia en français, la recommandation sur la notoriété établit une exigence supplémentaire, à savoir que les sources fiables reconnues doivent être avant tout des sources «d'envergure nationale ou internationale» de façon à exclure «des sujets dont la notoriété serait purement locale ou restreinte à un groupe spécifique de personnes». Cette clause rend contestable la citation de sources québécoises, et ce, parce qu'elles n'ont pas une audience à l'échelle du pays, alors que le Québec est lui-même défini comme une nation, avec des médias «nationaux». La difficulté est encore plus grande pour des cultures infranationales, comme le montre l'exemple de l'article «Littérature franco-ontarienne» (Casemajor et Couture, 2020). Un utilisateur français a proposé sa suppression au motif que cette littérature n'était pas assez singulière pour se distinguer de l'article «Littérature canadienne-française». L'article a été maintenu non pas pour des raisons politiques ou épistémologiques de reconnaissance des cultures minoritaires, mais parce que des sources de qualité ont pu être citées à l'appui de son maintien dans l'encyclopédie. Plus généralement, les nombreux débats autour de la suppression d'articles traitant de cultures minoritaires montrent que les sources à leur sujet jugées légitimes peuvent venir à manquer, compte tenu de leur marginalité et

de leur faible représentation dans les institutions culturelles, académiques et politiques canadiennes.

Certains protocoles viennent pallier ces difficultés. Ils prennent la forme de conventions de notoriété applicables à un seul pays ou groupe culturel. Dans la Wikipédia en anglais, c'est le cas de la recommandation du manuel de style concernant les sujets canadiens (la page «Wikipedia: Manual of Style/Canada-related articles»), où l'on retrouve des directives sur la notoriété spécifiques à ce pays. De même, dans la Wikipédia en français, le wikiprojet Suisse a élaboré une recommandation selon laquelle «du fait des régions linguistiques, il n'existe pas de médias couvrant l'ensemble du pays. Le projet Suisse considère donc que les régions linguistiques (Suisse alémanique, Suisse romande, Suisse italienne, Suisse rhéto-romane) sont assimilées à une envergure nationale». L'objectif des protocoles de localité applicables à la notoriété des sujets encyclopédiques est donc de favoriser l'inclusion d'articles dont l'admissibilité pourrait être débattue. Ce faisant, ils contribuent à la pluralisation culturelle en reconnaissant la légitimité encyclopédique de sujets liés à des cultures localisées sur un territoire particulier.

Quant à la Wikipédia en atikamekw nehiromowin, ses conventions de notoriété et de vérifiabilité des sources sont très peu formalisées. Elles incluent toutefois certaines dispositions originales, adaptées à la singularité des cultures autochtones et de leur rapport au savoir (Casemajor *et al.*, 2019). Ainsi, lors des discussions qui ont accompagné le lancement public de l'encyclopédie, les participants et participantes ont fait le choix d'exclure a priori les sujets qui concernent les rituels sacrés et les plantes médicinales.

Cette disposition relative à la notoriété vise d'abord à protéger certains pans du savoir atikamekw nehirowisiwok qui ne sont traditionnellement transmis que d'un aîné ou d'une aînée à son apprenti ou apprentie, et non via une forme écrite à tout un chacun. Cette disposition est aussi influencée par une certaine méfiance engendrée par la dépossession des peuples autochtones de leurs ressources dans l'histoire de la colonisation. Un autre protocole spécifique à la Wikipédia en atikamekw nehiromowin concerne la règle de vérifiabilité des sources. Dans cette encyclopédie, les sources orales font partie des sources les plus légitimes, étant donné la valeur primordiale accordée à la parole des aînés et aînées dans la transmission des savoirs, qui passe traditionnellement par le récit oral. Très singuliers comparés à ceux des autres Wikipédias, ces protocoles sont spécifiquement adaptés aux modes de transmission des savoirs caractéristiques des peuples autochtones, tout en s'intégrant dans le format de l'encyclopédie moderne caractéristique du monde occidental.

### *Internationaliser et particulariser*

Une deuxième série de protocoles de localité concerne les articles existants. Certains visent à introduire des points de vue culturels et linguistiques diversifiés à l'intérieur d'un même article dont la portée est internationale – par exemple, l'article « Médecin ». Dans la Wikipédia française, cet article a été critiqué pour ne représenter que l'état de la profession en France. Ce débat a donné lieu à la scission entre un article centré sur le point de vue français, intitulé « Médecin (France) », et un article générique à vocation internationale

« Médecin », qui mentionne que « la formation des médecins varie considérablement à travers le monde ». Il inclut d'ailleurs une section spécifique sur le Canada.

Le cas de l'article « Médecin (France) » illustre le privilège donné à un seul point de vue, légitime dans le cas d'un sujet d'intérêt strictement national. Dans ce cas, il s'agit de reconnaître l'intérêt particulier d'un article pour une culture ou un pays et, par conséquent, à y favoriser un point de vue culturel particulier. Dans la Wikipédia en anglais, quand un article entretient un « lien national fort avec le sujet », il peut en outre être rédigé suivant la variante nationale (ou régionale) de la langue caractéristique du groupe concerné. Toutefois, l'identification de ce lien national fort n'est pas toujours aisée. Elle est souvent l'objet d'une négociation, qui peut être pacifique ou conflictuelle.

L'article aujourd'hui intitulé « Canada jay » (« Mésangeai du Canada ») en est un bon exemple. Originellement intitulé « Gray jay » (selon la graphie états-unienne), il a d'abord été renommé « Grey jay » (graphie canadienne) à la suite de la désignation de cet oiseau comme symbole national du Canada par le magazine de la Société géographique royale du Canada. Puis au terme de nouvelles discussions qui ont duré deux ans, il a été renommé « Canada jay », mais seulement après que le Comité ornithologique international, autorité en matière de noms d'oiseaux, eut adopté officiellement ce nom. Cet exemple montre une limite des protocoles de localité de Wikipédia : pour des sujets où il existe une nomenclature scientifique bien établie avec des standards internationaux (chimie, zoologie), c'est ce standard externe qui tend à prévaloir sur les négociations de politique interne à Wikipédia.

Dans la Wikipédia en anglais, toute une série d'outils a été créée pour signaler et documenter l'usage des principales variantes de l'anglais. Comme mentionné plus haut, il existe une recommandation de guide stylistique spécifique à l'anglais canadien et une bannière à afficher dans les articles :

*This article is written in Canadian English, which has its own spelling conventions (colour, centre, travelled, realize, analyze), and some terms that are used in it may be different or absent from other varieties of English. According to the relevant style guide, this should not be changed without broad consensus.*

La page du guide stylistique canadien décrit dans le menu détail un ensemble de conventions et de consensus en vigueur sur la manière de nommer et de rédiger les articles qui concernent des sujets canadiens. Ces recommandations vont plus loin que de simples conseils stylistiques : elles proposent une interprétation « canadienne » des politiques et recommandations générales de l'encyclopédie. Ainsi, ce guide est un moyen crucial de localisation des savoirs : il institutionnalise la reconnaissance d'une particularité culturelle et linguistique des contenus canadiens, tout en étant le lieu d'exercice d'un certain pouvoir d'action des éditeurs et éditrices canadiens en termes de gouvernance sur les sujets qui les concernent.

Il n'existe pas d'équivalent à ce guide stylistique spécifiquement canadien dans la Wikipédia en français. La recommandation sur les conventions de style inclut simplement une section sur les « variantes géographiques » de vocabulaire, laquelle mentionne que « certains termes de la langue française pouvant varier selon le pays ou la région où ils

sont employés, on privilégiera la version la plus adaptée au contexte. Par exemple, dans un article lié spécifiquement au Québec, c'est le terme québécois qui sera privilégié». Dans tous les autres cas, c'est le français parlé en France qui serait privilégié. La Wikipédia francophone a tenté de se doter de conventions plus formelles sur la manière d'inclure les variantes linguistiques, mais cette tentative a échoué. Un sondage mené en 2008 a conclu qu'une majorité de Wikipédiens et Wikipédiennes pensait nécessaire de se doter de telles règles, mais ils n'ont pas réussi à s'entendre sur la nature de ces conventions. Seul le domaine des sports a bénéficié d'une prise de décision claire en 2006. Cette dernière établit que «les articles sur les clubs de sports nord-américains auront un titre francisé "à la canadienne"». Ainsi, le nom de l'article «Celtics de Boston» reflète la forme francisée à la québécoise, contrairement à l'usage en France où la forme anglaise *Boston Celtics* est plus courante. Mais de manière générale, la préséance de la variante linguistique française reste le choix par défaut, reflétant ainsi le poids politique et symbolique de la France dans la francophonie.

Dans le cas de la Wikipédia en atikamekw nehiromowin, c'est encore une fois un protocole original qui a été mis en place pour reconnaître et inclure les variantes de la langue. En effet, les trois communautés qui composent la nation ont chacune leurs propres usages linguistiques. Cette donnée a été prise en compte très tôt dans les discussions qui ont mené au lancement public de l'encyclopédie. Il a été convenu que les articles comporteraient trois sections différentes, chacune dédiée à une communauté qui serait libre d'y rédiger un contenu reflétant son point de vue et son usage de

la langue ou de fusionner les contenus. Par exemple, l'article « Kwekweciw » (« Mésangeai du Canada ») comporte une seule section fusionnée pour les communautés de Manawan et Wemotaci, et une autre pour Opitciwan. Cette structure des contenus reflète la structure politique de la nation, dans la mesure où Manawan et Wemotaci participent ensemble au Conseil de la nation Atikamekw, tandis qu'Opitciwan s'en est retiré.

### **Bilan : différences entre les Wikipédias**

Dans chacune des éditions linguistiques de Wikipédia, les Canadiens et Canadiennes négocient leur place parmi les autres nations en fonction de principes et règles communs à tous, et de recommandations spécifiques aux cultures nationales. Ces deux types de protocoles (communalité et localité) sont complémentaires dans la pluralisation culturelle. Ils offrent des cadres de négociation éditoriale entre des groupes aux perspectives divergentes, voire antagonistes, sans pour autant empêcher les conflits. La majorité de ces protocoles sont des addendas au principe cardinal de la neutralité de point de vue : ils viennent spécifier les conditions de ce qui est inclus et exclu de l'encyclopédie en prenant en compte les particularités des différents contextes culturels locaux.

En définitive, plusieurs différences majeures distinguent les Wikipédias en anglais, en français et en atikamekw nehiromowin. Les encyclopédies en anglais et en atikamekw nehiromowin se rejoignent dans une forme de pluralisation qui institue une forte expression des particularismes culturels locaux (surtout linguistiques) des divers groupes,

nations et pays. La Wikipédia en français institue le particularisme français comme standard dominant, avec en corollaire une faible reconnaissance officielle de l'expression singulière des autres cultures francophones. La Wikipédia en atikamekw nehiromowin se distingue par des protocoles nettement originaux comparativement aux autres encyclopédies : moins formels, ils légitiment l'autorité des sources orales, excluent les sujets sacrés et aménagent d'emblée différentes sections des articles pour y refléter les variantes locales de la langue.

En comparant les versions de l'encyclopédie en français et en anglais, on constate que la reconnaissance des variantes canadiennes de chaque langue est beaucoup plus formalisée du côté anglophone que francophone. Cette reconnaissance découle premièrement d'une règle claire selon laquelle aucune variété de l'anglais ne devrait avoir préséance sur les autres. Le choix d'une variante de la langue y est le plus souvent négocié à la pièce, au sein de chaque article, en fonction du sujet et du contexte de rédaction. Deuxièmement, il existe dans l'encyclopédie en anglais une claire reconnaissance des biais systémiques qui tendent à privilégier malgré tout un point de vue états-unien. Ces conditions ont permis l'élaboration d'un guide stylistique canadien, qui institutionnalise la particularité culturelle et linguistique des contenus relatifs au Canada.

À l'inverse, dans la Wikipédia francophone, la priorité est donnée à la France en termes de variante de langue à privilégier. Le terrain linguistique commun y est modelé sur la particularité française, selon une convention de moindre surprise pour une majorité de lecteurs et lectrices de France. Cette préséance donnée à la France est toutefois rarement

énoncée de manière explicite dans les recommandations officielles de l'encyclopédie francophone. Elle s'y applique le plus souvent par défaut. Ses pages de règles et recommandations évitent de reconnaître l'existence d'un « francocentrisme » dans l'encyclopédie. Par effet de retournement, c'est le fait de nommer un biais ethnocentriste en faveur de la France qui est jugé contraire à la neutralité.

De façon générale, dans la Wikipédia francophone, l'expression des variantes culturelles de la langue tend à être traitée comme un problème à encadrer, et non comme une condition de neutralité. S'il existe de rares dispositions établissant l'usage du français québécois dans quelques cas précis, la négociation se fait essentiellement lors de conflits d'édition, où le francocentrisme est abondamment discuté. Les enjeux politiques de la pluralisation culturelle tendent donc à être invisibilisés dans les pages officielles, et relégués à l'arrière-boutique de l'encyclopédie, dans les pages de discussion où les conflits d'édition ressurgissent périodiquement. Bien qu'il existe une volonté d'une majorité de contributeurs et contributrices francophones d'établir des protocoles plus clairs sur l'emploi des variantes culturelles de la langue, on observe à ce jour une incapacité à les formaliser en recommandations officielles.

Ces différences entre encyclopédies s'expliquent par un ensemble de facteurs internes et externes à Wikipédia. Le facteur qui semble le plus influent est d'ordre géopolitique : il tient au poids relatif des puissances nationales (ou locales pour l'atikamekw nehiromowin) dans chaque espace linguistique. La sphère culturelle anglophone est certes dominée par les États-Unis du point de vue économique et culturel, mais d'autres puissances majeures y cohabitent

néanmoins avec une certaine influence (notamment le Royaume-Uni, et à une moindre échelle le Canada et l'Australie). L'espace francophone est davantage unipolaire : la France y domine largement sur les autres pays et groupes infraétatiques, dont le Québec.

Ce facteur externe est redoublé par un facteur interne, lié à la démographie des Wikipédiens et Wikipédiennes : dans l'encyclopédie francophone, les contributeurs et contributrices français sont surreprésentés comparativement au nombre de lecteurs et lectrices de ce pays. Lors de prises de décision, les contributeurs et contributrices français ont ainsi l'avantage du nombre.

Enfin, ces différences s'expliquent par des cultures politiques distinctes. Chaque Wikipédia a ses propres habitudes de gouvernance, en bonne partie influencées par les philosophies politiques des différentes nations qui y participent : d'un côté le multiculturalisme anglo-saxon, de l'autre l'universalisme français. L'encyclopédie atikamekw nehiromowin est quant à elle influencée par une culture du consensus entre les différentes composantes de la nation. Ces différentes structures et cultures politiques ont une grande influence sur la manière dont s'organisent les rapports de force entre groupes minoritaires et majoritaires au sein de chaque encyclopédie.

Quelle que soit l'encyclopédie, les processus de pluralisation culturelle qui y ont cours restent marqués par des limites et des zones d'ambiguïté. En font partie la faible disponibilité des sources de référence au sujet des cultures minoritaires et les enjeux de représentation de la diversité culturelle interne au Canada. Plus fondamentalement, le statut de Wikipédia comme sphère publique globale est une

source d'ambiguïté pour la direction du projet encyclopédique. Vise-t-il à produire une encyclopédie internationale écrite en différentes langues, ou une collection d'encyclopédies des mondes anglophone, francophone et atikamekw nehiromowin, centrées sur leurs perspectives? Toujours en cours, ce débat promet d'animer encore longtemps les forums wikipédiens.

### **Remerciements**

L'autrice remercie Thérèse Ottawa pour sa traduction d'articles en atikamekw nehiromowin, ainsi que Pierre-Yves Beaudouin, Dipsacus fullonum, Tagishsimon, Jim Hayes, Tilman Bayer, Luc Patin et Étienne Beaulé pour leur partage d'informations.

## CHAPITRE 2

# Le portail Acadie de Wikipédia

### *Un portrait juste ?*

GABRIEL ARSENAULT ET MATHIEU WADE

La nation acadienne se laisse difficilement appréhender. Les individus et les institutions qui se réclament de ce peuple sans État n'ont pas tous la même réalité en tête. Les espaces-temps de l'Acadie font débat jusque dans le monde universitaire. Qu'en est-il alors dans Wikipédia ? Informé par les lignes de force qui structurent les études acadiennes, ce chapitre cherche à préciser les contours de l'Acadie telle qu'elle se donne à lire dans la version francophone de l'encyclopédie libre. Les auteurs, tous deux professeurs à l'Université de Moncton, une institution vitale du peuple acadien, ont créé en 2018 le Wikiclub Acadie pour stimuler la contribution aux articles consacrés à l'Acadie.

Les études montrent que le contenu disponible sur Wikipédia est globalement assez fiable (Giles, 2005). Mais les auteurs et autrices de la littérature savante portant sur Wikipédia déplorent que seule une faible minorité des

articles soient de haute « qualité ». La « qualité » d'un article (ou d'un portail) Wikipédia est jugée par la communauté wikipédienne, invitée à voter afin de lui attribuer une certification, selon différents critères (p. ex., le style d'écriture encyclopédique, la qualité des références, la clarté de l'organisation ou le choix des images). La certification la plus prestigieuse est celle d'« article de qualité », suivie par celle de « bon article ». En juin 2024, fr.wikipedia comptait 2 151 « articles de qualité » et 3 999 « bons articles », représentant respectivement seulement 0,08 % et 0,15 % de l'ensemble des articles de l'encyclopédie en français. Une étude se penchant sur les huit principaux Wikipédia révélait toutefois il y a quelques années que fr.wikipedia était le portail le plus exigeant en matière de certification de qualité des articles (Jemielniak et Wilamowski, 2017).

Les chercheurs et chercheuses universitaires s'intéressant à la qualité de l'encyclopédie reprochent aussi à ses articles de souffrir d'un important biais patriarcal. Martel et Villeneuve (dans le présent volume) rapportent ainsi que les biographies de femmes ne représentent que 18,8 % de l'ensemble des biographies sur fr.wikipedia (et 18,3 % sur en.wikipedia). Cette sous-représentation des femmes dans l'encyclopédie ne reflète pas seulement le patriarcat sociétal : elle démontre également un biais patriarcal au sein même de Wikipédia, les femmes devant être en moyenne plus éminentes que les hommes pour obtenir un article Wikipédia (Adams *et al.*, 2019; Wagner *et al.*, 2016). Selon une étude encore non publiée, citée dans Adams *et al.* (2019), si le même niveau de notoriété était requis des femmes que des hommes pour avoir un article Wikipédia, il y aurait 70 000 biographies de femmes de plus dans l'encyclopédie.

Pour des raisons encore mal comprises, les femmes ne représentent d'ailleurs possiblement qu'environ 10 % de l'ensemble des personnes contribuant à l'encyclopédie (voir encore Martel et Villeneuve dans le présent volume).

Nous proposons ici une réflexion sur les mérites et les limites de l'encyclopédie Wikipédia à partir d'un des quelque 1 800 portails de Wikipédia en français: le portail Acadie<sup>1</sup>. Avec ses 2 013 articles en date de juin 2024<sup>2</sup>, il n'existe nulle part ailleurs autant d'information aussi accessible sur l'Acadie. Le portail ne compte pourtant aucun article «de qualité», bien qu'il compte 24 «bons articles» (1,2 % du total). Bien qu'utiles, ces statistiques ne constituent que des indicateurs très superficiels de la qualité du portail. Comment aller plus loin ?

Notre analyse du portail Acadie, bien qu'elle porte sur une population minoritaire et globalement marginale – les Acadiens –, ne s'inscrit pas dans une réflexion sur les enjeux liés à la pluralisation et à l'internationalisation de Wikipédia (voir le chapitre de Casemajor dans ce volume). De nombreuses études se sont intéressées au processus de pluralisation des savoirs dans Wikipédia à partir d'une grille théorique postcoloniale et féministe (Bjork-James, 2021; Casemajor et Couture, 2020; Ford et Wajcman, 2017). Ces approches tendent à focaliser sur les conflits entourant la représentation des groupes minoritaires et sur les protocoles d'édition susceptibles de produire et de reproduire des biais systémiques leur étant défavorables. Ces situations

---

1. Pour une analyse du portail des Franco-Ontariens, voir Casemajor et Couture (2020).

2. Sauf indication contraire, toutes les données relatives au portail Acadie ont été extraites en juillet 2021.

renvoient à ce que Godrie et Dos Santos appellent des inégalités épistémiques, à savoir « un type particulier d'inégalité qui se manifeste dans l'accès, la reconnaissance et la production des savoirs et des différentes formes d'ignorance » (Godrie et Dos Santos, 2017, p. 7). Il serait possible d'étudier le portail Acadie en tant que lieu marginal où un groupe minoritaire lutte pour faire reconnaître la légitimité de sa culture et de son identité. Nathalie Casemajor et Stéphane Couture ont proposé une analyse du portail franco-ontarien qui va dans ce sens, en étudiant « les pratiques de savoir concernant les Franco-Ontariens et plus précisément [...] la négociation des savoirs concernant l'identité de ce groupe tel qui est représenté dans Wikipédia » (Casemajor et Couture, 2020). Cependant, ce n'est pas l'objectif de notre recherche.

Nous proposons plutôt de considérer le portail Acadie comme un lieu important de production de contenu sur l'Acadie, bien qu'il n'ait jusqu'ici jamais été étudié. Nous ancrons donc notre analyse dans le champ des études acadiennes afin de voir quelles réponses apporte le portail aux deux principales interrogations qui structurent ce champ : (i) L'Acadie renvoie-t-elle à un territoire relativement circonscrit des provinces canadiennes de l'Atlantique ou à une diaspora sans frontières ? (ii) Renvoie-t-elle à un territoire administratif disparu après le Grand Dérangement ou à une nation bien vivante ? La première interrogation est proprement politique et concerne les modalités d'appartenance au groupe. La question de savoir *où* se trouve l'Acadie, et par conséquent *qui* est Acadien ou Acadienne est fréquemment débattue tant au sein de la société civile qu'au sein de la communauté scientifique. Quelles réponses à ces questions

nous donne le portail Acadie si nous l'interrogeons ? La seconde interrogation relève davantage des recherches savantes. L'enjeu n'est pas de savoir si l'Acadie existe toujours, la réponse est évidemment positive. Cependant, l'histoire coloniale de l'Acadie et les sciences sociales acadiennes sont deux champs distincts qui fonctionnent en grande partie de façon parallèle. Le champ des études coloniales est davantage internationalisé et jouit d'une plus grande légitimité. Les sciences sociales acadiennes, qui portent généralement sur l'Acadie nationaliste qui émerge au XIX<sup>e</sup> siècle, sont quant à elles plus périphériques et locales. Parmi ces deux champs, lequel domine au sein du portail ? La prémisse de notre étude est que l'Acadie est un objet aux frontières géographiques et temporelles floues et que la définition de ces frontières constitue pour elle l'un des principaux enjeux politique et scientifique. Nous proposons une analyse systématique du portail afin de voir quelles représentations de l'Acadie s'en dégagent. Dans un premier temps, nous nous intéresserons à la géographie de l'Acadie proposée par le portail. Ensuite, nous analyserons les époques couvertes par les articles. Enfin, nous présenterons un portrait des articles les plus consultés et des principaux collaborateurs, démontrant que le biais patriarcal caractérisant l'encyclopédie dans son ensemble caractérise aussi le portail Acadie.

Une analyse du portail Acadie à l'aune des études acadiennes permet de confronter des interrogations théoriques et politiques à des réponses proposées par un espace collaboratif et permet de mettre en évidence un ensemble de biais, dont certains sont typiques au sein de Wikipédia, et d'autres sont spécifiques au contexte acadien.

## L'espace-temps acadien

Le portail Acadie a la particularité de concerner une entité aux frontières spatio-temporelles floues et dont la définition est sujette à de nombreux débats. «L'Acadie», ne possédant pas d'État ou de frontières administratives claires, peut prendre des sens divers. Ceux-ci donnent lieu à des champs scientifiques distincts, mais également à des controverses publiques portant sur le sens de l'acadianité, le rapport aux institutions, à la langue et au passé.

## L'espace acadien

En l'absence de frontières administratives formelles, l'Acadie, et tout particulièrement l'Acadie moderne, s'inscrit dans une géographie contentieuse et éclatée. Cette indétermination vaut à la fois pour l'époque coloniale, où les frontières sont approximatives et contestées (Desbarats et Greer, 2011), et pour l'époque moderne, où différentes conceptions de l'identité acadienne et modalités d'appartenance s'opposent. Les débats les plus vifs concernent toutefois l'époque moderne. Schématiquement, on peut distinguer une Acadie diasporique et une Acadie territoriale. Ces deux Acadie proposent des géographies considérablement distinctes, ainsi que des définitions de l'acadianité en partie incompatibles.

Pour certains, on ne saurait circonscrire un territoire acadien moderne. L'Acadie serait un peuple diasporique qui prend la forme d'un archipel de communautés, initialement dispersées par le Grand Dérangement, puis par des vagues d'émigration depuis le XIX<sup>e</sup> siècle (Bérubé, 1987; Magord et Belkhodja, 2005). On trouve des communautés acadiennes

de la Louisiane (Basque, 2009; Bruce et Urbain, 2021) à la Nouvelle-Angleterre, en passant par la France (Magord, 2010) et le Québec (Bergeron *et al.*, 2008). Cette Acadie se définit essentiellement par la généalogie et s'exprime aujourd'hui dans un mode surtout ponctuel et festif (Lefebvre, 2012). L'Acadie diasporique trouve son expression la plus éloquente dans le Congrès mondial acadien (CMA). Le CMA, lancé en 1994, met explicitement de l'avant une conception diasporique de l'identité acadienne et fait une place centrale aux réunions de famille, réaffirmant du coup une vision généalogique de l'acadianité. Il suscite aussi invariablement des débats au sein de la sphère publique acadienne des Maritimes (Allain, 1997; Bruce, 2018). L'Acadie diasporique mise davantage sur l'individu que sur un quelconque territoire et se compose davantage de réseaux et d'événements ponctuels que d'institutions.

Certains s'opposent à cette conception de l'Acadie, estimant qu'elle pose une menace au plein déploiement politique de la nation acadienne en diluant les effectifs et en mettant de l'avant une définition généalogique, plutôt que civique, de l'acadianité (Thériault, 1996). Pour plusieurs, l'Acadie renvoie plutôt à une volonté de « faire société » en français dans les provinces maritimes (Daigle, 1994; Thériault, 1995). Cette Acadie s'articule autour d'institutions francophones, permettant à la fois l'action collective et la transmission culturelle et linguistique. Elle est représentée par des organismes provinciaux (Société de l'Acadie du Nouveau-Brunswick, Fédération acadienne de la Nouvelle-Écosse, Société acadienne et francophone de l'Île-du-Prince-Édouard, Fédération des francophones de Terre-Neuve-et-Labrador) et régionaux (Société nationale de l'Acadie), et elle

se définit davantage par la langue et le territoire que par la filiation.

Par rapport à ce débat, la position du portail est nuancée. D'une part, qu'on s'intéresse aux articles sur les lieux, à la géographie des groupes et organismes retenus ou encore aux sites de naissance des personnalités du portail, le constat est similaire: les articles portent très majoritairement sur les provinces de l'Atlantique. C'est le cas de 77,3 % des articles sur les lieux, de 76,6 % des articles sur les personnalités et de 85 % des articles sur les groupes et organismes. D'autre part, force est d'admettre que le portail contient un nombre important d'articles relatifs à l'extérieur de la région atlantique, notamment de la Louisiane, du Maine et du Québec, ainsi que d'autres États américains, provinces canadiennes et régions françaises.

### **Le temps acadien**

Une autre controverse concerne la temporalité de l'Acadie. S'agit-il d'une entité historique ou contemporaine ? On peut schématiquement distinguer une Acadie coloniale et une Acadie moderne. Or les deux Acadie portent le même nom et font partie du même portail, contrairement à la Nouvelle-France et au Québec, par exemple, qui occupent des espaces distincts dans Wikipédia. L'Acadie coloniale renvoie à un territoire administratif correspondant en gros aux provinces maritimes et plus spécifiquement à la Nouvelle-Écosse actuelle. Établie en 1604, elle fut au cœur de nombreux conflits entre la France et la Grande-Bretagne jusqu'à la déportation, qui s'échelonna de 1755 à 1763, dans le contexte de la guerre de Sept Ans. La déportation marque une rupture

temporelle et spatiale importante pour l'Acadie. L'Acadie moderne fut construite par les rescapés de la déportation et s'est institutionnalisée à partir de la seconde moitié du xviii<sup>e</sup> siècle. Elle s'est par la suite organisée autour d'une idéologie et de symboles nationaux élaborés lors des Conventions nationales acadiennes à partir de la fin du xix<sup>e</sup> siècle. Ces deux Acadie constituent des champs d'étude scientifique distincts. L'étude de l'Acadie coloniale s'inscrit aujourd'hui dans des réseaux de recherche internationaux, majoritairement de langue anglaise (Farragher, 2005 ; Griffiths, 2004 ; Hodson, 2012 ; Kennedy, 2014 ; Reid, 2004), alors que les travaux scientifiques portant sur l'Acadie moderne sont généralement produits en français dans les provinces maritimes (Belliveau, 2014 ; Boudreau, 2016 ; Landry, 2015 ; Poplyansky, 2018 ; Thériault, 1995).

Le portail Acadie se positionne-t-il par rapport à cette dichotomie ? La classification du portail lui-même sur fr.wikipedia indique clairement une adhésion à la conception de l'Acadie comme colonie du passé. Le portail Acadie est classé de deux manières au sein de la liste des portails de fr.wikipedia. Selon la première catégorisation, l'Acadie se trouve dans la section « Histoire » et dans la sous-section « Période moderne (xvi<sup>e</sup> au xviii<sup>e</sup> siècle) », à l'instar du portail sur la Nouvelle-France. Selon la seconde catégorisation, l'Acadie se trouve dans la section « Géographie » et dans la sous-section « Territoires historiques », encore une fois à l'instar de la Nouvelle-France.

De la même façon, la déportation est sans équivoque l'événement de l'histoire acadienne attirant le plus l'attention des internautes.

Pourtant, une très forte majorité des personnalités retenues dans le portail sont nées à partir de 1764 et appartiennent donc à l'Acadie moderne plutôt qu'à l'Acadie coloniale (1604-1763). En clair, le portail se fait l'écho du débat entourant les différentes temporalités acadiennes. Notons par ailleurs le clivage entre l'offre, correspondant davantage à la définition moderne de l'Acadie, et la demande, nettement plus tournée vers sa définition coloniale.

## Le genre

Sur le plan de la distribution sexuée des personnalités au sein du portail, on observe, à l'instar de Wikipédia de façon générale, une surreprésentation d'hommes. Sur les 679 personnalités du portail, 569 sont des hommes (83,8 %), 106 sont des femmes (15,6 %) et 4 s'identifient comme queers ou non binaires (0,6 %).

Il apparaît difficile de ne pas faire l'hypothèse d'un lien causal entre cette domination masculine sur le plan du contenu et celle, aussi typique de l'encyclopédie Wikipédia dans son ensemble, qui caractérise la production des articles. Aucune des principales personnes contribuant au portail ne s'identifie comme femme. En effet, on recense 18 personnes figurant comme auteur principal pour au moins 10 articles du portail : parmi elles, 12 s'identifient comme homme, alors que 6 ne révèlent pas leur genre. On constate en outre le rôle prépondérant de Red Castle, contributeur principal de 39 % des articles du portail et de 19 de ses 24 « bons articles ».

## Savoirs légitimes et « bons articles »

Plusieurs études qui abordent Wikipédia par un angle critique et postcolonial font état des inégalités épistémiques subies par les groupes minoritaires et marginalisés (Godrie et Dos Santos, 2017). Les savoirs produits par ces groupes ne seraient pas forcément reconnus comme étant légitimes. Ces communautés seraient ainsi soumises à des productions discursives exogènes, elles seraient contraintes à se représenter avec des catégories qui ne sont pas les leurs. Le concept d'exiguïté, initialement développé par François Paré (1992) pour comprendre la production littéraire francophone minoritaire au Canada, et appliqué ensuite aux sciences sociales par Mourad Ali-Khodja (2003), décrit précisément ce phénomène en francophonie canadienne. Selon cette théorie, les savoirs exogènes sur l'Acadie devraient être dominants au sein du portail.

Le champ des études acadiennes est divisé en deux : les études coloniales, produites en anglais, davantage internationalisées et légitimes, et les sciences sociales acadiennes contemporaines, produites en français et plus marginales. Or, malgré sa légitimité plus importante, la domination des études coloniales dans le portail Acadie n'est pas claire. D'une part, il est vrai que la description du portail correspond à une définition coloniale de l'Acadie, mais l'Acadie contemporaine y est nettement plus représentée dans son contenu – qu'on se fie à la temporalité aussi bien qu'à la géographie des personnalités retenues, ainsi qu'à la géographie des lieux et des groupes et organismes figurant dans le portail. D'autre part, les 24 « bons articles » du portail portent pour la grande majorité sur la période contemporaine

plutôt que sur la période coloniale. On ne peut en effet pas parler d'inégalité épistémique. Le rapport aux savoirs qui se dégage du portail est considérablement plus nuancé et complexe.

### **Comment redynamiser le portail Acadie ?**

Le portail Acadie donne à voir une Acadie à la fois coloniale et moderne, diasporique et ancrée dans les provinces de l'Atlantique. Cette tension constitutive de l'Acadie contemporaine est très clairement exprimée par le portail. Il nous apparaît significatif que les personnalités acadiennes retenues dans le portail appartiennent très majoritairement à l'époque moderne, alors que la quasi-totalité des événements de l'histoire acadienne suscitant l'intérêt des internautes concerne l'époque coloniale. Le relativement faible intérêt porté aussi bien par les contributeurs et contributrices que par les internautes à des événements qu'on pourrait associer à la « Révolution tranquille » acadienne, comme la mise en œuvre dans les années 1960 du programme du gouvernement de Louis J. Robichaud, *Chances égales pour tous* (faisant actuellement l'objet d'un article assez superficiel), apparaît quand même étonnant. En quantifiant cette absence relative d'articles portant sur des moments phares de la modernité politique acadienne, notre étude fournit des données inédites à l'appui d'une critique sociologique de l'Acadie déjà bien articulée : s'il y a bien aujourd'hui des individus acadiens, l'agentivité du sujet politique acadien est bien faible. En l'absence de frontières politiques clairement définies, l'Acadie est contrainte d'agir par l'intermédiaire d'un réseau associatif diversifié, mais néanmoins précaire (Allain, 2003).

L'agir politique acadien s'en trouve ainsi fragmenté. Et en l'absence d'institutions représentatives, le discours acadien est divisé en organismes sectoriels marqués par un certain corporatisme (Landry, 2015). Enfin, les organismes acadiens se trouvent en situation de dépendance économique vis-à-vis de l'État canadien, limitant la marge de manœuvre des instances représentatives de la communauté (Léger, 2012). Pour plusieurs, cette fragmentation et cette dépendance minent la capacité de l'Acadie à se projeter dans l'avenir, à se donner des projets collectifs et à se penser comme sujet politique. Cette difficulté à se représenter comme sujet politique structure en partie le contenu du portail.

Aussi, le portail affiche deux grands défauts caractérisant malheureusement de manière générale le contenu sur Wikipédia : la quasi-totalité des articles en est encore à l'état d'ébauche et l'Acadie mise en valeur est dominée par les hommes. Pour améliorer la qualité des articles du portail et le rendre plus représentatif de la diversité de la population acadienne, une solution s'impose : stimuler la participation de la société civile au développement du portail Acadie. Des efforts de mobilisation collective, tels les ateliers wiki mensuels de Bibliothèque et Archives nationales du Québec, ou encore l'initiative de la municipalité de Monmouth, dans le Pays de Galles – consistant à mettre en place quelque 1 000 codes QR liant à un article Wikipédia différents sites de la ville (Kooser, 2012) –, nous ont ainsi inspirés à créer le Wikiclub Acadie. En octobre 2018, suivant une conférence de Patrick Degrâce, alias Red Castle, à l'Université de Moncton, un premier Wikiclub informel se met en place (Radio-Canada, 2018; Doiron, 2018). Entre l'automne 2018 et le printemps 2019 se tiennent une demi-douzaine de

séances de rédaction réunissant des membres de la communauté universitaire. Nous avons toutefois alors conclu que sans structure plus formelle, le « club » n'était pas destiné à un grand dynamisme.

C'est pour sortir de cette impasse qu'à l'été 2020, nous avons établi un partenariat avec la Société nationale de l'Acadie (SNA). Grâce à son appui, nous avons publié un rapport de recherche sur le portail Acadie (Arsenault et Wade, 2020) visant notamment à cerner les priorités pour les futurs contributeurs et futures contributrices. Le rapport contient, par exemple, une liste de 40 personnalités féminines acadiennes, identifiées par l'historien Maurice Basque, qui pourraient faire l'objet d'une biographie sur Wikipédia. À la suite de la publication de ce rapport, la SNA prend formellement en main en septembre 2020 le projet d'un Wikiclub (Boudreau, 2020). Au printemps 2021, des ateliers de formation sont donnés auprès des membres du Regroupement féministe du Nouveau-Brunswick et de la Fédération des femmes de la Nouvelle-Écosse. À l'été 2021, la SNA reçoit de Patrimoine canadien une importante subvention, qui lui permet notamment d'organiser une série d'activités virtuelles en mars 2022 ciblant spécifiquement la communauté de l'Université de Moncton : une table ronde, un atelier de formation et une séance de contributions (Radio-Canada, 2022). La SNA s'engage pour les années à venir à ce que Wikiclub Acadie garde son dynamisme, en organisant d'autres activités de formation dans les communautés francophones des quatre provinces de l'Atlantique. Espérons que cette mobilisation sera fructueuse et aboutira à une bonification concrète du portail.

## CHAPITRE 3

# La quantification du fossé des genres dans les ressources biographiques avec Wikidata

MARIE D. MARTEL ET SIMON VILLENEUVE

À l'aide de la base de connaissances libre Wikidata, les auteurs ont analysé le contenu biographique en ligne de dizaines de dictionnaires et d'encyclopédies afin d'établir la proportion d'articles dédiée aux personnalités féminines et masculines et de quantifier la présence relative du fossé des genres (*gender gap*). Ce chapitre présente les résultats portant sur 80 publications totalisant un peu plus de 2 millions d'articles biographiques. En moyenne, il est possible de dénombrer une proportion de 14 % d'articles sur des personnalités identifiées comme femmes et 86 % sur des personnalités identifiées comme hommes.

En tant que source d'information, Wikipédia appartient à la catégorie des ouvrages de référence de type encyclopédique. L'encyclopédisme vise à « rassembler le plus grand nombre possible de savoirs, les mettre en relation, les transmettre, les partager, les soumettre à la discussion » (Melançon 2018). Depuis 2001, l'encyclopédie Wikipédia, qui est sous licence libre, a contribué à mettre en place le projet et le mouvement Wikimedia, lequel a engendré un écosystème de connaissances libre en ligne. Ces contenus permettent aujourd'hui de partager et de quantifier certaines informations qui étaient autrefois inédites, confidentielles ou disparates et de discuter à leur sujet.

L'ancrage culturel, les idéologies, les valeurs façonnent les encyclopédies de même que la représentation de la réalité et les connaissances qu'elles mobilisent ou légitiment (Rey, 2007 ; McDowell et Vetter, 2018). À cet égard, comme tant d'autres encyclopédies, Wikipédia est aussi réputée pour perpétuer des biais sexistes tant en matière de participation éditoriale que de couverture biographique, ce qui n'a pas manqué d'attirer l'attention des médias (Cohen, 30 janvier 2011 ; Nadeau, 4 mars 2017). Sans pour autant proposer de définition, Noam Cohen, dans le *New York Times*, suggère de définir le concept de « fossé des genres » à partir du cas de Wikipédia, qui tend à décourager les contributrices. Plus précisément, le fossé des genres, dans ce contexte, fait référence aux disparités en matière de participation, de contenu et de représentation entre les genres sur les différentes plateformes wikimédiennes. Dans l'article *Fossé des genres* sur Méta-Wiki, qui sert de plateforme de discussion pour les différents projets wikimédiens, deux types de fossé des genres sont décrits : « (a) un fossé des genres du contenu :

Dans l'espace principal de contenu de nos wikis, plus de contenus sont dédiés aux hommes qu'aux femmes et (b) un fossé des genres de la participation : plus d'hommes que de femmes participent à la communauté Wikimedia créatrice de contenus».

Le nombre d'éditrices sur Wikipédia à travers les différentes versions linguistiques serait actuellement de 13 % en moyenne alors que celui des genres autres qu'homme et femme (cisgenre) serait de 4 % selon le Community Insights Report 2023 (Wikimedia Foundation, s.d.). Il semble également que, entre 2019 et 2022, la proportion des éditeurs actifs de genres autres qu'homme et femme (cisgenre) a doublé. Pendant ce temps, les proportions d'éditeurs hommes et femmes sont restées globalement les mêmes durant cette période, bien qu'il y ait eu une augmentation significative du nombre d'éditrices depuis 2019. Selon plusieurs études menées sur cette question, l'écart de la participation entre les genres se reflète dans les articles biographiques et leur contenu.

Dans ce chapitre, nous nous intéressons au fossé des genres touchant le contenu. Nous souhaitons présenter une méthode pour documenter les descriptions du genre à partir d'un échantillon de 80 ressources biographiques qui sont fortement liées sur Wikidata, la base de connaissances qui alimente Wikipédia. Cet exercice nous permettra de produire de nouvelles données sur le pourcentage de biographies consacrées aux femmes dans ces ressources, incluant des ouvrages de référence canadiens notables, ainsi que dans les projets Wikimedia tels que Wikipédia en français et en anglais, de même que Wikidata. L'analyse de ces données fera émerger certains constats, basés sur des données probantes,

au sujet de la couverture biographique dans ces ressources en plus d'une interprétation actualisée du fossé des genres dans les ouvrages de référence. Cette quantification des articles biographiques est susceptible de contribuer à une meilleure compréhension du fossé des genres et à une conscience plus grande des inégalités de genre et du sexisme systémique dans le monde documentaire en particulier et la société en général.

### **La diversité de genre en question**

Les préoccupations concernant la diversité sociale dans le projet Wikipédia, laquelle se manifeste par un déséquilibre dans la couverture thématique, ont été soulevées assez tôt par la communauté wikipédienne. Les participants et participantes au projet Wiki Countering Systemic Bias, lancé en 2004, s'inquiètent de la manière dont les intérêts et le profil sociodémographique des éditeurs et éditrices pourraient peser sur le contenu et engendrer des biais :

Le projet Wikipédia souffre de biais systémiques qui découlent naturellement des groupes démographiques de ses contributeurs, se traduisant par une couverture déséquilibrée de certains sujets, laissant ainsi des groupes démographiques moins représentés sans couverture adéquate [...]. Ce projet vise à éliminer les lacunes, en termes de perspective culturelle, engendrées par ces biais systémiques, en se concentrant consciemment sur des sujets et des points de vue négligés dans l'ensemble de l'encyclopédie. [traduction]

Un des sous-groupes du WikiProject Countering systemic bias, le Gender gap task force, est actif depuis 2013, le WikiProject Gender studies depuis 2005 et le WikiProject

Women in Red depuis 2015. Ces projets visent, à travers diverses formes d'engagement, à éliminer les biais systémiques de genre. Au moins deux de ces projets ont des groupes correspondants dans Wikipédia en langue française<sup>1</sup>.

Dans un article publié en 2020, Pierre-Yves Beaudouin, qui est alors administrateur de Wikimedia France, rapporte que Wikipédia en langue française compte 591 491 biographies, dont 18 % sont consacrées aux femmes. Il précise que l'information concernant plusieurs dictionnaires et encyclopédies dont les contenus sont harmonisés avec Wikipédia et Wikidata (la base de données qui alimente Wikipédia) permet de présenter des statistiques concernant la proportion de biographies dédiées à des femmes. Quelques exemples indiquent que la performance de Wikipédia en matière de biais de genre se compare à celle d'autres ressources biographiques de façon avantageuse dans certains cas; mais toutes, précise l'auteur en guise d'explication, y compris Wikipédia, sont tributaires des inégalités existant dans la société. Il soutient qu'un des mérites de Wikipédia, et plus spécialement de Wikidata, réside dans l'opportunité technologique de quantifier les biographies de plusieurs ressources en produisant des statistiques sur les écarts de genre dans les contenus biographiques.

Du côté de Wikipédia en anglais, les chiffres s'avèrent comparables, quoique légèrement supérieurs: en juin 2024, on retrouve 398 298 biographies consacrées aux femmes, ce

---

1. Le projet Les sans pagEs est la version en langue française de WikiProject Women in Red; le projet Femmes est la version correspondant au WikiProject Gender studies. Les projets WikiProject Countering systemic bias/Gender gap task force et WikiProject Countering systemic bias n'ont pas d'équivalent en français.

qui représente 19,8% des 2 006 964 biographies de cette encyclopédie à ce moment sur Humaniki qui calcule le pourcentage des genres parmi toutes les biographies des différents projets<sup>2</sup>.

À partir d'un échantillon de 80 ressources, dictionnaires et encyclopédies, incluant des ouvrages canadiens, nous souhaitons montrer comment Wikipédia et Wikidata peuvent enrichir notre compréhension d'enjeux sociaux, comme les inégalités de genre, malgré les limites qui les caractérisent en termes de biais de genre. Cette exploration quantitative sera conduite en investiguant les questions suivantes.

1. Comment pourrait-on utiliser Wikidata pour quantifier les contenus de sources biographiques telles que les dictionnaires et les encyclopédies ?
2. Quelles nouvelles données peut-on produire concernant la proportion de biographies consacrées aux femmes au sein de Wikipédia et d'autres ressources biographiques, grâce à Wikidata ?
3. Quelles interprétations et comparaisons peut-on dresser au sujet de ces ressources biographiques en considérant, plus spécifiquement, les données liées aux publications canadiennes, et qu'est-ce que cette quantification nous apprend sur le fossé des genres dans ces ouvrages de référence ?

---

2. *Humaniki: Exploring Wikimedia demography*. <https://humaniki.wmcloud.org/>.

Notre recherche se situe dans le prolongement des travaux sur le contenu de Wikipédia et porte plus particulièrement sur le fossé de contenu lié à l'asymétrie de genre dans la couverture biographique des ouvrages de référence. À notre connaissance, aucune étude rigoureuse sur l'évaluation comparative de la couverture biographique des ouvrages de référence n'a été entreprise depuis les travaux de Reagle et Rhue (2011). De plus, nous proposons une méthode qui diffère de l'approche utilisée par Reagle et Rhue, laquelle repose principalement sur la méthode de la liste de vérification. Cette dernière conduit les chercheurs à comparer la couverture en termes absolus des différentes sources biographiques, soit le ratio et le nombre respectif total de biographies d'hommes et de femmes. Elle leur sert aussi à comparer le comportement de Wikipédia et de *Britannica* relativement à différentes listes préétablies de personnes notables.

Notre approche, dans la foulée de recherches récentes (Klein et Konieczny, 2015a; 2015b; Konieczny et Klein, 2018), s'appuie sur la capacité de Wikidata de mesurer l'écart entre les genres dans les sources biographiques qui sont harmonisées sur cette dernière. Comme nous le décrivons dans la section suivante, cette méthode nous permet d'évaluer le fossé des genres dans un ensemble plus étendu d'encyclopédies et de dictionnaires que Reagle et Rhue ne l'ont fait. Ces ouvrages proviennent de divers pays, dont le Canada. Cette méthode quantifie l'écart de manière relativement plus précise en comparant ces sources biographiques entre elles. Nous soulignons enfin la rareté de la production de la recherche en français dans ce secteur auquel nous contribuons ici.

## Une méthodologie basée sur Wikidata

### Comment utiliser Wikidata pour quantifier les contenus de sources biographiques (dictionnaires ou encyclopédies) ?

La proposition méthodologique constitue une première réponse aux questions posées dans le cadre de la présente recherche. Pour mener celle-ci, nous nous sommes intéressés au genre associé aux sujets des publications biographiques de dictionnaires et encyclopédies possédant des identifiants uniques (UID) qui sont rattachés à des propriétés spécifiques dans la base de connaissances libre Wikidata<sup>3</sup>. Le contenu de Wikidata étant publié sous licence libre zéro (CCo), il est aisé de le parcourir, de le trier, de le réutiliser, de le modifier et de le publier de nouveau, selon des critères spécifiques à l'aide, notamment, du langage de requête SPARQL.

Ainsi, une première requête nous permet, en juin 2024, de recenser 465 propriétés distinctes reliées à des UID de publications contenant des articles biographiques grâce à leur classification utilisant l'élément de propriété Wikidata relative aux encyclopédies (Q55452870) ou aux sous-catégories de celui-ci. En éliminant d'abord les faux positifs manifestes,

---

3. En effet, les quelque 125 millions d'éléments (Q) de Wikidata, dont plus de 10 millions dédiés aux humains, sont reliés entre eux par une ou plusieurs des plus de 10 000 propriétés (P) sous la forme de triplets RDF. De ces propriétés, plus de la moitié sont consacrées à des UID de publications externes. Ainsi, selon l'exemple précédent, sur Wikidata, la propriété P5395 est dédiée aux identifiants uniques des entrées l'Encyclopédie canadienne (EC). Cette propriété permet donc d'associer chaque notion de l'EC à l'élément équivalent de la base de connaissances libre. Par exemple, l'entrée Émile Nelligan de l'EC est associée à l'élément Q2392492 de Wikidata (élément dédié au poète) à l'aide de la propriété P5395.

les doublons et les publications centrées sur le genre, nous avons sélectionné des publications dont l'échantillon compte au minimum 1000 entrées en essayant d'assurer une couverture géographique mondiale. Ce critère quantitatif nous paraît représenter un certain seuil d'impact documentaire sur le Web ainsi qu'un échantillon statistique significatif.

Une seconde requête SPARQL sur Wikidata a été appliquée pour mesurer le fossé des genres biographiques. Notre sélection finale comprend donc 80 publications issues de tous les continents, et pour lesquelles nous avons réalisé une évaluation quantitative du genre des biographies.

Il faut noter que pour la majorité des ressources concernées, l'échantillon associé à Wikidata qui les rassemble ne permet pas de quantifier de manière *exacte* le fossé des genres biographiques. Cependant, le suivi de l'évolution d'une cinquantaine de cas entre mars 2019 et avril 2021 révèle une tendance lourde sans changement significatif. Enfin, nous supposons que l'association des échantillons analysés n'a pas été faite expressément selon un critère de genre.

Le nombre total d'articles biographiques de chaque publication correspond généralement à celui donné par le site officiel de la publication. Ajoutons que si cette information est habituellement disponible pour les publications spécialisées en biographies, il en est autrement des publications généralistes. Pour la plupart de celles-ci, des données ont été déduites en explorant le moteur de recherche interne des sites concernés. Quelques autres ont été fournies par des personnes représentant les publications après les avoir sollicitées par courriel.

L'intervalle d'incertitude ne tient compte que de l'échantillon associé et représente une valeur maximale avec un

niveau de confiance de 95 %. Ainsi, pour les échantillons correspondant à 100 % du contenu d'une publication donnée, il n'y a pas d'incertitude sur les pourcentages. Les résultats sont présentés dans la section suivante.

Le contenu entier des trois wikis étudiés est disponible pour l'analyse à l'aide d'autres requêtes SPARQL et du site Humaniki. Les chiffres donnés sont exacts, mais toujours provisoires considérant qu'il s'y ajoute du nouveau contenu biographique chaque *heure*. Ainsi, les chiffres donnés ici pour Wikipédia en français et en anglais sont tirés du contenu en juin 2024, alors que ceux de Wikidata proviennent du site Qlever utilisant une version (ou *dump*) de Wikidata réalisé le 6 juin 2024.

## **Des données plus précises concernant les biographies de femmes**

**Quelles nouvelles données peut-on produire, grâce à Wikidata, concernant la proportion de biographies consacrées aux femmes au sein de Wikipédia et d'autres ressources biographiques ?**

L'objectif empirique de notre étude est d'évaluer le fossé des genres biographiques en considérant simultanément le nombre d'articles associés à des individus identifiés comme femme (F) et homme (H) dans l'échantillon de 80 ressources biographiques, encyclopédies et dictionnaires retenus. Ces ressources totalisent plus de 2 millions d'articles biographiques (2 014 479 articles).

Ces premiers résultats présentent ainsi des statistiques sur la répartition des genres dans ces diverses ressources biographiques en mettant en évidence le pourcentage d'ar-

ticles associés à des individus identifiés comme F et H. L'échantillon présente en moyenne une proportion de 14 % d'articles biographiques sur des personnalités F et de 86 % sur des personnalités H.

Les deux plus importantes ressources biographiques, soit *Deutsche Biographie* (622 800 entrées) et *China Biographical Database Project* (417 931 entrées), rassemblent 1 040 732 articles, soit près de 50 % des articles de l'échantillon. Or, le fossé des genres est relativement moins accentué dans le cas de ces ressources – comparativement à d'autres – et surtout en ce qui concerne le *Deutsche Biographie* qui présente 20,8 % de biographies F. Si on exclut ces deux publications afin d'avoir une idée plus représentative de l'ensemble des publications analysées, nous obtenons un fossé des genres encore plus creux, avec 11 % de biographies de personnalités F et 88 % de biographies de personnalités H.

**Quelles interprétations et comparaisons peut-on dresser au sujet de ces différentes ressources biographiques en considérant plus spécifiquement les données liées aux publications canadiennes, et qu'est-ce que cette quantification nous apprend sur le fossé des genres dans les ouvrages de référence ?**

D'abord, quelques remarques au sujet des données présentées dans le tableau complet des ouvrages de référence de l'échantillon que l'on peut consulter sur la plateforme en ligne qui accompagne le livre. Les données encyclopédiques sur les biographies dans Wikidata varient d'une ressource à l'autre ainsi qu'en fonction d'un historique de développement allant de la saisie de données manuelles à l'importation

en lot des bases de données qui sont de plus en plus répandues. En ce qui concerne l'*Encyclopédie canadienne* (EC) et le *Dictionnaire biographique du Canada* (DBC), nous avons réalisé une première analyse du contenu à l'automne 2018<sup>4</sup>. Les chiffres ont peu évolué depuis, le contenu associé à Wikidata de ces publications n'ayant que peu progressé. Le fossé des genres constaté a ainsi varié d'environ 1 % pour les deux.

À noter également que nous analysons le contenu biographique de l'*Encyclopædia Britannica* (EB) et de l'*Encyclopædia Universalis* (EU) associé à Wikidata depuis décembre 2017<sup>5</sup>. Les proportions données pour ces deux publications sont exactes au pourcentage près. Précisons que le nombre d'éléments associés à des identifiants de l'EB est beaucoup plus grand que le nombre de biographies dévoilé par les personnes représentant cette publication. Cela s'explique par l'association sur Wikidata des redirections d'URL de l'EB.

Certaines bases de données montrent une plus grande disparité de genres, comme le *Biographical Directory of the United States Congress* avec seulement 3 % de biographies F, contre 97 % de biographies H. C'est aussi le cas du *Dizionario biografico degli Italiani* qui a une faible proportion de biographies F, équivalant à 4 %. D'autres offrent une représentation un peu plus équilibrée comme le *Dictionary of New Zealand Biography* avec 27 % de biographies F, et 73 % de biographies H. Un autre exemple, l'*American National*

---

4. Villeneuve, S. (s. d.). *PoV – Wikipédia*. [https://fr.wikipedia.org/wiki/Utilisateur:Simon\\_Villeneuve/PoV#Quantifier\\_les\\_centrismes\\_de\\_l'EC\\_et\\_du\\_DBC](https://fr.wikipedia.org/wiki/Utilisateur:Simon_Villeneuve/PoV#Quantifier_les_centrismes_de_l'EC_et_du_DBC)

5. Villeneuve, S. (s. d.). *PoV – Wikipédia*. [https://fr.wikipedia.org/wiki/Utilisateur:Simon\\_Villeneuve/PoV#Quantifier\\_les\\_centrismes\\_de\\_Britannica\\_et\\_Universalis](https://fr.wikipedia.org/wiki/Utilisateur:Simon_Villeneuve/PoV#Quantifier_les_centrismes_de_Britannica_et_Universalis)

*Biography* dispose de 16 % d'articles portant sur des personnalités F et 83 % sur des personnalités H; ce qui est au-dessus de la moyenne des biographies F contenues dans les ressources de l'échantillon.

Plus de la moitié de ces publications (44 sur 80, soit 55 %) contiennent 10 % et moins de biographies F. Le *Dictionnaire biographique du Canada* qui en compte 6 % se trouve dans cette catégorie. Un peu plus d'un tiers des ressources biographiques (25 sur 80) en possède entre 11 % et 19,9 %.

Aucune ressource ne présente 50 % d'articles biographiques F et seulement 10 en contiennent 20 % et plus. C'est *Babelnet*, *TDV İslam Ansiklopedisi* et le *Dictionary of New Zealand Biography* avec respectivement 45 %, 37 % et 27 % de biographies F; ces ressources montrent la couverture biographique qui apparaît, de façon relative, la moins déséquilibrée de l'ensemble. La seconde ressource canadienne qui figure dans l'échantillon, soit l'*Encyclopédie canadienne* (EC), appartient à cette catégorie avec 21 % de biographies F.

Comment se comportent les trois ressources biographiques traditionnelles qui figurent communément dans les listes d'ouvrages de référence institutionnels français et anglais ou américains, soit l'*Encyclopædia Britannica*, l'*Encyclopædia Universalis* et l'*Encyclopédie Larousse*? L'*Encyclopædia Britannica* (45 296 éléments associés) montre une sous-représentation des entrées portant sur des personnalités F (11 %) par rapport aux personnalités H (89 %). L'*Encyclopædia Universalis* (17 589 éléments) présente également une forte disparité de genre, avec une faible proportion d'articles portant sur des personnalités F (9 %) et 91 % sur des personnalités H. L'*Encyclopédie Larousse* (3 164 éléments) affiche une proportion similaire à celle d'*Encyclopædia Universalis*,

avec 9 % d'articles biographiques F et 91 % d'articles biographiques H, soulignant un déséquilibre important entre les genres. Plus généralement, ces trois publications ont des résultats sous la moyenne (14 %) en termes de pourcentage de biographies F.

Par ailleurs, le pourcentage indiquant le fossé des genres dans les deux sous-classes de l'échantillon correspondant aux ressources biographiques de langue française et anglaise est relativement similaire : 14 % de biographies de personnalité F pour les premières et 13 % pour les secondes.

Les trois wikis hébergés par la Fondation Wikimedia analysés se situent en haut de la moyenne du pourcentage d'articles biographiques F (14 %) dans les ouvrages encyclopédiques de l'échantillon. Wikidata se place en tête des trois avec un pourcentage de 21 % de contenus biographiques F par rapport à Wikipédia en français et en anglais (20 %); ceci est moindre toutefois que *Babelnet* (45 %), *TDV İslam Ansiklopedisi* (37 %) et *le Dictionary of New Zealand Biography* (27 %). On note également que, en nombre absolu, la couverture biographique de Wikipédia en français (712 906) est largement supérieure à la couverture de chacune des ressources biographiques françaises, et même à la plus considérable, qui est *Who's Who in France* (24 513); à l'instar de Wikipédia en anglais (2 006 964) pour les ressources anglaises, qui dépasse de loin *Britannica*, qui compte 45 296 entrées.

Nous constatons enfin que le pourcentage moyen de biographies F dans les publications analysées (14 %) se rapproche du pourcentage de participantes dans les projets Wikimedia, qui oscille entre 11 et 15 %. En d'autres termes, il semble que le fossé des genres biographique des publications analysées présente le même ordre de grandeur que le

fossé des genres des contributrices et contributeurs aux projets Wikimedia.

### **Quelques éléments de réflexion et de discussion**

Plusieurs recherches sur le biais de genre se sont intéressées à la participation, d'autres au contenu, certaines aux deux à la fois. Dans notre étude, nous avons exploré le fossé des genres dans le contenu d'un ensemble de ressources biographiques de type encyclopédique. À la différence des travaux précédents sur la couverture biographique de ces dictionnaires et encyclopédies, le fossé des genres a été quantifié en ayant recours à Wikidata, la base de données en ligne qui supporte Wikipédia. Grâce à cette méthode s'appuyant sur les données massives, nous avons pu mettre à jour la situation en matière de genre de 80 ressources biographiques totalisant plus de 2 millions d'articles, et ce, de manière rigoureuse, efficiente et robuste – des résultats qu'aucune méthode auparavant n'avait permis de produire. Ainsi, et comme le reconnaissaient déjà Reagle et Rhue, il n'y avait pas eu encore, jusqu'à ce jour, de comparaisons à grande échelle de la couverture biographique et du genre dans les ouvrages de référence, incluant Wikipédia, en 2011. Or, cette affirmation est toujours valable et c'est cette lacune que notre recherche vise à combler. À noter que les travaux de Reagle et Rhue avaient cependant comparé Wikipédia et *Britannica* avec six autres ressources biographiques couvrant quelques milliers d'articles.

Du point de vue méthodologique, le critère de sélection pour le choix des ressources biographiques que nous avons retenu – un seuil minimal de 1000 entrées – a contribué à

réduire de manière significative l'échantillon à celles ayant le plus d'impact documentaire. Il a aussi assuré une plus grande précision des résultats, puisque l'alignement de ces ressources sur Wikidata était dans la plupart des cas très prononcé. Le nombre de ressources biographiques et d'articles impliqués dans l'échantillon a ainsi permis de détecter et de mieux établir une tendance qu'il serait difficile de nier désormais au sujet du fossé des genres et de la présence de biais sexistes systémiques à l'œuvre dans les ouvrages de référence de type encyclopédique. Les principaux résultats quantitatifs qui ont mis en évidence ces inégalités sociales de genre dans le domaine des savoirs encyclopédiques sont de trois ordres et portent sur :

1. **Le fossé des genres dans les ouvrages de référence de type encyclopédique.** La couverture biographique du genre indique une sous-représentation considérable des personnalités F, à l'échelle mondiale, avec une moyenne de 14 % d'articles biographiques concernant des personnalités féminines. En outre, plus de la moitié des ressources compte 10 % et moins de biographies F. On constate enfin que près de 88 % (87,5 %) des ressources encyclopédiques présentent moins de 20 % de biographies F.
2. **Le fossé des genres dans les ouvrages de référence de type encyclopédique en anglais et en français.** Dans l'ensemble, les ressources encyclopédiques anglaises et françaises se comparent à la moyenne. Parmi ces ressources, celles qui sont le plus généralement accessibles, soit l'*Encyclopædia Britannica*, l'*Encyclopædia Universalis*, le *Larousse*, se trouvent sous la moyenne, et deux d'entre elles se situent sous la barre des 10 %.

3. **Le fossé des genres dans les ouvrages de référence de type encyclopédique canadien.** Parmi les dix ressources qui se positionnent le plus favorablement, on en compte trois en anglais, une exclusivement en français et trois autres multilingues ou bilingues, comme c'est le cas de l'*Encyclopédie canadienne*. Cette dernière se démarque de l'ensemble de l'échantillon avec 21,4 % de biographies F. Le *Dictionnaire biographique du Canada*, qui est également bilingue, fait plutôt partie de la catégorie des publications qui présentent le fossé des genres le plus accentué avec 6,1 % de biographies F.
4. **Le fossé des genres wikimédien.** La comparaison entre la couverture biographique de ces diverses ressources et les trois wikis de la Fondation Wikimedia est avantageuse pour ces derniers. En revanche, les résultats confirment la constance et l'étendue du fossé des genres existant dans les contenus de type encyclopédique qui sont à la disposition des internautes.

Cette situation concernant le fossé des genres dans les ressources encyclopédiques n'est pas récente. Quelques auteurs et autrices ont déjà écrit au sujet des biais sexistes dans les encyclopédies. L'archiviste et féministe réputée Mary Ritter Beard avait produit en 1942 une étude intitulée «A Study of the *Encyclopædia Britannica* in Relation to Its Treatment of Women», dans laquelle ses collègues et elle soulignaient avec force qu'en termes de participation à la production des ouvrages de référence, les femmes étaient en quantité négligeable, à l'inverse des biais et des lacunes identifiés en termes de contenu et dont le dénombrement s'avérait substantiel. Une autre étude dans les années 1990,

celle de Gillian Thomas, a démontré que le traitement biographique des femmes les plaçait en position de figurantes dans une histoire de la connaissance humaine où même Marie Curie est surtout décrite comme l'épouse de Pierre Curie. Il est juste de constater, en particulier à la lumière des résultats de cette recherche, que la situation et les pratiques dans le domaine des ouvrages encyclopédiques ont peu évolué depuis l'époque de Beard ou de Thomas, et qu'elles apparaissent aujourd'hui essentiellement inchangées jusqu'à preuve du contraire. Tout en reconnaissant que Wikipédia continue la tradition des grandes encyclopédies qui a débuté avec Diderot et D'Alembert, sans laisser de côté ses objectifs ambitieux et une volonté de couvrir de manière exhaustive le savoir humain dans un projet quasi utopique, il faut aussi noter que cette encyclopédie en ligne perpétue certains préjugés sexistes qui existaient déjà dans ces œuvres plus anciennes. À première vue, ces biais existent en dépit même du principe de neutralité qui est pourtant un des piliers du projet Wikipédia.

Le principe de neutralité constitue un élément de réflexion pertinent qui est régulièrement mis en cause dans les discussions concernant le fossé des genres (McDowell et Vetter, 2018). D'entrée de jeu, il importe de préciser que le principe de neutralité ne réfère pas à un principe d'équilibre (par exemple, s'il y a 50 % d'hommes et de femmes, il devrait y avoir 50 % de biographies pour chacun). Le rôle du principe de neutralité dans le fossé des genres est lié à sa fonction épistémique. Comme les cinq principes fondateurs de Wikipédia qui sont « impératifs et non négociables », celui-ci constitue un fondement « intangible », s'appliquant à tous les articles. Le principe de neutralité, auquel les contribu-

teurs et contributrices réfèrent généralement par l'acronyme « NPOV » tiré de l'expression anglaise *neutral point of view* (point de vue neutre), affirme que :

les articles ne doivent pas promouvoir de point de vue particulier. Parfois, cela suppose de mentionner plusieurs points de vue et de représenter chacun de ces points de vue aussi fidèlement que possible, en tenant compte de leur importance respective dans le champ des savoirs. Cela suppose également de fournir le contexte nécessaire à la compréhension de ces points de vue selon les sources qui les portent, et de ne représenter aucun point de vue comme étant la vérité ou le meilleur point de vue. Ces conditions permettent la vérification des informations en citant des sources faisant autorité sur le sujet (particulièrement dans le cas de sujets controversés).

Ce principe, qui vise à ne pas promouvoir de parti pris, cherche fondamentalement à éviter les biais. Pour ce faire, il s'appuie sur deux autres règles : la vérifiabilité et l'interdiction de publier des travaux inédits. Ces conditions déterminent conjointement le contenu encyclopédique : « ce qui peut ou non être publié dans Wikipédia ». La première règle de la vérifiabilité exige que toute information doive, pour être utilisée, pouvoir être vérifiée par les lecteurs ou lectrices à l'aide d'une « source ou référence de qualité ». La seconde règle, qui interdit les travaux inédits ou les recherches personnelles, exclut plus précisément les « recherches qui n'ont jamais été publiées en dehors de Wikipédia ou bien qui représentent une “révolution” non encore connue ou débattue dans le domaine, une opinion “excessivement” minoritaire ou que l'on ne peut associer qu'à des sources jugées confidentielles et/ou peu fiables voire, plus simplement, des interprétations, déductions ou

intuitions personnelles du rédacteur de l'article». Ces trois règles sont aussi considérées comme codépendantes : elles « doivent être interprétées les unes par rapport aux autres » pour prescrire le contenu admissible et partant le savoir ou ce qui compte en tant que tel au sein de Wikipédia. Or, ces trois piliers qui forment une barrière contre la subjectivité et les biais tendent également à défavoriser de manière significative l'admissibilité des articles biographiques portant sur des femmes pour lesquelles, dans certains cas, des sources secondaires vérifiables ne sont pas disponibles ou qui sont, pour d'autres, insuffisantes ; ce qui compromet d'autant les opportunités d'atténuer le fossé des genres. Le paradoxe, que l'on pourrait appeler le *paradoxe du biais wikipédien*, tient à ce que les dispositifs (NPOV, vérifiabilité, pas de travaux inédits) mis en place pour contrer les biais dans les pratiques contributives individuelles exposent l'encyclopédie et la rendent vulnérable aux biais systémiques qui ne peuvent être contrés même lorsque les personnes participantes sont conscientes de ceux-ci et souhaitent intentionnellement y remédier. Dans cette perspective, ce principe de neutralité semble plutôt constituer un obstacle pour l'objectif de réduction du fossé des genres.

Par ailleurs, et comme le souligne Tkacz (2014), cette conception de la neutralité vise également à préserver l'encyclopédie des luttes au nom de la vérité : « La neutralité [...] tente de mettre à distance les batailles concernant la vérité du monde extérieur, c'est-à-dire des contestations de la vérité qui ont lieu en dehors de l'enceinte de Wikipédia. » À ce titre, Wikipédia diffère des encyclopédies historiques qui n'échappent pas aux controverses de leurs époques et vis-à-vis desquelles les auteurs assument un point de vue qui fait

autorité et qui reflète le plus souvent une vision du monde « relativement homogène » (Reagle, 2013), cité dans Tkacz. C'est aussi pour mieux soutenir la collaboration et établir le consensus collaboratif que ce choix épistémique en faveur de la neutralité de point de vue a été initialement revendiqué: « Le concept de neutralité de point de vue, tel que je l'ai envisagé à l'origine, reposait sur l'idée d'un concept social, visant à aider les gens à mieux s'entendre afin d'éviter ou de contourner de nombreux débats philosophiques. Quelqu'un qui croit que la vérité est socialement construite, et quelqu'un qui croit que la vérité correspond aux faits de la réalité peuvent toujours travailler ensemble dans ces conditions » [traduction] (Wales cité dans Reagle, 2013). À ce titre, comme le suggère Tkacz, ce ne sont pas seulement des batailles particulières au sujet de la vérité qui auront été abandonnées, mais bien la notion de vérité elle-même.

Si Wikipédia a abandonné la vérité du monde extérieur, ce projet n'aura pas été, pour autant, à l'abri de batailles et de controverses qui dépassent largement les débats et délibérations en matière de construction des savoirs. Dans les limites de ce qui relève de sa vérité interne<sup>6</sup>, des luttes se déroulent au quotidien, toutes aussi féroces que celles dont on voulait, vraisemblablement, se soustraire à l'origine. De nombreux

---

6. Par vérité interne, nous référons, à l'instar de Tkacz, à l'explicitation de la neutralité de point de vue qui sert de cadre de justification: « Aucun point de vue ne doit être présenté comme étant la vérité. Il ne faut jamais affirmer, sous-entendre ou même laisser croire qu'un des points de vue est d'une quelconque façon meilleure, égal ou moins bon qu'un autre. Ni imposer systématiquement un point de vue, en révoquant catégoriquement ce qui pourrait sembler divergeant de sa propre opinion (même si le point de vue que l'on voudrait imposer est scrupuleusement sourcé, ou trivialement répandu au sein de l'opinion publique). » Wikipédia: « Neutralités ». Dans *Wikipédia* [https://fr.wikipedia.org/wiki/Wikipédia:Neutralité\\_de\\_point\\_de\\_vue](https://fr.wikipedia.org/wiki/Wikipédia:Neutralité_de_point_de_vue)

débordements concernent des instances de biographies de femmes dont les sources secondaires vérifiables sont contestées. Ces controverses ne se résolvent pas toujours au bénéfice d'une délibération aboutissant à un consensus collaboratif suivant l'image idéalisée suggérée par Wales. Heather Ford (2015) a étudié les dynamiques à l'œuvre dans ces cas. Elle a montré que le processus de décision fonctionne en rhizome, dans les périphéries où est détenu le pouvoir de ce qui sera représenté dans Wikipédia. Or, ces wikipédiens, qui ne correspondent pas à l'image médiatisée des *amateurs*, disposent des savoirs et des savoir-faire spécialisés, ils « comprennent comment se comporter et parler selon le vocabulaire technique, symbolique et politique complexe de Wikipédia et ont tendance à l'emporter ». Plusieurs exemples illustrent comment ces pratiques ne sont pas avantageuses dans le cas de controverses liées à des biographies de femmes. En outre, comme le souligne Doutreix (2017), l'étude de ces controverses indique qu'elles procèdent au détriment des paroles expertes ou militantes.

On peut suggérer, par ailleurs, dans une perspective intersectionnelle, que la situation est aggravée lorsqu'il s'agit d'articles ou d'informations qui touchent des femmes qui combinent plusieurs motifs d'oppression tels que la race, la couleur ou l'origine ethnique. Cette situation pourrait affecter également des articles de femmes qui sont d'origine franco-québécoise ou franco-canadienne et dont les contenus sont remis en question dans les pages de discussion par des wikipédiens issus de la majorité sur fr.wikipedia, lesquels vont privilégier certaines idéologies sur le genre et aussi certaines sources au détriment d'autres qui sont associées aux cultures francophones « périphériques »

ou n'ayant pas un statut reconnu, selon les normes admises par la majorité, dans la hiérarchie des sources. Cette hypothèse rejoint les constats de Heather Ford (2017) selon lesquels ces wikipédiens « servent également à réaffirmer les centres d'autorité traditionnels, tout en amplifiant les nouveaux centres de connaissance et en niant l'autorité des connaissances qui ne sont pas codifiées dans la pratique ». En d'autres termes, le cadre épistémique relié à la neutralité empêche la production d'articles portant sur des personnes identifiées comme femmes pour lesquelles des sources secondaires vérifiables ne peuvent être fournies, mais il apparaît également préjudiciable dans les discussions touchant des articles dont les sources secondaires vérifiables pourraient être considérées comme suffisantes, mais qui sont rejetés par les wikipédiens qui détiennent le pouvoir sur les représentations et, par conséquent, sur ce qui est tenu pour vrai dans Wikipédia.

Avant de conclure cette discussion, il nous faut encore formuler quelques remarques au sujet des limites de cette étude. Ainsi, le fait que la déclaration de l'élément de propriété Wikidata relative aux encyclopédies (Q55452870) soit volontaire constitue peut-être une limite de l'outil, puisqu'il est possible de catégoriser des ressources dont le statut de publication ou d'ouvrage de référence de type encyclopédique pourrait être remis en question – c'est le cas, par exemple, de l'agrégateur *BabelNet* ou du *Notable Name Database*. Cependant, il s'agit des seuls cas de l'échantillon dont le statut apparaît possiblement contestable.

Par ailleurs, l'approche synchronique utilisée pourrait représenter une autre limite de la stratégie développée. L'idée de base de l'analyse actuelle consiste à mettre en évidence le

fossé des genres en fonction du contenu *qui est présenté* aux internautes par les différentes publications, donc le fossé pour lequel un effort de numérisation et d'accessibilité en ligne a été fait. Il serait fort intéressant d'analyser les données selon une approche diachronique en fonction des périodes historiques des personnalités concernées, surtout en considérant l'hypothèse d'une réduction du fossé des genres s'opérant dans les dernières décennies. L'outil Humaniki, par exemple, rend possible cette différenciation.

Toutefois, une des limites les plus préoccupantes de notre étude concerne son parti pris en faveur d'une conception binaire du genre. En mettant exclusivement l'accent sur le fossé des genres en tant que catégorie binaire (homme/masculin et femme/féminin), nous avons négligé le statut des identités non binaires. Cette conjecture contribue aux biais, aux discriminations et à l'invisibilisation systémiques que subissent les personnes s'identifiant comme non binaires ou non conformes à une conception binaire (personnes trans, intersexes, agenes, bispirituelles, gender queer, fluides, etc.). À ce titre, la définition même du fossé des genres est appelée à être réévaluée et étendue au-delà de l'approche fondée sur la conception occidentale binaire du genre.

### **Que faire maintenant ?**

Faute de pouvoir intervenir directement de manière substantielle et radicale sur les biais systémiques en raison des obstacles épistémiques découlant du NPOV, quelles sont encore les options possibles pour atténuer, de façon relative, le fossé des genres ? Comme le suggère Beaudouin (2020), la

production intensive de biographies de femmes qui respectent le critère de vérifiabilité, lequel requiert que toute information puisse être vérifiée par une source ou une référence de qualité, constitue une avenue pour réduire le fossé des genres. Le soutien et la participation à des initiatives telles que Les sans pagEs et les ateliers Art et féminisme, qui se déroulent un peu partout au Québec et au Canada, favorisent cette production de contenu. Il reste encore plusieurs dizaines de milliers d'articles à créer qui répondent à ces conditions et à ce régime.

De plus, on peut penser qu'il faut mettre en œuvre un chantier éditorial pour les articles biographiques de qualité touchant les femmes, qui présentent aussi une asymétrie de genre sur Wikipédia. Toutefois, pour de nombreuses personnalités féminines, les sources qui permettraient de satisfaire le critère de vérifiabilité ne sont tout simplement pas disponibles. Une autre piste est alors celle de la production de connaissances en amont, comme il est suggéré du côté de Wikimedia France : « Décernez des prix, faites des portraits, rédigez des entrées de dictionnaires bibliographiques, ou publiez des articles de recherche et des monographies sur des femmes méconnues. Wikipédia pourra ensuite traiter du sujet en faisant le travail de synthèse et de diffusion des connaissances. »

Dans une perspective régionale, une autre action consiste encore à accroître la création de liens entre les ressources bibliographiques canadiennes, québécoises et françaises, et Wikidata. Ainsi, l'ajout récent de 329 entrées du *Japanese Canadian Artists Directory* sur Wikidata montre qu'on y retrouve une proportion de 51,4 % de personnalités féminines.

On peut espérer le lancement d'un mouvement pour lier les contenus de ces ouvrages de référence – comme ceci a déjà eu lieu dans le cas d'encyclopédies telles que *Britannica* – afin de pouvoir mieux évaluer le fossé des genres existant dans ces dernières et y puiser également du matériel de contribution pour Wikipédia. Nous pensons notamment à un effort de mobilisation autour des encyclopédies québécoises suivantes : *Biographies canadiennes-françaises*, le *Dictionnaire des auteurs de langue française en Amérique du Nord*, l'*Encyclopédie du patrimoine culturel de l'Amérique française*, en plus de celles comprises dans notre recherche, soit le DBC et l'EC.

Notre étude nous invite à renouer avec la conclusion suivante : dans l'environnement wikimédien et au-delà, le problème n'est pas fondamentalement relié à l'importance du fossé des genres, que ce soit en termes de couverture biographique ou de participation éditoriale, mais il réside plutôt dans ce déséquilibre manifeste relié au sexisme systémique. Comme le suggère Jemielniak (2016), Wikipédia reflète les préjugés et les biais présents dans la société, et le sexisme, accentué par la culture techno-informatique, s'avère l'un des plus puissants leviers. Des résultats comme ceux de la présente recherche peuvent contribuer à une nouvelle explicitation et une meilleure compréhension du fossé des genres et du sexisme systémique. Toutefois, et comme d'autres travaux l'ont souligné, il faut désormais, de manière urgente et consciente, s'intéresser aux biais de genre en lien avec d'autres biais sociaux, selon une approche non binaire et intersectionnelle, dans le cadre des prochaines recherches avec l'aide de nouveaux outils, mais aussi aux politiques wikimédiennes inclusives afin de

réduire, dans une perspective plus large et fondamentale, les injustices épistémiques.

## **Remerciements**

Nous tenons à remercier Michèle Lefebvre, bibliothécaire à BANQ pour son aide et ses lumières au sujet des ouvrages de référence canadiens. Nous remercions également les wiki-médiens TomTom et Nicolas Vigneron qui ont développé plusieurs outils d'analyse d'éléments et d'articles wikis ainsi que le mathématicien Maxime André pour l'aide dans le traitement de certaines données. Merci enfin au wikipédien Cantons-de-l'Est et à Pierre-Yves Beaudouin pour la reprise et la diffusion des travaux réalisés antérieurement.



AXE II

**Institutions**





## CHAPITRE 4

# Le festival folk Mariposa et Wikidata au Canada

*Un projet de données liées*

STACY ALLISON-CASSIN

*Traduction de Pierre Gabriel Dumoulin*

Le festival Mariposa, un événement majeur de la musique folk au Canada, se transforme en laboratoire de données ouvertes grâce à un projet de longue haleine piloté par Stacy Allison-Cassin de l'Université Dalhousie. Elle propose un retour critique sur cette initiative visant à rendre les archives du festival, qui s'étendent sur plusieurs décennies, accessibles au public via Wikidata, la base de données sémantiques multilingue et collaborative lancée en 2012 par la Fondation Wikimedia. En revenant sur les difficultés, les enjeux et les opportunités rencontrés en cours de route, l'autrice souligne l'importance de la représentation équitable des cultures qui composent le Canada, en particulier celles des peuples autochtones. En filigrane, on apprend également comment le milieu des archives et des bibliothèques s'est approprié Wikidata pour documenter sur le Web la diversité des arts et cultures au pays.

Offrir la possibilité de rendre les données disponibles, réutilisables et transparentes, d'une part, et assurer l'accessibilité des données financées par des fonds publics, d'autre part, sont des caractéristiques estimées essentielles dans une société civile équitable. Les chercheurs, les organisations et les institutions non gouvernementales sont invitées à rendre leurs données plus accessibles. Au Canada, depuis plusieurs années, de nombreux ordres de gouvernement s'efforcent d'augmenter la disponibilité et d'accroître la sensibilisation de la population aux relations entre les données ouvertes et une société engagée, informée, innovante et juste. Pour le gouvernement canadien, les données ouvertes «sont définies comme étant des données structurées, lisibles par machine, qui peuvent être librement partagées, utilisées et mises à profit par quiconque, sans restriction» (Statistique Canada, 2023). En outre, il souligne que ces données offrent plusieurs bénéfices, comme l'innovation, la prise de décisions éclairée, la mise à profit des informations du secteur public, en plus d'accroître la responsabilité des gouvernements. Les critères qui rendent les données ouvertes – la structure, la lisibilité par machine, l'accès sans restriction – ne sont pas des éléments typiques des données commerciales exclusives, lesquelles ne sont généralement pas réutilisées en dehors des limites de leur organisation propriétaire.

Les bibliothèques, les archives, les musées, les galeries d'art, les centres culturels ainsi que d'autres organisations rattachées à la culture, ayant par exemple des projets musicaux ou historiques, se dotent également de mandats visant à avoir des impacts positifs sur la société canadienne. En général, ces organisations ont des besoins liés à la gestion de

l'information et des données, plus particulièrement sur la gestion des collections de culture matérielle. Par conséquent, elles sont aussi très impliquées dans la création et la mise à jour de données structurées. Dans le cas des bibliothèques, des archives et parfois d'autres organisations, ces données peuvent être lues par des machines, en plus d'avoir été créées, conservées ou échangées selon des normes locales, nationales ou internationales. Habituellement, ces organisations ont également intérêt à créer leurs propres données, qu'il s'agisse de rendre des collections accessibles pour un fonds d'archives ou d'augmenter l'impact de la recherche d'une université. Partager, améliorer et rendre accessibles les données structurées sont d'excellentes façons d'atteindre ces plus larges objectifs organisationnels et sociaux. Le partage ouvert de données peut aussi aider à rendre la société plus juste et équitable en augmentant l'impact de la culture au Canada et à l'international, comme dans le cas des données ouvertes du gouvernement. L'intérêt croissant pour l'intelligence artificielle (IA) et pour les nouveaux modes de recherche et de découverte souligne la nécessité d'avoir de plus vastes jeux de données très structurées pour soutenir le développement continu en apprentissage automatique (*machine learning*) et l'innovation en matière de traitement (*processing innovation*). Bien que la création et l'utilisation de données structurées disponibles en libre accès aient été bénéfiques, beaucoup de bibliothèques, d'archives et d'organisations culturelles sont souvent confrontées à un manque de ressources techniques pour rendre leurs données ouvertement accessibles (Allison-Cassin et Scott, 2018). De plus, de nombreux systèmes utilisés pour la gestion des collections au sein des organisations

ne sont pas disponibles sur Internet, ce qui limite les possibilités de rendre les informations visibles et réutilisables associées aux collections et aux personnes qui créent du contenu culturel. Les données culturelles ne sont pas souvent des données ouvertes (Zhu *et al.*, 2023).

Mon parcours dans Wikipédia et dans Wikidata a débuté par nécessité. En 2015, j'ai été nommée titulaire de la chaire W. P. Scott en bibliothéconomie numérique et j'ai commencé à mener un projet de recherche axé sur l'utilisation des technologies des données liées pour mieux décrire les matériaux musicaux (York University, s. d.). Plus spécifiquement, je me suis intéressée aux matériaux du festival folk Mariposa (Allison-Cassin *et al.*, 2015; Proffitt, 2018), l'un des plus anciens festivals de musique folk en Amérique et dont les archives sont conservées à l'Université York, dans les Clara Thomas Archives and Special Collections (St. Onge et Cohen-Palacios, 2010-2024). À l'origine, mon plan de recherche était de recueillir le contenu des programmations du festival, les artistes, les lieux, les dates et les personnes organisatrices par exemple, sous la forme de données liées afin de tester les descriptions enrichies et de révéler comment les informations sont liées entre elles dans le festival. Puisque les méthodes standards de description en bibliothèque ne facilitent pas le traitement adéquat de la musique, hormis le genre classique, me concentrer sur la musique folk canadienne me semblait être une étude de cas idéale pour examiner les concepts associés au passage d'une focalisation axée sur les œuvres musicales à une focalisation centrée sur le réseau de relations au sein d'une scène culturelle (Allison-Cassin, 2012). J'ai rapidement été confrontée à plusieurs défis dans le projet. Le premier problème était la

grande difficulté causée par un accès inadéquat aux ressources pour créer et publier des données liées. Un second problème concernait mon désir d'intégrer le contenu de Wikipédia dans mon projet, notamment la biographie des artistes : très peu d'articles sur Wikipédia traitent des artistes ou des lieux associés au festival Mariposa ou du renouveau de la musique folk sur la scène torontoise. J'ai alors commencé à écrire des articles Wikipédia sur des personnes ou des lieux importants du renouveau de la musique folk à Toronto et au Canada, comme la Riverboat Coffee House (Riverboat Coffee House, 2024). En voulant combler les lacunes sur le sujet, je me suis retrouvée de plus en plus impliquée dans l'utilisation de la plateforme et dans son soutien.

Dans ce chapitre, j'aborde plusieurs projets, initiatives et recherches dans le domaine des bibliothèques, des archives et des applications de données culturelles qui mobilisent la base de données Wikidata en contexte canadien. À partir de ces exemples, j'aborde les enjeux critiques de l'utilisation et du développement continu de Wikidata en tant que base de connaissances structurée ouverte et en libre accès au Canada, mais aussi les possibilités qu'une telle plateforme peut offrir. En tant qu'utilisatrice active et membre de la communauté Wikidata, j'adopterai une approche personnelle dans cette discussion, en me concentrant sur plusieurs thèmes et exemples. Pour une approche plus systématique de Wikidata, Zhao (2023) et Tharani (2021) ont publié des articles savants sur le sujet. J'ouvre ce chapitre sur mes débuts dans Wikidata avec le projet *Music in Canada @ 150*; les sections subséquentes suivront des regroupements par types d'activités.

En fin de compte, si Wikidata peut être utilisé pour soutenir des activités rattachées aux données culturelles et aux mécanismes et être employée pour rendre disponibles et accessibles les données structurées, il existe d'importants enjeux au sein de la plateforme Wikidata et de ses relations avec l'écosystème plus large qui doivent être pris en considération avant de mobiliser des ressources dans ce domaine. Au même moment, le besoin de développer une pratique éthique s'impose dans le paysage des données structurées afin de répondre à un large éventail de préoccupations, telles que les données autochtones, la vie privée et la discrimination. Pour plusieurs bibliothèques, archives et organisations, Wikidata offre un moyen de s'engager dans des projets qui visent à augmenter la visibilité des contenus sur des sujets canadiens. Cependant, des enjeux majeurs subsistent, comme le manque de soutien pour les données autochtones, la faible disponibilité des données et des contenus canadiens ainsi que les problèmes liés à la structure fondamentale du projet, et rendent improbable que Wikidata soit utilisé comme une source fiable de données ouvertes. Toutefois, compte tenu du rôle que jouent les données structurées dans le développement, actuel et à venir, de l'IA, Wikidata pourrait être appelé à devenir un acteur important de cette scène.

## **Contexte**

Projet de Wikimedia lancé en 2012, Wikidata évolue rapidement en une source importante de données structurées gratuites et en libre accès sur Internet. Comme son nom l'indique, Wikidata est un projet qui crée et soutient le

développement de données sémantiquement structurées en libre accès et auquel tout le monde peut contribuer. C'est un effort collaboratif semblable à celui de Wikipédia qui a un profond impact sur les lieux d'information et de connaissance, soulignant de fait l'importance de chaque contribution. Plateforme multilingue disponible gratuitement sous licence libre, Wikidata offre aux individus et aux organisations la possibilité de façonner les données structurées. Au Canada, Wikidata a trouvé sa place au sein de l'espace des données culturelles, s'alignant sur les objectifs organisationnels relatifs aux données ouvertes structurées (Allison-Cassin et Scott, 2018).

Wikidata est décrit sur la page d'introduction au projet comme «une base de données libre, collaborative, multilingue, et secondaire qui collecte des données structurées pour alimenter Wikipédia, Wikimedia Commons, les autres wikis du mouvement Wikimédia et pour n'importe qui de par le monde». C'est une base de connaissances, un dépôt d'informations structurées qui sont lues et traitées par des machines. Plus spécifiquement, les données structurées sont des informations (données) hautement organisées qui peuvent être facilement exploitables par des machines, comme des adresses ou des dates de naissance qui ont été normalisées. En d'autres mots, chaque élément dans le jeu de données est défini et utilisé de la même façon. Les données non structurées, quant à elles, peuvent être difficilement mobilisées par des machines ou être réutilisées. Wikidata est déjà intégré dans des applications ou des utilisations qui dépassent le cadre des projets Wikimédia. Contrairement à Wikipédia, la base de connaissances Wikidata est appréciée et valorisée pour ces utilisations

« au-delà des projets » de Wikimedia, ce qui la rend unique parmi leurs projets.

À l'origine, Wikidata a été développé pour répondre aux enjeux techniques que rencontrait Wikipédia. L'encyclopédie en ligne compte de nombreuses versions linguistiques qui sont uniques, et « l'un des premiers objectifs de Wikidata était de soutenir Wikipédia en centralisant les données contenues dans les infobox et les liens entre les versions linguistiques. Ainsi, Wikidata réduit la tâche des wikipédistes tout en augmentant la qualité et la cohérence des contenus » [traduction] (Vrandečić, 2013, p. 90). Avant sa création, il n'existait aucune relation entre les versions linguistiques de Wikipédia, ce qui rendait difficile la navigation entre les versions d'un article sur un artiste, comme passer de la version française à la version allemande de la page du musicien et poète Leonard Cohen. En tant que connecteur, Wikidata joue un rôle d'intermédiaire en s'assurant que tous les articles sur Leonard Cohen soient liés entre eux, quelle que soit la langue. Sur la plateforme, au lieu d'avoir « des articles liés entre eux, Wikidata conserve des listes de tous les articles sur un certain sujet dans les différentes versions linguistiques. Lorsqu'un article est soumis, le logiciel interroge Wikidata pour obtenir cette liste et l'afficher. Les versions linguistiques de Wikipédia sont complètement déchargées de la fonction d'assurer l'entretien de ses listes » [traduction] (Vrandečić, 2013, p. 91). La création de ces connexions a grandement amélioré la fonctionnalité de Wikipédia.

Alors que son rôle initial était la prise en charge des structures techniques de Wikipédia, la base de données Wikidata a rapidement commencé à soutenir d'autres types

de projets portant sur les données structurées. D'un côté, Wikipédia est structurée autour de textes écrits souvent sous la forme d'un récit pour être utilisables d'un point de vue humain ; de l'autre, Wikidata est adapté aux traitements automatiques de données. Toutefois, comme Wikipédia, « Wikidata est aussi fondé selon un modèle d'édition communautaire, c'est-à-dire qu'il bénéficie d'une communauté internationale de contributeurs, laquelle inclut autant des spécialistes que des dresseurs de *bot*. Tout le monde peut ajouter de nouvelles déclarations, des ajouts factuels à l'importation de données à grande échelle » [traduction] (Waagmeester *et al.*, 2020). Cependant, l'édition de données à plus grande échelle de Wikidata, par le téléversement de grands jeux de données, est étroitement liée à leur utilité dans le cadre de projets techniques. Celle-ci est davantage amplifiée par la structure de sa licence. Contrairement à de nombreuses bases de données commerciales propriétaires, « les données de Wikidata sont publiées sous la licence Creative Commons CC-0, laquelle permet à quiconque de réutiliser et de republier les données de quelque manière que ce soit. Les données sont gratuites et sans aucune condition ni exigence » [traduction] (Vrandečić, 2013, p. 90). Ces facteurs rendent l'énorme entrepôt de données intéressant à mobiliser dans le cadre d'autres applications. La taille des données et la licence utilisée sont essentielles à leur succès, notamment dans les secteurs du patrimoine culturel et de l'éducation.

## Wikidata et les bibliothèques, les archives et les musées

Les organisations œuvrant dans le patrimoine culturel et la culture, comme les bibliothèques, les archives et les musées, ont été actives dans Wikidata dès le début (Tharani, 2021). Lorsque Wikidata a été lancé, de nombreuses organisations culturelles, parfois regroupées sous l'acronyme anglais GLAM (pour *Galleries, Libraries, Archives and Museums*, soit les galeries d'art, les bibliothèques, les archives et les musées), participaient déjà à d'autres projets Wikimedia, comme Wikipédia et Wikimedia Commons, tirant parti d'initiatives de sensibilisation, comme la campagne Art + Feminism (Proffitt, 2018). De même, les organisations culturelles au Canada relevant des GLAM se sont également engagées dans l'utilisation de Wikipédia au cours des activités de sensibilisation, de diffusion et de communication (*outreach*). Par exemple, l'Art Gallery of Ontario organise des marathons d'éditions (*edit-a-thon*) depuis 2015 (Art Gallery of Ontario, 2015). Les organisations GLAM développent des liens étroits avec les métadonnées en raison de leurs besoins en gestion de collection et en personnel hautement qualifié dans l'utilisation des métadonnées. Pour cette raison, la barrière à l'entrée pour participer à Wikidata est plus basse qu'auprès d'autres communautés qui ont moins de ressources ou de connaissances techniques. Des facteurs tels que le développement d'un vocabulaire approprié et d'une interface facile à utiliser, ou la possibilité d'intégrer Wikidata dans les catalogues des bibliothèques sont cités parmi quelques-uns des aspects positifs de Wikidata (Allison-Cassin et Scott, 2018; ARL, 2018).

Wikidata est apparu comme un outil permettant de créer des données liées et d'améliorer l'accès aux organismes GLAM (Allison-Cassin et Scott, 2018; Tharani, 2021; Ansovi *et al.*, 2022). Toutefois, des interrogations subsistent quant à sa viabilité en tant qu'élément fiable de l'écosystème du secteur en raison de problèmes avec la qualité de ses données, de son manque de cohésion dans son modèle de données et de ses problèmes sous-jacents liés à la disponibilité des données et de l'information au Canada. Les organismes culturels et les secteurs du domaine culturel canadien utilisent Wikidata dans le cadre de leurs efforts pour gérer et rendre accessibles leurs collections. Notamment, la plateforme est une base de connaissances utilisée par des systèmes de traitement automatisé comme Google ou dans le développement d'outils d'IA. Elle permet la recherche contextuelle et la mise en relation d'informations dans la localisation de liens en relation, ce qui contribue à en augmenter la visibilité (Gertner *et al.*, 2023). Bien que la création de ses liens vers des fichiers d'autorité ou vers d'autres sources de données a pour objectif de répondre aux besoins en matière de données et d'information ainsi qu'aux processus sur Wikipédia et dans l'ensemble des projets Wikimedia, Wikidata soutient également des cas d'utilisation externes aux secteurs culturels, notamment sur le contrôle d'autorité (Veen, 2019; Bianchini *et al.*, 2021), sur l'influence de la recherche et des profils de chercheurs (Nielsen *et al.*, 2017; Odell *et al.*, 2022) et sur l'enrichissement des informations sur les collections (Colla *et al.*, 2021; Ansovini *et al.*, 2022; Hawkins, 2022).

Wikidata joue un rôle essentiel dans l'interconnexion des contenus axés sur le Canada, améliorant ainsi la visibilité

des ressources culturelles tant au Canada qu'à l'international. Par exemple, Wikidata connecte l'élément de l'autrice, poète et actrice Pauline Johnson à l'élément de sa notice dans les catalogues des bibliothèques et dans les ressources d'information, incitant la création d'une carte de connaissances de Google. Tant Wikipédia que Wikidata alimentent les graphes de connaissances de Google, mais les données structurées et lisibles par les machines de Wikidata facilitent ce processus, en plus de permettre également aux wikidatistes d'améliorer plus efficacement les connexions du graphe de connaissances avec d'autres sources d'information. En tant que plateforme de métadonnées qui offre la possibilité de transférer des informations d'une base de données à l'autre, Wikidata a la capacité d'amplifier les collections et les contenus, en plus de solidifier les liens entre les sites de données. Elle augmente l'impact des données, rend des améliorations aux données plus importantes et devient un outil encore plus efficace pour le déplacement des données, renforçant de fait les liens avec notre patrimoine culturel.

### **Wikidata et les données du festival folk Mariposa**

Comme mentionné au début de ce chapitre, j'ai commencé à travailler avec Wikidata dans le cadre d'un projet lié à mon poste de titulaire de la chaire W. P. Scott en bibliothéconomie numérique. Le projet de données liées du festival Mariposa visait à modéliser un réseau de relations entre les artistes du festival sur la base de jeux de données consolidés à partir des programmations. Pour créer ce jeu de données, j'ai commencé par développer un modèle de données idéal fondé sur les

catégories d'éléments contenus dans les programmations, puis j'ai poursuivi par l'examen des attributs les plus utiles pour chaque catégorie. Le modèle de données initial incluait plusieurs catégories, comme les artistes, les artisanes, les danseurs, les lieux, les dates et le personnel administratif du festival. Un ensemble d'attributs a été mis sur pied pour chacune de ces catégories de données: pour les artistes, il s'agissait de propriétés comme le nom, le genre musical, le médium avec lequel l'artiste a travaillé et la date à laquelle la performance a eu lieu. L'application Google Sheets a été utilisée pour enregistrer les données dans des feuilles de calcul, et chaque groupe avait sa propre feuille de calcul. Par exemple, une feuille distincte de Google Sheets a été utilisée pour les musiciens et musiciennes. Étant donné qu'il s'agissait d'un projet de données liées et que je désirais établir des liens entre le jeu de données du festival Mariposa et les données ouvertes et en libre accès sur Internet, des sources d'identifiants pérennes (*sources of persistent identifiers*) ont été incluses. Ces sources comprenaient le fichier d'autorité VIAF (*Virtual International Authority File*), qui est une source d'informations liée aux données d'autorité des bibliothèques, ainsi que des sources d'identifiants spécifiques au milieu musical, comme MusicBrainz et Discogs. Dans le cadre de son initiative Un avenir numérique lié, l'Association canadienne des organismes artistiques (CAPACOA) s'est intéressée à la création de données structurées pour soutenir la visibilité des artistes, des lieux et des événements des arts de la scène au Canada. Elle a signalé la nécessité d'augmenter la disponibilité des données liées, et par extension des identifiants de ressource uniforme (*Uniform Resource Identifier*, URI), pour développer rapidement et efficacement des stratégies numériques au diapason avec

les arts de la scène canadiens. En l'absence d'une culture de métadonnées forte, ouverte et stable au Canada, la santé générale des secteurs de la culture et des arts de la scène est affaiblie: «les métadonnées des arts de la scène sont devenues l'un des enjeux les plus pressants pour les secteurs des arts du spectacle au Canada» (CAPACOA, s. d.).

Tout en développant les données relatives au festival Mariposa, j'ai exploré les méthodes disponibles pour transformer les données de la feuille de calcul en données liées et les rendre disponibles sur Internet. Faute de logiciels, d'outils et de ressources accessibles, incluant l'accès à des spécialistes en logiciels de créations de données, je me suis tournée vers Wikidata comme outil et méthode pour créer, puis publier des données relatives au festival. L'excellente compatibilité entre les outils logiciels, comme Google Sheets et OpenRefine et Wikidata, a rendu relativement simple le téléversement des données. De plus, j'ai pu faire tout le travail moi-même, puisque Wikidata est en libre accès, gratuit et soutenu par une communauté active. Utiliser Wikidata pour publier les données était une solution utile qui a aussi permis de souligner plusieurs enjeux d'intérêts sous-jacents à l'utilisation de Wikidata dans le contexte canadien.

### **Le manque de couverture de la musique canadienne**

Le festival Mariposa est l'un des principaux festivals de musique folk au Canada, et le fait d'y participer est un gage de notoriété. Cependant, il s'est avéré difficile de trouver des informations pour compléter les données sur les artistes, les groupes et toute personne qui y a participé. S'il est possible d'utiliser Wikipédia, une encyclopédie ou toute autre source

de notoriété pour des artistes comme Gordon Lightfoot et Joni Mitchell, il n'en va pas de même pour la majorité des artistes, d'autant plus que ce festival se distancie des artistes grand public (Mariposa Folk Festival, 2014). Les programmations ne contiennent pas assez de détails biographiques pour remplir convenablement les données requises selon les standards établis par la communauté. Les sources de référence habituelles telles que l'*Encyclopédie de la musique au Canada*, laquelle a été intégrée à l'*Encyclopédie canadienne* en 2003, continuent de privilégier la musique classique établie par des musiciens et musiciennes jazz. Malgré qu'elle cherche à étendre sa couverture à la musique populaire, l'*Encyclopédie canadienne* a encore beaucoup de retard à rattraper pour se défaire des préjugés hérités du comité éditorial de l'*Encyclopédie de la musique* sur ce qui justifie la notoriété ou l'importance d'artistes ou genres musicaux.

Les journaux locaux et nationaux disposant d'une rubrique ou d'une section régulière de critique musicale peuvent souvent être une bonne source de couverture des artistes. Toutefois, les recherches dans les journaux n'ont pas donné de bons résultats. Par exemple, bien que le groupe torontois The Dirty Shames apparaisse dans les programmations du festival Mariposa et dans les photographies du journal *Toronto Telegram* (McFadden *et al.*, 1966), ces sources donnent trop peu d'information pour renseigner l'élément Wikidata (The Dirty Shames, 2021). Certains détails sur la vie des artistes dans le jeu de données du festival Mariposa ont été trouvés dans des archives de journaux numériques, mais un abonnement payant était nécessaire auprès de tels périodiques, comme le *Toronto Star* et le *Globe and Mail*. La nature «verrouillée» de ces ressources signifie qu'elles ne

sont plus accessibles au public et que les données ne sont pas en libre accès sur Internet. D'autres détails n'étaient toutefois pas disponibles sous forme numérique et n'ont pu être trouvés. Les journaux ou publications plus modestes peuvent être essentiels pour couvrir des scènes musicales locales, mais les archives de ces publications ne sont pas toujours disponibles, et nombre d'entre elles ne sont plus en activité et leurs sites Internet ne sont plus accessibles. Les médias locaux, tels que les bulletins de nouvelles et les journaux locaux, disparaissent rapidement au Canada; en 2023, CBC News signalait la fermeture de 70 journaux locaux en Ontario. L'absence de couverture communautaire et la disparition des petits médias locaux ne signifient pas seulement que les populations locales ne sont plus informées sur certains sujets: c'est une perte critique du patrimoine culturel canadien, en plus d'avoir de graves conséquences sur la capacité des wikipédistes de travailler à la documentation de la culture canadienne dans des projets Wikimédia (CBC News, 2023). Dans le même temps, les journaux communautaires numériques indépendants, comme le *West End Phoenix* de Toronto, reconfigurent la couverture de l'actualité (West End Phoenix, s. d.).

Le manque d'information sur la musique canadienne sous format numérique est un défi non seulement pour mener des recherches sur des éléments individuels, mais aussi pour s'assurer de la fiabilité des données. Dans les premières années de Wikidata, référencer les déclarations était une bonne pratique pour développer la connaissance sur la plateforme, bien qu'elle demeurât optionnelle. Par conséquent, il aurait dû être essentiel d'inclure des références lors de l'ajout de données à Wikidata dans le cadre du

projet du festival folk Mariposa. Depuis la création de ce projet, les références sont devenues essentielles pour établir la notoriété. Les métadonnées de provenance sont cruciales pour la plateforme Wikidata. Elles ont la même fonction que les citations: «Les références sont utilisées pour indiquer quelles sources spécifiques peuvent soutenir l'information ajoutée dans une déclaration.» [traduction] (Help: Sources, 2023) Une déclaration qui suggère que le lieu de naissance de Céline Dion est Charlemagne doit être étayée par des références de sources fiables dans lesquelles l'information peut être vérifiée. Comme dans le cas de Wikipédia, les sources secondaires et tertiaires fiables et de haute qualité sont à prioriser. L'ajout de références aux déclarations Wikidata augmente la qualité des données liées pour de nombreuses utilisations, en plus de fournir des données de meilleure qualité pour des applications avec une IA. Un manque de références peut avoir des incidences sur l'utilité même des données (Beghaeiraveri, Gray et McNeill, 2024).

Le manque de références facilement accessibles pour les déclarations sur les artistes du festival folk Mariposa a un effet aggravant et met en évidence un problème le plus important de la disponibilité des informations sur la musique et les arts. Dans les projets ultérieurs que j'ai déployés à l'aide de Wikidata, sur la documentation de la littérature, le patrimoine bâti ou les films par exemple, l'incapacité à fournir des références et une provenance pour les données canadiennes sur Wikidata pose de sérieux problèmes. Comme cela a été noté pour Wikipédia, les sources établies et fiables doivent avoir prouvé leur notoriété, et le manque de documentation est connu (McCracken, 2018). Il peut donc être difficile de rédiger des articles et de veiller à

ce qu'ils ne soient pas supprimés. Le lien entre la notoriété et le besoin de références pour les déclarations rend difficile la création de données sur des éléments canadiens, puisqu'ils risquent d'être supprimés. De manière générale, le manque de documentation sur les activités musicales au Canada rend difficile la constitution de jeux de données exhaustifs et bien référencés sur la musique dans Wikidata.

### **Wikidata et la campagne Music and Belonging: Canada at 150**

En réponse au peu d'informations sur la musique canadienne, autant sur Wikipédia que sur Wikidata, j'ai lancé la campagne Wikimedia «Music and Belonging: Canada at 150» (Musique et appartenance: le Canada à 150 ans). Celle-ci avait pour but de produire plus de contenus et de sensibiliser une plus large communauté au manque d'information sur la musique canadienne. Elle a été développée pour concentrer les efforts pancanadiens sur la création de contenus sur la musique au Canada (Allison-Cassin et Scott, 2018). En tant que membre de l'Association canadienne des bibliothèques, archives et centres de documentation musicaux (ACBM), je trouvais logique de faire appel à la communauté des bibliothèques musicales au Canada, laquelle dispose d'un réseau de personnes intéressées et bien informées, tant sur la musique que sur l'information. À l'origine, cette initiative a été conçue pour s'intégrer aux activités du 150<sup>e</sup> anniversaire de la Confédération canadienne et a repris le thème de la conférence annuelle de la Société canadienne pour les traditions musicales, de l'Association canadienne des bibliothèques, archives et centres de documentation

musicaux, et de la branche canadienne de l'Association internationale pour l'étude de la musique populaire. Elle a aussi été financée par Wikimedia et par le fonds pour les événements spéciaux de l'Université York. Cette campagne incluait un atelier en personne à l'Université de Toronto, deux ateliers virtuels, et plusieurs événements disséminés sur une année. Les événements ont eu lieu à l'Université de l'Île-du-Prince-Édouard à Charlottetown, à l'Université Memorial de Terre-Neuve, à l'Université Western Ontario à London, à la Hamilton Public Library à Hamilton, à l'Université York et à l'Université de Toronto à Toronto, à l'Université Laurentienne à Sudbury, à l'Université du Manitoba à Winnipeg, à l'Université de la Saskatchewan à Saskatoon et à l'Université MacEwan à Edmonton. Les événements ont été organisés par Dan Scott (Université Laurentienne), Caroline Doi (Université de Saskatchewan) et Monica Fazekas (Université Western). Pour documenter tout ce travail, un tableau de bord des événements Wikimedia a été utilisé. Il couvre 11 programmes et le travail de 124 wikipédistes, en plus de documenter la création de 24 nouveaux articles et la modification de 187 articles.

Le coup d'envoi de la campagne a été donné lors d'un atelier d'une journée à l'Université de Toronto. Comme celui-ci coïncidait avec la réunion annuelle de plusieurs associations professionnelles liées à la musique, des bibliothécaires et des chercheurs en musique ont participé à l'atelier. Ce dernier a été conçu pour « former les formateurs » en donnant des instructions sur le processus d'édition de Wikipédia et de Wikidata, en plus d'ouvrir des discussions sur la manière de préparer des marathons d'édition. L'objectif du projet était de venir en aide aux personnes

qui allaient organiser de telles journées durant la campagne. L'événement comprenait une session de discussion avec des personnalités locales, comme Amy Furness, archiviste aux collections spéciales Rosamon Ivey et cheffe de la bibliothèque et des archives du Musée des beaux-arts de l'Ontario, et John Dupuis, bibliothécaire scientifique à l'Université York avec une expérience dans l'organisation de marathons d'édition lors de campagnes, comme Art + Feminism et Ada Lovelace Day. L'atelier s'est révélé être un mécanisme essentiel pour que les personnes participantes puissent acquérir de l'expérience et des compétences, en plus d'engager des discussions sur les sujets susmentionnés sur la disponibilité des sources d'information concernant la musique canadienne. Parmi les sujets de conversation animée, il y avait la question de savoir « pourquoi Wikipédia » plutôt qu'un autre site Internet institutionnel ou une autre forme de publication. Cela a donné lieu à des débats sur le libre accès et l'utilisation des licences libres.

Ces événements se sont déroulés de 2016 à 2017, et Wikidata était encore relativement inconnu auprès des bibliothèques, archives et musées canadiens. Peu de contenus étaient donc disponibles lors des ateliers. Dan Scott, bibliothécaire à l'Université Laurentienne, a joué un rôle déterminant dans la création et l'animation des ateliers sur Wikidata et la création de matériel didactique (Scott, 2017a). Grâce à ce travail et au projet du festival folk Mariposa, Scott et moi avons travaillé à la modélisation d'un moyen de capturer les données des festivals de musique à l'aide de Wikidata. Plus précisément, Scott s'est concentré sur le Northern Light Festival Boréal qui met l'accent sur des considérations bilingues (Scott, 2017b). Le travail collabo-

ratif sur Wikidata, à partir du projet Music in Canada @ 150, a tôt offert un engagement dans la création de données liées à la musique et est devenu la base pour les futurs efforts de collaboration entre Scott et moi, y compris un atelier lors de la conférence à l'Association des bibliothèques de l'Ontario sur l'ajout de bibliothèques aux projets Wikimedia et un autre lors de la conférence Semantic We in Librairies sur Wikibase. La campagne a également donné lieu à la création de la conférence GLAM-Wiki à Toronto en 2019 qui a attiré plus de 100 personnes participantes des bibliothèques, archives et musées de la région de Toronto.

La campagne Music in Canada @ 150 est née en réponse aux lacunes que j'ai observées en travaillant sur le projet de données du festival folk Mariposa. Dans une large mesure, la campagne a réussi à mobiliser les bibliothécaires des bibliothèques musicales intéressées par les épineuses questions liées à l'information sur la musique. Même si elles existaient avec le projet, les communautés de bibliothèques musicales impliquées dans le projet ont prouvé qu'il est possible de participer au développement de Wikidata, puisque leurs membres désirent trouver des mécanismes pour soutenir la visibilité de la musique au Canada. Scott a résumé succinctement les efforts déployés: «Notre argument central était que, plutôt que de se concentrer sur l'amélioration directe de nos propres silos de dépôts de données locaux (par exemple les catalogues de bibliothèques, les expositions numériques), les bibliothèques et les archives devraient plutôt investir leurs ressources limitées dans l'enrichissement de Wikidata, un dépôt de données centralisé, pour maximiser la visibilité de ces entités et la réutilisation des données dans le monde entier [...] et

ensuite extraire ces mêmes données pour nos dépôts locaux afin d'enrichir nos affichages et notre intégration au monde plus large des données.» [traduction] (Scott, 2017b) Toutefois, il est difficile d'évaluer les impacts à long terme de ces campagnes sur l'utilisation de Wikipédia et de Wikidata pour le contenu musical, et nous n'avons pas effectué de suivi auprès des personnes participantes. Évidemment, organiser des campagnes d'édition ne permettra pas de résoudre le problème du manque de sources secondaires et tertiaires sur la musique canadienne.

Comme avec Wikipédia, Wikidata est un projet organisé par des bénévoles, et les propriétés utilisées et disponibles sont choisies par l'obtention d'un consensus auprès des membres de la communauté. S'il peut être très avantageux d'offrir une plus grande flexibilité dans la création de données, cela aura pour conséquence d'avoir des données moins standardisées que celles utilisées par les systèmes qui utilisent des ontologies telles que le CIDOC CRM ou par les systèmes descriptifs normalisés des bibliothèques. Wikidata n'a pas de structure hiérarchique stricte; par exemple, les concepts plus larges ou plus étroits ne sont pas nécessairement respectés. Un autre problème connu est relié à la nature multilingue de Wikidata. La traduction ne permet jamais une transposition univoque et les concepts sont complexes à saisir. Augmenter la disponibilité des informations en ligne sur des sujets canadiens n'est pas un problème qui doit être réglé uniquement par les membres de la communauté Wikidata. La tâche d'envisager des projets de numérisation et d'archivage sur Internet revient plutôt aux gouvernements, aux bibliothèques, aux archives et aux organisations musicales.

## Wikidata et les événements

Le projet de données du festival folk Mariposa a permis d'explorer des aspects de la description des données culturelles qui seraient autrement impossibles avec les pratiques descriptives traditionnelles des bibliothèques et des archives. Une dimension importante de ce travail était la création d'éléments Wikidata sur des événements ou des objets qui durent dans le temps, généralement le matériel des bibliothèques ou des archives qui est inclus dans des collections d'objets, comme des livres, des manuscrits ou des enregistrements. Le besoin ou l'intérêt pour une collection axée sur des objets crée des limites dans la façon qu'il est possible de le décrire, restreint son accès et nuit à la compréhension des œuvres d'art qui s'inscrivent dans le temps comme la musique. Wikidata rend possible la description d'événements et fournit d'excellentes cotes (*dimension*) en lien avec les données. La capacité de décrire des événements a été un des atouts de Wikidata pour le projet du festival Mariposa.

Wikidata permet la représentation des éléments dans le temps. La création d'outils de génération de frises chronologiques est devenue une méthode populaire de visualisation des données, et un certain nombre d'outils ont été développés pour tirer parti des éléments de Wikidata qui contiennent des données liées aux dates. Toutefois, la modélisation des événements du festival Mariposa a été traitée de manière plus détaillée en associant les personnes participantes aux événements. Les projets Wikidata constituent un moyen efficace pour apprendre la meilleure façon de modéliser et de créer des données en relation avec des sujets

spécifiques et ils peuvent souvent fournir le support nécessaire pour comprendre un sujet. Les projets Wikidata, comme celui intitulé WikiProjet Lieux culturels, sont utiles pour passer en revue les enjeux essentiels liés à la description d'événements culturels. Pour le projet de données du festival Mariposa, il est devenu clair que la création manuelle de données sur les événements était un effort laborieux : chaque artiste-interprète devait avoir sa propre entrée Wikidata, ce qui représentait beaucoup de temps et d'effort. Ainsi, bien que la création de données granulaires soit une représentation plus complète du festival lui-même, elle est devenue laborieuse. Et même si plusieurs années se sont écoulées depuis la création du projet de données du festival Mariposa, la question du rapport entre l'investissement en main-d'œuvre et les impacts positifs perçus demeure pertinente, compte tenu des défis contemporains dans la création de données structurées. La disponibilité croissante d'outils axés sur le traitement du langage naturel peut réduire le besoin de saisir manuellement des données, mais une dépendance subsiste quant à la création de données lisibles par machine.

### **Représentation des collections d'archives au Canada**

La création de données relatives aux collections et aux fonds d'archives sur Wikidata peut contribuer à donner une vision d'ensemble plus claire des collections des institutions canadiennes et peut rendre les sources primaires plus faciles à trouver. Le projet de données du festival folk Mariposa s'est concentré sur une collection d'archives conservées aux

Clara Thomas Archives and Special Collections de l'Université York. L'une des façons de relier les collections uniques détenues par les institutions canadiennes et Wikidata est d'utiliser la propriété « Archivé par ». Une propriété sur Wikidata, peut-on lire sur la page d'aide dédiée à cette notion, « décrit la valeur des données d'une déclaration et peut être considérée comme une catégorie de données, par exemple "couleur" pour la valeur "bleu". Les propriétés, lorsqu'elles sont associées à des valeurs, forment une déclaration dans Wikidata ». En grande partie, les propriétés définissent la structure de Wikidata et doivent être adoptées par un consensus de la communauté. Wikidata est constitué de déclarations qui incluent une propriété et une valeur, sous la forme d'un élément. Les valeurs dans Wikidata peuvent être des liens internes à d'autres éléments de Wikidata ou être des liens externes associés à des adresses Internet permanentes, nommément un URI. En juin 2024, Wikidata répertoriait 140 propriétés liées aux archives, aux personnes et aux organisations (Liu, 2018). La plupart des propriétés répertoriées sont des identifiants externes, mais un petit nombre d'entre elles sont utilisées pour décrire des collections. Par exemple, la propriété P485 « Archivé par » relie l'élément Wikidata décrit aux institutions qui détiennent les archives pour cet élément. Katrina Cohen-Palacios (2019), archiviste à l'Université York, a présenté et fourni des documents d'atelier sur l'ajout de fonds d'archives à Wikidata, comme la méthode pour ajouter un lien à l'élément Wikidata « Toronto Telegram Newspapers » vers l'élément « Clara Thomas Archives and Special Collections ». Cohen-Palacios (2019) a suggéré que la création d'informations sur les instruments de recherche dans Wikidata peut

faire gagner du temps aux archivistes en automatisant certaines tâches qui pourraient autrement devoir être maintenues manuellement.

Suivant les exemples présentés par Cohen-Palacios (2019), Ansovini et ses coauteurs (2022) ont lancé une initiative visant à ajouter à Wikidata certaines informations provenant des fonds d'archives de l'Université de Toronto. L'ajout d'informations sur les archives peut aider les individus (et les machines) du monde entier à trouver et à améliorer les collections et documents liés à la culture canadienne. En outre, l'ajout d'informations provenant des fonds d'archives relie les institutions au réseau Wikidata, ce qui permet d'effectuer des requêtes et des visualisations de données pour obtenir des vues différentes sur une myriade de connexions. À partir de leur travail sur l'autrice canadienne Margaret Atwood, Ansovini et ses coauteurs (2022) ont constaté que «le simple ajout d'un triplet "Archivé par", qui lie les documents à un article Wikidata, permet les types de connexions qui peuvent avoir été faites dans la salle de lecture, où les archivistes utilisent leurs connaissances contextuelles pour suggérer aux chercheurs des ressources liées aux personnes, aux dates importantes ou aux publications notables. Wikidata peut fournir des points d'accès générés par la communauté et exploitables par des machines qui rendent les recherches exploratoires possibles pour visualiser, rechercher et explorer les relations entre entités» [traduction]. Le projet de données du festival folk Mariposa a permis quelques expérimentations en ajoutant à Wikidata des informations sur les fonds d'archives; les Clara Thomas Archives and Special Collections ont également été mobilisées pour ajouter du contenu en lien avec le festival Mariposa.

En tant qu'activité, l'ajout d'une déclaration à des éléments Wikidata existants en lien avec des fonds d'archives est plus facile à réaliser que la création d'entrées pour des festivals, car elle est limitée dans son champ d'application et ne nécessite pas de connaissances techniques approfondies. Alors que l'ajout de fonds d'archives pourrait avoir un impact dans la création des données structurées en lien avec les archives plus larges d'Internet, les fonds d'archives disponibles sur Wikidata sont irréguliers. Cohen-Palacios (2019) a mis en évidence cette disparité à l'aide d'une visualisation montrant que certaines institutions, principalement en Ontario et au Québec, ont entre 300 et 550 fonds répertoriés, tandis que d'autres sont moins représentées. La création d'entrées pour les fonds d'archives pourrait bénéficier d'une sensibilisation et d'un soutien afin d'assurer une plus grande équité dans la couverture, sans quoi l'impact de ces listes risque de rester relativement faible. D'autres sites de patrimoine culturel utilisent aussi Wikidata comme outil de travail. Par exemple, le Réseau canadien d'information sur le patrimoine (RCIP) utilise Wikidata pour leur projet Nomenclature. Le projet vise à créer un vocabulaire de données liées à utiliser dans le secteur des musées et le RCIP s'est concentré, à ce stade du projet, sur la description des œuvres d'art. Tandis que RCIP s'aligne sur le CIDOC CRM en tant que norme de contenu et de schématisation, il envisage Wikidata comme un moyen de relier les collections entre elles et de les rendre davantage visibles.

## Wikidata et les données autochtones au Canada

Créée en 1968, la zone pour les peuples autochtones (*Native People's Areas*) était une autre facette du projet de données du festival folk Mariposa. Un espace physique dans l'enceinte du festival était dédié à celle-ci. Des musiciennes, des danseurs et des conteuses faisaient partie de la programmation. La cinéaste et artiste abénaquise Alanis Obomsawin a assuré le commissariat de la zone des peuples autochtones de 1970 à 1976, et la programmation témoigne de l'influence de la montée des mouvements de défense des droits des Autochtones au Canada et aux États-Unis. Par exemple, des groupes comme North American Indian Travelling College et des artistes comme le poète Duke Redbird ont participé au festival (Mariposa Folk Festival, 2014). Dans le cadre de la production de données en lien avec le festival folk Mariposa, à la zone des peuples autochtones correspondait son propre jeu de données. À l'instar des autres dimensions du projet, travailler avec les données du festival Mariposa sur Wikidata a mis en évidence des enjeux pertinents à la discussion plus large sur l'utilisation de Wikidata dans le contexte canadien.

Comme souligné plus tôt dans ce chapitre, il a été difficile de trouver des informations sur les individus inscrits dans les programmations, une situation qui s'est révélée encore plus difficile pour les artistes de la zone des peuples autochtones. Si certaines personnes connues apparaissaient dans la programmation, comme Shingoose, ce n'était pas le cas pour beaucoup d'autres dont les noms apparaissaient seulement sous la forme d'un nom général, comme « Six Nations Reserve Dancers » ou « Metis Group ». Les recherches dans les archives du festival Mariposa n'ont pas permis d'obtenir

plus d'information : en l'absence d'informations plus spécifiques, il était impossible d'ajouter ces personnes à Wikidata. Le manque d'information en lien avec la zone des peuples autochtones pourrait être un exemple de manque d'attention ou de connaissances de la part des personnes organisatrices du festival. De plus, ce même manque d'information sur ces groupes sous-entend généralement un manque d'information au sein du vaste paysage canadien. Bien qu'il puisse sembler approprié d'ajouter le nom des personnes participantes dans une base de données interne malgré leurs noms vagues ou incertains, de telles informations sont toutefois inappropriées pour une base de connaissances globale et ouverte, car leur signification est imprécise. Il est important de soigneusement réfléchir aux usages de Wikidata par rapport à un dépôt ou une base de connaissances contrôlée et hébergée en interne.

En outre, la pertinence d'utiliser Wikidata au Canada pour examiner les données autochtones est un sujet de préoccupation crucial. Les données autochtones peuvent être définies comme « des données, de l'information, des connaissances, sous quelque forme que ce soit, qui ont un impact sur les peuples, nations et communautés autochtones aux niveaux collectif et individuel; des données sur leurs ressources et leurs environnements, sur eux en tant qu'individus et en tant que collectivités » [traduction] (Carroll *et al.*, 2021). Cette définition de grande envergure est conforme à la compréhension de l'expression autochtone telle qu'utilisée en droit, notamment dans la Déclaration des Nations Unies sur les droits des peuples autochtones (DNUDPA). Compte tenu de cette compréhension, une grande partie des données sur Wikidata relatives aux territoires, environnements et peuples

autochtones dans le pays, aujourd'hui connu sous le nom de Canada, sont des données autochtones. Le 21 juin 2021, le gouvernement fédéral canadien a adopté le Projet de loi C-15, la *Loi sur la Déclaration des Nations Unies sur les droits des peuples autochtones*. Sa mise en application « prévoit pour le gouvernement du Canada de prendre les mesures nécessaires pour veiller à ce que les lois fédérales soient compatibles avec la Déclaration des Nations Unies sur les droits des peuples autochtones et d'élaborer et de mettre en œuvre un plan d'action visant à atteindre les objectifs de la Déclaration » (Projet de loi C-15, 2020). Bien que la DNUDPA et le projet de loi C-15 aient été critiqués par certaines communautés autochtones, ils fournissent néanmoins un cadre et des moyens de mettre en œuvre les mesures nécessaires pour passer à l'action dans la reconnaissance des droits autochtones, dont les connaissances, les cultures, les données et les informations. La DNUDPA et le projet de loi C-15 donnent l'élan nécessaire aux gouvernements municipaux et provinciaux, de même qu'aux organisations, pour envisager des mesures pour harmoniser les politiques et pratiques.

Le mouvement mondial grandissant visant la souveraineté des Autochtones appelle à la nécessité d'une reconnaissance, d'une gouvernance et de systèmes qui garantissent aux peuples autochtones le contrôle de leurs données, y compris les données liées aux territoires et aux cultures. En tant que projet ouvert et en libre accès, Wikidata adhère aux principes FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable) qui engendrent des impacts positifs sur la communauté des savoirs libres (Odel, Lemus-Roja et Brys, 2022). Toutefois, les principes FAIR ignorent « les différences de pouvoir et les contextes historiques » [traduction] et soutiennent potentiel-

lement des pratiques extractives qui continuent de nuire aux peuples autochtones (Hudson, 2020). Le caractère absolu et non spécifique du libre accès est largement reconnu comme problématique (Anderson et Christen, 2019). Au moment de la rédaction de ce chapitre, Wikidata n'avait aucun mécanisme pour garantir que les données autochtones soient traitées de manière appropriée, y compris assurer la mise en place de mécanismes pour éviter que des données soient ajoutées de manière non éthique. Idéalement, la communauté de Wikidata et la Fondation Wikimedia travailleraient à la mise en œuvre des principes CARE pour les données autochtones : Collectivement bénéfique, Autorité de contrôle, Responsabilité et Éthique (Hudson, 2020). La Global Indigenous Data Alliance (GIDA) appelle les organisations à respecter le principe du #BeFAIRandCARE (être juste et attentionné). Compte tenu de l'absence de mécanismes pour traiter et gérer de manière éthique les données autochtones, il est difficile d'envisager comment Wikidata pourrait être librement employée pour utiliser ces données.

Le modèle de données pose également un problème pour les données autochtones et la représentation respectueuse des peuples, territoires et cultures autochtones sur Wikidata. Un exemple typique est le manque de ressources pour identifier adéquatement les peuples autochtones dans la plateforme. En effet, ceux-ci peuvent s'identifier de diverses façons et leur identité peut être liée autant à leur communauté qu'à leur nation. Wikidata n'a pas les moyens requis pour présenter les nations autochtones en tant que nations. Les nations autochtones ont plutôt été catégorisées selon leur appartenance ethnique. L'identité autochtone n'est pas une ethnie, et les travaux récents de Allison-Cassin (2023) et

de Johnston, Julian et Singh (2022) suggèrent que les problèmes structurels du modèle de données de Wikidata pour documenter les identités autochtones constituent de sérieux obstacles pour considérer Wikidata comme une plateforme appropriée pour traiter de l'information sur les peuples autochtones. Un autre problème lié à l'identification d'une communauté est le manque de liens entre les nations autochtones et les items Wikidata. Dans *Elements of Indigenous Style*, Greg Younging souligne à plusieurs occasions l'importance de reconnaître l'identité nationale autochtone pour parler des peuples autochtones. Il avance que la nation «est devenue généralement acceptée par les peuples autochtones pour décrire différents groupes autochtones distincts en tant qu'entités politiques» [traduction] (2018). De plus, Younging affirme que certains peuples autochtones s'identifient à une nation officielle, comme la Nation métisse d'Alberta. Cependant, Wikidata ne permet pas de lier les peuples autochtones à des nations, ce qui rend l'utilisation de Wikidata inappropriée ou discutable, spécifiquement en regard des interprétations issues de la DNUDPA et des diverses initiatives et pratiques liées à l'autodétermination.

Qui plus est, Wikipédia entraîne et amplifie certains problèmes structurels de Wikidata quant aux données autochtones. Bien que ces problèmes puissent être involontaires, ils représentent néanmoins des obstacles supplémentaires au traitement adéquat des données autochtones, engendrant ainsi un problème spécifique au contexte de Wikidata Canada. Un exemple marquant est la façon dont les zones géographiques sont structurées dans Wikidata, ce qui devient d'autant plus frappant pour les territoires

autochtones. En raison de la *Loi sur les Indiens*, de nombreux membres des Premières Nations ont été expulsés de leurs territoires traditionnels, qui ont été remplacés par un système de réserves. Généralement, ces réserves sont des zones géographiques occupées par les membres d'une seule Première Nation, et chaque réserve possède également son propre système de gouvernance, le conseil de bande, imposé par le gouvernement fédéral du Canada. Dans Wikidata, une confusion existe entre les « types » d'entités pour décrire les communautés des Premières Nations. Par exemple, pour décrire une Première Nation, la propriété « Nature de l'élément » pourrait être une bande de Premières Nations, un lieu géographique, voire une autre entité. Cette confusion dans la structure de données est souvent due à la création automatique d'éléments dans Wikidata à partir des articles de Wikipédia, où, sous forme narrative, il y a peu de confusion. Brown affirme que « [l]es logiques de description sont hostiles à l'ambiguïté » [traduction] (Brown, 2022, p. 3). L'ambiguïté qui subsiste autour de ce que l'on entend par « Première Nation » est un problème structurel qui engendre de graves problèmes dans la gestion des données. Pour une discussion approfondie sur ce sujet, voir Allison-Cassin (2023).

En ce qui concerne les données du festival folk Mariposa, il s'est avéré difficile de documenter l'identité des individus et de tenir compte des considérations discutées ci-dessus. Dans le rapport de la CAPACOA, *Indigenous Artists and Wikidata: Explorations and Consultations* (2022), les préoccupations générales exprimées par les personnes participantes quant à la pertinence de Wikidata pour leurs données de profils sont graves, malgré la taille restreinte du groupe de

consultation. Malgré son apparente ouverture sans parti pris, Wikidata a été conçu et structuré en fonction des conceptions occidentales du monde. Son modèle peut être changé et modifié par consensus communautaire, mais il n'en demeure pas moins que les membres de cette communauté n'incluent probablement que très peu de personnes autochtones. La portée de certains des problèmes est d'une telle ampleur que des interventions majeures pourraient être nécessaires. De plus, peu de personnes sont sensibilisées à connaître et comprendre les différents systèmes de connaissances, et il en va de même quant à l'importance de bien saisir les enjeux liés aux régimes de propriété intellectuelle.

## **Conclusion**

Le projet de données liées du festival folk Mariposa a permis d'expérimenter la création d'un réseau de données liées à partir de Wikidata. D'un point de vue canadien, les défis rencontrés en travaillant avec Wikidata soulignent des enjeux qui dépassent l'utilisation de la base de données pour des données culturelles canadiennes. Par exemple, d'importants enjeux persistent sur Wikidata quant à la place qu'offre la plateforme aux données autochtones, ce qui exige une intervention opportune et essentielle. De plus, la fermeture de journaux communautaires et de médias locaux se poursuit et nuit aux efforts de documentation des sujets canadiens des personnes qui travaillent sur les plateformes de Wikimédia. Au fur et à mesure que la plateforme Wikidata se développe et devient une plaque tournante pour les données structurées, les personnes au Canada devront élaborer des moyens pour offrir et promouvoir de nouvelles façons

d'ajouter des données sur la plateforme, plus spécifiquement en ce qui concerne les communautés qui ne sont pas bien représentées dans les données traditionnelles. Peut-être que Wikidata ne jouera jamais un rôle clair dans la vie quotidienne des gens; néanmoins, la plateforme joue un rôle important dans la transition vers des données ouvertes au Canada, comme en témoignent ses nombreuses et différentes utilisations au Canada. Pour que les institutions et les individus puissent y participer plus efficacement, une grande attention doit être portée aux métadonnées et aux pratiques numériques des organisations afin d'améliorer la disponibilité des identifiants stables, des données ouvertes et des sources de référencement pour les éléments Wikidata. En reconnaissant la nécessité de soutenir les initiatives de données ouvertes qui serviront toutes les personnes au Canada, plusieurs agences culturelles du gouvernement fédéral canadien utilisent déjà Wikidata: «Ceci permet de réduire le travail entourant plusieurs facettes de la gestion numérique des collections en mobilisant l'expertise et les mises à jour faites par d'autres équipes dont l'autorité est établie» (Gouvernement du Canada, 2023). Il convient aussi de rappeler que, «en tant que la plus grande base d'informations jamais rassemblées, Wikipédia reste à la fois une réalisation humaine stupéfiante et un puits sans fond d'améliorations, tant au niveau du contenu que de la communauté» [traduction] (McDowell et Vetter, 2021, p. 91). Wikidata continuera sans doute à se développer au Canada en développant et en soutenant la communauté. Comme la chercheuse canadienne en humanités numériques Susan Brown le suggère, «bâtir sa communauté et développer des collaborations intersectorielles sont probablement les

points les plus importants pour aborder efficacement les enjeux liés à la diversité, compte tenu des ressources substantielles et des infrastructures requises pour travailler avec les données ouvertes et liées» [traduction] (Brown, 2022). Un avenir fort et diversifié pour Wikidata au Canada renforcera la présence numérique du pays.

## CHAPITRE 5

# Wikimédia dans un musée d'art québécois

*Exploration d'un modèle d'institution culturelle ouverte*

NATHALIE THIBAUT

Le Musée national des beaux-arts du Québec (MNBAQ) s'est engagé, par une série d'initiatives et d'activités, à l'exploration de pratiques plus ouvertes et collaboratives par le biais de plusieurs chantiers wikimédiens dans la foulée des Open GLAM. Ces projets favorisent un accès à ses contenus, ses collections et ses données plus accessibles à l'aide des licences ouvertes ou libres. Le musée soutient la participation et la collaboration avec des publics élargis de chercheurs, d'éducateurs et d'autres institutions culturelles qui peuvent réutiliser, partager et modifier ces contenus et ses données.

Le Musée national des beaux-arts du Québec (MNBAQ) s'est engagé dans une démarche pour accroître l'accès à ses contenus et à ses données par le biais de trois chantiers en lien avec le mouvement Wikimédia. Ces travaux qui ont

débuté en 2018 sont toujours en développement. Il s'agit du chantier Wikipédia (rédaction d'articles prioritairement en français), du chantier Wikidata et du chantier Wikimedia Commons. La finalité principale de ces travaux est de valoriser et de partager l'information ainsi que des images libres de droits détenues par le MNBAQ sur le Web tout en acquérant, au plan organisationnel, de nouvelles compétences numériques. Ces initiatives rejoignent les différents objectifs du mouvement du libre accès, celui des Open GLAM ainsi que le modèle du musée participatif. La notion de « chantier » souligne l'engagement du musée d'agir comme maître d'œuvre d'un ensemble d'ouvrages interdépendants sur lequel reposeront la vision et l'infrastructure d'une institution culturelle ouverte.

De manière générale, les projets wikimédiens favorisent l'accès au contenu culturel numérique par le recours à des licences plus ouvertes. Par « ouvert », nous entendons le fait que « n'importe qui puisse librement accéder, utiliser, modifier et partager à n'importe quelle fin (sous réserve, des exigences qui préservent la provenance et l'ouverture) » [traduction] (Open Definition, OKF). Cet accès ouvert au patrimoine culturel numérique en tant que commun permet désormais au Musée de mieux répondre aux besoins et aux attentes de différents publics et de mieux soutenir le travail des chercheurs et chercheuses, comme celui de la communauté muséale, artistique et documentaire. La Commission européenne sur les Open GLAM a illustré les impacts diversifiés de l'ouverture des contenus et des données en soulignant :

l'importance du partage des connaissances pour la recherche, l'innovation et la créativité. Par exemple :

- Plus de contenu culturel sous licence ouverte permet aux enseignants à travers le monde de réutiliser ce travail en classe ;
- Plus de données culturelles ouvertes permettent aux chercheurs d'établir des liens entre les personnes, les choses et les événements grâce à l'utilisation de techniques innovantes telles que l'exploration ou la fouille de texte et la visualisation ;
- Plus de contenu culturel ouvert permet aux citoyens du monde entier de profiter de ce matériel, de comprendre leur héritage culturel et de réutiliser ces contenus pour produire de nouvelles œuvres d'art. [traduction]

### **Brève revue de la littérature et cadre conceptuel**

En 2006, alors que l'inquiétude au sujet de la validité des contenus wikipédiens dominait le débat public et scientifique, un article de recherche proposait pourtant un argumentaire et une méthode afin d'accroître la visibilité et la promotion des institutions de mémoire : on y encourage par le détail les musées à se doter d'une entrée dans Wikipédia (Bowen, 2006). C'est quelques années plus tard que le mouvement des Open GLAM a véritablement émergé, à la suite de la conférence Wikimania qui s'est tenue à la bibliothèque Alexandrina en 2008, un lieu et un contexte symboliques porteurs où Liam Wyatt, éditeur australien, a entrevu le potentiel de structurer dans la durée les relations entre les différentes institutions culturelles et Wikipédia. L'expression et la communauté GLAM-Wiki ont été forgées de concert, en 2009 à Canberra, à l'occasion d'une séance organisée par Wyatt et intitulée

«Finding The Common Ground» réunissant les acteurs et actrices des bibliothèques, musées, archives (Lih, 2018). Cet événement a permis de générer une conversation visant à mutualiser les capacités des institutions culturelles tout en convergeant avec le mouvement des communs. Peu après, Liam Wyatt devient en 2010 le premier wikipédien en résidence dans un musée, le British Museum, et s'investit à bâtir la relation et le partage d'expertise entre les univers muséaux et wikimédiens. Au British Museum, l'intérêt d'une telle collaboration avait été rapidement comptabilisé puisque l'article consacré à la vedette du musée, la pierre de Rosette, était déjà cinq fois plus consulté sur Wikipédia que sur le site Web du musée (Cohen, 2010).

Naguère, les publics, notamment les chercheurs et chercheuses, venaient en personne dans les musées, les centres d'archives et les bibliothèques consulter des documents et des images pour leurs travaux. Dans l'environnement numérique, les stratégies des musées, archives, bibliothèques visent désormais à élargir l'accès; elles proposent de nouveaux services en ligne et des pratiques de plus en plus ouvertes (Klein et Cardin, 2018, p. 101). Les approches se veulent plus participatives; les institutions culturelles explorent les possibilités de collaboration et, éventuellement, de cocréation en renouvelant la manière dont celles-ci assurent traditionnellement l'accès des utilisateurs et des visiteurs aux collections. Le concept de «musée participatif» devient un référentiel de transformation où les projets wikimédiens servent régulièrement de tremplin (Simon, 2010). Par conséquent, le concept et l'opérationnalisation de l'accès se déplacent et passent d'une dimension physique à une présence en ligne de même qu'à la prise en compte des capacités de collaboration et de

participation qui s'inscrivent dans le mouvement des musées participatifs, inclusifs, ouverts et dans celui des Open GLAM (Evans, 2024; Salgado et Marttila, 2013).

Pour une institution culturelle, le partage et l'ouverture des collections et de l'information dans l'environnement numérique apportent son lot d'enjeux et de défis ayant des effets qui dépassent la gestion du changement au sein de l'organisation. Les dimensions de gouvernance, de pouvoir, d'autorité et de contrôle qui caractérisent la tradition des musées, surtout nationaux, continuent de jouer un rôle prépondérant au plan de la gestion des collections numériques, bien que cette culture des institutions culturelles tende à bouger peu à peu en dépit des résistances (Fouseki et Vacharopoulou, 2013). Par ailleurs, il semble que les préoccupations anticipées au sujet de la perte de contrôle sur les collections ou d'une éventuelle réduction du nombre de visites en personne, et qui seraient tributaires de l'ouverture des contenus et des données, ne seraient pas fondées (Kapsalis, 2016). L'exploration progressive de différentes plateformes wikimédiennes offre alors l'occasion de repenser et d'actualiser le positionnement éthique des musées, tout particulièrement en ce qui concerne les questions de propriété et de droit d'auteur dont dépend cet accès reconçu et élargi pour les publics. Plus précisément, l'expérience wikimédienne entraîne des débats et des négociations sur la propriété des objets, des images, des données numériques ainsi que sur la place de leurs utilisations commerciales, publiques ou encore en tant que communs. La collaboration au sein des projets wikimédiens favorise ces usages émergents qui vont refléter « la nature participative, la philosophie et l'idéologie de chaque institution muséale » [traduction] et l'état ou le degré

de la transition institutionnelle (Fouseki et Vacharopoulou, 2013). Les différentes initiatives présentées, les trois chantiers, s'inscrivent dans ce cheminement et deviennent des marqueurs de cette identité transitionnelle du Musée à l'aune de l'ouverture et de la participation. Les trois chantiers constituent autant de cas qui témoignent des stratégies reliées à la façon d'ouvrir les collections, les informations, les données du musée ainsi qu'à la façon d'adapter les pratiques professionnelles en termes de participation, d'inclusion sociale, d'ouverture face à de nouveaux usages.

Au plan de l'évolution des besoins et des usages touchant les informations des musées, d'autres constats émergent. D'abord, le recours aux plateformes wikimédiennes et aux licences plus ouvertes qu'elles préconisent pour le contenu du patrimoine culturel numérique amplifie les possibilités pour les chercheurs et chercheuses issus de différents domaines en arts, en sciences humaines et sociales, tant pour l'accès que pour l'étude du patrimoine culturel (Terras, 2015). Ensuite, on observe que la consommation culturelle numérique est également susceptible d'étendre le registre des usages « en dehors du contexte artistique attendu, permettant de nouvelles formes de consommation hédonique et utilitaire » [traduction] (Navarrete et Villaespesa, 2020). En d'autres termes, les usages wikimédiens soutenus par les musées, comme le suggère le cas du Metropolitan Museum of Arts par exemple, amène une conceptualisation inédite du goût authentique qui prend en compte la consommation numérique, laquelle valorise les possibilités des usages plus informatifs d'une collection (Navarrete et Villaespesa, 2020).

En outre, le MNBAQ explore également diverses collaborations afin d'élargir l'accès à la réutilisation de ses données

et contenus par d'autres institutions principalement de type GLAM. Ces collaborations représentent un défi non négligeable où une stratégie de données ouvertes et liées, par l'entremise de Wikidata par exemple, apparaît comme un dispositif facilitant la recherche en « permettant aux chercheurs en humanités numériques d'établir des liens entre (et de donner un sens à) la multitude de patrimoines culturels numérisés disponibles sur la toile » [traduction] (Alexiev, 2018). Or, en se fondant sur ces différentes considérations qui s'appuient sur les missions renouvelées des musées et leurs objectifs explicités en matière de démocratisation, ainsi que sur les nouveaux usages individuels, communautaires comme institutionnels, cette exploration des facettes de Wikimédia en système ouvert s'est avérée pour le MNBAQ un choix opportun, productif et durable.

### **La transition du Musée vers un modèle d'institution culturelle ouverte**

Les trois chantiers du MNBAQ ont nécessité l'apport d'une équipe multidisciplinaire variée provenant de plusieurs secteurs au sein de l'institution : affaires juridiques, technologies de l'information, conservation, gestion des collections, conservation préventive, restauration, photographie, archives et documentation. Comme mentionné, la finalité première était de rendre accessibles en ligne de nouveaux contenus culturels numérisés à différents publics. Très rapidement, il est apparu nécessaire de diffuser des données synchronisées, fiables et vérifiées concernant les œuvres et les artistes représentés dans les collections du MNBAQ : collection permanente, collection d'étude, collection prêt d'œuvres

d'art, sans oublier le vaste volet des archives privées. Le Musée a une longue tradition de collaboration avec ses usagers et répond annuellement à un nombre grandissant de demandes d'information provenant de chercheurs ou chercheuses, d'étudiantes ou d'étudiants et aussi du grand public. Toutes ces personnes souhaitent avoir accès à des images et à des données descriptives des collections nationales et publiques. La diffusion de l'ensemble de la collection patrimoniale permet avantagement de rapprocher les citoyens et citoyennes des œuvres moins exposées ou n'ayant pas nécessairement fait l'objet de recherches approfondies. On estime à environ 25 % le nombre d'œuvres et d'objets déjà exposés ou reproduits dans une publication. La diffusion en ligne de contenus permet de découvrir des éléments archivés et souvent méconnus.

En 2016, l'ajout du pavillon Pierre Lassonde a généré une fréquentation importante sur place et en ligne. C'est à l'occasion des préparatifs d'ouverture de ce pavillon que le MNBAQ a repensé l'accès à ses collections en ligne. L'intention d'élargir et d'ouvrir l'accès à nos données et à nos informations s'est combinée à ce nouvel élan architectural.

Le MNBAQ est un musée d'État québécois qui possède à ce jour plus de 42 000 œuvres d'art, des milliers de documents d'archives, d'objets et de livres rares. Le catalogue répertorie plus de 4 500 artistes, collectifs d'artistes, artisans, artisanes, auteurs, autrices et autres représentés au sein de ses collections depuis son ouverture en juin 1933. Étant aussi éditeur, il a publié plus de 1000 catalogues d'expositions et de livres en lien avec ses acquisitions, ses expositions et ses autres opérations de diffusion. Ces activités d'édition ont généré un nombre important de contenus, de

données, de textes, de photographies et de documents d'archives à conserver et à partager dans l'avenir selon la législation entourant les droits d'auteur en vigueur.

La mission du Musée est de promouvoir, de faire connaître et de conserver l'art québécois de toutes les périodes, de l'art ancien à l'art actuel, et d'assurer une présence de l'art international par des acquisitions, des expositions et d'autres activités d'animation. Le Musée possède plusieurs œuvres d'art relevant du domaine public ainsi qu'un patrimoine artistique et documentaire important. Ce corpus est actuellement en processus de numérisation pour une future diffusion et un accès direct en ligne. Il paraissait logique pour le MNBAQ d'explorer des projets de diffusion Web dont les orientations et les objectifs seraient compatibles avec une approche favorisant une réutilisation par les citoyens et citoyennes ainsi que la communauté muséale, artistique et documentaire. Nous visions des projets numériques favorisant une actualisation et une synchronisation de l'information presque en temps réel ainsi qu'une souplesse de gestion et de suivi des processus au sein des équipes du MNBAQ. Comme d'autres organismes culturels, le Musée traverse depuis quelques années une période de transformation organisationnelle et technologique. Il veut atteindre une maturité numérique et une efficience par des accès multipliés grâce aux technologies et aux sciences de l'information. Il doit prendre en considération les normes documentaires, les standards d'interopérabilité, le développement du Web sémantique et des ontologies afin de parvenir à échanger et à agréger des informations à travers la réalisation de plateformes fédératrices présentement en révision (Réseau canadien d'information sur le patrimoine, Société des musées du Québec, Répertoire sur le

patrimoine culturel du Québec, Artistes au Canada, etc.). Nous préparons donc le terrain en explorant le potentiel des outils Wikimédia.

À la suite des réflexions amorcées en 2018 pour l'élaboration de sa stratégie numérique et grâce au financement du Plan culturel numérique du Québec, le MNBAQ a décidé d'explorer et d'expérimenter des projets de contenus en libre accès ainsi qu'en données ouvertes, considérant qu'il était un acteur muséal, documentaire et de recherche au sein de l'écosystème artistique et culturel du Québec. Le Musée gère une bibliothèque, un centre de documentation ainsi que des collections et des fonds d'archives privés dont il souhaite orienter la diffusion et la consultation vers le plus grand nombre possible d'utilisateurs. Avant de diffuser sur des plateformes Web, il considère le cadre juridique et gouvernemental québécois et canadien : le droit d'auteur, le respect et la protection des renseignements personnels, l'accès à l'information, la divulgation proactive, etc. Un important chantier de numérisation des collections en cours depuis 2018 a déjà entraîné la diffusion de milliers d'images sur le site des collections du MNBAQ.

Comme le mentionne Christelle Molinié du Musée de Bretagne : « En s'ouvrant aux pratiques collaboratives du Web 2.0, le travail de médiation documentaire revêt une nouvelle dimension qui favorise l'interaction, l'appropriation et la réutilisation des œuvres par le public, tout en poursuivant sa participation aux missions fondamentales muséales en termes d'étude, d'accessibilité et de diffusion des collections du domaine public et de démocratisation culturelle » (Frayse, 2015). Nous avons donc tenu compte de l'importance de l'accès gratuit et de la possible réutilisa-

tion de certaines données, d'informations publiques ou d'images libres de droits dans l'approche numérique de l'ensemble de ces chantiers au MNBAQ.

Les expériences récentes et les innovations numériques de Bibliothèque et Archives nationales du Québec (BA nQ) et de la Cinémathèque québécoise ont fortement influencé la réflexion du MNBAQ sur les chantiers documentaires à planifier et à tester au sein de l'organisation et pour la communauté sur le Web. Cette réflexion a mené à une approche d'ouverture, de partage libre de l'information, laquelle était jusque-là accessible uniquement dans les bases de données de l'organisation. Comme le mentionne Jessica de Bideran : « La présence sur Wikipédia repose sur deux mouvements a priori antagonistes pour une institution culturelle : d'une part, le désir d'atteindre des objectifs tangibles (de rédaction d'articles, de valorisation de fonds patrimoniaux et de visibilité de l'institution), d'autre part, la prise en compte d'une construction collaborative et évolutive de connaissances résultant des choix d'une communauté libre et autonome. » (Bideran et Wenz, 2020) Les différentes initiatives ont permis de développer de nouvelles compétences numériques et d'approfondir la sensibilisation aux enjeux de la culture libre, facilitant ainsi une meilleure articulation entre les approches muséales traditionnelles et participatives.

### **Premier chantier : exploration d'articles Wikipédia en français**

En 2018, le MNBAQ a exploré le potentiel de Wikipédia en français en intégrant directement les articles existants sur des artistes dans la nouvelle version de son site des collec-

tions. Un survol des sites de collections de plusieurs musées d'art a démontré l'intérêt d'exploiter Wikipédia en différentes langues. Le musée Tate Modern intègre directement sur son site un lien pour chaque artiste afin de diriger l'utilisateur vers l'article Wikipédia en version anglaise. L'approche multilingue est un réel avantage, car elle nous permet d'afficher les noms d'artistes dans leur langue d'origine – innue, japonaise, etc. – en plus du français.

Les essais et les preuves de concept ont consisté à tester un script informatique quotidien et automatisé. Ce script a permis d'actualiser une sélection de données publiques, et ce, au gré des mises à jour des contributeurs et contributrices à partir de Wikipédia en français. Une première expérience a été menée à l'été 2018 avec une sélection de 300 artistes québécois et canadiens dont les œuvres d'art sont exposées dans les pavillons du MNBAQ. Nous avons d'abord étudié le nombre et la qualité des articles présents dans Wikipédia en français. Le suivi, la veille et l'examen de ces articles ont été concluants en matière d'accès, de stabilité du script, de surveillance du vandalisme et surtout de la fiabilité des sources. Le chantier a pris son envol à l'automne 2018. Dès le départ, nous avons constaté l'importance d'offrir des formations aux membres du personnel au sujet de ce projet collaboratif afin de connaître leurs appréhensions et leurs perceptions, et d'orienter le projet selon certaines priorités : notoriété des artistes, sources publiées, œuvres exposées dans les pavillons du Musée, etc. Depuis 2019, le Musée ajoute des ébauches d'articles ainsi que des articles sur des artistes québécois et canadiens émergents avec la complicité et la collaboration de la communauté wikipédienne francophone à travers le monde. À ce jour, plus de 1000 articles en

français sont liés directement à notre site des collections. Ces articles sont évolutifs et consultés par des centaines d'internautes chaque jour. On trouve autant d'articles en anglais qu'en français sur les artistes des collections du MNBAQ. Certains artistes sont représentés dans plus d'une langue. Au total, 128 langues distinctes y figurent. Cet apport multilingue est précieux afin de favoriser la visibilité des artistes québécois et canadiens. L'organisation d'ateliers animés sur la façon de contribuer à Wikipédia a débuté en 2019. Tout en discutant des mythes entourant l'encyclopédie, nous voulions rendre accessibles plusieurs sources fiables et publiées – ouvrages, périodiques, articles de journaux, etc. – de la bibliothèque du MNBAQ pour documenter les sujets ciblés par les participants et participantes. Malheureusement, la pandémie a freiné ce volet sur place des ateliers. Il sera possiblement repris dans le futur avec des thèmes pour chacun. Une collaboration de rédaction d'articles en français dans le cadre d'un cours au collégial a été lancée. Comme le mentionne Martha Severo: « Cette tendance fait partie d'un phénomène appelé *sciences citoyennes*. Ce terme est devenu particulièrement à la mode ces dernières années; il vise à inclure toutes les initiatives facilitant la participation de non-professionnels à la recherche scientifique et à la construction des connaissances scientifiques. » (Barbe et Severo, 2021, p. 88)

## **Deuxième chantier : exploration de Wikidata**

Un deuxième chantier lié à l'ouverture des contenus s'est ajouté en juin 2019 et visait à combiner au chantier Wikipédia la révision de certaines données biographiques publiques

détenues par le MNBAQ dans son système de gestion des collections ainsi que dans son système intégré de gestion de bibliothèque. La comparaison, la normalisation, le nettoyage et l'actualisation des données de ces deux systèmes ont nécessité la collaboration de plusieurs collègues afin d'harmoniser celles-ci et de les comparer à celles déjà présentes dans la base de connaissances libre Wikidata. Cette base est éditée de manière collaborative et hébergée par la Fondation Wikimedia. L'élément Wikidata de chaque artiste débutant ou débutante a été intégré directement dans un des champs du système de gestion des collections du MNBAQ. Cet identifiant Wikidata permet la liaison directe des données extraites et visibles sur le site des collections et l'identification efficace des artistes grâce à des référentiels communs. Chaque élément Wikidata d'un ou d'une artiste rassemble et collige des identifiants uniques et pérennes (VIAF, ISNI), des autorités pérennes de la Bibliothèque du Congrès, de la Bibliothèque nationale de France, du Online Computer Library Center, de l'Union List of Artist Names, etc., ainsi que des liens directs vers le Répertoire du patrimoine culturel du Québec, l'identifiant du *Dictionnaire biographique du Canada*, le site Web officiel de chaque artiste, ses réseaux sociaux, etc.

Ces données ouvertes et structurées de Wikidata sont utilisées de façon bidirectionnelle par le MNBAQ, qui importe et verse de nouvelles données au gré de ses acquisitions d'œuvres d'art. Un outil de visualisation sous forme de tableau de bord aide à suivre l'évolution du projet. Il facilite l'ajout de nouvelles autorités d'artistes à créer au gré des semaines tout en conservant un historique des actions. Cet outil est très important, car il améliore la gestion de l'information diffusée, la traçabilité et la journalisation des actions sur

le plan de la mise à jour, tant des contributeurs et contributrices externes de données que de celles du personnel du MNBAQ. Le Musée a profité de la récente mise en place d'une agence ISNI au sein de BAnQ pour créer, avec la collaboration de cette institution, des identifiants ISNI pour certains artistes décédés. Nous préparons le terrain, car cet identifiant unique à portée internationale se révèle fort utile dans les projets de plateformes d'interopérabilité pour bien désigner les artistes et les collectivités et offrir la possibilité d'une chaîne de rétribution financière dans un avenir pas si lointain.

Les professionnels et professionnelles de l'information, l'équipe du registriat, les conservateurs et conservatrices du Musée ont ainsi révisé et approuvé le versement de milliers d'éléments et de propriétés dans la base de connaissance Wikidata sous licence libre. Cette tâche réalisée avec rigueur et minutie était nécessaire afin d'harmoniser les données publiques avec des sources publiées et des connaissances actualisées sur les artistes ciblés par ce chantier Wikidata. De plus, le potentiel de réutilisation de ces données par d'autres organismes ou personnes à long terme est considérable. À ce jour, les données versées dans Wikidata relient les artistes des collections du MNBAQ à plus de 1000 musées et collections à travers le monde.

Par ailleurs, ce chantier explore et précise la qualité de données et de mutualisation de contenus en accès libre. Il a suscité des échanges professionnels avec d'autres musées en Europe et aux États-Unis, et nous avons pu confirmer l'authenticité de certaines œuvres d'art en croisant les données avec celles de musées internationaux. Il assure une veille quotidienne sur plus de 7000 éléments dans Wikidata et facilite la mise à jour des données chaque jour dans nos sys-

tèmes de gestion des collections et de bibliothèque. De plus, des propriétés liées à l'existence de fonds d'archives et de dossiers documentaires sur des artistes québécois et canadiens ont été intégrées dans ce chantier Wikidata, tout comme la mention des musées québécois et canadiens qui détiennent des œuvres des artistes représentés au MNBAQ. Enfin, plusieurs éléments ont été créés pour des historiens et historiennes de l'art, des musées, des galeries, des expositions, etc., afin de bien lier les données décrites dans Wikidata.

En juin 2020, un contributeur wikimédien ainsi que la communauté de Wikidata ont proposé et obtenu la création d'un identifiant unique d'un ou d'une artiste du MNBAQ. Cette propriété englobe plus de 4000 éléments et permet désormais au Musée de réaliser une veille efficace des données sur les artistes représentés dans ses collections. De plus, le MNBAQ rejoint ainsi plusieurs musées d'art qui s'intéressent au partage de contenus et à la divulgation proactive de leurs données sur Wikidata et Wikipédia, dont le Musée des beaux-arts du Canada, le Cleveland Museum of Art, l'Art Institute of Chicago, le Smithsonian, l'American Museum of Art, le Whitney Museum of American Art, le Musée d'art Nelson-Atkins et le Musée d'art de Saint-Louis, pour ne nommer que ceux-ci.

Le fait que la communauté Wikimedia s'implique à son tour dans les projets a été un élément de motivation pour plusieurs collègues. Après trois ans de chantiers, il est possible de constater l'ampleur du travail collaboratif effectué sur les articles Wikipédia en français et sur les données Wikidata entamé par le MNBAQ. Chaque jour, plusieurs ajouts, modifications et ajustements sont réalisés sur les données, les propriétés et les textes.

Le MNBAQ collabore aussi au projet Wikidata *Sum of all paintings*, qui répertorie plus de 95 000 peintures à partir des collections de musées, d'institutions publiques et de collections privées à travers le monde entier. Ce projet lie les données des artistes à leurs peintures et devient parfois l'amorce d'un catalogue raisonné où l'on trie les données chronologiquement, ou par thématique ou sujet représenté.

Ce projet a conduit en décembre 2020 à l'usage d'un nouvel identifiant sur Wikidata pour le MNBAQ : l'identifiant numérique d'une œuvre d'art sur le site web du Musée national des beaux-arts du Québec<sup>1</sup>. Cet identifiant répertorie des données descriptives, des images, etc., pour chaque peinture détenue par le MNBAQ. Encore une fois, c'est la communauté qui a créé cet identifiant et rendu possible la réutilisation des données. Nous effectuons des ajustements, des modifications selon les recherches de l'équipe de la conservation. Ce projet spécifique a permis de découvrir les images versées dans Wikimedia Commons au cours des années. L'intérêt des contributeurs et contributrices pour des images libres de droits de certaines œuvres de nos collections est constant. Nous versons régulièrement de nouvelles images selon les besoins des usagers et les sujets de recherche en cours.

### **Troisième chantier : Images dans Wikimedia Commons**

Un troisième chantier est lié au potentiel de diffusion et de reproduction d'images sous licence libre. Le Musée a exploré

---

1. Identifiant MNBAQ d'une œuvre (P9091). Dans *Wikidata*. <https://www.wikidata.org/wiki/Property:P9091>.

et ciblé le versement de certaines images d'œuvres d'art, de documents d'archives, etc., dans le domaine public à partir de la médiathèque Wikimedia Commons. Le fait de permettre le téléchargement, la diffusion et la reproduction de certaines œuvres sans avoir à demander l'autorisation du MNBAQ est un grand test pour l'institution. La sélection des images pour ce chantier a été orientée vers les besoins des utilisateurs internes et externes du Musée. Avant ce projet d'intégration, l'équipe avait déjà versé des images d'œuvres d'art prises lors de leur exposition ou présentées dans les espaces publics du Musée. Les contributeurs et contributrices de Wikimedia Commons sont des photographes passionnés qui consacrent du temps à visiter des lieux et des espaces publics pour ensuite partager le fruit de leur travail pour le bien commun. Comme le mentionne Simon Côté-Lapointe dans son article « Les documents audiovisuels numériques d'archives » : « Tout un chacun est potentiellement un expert, un créateur, un collaborateur grâce à la démocratisation des outils par le numérique comme en font foi des initiatives telles que Wikipédia, YouTube et Internet Archive. Et ceci touche tout particulièrement l'audiovisuel grâce à la démocratisation des moyens de manipulations » (Côté-Lapointe, 2019).

Pour lancer de tels chantiers de divulgation et de partage d'informations, nous avons dû accepter que les données partagées ne soient pas complètes lors des premières itérations. Le travail de documentation et d'observation a permis d'acquérir de nouvelles compétences en termes de normalisation et de standardisation des données. L'approche multidisciplinaire de l'équipe du Musée a facilité le respect de l'échéancier, l'appropriation des chantiers et l'intégration

d'une nouvelle façon de travailler. Certains champs de notre base de données et leur concordance avec des propriétés dans Wikidata sont toujours en exploration : identité de genre, orientation sexuelle, nationalité versus citoyenneté, appartenance à un groupe ethnoculturel. Le réflexe est de se tourner vers les artistes concernés et de respecter leurs choix dans la divulgation ou non de certaines informations.

La portée de ces trois chantiers est tangible, car un pourcentage de référencement provient des actions numériques menées sur Wikipédia, Wikidata et Wikimedia Commons. On constate, grâce à des requêtes et des outils d'analyse, qu'un nombre croissant de consultations du site des collections proviennent directement de ces projets.

## **Discussion et conclusion**

En résumé, ces chantiers exploratoires mettent en place une base de connaissances numérique chez le personnel afin qu'ils puissent s'initier à l'éditorialisation de contenus sur le Web. Il faut prévoir un alignement pour le Web sémantique et des ontologies qui devront s'implanter dans le milieu patrimonial et culturel au cours des prochaines années. Un autre bénéfice est notable : la conversation avec les internautes est plus riche depuis la mise en ligne des collections, car ceux-ci participent directement à la documentation par leurs commentaires et leurs échanges avec le Musée. De plus, certaines données versées par le MNBAQ dans Wikidata rejoignent déjà d'autres projets en cours au sein de la communauté culturelle du Québec et du Canada sur les arts de la scène, les arts littéraires, dont certains projets collaboratifs dans des régions du Québec. Cet effort des communautés de pratique montre la volonté

d'assurer pour chacun et chacune – à partir de leurs bases de données, de la documentation et des sources publiées – une meilleure visibilité des contenus culturels québécois tout en s'appropriant des principes de normalisation et de futures interopérabilités à travers différentes disciplines. Des expérimentations comme celles des chantiers wiki au MNBAQ rapprochent petit à petit l'utilisateur d'une nouvelle mutualisation, des humanités numériques et du potentiel de visualisation et d'édition de ses données et métadonnées. La valeur de l'information dans une institution muséale ne fait que s'accroître grâce à ce type de projets en contenu libre à travers les apprentissages numériques réalisés. Ces chantiers tracent la voie vers l'intelligence collective et collaborative tout en valorisant les savoirs détenus par une organisation et en consolidant l'engagement dans le mouvement du libre accès et des Open GLAM.

## CHAPITRE 6

# Gouvernements ouverts

### *Un wiki pour les lier tous*

MIGUEL TREMBLAY

Wikidata, la base de données libre de la Fondation Wikimedia, permet de centraliser des métadonnées dans le cadre des gouvernements ouverts, ce qui facilite la fusion des catalogues de données ouvertes. Les avantages de telles initiatives entre les gouvernements et Wikidata sont nombreux, notamment en favorisant une meilleure accessibilité et la réutilisation des données, une uniformité accrue de celles-ci ainsi qu'une plus grande collaboration internationale. Le gouvernement du Canada offre un exemple concret de cet engagement avec le projet de téléversement des métadonnées météorologiques, lequel permet de préciser les phases qui mènent à ce modèle de gouvernement ouvert à l'aide de Wikidata.

Wikidata, la base de données libre de la Fondation Wikimedia, est un lieu de convergence idéal pour les métadonnées des gouvernements ouverts du monde entier (Krabina et Polleres, 2021; Pellissier *et al.*, 2016). En effet, regrouper des catalogues de données ouvertes permet de croiser les informations de ces jeux de données entre elles, mais aussi celles d'une centaine

de millions d'éléments déjà présents sur Wikidata (Hitz-Gamper *et al.*, 2019; Krabina et Polleres, 2021). Cette solution clés en main offre une infrastructure informatique, des méthodes et des accès logiciels pour combiner les données, et peut compter sur la participation bénévole d'une communauté active<sup>1</sup>. Une telle centralisation, visant les catalogues de données ouvertes des gouvernements, n'a pas été, à notre connaissance, explorée dans la littérature scientifique (Mora-Cantallops *et al.*, 2019).

Comment une telle centralisation pourrait-elle concrètement s'effectuer ? Nous identifions les premiers jalons pour progresser dans cette direction, nous listons également une série d'avantages et d'enjeux pour cette solution. Nous prenons en exemple un jeu de métadonnées du gouvernement du Canada qui a déjà été versé dans Wikidata. En nous inspirant de ce cas réel et récent, nous explorons la possibilité et la faisabilité d'ajouter d'autres métadonnées provenant de catalogues de données ouvertes.

## **Gouvernement ouvert, données ouvertes**

Les principes de la loi de Moore, selon laquelle la puissance des puces informatiques double tous les deux ans, s'appliquent également aux volumes de données informatiques.

Devant ces défis, les gouvernements se tournent vers l'infonuagique, en amenant l'utilisateur où se trouvent les données, plutôt que l'inverse.

---

1. L'une des particularités de Wikidata est de reposer sur une communauté bénévole. C'est cette force communautaire qui a fait en sorte que Google a migré les données de Freebase vers Wikidata. Voir Pellissier *et al.* (2016).

En parallèle, plusieurs administrations publiques s'inscrivent dans le mouvement de gouvernement ouvert (Estermann, 2018). Le gouvernement du Canada lance dès 2011 son premier plan d'action pour un gouvernement ouvert. Il compte parmi les premiers adhérents au Partenariat pour un gouvernement ouvert (PGO), un partenariat multilatéral entériné en 2011. En date de juin 2024, 75 pays et 150 gouvernements locaux – représentant plus de deux milliards de personnes – ainsi que des milliers d'organisations de la société civile sont membres du PGO. On peut parler d'un mouvement mondial.

Le premier axe développé lors de la mise en place des gouvernements ouverts est l'accès aux données, communément appelées *données ouvertes*.

### **Multitude de portails de données des administrations publiques**

Les données ouvertes des administrations publiques sont habituellement regroupées et rendues disponibles sous un portail unique. Elles sont répertoriées dans un catalogue qu'il est possible de consulter grâce à une nomenclature et à des outils de recherche spécifiques au site Internet (Tygel *et al.*, 2016). Le gouvernement du Canada présente son catalogue de données ouvertes sur son site Internet, où l'on peut faire des recherches parmi les quelque 40 000 métadonnées disponibles.

Chaque portail gouvernemental propose des outils pour identifier, et parfois visualiser, des jeux de données. Ceux-ci sont souvent regroupés par thématiques, comme la santé, la culture ou la technologie (Tygel *et al.*, 2016).

Les portails gouvernementaux regroupent parfois les descriptions de jeux de données de plusieurs administrations. Rien qu'au Canada, on décompte au minimum 66 portails de données ouvertes différents.

En plus de souffrir du manque d'uniformité de l'interface, des fonctionnalités et des catégories, plusieurs de ces portails ne sont disponibles qu'en langue anglaise, limitant la découvrabilité et l'accès aux données pour les personnes ne maîtrisant pas cette langue.

Pour trouver un jeu de données spécifique à chacun des gouvernements, un usager doit donc découvrir les portails, se familiariser avec les outils de recherche, identifier le jeu de données en question, et ce, seulement s'il maîtrise la langue anglaise. Bien que l'ouverture des données soit un pas dans la bonne direction pour le gouvernement ouvert, force est de constater que l'existence d'une multitude de portails crée des silos restreignant l'accès aux données.

### **Centralisation des métadonnées sur Wikidata**

En 2003, deux ans après la naissance de Wikipédia, la Fondation Wikimedia est créée. Son but : financer le soutien technique de l'encyclopédie libre. Progressivement, plusieurs projets s'y greffent, toujours dans l'objectif de rendre les connaissances accessibles à toutes et à tous (Yoakim, 2019).

Lancée en 2012 par le chapitre allemand de la Fondation Wikimedia, Wikidata est une base de connaissances libre, multilingue, éditée de manière collaborative, qui permet de répondre à des requêtes complexes et s'inscrit dans un contexte de données liées, à la base du Web sémantique. Son

contenu est placé sous la licence libre Creative Commons Zero (CCo)<sup>2</sup> favorisant la libre réutilisation des connaissances et des informations qu'elle contient.

En juin 2024, plus de 10 ans après sa création, Wikidata comporte plus de 110 millions d'éléments et plus de 23 000 utilisateurs et utilisatrices actifs. Ses données peuvent être visualisées au moyen de 50 outils externes. Plus de 500 articles ont été publiés à propos de Wikidata.

Plusieurs ensembles de données ouvertes ont été intégrés dans Wikidata, dont celles de Wikipédia, toutes langues confondues. Wikidata est propice à la centralisation des données ouvertes gouvernementales, en raison notamment de ses capacités de Web sémantique (Ajose-Ismail et Osanyin, 2019).

Ces capacités sont d'autant plus intéressantes que l'on trouve dans Wikidata des informations autant sur les entités territoriales administratives (1,9 million d'entrées) que sur les composés chimiques (1,2 million), ou sur les voies de communication (630 000 entrées), par exemple. Les métadonnées des publications scientifiques composent quant à elles plus de 30 % des entrées avec 22,6 millions de fiches descriptives.

En plus d'offrir l'ajout de données structurées à l'aide de script, Wikidata propose un protocole (SPARQL) pour rechercher, ajouter, modifier ou supprimer des données.

Wikidata est une infrastructure solide qui, comme le démontre l'exemple pour les articles scientifiques, est en mesure d'absorber des millions de métadonnées.

---

2. La licence CCo (Creative Commons Zero) est une licence libre Creative Commons permettant au titulaire de droits d'auteur de renoncer au maximum à ceux-ci dans la limite des lois applicables, afin de placer son œuvre au plus près du domaine public.

## Exemple du Service météorologique du Canada

L'échange de données météorologiques est historiquement un domaine où la collaboration entre les pays et les scientifiques a toujours prédominé. En effet, une prévision météorologique de plus de 48 heures doit ingérer des données provenant de partout sur la Terre. Et comme cela est vrai pour tous les endroits, cela favorise l'échange des données entre pays, car tous ont besoin de données pour effectuer une prévision sur leur territoire. Même au cœur de la guerre froide, les données météorologiques circulaient entre les pays du bloc de l'Est et du bloc de l'Ouest (Zillman, 2020). C'est notamment pour cette raison que ce secteur de connaissance est particulièrement bien adapté à l'idéal du partage de la connaissance de la Fondation Wikimedia.

En 2019, le Service météorologique du Canada accorde un financement à Wikimedia Canada pour téléverser les données d'observation et les métadonnées des stations météorologiques dans l'écosystème de la Fondation Wikimedia. Ces données ont été téléversées respectivement dans Wikimedia Commons et dans Wikidata (voir la carte des stations sur <http://surl.li/hvjswd>). Les métadonnées incluent un lien vers les données de la station météorologique correspondante dans Commons. Les métadonnées des stations météorologiques du Canada s'ajoutent à celles de 12 autres pays ayant plus de 10 stations dans Wikidata.

Il devient possible d'utiliser les outils et d'exploiter les différentes capacités de calculs et de requêtes de Wikidata. On peut ainsi croiser les métadonnées des stations entre elles, et créer ensuite des catégories de stations par altitude. On peut également croiser les métadonnées avec d'autres

informations déjà présentes dans Wikidata et, par exemple, identifier les stations météorologiques du Québec situées dans un rayon de 1 kilomètre d'un pont d'une longueur supérieure à 100 mètres. L'information sur les ponts existait déjà dans Wikidata lors de l'import des métadonnées des stations. Ces quelques exemples illustrent la puissance des outils de recherche de Wikidata.

À l'été 2021, en lien avec ce projet, Wikimedia Canada et l'Acfas ont accueilli deux stagiaires boursiers IVADO du programme Des données pour raconter.

Deux articles sont issus du travail collaboratif entre Laurence Taschereau (UQAM, journalisme) et Ali Akbar Sabzi Dizajyekan (Polytechnique Montréal, visualisation de données) sur les données météorologiques. Leur travail permet de sensibiliser le grand public aux effets des changements climatiques sur la population. Le premier article traite de résilience agricole (Taschereau et Sabzi Dizajyekan, 2021b) et le deuxième porte sur les îlots de chaleur (Taschereau et Sabzi Dizajyekan, 2021a).

Wikimedia Canada et l'Organisation météorologique mondiale (OMM) réfléchissent ensemble à la façon de téléverser l'entièreté des métadonnées des stations météorologiques enregistrées à l'OMM dans Wikidata. Des échanges ont également lieu avec des usagers du Cameroun et du Nigéria, pays où les capacités d'infrastructure informatique sont moins développées, afin de créer des produits d'analyses météorologiques.

## **Première phase à considérer pour un import dans Wikidata : le ciblage des métadonnées et la standardisation**

Chaque gouvernement ouvert suit une certaine standardisation pour décrire ses jeux de données. Les informations regroupent souvent, mais sans s'y limiter, la licence, l'entité (ministère) responsable du jeu de données, l'URL pour télécharger les données et la date de la dernière mise à jour. L'Open Data Charter travaille à établir une charte internationale des données ouvertes, qui couvre notamment l'interopérabilité afin que les données puissent être comparées et croisées, peu importe la source d'origine.

L'objectif est de convertir les métadonnées des catalogues, standardisées localement, en une forme qui correspond à ce que l'on retrouve dans Wikidata. Certaines métadonnées se prêtent mieux que d'autres à un import dans Wikidata, car elles adoptent une standardisation internationale, en plus d'une standardisation locale. Les convertir dans un format commun dans Wikidata est donc moins complexe.

Tout standard international est appuyé par une communauté composée d'utilisateurs et de contributeurs ou contributrices susceptibles de participer au projet d'adaptation du catalogue pour l'import dans Wikidata. En plus des données météorologiques citées plus haut, qui correspondent à ce critère, les données géoréférencées sont un autre exemple de données standardisées. Ces métadonnées incluent celles décrivant les zones administratives (politiques et de recensement) ainsi que les routes pour différents moyens de transport.

Rassemblées, ces métadonnées pourraient favoriser le déploiement d'une foule d'applications à l'échelle mondiale, ce qui est présentement impossible sans un fastidieux travail manuel.

## **Avantages de la centralisation des catalogues de données ouvertes**

### **Endroit unique pour accéder aux métadonnées**

Les gouvernements ouverts ont centralisé les données de leurs ministères et organismes sous un portail unique imposant une uniformisation des métadonnées des jeux de données, comm expliqué précédemment. La centralisation sur Wikidata est le simple prolongement de ce principe : créer un endroit unique où se trouvent non pas un catalogue, mais tous les catalogues des gouvernements ouverts, les rendant accessibles aux citoyens et aux citoyennes du monde entier.

### **Bonification d'une capacité existante**

La copie des métadonnées sur Wikidata ne vise pas à remplacer les différents portails de données ouvertes des gouvernements. Il s'agit d'un ajout de capacité aux portails existants. Wikidata ne doit pas devenir un point de défaillance unique en étant le dépôt officiel des métadonnées. Les gouvernements doivent idéalement conserver une copie, faisant autorité, de leurs métadonnées.

Les gouvernements demeurent responsables des données ouvertes qui resteraient sur le serveur informatique d'origine. Ceci permet d'effectuer des mesures de performance et de

bénéficier des mesures de contrôle habituelles. Wikidata, tout comme Wikipédia, requiert de citer les sources des informations. Cette fonctionnalité préserve le lien entre la copie sur Wikidata et la source gouvernementale originale.

Les métadonnées, une fois recopiées dans Wikidata, bénéficient d'un plus grand rayonnement en étant décrites à plus d'un endroit sur Internet, où sont colocalisées les métadonnées d'autres gouvernements.

### **Surveillance et révision des métadonnées**

Le principe du wiki est de conserver chaque version des documents et de rendre possibles la visualisation et la surveillance des changements. Wikidata applique ce principe. La personne responsable des métadonnées peut ainsi avoir dans sa liste de surveillance associée à son compte Wikidata toutes les entrées correspondant à un catalogue. Elle peut être avisée immédiatement par courriel de toute modification et, le cas échéant, la réviser. Les wikis permettent donc à n'importe qui de contribuer, mais toutes les modifications peuvent faire l'objet d'un contrôle.

Dans le cadre de l'import des données d'observations météorologiques du Service météorologique du Canada, aucune modification inutile ni vandalisme n'ont été constatés depuis cinq ans. En fait, les dépositaires des données ont plutôt été avisés d'erreurs dans les données, leur offrant ainsi la possibilité de les corriger.

## Information multilingue et standardisée

L'interface de Wikidata est disponible en plus de 300 langues, chacune bénéficiant d'une communauté qui discute des termes à utiliser pour décrire les entrées encyclopédiques. Ceci constitue plus qu'une traduction, comme en témoigne le déploiement de l'interface en langue atikamekw, qui a impliqué la communauté et les gardiens de la langue atikamekw afin de choisir ou de créer de nouveaux mots (Casemajor *et al.*, 2017). En plus de l'interface, les éléments eux-mêmes peuvent être traduits en plus de 300 langues. Les métadonnées du libellé « stations météorologiques » peuvent, par exemple, être vues en arabe (محطة رصد جوي) ou en biélorusse (метэаралагічная станцыя).

L'utilisation de noms alternatifs (alias) pour désigner les éléments dans chacune des langues est une fonctionnalité qui distingue Wikidata de Wikipédia. Cette caractéristique permet de réconcilier des nomenclatures différentes dans la gestion des étiquettes, évitant un problème identifié dans d'autres bases de données ouvertes (Tygel *et al.*, 2016).

Le multilinguisme de Wikidata est vu comme une valeur ajoutée pour les organisations qui souhaitent y partager leurs données (Evans, 2024). L'interface multilingue de Wikidata constitue un avantage considérable par rapport aux portails des gouvernements ouverts, habituellement disponibles dans une langue seulement (deux dans le cas du Canada). Cela favorise en outre la collaboration entre les gens qui ne possèdent pas de langue commune. Les locuteurs des différentes langues du monde peuvent chercher et utiliser les jeux de données de leur choix en utilisant une

interface dans leur langue (voir l'exemple sur Wikipedia : <http://surl.li/wgygeb>).

### **Infrastructure infonuagique et communauté déjà en place**

Wikidata est une base de données disponible sur une infrastructure fiable, financée et gérée par la Fondation Wikimédia. La plateforme fait l'objet d'une surveillance régulière et les capacités sont augmentées en fonction des besoins, incluant l'intégration de nouvelles technologies. Les gouvernements désirant apporter leur contribution avec leurs métadonnées bénéficient ainsi de serveurs déjà existants et les usagers, d'un accès gratuit aux données.

Bien qu'il soit difficile d'évaluer combien d'éléments contiendrait Wikidata si des gouvernements ajoutaient leurs métadonnées – à titre d'exemple, le gouvernement du Canada dispose de 40 000 entrées dans son catalogue de données ouvertes –, il est raisonnable de penser que l'infrastructure serait en mesure de tolérer ces imports massifs. À des fins de comparaison, le nombre d'entrées dans Wikidata pour les articles scientifiques est supérieur à 22 millions. Le dialogue avec la communauté Wikidata demeure impératif afin de prévenir une levée de boucliers à la suite d'un import massif d'éléments. Le caractère incrémental du processus laissera de la place à la négociation afin de satisfaire la communauté.

Les résultats des requêtes des usagers dans Wikidata peuvent être visualisés de plusieurs manières : dans des tableaux, dans des graphiques ou sur des cartes. L'exportation des résultats sous différents formats est possible de même que la réutilisation des données dans d'autres systèmes.

Enfin, il existe plusieurs applications externes pour effectuer des recherches ou des calculs dans Wikidata. Une page répertorie ces applications, qui fonctionnent souvent par champ thématique. Il est possible de développer des outils spécifiques aux métadonnées des gouvernements ouverts.

## **Défis de la centralisation**

### **Correspondance des informations, transfert de données et mises à jour**

La première étape pour copier des métadonnées dans Wikidata, et potentiellement la plus importante, consiste à créer une correspondance entre la typologie des catalogues de données ouvertes et leur représentation dans Wikidata. Ceci nécessite une connaissance et une maîtrise du contenu des catalogues et de Wikidata. Une telle tâche requiert des ressources humaines et du temps, et peut s'étendre à la création de propriétés dans Wikidata en collaboration avec la communauté.

Une fois les correspondances créées, on téléverse les métadonnées dans Wikidata, idéalement à l'aide de scripts automatisés. Cette tâche nécessite des connaissances en programmation, un élément clé pour brancher un catalogue à Wikidata.

Lorsque le catalogue est mis à jour du côté du portail gouvernemental, il faut assurer la synchronisation des informations dans Wikidata, et donc téléverser les métadonnées ajoutées ou modifiées. Ceci peut se faire à intervalle régulier (toutes les semaines, mois ou années), ou au besoin. Inversement, un mécanisme de rétroaction doit être mis en

place afin de communiquer les changements effectués dans Wikidata aux responsables des métadonnées gouvernementales. La communauté peut déceler des erreurs ou des incongruences dans les métadonnées lors de leur utilisation (Krabina, 2023). Le système de suivi des éléments, déjà présent sur Wikidata et mentionné précédemment, peut être utilisé pour effectuer cette rétroaction.

La clé de voûte de l'import massif de données gouvernementales dans Wikidata est donc la communication régulière entre la communauté et ses homologues gouvernementaux. L'idéal est d'obtenir le soutien d'une communauté de pratique avec laquelle établir des liens de confiance. Le code source des scripts devrait être libre pour partager les meilleures pratiques.

### **Soutien aux usagers**

Un plus grand partage de métadonnées pourrait amener un plus grand nombre d'utilisateurs à utiliser les données ouvertes. C'est d'ailleurs l'objectif ultime. Le revers de la médaille d'une augmentation de popularité est l'augmentation de la demande de soutien technique aux usagers, demande à laquelle les gouvernements devraient en théorie répondre par des outils adéquats.

### **SPARQL ou l'art de la requête**

Le langage de requête dans Wikidata, nommé SPARQL, demande un niveau d'expertise avancé pour être utilisé à son plein potentiel et pour qu'on réussisse à identifier les informations recherchées. Bien que des efforts soient

déployés pour simplifier l'interface, notamment grâce à l'outil en ligne de construction de requête pour Wikidata (Wikidata Query Builder), il est nécessaire de posséder des compétences spécifiques en programmation informatique pour rédiger des requêtes complexes. Ceci dit, il est possible de demander de l'aide à la communauté d'utilisateurs et d'utilisatrices qui possèdent ces connaissances, mais il peut s'écouler plusieurs semaines avant d'obtenir une réponse.

Pour pouvoir recouper les métadonnées des catalogues de données ouvertes avec d'autres données dans Wikidata, les responsables des métadonnées doivent mettre au point une documentation incluant notamment des exemples typiques de requêtes SPARQL. Cette documentation peut être hébergée directement dans Wikidata.

Les outils d'intelligence artificielle génératifs, comme ChatGPT, ont grandement aidé à démocratiser la rédaction de requêtes SPARQL. En posant une question en langage naturel, l'outil traduit cette question en code SPARQL adapté pour Wikidata. Cela permet aux utilisateurs, même sans compétences avancées en programmation, d'accéder facilement aux données recherchées et de tirer pleinement parti des capacités de Wikidata (Lubiana, 2023).

Afin d'optimiser l'expérience utilisateur, des outils sur mesure peuvent être créés dans l'écosystème de la Fondation Wikimedia ou ailleurs sur le Web. Les usagers bénéficient alors d'une interface intuitive et plus facile à utiliser que la formulation d'une requête sous forme de code.

## Licence des métadonnées

Le contenu de Wikidata est placé sous la licence CCo, sous laquelle les propriétaires renoncent à leurs droits d'auteur au maximum permis par la loi, pour se rapprocher des caractéristiques du domaine public.

Les gouvernements qui téléversent les métadonnées dans Wikidata doivent renoncer à leurs droits d'auteur. Ceci ne devrait pas, en principe, être un enjeu, car il est question de la description des jeux de données, et non des données en elles-mêmes. Les métadonnées représentent une description factuelle d'information plutôt qu'une plus-value reliée à la création des données. Il y a souvent peu de valeur ajoutée dans ces informations. Cependant, des embûches administratives sont à prévoir selon les législations.

## Absence de catégories standardisées

La plupart des portails de données ouvertes présentent les jeux de données par thématiques. Pourtant, il n'existe pas de catégorisation standardisée (Pellissier Tanon et Kaffee, 2018). À cette absence d'uniformité conceptuelle s'ajoute le problème de leur traduction dans différentes langues (Farda-Sarbas et Müller-Birn, 2019; Pellissier Tanon et Kaffee, 2018).

Les métadonnées doivent être classées selon des propriétés spécifiques à Wikidata. Cet effort de classification soulève des défis de nomenclature, de logique et d'ontologie. On peut travailler en collaboration avec une communauté de champs d'intérêt pour tenter de répondre à ces questions et pour faire évoluer les réflexions. Il existe, par exemple,

une communauté s'intéressant aux arts de la scène qui pourrait répondre aux questions reliées au milieu culturel.

## **Granularité**

Dans la gestion de métadonnées, la question du niveau de granularité est omniprésente. À quel niveau doit-on représenter les jeux de données? Par regroupement en utilisant le plus bas niveau d'abstraction possible ou, au contraire, en allant au plus haut niveau de granularité possible? Les solutions mitoyennes sont également valables.

Pour reprendre l'exemple des stations météorologiques, doit-on publier une seule métadonnée, contenant les quelque 8 000 stations répertoriées, ou au contraire une entrée par station? Les deux? La différence de volume de métadonnées à entrer, dans ce cas, est de plusieurs ordres de grandeur.

Un des avantages liés au plus haut niveau de granularité des métadonnées est l'utilisation des outils de Wikidata à leur plein potentiel dans l'identification des données. Cela exige un travail d'envergure pour établir la correspondance des métadonnées entre le catalogue des gouvernements et Wikidata.

Il serait surprenant qu'une règle générale puisse être identifiée. Les propriétaires des métadonnées doivent effectuer des choix. À l'instar de la catégorisation, cette tâche pourrait se faire en consultation avec la communauté de champs d'intérêt, lorsque celle-ci existe.

## Conclusion

L'intérêt initial pour le projet des métadonnées météorologiques a été surprenant. Nous avons été contactés par des organisations gouvernementales ou paragouvernementales, et des institutions de la société civile intéressées par la connaissance libre (bibliothèques, OBNL, entreprises privées). Cet intérêt grandissant nous a encouragés à étendre la réflexion à d'autres jeux de données, car les principes sous-jacents à l'import massif et à la diffusion des données météorologiques peuvent s'y appliquer.

La notoriété de Wikipédia et des autres projets de la Fondation Wikimédia favorise la compréhension et l'acceptation par les responsables étatiques. L'attrait de l'utilisation de ressources informatiques déjà existantes est un atout, ressources dont les méthodes d'accès, de recherche et de diffusion sont reconnues par les administrations publiques et les utilisateurs de données ouvertes en général.

Une telle infrastructure facilite en outre la collaboration entre différents paliers de gouvernement, ou encore entre les États, car les questions de coûts d'infrastructure et, surtout, de politique de gestion et de financement de cette infrastructure sont entre les mains d'un tiers : la Fondation Wikimédia. Il ne reste alors pour les gouvernements qu'à investir le capital humain nécessaire pour utiliser cette infrastructure, et bien sûr répondre aux incontournables questions ontologiques.

Nous avons esquissé un portrait des avantages et des défis liés à une centralisation des métadonnées dans Wikidata. La prochaine étape sera d'identifier les métadonnées dont la mise en commun pourrait bénéficier à toute la société. Il

s'agira ensuite de s'investir à plus long terme pour explorer et baliser le sentier conduisant à une plus grande diffusion des données ouvertes.

Les enjeux de découvrabilité sont les principaux freins à la diffusion des données ouvertes. Investir dans de nouvelles solutions et se joindre à une communauté ou à des projets déjà existants sont des actions en accord avec les valeurs à la base du mouvement de démocratie ouverte. Il ne reste qu'à effectuer le deuxième pas.

Les outils d'IA générative, tels que ChatGPT, vont aider à explorer les données pour répondre à diverses questions. Cependant, pour que ces outils soient réellement efficaces, il est indispensable que les informations de base soient accessibles. Il est crucial que ces données soient fiables et sourcées afin d'assurer que les réponses fournies ne reposent pas uniquement sur des inférences statistiques, mais sur des données précises et vérifiées. Ainsi, l'accessibilité et la qualité des données sont des prérequis incontournables pour tirer pleinement parti des capacités des IA génératives dans la gestion et l'analyse des données ouvertes.



AXE III

**Littératies**





## CHAPITRE 7

# Les savoirs publics lors de l'infodémie COVID-19

*La littératie en santé et l'effet de Wikipédia*

DENISE SMITH

*Traduction de Nina Duque*

La pandémie de COVID-19 s'est accompagnée d'une « info-démie » de désinformation en ligne. Or, Wikipédia s'est affirmé durant cette période trouble comme un espace numérique exempt de fausses informations en dépit du caractère ouvert du site, que chacun est invité à améliorer. Ce chapitre avance que l'encyclopédie collaborative a un rôle important et méconnu à jouer pour la littératie en matière de santé. Sans être une panacée, le fonctionnement de Wikipédia et la synthèse d'information relative à la santé qu'il donne à lire s'avèrent plus instructifs pour améliorer sa littératie que ne saurait l'être le recours à Docteur Google.

La pandémie de COVID-19 est arrivée à une époque où la production d'information atteint une fréquence et un volume sans précédent. Parallèlement, le flux continu d'informations et la diversité des plateformes de médias sociaux

ont intensivement focalisé notre attention sur la pandémie. Ces deux éléments du paysage informationnel ont permis au public de suivre en temps réel l'évolution des savoirs sur cette nouvelle maladie infectieuse. Cependant, cette exposition aux processus habituels du développement des connaissances scientifiques et médicales, avec ses hauts et ses bas, a suscité une grande méfiance parmi la population.

Au début de la pandémie de COVID-19, les politiciens et les responsables de la santé publique ont ouvertement condamné la mésinformation circulant sur le sujet (CBC Ideas, 2021; Patel, 2021), à tel point que le terme « info-démie » a refait surface, rappelant son utilisation initiale en 2003 lors de l'épidémie de SRAS (Merriam-Webster, s. d.). Malgré les défis posés par l'omniprésence de la mésinformation sur la COVID-19, ce chapitre mettra en lumière un autre aspect souvent négligé, mais déterminant révélé par la pandémie : la littératie en santé. Bien que la crise de désinformation liée à la COVID-19 ait eu des effets négatifs sur la santé publique (Islam *et al.*, 2020), il est irréaliste de s'attendre à éliminer les fausses informations du paysage médiatique. Une réponse plus efficace consiste à fournir au public les compétences et les outils nécessaires pour identifier et détecter la mésinformation. Néanmoins, une initiative d'éducation publique massive visant à enseigner à la population comment évaluer de manière critique les informations en santé dans divers contextes présente inévitablement des défis uniques. C'est ici que Wikipédia entre en jeu.

Wikipédia est devenu un leader dans la communication d'informations sur la santé, s'imposant comme une ressource clé pour les données liées à la COVID-19 (Benjakob, 2020; Benjakob *et al.*, 2022; Cohen, 2020). En plus de son potentiel

à atténuer la crise de mésinformation, le dispositif technique de Wikipédia, notamment l'archivage de toutes les versions, permet de documenter de manière exhaustive la pandémie (Harrison, 2020; Yang et Tanaka, 2023). Wikipédia a pareillement été exploré comme une ressource pour la surveillance épidémiologique et une mesure de l'intérêt du public pour la maladie (Chrzanowski *et al.*, 2021; Lozzi *et al.*, 2020).

Il est bien connu que le public s'appuie fortement sur Internet pour obtenir des informations sur la santé (Heilman *et al.*, 2011), et que les connaissances en la matière varient considérablement, influencées par de nombreux facteurs. Le potentiel de Wikipédia à réduire les inégalités dans les connaissances relativement à la santé au sein de la population générale ne doit pas être sous-estimé. Souvent actualisé, Wikipédia permet de comprendre facilement une situation en évolution et fournit des informations qui peuvent soit clarifier une situation, soit aider à la prise de décisions. Le lectorat peut raisonnablement s'attendre à y trouver des renseignements objectifs et fondés sur des données probantes, rédigés dans un langage simple et accessible. De plus, les directives d'édition de Wikipédia, surveillées par la communauté, œuvrent à créer une source d'informations qui résume les données médicales de haute qualité provenant de sources fiables et sûres.

À ce titre, les quatre arguments suivants guideront ce chapitre :

1. Les connaissances en matière de santé sont un facteur médiateur déterminant l'état de santé.
2. Le public dépend d'Internet pour obtenir des informations sur la santé. Toutefois, nous devons posséder des connais-

- sances approfondies en matière de santé pour évaluer et utiliser correctement les informations trouvées.
3. Les niveaux de connaissances relativement à la santé varient, et ces inégalités sont influencées par des facteurs socioéconomiques.
  4. Wikipédia peut réduire le besoin de littératie critique et interactive en matière de santé nécessaire pour distinguer les preuves fiables de la mésinformation.

Le présent chapitre introduira le concept de littératie en santé et expliquera l'importance de Wikipédia dans la lutte contre les inégalités sur le plan des connaissances en santé révélées par la pandémie de COVID-19. Nous examinerons la place de Wikipédia comme source principale d'information sur la pandémie et évaluerons son application dans le contexte canadien. Le chapitre se conclura par un résumé des aspects encore méconnus de Wikipédia comme ressource d'information sur la santé et décrira les recherches futures sur la relation inexplorée entre les personnes qui consomment des informations sur la santé et Wikipédia.

## **La littératie en santé**

La littératie en santé n'est pas un nouveau concept. Cependant, avec l'arrivée de la COVID-19, elle a été propulsée à l'avant-scène en raison des discussions autour de la mésinformation. Souvent réduite à sa dimension fonctionnelle, la littératie en santé se réfère à la capacité d'une personne à appliquer des compétences premières pour adopter des comportements élémentaires en matière de santé, comme comprendre les formulaires de rendez-vous ou les

directives d'ordonnance. Cependant, ce concept englobe bien plus que cela. Ce chapitre se concentrera plus particulièrement sur la littératie interactive et critique en santé (Nutbeam, 2000, 2008). La littératie interactive en santé fait référence à l'habileté d'extraire de l'information, d'en tirer du sens à partir de diverses formes de communication et d'appliquer ces nouveaux renseignements à des circonstances changeantes (Nutbeam, 2000). La littératie critique en santé, quant à elle, suppose des compétences cognitives plus avancées afin d'analyser l'information et l'utiliser dans le but d'exercer un plus grand contrôle sur les événements et situations de la vie (Nutbeam, 2000).

Une personne ayant une littératie en santé avancée est plus autonome dans la recherche, la compréhension, l'évaluation et l'application des informations sur la santé dans sa vie quotidienne. Elle dispose des outils requis pour déterminer le type d'information dont elle a besoin, trouver et récupérer cette information, puis évaluer sa fiabilité et sa crédibilité. Elle est également capable de comparer l'information trouvée avec d'autres sources pour choisir les preuves les plus fiables. Dans le contexte de la COVID-19, cela signifie qu'une personne ayant des compétences interactives et critiques en littératie en santé peut explicitement identifier les questions qu'elle se pose, puis suivre un cheminement réfléchi afin de repérer des informations pertinentes. De la même manière, de nombreuses personnes rencontrent des informations sur la COVID-19 sans les chercher activement. Quel que soit le chemin emprunté pour découvrir l'information, l'individu doit être capable de vérifier la crédibilité et la fiabilité de l'information, de l'appliquer à ses savoirs préexistants et de comprendre son rôle potentiel.

La littératie en santé est au cœur de l'infodémie liée à la COVID-19. Toutefois, cette crise ne réside pas uniquement dans la prévalence de la mésinformation, mais également dans le manque de compétences du public pour identifier et utiliser correctement les informations appropriées sur la santé. En outre, la littératie en santé est reconnue comme un déterminant social de la santé. Des données récentes montrent que renforcer la littératie en santé, parallèlement à d'autres initiatives, pourrait réduire les disparités et promouvoir l'équité dans le domaine (Nutbeam et Lloyd, 2021). Une revue de la littérature suggère par ailleurs que le niveau de littératie en santé d'une personne reflète son expérience dans l'utilisation de l'information, son aisance avec la technologie, son niveau de scolarité et son statut socioéconomique (Martzoukou et Sayyad Abdi, 2017). Internet constitue donc une importante source d'informations sur le sujet pour différents groupes de personnes, et le besoin de littératie en santé s'accroît avec l'accès du public à l'information en ligne (Martzoukou et Sayyad Abdi, 2017). Ces résultats s'accordent avec les connaissances actuelles sur les fractures numériques et sur la façon dont les structures sociales existantes, ainsi que le statut socioéconomique d'une personne, peuvent influencer autant l'accès à des informations de santé adéquates, que la perception de la capacité des individus à accéder à cette information (Bodie et Dutta, 2008; Smith, 2021).

## **Wikipédia et la COVID-19 : réduire le déficit de littératie en santé**

Wikipédia montre un potentiel prometteur pour réduire les disparités en matière de littératie en santé. La littératie en santé interactive et critique nécessite des compétences cognitives avancées afin que les chercheurs ou les personnes qui utilisent de l'information puissent évaluer la qualité, la pertinence et l'utilité des informations disponibles. Dans le contexte de la COVID-19, toute personne disposant d'une connexion Internet a probablement été submergée par une surcharge d'informations. De nouvelles connaissances liées à la pandémie sont générées et partagées à une vitesse qui dépasse la compétence de quiconque à suivre le rythme. En outre, le bruit de la désinformation peut étouffer les connaissances scientifiques méthodologiquement solides, créant ainsi une expérience accablante. Cela complique la compréhension de la pandémie, du virus ou de la maladie, même pour ceux qui possèdent des connaissances en santé adéquates ou avancées.

Malgré tous les efforts déployés par des organisations et des particuliers bien intentionnés pour créer des sites Web de haute qualité et crédibles (par exemple, Santé Canada, MedlinePlus, MayoClinic, NHS), Wikipédia reste l'une des ressources d'information sur la santé les plus souvent consultées sur Internet (Birkun, 2024; Heilman *et al.*, 2011; Heilman et West, 2015; Shafee *et al.*, 2017). Elle est considérée comme une ressource précieuse et largement utilisée dans divers contextes, tant pour le public que pour les médecins, les personnes enseignantes ou étudiantes ainsi que les chercheurs (Smith, 2020). Les premières recherches

sur l'utilisation de Wikipédia pour obtenir des informations en santé montrent que la familiarité générale avec cette plateforme – les internautes la connaissent et la reconnaissent dans les résultats des moteurs de recherche – contribue à l'étendue de son audience et à la popularité de son contenu médical et de santé (Huisman *et al.*, 2020; Smith, 2023a).

La qualité du contenu médical de Wikipédia ne peut pas être évaluée dans son ensemble. En effet, « les articles de Wikipédia sont des morceaux individuels d'un ensemble plus large [...] une mosaïque [...] où certains éléments sont plus complets que d'autres » [traduction] (Smith, 2020). Cependant, son potentiel à combler les lacunes en matière de littératie en santé repose sur ses lignes directrices communautaires autogérées pour la création et la maintenance du contenu. Les bénévoles qui contribuent doivent se conformer aux directives strictes de la communauté, auxquelles ils sont tenus responsables par d'autres bénévoles. Parmi ces directives figure l'exigence de maintenir un point de vue neutre dans les articles. Le respect de cette neutralité peut minimiser la nécessité pour le lectorat d'évaluer l'article pour des biais ou de rechercher des contrepoints. Idéalement, toutes les informations pertinentes se trouvent dans l'article, offrant aux personnes qui lisent une image claire des connaissances disponibles, présentée de manière impartiale.

L'exigence de vérifier les informations ajoutées avec des sources fiables de haute qualité et des directives spécifiques pour les sources médicales limite la nécessité pour le lectorat d'évaluer la qualité des sources utilisées dans les articles de Wikipédia. L'évaluation des sources médiatiques et académiques citées dans un article est présumée être

effectuée par les wikipédistes qui les ajoutent. De plus, la crédibilité et la fiabilité des sources sont négociées par la communauté éditoriale, créant ainsi un environnement informationnel où le lectorat consomme des informations qui ont déjà fait l'objet d'un certain degré d'examen. Wikipédia n'est pas non plus un lieu pour créer de nouvelles connaissances. Comme telle, elle se présente comme une ressource pour apprendre sur les connaissances existantes, sans obliger le lectorat à évaluer rigoureusement son contenu comme il le ferait pour des études primaires ou secondaires, concernant des biais potentiels ou de rigueur méthodologique.

Wikipédia est une ressource unique qui permet aux individus d'avoir une vue d'ensemble sur le virus de la COVID-19, les vaccins, la pandémie et bien plus encore. Dès le début de la pandémie, Wikipédia a retenu l'attention des médias pour plusieurs raisons. D'abord, la rapidité avec laquelle de nouvelles informations ont été ajoutées aux articles pertinents était impressionnante. Ensuite, la couverture mondiale de la pandémie était vaste et exhaustive. Enfin, le dévouement constant de ses bénévoles à la mise à jour des contenus a également été remarqué (Bedirian, 2021). Ce qui distingue Wikipédia, c'est son engagement à créer du contenu «au service de l'intérêt public» [traduction] (Keton et Humborg, 2021). En juin 2021, une source médiatique canadienne a publié un article remettant en question la réputation de Wikipédia, perçu depuis des décennies comme peu fiable. L'article soulignait que toutes les sources d'information devraient être soumises à un examen équitable (Ibrahim, 2021). Les informations sur la COVID-19, résumées sur Wikipédia, sont constamment

surveillées par des wikipédistes, mises à jour à mesure que nos connaissances évoluent et sont appuyées par des citations sûres et vérifiées. En 2020, après la déclaration de la COVID-19 comme pandémie mondiale, des études ont montré une augmentation significative du nombre de bénévoles qui contribuent et de leur productivité sur Wikipédia (Rupprechter *et al.*, 2021). Cette mobilisation a permis de maintenir des informations précises et à jour, renforçant ainsi l'encyclopédie comme ressource précieuse pour comprendre la pandémie et ses impacts.

Wikipédia peut atténuer les inégalités en matière de littératie en santé en aidant les individus à distinguer les différentes sources d'information et à évaluer les renseignements sur la santé et les risques personnels (Nutbeam et Lloyd, 2021). Cependant, elle ne peut pas résoudre entièrement ces inégalités. Elle n'a jamais prétendu être une solution aux iniquités concernant la littératie en santé au sein de la population générale. Néanmoins, son modèle éditorial en fait une ressource prometteuse, accessible et axée sur la santé des personnes. Idéalement, Wikipédia fournit des renseignements médicaux impartiaux et fondés sur des faits, soumis à un système interne de révision par les pairs afin de garantir leur vérifiabilité et leur fiabilité, réduisant ainsi une grande partie du travail mental pour les personnes qui la consomment (Smith, 2023b). Bien qu'il incombe toujours à celles-ci de vérifier chaque page de Wikipédia individuellement (Epstein, 2022), le travail effectué dans le cadre du Wiki-Project COVID-19 allège cette responsabilité.

Certes, d'autres sites Web sur la santé concernant la COVID-19 sont souvent actualisés, citent des preuves de haute qualité et utilisent un langage simple et engageant.

Cependant, le fait que les personnes utilisatrices choisissent fréquemment Wikipédia parmi ces ressources ne doit pas être négligé. Les personnes ont tendance à se tourner vers Wikipédia, même si ce n'est pas toujours leur source d'information préférée (Sebelefsky *et al.*, 2015; Smith, 2023a). Wikipédia est généralement sélectionné à partir des résultats des moteurs de recherche par ceux qui recherchent des informations sur la santé, en raison de sa familiarité, de sa facilité de lecture et de sa couverture complète (Huisman *et al.*, 2020; Smith, 2023b). Cette préférence s'aligne sur les théories sur la recherche d'information, en particulier le principe du moindre effort. Ce principe stipule que « les personnes investissent peu dans la recherche d'informations, préférant des sources accessibles et faciles à utiliser à des sources de haute qualité qui sont moins faciles à utiliser, ou bien moins accessibles » [traduction] (Fisher *et al.*, 2005, p. 4).

Enfin, malgré les grands progrès réalisés par Wikipédia afin de documenter l'évolution d'une pandémie mondiale et résumer les connaissances disponibles à mesure que la situation évolue, les inégalités au sein du contenu de l'encyclopédie ne doivent pas être ignorées. La documentation de la COVID-19 sur Wikipédia a profité à un large éventail de personnes, mais comme dans le monde hors ligne, elle est réalisée par un groupe d'individus privilégiés aux expériences de vie similaires. Il est largement reconnu que les wikipédistes de la version anglaise de l'encyclopédie forment un groupe homogène, avec une diversité limitée. Cela signifie que des voix importantes sont absentes de la communauté d'édition.

Un exemple marquant est la mauvaise documentation des expériences uniques des populations autochtones du

Canada dans l'article sur la pandémie au pays. De nombreux articles liés à la pandémie dans les provinces ou les territoires canadiens contiennent peu ou pas d'informations sur les Premières Nations, les Inuits ou les Métis. Pourtant, ces populations ont été prioritaires pour la vaccination dès le début de la pandémie, continuent de faire face à des défis uniques en matière de contrôle de la transmission du virus et, dans certaines collectivités nordiques ou éloignées, rencontrent encore des difficultés à encadrer un afflux de nouveaux cas en raison de leur capacité limitée. Plusieurs facteurs contribuent à cette lacune. S'ils dépassent la portée de ce chapitre, ils n'en sont pas moins importants pour la discussion. Parmi eux, on peut citer le manque de représentation autochtone parmi les personnes qui contribuent à Wikipédia au Canada, ainsi que des perceptions incompatibles quant à ce qui constitue la fiabilité d'une source entre, d'un côté, les recommandations générales de Wikipédia sur la fiabilité des sources et celles spécifiques au domaine biomédical et, de l'autre, les modes de connaissance autochtones.

### **Wikipédia et la COVID-19 dans le contexte canadien**

À une époque où les compétences en littératie en santé sont souvent limitées, Wikipédia peut jouer un rôle décisif dans la communication de santé publique. Cependant, Wikipédia a-t-elle vraiment été bénéfique au Canada durant la pandémie? Explorons son impact sur la COVID-19 dans le contexte canadien.

En décembre 2020, Wikipédia comptait 6 950 articles sur la COVID-19, rédigés en 188 langues, grâce aux contribu-

tions de plus de 97 000 wikipédistes éditeurs (Wikimedia Foundation, 2020). Ces pages ont été vues près de 580 millions de fois, avec un record de plus de 10 millions de vues en une journée, le 23 mars 2020 (Wikimedia Foundation, 2020). En juin 2024, plus de 2 400 articles en anglais sur Wikipédia étaient consacrés à la COVID-19 ou faisaient partie du WikiProject COVID-19 (Wikipedia, 2021). Les articles les plus consultés portaient sur la COVID-19 (la maladie), le SARS-CoV-2 (le virus), les vaccins contre la COVID-19 et la pandémie de COVID-19. Communément appelés les « big four », nous les désignerons par l'expression les « quatre grands articles » dans la suite de notre propos. Ensemble, ces quatre articles ont été consultés plus de 60 millions de fois entre janvier 2020 et décembre 2022, avec une moyenne de plus de 86 000 vues par jour lors de la première année de la pandémie. Ces chiffres ne tiennent pas compte des traductions de ces articles dans d'autres langues. Par exemple, les articles en français et en espagnol traitant des mêmes sujets ont cumulé respectivement plus de 7,7 millions et 21,2 millions de vues entre janvier 2020 et décembre 2022.

L'article sur la pandémie de COVID-19 au Canada, dans la version de langue anglaise (COVID-19 pandemic in Canada), est semi-protégé, ce qui signifie que seuls les wikipédistes possédant un compte ayant effectué au moins 10 modifications peuvent y contribuer. Le WikiProject COVID-19 lui a attribué une note de qualité de classe B et d'importance moyenne. Au moment de la rédaction de ce chapitre (juin 2024), l'article nécessitait des mises à jour étant donné qu'il avait reçu la majeure partie de ses contributions au début de la pandémie. Créé le 26 février 2020, avant que l'OMS ne

déclare l'épidémie de coronavirus une pandémie, cet article s'intitulait originalement « 2020 Coronavirus Outbreak in Canada ». Les statistiques de la page indiquent que, jusqu'en juin 2024, 604 wikipédistes ont contribué à rédiger l'article avec un total de 4 500 modifications. L'article a été consulté plus de 1,8 million de fois, dont 1,7 million de vues entre sa création et décembre 2022. Au début de la pandémie, en mai 2020, la page a enregistré une moyenne de 10 000 vues par jour. L'article compte 6 000 mots et utilise 295 références uniques. Cette page a également servi de base de données pour Google. En novembre 2021, une recherche d'informations sur le thème « Covid-19 Pandemic in Canad » a permis de générer des statistiques sur les nouveaux cas et les cas actifs de COVID-19 grâce, entre autres, aux données de Wikipédia.

L'Ontario est la province la plus peuplée du Canada. L'article sur la pandémie de COVID-19 dans la version de langue anglaise en Ontario (COVID-19 pandemic in Ontario) est également semi-protégé. Il a obtenu une note de qualité de classe B, mais a été classé de « faible importance » par le WikiProjet COVID-19. Il a été créé le 13 mars 2020, deux jours après que l'Organisation mondiale de la Santé a déclaré la COVID-19 une pandémie mondiale. Selon les statistiques de la page, 344 bénévoles ont contribué à éditer cet article plus de 6 200 fois. Consulté plus de 970 000 fois, il recevait en moyenne plus de 4 800 vues par jour en mai 2020. En juin 2024, l'article comptait 4 500 mots et contenait des citations de 196 références uniques. Comme pour l'article sur la pandémie de COVID-19 au Canada, une recherche Google menée en novembre 2021 sur le thème « Covid-19 pandemic in Ontario » a permis de générer des statistiques établies à

partir des données de Wikipédia, en collaboration notamment avec d'autres sources telles qu'Our World in Data.

Il est juste de dire que Wikipédia était fréquemment consulté pour obtenir des informations sur la pandémie, tant dans le contexte canadien qu'ontarien. Cette affirmation est soutenue par le nombre élevé de vues sur ces articles et par l'engagement des personnes utilisatrices, stimulé par les résultats de recherche Google qui exploitaient les données de Wikipédia. Les deux articles de Wikipédia discutés précédemment montrent un certain niveau d'engagement dans le contexte canadien, mais ils se limitent à résumer les connaissances sur la pandémie. Il est probable que les personnes résidant au Canada figuraient également parmi les utilisateurs et utilisatrices fréquentant les « quatre grands articles » sur la COVID-19. Cette tendance se reflète par ailleurs dans les statistiques de consultation de la page sur les vaccins contre la COVID-19. Bien que cela ne soit pas la seule raison de la hausse des consultations, à la fin du mois de mai 2021, le gouvernement de l'Ontario a annoncé qu'il raccourcirait l'intervalle entre la première et la seconde dose des vaccins, passant d'un intervalle de trois mois à seulement huit semaines (CBC News, 2021). À l'international, le vaccin de la pharmacie Johnson & Johnson avait été approuvé au Royaume-Uni (Taylor, 2021) et les médias américains avaient également proposé l'idée que les employeurs imposent des obligations vaccinales (Diaz, 2021). Lors de cette période, les articles sur la pandémie de COVID-19 au Canada et en Ontario sont restés relativement stables, tandis que l'article sur le vaccin contre la COVID-19 avait connu une augmentation significative des visites, passant de 20 033 vues le 28 mai 2021 à 47 282 vues le 29 mai 2021. Tous ces changements sur

les vaccins, amplement couverts par les médias, pourraient avoir contribué à cette augmentation. Cependant, il ne s'agit que d'une corrélation positive, et une relation causale ne peut pas être établie, car les données de consultation sur Wikipédia manquent de contexte.

Les données et statistiques de Wikipédia, bien qu'utiles, sont limitées. Elles ne fournissent un aperçu de l'engagement que par l'activité de consultation ou de contribution des personnes utilisatrices. On pourrait facilement soutenir que, puisque Wikipédia était souvent consulté, il pourrait jouer un rôle dans l'atténuation de certaines inégalités en matière de littératie en santé, inégalités qui peuvent perpétuer le phénomène de l'infodémie. Toutefois, si nous ignorons comment les individus interagissent avec le contenu de Wikipédia (par exemple, ce qu'ils font des informations auxquelles ils ont accès), notre compréhension du potentiel de Wikipédia à atténuer l'infodémie se résumera à faire des rapprochements entre ce que l'on connaît de la littératie en santé et des données parfois difficiles à contextualiser.

### **Ce que nous ne savons pas**

Les motivations et les habitudes d'utilisation du contenu de santé et médical de Wikipédia sont peu étudiées. Les théories sur le comportement et les pratiques informationnels pourraient nous éclairer sur ce qui incite une personne à rechercher ou à utiliser des informations en ligne. En 2012, Johnson et Case ont décrit le profil type des grands chercheurs d'information comme étant des femmes blanches d'âge moyen, membres de groupes socioéconomiques élevés et très instruits. Ce profil est particulièrement pertinent

lorsque l'on considère la corrélation entre la littératie en santé et le statut social, telle qu'elle a été décrite par Nutbeam et Lloyd (2021). Bien qu'il soit important de mieux comprendre comment les personnes utilisent le contenu de santé et médical de Wikipédia, il est tout aussi important de noter que toute exploration de cette utilisation restera limitée au segment de la population ayant le privilège d'accéder à cette technologie et possédant des compétences adéquates en littératie et en compréhension de lecture.

Le comportement informationnel en santé peut être défini comme les variables clés qui influencent la recherche, la réception, l'évitement, le partage, la gestion ou l'utilisation des informations de santé, ainsi que les résultats qui en découlent (Smith, 2021). Une étude a découvert de manière fortuite des données riches sur l'utilisation de Wikipédia pour les informations de santé (Huisman *et al.*, 2020). Cependant, au moment de la rédaction, la seule enquête intentionnelle de ce type avait été menée par l'auteur de ce chapitre (Smith, 2023a). Cette étude, menée à partir d'entretiens avec des individus ayant utilisé les informations de santé et médicales de Wikipédia, a été réalisée dans le but de contextualiser qualitativement cette utilisation. Bien que les résultats n'apportent pas de nouvelles connaissances sur l'état actuel de la littératie en santé et ne positionnent pas Wikipédia comme une solution potentielle aux disparités en matière de santé, ils constituent toutefois une première étape importante. Cette recherche permet de mieux comprendre comment et pourquoi le contenu de santé de Wikipédia est utilisé, offrant ainsi la possibilité d'approfondir notre compréhension de la littératie en santé et du comportement informationnel.

## Conclusion

Wikipédia n'est pas la solution aux inégalités en matière de littératie en santé ni, par extension, aux disparités observées en matière de santé et de soins causées par des facteurs socioéconomiques. Elle ne génère pas directement de compétences en littératie en santé. Cependant, elle peut effectivement réduire certains obstacles significatifs à l'accès à l'information sur la santé, obstacles qui sont souvent engendrés par une littératie en santé inadéquate.

Cette compétence est un déterminant clé de la santé, souvent influencée par diverses variables sociales et économiques. Compte tenu des facteurs qui contribuent aux compétences avancées en littératie en santé, de la disponibilité et de l'accessibilité de Wikipédia au grand public, de ses processus éditoriaux rigoureux et de son émergence comme leader de l'information sur la santé pendant la pandémie de COVID-19, nous pouvons avancer que Wikipédia a un rôle important à jouer afin d'assurer un accès équitable à des informations de santé, surtout à une époque où les disparités socioéconomiques sont amplifiées. Les « quatre grands articles » de Wikipédia liés à la COVID-19, combinés à son contenu fréquemment consulté et à sa couverture consciencieuse de la pandémie dans le contexte canadien, montrent que Wikipédia s'est positionné comme une source fiable d'informations sur la santé, exprimées dans un langage clair. Par ailleurs, les informations publiées sont soutenues par des données de haute qualité provenant de sources réputées.

La nature même de la création et de l'organisation du contenu de Wikipédia la positionne comme un acteur clé

dans la lutte contre la désinformation. En matière de littératie en santé, notre plus grand défi n'est pas seulement l'omniprésence et la création de désinformation, mais également le déséquilibre du public dans sa capacité à évaluer de manière critique l'information en santé, à s'engager et à prendre des décisions fondées sur ces informations. Des recherches approfondies sur la manière dont le grand public utilise Wikipédia sont nécessaires. Elles pourraient fournir un aperçu plus clair du potentiel de Wikipédia à s'attaquer efficacement aux inégalités en matière de littératie en santé et d'accès à l'information. Ainsi, Wikipédia pourrait partiellement combler les lacunes en la matière, surtout à une époque où l'accès équitable à l'information est plus crucial que jamais.



## **Postface**

PIERRE LÉVY

### **La valeur du savoir vérifié à l'heure de l'IA générative**

La plupart des utilisateurs quasi quotidiens de Wikipédia, dont je suis, n'ont généralement pas conscience de l'importance, aussi bien théorique que pratique, des débats qui ont lieu en interne parmi les contributeurs et qui touchent à des enjeux sociaux, culturels et épistémologiques essentiels. Après lecture de cet ouvrage sur la culture libre et le mouvement Wikimedia au Canada, on aura pris connaissance de fascinantes analyses sur les problèmes complexes que les auteurs et éditeurs de Wikipédia doivent affronter. On aura aussi découvert un aperçu de la mine d'information en croissance qu'est Wikidata, un entrepôt de métadonnées sémantiques qui sert aujourd'hui de squelette à de nombreux systèmes d'intelligence artificielle (IA) symbolique<sup>1</sup>. Je voudrais justement commencer cette postface par une citation

---

1. L'IA symbolique repose sur l'utilisation de règles logiques et de réseaux sémantiques explicites (par opposition à l'utilisation de méthodes statistiques).

de Denny Vrandečić, un des initiateurs de Wikidata, qui avait travaillé sur le graphe de connaissances de Google comme ontologiste et qui est aujourd’hui le chef de file du projet « Abstract Wikipedia » visant à rendre les données des articles de Wikipédia indépendants des langues (c’est-à-dire traduisibles dans toutes les langues). Denny a déclaré dans sa communication à la Knowledge Graph Conference de mai 2023 que « [d]ans un monde au contenu infini, la connaissance prend toute sa valeur » [traduction]. Ce monde où les contenus sont potentiellement infinis résulte évidemment de l’usage désormais massif de l’intelligence artificielle générative. Parmi tous les problèmes que posent cette nouvelle situation, citons-en deux, particulièrement prégnants du point de vue de l’accès à la connaissance. Premièrement, malgré les usages qui sont faits des modèles génératifs pour obtenir rapidement des réponses directes, il faut réaffirmer que, contrairement à l’IA symbolique classique du xx<sup>e</sup> siècle, l’IA statistique (dite aussi neuronale) d’aujourd’hui – limitée à ses seules capacités – n’offre aucune garantie de vérité. GPT-4 comme les autres modèles du même genre *ne sont pas des bases de connaissances*. Les erreurs de fait et de raisonnement sont nombreuses et il suffit d’être spécialiste d’un domaine pour constater les faiblesses de ces IA, comme nous le faisons d’habitude lorsque nous lisons un article portant sur un de nos domaines de compétence lorsqu’il est rédigé par un journaliste pressé qui se contente de reformuler la doxa dans laquelle il baigne. Les réponses probabilistes de ChatGPT sont seulement *vraisemblables*. Deuxièmement, comme les modèles d’IA générative sont entraînés sur les données du Web et que celles-ci sont de plus en plus rédigées et illustrées par les modèles en question, on se trouve en

présence d'un dangereux cercle vicieux, et cela d'autant plus que les travailleurs au rabais chargés d'aligner les modèles et de redresser leurs biais ou erreurs utilisent eux-mêmes des IA génératives pour accomplir leur tâche ! Pour s'extraire de ces sables mouvants épistémologiques, il est donc nécessaire d'investir plus que jamais dans la construction de sources d'information fiables. En d'autres termes, l'explosion des usages de l'IA générative, loin de nous dispenser d'alimenter et d'utiliser Wikipédia, Wikidata et d'autres bases de connaissances vérifiées, rend l'effort d'y contribuer et le plaisir de les consulter encore plus nécessaires ! Ceci dit, comme nous le verrons dans la suite de ce texte, l'IA neuronale a néanmoins vocation à jouer un rôle positif dans le partage du savoir cher aux wikipédiens.

Comme je ne suis pas un chercheur spécialisé dans l'objet Wikimédia, mais plutôt un philosophe qui s'attache à penser le numérique, je me contenterai de quelques réflexions en guise de coda sur la relation triangulaire entre l'intelligence collective, l'intelligence artificielle et la noble finalité de mettre la connaissance à la portée de tous.

## **L'intelligence collective**

Étroitement définis comme des moyens de résoudre des problèmes, les processus d'intelligence collective se présentent sous de multiples formes, dont les plus étudiées sont les espèces statistique, délibérative et stigmergique (Lévy, 1994 ; Baltzersen, 2022).

Au début du xx<sup>e</sup> siècle, en Angleterre, le savant Francis Galton visitait une foire agricole. Un concours avait été organisé : on demandait aux 800 participants, la plupart

éleveurs, de deviner le poids d'un bœuf. Mais aucun d'eux n'avait trouvé le poids exact. Galton fit la moyenne de toutes les évaluations et trouva qu'elle était beaucoup plus proche du poids réel qu'aucune des estimations individuelles. La «sagesse» de la foule était supérieure à chacune des intelligences isolées (Galton, 1907). Cette forme d'intelligence collective statistique – ou comptable – suppose que les individus ne communiquent pas entre eux et ne se coordonnent d'aucune manière. Elle fonctionne d'autant mieux que la distribution des choix ou des prédictions s'étend sur un large spectre, de sorte que les erreurs et les biais individuels se compensent. Ce type d'intelligence collective suppose paradoxalement l'ignorance mutuelle. Elle s'exprime dans les sondages ou dans les élections, lorsqu'il est interdit de communiquer les résultats partiels avant que tout le monde ait voté. Cette approche de l'intelligence collective statistique sans connexion entre ses membres a été notamment popularisée par James Surowiecki dans son livre *La sagesse des foules*.

Une seconde forme d'intelligence collective, délibérative, repose au contraire sur la communication directe entre les membres d'une communauté. Elle résulte de l'échange des arguments et des points de vue. Face à un problème commun, elle peut converger sur un consensus ou se partager entre quelques solutions dont on a tous ensemble pesé le pour et contre. Pourvu que l'écoute soit au rendez-vous, chacun amène son point de vue, sa compétence particulière, qui enrichit le débat général (Zara, 2021 ; Mulgan, 2017). Ce type d'intelligence collective, apparemment idéale puisqu'elle est ouverte et réflexive, est d'autant plus difficile à mettre en œuvre que la collectivité est étendue. Il faut alors

établir des formes de hiérarchie et de délégation, certes indispensables, mais qui troublent la transparence de l'intelligence collective à elle-même.

Je voudrais maintenant introduire une forme d'intelligence collective moins connue, mais qui n'en est pas moins à l'œuvre dans nombre de sociétés animales et qui a trouvé son plus haut degré d'achèvement dans l'humanité: la communication stigmergique. L'étymologie grecque explique assez bien le sens du mot « stigmergie »: des marques (*stigma*) sont laissées dans l'environnement par l'action ou le travail (*ergon*) de membres d'une collectivité, et ces marques guident en retour – et réciproquement – leurs actions (Heylighen, 2016). Le cas classique est celui des fourmis qui laissent une traînée de phéromones sur leur passage lorsqu'elles ramènent de la nourriture à la fourmilière. L'odeur des phéromones incite d'autres fourmis à remonter leurs traces pour découvrir le butin et ramener des vivres à la ville souterraine en laissant par terre à leur tour un message parfumé. Le langage confère à l'humanité un haut degré d'intelligence collective, supérieur à celui des autres mammifères et comparable à celui des abeilles ou des fourmis. Comme d'autres espèces eusociales, nous communiquons en grande partie de manière stigmergique, mais au lieu de marquer un territoire physique au moyen de phéromones ou d'autres types de signaux visuels, sonores ou olfactifs, nous laissons des traces symboliques. Au fur et à mesure de l'évolution culturelle, les signifiants s'accumulent dans des *mémoires externes* de plus en plus perfectionnées: pierres levées, totems, paysages sculptés, monuments, architectures, signes d'écriture, archives, bibliothèques, bases de données. On peut prétendre que toute forme d'écriture qui n'est pas précisément adressée est une forme de communication stigmergique: des

traces sont déposées pour une lecture à venir et font office de mémoire externe d'une communauté.

Les différents processus d'intelligence collective qui viennent d'être évoqués, statistiques (sans communication), délibératifs (à communication directe) et stigmergiques (à communication indirecte) ne s'excluent évidemment pas l'un l'autre et peuvent fort bien se succéder ou se combiner. Par exemple, les wikipédiens se coordonnent par l'intermédiaire de bases de données communes, délibèrent *et* votent.

À l'échelle de l'espèce, l'intelligence collective humaine se situe dans la continuité de l'intelligence collective animale, mais elle est plus perfectionnée à cause du langage, des techniques et des institutions politiques, économiques, légales et autres qui nous caractérisent. La principale différence entre les intelligences collectives animale et humaine tient à la culture. Dans une dimension diachronique, notre espèce est entraînée par une vitesse d'apprentissage supérieure à celle de l'évolution biologique. Nos savoir-faire s'accumulent et se transmettent d'une génération à l'autre grâce à nos mémoires externes, au moyen de systèmes de signes, de conventions sociales et d'outils. Aucun individu ne serait « intelligent » s'il n'héritait pas des connaissances créées par les ancêtres. Dans une dimension synchronique, nous participons à une intelligence collective coordonnée où résonnent et se relancent l'architecture conceptuelle de nos mémoires communes et l'organisation sociale de nos collectivités. La définition réciproque des identités et la reconnaissance des problèmes se décident à ce métaniveau de la culture. Au-delà des procédures utiles (stigmergique, statistique et délibérative) pour résoudre des problèmes, il

existe donc une intelligence collective plus holistique, qui circonscrit les capacités cognitives d'une société.

L'évolution culturelle a déjà franchi plusieurs seuils d'intelligence collective. En effet, les inventions de l'écriture, de l'imprimerie et des médias électroniques (enregistrement musical, téléphone, radio, télévision) ont déjà augmenté de manière irréversible nos capacités de mémoire et de communication sociale. Le surgissement d'une communication globale par l'intermédiaire de la mémoire numérique est probablement le plus grand changement social des 25 dernières années. Cette nouvelle forme de communication par lecture-écriture distribuée dans une mémoire numérique collective représente une mutation anthropologique de grande ampleur. Plongés dans le nouvel environnement numérique, nous interagissons par le moyen de la masse océanique de données qui nous rassemble. Les encyclopédistes de Wikipédia et les programmeurs de GitHub collaborent par l'intermédiaire d'une même base de données. À notre insu, chaque lien que nous créons, chaque étiquette ou mot-clic apposé sur une information, chaque acte d'évaluation ou d'approbation, chaque « j'aime », chaque requête, chaque achat, chaque commentaire, chaque partage, toutes ces opérations modifient subtilement la mémoire commune, c'est-à-dire le magma inextricable des rapports entre les données. Notre comportement en ligne émet un flux continu de messages et d'indices qui transforment la structure de la mémoire et contribuent à orienter l'attention et l'activité de nos contemporains. Nous déposons dans l'environnement virtuel des phéromones électroniques qui déterminent en boucle

l'action des autres internautes et qui entraînent par-dessus le marché les neurones formels des IA.

### **L'intelligence artificielle comme augmentation de l'intelligence collective**

Abordons maintenant le thème de l'intelligence artificielle, mais sous l'angle – qui paraîtra peut-être insolite à quelques lecteurs – de l'intelligence collective. Les journalistes et le grand public ont tendance à classer dans « l'intelligence artificielle » les applications considérées comme avancées à l'époque où elles apparaissent. Mais quelques années plus tard, ces mêmes applications, devenues banales et quotidiennes, seront le plus souvent réinterprétées comme appartenant à l'informatique ordinaire. Par-delà les titres apocalyptiques et les images de jeunes femmes au cerveau chromé censées illustrer l'intelligence artificielle, nous assistons depuis le milieu du xx<sup>e</sup> siècle à un processus de réification formelle et d'extériorisation des fonctions cognitives. L'augmentation de puissance et la baisse des coûts du matériel distribuent ces fonctions cognitives objectivées dans l'ensemble de la société. Des machines interconnectées enregistrent et retrouvent de l'information, effectuent des calculs arithmétiques ou algébriques, simulent des phénomènes complexes, raisonnent logiquement, respectent des syntaxes et des systèmes de règles, extraient des formes à partir de distributions statistiques enchevêtrées... L'informatique automatise et socialise nos capacités de communication, nos facultés de mémoire, de perception, d'apprentissage, d'analyse et de synthèse.

Du fait même de son nom, l'intelligence artificielle évoque naturellement l'idée d'une intelligence autonome de la

machine, qui se pose en face de l'intelligence humaine, pour la simuler ou la dépasser. Mais si nous observons les usages réels des dispositifs d'IA, force est de constater que, la plupart du temps, ils augmentent, assistent ou accompagnent les opérations de l'intelligence humaine. Déjà, à l'époque des systèmes experts – lors des années 1980 et 1990 –, j'observais que les savoirs critiques de spécialistes au sein d'une organisation, une fois codifiés sous forme de règles animant des bases de connaissances, pouvaient être mis à la portée des membres qui en avaient le plus besoin, répondant précisément aux situations en cours et toujours disponibles. Plutôt que d'intelligences artificielles prétendument autonomes, il s'agissait de médias de diffusion des savoir-faire pratiques, qui avaient pour principal effet d'augmenter l'intelligence collective des communautés utilisatrices.

Dans la phase actuelle du développement de l'IA, le rôle de l'expert est joué par les foules qui produisent les données et le rôle de l'ingénieur cognitif qui codifie le savoir est joué par les réseaux neuronaux. Au lieu de demander à des linguistes comment traduire ou à des auteurs reconnus comment produire un texte, les modèles statistiques interrogent à leur insu les multitudes de rédacteurs anonymisés du Web pour en extraire automatiquement des structures types qu'aucun programmeur humain n'aurait pu tirer au clair. Conditionnés par leur entraînement, les algorithmes peuvent alors reconnaître et reproduire des données correspondant aux formes apprises. Mais parce qu'ils ont abstrait des structures plutôt que de tout enregistrer, les voici capables de conceptualiser correctement des formes (d'image, de texte, de musique, de code, etc.) qu'ils n'ont jamais rencontrées et de produire une infinité d'arrangements symboliques nouveaux. C'est pourquoi l'on parle

d'intelligence artificielle *générative*. L'IA neuronale synthétise et mobilise la mémoire commune. Bien loin d'être autonome, elle prolonge et amplifie l'intelligence collective. Des millions d'utilisateurs contribuent au perfectionnement des modèles en leur posant des questions et en commentant les réponses qu'ils en reçoivent. On peut prendre l'exemple de Midjourney, dont les utilisateurs s'échangent leurs consignes (*prompts*) et améliorent constamment leurs compétences en IA. Les serveurs Discord de Midjourney sont aujourd'hui les plus populaires de la planète, avec plus d'un million d'utilisateurs. Une nouvelle intelligence collective stigmergique émerge de la fusion des médias sociaux, de l'IA et des communautés de créateurs.

L'IA contemporaine fonctionne ainsi comme le conduit d'une boucle de rétroaction entre la mémoire numérique commune et les productions individuelles qui l'exploitent et s'accumulent à leur tour dans les centres de données. Derrière la machine, il faut entrevoir l'intelligence collective qu'elle réifie et mobilise.

### **Le partage du savoir : vers une intelligence collective neurosymbolique**

L'intelligence collective aujourd'hui supportée par l'intelligence artificielle n'est encore que partielle. En effet, l'utilisation des données de l'Internet pour entraîner les modèles mobilise les intelligences collectives stigmergique (la boucle de rétroaction entre les comportements individuels et la mémoire commune) et statistique (l'apprentissage neuronal). Au début des années 2020, la connexion et le renforcement mutuel de ces deux formes d'intelligence

collective par les nouveaux dispositifs d'IA ont provoqué un choc intellectuel – et des émotions fortes – chez ceux qui en ont aperçu la puissance. Mais une intelligence collective délibérative et réflexive manque encore à l'appel. À l'échelle où nous nous situons, cette intelligence collective délibérative doit porter sur l'organisation des données, c'est-à-dire sur la structure conceptuelle de la mémoire, inévitablement couplée aux pratiques des communautés. Comment faire en sorte que les réseaux de concepts qui informent la mémoire numérique puissent faire l'objet d'une conversation ouverte, transparente, attentive aux conséquences de ses choix ? Le Web sémantique et son empilement de standards (XML, RDF, OWL, SPARQL) ont certes établi une interopérabilité de formats, mais non pas l'interopérabilité proprement sémantique – celle des architectures de concepts – dont nous avons besoin. Les géants du Web ont leurs graphes de connaissances, mais ces derniers sont malheureusement privés et secrets. Wikidata propose un exemple de graphe de connaissances ouvert, mais il est encore bien difficile à explorer et à utiliser quotidiennement par le grand public. Il se présente de plus comme *une* ontologie, celle de l'encyclopédie Wikipédia, alors qu'il faudrait mettre en harmonie et en dialogue la multitude des ontologies qui émergent de pratiques aussi diverses que l'on voudra.

C'est pour résoudre ce problème de l'émergence d'une intelligence collective délibérative (ou réflexive) à support numérique que j'ai inventé IEML (Information Economy MetaLanguage): une langue artificielle pourvue d'une structure algébrique régulière, dont la sémantique est calculable, qui permet de tout dire et qui peut traduire n'importe quel réseau de concept (Lévy, 2023). IEML est un langage à source

ouverte, qui se place dans la perspective d'une augmentation des communs de la connaissance, et dont le développement doit faire l'objet d'une gouvernance décentralisée. Aussi hétérogènes ou divers qu'il soit, les IEML projette les ontologies, les graphes de connaissances, les collections d'étiquettes et les modèles de données sur le même système de coordonnées sémantique: un univers virtuellement infini de différences conceptuelles donnant prise aux algorithmes. IEML peut servir de *langage pivot* entre les langues naturelles, entre les humains et les machines, entre les modèles d'IA. Il va sans dire que la plupart de ses bénéficiaires n'auront pas à l'apprendre puisque les interfaces des applications, y compris l'éditeur lui-même, seront en langues naturelles ou sous forme iconique. La face «code» d'IEML n'est destinée qu'aux ordinateurs. On peut dès lors envisager qu'une multitude de bases de connaissances aux architectures conceptuelles singulières puissent échanger des modules ontologiques et des informations grâce à l'interopérabilité sémantique assurée par ce *langage de métadonnées* commun.

Revenons maintenant à ce qui réunit tous les wikipédiens – ses auteurs, ses éditeurs et ses lecteurs –, non pas l'objet «encyclopédie», qui n'est finalement qu'un moyen particulier adapté aux possibilités techniques et culturelles d'une époque, mais la finalité beaucoup plus vaste, et qui retentit jusque dans un avenir encore inimaginable, de mettre la connaissance à la portée de tous. Ce concept se décompose en deux exercices complémentaires: (a) celui de permettre à toutes les connaissances de s'exprimer, de s'accumuler et de communiquer; (b) celui de faciliter l'exploration et l'appropriation des connaissances selon la gamme

étendue des situations pratiques, des parcours d'apprentissage et des styles cognitifs. On voit l'affinité de cet idéal avec celui d'une intelligence collective – diamétralement opposée au « *group think* » – qui vise à maximiser simultanément la liberté créatrice et l'efficacité collaborative.

On pardonnera au philosophe que je suis l'évocation d'une utopie concrète, sans doute techniquement réalisable, mais qui – à court terme – vise d'abord à faire penser. Imaginons donc un dispositif destiné au partage des connaissances et qui tire le maximum des possibilités techniques contemporaines. Au cœur de ce dispositif évolue *un écosystème ouvert de bases de connaissances* catégorisées en IEML, qui émergent d'une multitude de communautés de recherche et de pratique. Entre ce noyau de bases de connaissances interopérables et les utilisateurs humains vivants s'interpose une *interface neuronale* (un écosystème de modèles) « *no code* » qui donne accès au contrôle, à l'alimentation, à l'exploration et à l'analyse des données. Tout se passe de manière intuitive et directe, selon les modalités sensori-motrices sélectionnées. C'est aussi par l'intermédiaire de ce gigaperceptron – un métavers immersif, social et génératif – que les collectifs échangent et discutent les modèles de données et réseaux sémantiques qui organisent leurs mémoires. En bonne gestion des connaissances, le nouveau dispositif de partage des savoirs favorise l'enregistrement des créations, accompagne les parcours d'apprentissage et présente les informations utiles aux acteurs engagés dans leurs pratiques.

Pour ce qui est commun, chaque base de connaissances – personnelle ou collective – affiche son univers de discours, ses données et ses statistiques, aussi transparente aux algorithmes

qu'elle l'est aux regards humains. Mais pour ce qui est privé, notre dispositif de partage des connaissances assure la *souveraineté* pratique et légale des individus et des groupes sur les données qu'ils produisent et qu'ils ne divulguent qu'aux acteurs choisis.

L'augmentation décisive de la dimension délibérative de l'intelligence collective grâce à l'utilisation d'un langage de métadonnées commun a des effets multiplicateurs sur les intelligences collectives statistique et stigmergique déjà à l'œuvre aujourd'hui. Une nouvelle infrastructure neuro-symbolique plonge l'intelligence collective du futur dans l'univers explorable émanant de ses propres activités cognitives. Il faut cependant bien distinguer l'intelligence collective qui anime les personnes et les collectivités humaines vivantes des extensions mécaniques et des représentations médiatiques qui l'augmentent. Ne faisons pas une idole de l'intelligence artificielle.

Citant Ibn Roshd (l'Averroès des Latins), Dante écrit au chapitre I, 3 de sa *Monarchie*: « Le terme extrême proposé à la puissance de l'humanité est la puissance, ou vertu, intellectuelle. Et parce que cette puissance ne peut, d'un seul coup, se réduire toute entière en acte par le moyen d'un seul homme ou d'une communauté particulière, il est nécessaire qu'il règne dans le genre humain une multitude par le moyen de laquelle soit mise en acte cette puissance tout entière. » Que cette multitude devienne transparente à elle-même dans le nouveau médium algorithmique et nous serons passés de la fourmilière à la cité.

# Table des matières

<b>Introduction</b>	<b>7</b>
<i>Jean-Michel Lapointe et Marie D. Martel</i>	

## **AXE I** **Identités**

<b>CHAPITRE 1</b>	
<b>Les protocoles de pluralisation</b>	<b>29</b>
<i>Négocier la cohabitation culturelle dans Wikipédia</i>	
<i>Nathalie Casemajor</i>	

<b>CHAPITRE 2</b>	
<b>Le portail Acadie de Wikipédia</b>	<b>59</b>
<i>Un portrait juste ?</i>	
<i>Gabriel Arsenault et Mathieu Wade</i>	

<b>CHAPITRE 3</b>	
<b>La quantification du fossé des genres dans les ressources biographiques avec Wikidata</b>	<b>73</b>
<i>Marie D. Martel et Simon Villeneuve</i>	

## **AXE II** **Institutions**

<b>CHAPITRE 4</b>	
<b>Le festival folk Mariposa et Wikidata au Canada</b>	<b>103</b>
<i>Un projet de données liées</i>	
<i>Stacy Allison-Cassin</i>	

**CHAPITRE 5**

**Wikimédia dans un musée d'art québécois** **139**

*Exploration d'un modèle d'institution culturelle ouverte*

*Nathalie Thibault*

**CHAPITRE 6**

**Gouvernements ouverts** **159**

*Un wiki pour les lier tous*

*Miguel Tremblay*

**AXE III**

**Littératies**

**CHAPITRE 7**

**Les savoirs publics lors de l'infodémie COVID-19** **181**

*La littératie en santé et l'effet de Wikipédia*

*Denise Smith*

**Postface** **201**

*Pierre Lévy*