# The Wikimedia Movement in Canada

*This page intentionally left blank.*

# The Wikimedia Movement in Canada

## Communities, Institutions, and Open Culture

Edited by Jean-Michel Lapointe and Marie D. Martel

AU PRESS

Published by AU Press, Athabasca University
1 University Drive, Athabasca, AB T9S 3A3

Originally published in French as *Le mouvement Wikimédia au Canada: communautés, institutions et culture libre* (Les Presses de l'Université de Montréal, 2025)

Translation by George Tombs

https://doi.org/10.15215/aupress/9781771994521.01

Cover design by Will Brown
Printed and bound in Canada

CERTIFIED
CANADIAN
PUBLISHER

Library and Archives Canada Cataloguing in Publication

Title: The Wikimedia movement in Canada : communities, institutions, and open culture / edited by Jean-Michel Lapointe and Marie D. Martel.
Other titles: Mouvement Wikimédia au Canada. English
Names: Lapointe, Jean-Michel, editor. | Martel, Marie D., editor
Description: Translation of: Le mouvement Wikimédia au Canada: communautés, institutions et culture libre. | Includes bibliographical references.
Identifiers: Canadiana (print) 20250324954 | Canadiana (ebook) 20250325047 | ISBN 9781771994521 (softcover) | ISBN 9781771994538 (PDF) | ISBN 9781771994545 (EPUB)
Subjects: LCSH: Electronic encyclopedias—Canada. | LCSH: Online databases—Canada. | LCSH: Access to knowledge movement—Canada. | LCSH: Wikipedia. | LCSH: Wikidata.
Classification: LCC Z674.75.W55 M6813 2026 | DDC 030.971—dc23

We acknowledge the financial support of the Government of Canada through the Canada Book Fund (CBF) for our publishing activities and the assistance provided by the Government of Alberta through the Alberta Media Fund.

Canada Alberta Government

# Contents

## Focus III Literacies

# Introduction

There is abundant literature about Wikipedia and open culture, but there has rarely been a work focused on describing and understanding the specific participation of a country's population in Wikimedia platforms. The present collective work aims to examine Wikimedia initiatives in Canada, presenting a series of studies that shed light on how the ideals of sharing, open access, and collaboration—the pillars of open culture—manifest themselves in Canada.

This work raises a fundamental question: How do Canadians participate in Wikimedia platforms, and what are the issues and dynamics of their participation? In the introduction, we aim to set out a few benchmarks and milestones for this question and to point out some of the limits constraining it. From the outset, we would like to acknowledge that there is some peril in producing an all-encompassing portrait of Canada, whatever the theme, given the country's complex regionalisms, linguistic and ethnocultural divisions, and historical presence of its First Peoples. Canada is a country whose identity and history benefit from being seen as plural (Wright, 2020). This country therefore serves as a somewhat troubling mirror of the Wikimedia movement itself, where each point in the network of people connected to the platform can add its own touch to the projects, based on its contribution interests. So much so that the Wikipedia encyclopedia has already been described as a large dynamic territory where different projects and communities are developing autonomously, but the accent should not be put on separation from one another (McDowell & Vetter, 2021, p. 77). The multiple authors' perspectives in this collective work reflect this intersecting complexity of the Canadian and Wikimedian territories.



https://doi.org/10.15215/aupress/9781771994521.01

Furthermore, the policies for collecting data on the people contributing to various Wikimedia projects do not seem to allow for their country of residence to be identified. The culture of the community is largely based on respect for anonymity and does not require any link to civil identity (Konieczny, 2023). As a result, we can hardly characterize the Canadian contributor population by identifying where each editor resides. This lack of specific data on the profile of Canadian editors creates other challenges in fully understanding the participation dynamics that interest us. That said, some anonymized datasets made available by the Wikimedia Foundation's technology team make it possible to geolocate contributions to Wikipedia, which is the movement's most active project. Between September and October 2021, 3,101 active contributors located in Canada participated in the English-language Wikipedia, compared to 391 in the French-language version.[1] In addition to these two main language versions, there is steady, although smaller-scale, participation in the Farsi, Hebrew, Russian, Serbian, Ukrainian, and Mandarin versions. This suggests that Canadians of immigrant background maintain a connection to their culture and language of origin by contributing to the free encyclopedia.

This previously mentioned question can also be considered from the point of view of the birth and growth of the Canadian Wikimedia "chapter." This approach helps reveal the motivations and challenges arising from the very first initiatives while offering a crucial entry point for proposing some answers; the approach also highlights the importance of local initiatives and national chapters such as Wikimedia Canada, which provide valuable insight into regional contributions and specific issues.

Since 2011, the Wikimedia Foundation, the movement's parent organization, has officially recognized the nonprofit organization Wikimedia Canada. Wikipedia itself was founded in 2001, and 2 years later, the foundation was created to support the contribution of volunteers by administering, developing, and maintaining the technological infrastructure of all the projects revolving around Wikipedia. As such, the foundation is the custodian of the Wikimedia brand. What's more, the foundation provides financial

1 Data collated from the months of January and June for the years 2021, 2022, and 2023 geolocation dumps. Note that the constant participation of language versions other than French and English means more than 100 modifications per month made by an indeterminate number of people ranging from 1 to 10 (stats.wikimedia.org).



https://doi.org/10.15215/aupress/9781771994521.01

support to 37 chapters around the world—that is, organizations whose aim is to support and stimulate the development of Wikimedia projects on their territory.

Starting in 2005, a handful of volunteers in Canada from Ontario, Québec, British Columbia, and Alberta took steps to set up a structure that would publicize and encourage contributions to Wikipedia in this country. These volunteers also share the ideals of the free- and open-culture social movement behind Wikipedia, promoting the free circulation of knowledge and information so that it can be legally copied and redistributed. Under the terms of its free licence, all of Wikipedia's content can be used, modified, and shared, as long as the source of this content is cited and its free nature is preserved.

From the very start of their collective efforts, the activists laying the organizational foundations of what would become Wikimedia Canada called for the abolition of section 12 of the Copyright Act, which deals with what is known as Crown copyright. This section of the law, which has been in force in Canada since 1921 and is common to several Commonwealth countries, stipulates that documents produced by the federal government only enter the public domain 50 years after being created, which has the effect of preventing publicly funded content from being reused in Wikipedia and the movement's other projects. For example, official photographs of Members of Parliament cannot be used to illustrate articles in the encyclopedia, as such photographs belong to His Majesty. This law contrasts with practice in the United States, where all documents produced by the US federal government are placed in the public domain from the outset and can be freely reused. In addition to members of the Wikimedia community, the main associations of archivists, librarians, researchers, and teachers in Canada are also calling for the abolition or reform of this section of the law, which would make it possible to preserve and enhance the value of documents produced by federal public institutions (Wakaruk, 2020).

In the summer of 2011, the long-term efforts of the first volunteers to create a Canadian organization that would be recognized by the American foundation came to fruition. When Wikimedia Canada (WMCA) officially became an affiliate of the Wikimedia Foundation, James Heilman, MD, became the organization's first president. Since then, WMCA has become an official affiliate of the Wikimedia Foundation and is run by people who aim to stimulate the development of Canadian content projects by increasing the number of opportunities for meetings and contribution events.



https://doi.org/10.15215/aupress/9781771994521.01

The chapter's first official public activity took place on August 28, 2011, and illustrates this approach. The aim was to document more than 375 places of historical interest so that the encyclopedic articles devoted to them could be illustrated with freely licensed images. Known as Wikipedia Takes Montréal, the project was significant in more ways than one. It aimed, on the one hand, to mark Wikipedia's 10th anniversary and, on the other, to stimulate unprecedented participation in the production of Canadian commons in the country's information landscape. The event attracted more than 200 people thanks to media coverage across Canada in both French and English. This local initiative also had an international dimension: It was prepared in advance along with the municipal administration of the City of Montréal, and this collaboration with a public institution would become a hallmark of WMCA once it was officially affiliated with the Wikimedia Foundation.

WMCA received logistical support from another chapter of the movement, Wikimedia France. This activity suffered a setback that was both unforeseen and highly memorable when the remnants of Hurricane Irene arrived in Montréal on the same day, compromising the quality of most of the thousands of photos taken. Relatively speaking, this was a disappointment, but it was far from demotivating the organizers; it actually had some impact within the movement.

The establishment in 2013 of a partnership agreement with Bibliothèque et Archives nationales du Québec (BAnQ) marked a decisive moment in the recognition of the Canadian chapter of the movement by institutions across the country, particularly those belonging to GLAM (galleries, libraries, archives, and museums). This agreement is a true "public-common partnership," it remains in force a decade later, and it is proving beneficial to both parties and to more besides. As described by David Bollier and Silke Helfrich (2022), a public-commons partnership is a form of collaboration between public entities (governments and public institutions) and the commons (resources managed collectively by a community according to community-established rules). In this case, Wikimedia content created from institutional documentary resources is more precisely designed as a knowledge commons.

On the one hand, WMCA is officially affiliated with the Wikimedia Foundation, and the movement of volunteers it is seeking to develop can count on a free physical place of convergence and offline socialization for the Montréal community through training evenings entitled "Mardi, c'est Wiki" (Wiki Tuesdays) following a tried and tested formula in the French-speaking world. On



https://doi.org/10.15215/aupress/9781771994521.01

the other hand, BAnQ is encouraging archivists and librarians at this institution, which is present throughout Québec, to broaden their documentary expertise by promoting the collections they hold on Wikimedia platforms and by involving their audiences, who are offered training in how to contribute to Wikimedia projects (Boudreau et al., 2016). Canada's chief librarian, Guy Berthiaume, who was head of BAnQ when the partnership was established and then replicated it to some extent when he moved to Library and Archives Canada (LAC) in 2017, had this to say in his opening speech at the Ontario Museum Association Conference that same year:

> So, while museums and other GLAMs have the knowledge and the records, Wikipedia has a reach that none of our institutions could ever match on its own. . . .
>
> We are working with Wikimedia Canada, the non-profit body that aims to increase Canadian content in Wikipedia. The idea was that our expanded presence on Wikipedia would attract new users to our website and our collections. LAC now has over 3,000 images embedded on Wikipedia article pages, and in turn, this results in roughly 30 million hits a month on LAC materials. (Berthiaume, 2017)

WMCA was now officially affiliated with the Wikimedia Foundation, and the partnerships created with memory institutions were acting as a catalyst to "wikify" the country's noncommercial institutional ecosystem. Museums, learned societies, universities, public and research libraries, and cultural and community organizations have all hosted contribution and training activities or incorporated a Wikimedia component into their organizational strategy. This sometimes involves the occasional hiring of a Wikipedian in residence, although often the Wikipedian expertise of the staff is deployed in the shadows, without there being any institutional recognition of the contributory work, which is often the case within the Wikimedia movement (Stintson & Evans, 2018).

Even before the partnership with BAnQ, however, the first interactions with libraries took place as part of an edit-a-thon. The first of these interactions took place on February 18, 2012, at the Université Laval Library in Québec City, with a workshop devoted to Jean Talon, the first intendant of New France. A year later, on April 6, 2013, Mile End Library (now called the Mordecai-Richler Library) hosted the first edit-a-thon in a public library in Québec and, as far as we know, in all of Canada. The event was part of the



https://doi.org/10.15215/aupress/9781771994521.01

Mile End Project, launched in February by Mile End Memories, a local history organization, in collaboration with Wikimedia Canada. In Montréal, public libraries would be incorporating contribution workshops at Wikipedia into their regular programming. Between 2016 and 2018, more than 20 workshops were held, most of them in partnership with the Café des savoirs libres, a collective of librarians and open-culture advocates committed to creating knowledge commons (Martel, 2021). By setting up partnerships based at the local library site, it has proved possible to bring together players from public institutions as well as civil society (historical societies and community groups), creating locally based Wikimedia networks. A number of workshops have also been set up to promote the cinema commons at the Cinémathèque québécoise, in the presence of key figures in the Canadian animation film community.

Many of these institutions have focused their Wikimedia activities on combating systemic bias, narrowing the gender gap, and raising the profile of certain minority communities or marginalized groups in line with the public policies of reconciliation, equity, diversity, and inclusion (REDI) that have now been adopted by these institutions across Canada.

In terms of integrating the movement into higher education, the many roles played by the country's librarians should be underlined in opening up and paving the way for building bridges with an academic community that was initially wary of Wikipedia's potential to disrupt established uses of documentation in the digital environment.

The contribution of universities to the movement is considered so central to WMCA that the people active in the Canadian chapter have taken to building on the acronym GLAM by adding a U at the end. Benoit Rochon, president of WMCA from 2016 to 2019, explains the origin of the expression GLAMU in these terms: "For the record, I spoke at a conference in 2011 at the Mushroom Museum in Amsterdam, where a few university professors from the Netherlands were in the audience. I wrote on a slide 'We will GLAM•U, universities' recalling Queen's song *We Will Rock You*" (Rochon & Phan, email exchange, May 26, 2023). Ha-Loan Phan, a volunteer since 2012 and president of WMCA from 2022 to 2024, explains why she set out to popularize the expression afterward:

> I'd learnt to publish information on the opossum on my own after doing my Masters on the animal, and I felt that GLAM was missing the presence of universities, even though they are the places where



https://doi.org/10.15215/aupress/9781771994521.01

> knowledge is produced and circulated. University librarians have always been privileged allies for our training workshops: as specialists in sources, they gave us the assurance of their expertise, which reassured the professors! What's more, university libraries are perfect places to host this kind of "extracurricular" activity, with the sources close at hand. (Rochon & Phan, email exchange, May 26, 2023)

Another highlight in the history of the Canadian movement was the annual international Wikimania Convention held in downtown Montréal in 2017, which helped attract new members and deepen the roots of the Canadian chapter in Québec's leading city. This served as an opportunity for Wikimedians from all over the world to discover Canada's contribution to the movement but also to appreciate the country's linguistic duality and the importance of its Indigenous cultures. Wikimania 2017 was an opportunity to celebrate the public launch earlier that year of Wikipetcia, the version of the encyclopedia in Atikamekw Nehiromowin, the Atikamekw native language.

The presence of Canadian indigenousness in Wikipedia is also worth highlighting, since it indicates the role and underlying issues of participation in the encyclopedia in working toward the diversification and neutrality of the information available online. The Wikimedia movement makes it possible to include the plurality of Canadian realities within one of the most visited platforms on the web, helping people to participate in shaping knowledge and also disseminating knowledge in a variety of languages. That said, to date, there are only three Indigenous language versions: Inuktitut, Cree, and Atikamekw Nehiromowin, the last of which is very dynamic. The 8,000 members of the Atikamekw Nehirowisiwok nation are spread across three communities in the Mauricie region of central Québec and still speak their ancestral language. The Atikamekw Nehirowisiwok version of Wikipedia was created in 2013 and launched in 2017 and has enabled a culture with an oral tradition to appropriate and adapt to its needs a Western textual form: the encyclopedia. This digital presence offers a means of stimulating the revitalization of the Atiamekw language, in particular through a system of intergenerational transmission between the elders who hold the knowledge that the language carries and the younger people who gravitate to the contemporary technological universe.

This encyclopedic sharing project has also served as an opportunity to redefine what the principle of openness, so dear to the movement, means for the



https://doi.org/10.15215/aupress/9781771994521.01

Atikamekw nation, which has led them to exclude knowledge regarded as sacred from their Wikipedia (Casemajor et al., 2019). Throughout history, dispossession has long structured relations between settlers of European origin and the Indigenous peoples of Canada, and this explains the legitimate skepticism that Indigenous people may have toward the ideal of openness. Science, the media, and Wikipedia are inherently Western spaces of knowledge, and Indigenous people do not recognize them as being conducive to a genuine consideration of their own perspectives on the world. Wikipedia is nonetheless distinctive in that it is increasingly seen as a space for the production of knowledge, which, because of its decentralized mode of operation, can be used by Indigenous people in Canada to rewrite history in a more neutral way (Lugosi et al., 2022). The Canadian chapter of the Wikimedia movement is promoting such an approach and hopes to stimulate the development of new Indigenous language versions over the next decade.

Holding Wikimania 2017 in the Québec metropolis marks a decisive turning point for WMCA, consolidating its presence in Montréal and within French-speaking culture, despite the predominance of English in Canada. This characteristic is reflected in the content of this book, which was originally written mainly in French, and can be explained not only by the vigorous commitment of Québec documentary institutions to Wikimedia projects but also by the steadfast dedication of one volunteer in particular. Working behind the scenes, Benoit Rochon has played a crucial role in the establishment and growth of the Canadian chapter of the movement. He has been contributing to Wikipedia since 2003 and is a founding member of WMCA and of WikiFranca—an organization founded in 2012 to facilitate collaboration between volunteers in the movement from all over the French-speaking world.

He is also the driving force behind the partnerships with BAnQ and LAC, as well as countless edit-a-thons across the country, many of them in the Montréal area, including the organization of the Wikimania Conference in Montréal. Collaborative culture likes to promote collective intelligence as a process that upholds anonymity, but it should be remembered that contributions are unequally distributed between individuals. In fact, contributions are subject to a dynamic where a majority of people participate little, while a tiny minority of people are responsible for the vast majority of contributions (Shirky, 2008). Benoit Rochon belongs to this minority of people who have played a critical role in initiating collaboration and in developing local,



https://doi.org/10.15215/aupress/9781771994521.01

provincial, pan-Canadian, and international projects within the Wikimedia movement.

When Benoit Rochon took over as president of WMCA in 2016, he sought to surround himself with volunteers on the Board of Directors who, like him, had a workforce capable of taking WMCA further by developing new partnerships and tapping new sources of funding. Their efforts have been crowned with success, but the workload rests exclusively on the shoulders of a team of volunteers, which means that a strategy needs to be developed to professionalize the chapter's organization by providing it with salaried staff capable of tackling the issues that require sustained attention, particularly when working with institutions. Compared with the movement's chapters in France and Germany, which have more than 12 and 90 employees, respectively, Wikimedia Canada's workforce is based entirely on the free time that its members and board members manage to find, along with occasional mandates given to contractors. This way of doing things, which has been the case since WMCA was founded, began to change in 2022.

The Canadian chapter underwent a major transformation, setting up its first operational team and hiring its first-ever executive director, Louis Germain, who until then had been executive director of the Association des archivistes du Québec. This may be a sign of growing maturity, or it may illustrate the wish to inject new dynamism in order to stimulate the movement's activities in Canada, but it surely marks a new departure. In November 2023, the team took on the challenge of organizing the Wikiconvention North America at the Toronto Public Library, bringing together enthusiasts from across North America to exchange ideas, strengthen collaborations, and celebrate the movement. Among other things, the convention examined the vital question of the impact of artificial intelligence on Wikimedia projects, culture, and free knowledge but also on democracy, considering the risks of information manipulation, mass surveillance, algorithmic bias, political microtargeting, and the concentration of power.

The other challenges that will continue to guide Canadian open culture in the future, particularly within the Wikimedia movement, include the need to broaden and renew the communities contributing to the projects beyond the circles historically associated with the movement (free-software developers, librarians, archivists, educators, and researchers). The COVID pandemic increased people's isolation and encouraged the emergence of remote socialization practices, making it more difficult to develop strong social interactions.



https://doi.org/10.15215/aupress/9781771994521.01

In this context, the in-person contributions and networking activities that have been important in structuring and fostering local Wikimedia communities in various Canadian cities need to be revived. What's more, the evolving role of the movement's Canadian partner institutions needs to be given particular attention, as there is no guarantee of sustainability because of changes in personnel, particularly in the management teams that define organizational priorities.

Finally, three key issues are likely to have an impact on the vitality of the movement in Canada over the next few years. Firstly, the strong comeback of concerns for open education and open educational resources and practices, which are directly aligned with the principles of the Wikimedia movement, are the subject of marked interest within provincial government ministries of education, as well as renewed advocacy within library associations (CARL, 2023). Secondly, the fact that open science is now public policy at both the federal and the granting agency levels, which has the potential to transform the practices of researchers and, more generally, the scholarly publication ecosystem, bringing it closer to a self-organized governance model similar to that of the Wikimedia movement itself. And finally, the legal framework in place in Canada, which places significant limits on the reuse of works, whether produced within the federal government under Crown copyright or not but also because of the decline in the Canadian public domain since January 1, 2023, which has been extended from 50 years to 70 years after the death of the author.

As a result, given their concerns about the possible exhaustion of the open movement, other internal critics are calling for ways to revitalize it, such as promoting the contribution of other voices but also of other narratives, such as that of the "commons." The Wikimedia movement and the concepts associated with the commons converge in terms of participatory and collaborative engagement, free access to knowledge, and the promotion of collective management of resources.

* * *

This present work is organized around three key themes linking the texts together. We have grouped them into three main sections according to the importance attached to them by their authors: (1) identity issues; (2) collaboration with institutions; and (3) the literacies developed by taking part in or interacting with projects.



https://doi.org/10.15215/aupress/9781771994521.01

We can see that Wikipedia remains the main object of investigation in the analyses, closely followed by Wikidata, which now plays a central role within the Wikimedia ecosystem and beyond; finally, two studies in the present work also look at the use of the Wikimedia Commons image bank. It should also be noted that all the people authoring a study in this book have a contribution practice linked to the projects on which they are developing knowledge.

Starting with the first theme, which looks at identity issues within the platforms of the Wikimedia ecosystem, Nathalie Casemajor analyzes the intercultural cohabitation that takes place within three language versions of Wikipedia: English, French, and Atikamekw Nehiromowin. She conceptualizes the pluralization protocols that these three encyclopedias have developed over time to ensure the existence of Canadian linguistic and cultural particularities in these globalized digital spaces. Gabriel Arsenault and Mathieu Wade explore the portrayal of the Acadian nation in the French version of Wikipedia. Their work is guided by two questions. On the one hand, is Acadia a territory confined to Canada's Maritime provinces, or is it instead a diaspora without borders? On the other hand, are the Acadian people a reality relying purely on the written word and belonging to the past, or, on the contrary, do they constitute a living collective subject that is still full of vitality? This scholarly exploration of Acadian time and space ends with a vibrant call for Acadia, past and present, to bring to life its diversity in the free encyclopedia.

Finally, Marie D. Martel and Simon Villeneuve use massive data from Wikidata to assess the representation of people identifying themselves as women and men and the gender gap in biographical articles in over 80 reference dictionaries and encyclopedias from a wide range of countries, from Canada to China, Germany, and New Zealand. These two scholars show there is a major underrepresentation of people who identify as women in all these works, including in the Canadian publications identified and—albeit to a lesser extent—in the biographical resources of the Wikimedia movement.

The second theme deals with the integration of Wikimedia projects within Canadian institutions, a subject we began to describe earlier in this introduction. These collaborations give rise to mutual enrichment, but they also require institutions to overcome certain misgivings, such as admitting the contribution and expertise of amateurs, understanding the value of making accessible the content they possess or of which they are the custodians, and, more generally, adopting practices of openness and transparency, forcing them to develop new ways of working and making decisions. First, Stacy



https://doi.org/10.15215/aupress/9781771994521.01

Allison-Cassin reflects on the presence of Canadian music online by offering a critical review of an initiative she led, which aimed to make the Mariposa Folk Festival archives accessible to the public via Wikidata. In doing so, she paints a portrait of the appropriation of Wikidata within the Canadian library community and, more broadly, within GLAM in this country. This contribution is followed by two other case studies. Nathalie Thibault explains how and why the Musée national des beaux-arts du Québec (MNBAQ) has embarked on open GLAM–type collaborative initiatives, setting up Wikimedia workgroups to make its collections more accessible thanks to free licences and to allow the reuse, sharing, and modification of its cultural data and content. As far as government institutions are concerned, Miguel Tremblay's article studies an innovative project for uploading meteorological metadata produced by the Government of Canada, a case that illustrates an application of the open government model using Wikidata.

The third theme in the present work concerns literacy. One of the leitmotifs working its way through the studies presented here is highlighting the various skills that are developed or called upon when taking part in Wikimedia projects and the movement's community life. Literacy comes in a variety of forms, which often tend to converge. In many ways, information literacy is also digital literacy, which also calls on skills that can become more specifically legal and hermeneutic but also algorithmic, data related, organizational, relational, and technological. As individuals, communities, and institutions advance within the movement, they gradually incorporate most of these forms of literacy, which they need in order to successfully carry out the initiatives they are involved in. This multifaceted learning process attests to the undeniably educational value of the Wikimedia movement's projects. That said, while the majority of the studies in the present work focus on amateur or expert production practices, Denise Smith's chapter stands out for its anchoring in the field of knowledge dealing with health literacy, from the point of view of both the knowledge-producing community and that of the general public seeking to make informed decisions about their health. Smith places her comments in the context of the misinformation *infodemic* that raged during the COVID-19 pandemic, and she shows the extent to which Wikipedia has proved to be a major source of information for Canadian internet users.

For the afterword, we have asked the philosopher Pierre Lévy to extend and broaden the thinking about collective intelligence processes that are examined throughout this book. Lévy starts from a consideration of knowledge



https://doi.org/10.15215/aupress/9781771994521.01

production and sharing practices that take place within the Wikimedia movement, then places them in a broader context rooted in both human and animal history but also in the most recent technological developments in generative artificial intelligence.

This introduction offers an outline of the content in the present work, but it also aims to capture the main thrust of the Wikimedia movement in Canada with the help of a brief chronology and a map of the institutional poles linked to it.

## References


Berthiaume, G. (2017). *Faire plus avec plus: La valeur réelle des institutions de mémoire* [Discours]. Conférence de l'Association des musées de l'Ontario. https://www.canada.ca/fr/bibliotheque-archives/nouvelles/2017/12/faire_plus_avec_pluslavaleurreelledesorganisationsdelamemoire.html

Bollier, D., & Helfrich, S. (2022). *Le pouvoir subversif des communs*. ECLM.

Boudreau, D., Daveau, F., & Giuliano, F. (2016). Diffuser, partager et s'approprier le patrimoine documentaire québécois. Le projet collaboratif de BAnQ sur Wikimédia: une première au Canada. *Archives*, *46*(1), 61–81. https://doi.org/10.7202/1035723ar

Canadian Association of Research Libraries. (2023). *National advocacy framework for open educational resources in Canada*. https://www.carl-abrc.ca/wp-content/uploads/2023/05/A-National-Advocacy-Framework-for-Open-Educational-Resources-in-Canada.pdf

Casemajor, N., Gentelet, K., & Coocoo, C. (2019). Openness, inclusion and self-affirmation: Indigenous knowledge in open knowledge projects. *Journal of Peer Production*, *13*, 1–17.

Konieczny, P. (2023). European Wikipedia platforms, sharing economy and national differences in participation: A case study. *Innovation: The European Journal of Social Science Research*, *38*(3), 1198–227. https://doi.org/10.1080/13511610.2023.2195584

Lugosi, N. V. T., Patrie, N., & Cromwell, K. (2022). Theorizing and implementing meaningful Indigenization: Wikipedia as an opportunity for course-based digital advocacy. *Critical Studies in Education*, *64*(3), 201–17. https://doi.org/10.1080/17508487.2022.2074489

Magord, A. (2010). *Le fait acadien en France: Histoire et temps présent*. Geste.

Martel, M. (2021). Apprendre "en Commun": L'expérience des ateliers de contribution à Wikipédia dans les bibliothèques publiques de Montréal. *Revue électronique suisse de science de l'information*, *21*(1). https://oap.unige.ch/journals/ressi/article/view/2217




https://doi.org/10.15215/aupress/9781771994521.01

McDowell, Z. J., & Vetter, M. A. (2021). *Wikipedia and the representation of reality*. Routledge.

Shirky, C. (2008). *Here comes everybody: The power of organizing without organizations*. Penguin.

Stintson, A., & Evans, J. (2018). Bringing Wiki(p/m)edians into the conversation at libraries. In M. Proffitt (Ed.), *Leveraging Wikipedia: Connecting communities of knowledge* (pp. 31–54). American Library Association.

Wakaruk, A. (2020). Reforming Crown copyright in Canada. *DttP: Documents to the People*, *48*, 12. https://heinonline.org/HOL/Page?handle=hein.journals/dttp48&id=66&div=&collection=

Wright, D. (2020). *Canada: A very short introduction*. Oxford University Press.



https://doi.org/10.15215/aupress/9781771994521.01

Focus I

# Identities

*This page intentionally left blank.*

# 1 Protocols of Pluralization

## Negotiating Cultural Cohabitation in Wikipedia

Nathalie Casemajor

Wikipedia is a melting pot where Canadians share space with other internet users from around the world. The encyclopedia is organized by language rather than by country; it is a resource for internet users who share the same language but not the same customs or references. For example, French-speaking Canadians play hockey with a *rondelle* (puck), while in France, people play with a *palet*. English-speaking Canadians, meanwhile, are issued a *driver's licence*, whereas the British have a *driving licence*. These cultural differences are what make the encyclopedia so rich, but they also lead to numerous editing conflicts. How do Canada's different linguistic and cultural groups negotiate their presence in Wikipedia? The aim of this study is to analyze the cohabitation, within the same encyclopedic space, of members of the Wikipedia community from different cultural backgrounds. I propose to compare the cases of the three main Wikipedias that exist for Canada—that is, the encyclopedias in English, French, and Atikamekw Nehiromowin, the language of the Atikamekw First Nation.

By combining two kinds of analyses—cultural pluralization and governance of the Wikimedia platform—I identify two types of editing protocols that influence the pluralization of knowledge in these spaces: the establishment of provisions common to all groups (*commonality protocols*) and the introduction of clauses that allow for the particular expression of different cultures (*locality protocols*). This work draws on a series of ethnographic studies, participant observations, and action research conducted since 2011 on the



https://doi.org/10.15215/aupress/9781771994521.01

subject of participation in Wikipedia (Casemajor & Couture, 2020; Casemajor et al., 2019). The observations presented in this chapter were mainly made in 2021. They were enriched by an updated analysis in 2024, which revealed that the orders of magnitude in the contribution and readership statistics remain broadly unchanged.

## Political Culture and the Politics of Culture

### Cultural Pluralization

The notion of pluralization refers to the processes of coexistence, inclusion, and exclusion between different social groups within the same public space. The groups that coexist in Wikipedia are positioned at different levels of power and legitimacy in the field of knowledge production. These differences give rise to political negotiation involving relations of power and minority status, which are expressed in the form of concerted action and encyclopedic publishing conflicts (Auray et al., 2009). From this point of view, the analysis of pluralization emphasizes the evolving nature of a process that involves constant adjustments between different heterogeneous stakeholders.

Many academic works have documented the gaps between social groups in Wikipedia. These gaps are addressed from the point of view of gender differences (Berson et al., 2021; Hill & Shaw, 2013), literacy (Ford & Geiger, 2012), or the level of experience in the encyclopedia (Hargittai & Walejko, 2008). Cultural diversity issues in Wikipedia have also been studied from the point of view of the origins of contributors (Massa & Scrinzi, 2012) and obstacles to the inclusion of minority cultures and non-Western voices (Reagle & Koerner, 2020). However, these issues are rarely examined from the point of view of cultural pluralization within a single Wikipedia. I aim to contribute here to this latter angle of analysis by focusing on the expression of viewpoints and uses of language that are culturally situated in national or local territories.

I will use the term *cultural pluralization protocols* in referring to a set of formalized procedures and consensuses concerning the cohabitation of different cultures. These protocols are embodied in the encyclopedia's uses, decision-making, and technical organizational systems. Applying the term *protocol* to the field of digital platforms suggests that the technical and sociopolitical dimensions of the activity are intertwined in a single computerized device. To what extent are Wikipedia's governance protocols agents of cultural



https://doi.org/10.15215/aupress/9781771994521.01

pluralization? I propose to distinguish here two parallel and complementary types of protocols that take part in this process: *protocols of commonality* and *protocols of locality*. Both play a significant role in the governance of Wikipedia's cultural pluralization.

## Wikipedia Governance

Wikipedia's editing activity is governed by a series of pillars, policies, and guidelines that frame the activities of its contributors. These governance frameworks are organized according to a hierarchy in which the directives that are most fixed and broadly shared across most Wikipedias take precedence over those that are more specific to each encyclopedia and easier to modify. These frameworks shape a form of governance that has been described as procedural (Cardon & Levrel, 2009), in the sense that arguments aimed at compliance with established conventions tend to take precedence over those that take account of the wider sociocultural context of an issue.

At the top of the hierarchy of Wikipedia conventions are the *pillars* (*principes fondateurs* in French), defined by the English edition of Wikipedia as "the fundamental principles of Wikipedia." These five pillars are (1) encyclopedism (comprehensive knowledge), (2) neutrality of point of view, (3) free licences, (4) community spirit (respect and civility), and (5) flexibility of rules. In contrast, the other rules described later are determined by each project community, which retains the right to establish and adapt them to the specific needs of its language edition of the encyclopedia. Secondly, there are the *policies* (*règles*)—that is, the standards widely accepted by Wikipedians. These standards are formally adopted by a collective decision of the members. Thirdly, at the bottom of the hierarchy are the *guidelines* (*recommandations*), which crystallize established practices; adopted by consensus (discussion and polling), they may be followed less strictly than the upper-level rules.

The English-language edition of Wikipedia is much more formalized than its French-language counterpart and has almost three times as many policies and guidelines. This difference is mainly attributable to the founding status of the English-language encyclopedia, the first one to be created in 2001, and still by far the most developed. The policies and guidelines of the English-language edition of Wikipedia often serve as a model for other Wikipedias, which have adopted and adapted many, but not all, of them, though not in their entirety. The differences between these diverse language editions also


https://doi.org/10.15215/aupress/9781771994521.01

stem from the political philosophies, governance habits, and types of editors that are specific to each one.

In 2021, the English-language edition of Wikipedia had 66,700 active contributors (table 1.1). The main contributing countries are the United States (35.4%), the United Kingdom (11.3%), India (9.7%), Canada (4.6%), and Australia (4.2%). This distribution is broadly the same for contributors as for readers, with slightly lower percentages: 30.2% of readers are from the United States and 3.6% from Canada. In the English-language edition of this encyclopedia, the United States, as the country where the Wikipedia project was born and where the Wikimedia Foundation has its head office, has a predominant position. Canada is in a minority position among a number of Commonwealth countries. Nevertheless, both the United States and Canada tend to be overrepresented in relation to their demographic weight in the English-speaking world.

In 2021, the French-language edition of Wikipedia had around 7,800 active contributors, eight times fewer than its English-language counterpart. The main contributing countries are France (71.4%), Belgium (5.3%), Canada (5%), Switzerland (3.7%), and Algeria (1.7%). Here again, the distribution of readers broadly mirrors that of contributors, albeit on a smaller scale (54.2% for France, 3.5% for Canada). Compared with the English-language encyclopedia, the gap between the leading contributor (France) and other contributing countries is considerably greater in the French-language edition. France's position is much more dominant. As far as Canada's position is concerned, it is just as much in the minority overall in the French and English editions, but Canadians are rather underrepresented given their demographic weight in the French-speaking world.

The case of the Wikipedia in Atikamekw Nehiromowin is quite different from the other previously mentioned language editions of Wikipedia. It was created more recently (in 2013), its content is still very limited (1,580 articles in 2021), and its editors are rare (an average of four active editors over the last 12 months in 2021). Even so, it is following a continuous development curve—quite the achievement given the small pool of speakers of the language (fewer than 7,000). The encyclopedia contains very few formalized policies and guidelines. In 2021, for example, the page describing the pillars and general policies of the Wikipetcia Atikamekw Nehiromowin was mainly limited to setting out the five founding pillars ("Witcihikoiin: Wikipetcia Atikamekw Nehiromowin," 2017). This less formal approach is mainly because



https://doi.org/10.15215/aupress/9781771994521.01

**Table 1.1.**
*Key figures for the three main language versions of Wikipedia of interest to Canada (September–October 2021)*

| English | French | Atikamekw Nehiromowin |
|---|---|---|
| URL: en.wikipedia | URL: fr.wikipedia | URL: atj.wikipedia |
| Creation date: 2001 | Creation date: 2001 | Creation date: 2013 |
| Articles: 6.4 million | Articles: 2.4 million | Articles: 1,580 |
| Active contributors: 66,700 | Active contributors: 7,804 | Active contributors: 2 (4 avg. over last 12 months) |
| **Top countries of origin of active contributors** | | |
| 1. United States (35.4%) | 1. France (71.4%) | 1. Canada (100%) |
| 2. United Kingdom (11.3%) | 2. Belgium (5.3%) | — |
| 3. India (9.7%) | 3. Canada (5%) | — |
| 4. Canada (4.6%) | 4. Switzerland (3.7%) | — |
| 5. Australia (4.2%) | 5. Algeria (1.5%) | — |
| **Top countries of origin for readers (monthly page views by country)** | | |
| 1. United States (30.2%) | 1. France (54.2%) | 1. France (4.5%) |
| 2. United Kingdom (8.6%) | 2. Belgium (3.5%) | 2. Canada (4.5%) |
| 3. India (8%) | 3. Canada (3.5%) | 3. United States (3.4%) |
| 4. Canada (3.6%) | 4. United States (2.1%) | 4. Sweden (2.5%) |
| 5. Australia (2.9%) | 5. Switzerland (2%) | 5. Russia (2.5%) |
| **Total page views: 9 billion** | **Total page views: 841 million** | **Total page views: 89,000** |

*Note:* Data from stats.wikimedia.org. It is important to note the limitations of these statistics, which should be read with caution. For example, for active contributors, the number provided by WikiTech is a deliberately imprecise estimate when the number is less than 10 (displayed as 0–10). The number of pageviews also needs to be qualified: Notwithstanding the filter designed to exclude traffic generated by robots (notably for indexing pages), the calculations could still be influenced by robot activity, particularly in the case of "small" encyclopedias such as those for Indigenous languages. The data in this table should therefore be read as orders of magnitude rather than precise quantities (percentages are rounded off).


https://doi.org/10.15215/aupress/9781771994521.01

of the fact that the encyclopedia is still emerging. But it is also due to the different relationship that Indigenous peoples have with political organization and knowledge: Customs and knowledge are traditionally transmitted orally, whereas the formalization of written rules is more a feature of Western modernity (Casemajor et al., 2019).

In the next sections of this study, we will identify the elements of each language edition of Wikipedia's list of pillars, policies, and guidelines operating as protocols of community and locality. What do the various language editions of the encyclopedia have in common, and what are the differences? And how can these differences be explained? While it is possible to compare Wikipedia's governance protocols in English and French in detail, the distinctive character of the Atikamekw Nehiromowin edition makes such a comparison more delicate and potentially misleading. Instead, we propose to undertake a separate analysis while putting our findings into perspective with other encyclopedias. More broadly, these questions raise the issue of how knowledge is shared through Wikipedia—an arrangement shaped by political (and geopolitical), epistemological, and technological dynamics. For a comparison of the cultural pluralization protocols in French and English Wikipedia, see figure 1.1.

## Pluralization Protocols

### Commonality Protocols

By *commonality protocols*, I mean the provisions that apply commonly to all editing situations within a given Wikipedia. These provisions are applied to all content, whatever the subjects, contexts, contributors, or readers concerned, and do not admit any exceptions. In other words, these protocols apply regardless of the cultural worlds in which the knowledge, its production, and its interpretation are located. They aim to produce *communality* in the sense of "what is common" (as a counterpoint to *locality*, which is discussed later).

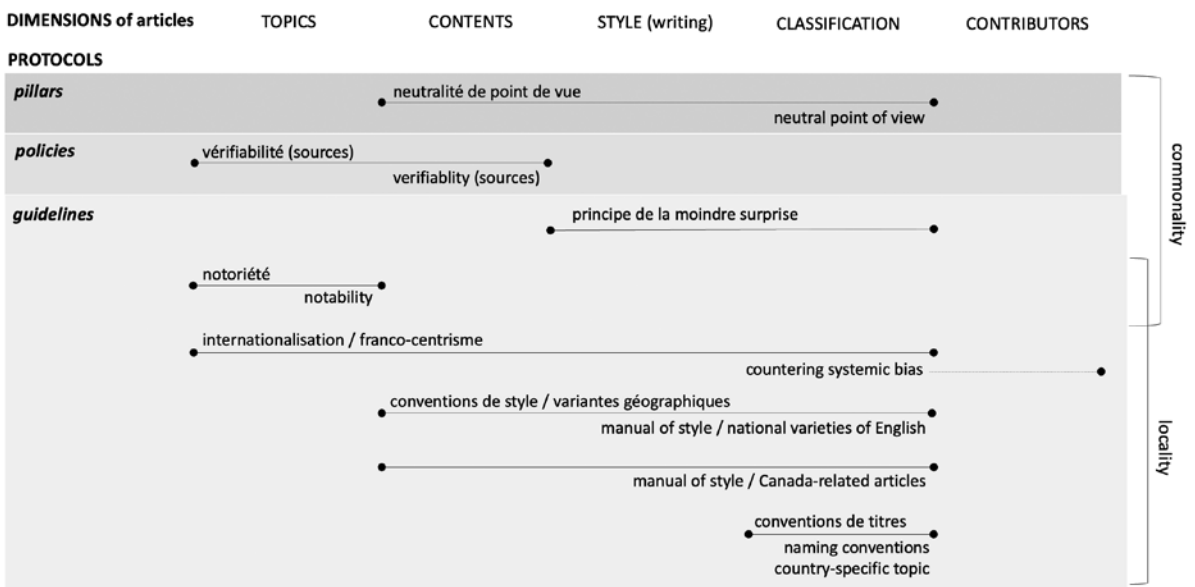
#### Neutral Point of View

The main commonality protocol, which is shared by most language versions, is one of the encyclopedia's founding pillars. Indeed, the neutral point of view is one of Wikipedia's oldest governance principles. The idea of neutrality



https://doi.org/10.15215/aupress/9781771994521.01

**Figure 1.1.**
*Cultural pluralization protocols in French and English Wikipedia*




*Note:* Comparison between the protocols of cultural pluralization in the French-language edition (names at the top of the line, where the protocol exists) and the English-language edition (likewise, at the bottom of the line), and their influence on the main dimensions of article editing.


https://doi.org/10.15215/aupress/9781771994521.01

is described as the fact of "representing fairly, proportionately, and as far as possible, without editorial bias, all the significant views" ("Wikipedia: Neutral Point of View," 2026). Neutral point of view operates as a principle of exclusion, since its function is to rule out partisan contributions, personal opinions, seriously disputed assertions, and ultraminority points of view. At the same time, however, it is a principle of inclusion, since the objective of neutrality is served by presenting a multiplicity of viewpoints (in proportion to their relative weight in the reliable sources published on the subject). In this way, the principle of neutral point of view acts as a pluralization protocol: It aims to create a common space that invariably excludes nonencyclopedic points of view while including a multiplicity of points of view that are deemed relevant.

At the heart of the encyclopedia project, this principle has nevertheless been subjected to much criticism (Tkacz, 2014). The internal policies of the English-language Wikipedia present several objections frequently raised against this principle of neutrality. One of these relates to the encyclopedia's "Anglo-American focus" ("Wikipedia: Neutral Point of View," 2026). This dominant cultural perspective stems from the predominance of American and (to a lesser extent) British contributors and is cited as a provision that runs counter to the principle of neutrality, and that needs to be corrected. The page of the French-language edition describing this principle is descriptive and general and makes no mention of these limitations. On the other hand, the French-language Wikipedia contains a recommendation called *principle of least surprise* (*principe de moindre surprise*), which applies to a similar cultural issue although in a completely different way.

**Principle of Least Surprise**

This guideline is listed among the style conventions and mainly concerns the editorial style of articles and, in particular, the choice of headlines. It "aims to ensure that the information presented is easily understood by the reader," based on arguments of coherence, clarity, and comprehension by the broadest possible readership ("Wikipédia: Principe de moindre surprise," 2025). This principle of least surprise was initially mentioned in a Wikimedia Foundation resolution on controversial and offensive subjects and was then incorporated by Wikipedia into a manual to help edit articles ("Wikipédia: Principe de moindre surprise," 2025). It was adopted by the French-language edition of Wikipedia and formalized in 2011 as an autonomous guideline following acrimonious discussions between people in France and Québec about the



https://doi.org/10.15215/aupress/9781771994521.01

use of Québécois expressions, which are not commonly used in France. For example, in 2008, the (unsuccessful) attempt to rename the French article "Palet (Sport)" to "Rondelle"—a term used in Québec for ice hockey—led to a memorable editing war in the history of relations between editors from Québec and France within Wikipedia.

In the English-language edition of Wikipedia, the principle of least surprise is not intended to arbitrate conflicts linked to cultural differences in the relationship to knowledge, such as the different ways of article editing depending on the usage of the language in each country. In the French-language encyclopedia, on the other hand, this principle of least surprise has been formalized into a guideline that serves in practice to regulate the use of variations in the French language. More specifically, it leads to the use of "French as spoken in France" as the default common standard, to the detriment of other national or regional variants. Given that the majority of (French-language) Wikipedia readers are located in France, "the most common terminology [is] generally that used in France for statistical reasons" in order to cause the least surprise to the French readership. This principle of least surprise applies "to all readers, whatever their country and culture." As a result, it establishes a protocol of linguistic commonality based on the distinctive character of France ("Wikipédia: Conventions de style," 2025).

This principle of least surprise is widely applied for the France-based readership, but it is also regularly contested on the grounds that it imposes "Francocentrism" (a mindset focused on France and the French people), which critics view as a breach of the neutrality principle, insofar as it imposes a specifically French cultural lens on others. On the discussion pages where contributors debate this issue, one finds various arguments that nuance the "Francocentrism" of the least surprising principle, putting into practice exceptions to the rule. But these reflections and exceptions are not clearly indicated on the page setting out the guideline. They therefore remain relatively fragile when publishing conflicts arise, where the France-focused argument tends to prevail.

In the English-language version of Wikipedia, the guideline of less surprise is tempered by another convention called "not what first comes to [your] mind" ("Wikipedia: Disambiguation," 2026). According to this convention, the terms that seem most common to us are inevitably interpreted according to our geographical location, and these partial readings can lead to systemic biases that are detrimental to the encyclopedia's project.


https://doi.org/10.15215/aupress/9781771994521.01

More fundamentally, a series of guidelines from the English-language edition of Wikipedia states that no country or group has priority in terms of language: "The English Wikipedia does not prefer any national variety of English to any other." The United States constitutes a statistical majority group but does not, in theory, have priority over other groups: It is specified that "American English spelling must not be transformed into British English spelling, and vice versa" ("Wikipedia: Manual of Style," 2026).

A substantial difference can thus be observed between the English and French editions of Wikipedia with respect to their communal protocols. In the English-language edition, the Anglo-American orientation of the encyclopedia is immediately identified as a systemic bias to be corrected insofar as the prevalence of one cultural point of view over the others is seen as a breach of the cardinal principle of neutrality. In the French-language edition, the precedence of the French point of view is seen as a guarantee of neutrality in the name of a principle of least surprise for the majority of readers based in France.

### Locality Protocols

Unlike commonality protocols, locality protocols are governance clauses that aim to allow the expression of cultural difference. These protocols contribute to cultural pluralization insofar as they involve a political process of recognizing the value of distinct cultures (Honneth, 2004). In the context of Wikipedia, three dynamics of recognition are at play.

The first concerns the distinctive character of cultures—that is, the existence of particular traits that are unique to each one. In terms of the encyclopedia, cultural groups are distinguished from each other by their use of language, their worldview, their pantheon of famous people, and their specific historical experience.

A second dynamic concerns the recognition of the encyclopedic legitimacy of cultural characteristics. The protocols put in place in Wikipedia for this purpose concern the following elements: the choice of article subjects linked to a particular culture (notability guideline; verifiability guideline); the expression of points of view culturally situated within articles, titles, and indexing categories (title conventions); the use of language variants in the writing of articles (style conventions). Most of these protocols are at the bottom of the hierarchy of Wikipedia standards: They are exceptions to upper-level rules and



https://doi.org/10.15215/aupress/9781771994521.01

guidelines and, as such, are more flexible and defined within each encyclopedia. Each language edition of Wikipedia therefore has its own locality protocols, adapted to the cultural differences observed in its language area.

Thirdly, cultural pluralization in Wikipedia requires the recognition that there are certain obstacles to the expression or inclusion of cultural diversity in the encyclopedia. On the one hand, these obstacles are linked to internal factors—in particular, the demography of contributors, which is unbalanced in terms of geography, social class, and cultural origin. On the other hand, the obstacles to pluralization are caused by external factors. These include the unequal distribution of material and symbolic resources between countries or social groups with respect to the production of knowledge (Casemajor & Couture, 2020; Fraser, 2004). But these obstacles also stem from cognitive, institutional, or political biases that favour certain points of view to the detriment of others. All of these factors can hamper a social group's ability to have its linguistic and cultural uniqueness recognized in the encyclopedia or to make its point of view visible and accessible. These obstacles lead to underrepresentation of certain population groups (and overrepresentation of others), a situation that is considered detrimental to the objective of neutrality.

### "Systemic Bias" vs. "Internationalization"

The English-language and French-language editions of Wikipedia show a different understanding of the obstacles to cultural pluralization. In the English edition, one of the first discussions on this subject escalated into a particularly intense and defining edit war, which occurred only a few months after the encyclopedia was created ("Systemic Bias of Wikipedia," 2023). A contributor named User 24 (who has since been banned for an aggressive attitude) challenged the deletion of an article entitled "Viral Licence," which was intended to reflect an international perspective on copyright. The article had been removed on the grounds that the article entitled "Copyleft" already covered the subject, albeit from a US perspective, but that this cultural perspective was the majority by default given the US-based readership. What followed was an epic and highly learned epistemological confrontation on a neutral point of view and ethnocentrism.

Subsequently, discussion was formalized in two directions. On the one hand, a series of WikiProjects have brought together users in interest groups around identity, cultural, or national subjects. For example, the Ethnic


https://doi.org/10.15215/aupress/9781771994521.01

Groups (2004), Canada (2004), and Indigenous Peoples of North America (2006) projects aim to develop the quality of articles on these particular subjects. On the other hand, WikiProjects were specifically created to work on the cohabitation of different cultures in Wikipedia, such as the Countering Systemic Bias Project in 2004 ("Wikipedia: WikiProject Countering Systemic Bias," 2023). The latter played a structuring role, since it gave rise to an essay under the same name ("Systemic Bias of Wikipedia," 2023). The discussion has become relatively stable since then, although it is still the subject of more or less heated debate. This essay on systemic bias identifies obstacles to the neutrality of Wikipedia's points of view: The points of view represented in the encyclopedia are considered insufficiently diversified across social class, gender, and cultural or geographical origin.

In the French-language edition of Wikipedia, there are also local interest groups, such as Projet: Québec (2005). But the general subject of cultural and linguistic diversity is dealt with in a relatively different way. For example, the English page "Countering Systemic Bias" is translated into French by the term *internationalisation*, which originally stems from the *globalizing* injunction of the English Wikipedia: the idea that Wikipedia is a global project and that it should be able to be "read and understood by any French-speaking reader," as articulated on the internationalization page ("Wikipédia: Internationalisation," 2023). The French version of this page is not particularly developed, but it nevertheless constitutes a guideline and is more formalized than the English page on the same subject. However, the scope of the French page is subject to confrontation between a minimalist interpretation seeking to restrict it to a convention of style (form) and a maximalist interpretation seeking to make it a convention of encyclopedic content (substance).

When the "Internationalisation" page was created in 2005, it was called "Guide de défrancocentrage" (or Guide to removing Francocentrism from Wikipedia articles): in other words, it sought ways to mitigate France's dominant cultural viewpoint (known as *Francocentrism*). The French term *Francocentrism* was subsequently removed from the body of the text.[1] On the discussion page, one user even defends this position: "The fr [of the web address fr.wikipedia] refers mainly to France for me, and not to the French-speaking world. . . . Personally, as a French person, I couldn't care less about the 'situation in Québec'" ("Discussion Wikipédia: Internationalisation," 2025).

1 It only appears in the hypertext links at the bottom of the page.



https://doi.org/10.15215/aupress/9781771994521.01

For this contributor, it is quite simply the transnational status of the French-language edition of Wikipedia that is being challenged. As it stands, the guideline in favour of internationalization does not explicitly identify the existence of any particular bias linked to the dominant status of France. Instead, it states in general terms the need to "include different cultural perspectives in an article in order to take cultural diversity into account." In a banner linked to the page, there is nevertheless a significant note to the effect that "the French-language edition of Wikipedia is not the Wikipedia of France, Québec or Gabon" ("Catégorie: Article à internationaliser," 2025). In other words, it is not the exclusive project of any particular country. The guideline in favour of internationalization is still widely discussed: Its discussion page is twice as long as the content of the guideline itself, which shows that the consensus has not yet stabilized.

Once again, an important difference is worth noting between the French and English editions of Wikipedia. In the French-language edition, some users consider that acknowledging the existence of biases linked to France's dominant position is nonneutral, and they favour restricting the guideline on internationalization to formal characteristics, without any political recognition of the underlying issues. In the English-language edition, the vocabulary of systemic bias points explicitly toward the recognition of imbalances linked to internal and external factors, including geopolitical power relationships that undermine the objective of neutrality. The article on "Systemic Bias" goes further, proposing that underrepresented subjects and perspectives be rebalanced. However, the extent of the corrections to be made remains debated: "Should Wikipedia reflect the world as it is, or as Wikipedians hope it might be?" The current consensus is that Wikipedia "does not right great wrongs," because the encyclopedia is designed to synthesize the current state of information, not to produce original thought: "Wikipedia does not guide, we follow" ("Wikipedia: Tendentious Editing," 2026). As a result, the protocols put in place to encourage the expression of local cultural characteristics in Wikipedia essentially concern a rebalancing within the encyclopedia and not a political action aimed at influencing the state of the world.

### Diversifying the Subjects of Articles (Eligibility)

In addition to recognizing the obstacles to cultural pluralization, Wikipedia contains a series of protocols that encourage the expression of particular cultural characteristics. The first of these concerns the creation of articles in


https://doi.org/10.15215/aupress/9781771994521.01

the encyclopedia. These protocols aim to diversify the subjects of articles by mitigating some of the effects of the notability guideline, which is considered to be detrimental to cultural diversity. This guideline was first introduced in 2006 in the English-language edition of Wikipedia to exclude irrelevant subjects. It is closely linked to the verifiability policy and stipulates that only topics "that have received sufficient attention" should be included, taking into account "evidence from reliable and independent sources" ("Wikipedia: Notability," 2026).

While the guideline is broadly accepted in principle, its implementation has been the subject of ongoing debate. The controversies pit an "inclusionist" party, which advocates a broad retention of poorly supported articles, against a "deletionist" party,[2] which advocates strict selection criteria for articles, resulting in the deletion of several articles dealing with minority cultures. In fact, the eligibility of articles dealing with these cultures is more difficult to establish because of the lesser availability of reference sources in English or the lack of sources altogether. In French, the inclusion of a diversity of international perspectives is perceived by some contributors as an even greater challenge: According to one contributor, "In many fields involving distant cultural horizons, there are few French-language sources compared with English-language resources" ("Discussion Wikipédia: Internationalisation," 2025). The effects of the unequal availability and legitimacy of reference sources have been widely documented by academic studies that conclude that certain marginalized subjects and voices are excluded (Lemieux et al., 2023; McDowell & Vetter, 2021).

Subjects dealing with Canada tend to be overrepresented in the English and French Wikipedias in relation to the country's overall demographic weight. However, their proportion is still very much in the minority compared with that of the United States and France, which respectively dominate the number of articles per country in the two encyclopedias. But there are still some debates about the coverage of subjects relating to internal cultural diversity in Canada. As a multicultural and multilingual country,[3] Canada is itself characterized

2 See the pages "Inclusionnisme" and "Suppressionnisme," policy principles as defined on the Meta site of the Wikimedia platform in French.

3 The Canadian government adopted the Canadian Multiculturalism Act in 1971. According to the Official Languages Act of 1969, Canada is bilingual (English and French), but over 70 Indigenous languages are spoken in the country. Although these



https://doi.org/10.15215/aupress/9781771994521.01

by great cultural diversity within its borders. For example, there have been requests that several articles on the history of Chinese immigration to different Canadian provinces or municipalities be deleted or merged, on the grounds that the subject was already covered by other articles and that the sources cited were insufficient or biased. For example, the article "Chinese Canadians in Greater Vancouver" was retained after a lengthy debate, based on the existence of four scholarly works and a master's degree on the subject (2024).

In the French edition of Wikipedia, the guideline on notoriety sets an additional requirement—namely, that reliable sources must above all be sources "of national or international scope" so as to exclude "subjects whose notoriety would be purely local or restricted to a specific group of people" ("Wikipédia: Notoriété," 2025). This clause makes it questionable to quote Québec sources, because they do not have a readership across the country, even though Québec itself is defined as a nation, with "national" media. The difficulty this requirement presents is even greater for subnational cultures, as, for instance, in the case of the article "Franco-Ontarian literature" ("Littérature franco-ontarienne") (2025; Casemajor & Couture, 2020). A French user proposed the deletion of this article on the grounds that Franco-Ontarian literature was not distinctive enough to deserve its own article separately from the article "French-Canadian literature" ("Littérature canadienne-française"). Ultimately, the article was retained, not for political or epistemological reasons relating to the recognition of minority cultures but because quality sources were cited to justify its continued inclusion in the encyclopedia. More generally, the many debates surrounding the deletion of articles dealing with minority cultures show that there is often a lack of legitimate sources concerning them, given their marginality and poor representation in Canadian cultural, academic, and political institutions.

Some protocols have been established to overcome these difficulties. They take the form of notability conventions applicable to a single country or cultural group. In the English edition of Wikipedia, this is the case of the style manual guideline for Canadian subjects (the "Wikipedia: Manual of Style/Canada-Related Articles" page), which contains guidelines on notability specific to that country. Similarly, in the French edition of Wikipedia, the Swiss WikiProject has drawn up a guideline to the effect that "because of

latter languages have no official status, they make up the linguistic mosaic of the country, which can thus be described as multilingual.



https://doi.org/10.15215/aupress/9781771994521.01

the language regions, no media organizations cover the whole country. The Swiss project therefore considers the language regions (German-speaking Switzerland, French-speaking Switzerland, Italian-speaking Switzerland and Rhaeto-Romanic Switzerland) to be equivalent to national in scope" ("Projet: Suisse/Admissibilité," 2025). As a result, locality protocols applicable to the notoriety of encyclopedia subjects exist to encourage the inclusion of articles whose admissibility could be debated. In so doing, these locality protocols contribute to cultural pluralization by recognizing the encyclopedic legitimacy of subjects linked to cultures located in a particular territory.

As for the Wikipedia in Atikamekw Nehiromowin, its conventions of notability and verifiability of sources are not very formalized. They do, however, include some original provisions, adapted to the intrinsic character of Indigenous cultures and their relationship to knowledge (Casemajor et al., 2019). For example, during the discussions that accompanied the public launch of this edition of the encyclopedia, participants opted to exclude a priori subjects relating to sacred rituals and medicinal plants. The primary aim of this notability provision is to protect certain areas of Atikamekw Nehirowisiwok knowledge that are traditionally passed on only from an elder to their apprentice and not in written form to everyone. This arrangement is also influenced by a certain mistrust engendered by the dispossession of Indigenous peoples of their resources in the history of colonization.

Another protocol specific to the Atikamekw Nehiromowin language edition of Wikipedia concerns the policy of verifiability of sources. In this encyclopedia, oral sources are among the most legitimate, given the primordial value placed on the word of elders in the transmission of knowledge, which traditionally takes the form of oral accounts. Compared with other language editions of Wikipedia, these protocols are highly unusual, since they specifically adapted to the ways in which knowledge is transmitted that characterize Indigenous peoples while at the same time fitting into the modern encyclopedia format that is characteristic of the Western world.

## Internationalize and Particularize

A second series of locality protocols concerns existing articles. Some of these protocols aim to introduce diverse cultural and linguistic viewpoints within a single article with an international scope. A case in point is the article "Médecin" (Doctor). In the French edition of Wikipedia, this article has been criticized for representing only the state of the medical profession



https://doi.org/10.15215/aupress/9781771994521.01

in France. This debate has led to a split between an article focusing on the French point of view, entitled "Médecin (France)," and a generic article with an international focus, "Médecin," which mentions that "the training of doctors varies considerably around the world." This latter article also includes a specific section on Canada.

The case of the article "Médecin (France)" illustrates the privilege given to a single point of view, which is legitimate in the case of a subject of strictly national interest. In this case, an issue arises concerning the recognition of the particular interest of an article for a culture or a country and, consequently, favouring a particular cultural point of view. In the English edition of Wikipedia, when an article has a "strong national link with the subject," it may also be written in the national (or regional) variant of the language characteristic of the group concerned ("Wikipedia: Manual of Style," 2026). However, identifying this strong national link is not always easy. It is often the subject of negotiation, which may be consensual or conflictual.

The article now entitled "Canada Jay" serves as a good example of this kind of negotiation. The article was originally entitled "Gray Jay" (the American spelling) and was then renamed "Grey Jay" (the Canadian spelling) after this bird was designated as Canada's national symbol by the magazine of the Royal Canadian Geographical Society. After further discussions lasting 2 years, the bird was renamed "Canada jay," but only once the International Ornithological Committee, the authority on bird names, had officially adopted this nomenclature. This example highlights a limitation of Wikipedia's locality protocols: For subjects where there is a well-established scientific nomenclature with international standards (chemistry, zoology), it is this external standard that tends to prevail over internal Wikipedia policy negotiations.

In the English Wikipedia, a whole series of tools has been created to indicate and document the use of the main variants of English. As mentioned earlier, there is a recommended style guide specific to Canadian English and a banner to display in 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" ("Template: Canadian English," 2026).

The Canadian style guide page describes in great detail a set of conventions and consensuses in force on how to name and write articles that concern Canadian subjects. These guidelines go further than mere stylistic advice:


https://doi.org/10.15215/aupress/9781771994521.01

They offer a "Canadian" interpretation of the encyclopedia's general policies and guidelines. As such, the guide is a crucial means of locating knowledge: It institutionalizes recognition of the cultural and linguistic specificity of Canadian content while at the same time giving Canadian editors a degree of power to act in terms of governance over the subjects that concern them ("Wikipedia: Manual of Style/Canada-Related Articles," 2025).

There is no equivalent to this specifically Canadian style guide in the French-language edition of Wikipedia. The guideline on style conventions simply includes a section on "geographical variants" of vocabulary, which states that "as certain terms of the French language may vary according to the country or region where they are used, the version best suited to the context should be preferred" ("Wikipédia: Conventions de style," 2025). For example, in an article relating specifically to Québec, the Québec term will be preferred, and in all other cases, the French spoken in France would be preferred. The French edition of Wikipedia has tried to adopt more formal conventions on how to include linguistic variants, but this attempt has failed. A survey carried out in 2008 concluded that a majority of Wikipedians thought it necessary to adopt such rules, but they were unable to agree on the nature of these conventions. The field of sports was the only one to benefit from a clear decision in 2006. The decision in this respect states that "articles on North American sports clubs will have a title in Canadian French." Accordingly, the article on the "Celtics de Boston" reflects French usage in Québec, unlike usage in France, where the English form "Boston Celtics" is more common.

Even so, the French linguistic variant generally remains the default choice, reflecting France's political and symbolic weight in the French-speaking world.

In the case of the Atikamekw Nehiromowin edition of Wikipedia, once again, an original protocol has been put in place to recognize and include language variants. The three communities that make up the nation each have their own linguistic usages. This fact was taken into account very early on in the discussions leading up to the public launch of the encyclopedia. It was agreed that the articles would compose three different sections, each dedicated to a community that would be free to write content reflecting its point of view and use of the language, or to merge the content. For example, the article "Kwekweciw" (Canada jay) has a single merged section for the communities of Manawan and Wemotaci and another for Opitciwan ("Kwekweciw," 2022). This content structure reflects the political structure of the Atikamekw nation



https://doi.org/10.15215/aupress/9781771994521.01

insofar as Manawan and Wemotaci participate together in the Atikamekw Nation Council, while Opitciwan has withdrawn from it.

## Summary: Differences Between Wikipedias

In each of Wikipedia's language editions, Canadians negotiate their place among other nations on the basis of pillars and principles that are common to all and guidelines that are specific to national cultures. These two types of protocol (commonality and locality) are complementary in the context of cultural pluralization. They provide frameworks for editorial negotiation between groups with divergent or even antagonistic perspectives without preventing conflict. Most of these protocols are addenda to the cardinal principle of the neutral point of view: They specify the conditions of what is included and excluded from the encyclopedia, taking into account the distinct characteristics of different local cultural contexts.

There are several major differences between the English, French, and Atikamekw Nehiromowin Wikipedias. The encyclopedias in English and Atikamekw Nehiromowin share a form of pluralization that establishes a strong expression of the local cultural characteristics (especially linguistic) of the various groups, nations, and countries concerned. The French-language edition of Wikipedia establishes the particularism of France as the dominant standard, with the corollary that there is little official recognition of the singular expression of other French-speaking cultures. The Wikipedia in Atikamekw Nehiromowin differs from other encyclopedias in that its protocols are clearly original: They are less formal, they legitimize the authority of oral sources, they exclude sacred subjects, and from the outset, they arrange different sections of the articles to reflect local variants of the language.

By comparing the French and English editions of the encyclopedia, it is clear that recognition of the Canadian variants of each language is much more formalized on the English side than on the French side. This recognition stems firstly from the clear rule that no single variety of English should take precedence over the others. The choice of a language variant is usually negotiated piecemeal, within each article, depending on the subject and the context in which it is written. Secondly, in the English-language encyclopedia, there is a clear recognition of the systemic biases that tend to favour an American point of view in spite of everything. These conditions have enabled the


https://doi.org/10.15215/aupress/9781771994521.01

development of a Canadian style guide, which institutionalizes the cultural and linguistic particularity of content relating to Canada.

On the other hand, in the French-language edition of Wikipedia, priority is given to France in terms of the language variant to be favoured. The common linguistic terrain is modelled on France's particularism, following the convention of least surprise for the benefit of the majority of readers in France. However, this precedence given to France is rarely stated explicitly in the official guidelines of the French-language encyclopedia. It is usually applied by default. Its pages of policies and guidelines avoid acknowledging the existence of "Francocentrism" in the encyclopedia. In a reversal, the fact of naming an ethnocentric bias in favour of France is deemed to be contrary to neutrality.

Generally speaking, in the French edition of Wikipedia, the expression of cultural variants of the language tends to be treated as a problem to be managed rather than as a condition of neutrality. There are several rare provisions establishing the use of Québec French in a few specific cases, but negotiations mainly take place during editing conflicts, where Francocentrism is discussed at length. The political issues of cultural pluralization therefore tend to be absent from the official pages and relegated to the back of the encyclopedia, in the discussion pages where publishing conflicts periodically resurface. Although the majority of French-speaking contributors want to establish clearer protocols for the use of cultural variants of the language, to date, it has not been possible to formalize them in official guidelines.

These differences between encyclopedias can be explained by a range of factors both internal and external to Wikipedia. The most influential factor seems to be geopolitical: It has to do with the relative weight of national powers (or local powers in the case of Atikamekw Nehiromowin) in each linguistic area. The English-speaking cultural sphere is certainly dominated by the United States from an economic and cultural point of view, but other major powers nevertheless coexist there with some influence (particularly the United Kingdom and, to a lesser extent, Canada, Australia, and India). The French-speaking world is more unipolar, with France largely dominating other countries and subnational groups, including Québec.

This external factor is compounded by an internal one, linked to the demographics of Wikipedians: In the French-language encyclopedia, French contributors are overrepresented compared to the number of readers in that country. When it comes to decision-making, French contributors have the advantage of numbers.



https://doi.org/10.15215/aupress/9781771994521.01

Finally, these differences can be explained by distinct political cultures. Each edition of Wikipedia has its own governance habits, largely influenced by the political philosophies of the different nations that participate in it: on the one hand, Anglo-Saxon multiculturalism, and on the other, French universalism. For its part, the Atikamekw Nehiromowin edition of the encyclopedia is shaped by a culture of consensus between the various components of the nation. These different political structures and cultures have a major influence on the way in which the balance of power between minority and majority groups is organized within each encyclopedia.

The processes of cultural pluralization that take place within each edition of the encyclopedia are marked by limitations and zones of ambiguity. These include the limited availability of reference sources on minority cultures and the challenges of representing Canada's internal cultural diversity. More fundamentally, Wikipedia's status as a global public sphere is a source of ambiguity for the management of the encyclopedic project. Does it aim to produce an international perspective written in different languages or a collection of encyclopedias rooted in the English-speaking, French-speaking, and Atikamekw Nehiromowin worlds, each centred on its collection seen from its own perspective? This debate is still ongoing and promises to keep the Wikipedian forums lively for a long time to come.

## Acknowledgements

The author would like to thank Thérèse Ottawa for translating articles in Atikamekw Nehiromowin, as well as Pierre-Yves Beaudouin, Dipsacus fullonum, Tagishsimon, Jim Hayes, Tilman Bayer, Luc Patin, and Étienne Beaulé for sharing information.

## References

Auray, N., Hurault-Plantet, M., Poudat, C., & Jacquemin, B. (2009). La négociation des points de vue: Une cartographie sociale des conflits et des querelles dans le Wikipédia francophone. *Réseaux*, *154*(2), 15–50. https://doi.org/10.3917/res.154.0015

Berson, A., Sengul-Jones, M., & Tamani, M. (2021). Unreliable guidelines: Reliable sources and marginalized communities in French, English and Spanish Wikipedias. *Art + Feminism*. https://artandfeminism.org/resources/research/unreliable-guidelines/



https://doi.org/10.15215/aupress/9781771994521.01

Cardon, D., & Levrel, J. (2009). La vigilance participative: Une interprétation de la gouvernance de Wikipédia. *Réseaux*, *154*(2), 51–89. https://doi.org/10.3917/res.154.0051

Casemajor, N., & Couture, S. (2020). Pluralisation des savoirs et cultures minoritaires: L'identité franco-ontarienne dans Wikipédia. *Revue d'anthropologie des connaissances*, *14*(4). https://journals.openedition.org/rac/14581

Casemajor, N., Gentelet, K., & Coocoo, C. (2019). Openness, inclusion and self-affirmation: Indigenous knowledge in open knowledge projects. *Journal of Peer Production*, *13*, 1–17.

Catégorie: Article à internationaliser. (2025, November 20). In *Wikipédia*. https://fr.wikipedia.org/w/index.php?title=Cat%C3%A9gorie:Article_%C3%A0_internationaliser&oldid=230813094

Chinese Canadians in Greater Vancouver. (2024, December 14). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Chinese_Canadians_in_Greater_Vancouver&oldid=1327513130

Discussion Wikipédia: Internationalisation [Discussion page]. (2025, January 29). In *Wikipédia*. https://fr.wikipedia.org/w/index.php?title=Discussion_Wikip%C3%A9dia:Internationalisation&oldid=222519980

Ford, H., & Geiger, R. S. (2012). Writing up rather than writing down: Becoming Wikipedia literate. In *Proceedings of the eighth annual international symposium on Wikis and open collaboration (WikiSym '12)* (pp. 1–4). Association for Computing Machinery. https://doi.org/10.1145/2462932.2462954

Fraser, N. (2004). Justice sociale, redistribution et reconnaissance. *Revue du MAUSS*, *23*(1), 152–64. https://doi.org/10.3917/rdm.023.0152

Hargittai, E., & Walejko, G. (2008). The participation divide: Content creation and sharing in the digital age. *Information, Communication & Society*, *11*(2), 239–56. https://doi.org/10.1080/13691180801946150

Hill, B. M., & Shaw, A. (2013). The Wikipedia gender gap revisited: Characterizing survey response bias with propensity score estimation. *PLOS ONE*, *8*(6), e65782. https://doi.org/10.1371/journal.pone.0065782

Honneth, A. (2004). La théorie de la reconnaissance: Une esquisse. *Revue du MAUSS*, *23*(1), 133–36. https://doi.org/10.3917/rdm.023.0133

Kwekweciw. (2022, May 4). In *Wikipedia*. https://atj.wikipedia.org/w/index.php?title=Kwekweciw&oldid=14612

Lemieux, M. E., Zhang, R., & Tripodi, F. (2023). "Too soon" to count? How gender and race cloud notability considerations on Wikipedia. *Big Data & Society*, *10*(1). https://doi.org/10.1177/20539517231165490

Littérature franco ontarienne. (2025, July 7). In *Wikipédia*. https://fr.wikipedia.org/w/index.php?title=Litt%C3%A9rature_franco-ontarienne&oldid=227084536



https://doi.org/10.15215/aupress/9781771994521.01

Massa, P., & Scrinzi, F. (2012). Manypedia: Comparing language points of view of Wikipedia communities. In *Proceedings of the eighth annual international symposium on Wikis and open collaboration (WikiSym '12)* (pp. 1–9). Association for Computing Machinery. https://doi.org/10.1145/2462932.2462960

McDowell, Z. J., & Vetter, M. A. (2021). *Wikipedia and the representation of reality*. Routledge.

Projet: Québec. (2022, June 12). In *Wikipédia*. https://fr.wikipedia.org/w/index.php?title=Projet:Qu%C3%A9bec&oldid=194461981

Projet: Suisse/Admissibilité. (2025, June 21). In *Wikipédia*. https://fr.wikipedia.org/w/index.php?title=Projet:Suisse/Admissibilit%C3%A9&oldid=226681793

Reagle, J., & Koerner, J. (2020). *Wikipedia @ 20: Stories of an incomplete revolution*. MIT Press.

Systemic bias of Wikipedia. (2023, August 12). In *Meta-Wikimedia*. https://meta.wikimedia.org/w/index.php?title=Systemic_bias_of_Wikipedia&oldid=25448328

Template: Canadian English. (2026, January 24). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Template:Canadian_English&oldid=1334605667

Tkacz, N. (2014). *Wikipedia and the politics of openness*. University of Chicago Press.

Wikipédia: Conventions de style. (2025, December 6). In *Wikipédia*. https://fr.wikipedia.org/w/index.php?title=Wikip%C3%A9dia:Conventions_de_style&oldid=231231622

Wikipedia: Disambiguation. (2026, January 19). In *Wikipedia*. http://en.wikipedia.org/w/index.php?title=Wikipedia:Disambiguation&oldid=1333782459

Wikipédia: Internationalisation. (2023, September 7). In *Wikipédia*. https://fr.wikipedia.org/w/index.php?title=Wikip%C3%A9dia:Internationalisation&oldid=208227778

Wikipedia: Manual of Style. (2026, January 29). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Wikipedia:Manual_of_Style&oldid=1335504315

Wikipedia: Manual of Style/Canada-related articles. (2025, September 11). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Wikipedia:Manual_of_Style/Canada-related_articles&oldid=1310808315

Wikipedia: Neutral point of view. (2026, January 15). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Wikipedia:Neutral_point_of_view&oldid=1333125479

Wikipedia: Notability. (2026, January 30). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Wikipedia:Notability&oldid=1335653865

Wikipédia: Notoriété. (2025, December 17). In *Wikipédia*. https://fr.wikipedia.org/w/index.php?title=Wikip%C3%A9dia:Notori%C3%A9t%C3%A9&oldid=231524277



https://doi.org/10.15215/aupress/9781771994521.01

Wikipédia: Principe de moindre surprise. (2025, May 12). In *Wikipédia*. https://fr.wikipedia.org/w/index.php?title=Principe_de_moindre_surprise&oldid=225606187

Wikipedia: Tendentious editing. (2026, January 24). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Wikipedia:Tendentious_editing&oldid=1334530628

Wikipedia: WikiProject Countering systemic bias. (2025, November 25). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Wikipedia:WikiProject_Countering_systemic_bias&oldid=1324167268

Witcihikoiin: Wikipetcia Atikamekw Nehiromowin. (2017, June 28). In *Wikipedia*. https://atj.wikipedia.org/w/index.php?title=Witcihikoiin:Wikipetcia_Atikamekw_Nehiromowin&oldid=6774



https://doi.org/10.15215/aupress/9781771994521.01

# 2 Does Wikipedia's Acadia Portal Offer an Accurate Portrait?

Gabriel Arsenault and Mathieu Wade

Studies show that the content available on Wikipedia is generally fairly reliable (Giles, 2005). However, scholars writing about Wikipedia deplore the fact that only a small minority of articles are of high "quality." The "quality" of a Wikipedia article (or portal) is assessed by the Wikipedia community, which in turn is invited to vote on the article's ranking based on various criteria (e.g., encyclopedic writing style, quality of references, clarity of organization, or choice of images). The most prestigious ranking is "featured article," followed by "good article." In June 2024, fr.wikipedia had 2,151 "featured articles" and 3,999 "good articles," representing just 0.08% and 0.15%, respectively, of all the articles in the French-language encyclopedia. However, a study undertaken a few years ago of the eight main Wikipedias revealed that fr.wikipedia was the most demanding portal in terms of assessing the quality of articles (Jemielniak & Wilamowski, 2017).

Academic researchers interested in the quality of the encyclopedia also criticize its articles for suffering from a significant patriarchal bias. Martel and Villeneuve (see chapter 3) report that biographies of women represent only 18.8% of all biographies on fr.wikipedia (and 18.3% on en.wikipedia). This underrepresentation of women in the encyclopedia not only reflects societal patriarchy; it also demonstrates a patriarchal bias within Wikipedia itself, given that women, on average, have to be more prominent than men to get a Wikipedia article (Adams et al., 2019; Wagner et al., 2016). According to an as-yet-unpublished study cited in Adams et al. (2019), if the same level of prominence were required of women as men to have a Wikipedia article



https://doi.org/10.15215/aupress/9781771994521.01

devoted to them, there would be 70,000 more women's biographies in the encyclopedia. For reasons that are still not fully understood, women may only represent around 10% of all contributors to the encyclopedia (see chapter 3).

In this article, we propose to look at the merits and limitations of the Wikipedia encyclopedia, based on one of the 1,800 or so French-language Wikipedia portals: the portal devoted to Acadia ("Acadie" in French).[1] As of June 2024, Wikipedia devotes 2,013 articles to Acadia and is the leading source of accessible information on the subject.[2] Yet the portal has no "featured" articles, although it does have 24 "good" articles (1.2% of the total). These statistics are useful, yet they only serve as very superficial indicators of the Acadia portal's quality. How can we get beyond that?

Although our analysis of the Acadia portal focuses on a minority and globally marginal population—the Acadians—it is not related to thinking about the issues of the pluralization and internationalization of Wikipedia (see the chapter by Casemajor in the present volume). Many studies have looked at the process of knowledge pluralization in Wikipedia from a postcolonial and feminist theoretical perspective (Bjork-James, 2021; Casemajor & Couture, 2020; Ford & Wajcman, 2017). These approaches tend to focus on the conflicts surrounding the representation of minority groups as well as on publishing protocols likely to produce and reproduce systemic biases against them.

These situations refer to what Godrie and Dos Santos call epistemic inequalities—that is, "a particular type of inequality that manifests itself in the access, recognition and production of knowledge and different forms of ignorance" (Godrie & Dos Santos, 2017, p. 7). The Acadia portal could be studied as a marginal place where a minority group struggles for recognition of the legitimacy of its culture and identity. Nathalie Casemajor and Stéphane Couture have proposed an analysis of the Franco-Ontarian portal along these lines, studying "knowledge practises concerning Franco-Ontarians and more specifically . . . the negotiation of knowledge concerning the identity of this group as represented in Wikipedia" (2020). However, this is not the focus of our own research.

1 For an analysis of the portal devoted to Franco-Ontarians, see Casemajor and Couture (2020).

2 Unless otherwise indicated, all data for the Acadia portal was extracted in July 2021.



https://doi.org/10.15215/aupress/9781771994521.01

Instead, we propose to consider the Acadia portal as an important site for the production of content on Acadia, even though it has never been studied before. We therefore anchor our analysis in the field of Acadian studies in order to see what answers the portal provides to the two main questions that structure this field: (1) Does Acadia refer to a relatively circumscribed territory in the Atlantic provinces of Canada or to a diaspora without borders? (2) Does it refer to an administrative territory that disappeared after the Great Upheaval of 1755 (when British forces deported almost all Acadians to other territories, from the thirteen colonies to the Antilles and France itself) or to a living nation? The first question is properly political and concerns the ways in which people belong to a group. The question of where Acadia is, and consequently who Acadian is, is frequently debated in both civil society and the scientific community. What answers does Wikipedia's Acadia portal provide to these questions? The second question has more to do with scholarly research. The question is not whether Acadia still exists—the answer is obviously yes, it does. However, the colonial history of Acadia and Acadian social science are two distinct fields that operate largely in parallel.

The field of colonial studies is more internationalized and enjoys greater legitimacy. Acadian social sciences, which generally focus on the nationalist Acadia that emerged in the 19th century, are more peripheral and local. Which of these two fields dominates the portal on Wikipedia? Our study is based on the premise that Acadia is a subject with blurred geographical and temporal boundaries and that defining these boundaries is one of the main political and scientific issues at stake. We propose a systematic analysis of the portal to see what representations of Acadia emerge. First, we will look at the geography of Acadia as presented on the portal. We will then analyze the historical periods covered by the articles appearing there. Finally, we will develop a portrait of the most frequently consulted articles and the main contributors, which will serve to show that the patriarchal bias characterizing the encyclopedia as a whole also characterizes the Acadia portal.

By analyzing the Acadia portal from the perspective of Acadian studies, we will be able to compare theoretical and political questions with the answers proposed by a collaborative space and to highlight a series of biases, some of which are typical of Wikipedia, whereas others are specific to the Acadian setting.


https://doi.org/10.15215/aupress/9781771994521.01

## Acadian Space and Time

The Acadia portal is unique in that it deals with an entity with blurred space-time boundaries that are, moreover, the subject of much debate. Since "Acadia" has neither a politically recognized state of its own nor clear administrative boundaries, it can take on a variety of meanings. These meanings in turn lead us to distinct fields of scientific endeavour but also to public controversies about the very meaning of "acadianité" (the sense of being Acadian), the relationship with institutions, language, and the historical past.

### Acadian Space

In the absence of formal administrative boundaries, Acadia—particularly modern Acadia—is part of a contentious and fragmented geography. This indeterminacy applies to the colonial period, when boundaries were approximate and contested (Desbarats & Greer, 2011), but it also applies to the modern period, when different conceptions of Acadian identity and modes of belonging have clashed. The most heated debates, however, concern the modern period. Broadly speaking, we can distinguish between a diasporic Acadia and a territorial Acadia. These two Acadias present us with considerably different geographies as well as definitions of Acadian identity that are in part incompatible.

Some people hold that a modern Acadian territory simply cannot be defined. Acadia is said to be a diasporic people in the form of an archipelago of communities, initially dispersed by the Great Upheaval, then by waves of emigration since the 19th century (Bérubé, 1987; Magord, 2010). Acadian communities can be found from Louisiana (Basque, 2009; Bruce & Urbain, 2021) to New England, via France (Magord, 2010) and Québec (Bergeron et al., 2008). This Acadia is defined essentially by genealogy and is expressed today in a mostly ad hoc and festive mode (Lefebvre, 2012). Diasporic Acadia finds its most eloquent expression in the Congrès mondial acadien (CMA), which was launched in 1994. The CMA explicitly promotes a diasporic conception of Acadian identity and makes family reunions central, thereby reaffirming a genealogical vision of Acadian identity. It also invariably provokes debate within the Atlantic Acadian public sphere (Allain, 1997; Bruce, 2018). Diasporic Acadia focuses more on the individual than on any given territory and consists more of networks and one-off events than institutions.



https://doi.org/10.15215/aupress/9781771994521.01

Some people object to this conception of Acadia on the grounds that it threatens the full political development of the Acadian nation by diluting its membership and promoting a genealogical rather than a civic definition of the sense of being Acadian (Thériault, 2006). For many, Acadia refers more to a desire to "socialize" in French in the Atlantic provinces (Daigle, 1994; Thériault, 1995). This Acadia centres around French-speaking institutions that enable collective action and cultural and linguistic transmission. Indeed, this Acadia is defined more by language and territory than by parentage and is represented by organizations at the provincial level (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) as well as the regional level (Société nationale de l'Acadie).

As table 2.1 shows, the portal's position in this debate is somewhat nuanced. On the one hand, whether we look at the articles on places, the geography of the groups and organizations selected, or the birthplaces of personalities on the portal, we come to a similar observation: The vast majority of articles focus on the Atlantic provinces. This is the case for 77.3% of articles on places, 76.6% of articles on personalities, and 85% of articles on groups and organizations. On the other hand, the portal admittedly contains a significant number of articles relating to subjects outside the Atlantic region, in particular Louisiana, Maine, and Québec, as well as other American states, Canadian provinces, and French regions.

**Table 2.1.**
*Geography of personalities (birthplaces), groups and organizations, and locations selected for the portal (June 2024)*

| | Number of places | Number of personalities | Number of groups and organizations |
|---|---|---|---|
| Atlantic region | 675 | 505 | 113 |
| Outside the Atlantic region | 198 | 154 | 20 |

*Note:* Data from Wikipedia. We exclude here places, personalities, groups, and organizations whose pages do not explicitly specify a geographic location. We also exclude places that may be both in the Atlantic region and outside the Atlantic region, such as New France.


https://doi.org/10.15215/aupress/9781771994521.01

## Acadian Time

Another controversy concerns the temporality of Acadia: Does it belong mostly to the past or the present? Broadly speaking, we can distinguish between a colonial Acadia and a modern Acadia. These two Acadias have the same name and are part of the same portal, unlike New France and Québec, for example, which occupy separate spaces in Wikipedia. Colonial Acadia refers to an administrative territory roughly corresponding to the Maritime provinces and more specifically to present-day Nova Scotia. Established in 1604, colonial Acadia was at the heart of numerous conflicts between France and Great Britain until the Great Upheaval, which took place between 1755 and 1763, during the Seven Years' War. This deportation marked a major break in time and space for Acadia. Modern Acadia was built by the survivors of the deportation and became institutionalized starting in the second half of the 18th century. Subsequently, modern Acadia was organized around an ideology and national symbols, as developed at the Acadian National Conventions from the end of the 19th century onward. These two Acadias form distinct fields of scientific study. The study of colonial Acadia is now the focus of international research networks, mostly in English (Faragher, 2005; Griffiths, 2005; Hodson, 2012; Kennedy, 2014; Reid et al., 2004), whereas scientific studies on modern Acadia are generally produced in French in the Maritime provinces (Belliveau, 2014; Boudreau, 2016; Landry, 2015; Poplyansky, 2018; Thériault, 1995).

What position does the Acadia portal take with respect to this dichotomy? The portal on the English-language edition of Wikipedia clearly interprets Acadia as a colony of the historical past. The portal on the French-language edition of Wikipedia characterizes Acadia in two ways within the list of portals. First, Acadia is found in the "History" section and the "Modern Period (16th to 18th Centuries)" subsection, like the portal on New France. Second, Acadia is found in the "Geography" section and in the "Historical Territories" subsection, again following the example of New France.

Similarly, as table 2.2 illustrates, the deportation is unequivocally the single event in Acadian history that attracts the most attention from internet users.

However, as figure 2.1 illustrates, a very large majority of the personalities included in the portal were born after 1764 and therefore belong to modern Acadia rather than colonial Acadia (1604–1763). Clearly, the portal reflects the debate surrounding the different time periods in Acadia. There is also a clear



https://doi.org/10.15215/aupress/9781771994521.01

**Table 2.2.**
*Most popular events on the Acadia portal (June 6, 2019–June 6, 2024)*

| Events | Number of times consulted |
|---|---|
| Deportation of the Acadians | 237,727 |
| Great Upheaval | 94,068 |
| Acadia's national holiday | 21,096 |
| Treaty of Fontainebleau (1762) | 14,717 |
| World Acadian Congress | 13,075 |
| Battle of Fort Beauséjour | 10,282 |
| Moncton shooting in 2014 | 8,978 |
| Acadian Games | 8,456 |
| Deportation from Île Saint-Jean (Prince Edward Island) | 8,281 |
| Causes of the Great Upheaval | 7,969 |
| Louis Mailloux affair | 7,632 |
| Treaty of Utrecht (1713) | 5,943 |
| Battle of Cran (1758) | 4,974 |
| Acadian Renaissance | 3,266 |
| The first Acadian National Convention | 2,251 |

divide between supply, which is more in line with the modern definition of Acadia, and demand, which is much more in line with the colonial definition.

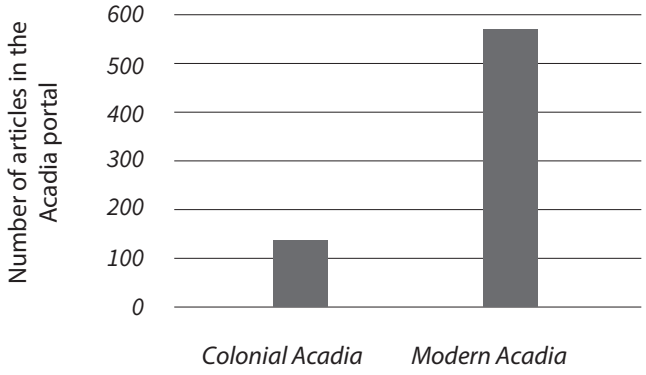
## Gender

In terms of the gender distribution of personalities on the Acadia portal, there is an overrepresentation of men, which is also the case with Wikipedia in general. Of the 679 personalities on the portal, 569 are men (83.8%), 106 are women (15.6%), and 4 identify as queer or nonbinary (0.6%).

It would appear difficult not to assume a causal link between this male domination of content and the domination of article production, which is also typical of the Wikipedia encyclopedia as a whole. As table 2.3 shows, none of


https://doi.org/10.15215/aupress/9781771994521.01

**Figure 2.1.**
*Temporality of the personalities (dates of birth) selected in the Acadia portal (June 2024)*




*Note:* We exclude here all personalities whose articles do not specify the birthdate.

the main contributors to the portal identify themselves as women. In fact, the table lists 18 people as the main authors of at least 10 articles on the portal: of these, 12 identify themselves as men, while 6 do not reveal their gender. The table also confirms the dominant role of Red Castle, the main contributor to 39% of the portal's articles and 19 of its 24 "good articles."

## Legitimate Knowledge and "Good Articles"

Several studies that approach Wikipedia from a critical and postcolonial angle point to the epistemic inequalities suffered by minorities and marginalized groups (Godrie & Dos Santos, 2017). The knowledge produced by these groups is not necessarily recognized as legitimate. These communities would thus be subject to exogenous discursive productions; they would be forced to represent themselves with categories that are not their own. The concept of "exiguïté" (smallness), initially developed by François Paré (1992) to understand minority francophone literary production in Canada and then applied to the social sciences by Mourad Ali-Khodja (2003), describes precisely this phenomenon in the Canadian French-speaking sphere. According to this theory, exogenous knowledge about Acadia


https://doi.org/10.15215/aupress/9781771994521.01

**Table 2.3.**
*Main contributors to the portal (June 2024)*

| Username | Number of portal articles where user is designated as main contributor | Gender |
|---|---|---|
| Red Castle | 784 | Male |
| Quéré | 120 | Male |
| Fralambert | 84 | Male* |
| Rc1959 | 42 | Unknown |
| Parigot | 40 | Male |
| Amqui | 26 | Male |
| Thierry Caro | 22 | Male |
| Passoat15 | 19 | Male |
| Jeangagnon | 17 | Male |
| ADM | 15 | Unknown |
| Shawn à Montréal | 16 | Male* |
| GJFraser | 16 | Male |
| Seanoconaill | 14 | Male* |
| Funnyhat | 14 | Male |
| Pierre5018 | 11 | Unknown |
| Patangel | 11 | Unknown |
| BoucherCL | 10 | Male |
| Jihaim | 10 | Unknown |

* The asterisk indicates that the gender of the person has been identified on the basis of the gender agreements used on their presentation page.

(i.e., developed from outside Acadia) can be expected to be dominant within the portal.

The field of Acadian studies comprises (1) colonial studies, produced in English and more internationalized and legitimate, and (2) contemporary Acadian social sciences, produced in French and more marginal. Colonial studies would appear to have greater legitimacy, yet it does not clearly dominate the Acadia portal. On the one hand, the portal's description corresponds


https://doi.org/10.15215/aupress/9781771994521.01

to a colonial definition of Acadia, whereas contemporary Acadia is clearly more represented in its content, and this is the case whether we consider the temporality and the geography of the personalities selected and the geography of the places and the groups and organizations featured on the portal. On the other hand, as table 2.4 shows, the vast majority of the 24 "good articles" on the portal relate to the contemporary period rather than the colonial period. There is no epistemic inequality here. The relationship to knowledge emerging from the portal is considerably more nuanced and complex.

## How Can the Acadia Portal Be Revitalized?

The Acadia portal shows that Acadia is both colonial and modern, diasporic and rooted in the Atlantic provinces. This tension, which goes to the very heart of contemporary Acadia, is clearly articulated on the portal. It strikes

**Table 2.4.**
*"Good articles" on the Acadia portal based on the preferred Acadian time frame (June 2024)*

| Colonial Acadia (n = 3) | Contemporary Acadia (n = 16) | Colonial and contemporary Acadia (n = 5) |
|---|---|---|
| Deportation from Île Saint-Jean | Louis Mailloux affair | Acadia |
| Françoise-Marie Jacquelin | Atholville | Acadian architecture |
| Gamaliel Smethurst | Caraquet | Sinclair Inn |
| | Chéticamp | History of Acadia |
| | Acadian flag ("Drapeau de l'Acadie") | Acadian theatre |
| | Île Brion | |
| | Matapedia | |
| | Luc Bourdon | |
| | Memramcook | |
| | Paul Carmel Laporte | |
| | Pierre-Amand Landry | |
| | Pokemouche | |
| | Port de Caraquet | |
| | Sainte-Anne-du-Bocage | |
| | UNI Coopération financière | |
| | Matapedia Valley | |



https://doi.org/10.15215/aupress/9781771994521.01

us as significant that the majority of Acadian personalities featured on the portal belong to the modern era, while almost all the events in Acadian history of interest to internet users concern the colonial period. It is nonetheless surprising to see the relatively low level of interest of both contributors and internet users in events that could be associated with the Acadian "Quiet Revolution," such as the implementation in the 1960s of Premier Louis J. Robichaud's government program, Chances égales pour tous (the New Brunswick Equal Opportunity program), which is currently the subject of a fairly superficial article.

By quantifying this relative absence of articles on key moments in Acadian political modernity, our study provides unpublished data in support of an already well-articulated sociological critique of Acadia: While there are indeed Acadian individuals today, the agency of the Acadian political subject is very weak. In the absence of clearly defined political boundaries, Acadia is forced to act through a diverse but nonetheless precarious network of associations (Allain, 2003). As a result, Acadian political action is fragmented. And in the absence of representative institutions, Acadian discourse is divided into sectoral organizations marked by a certain corporatism (Landry, 2015). Finally, Acadian organizations are economically dependent on the Canadian state, which limits the room for manoeuvre of the community's representative bodies (Léger, 2012). For many, this fragmentation and dependence undermine Acadia's ability to project itself into the future, to develop collective projects, and to think of itself as a political subject. This difficulty in representing itself as a political subject partly structures the content of the portal.

The portal also has two major shortcomings that unfortunately characterize Wikipedia content in general: Almost all of the articles are still in draft form, and the Acadia featured on the portal is dominated by men. To improve the quality of the portal's articles and make it more representative of the diversity of the Acadian population, the participation of civil society in the development of the Acadia portal needs to be stimulated. Collective mobilization efforts such as the Bibliothèque et Archives nationales du Québec's monthly wiki workshops or the initiative of the municipality of Monmouth in Wales—consisting of setting up some 1,000 QR codes linking various sites in the city to a Wikipedia article (Kooser, 2012)—thus inspired us to create the Wikiclub Acadie. In October 2018, following a lecture by Patrick Degrâce, a.k.a. Red Castle, at the Université de Moncton, a first informal Wikiclub was set up (Doiron, 2018; Radio-Canada, 2018). Between autumn 2018


https://doi.org/10.15215/aupress/9781771994521.01

and spring 2019, half a dozen writing sessions were held, bringing together members of the university community. At the time, however, we concluded that without a more formal structure, the "club" was not destined to be very dynamic.

To overcome this stalemate, we established a partnership with the Société nationale de l'Acadie (SNA) in the summer of 2020. Thanks to the SNA's support, we published a research report on the Acadia portal (Arsenault & Wade, 2020) that was aimed particularly at identifying priorities for future contributors. The report contains, for example, a list of 40 prominent Acadian women, identified by historian Maurice Basque, who could be the subject of a biography on Wikipedia. Following the publication of this report, in September 2020, the SNA formally took charge of the Wikiclub project (Boudreau, 2020). In the spring of 2021, training workshops were held for members of the Regroupement féministe du Nouveau-Brunswick and Fédération des femmes acadiennes de la Nouvelle-Écosse.

In the summer of 2021, the SNA received a major grant from Canadian Heritage, enabling it to organize a series of virtual activities in March 2022 specifically targeting the Université de Moncton community: a round table, a training workshop, and a contribution session (Radio-Canada, 2022). Over the coming years, the SNA is committed to ensuring that Wikiclub Acadie maintains its dynamism by organizing other training activities in the French-speaking communities of the four Atlantic provinces. Let's hope that this mobilization will be fruitful and lead to concrete improvements to the portal.

## References


Adams, J., Brückner, H., & Naslund, C. (2019). Who counts as a notable sociologist on Wikipedia? Gender, race, and the "professor test." *Socius*, *5*. https://doi.org/10.1177/2378023118823946

Allain, G. (1997). Le Congrès Mondial Acadien de 1994: Réseaux, conflits, réalisations. *Revue de l'Université de Moncton*, *30*(2), 141–59.

Allain, G. (2003). Fragmentation ou vitalité? Regard sociologique sur l'Acadie actuelle et ses réseaux associatifs. In S. Langlois & J. Létourneau (Eds.), *Aspects de la nouvelle francophonie canadienne* (pp. 231–54). Presses de l'Université Laval. https://www.erudit.org/en/books/culture-francaise-damerique/aspects-nouvelle-francophonie-canadienne/000637co/




https://doi.org/10.15215/aupress/9781771994521.01

Arsenault, G., & Wade, M. (2020). Une analyse du portail Acadie dans Wikipédia: Un rapport de recherche. *Société nationale de l'Acadie*. https://snacadie.org/images/PDF/Une_analyse_du_Portail_Acadie_dans_Wikipedia_SNA200917.pdf

Basque, M. (2009). Acadiens, Cadiens et Cajuns: Identités communes ou distinctes? In U. Mathis-Moser & G. Bischof (Eds.), *Acadians and Cajuns: The politics and culture of French minorities in North America* (pp. 24–37). Innsbruck University Press.

Belliveau, J. (2014). *Le "moment 68" et la réinvention de l'Acadie*. Presses de l'Université d'Ottawa.

Bergeron, J., Vézina, H., Houde, L., & Tremblay, M. (2008). La contribution des Acadiens au peuplement des régions du Québec. *Cahiers québécois de démographie*, *37*(1), 181–204. https://doi.org/10.7202/029644ar

Bérubé, A. (1987). De l'Acadie historique à l'Acadie à la nouvelle Acadie: Les grandes perceptions contemporaines de l'Acadie. In J. Lapointe & A. Leclerc (Eds.), *Les Acadiens: État de la recherche* (pp. 198–226). Conseil de la vie française en Amérique.

Bjork-James, C. (2021). New maps for an inclusive Wikipedia: Decolonial scholarship and strategies to counter systemic bias. *New Review of Hypermedia and Multimedia*, *27*(3), 207–28. https://doi.org/10.1080/13614568.2020.1865463

Boudreau, A. (2016). *À l'ombre de la langue légitime: L'Acadie dans la francophonie*. Classiques Garnier.

Boudreau, A. (2020). Wikiclub, un projet de la SNA pour bonifier la présence de l'Acadie sur Wikipédia. *Acadie Nouvelle*. https://www.acadienouvelle.com/actualites/2020/09/17/wikiclub-un-projet-de-la-sna-pour-bonifier-la-presence-de-lacadie-sur-wikipedia/

Bruce, C. (2018). L'oubli de l'Acadie politique? Le débat sur les Congrès mondiaux acadiens à la lumière de la question diasporique. *Minorités linguistiques et société / Linguistic Minorities and Society*, *10*, 100–132. https://doi.org/10.7202/1054098ar

Bruce, C., & Urbain, E. (2021). Discours du tourisme diasporique: L'exemple d'une visite louisianaise en Acadie. *Argumentation et Analyse du Discours*, *27*. https://doi.org/10.4000/aad.5483

Casemajor, N., & Couture, S. (2020). Pluralisation des savoirs et cultures minoritaires: L'identité Franco-Ontarienne dans Wikipédia. *Revue d'anthropologie des connaissances*, *14*(4). https://journals.openedition.org/rac/14581

Daigle, J. (1994). *L'Acadie des Maritimes: Études thématiques des débuts à nos jours*. Chaires d'Études acadiennes.

Desbarats, C., & Greer, A. (2011). Où est la Nouvelle-France? *Revue d'histoire de l'Amérique française*, *64*(3–4), 31–62. https://doi.org/10.7202/1017969ar

Doiron, J.-M. (2018, October 8). Un Acadien a fait plus de 60 000 contributions à Wikipedia. *Acadie Nouvelle*. https://www.acadienouvelle.com/actualites/2018/10/08/un-acadien-a-fait-plus-de-60-000-contributions-a-wikipedia/



https://doi.org/10.15215/aupress/9781771994521.01

Faragher, J. M. (2005). *A great and noble scheme: The tragic story of the expulsion of the French Acadians from their American homeland*. W. W. Norton.

Ford, H., & Wajcman, J. (2017). "Anyone can edit," not everyone does: Wikipedia's infrastructure and the gender gap. *Social Studies of Science*, *47*(4), 511–27. https://doi.org/10.1177/0306312717692172

Giles, J. (2005). Internet encyclopaedias go head to head. *Nature*, *438*(7070), 900–901. https://doi.org/10.1038/438900a

Godrie, B., & Dos Santos, M. (2017). Inégalités sociales, production des savoirs et de l'ignorance. *Sociologie et sociétés*, *49*(1), 7–31. https://doi.org/10.7202/1042804ar

Griffiths, N. E. S. (2005). *From migrant to Acadian: A North American border people, 1604–1755*. McGill-Queen's University Press.

Hodson, C. (2012). *The Acadian diaspora: An eighteenth-century history*. Oxford University Press.

Jemielniak, D., & Wilamowski, M. (2017). Cultural diversity of quality of information on Wikipedias. *Journal of the Association for Information Science and Technology*, *68*(10), 2460–70. https://doi.org/10.1002/asi.23901

Kennedy, G. M. W. (2014). *Something of a peasant paradise? Comparing rural societies in Acadie and the Loudunais, 1604–1755*. McGill-Queen's University Press.

Kooser, A. (2012). World's first "Wikipedia town" covered in QR codes. *CNET*.

Landry, M. (2015). *L'Acadie politique: Histoire sociopolitique de l'Acadie du Nouveau-Brunswick*. Presses de l'Université Laval.

Lefebvre, M. (2012). *Le rôle géographique de la fête: Le Congrès mondial acadien comme catalyseur identitaire et inhibiteur de frontières* [Doctoral dissertation, Université d'Ottawa / University of Ottawa]. http://hdl.handle.net/10393/22871

Léger, R. (2012). Le régime linguistique canadien à l'épreuve du désir de faire société. *International Journal of Canadian Studies / Revue internationale d'études canadiennes*, *45–46*, 187–98. https://doi.org/10.7202/1009901ar

Magord, A. (2010). *Le fait acadien en France: Histoire et temps présent*. Geste.

Paré, F. (1992). *Les littératures de l'exiguïté*. Le Nordir.

Poplyansky, M. (2018). *Le Parti acadien et la quête d'un paradis perdu*. Septentrion.

Radio-Canada. (2018, October 5). Faire rayonner l'Acadie sur Wiki. *L'heure de pointe–Acadie*. Radio-Canada. https://ici.radio-canada.ca/ohdio/premiere/emissions/l-heure-de-pointe-acadie/segments/reportage/89746/wikipedia-acadie-nouveau-brunswick

Radio-Canada. (2022, February 24). Une première formation pour le Wikiclub Acadie. *L'heure de pointe–Acadie*. Radio-Canada. https://ici.radio-canada.ca/ohdio/premiere/emissions/l-heure-de-pointe-acadie/segments/entrevue/391720/nb-wikiclub-acadie-sna-web



https://doi.org/10.15215/aupress/9781771994521.01

Reid, J. G., Basque, M., & Mancke, E. (2004). *The "conquest" of Acadia, 1710: Imperial, colonial, and Aboriginal constructions*. University of Toronto Press.

Thériault, J. Y. (1995). *L'identité à l'épreuve de la modernité: Écrits politiques sur l'Acadie et les francophonies canadiennes minoritaires*. Éditions d'Acadie.

Thériault, J. Y. (2006). Identité, territoire et politique en Acadie. In A. Magord (Ed.), *Adaptation et innovation: Expériences acadiennes contemporaines* (pp. 37–49). Peter Lang.

Wagner, C., Graells-Garrido, E., Garcia, D., & Menczer, F. (2016). Women through the glass ceiling: Gender asymmetries in Wikipedia. *EPJ Data Science*, *5*(1), 1–24. https://doi.org/10.1140/epjds/s13688-016-0066-4



https://doi.org/10.15215/aupress/9781771994521.01

*This page intentionally left blank.*

# 3 Using Wikidata to Quantify the Gender Gap in Biographical Resources

Marie D. Martel and Simon Villeneuve

As a source of information, Wikipedia belongs to the category of encyclopedic reference works. Encyclopedism aims to "bring together as much knowledge as possible, and connect it, transmit it, share it and submit it for discussion" (Melançon, 2018). The Wikipedia encyclopedia operates under a free licence. Since 2001, it has helped set up the Wikimedia project and movement, which in turn have spawned an ecosystem of free online knowledge. Today, this content makes it possible to share, quantify, and discuss information that was previously unpublished, confidential, or disparate.

Encyclopedias are shaped by their cultural roots, ideologies, and values, as well as by the representation of reality and the knowledge they mobilize or legitimize (McDowell & Vetter, 2021; Rey, 2007). In this respect, Wikipedia is known for perpetuating sexist biases in terms of both editorial participation and biographical coverage, like so many other encyclopedias, and Wikipedia's biases have not failed to attract media attention (Cohen, 2011; Nadeau, 2017).

Writing in *The New York Times* in 2011, Noam Cohen did not articulate a definition; instead, he suggested grounding the concept of the "gender gap" in the case of Wikipedia, which tends to discourage female contributors. More specifically, the gender gap, in this context, refers to the disparities in participation, content, and representation between genders on the various Wikimedia platforms. Meta-Wiki serves as a discussion platform for the various Wikimedia projects and describes two types of gender gap: "(a) a content gender gap (meaning that more men than women are covered in the main



https://doi.org/10.15215/aupress/9781771994521.01

space content of our wikis), and (b) a participation gender gap, meaning that more men participate in the peer production communities of Wikimedia" ("Gender Gap," 2025).

The number of female editors on Wikipedia across the different language versions is currently estimated at 13% on average, while the number of editors of genders other than male and female (cisgender) is estimated at 4% according to the *Community Insights 2023 Report* (Wikimedia Foundation, 2023). It also appears that between 2019 and 2022, the proportion of active editors of genders other than male and female (cisgender) doubled. Meanwhile, the proportions of male and female editors have remained broadly the same over this period, although there has been a significant increase in the number of female editors since 2019. According to several studies focusing on this issue, the gender gap in participation is reflected in biographical articles and their content.

In this chapter, we examine the gender gap as it affects content. We aim to present a method for documenting descriptions of gender from a sample of 80 biographical resources that are strongly linked on Wikidata, the knowledge base that powers Wikipedia. This exercise will allow us to produce new data on the percentage of biographies devoted to women in these resources, including notable Canadian reference works, and in Wikimedia projects such as the French-language and English-language versions of Wikipedia and Wikidata. Analysis of this data will bring to light some evidence-based findings about biographical coverage in these resources while enabling us to update the interpretation of the gender gap in reference works. This quantification of biographical articles is likely to contribute to a better understanding of the gender gap and a greater awareness of gender inequality and systemic sexism in the documentary world in particular and society in general.

## The Question of Gender Diversity

Concerns were raised early on within the Wikipedia community about social diversity in the Wikipedia project, which manifests itself in an imbalance in the coverage of specific topics. Participants in the WikiProject Countering Systemic Bias, launched in 2004, explicitly addressed the ways in which the interests and sociodemographic profile of editors could influence content and lead to bias. Notably, the introductory text written at the beginning of the project remains practically unchanged to this day:



https://doi.org/10.15215/aupress/9781771994521.01

> The Wikipedia project contains several types of WP:NPOV violations that arise from systemic bias in the demographics of the editor community. Encyclopedic coverage is imbalanced and often omits points of view from under-represented demographic groups. Systemic bias on Wikipedia may take the form of gender, geographical, racial, ideological, and other forms of bias. ("Wikipedia: WikiProject Countering Systemic Bias," 2025)

One of the subgroups of the WikiProject Countering Systemic Bias, the Gender Gap Task Force, has been active since 2013, whereas the WikiProject Gender Studies subgroup has been active since 2005 and the WikiProject Women in Red since 2015. Through various forms of engagement, these projects aim to eliminate systemic gender bias. At least two of these projects have corresponding groups in the French-language version of Wikipedia.[1]

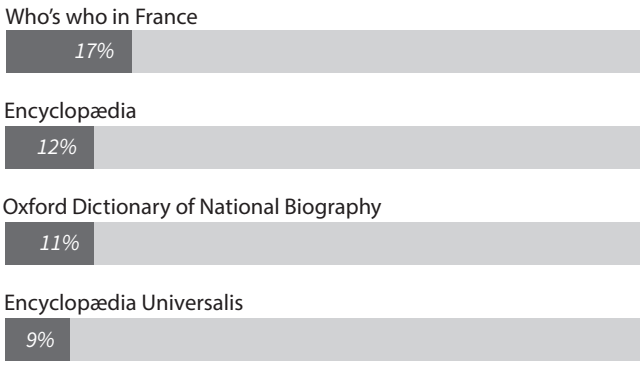
In an article published in 2020, Pierre-Yves Beaudouin, then administrator of Wikimédia France, reported that the French-language version of Wikipedia had 591,491 biographies, 18% of which were devoted to women (Beaudouin, 2020; see figure 3.1, which reproduces the one in the text). He pointed out that information on several dictionaries and encyclopedias whose contents were harmonized with Wikipedia and Wikidata (the database that feeds Wikipedia) makes it possible to present statistics on the proportion of biographies dedicated to women. Some examples show that Wikipedia's performance in terms of gender bias compares favourably in some cases with that of other biographical resources, but the author explained that all these resources, including Wikipedia, are shaped by the inequalities that exist in society. He maintained that one of the strengths of Wikipedia, and of Wikidata in particular, lies in the technological opportunity to quantify the biographies of several resources by producing statistics on gender gaps in biographical content.

The figures for the English-language version of Wikipedia are comparable, although slightly higher: In June 2024, 398,298 biographies were devoted to women, representing 19.8% of the 2,006,964 biographies in this encyclopedia

1 The Les sans pagEs project is the French-language version of WikiProject Women in Red; the Femmes project corresponds to the WikiProject Gender Studies. The projects WikiProject Countering Systemic Bias / Gender Gap Task Force and WikiProject Countering Systemic Bias have no equivalent in French.


https://doi.org/10.15215/aupress/9781771994521.01

**Figure 3.1.**
*Statistics on the proportion of biographies dedicated to women in different resources*




*Source:* Beaudouin, P.-Y. (2020, March). Wikimédia France. https://www.wikimedia.fr/biais-de-genre-wikipedia-aussi-imparfaite-que-la-societe/. CC BY-SA.

at that time on Humaniki, which calculates the percentage of genders among all the biographies in the various projects (Humaniki, n.d.).

Based on a sample of 80 resources, dictionaries, and encyclopedias, including Canadian works, we wish to show how Wikipedia and Wikidata can enrich our understanding of social issues such as gender inequality, despite the limitations of these knowledge resources in terms of gender bias. This quantitative exploration will be conducted by investigating the following questions:

1. How could Wikidata be used to quantify the content of biographical sources such as dictionaries and encyclopedias?
2. What new data can be produced on the proportion of biographies devoted to women within Wikipedia and other biographical resources, thanks to Wikidata?
3. What interpretations and comparisons can be made of these biographical resources by considering, more specifically, data relating to Canadian publications, and what does this quantification tell us about the gender gap in these reference works?

Our research follows on from work on Wikipedia content and focuses specifically on the content gap related to gender asymmetry in biographical


https://doi.org/10.15215/aupress/9781771994521.01

coverage of reference works. To our knowledge, no rigorous study on the comparative evaluation of biographical coverage of reference works has been undertaken since the work of Reagle and Rhue (2011). Furthermore, we propose a method that differs from the approach used by Reagle and Rhue, which relies primarily on the checklist method.[2] This leads researchers to compare the coverage in absolute terms of the different biographical sources—that is, the ratio and the respective total number of biographies of men and women. It is also used to compare the behaviour of Wikipedia and the *Encyclopædia Britannica* in relation to various preestablished lists of notable people.[3]

Our approach, following recent research (Klein & Konieczny, 2015; Konieczny & Klein, 2018), builds on Wikidata's ability to measure the gender gap in biographical sources that are harmonized with it. As we describe in the next section, this method allows us to assess the gender gap in a more extensive set of encyclopedias and dictionaries than Reagle and Rhue did. These works come from different countries, including Canada. This method quantifies the gap relatively more precisely by comparing these biographical sources with one another. Finally, we would like to emphasize that there is a scarcity of research in the French language in this sector, to which we are contributing here.

## A Methodology Based on Wikidata

*How can Wikidata be used to quantify the content of biographical sources (dictionaries or encyclopedias)?*

The proposed methodology provides an initial response to the questions raised in this research. To carry out this research, we focused on the gender associated with the subjects of biographical publications in dictionaries and encyclopedias with unique identifiers (UID)[4] that are attached to specific

2 The checklist is a mixed approach: Although classified as qualitative, it is a method whose data is primarily collected quantitatively and processed through descriptive and inferential statistical analysis (Johnson, 2018).

3 For example, Gale Biographical Resource Center, Wilson's *Current Biography Illustrated*, American National Biography Online, *Chambers Biographical Dictionary*, and *Le top 100 des personnes les plus influentes selon les magazines* The Atlantic et Time (Reagle & Rhue, 2011, pp. 1144–45).

4 For example, each entry in the *Canadian Encyclopedia* (CE) is associated with a unique identifier in the form of a word string. This identifier must be added to the



https://doi.org/10.15215/aupress/9781771994521.01

properties in the Wikidata free knowledge base.[5] As the content of Wikidata is published under Creative Commons Zero (CC0), it is easy to browse, sort, reuse, modify, and republish based on specific criteria using the SPARQL query language in particular.[6]

In June 2024, an initial query enabled us to identify 465 distinct properties linked to the UIDs of publications containing biographical articles, based on their classification using the Wikidata property element relating to encyclopedias (Q55452870) or its subcategories.[7] First we eliminated obvious false positives, duplicates, and gender-focused publications, then we selected publications with a minimum sample size of 1,000 entries, with a view to ensuring worldwide geographical coverage. This quantitative criterion seems to us to represent a certain threshold of documentary impact on the web as well as a statistically significant sample.

A second SPARQL query on Wikidata was applied to measure the biographical gender gap. Our final selection therefore includes 80 publications from all continents, for which we have carried out a quantitative evaluation of the gender of the biographies.

It should be noted that for the majority of the resources concerned, the sample associated with Wikidata that brings them together does not allow

---

end of the base URL—https://www.thecanadianencyclopedia.ca/en/article/ (English) or https://www.thecanadianencyclopedia.ca/fr/article/ (French)—to access a specific entry. For instance, the French-language article on Émile Nelligan can be accessed by adding the identifier *emile-nelligan*, resulting in the URL: https://www.thecanadianencyclopedia.ca/fr/article/emile-nelligan.

5 Wikidata's 125 million elements (Q), including more than 10 million dedicated to humans, are linked by one or more of its 10,000 properties (P) in the form of RDF triplets. Of these properties, more than half are dedicated to UIDs of external publications. As a result, according to the previous example, on Wikidata, property P5395 is dedicated to the unique identifiers of CE entries. This property therefore associates each notion in the CE with the equivalent element in the free knowledge base. For example, the CE entry Émile Nelligan is associated with Wikidata element Q2392492 (element dedicated to the poet) using the property P5395.

6 See, for example, the tens of thousands of items concerning female personalities who do not have a French-language Wikipedia article. These entries were identified during Women's History Month 2019 (see https://fr.wikipedia.org/wiki/Utilisateur:Simon_Villeneuve/femmes).

7 List of properties: https://w.wiki/37pf.



https://doi.org/10.15215/aupress/9781771994521.01

us to quantify the biographical gender gap accurately. However, monitoring the evolution of around 50 cases between March 2019 and April 2021 reveals a strong trend with no significant change ("Wikipédia: RAW/2019 03 01," 2025; "Wikipédia: RAW/2021 04 01," 2025). Lastly, we assume that the samples analyzed were not associated specifically based on a gender criterion.

The total number of biographical articles in each publication generally corresponds to the number given on the publication's official website. It should be added that while this information is usually available for publications specializing in biographies, this is not the case for general publications. For most of these, data were deduced by exploring the internal search engine of the sites concerned. Some other data were provided by people representing the publications after they were contacted by email.

The uncertainty interval takes into account only the associated sample and represents a maximum value with a confidence level of 95%. Thus, for samples corresponding to 100% of the content of a given publication, there is no uncertainty in the percentages. The results are presented in the following section.

The full content of the three wikis studied is available for analysis using other SPARQL queries and the Humaniki site. The figures given are accurate, but they are still provisional given that new biographical content is added every hour. As a result, the figures given here for the French- and English-language versions of Wikipedia are taken from the content from June 2024, while those for Wikidata come from the Qlever site using a version (or *dump*)[8] of Wikidata undertaken on June 6, 2024.[9]

8 A *dump* of Wikidata is a version exported from the Wikidata database at a given time and used for research or analysis.

9 See https://qlever.dev/wikidata/ehCgou for the query. At that time, around 19% of Wikidata's human elements had no registered gender. This therefore leads to uncertainty about the actual proportion of each gender on this site, although we have no reason to believe that the current sample is not representative of the whole. It should also be noted that over 40 gender identities are used on Wikidata.



https://doi.org/10.15215/aupress/9781771994521.01

## More Precise Data on Women's Biographies

*What new data can be produced using Wikidata on the proportion of biographies devoted to women within Wikipedia and other biographical resources?*

The empirical objective of our study is to evaluate the biographical gender gap by simultaneously considering the number of articles associated with individuals identified as female (F) and male (M) in the sample of 80 biographical resources, encyclopedias, and dictionaries selected. Table 3.1 presents the data for these resources, which total more than two million biographical articles (2,014,479 articles).

These initial results present statistics on the distribution of genders in these various biographical resources, highlighting the percentage of articles associated with individuals identified as F and M. On average, the sample contains 14% of biographical articles on F personalities and 85% on M personalities.

The two largest biographical resources, *Deutsche Biographie* (622,800 entries) and China Biographical Database Project (417,931 entries), bring together 1,040,732 articles, or close to 50% of the articles in the sample. However, the gender divide is relatively less pronounced in the case of these resources when compared to other resources and especially in the case of *Deutsche Biographie*, which presents 20.8% F biographies. If we exclude these two publications in order to have a more representative idea of all the publications analyzed, we obtain an even shallower gender divide, with 11% of biographies of F personalities and 88% of biographies of M personalities.

*What interpretations and comparisons can be made of these different biographical resources by looking more specifically at data relating to Canadian publications, and what does this quantification tell us about the gender gap in reference works?*

First, a few remarks about the data presented in the full table of reference works in the sample, which can be consulted on the online platform that accompanies the book. The encyclopedic data on biographies in Wikidata varies from one resource to another, as well as according to a development history ranging from manual data entry to the batch import of databases that are becoming increasingly widespread. For the *Canadian Encyclopedia* (CE) and the *Dictionary of Canadian Biography* (DCB), we conducted an initial content analysis in autumn 2018 (Villeneuve, 2025). The figures have changed



https://doi.org/10.15215/aupress/9781771994521.01

little since then, as the Wikidata-associated content of these publications has grown only slightly. The observed gender gap has thus varied by around 1% for both of these publications.

It should also be noted that we have been analyzing the biographical content of the *Encyclopædia Britannica* (EB) and the *Encyclopædia Universalis* (EU) associated with Wikidata since December 2017 (Villeneuve, 2025). The proportions given for these two publications are exact to the nearest percentage point. Note that the number of elements associated with EB identifiers is much greater than the number of biographies revealed by the people representing this publication. This is explained by the association on Wikidata of the EB URL redirections.[10]

Some databases show a greater disparity of genders, such as the *Biographical Directory of the United States Congress*, with only 3% F biographies, compared with 97% M biographies. This is also the case for the *Dizionario biografico degli Italiani*, which has a low proportion of F biographies, equivalent to 4%. Others offer a slightly more balanced representation, such as the *Dictionary of New Zealand Biography*, with 27% F biographies and 73% M biographies. Another example is the *American National Biography*, which has 16% articles on F personalities and 83% on M personalities, which is above the average for F biographies contained in the resources in the sample.

More than half of these publications (44 out of 80, or 55%) contain 10% or less F biographies. The *Dictionary of Canadian Biography*, with 6%, falls into this category. Slightly more than a third of biographical resources (25 out of 80) contain between 11% and 19.9% F biographies.

No resource has 50% F biographical articles, and only 10 have 20% or more (see table 3.1). BabelNet, *TDV İslam Ansiklopedisi*, and the *Dictionary of New Zealand Biography* have 45%, 37%, and 27% F biographies, respectively; these resources show biographical coverage that appears relatively the least unbalanced of the whole. The second Canadian resource in the sample, the CE, falls into this category, with 21% F biographies.

What of the three traditional biographical resources commonly appearing in lists of French and English institutional reference works? These resources are the *Encyclopædia Britannica*, the *Encyclopædia Universalis*,

10 For example, Gilles Vigneault has a unique identifier on *Britannica*, but this links to an article on French Canadian literature and not to a biographical article devoted to Vigneault.



https://doi.org/10.15215/aupress/9781771994521.01

**Table 3.1.**
*Ten resources with more than 20% F biographical articles with the corresponding proportion*

| | F biography | B biography |
|---|---|---|
| **Biographical resources (total)** | **14.4%** | **85.2%** |
| **BabelNet** | 44.7% | 55.1% |
| **Deutsche Biographie** | 20.8% | 79.3% |
| **Dictionary of New Zealand Biography** | 26.7% | 73.0% |
| **Encyclopaedia Itaù Cultural** | 22.6% | 75.9% |
| ***Encyclopedia of Fantasy*** | 21.2% | 78.0% |
| ***Canadian Encyclopedia*** | 21.4% | 78.5% |
| **La Documentation permanente** | 25.1% | 73.8% |
| **Notable Name Database** | 20.1% | 79.8% |
| **SIKART** | 19.9% | 73.2% |
| **TDV İslam Ansiklopedisi** | 36.7% | 63.2% |

*Note:* In this case, we have rounded off SIKART's data, which actually indicates 19.9% biographical F entries, to include it in the list of those 10 resources holding more than 20% biographical F articles.

and the *Encyclopédie Larousse*. The *Encyclopædia Britannica* (45,296 associated items) shows an underrepresentation of entries on F personalities (11%) compared with M personalities (89%). The *Encyclopædia Universalis* (17,589 items) also shows a strong gender disparity, with a low proportion of articles on F personalities (9%) and 91% on M personalities. The *Encyclopédie Larousse* (3,164 items) shows a similar proportion to the *Encyclopædia Universalis*, with 9% of biographical articles on F personalities and 91% on M personalities, highlighting a significant gender imbalance. More generally, these three publications have below-average results (14%) in terms of the percentage of F biographies.

Furthermore, the percentage indicating the gender gap in the two subclasses of the sample corresponding to French- and English-language biographical resources is relatively similar: 14% of biographies of F personalities for the former and 13% for the latter.



https://doi.org/10.15215/aupress/9781771994521.01

In the encyclopedic works in the sample, the three Wikimedia Foundation-hosted wikis analyzed are at the top of the average for the percentage of F biographical articles (14%). Wikidata comes out on top of the three with 21% F biographical content compared with Wikipedia in French and English (20%); this is lower, however, than BabelNet (45%), *TDV İslam Ansiklopedisi* (37%), and the *Dictionary of New Zealand Biography* (27%). It should also be noted that in absolute numbers, the biographical coverage of the French-language version of Wikipedia (712,906) is far greater than the coverage of each of the French biographical resources and even the most important of them—namely, *Who's Who in France* (24,513); likewise, the English-language version of Wikipedia (2,006,964) is far greater than the coverage of English resources, far exceeding *Britannica*, which has 45,296 entries.

Finally, we note that the average percentage of F biographies in the publications analyzed (14%) is close to the percentage of female participants in Wikimedia projects, which fluctuates between 11% and 15% (see table 3.2). In other words, it seems that the biographical gender gap in the publications analyzed is of the same order of magnitude as the gender gap for male and female contributors to Wikimedia projects.

## Some Food for Thought and Discussion

Much research on gender bias has focused on participation, some research has focused on content, and some has been undertaken to focus on both. In our study, we explored the gender gap in the content of a set of encyclopedic biographical resources. Unlike previous work on the biographical coverage of these dictionaries and encyclopedias, the gender gap was quantified using Wikidata, the online database that supports Wikipedia. Thanks to this method

**Table 3.2.**
*The biographical gender gap in Wikimedia resources (June 2024)*

| | F biography | M biography | Other genders |
|---|---|---|---|
| **Biographical resources (total)** | **12.6%** | **86.4%** | **Not available** |
| **en.wikipedia** | 19.8% | 80.0% | 0.13% |
| **fr.wikipedia** | 20.0% | 79.9% | 0.143% |
| **Wikidata** | 20.9% | 60.3% | 0.09% |


https://doi.org/10.15215/aupress/9781771994521.01

based on massive data, we were able to update the gender situation of 80 biographical resources, totalling more than two million articles, in a rigorous, efficient, and robust way. These are results that no previous method had been able to produce. Thus, as Reagle and Rhue (2011) had already acknowledged, until now, there had been no large-scale comparisons of biographical coverage and gender in reference works, including Wikipedia, in 2011. However, this statement remains valid, and this is precisely the gap that our research aims to fill. It should be noted that the work of Reagle and Rhue did, however, compare Wikipedia and Britannica with six other biographical resources covering a few thousand articles.

From a methodological point of view, the selection criterion we used for choosing the biographical resources—a minimum threshold of 1,000 entries—helped significantly reduce the sample to resources with the greatest documentary impact. It also ensured greater precision in the results, since the alignment of these resources with Wikidata was very pronounced in most cases. The number of biographical resources and articles involved in the sample thus made it possible to detect and better establish a trend that would be difficult to deny from now on regarding the gender gap and the presence of systemic gender bias at work in encyclopedic-type reference works. The main quantitative results that have highlighted these social gender inequalities in the field of encyclopedic knowledge fall into four categories:

1. *The gender gap in encyclopedic reference works.* Biographical coverage of gender shows a considerable underrepresentation of F personalities worldwide, with an average of 14% of biographical articles concerning female personalities. In addition, more than half of the resources have 10% or fewer F biographies. Finally, we note that almost 88% (87.5%) of encyclopedic resources have fewer than 20% F biographies.
2. *The gender gap in English- and French-language encyclopedic reference works.* On the whole, English- and French-language encyclopedic resources are comparable to the average. Among these resources, those that are most generally accessible—*Encyclopædia Britannica*, *Encyclopædia Universalis*, and *Larousse*—are below average, and two of them are below the 10% mark.
3. *The gender gap in Canadian encyclopedic reference works.* Among the 10 most favourably positioned resources, three are in English,



https://doi.org/10.15215/aupress/9781771994521.01

one is in French alone, and three others are multilingual or bilingual, such as the *Canadian Encyclopedia*. The latter stands out from the rest of the sample, with 21.4% F biographies. The *Dictionary of Canadian Biography*, which is also bilingual, is in the category of publications with the widest gender gap, with 6.1% F biographies.

4. *The Wikimedian gender gap.* The comparison between the biographical coverage of these various resources and the three Wikimedia Foundation wikis is to the advantage of the latter. The results confirm the consistency and extent of the gender gap in the encyclopedic content available to internet users.

The gender gap in encyclopedia resources is not a recent phenomenon. A number of authors have already written about gender bias in encyclopedias. The renowned archivist and feminist Mary Ritter Beard produced a study in 1942 entitled "A Study of the *Encyclopaedia Britannica* in Relation to Its Treatment of Women," in which she and her colleagues strongly emphasized that, in terms of participation in the production of reference works, women were a negligible quantity, unlike the biases and gaps identified in terms of content, which proved to be substantial. Another study undertaken by Gillian Thomas (1992) showed that the biographical treatment of women placed them in the position of extras in a history of human knowledge. Even Marie Curie is generally described as the wife of Pierre Curie. It is fair to say, particularly in the light of the results of this research, that the situation and practices in the field of encyclopedic works have changed little since the days of Beard or Thomas. They appear essentially unchanged to this day, until proven otherwise.

We may acknowledge that Wikipedia continues the tradition of the great encyclopedias that began with Diderot and D'Alembert in the 18th century and that it has not abandoned its ambitious aims and the desire to cover human knowledge exhaustively in an almost utopian project. But it should be noted that this online encyclopedia perpetuates certain gender biases that already existed in these previous works of knowledge. At first glance, these biases exist despite the principle of neutrality that is one of the pillars of the Wikipedia project.

The principle of neutrality is a relevant consideration that is regularly called into question in discussions about the gender gap (McDowell & Vetter, 2021). From the outset, it is important to make clear that the principle of neutrality does not refer to a principle of balance (for example, if there are


https://doi.org/10.15215/aupress/9781771994521.01

50% men and 50% women, there should be 50% biographies for each). The role played by the principle of neutrality in the gender divide is linked to its epistemic function. Like Wikipedia's five founding principles, which are "fundamental . . . and non-negotiable," the principle of neutrality constitutes an "intangible" foundation that applies to all articles ("Wikipédia: Neutralité de point de vue," 2025). According to the principle of neutrality, generally referred to by contributors as "NPOV" (neutral point of view),

> Articles should not promote any particular point of view. Sometimes this means mentioning several points of view and representing each of them as faithfully as possible, taking into account their respective importance in the field of knowledge. It also means providing the context needed to understand these points of view according to the sources that support them, and not representing any one point of view as being the truth or the best point of view. These conditions make it possible to verify information by citing sources that are authoritative on the subject (particularly in the case of controversial subjects). ("Wikipédia: Principes fondateurs," 2025)

This principle aims not to promote subjectivity and is fundamentally concerned with avoiding bias. To do this, this first rule relies on two further rules: verifiability and a ban on publishing unpublished works. These conditions jointly determine the encyclopedic content: "what can or cannot be published in Wikipedia." The first rule, verifiability, requires that for information to be used, it must be verifiable by readers using a "quality source or reference." The second rule, which prohibits unpublished work or personal research, specifically excludes "research that has never been published outside Wikipedia or that represents a 'revolution' not yet known or debated in the field, an opinion that is 'excessively' in the minority or that can only be associated with sources considered confidential and/or unreliable or, more simply, the personal interpretations, deductions or intuitions of the editor of the article" ("Wikipédia: Travaux inédits," 2025).

These three rules are also considered to be codependent: They "must be interpreted in relation to each other" in order to prescribe admissible content and therefore knowledge or what counts as such within Wikipedia ("Wikipédia: Vérifiabilité," 2026). However, these three rules, forming a barrier against subjectivity and bias, also tend to significantly disadvantage the admissibility of biographical articles about women for whom, in some cases,



https://doi.org/10.15215/aupress/9781771994521.01

verifiable secondary sources are not available or, in others, are insufficient; this further compromises opportunities for narrowing the gender gap. The paradox, which could be called the paradox of Wikipedian bias, stems from the fact that the mechanisms (NPOV, verifiability, no unpublished work) put in place to counter bias in individual contributory practices expose the encyclopedia and make it vulnerable to systemic bias, which cannot be countered even when participants are aware of it and intentionally wish to remedy it. From this point of view, this principle of neutrality seems to be more of an obstacle to the objective of reducing the gender gap.

Moreover, as Tkacz (2014) points out, this conception of neutrality also aims to protect the encyclopedia from struggles in the name of truth: "Neutrality . . . attempts to distance itself from the truth-battles of the outside world, that is, contests of truth that take place outside the Wikipedia formation." In this respect, Wikipedia differs from historical encyclopedias, which are not immune to the controversies of their times and whose authors assume an authoritative point of view that more often than not reflects a "relatively homogenous" view of the world (Reagle, 2013, quoted in Tkacz, 2014). This epistemic choice in favour of viewpoint neutrality was initially claimed to better support collaboration and build collaborative consensus:

> The whole concept of neutral point of view, as I originally envisioned it, was this idea of a social concept, for helping people get along: to avoid or sidestep a lot of philosophical debates. Someone who believes that truth is socially constructed, and somebody who believes that truth is a correspondence to the facts in reality, they can still work together. (Wales, quoted in Reagle, 2010, p. 53)

In this respect, as Tkacz suggests, it is not just particular battles over truth that will have been abandoned but the notion of truth itself.

Although Wikipedia has abandoned the truth of the outside world, the project has not been immune from battles and controversies that go far beyond the debates and deliberations on the construction of knowledge. Within the confines of its internal truth, battles are waged on a daily basis, and they have proven every bit as fierce as the battle from which it likely intended to escape in the first place.[11] Many of the outbursts concern biographies of

11 By "internal truth," we refer, following Tkacz, to the explicitness of point-of-view neutrality, which serves as a justification framework ("Wikipédia: Neutralité," 2025).


https://doi.org/10.15215/aupress/9781771994521.01

women whose verifiable secondary sources are contested. These controversies are not always resolved in favour of deliberation leading to a collaborative consensus—following from the idealized image suggested by Wales—to prescribe admissible content and therefore knowledge or what counts as such within Wikipedia (Ford, 2015). However, these three rules, forming a barrier against subjectivity and bias, also tend to significantly disadvantage the admissibility of biographical articles about women for whom, in some cases, verifiable secondary sources are not available or, in others, are insufficient; this further compromises opportunities for narrowing the gender gap.

Heather Ford (2015) has studied the dynamics at work in these cases. She has shown that the decision-making process operates in a rhizome, in the peripheries where the power of what will be represented in Wikipedia is held. These Wikipedians do not fit the media image of amateurs; they have specialized knowledge and skills. And those who "understand how to perform and speak according to Wikipedia's complex technical, symbolic and policy vocabulary tend to prevail" (Ford, 2015, p. 3). Several examples illustrate how these practices are not an advantage when it comes to controversies about women's biographies. Moreover, as Doutreix (2017) points out, the study of these controversies indicates that they work to the detriment of statements made by experts or activists.

From an intersectional perspective, it can also be maintained that the situation is aggravated when it comes to articles or information affecting women who combine several grounds of oppression, such as race, colour, or ethnic origin. This situation could also affect articles by women of French Québec or French Canadian origin whose content is questioned on the discussion pages by Wikipedians from the majority on en.wikipedia, given that this latter majority favours certain ideologies on gender and also certain sources to the detriment of others who are associated with "peripheral" francophone cultures or who do not have recognized status in the hierarchy of sources, according to the standards accepted by the majority. This hypothesis ties in with Heather Ford's (2017) findings that these Wikipedians also reaffirm "the already legitimated power of scientists and academic professionals on Wikipedia, but it also introduces new, unequal power relations among editors who do not share the same levels of expertise and access to what are considered credible knowledge sources."

In other words, the epistemic framework linked to neutrality prevents the production of articles about people identified as women for whom verifiable



https://doi.org/10.15215/aupress/9781771994521.01

secondary sources cannot be provided, but it also appears prejudicial in discussions about articles whose verifiable secondary sources could be considered sufficient but that are rejected by Wikipedians who hold power over representations and, consequently, over what is held to be true in Wikipedia.

Before concluding this discussion, we need to make a few more remarks about the limitations of the present study. The fact that the declaration of the Wikidata property element relating to encyclopedias (Q55452870) is voluntary is perhaps a limitation of the tool, since it is possible to categorize resources whose status as encyclopedic publications or reference works could be called into question. This is the case, for example, with the BabelNet aggregator or the Notable Name Database. However, these are the only cases in the sample where their status appears to be open to question.

Furthermore, the synchronic approach used could constitute another limitation of the strategy developed. The basic idea behind our analysis is to highlight the gender gap in terms of the content presented to internet users by the various publications and therefore the gap for which an effort has been made to digitize and make accessible online. It would be very interesting to analyze the data diachronically according to the historical periods of the personalities concerned, especially considering the hypothesis that the gender gap has narrowed in recent decades. The Humaniki tool, for example, makes this differentiation possible.

However, one of the most worrying limitations of our study concerns its bias in favour of a binary conception of gender. By focusing exclusively on the gender divide as a binary category (male/male and female/female), we have neglected the status of nonbinary identities. This conjecture contributes to the systemic bias, discrimination, and invisibilization experienced by people who identify as nonbinary or who do not conform to a binary conception (trans, intersex, agender, two-spirited, gender queer, fluid, etc.). For this reason, the very definition of the gender gap needs to be reevaluated and extended beyond the approach based on the Western binary conception of gender (Metilli & Paolini, 2021).

## What Do We Do Now?

We are unable to counteract systemic biases directly in a substantial and radical way because of the epistemic obstacles arising from the NPOV, so what are the possible options to mitigate the gender gap, relatively speaking? As


https://doi.org/10.15215/aupress/9781771994521.01

Beaudouin (2020) suggests, the intensive production of women's biographies that meet the verifiability criterion, which requires all information to be verifiable by a quality source or reference, is one way of reducing the gender gap. Supporting and taking part in initiatives such as Les sans pagEs and the Art + Feminism workshops, held across Québec and Canada, encourage the production of this content. Tens of thousands of articles remain to be created that meet these conditions.

In addition, it may be thought that an editorial project should be set up for quality biographical articles about women, who also present a gender asymmetry on Wikipedia.[12] However, for many female personalities, the sources needed to satisfy the verifiability are simply not available. Another option is to produce knowledge upstream, as suggested by Wikimedia France: "Award prizes, draw up portraits, write entries for bibliographic dictionaries, or publish research articles and monographs on little-known women. Wikipedia can then deal with the subject by synthesizing and disseminating knowledge" (Beaudouin, 2020).

From a regional perspective, another action is to increase the creation of links between Canadian, Québec, and French bibliographic resources and Wikidata. For example, the recent addition of 329 entries from the Japanese Canadian Artists Directory to Wikidata shows that 51.4% of the entries are about female personalities.[13]

We hope that a movement can be launched to link the contents of these reference works—as has already happened in the case of encyclopedias such as Britannica—so that we can better assess the gender gap that exists in the latter and also draw on it for material to contribute to Wikipedia. We are thinking particularly of a mobilization effort around the following Québec encyclopedias: *Biographies canadiennes-françaises*, the *Dictionnaire des auteurs de langue française en Amérique du Nord*, the *Encyclopédie du patrimoine culturel de l'Amérique française*, and those included in our research—that is, the *Dictionnaire biographique du Canada* and the *Encyclopédie canadienne*.

Our study leads us to the following conclusion: In the Wikipedian environment and beyond, the problem is not fundamentally related to the size of the gender gap, whether in terms of biographical coverage or editorial

12 In the French edition of Wikipedia, there are currently 294 M biographical articles and 53 M biographical articles; https://w.wiki/3pYE.

13 The statistics may be accessed by making the following query: https://w.wiki/AQvk.



https://doi.org/10.15215/aupress/9781771994521.01

participation, but rather in the obvious imbalance linked to systemic sexism. As Jemielniak (2016) suggests, Wikipedia reflects the prejudices and biases present in society; sexism, accentuated by technoculture, proves to be one of the most powerful levers. Research findings such as those presented here can contribute to rendering the gender gap and systemic sexism more explicit and also making them better understood. However, as other studies have pointed out, there is now an urgent and conscious need to look at gender bias in relation to other social biases, using a nonbinary and intersectional approach, as part of future research with the help of new tools but also at inclusive Wikimedia policies in order to reduce epistemic injustices from a broader and more fundamental perspective.

## Acknowledgements

We would like to thank Michèle Lefebvre, librarian at Bibliothèque et Archives nationales du Québec (BAnQ), for her help and insights into Canadian reference works. We would also like to thank Wikimedians TomTom and Nicolas Vigneron, who developed several tools for analyzing wiki elements and articles, and mathematician Maxime André for his help in processing certain data. Finally, we would like to thank the Wikipedian Cantons-de-l'Est and Pierre-Yves Beaudouin for reprinting and disseminating previous work.

## References

Beard, M., Edinger, D., Selig, J. A., & White, M. (1977). A study of the *Encyclopaedia Britannica* in relation to its treatment of women. In A. J. Lane (Ed.), *Mary Ritter Beard: A sourcebook* (pp. 215–24). Schocken Books.

Beaudouin, P.-Y. (2020). Biais de genre: Wikipédia aussi imparfaite que la société. In *Wikimédia France*. https://www.wikimedia.fr/biais-de-genre-wikipedia-aussi-imparfaite-que-la-societe/

Cohen, N. (2011, January 30). Wikipedia ponders its gender-skewed contributions. *New York Times*. https://www.nytimes.com/2011/01/31/business/media/31link.html

Doutreix, M.-N. (2017). Quel dialogisme dans l'écriture encyclopédique de Wikipédia? In M.-D. Popelard (Ed.), *La reprise en actes* (pp. 149–59). Presses universitaires de Rennes. https://doi.org/10.4000/books.pur.183897



https://doi.org/10.15215/aupress/9781771994521.01

Ford, H. (2015). *Fact factories: Wikipedia and the power to represent*. Kellogg College, University of Oxford. https://doi.org/10.13140/RG.2.1.4068.9361

Gender gap. (2025, November 18). In *Meta-Wikimedia*. https://meta.wikimedia.org/w/index.php?title=Gender_gap&oldid=29668667

Humaniki. (n.d.). *Humaniki: Wikimedia diversity data tool*. Retrieved January 30, 2026, from https://humaniki.wmcloud.org/

Jemielniak, D. (2016). Breaking the glass ceiling on Wikipedia. *Feminist Review*, *113*(1), 103–8. https://doi.org/10.1057/fr.2016.9

Johnson, P. (2018). *Fundamentals of collection development and management* (4th ed.). ALA Editions.

Klein, M., & Konieczny, P. (2015). Gender gap through time and space: A journey through Wikipedia biographies and the "WIGI" index. *arXiv*. https://doi.org/10.48550/arXiv.1502.03086

Konieczny, P., & Klein, M. (2018). Gender gap through time and space: A journey through Wikipedia biographies via the Wikidata human gender indicator. *New Media & Society*, *20*(12), 4608–33. https://doi.org/10.1177/1461444818779080

McDowell, Z. J., & Vetter, M. A. (2021). *Wikipedia and the representation of reality*. Routledge.

Melançon, B. (2018). Les vertus utopiques de l'encylopédisme. *Sens public*. https://doi.org/10.7202/1059033ar

Metilli, D., & Paolini, C. (2021). *Non-binary gender identities in Wikidata* [Video]. YouTube. WikidataCon 2021. https://www.youtube.com/watch?v=lIsmFUuGvCw

Nadeau, J.-F. (2017, March 4). L'histoire invisible des femmes. *Le Devoir*. https://www.ledevoir.com/societe/493173/les-grandes-oubliees-l-histoire-invisible-des-femmes

Reagle, J. M. (2010). *Good faith collaboration: The culture of Wikipedia*. MIT Press.

Reagle, J. (2013). "Free as in sexist?" Open culture and the gender gap. *First Monday*, *18*(1). https://doi.org/10.5210/fm.v18i1.4291

Reagle, J., & Rhue, L. (2011). Gender bias in Wikipedia and *Britannica*. *International Journal of Communication*, *5*, 21. https://ijoc.org/index.php/ijoc/article/view/777

Rey, A. (2007). *Miroirs du monde: Une histoire de l'encyclopédisme*. Fayard.

Thomas, Gillian. (1992). *A position to command respect: Women and the eleventh Britannica*. Scarecrow Press.

Tkacz, N. (2014). *Wikipedia and the politics of openness*. University of Chicago Press.

Villeneuve, S. (2025, November 30). PoV. In *Wikipédia*. https://fr.wikipedia.org/w/index.php?title=Utilisateur:Simon_Villeneuve/PoV&oldid=231048848

Wikimedia Foundation. (2023). *Community insights 2023 report*. Wikimedia Foundation. https://meta.wikimedia.org/wiki/Community_Insights/Community_Insights_2023_Report



https://doi.org/10.15215/aupress/9781771994521.01

Wikipédia: Neutralité de point de vue. (2025, October 11). In *Wikipédia*. https://fr.wikipedia.org/w/index.php?title=Wikip%C3%A9dia:Neutralit%C3%A9_de_point_de_vue&oldid=229682203

Wikipédia: Principes fondateurs. (2025, October 14). In *Wikipédia*. https://fr.wikipedia.org/w/index.php?title=Wikip%C3%A9dia:Principes_fondateurs&oldid=229755908

Wikipédia: RAW/2019 03 01. Regards sur l'actualité du mouvement Wikimedia. (2025, March 5). In *Wikipédia*. https://fr.wikipedia.org/w/index.php?title=Wikip%C3%A9dia:RAW/2019-03-01&oldid=223624218

Wikipédia: RAW/2021 04 01. Regards sur l'actualité du mouvement Wikimedia. (2025, March 6). In *Wikipédia*. https://fr.wikipedia.org/w/index.php?title=Wikip%C3%A9dia:RAW/2021-04-01&oldid=223624300

Wikipédia: Travaux inédits. (2025, October 16). In *Wikipédia*. https://fr.wikipedia.org/w/index.php?title=Wikip%C3%A9dia:Travaux_in%C3%A9dits&oldid=229815221

Wikipédia: Vérifiabilité. (2026, January 10). In *Wikipédia*. https://fr.wikipedia.org/w/index.php?title=Wikip%C3%A9dia:V%C3%A9rifiabilit%C3%A9&oldid=232315309

Wikipedia: WikiProject Countering systemic bias. (2025, November 25). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Wikipedia:WikiProject_Countering_systemic_bias&oldid=1324167268



https://doi.org/10.15215/aupress/9781771994521.01

*This page intentionally left blank.*

Focus II

# Institutions

*This page intentionally left blank.*

# 4 Wikidata in Canada and the Mariposa Folk Festival Linked Data Project

Stacy Allison-Cassin

Creating opportunities for data availability, reuse, and transparency and ensuring the accessibility of publicly funded data are considered essential hallmarks of an equitable civil society. Researchers, organizations, and institutions beyond government are encouraged to consider making data more available. In Canada, numerous levels of government have, for many years, been concerned with increasing the availability and awareness of the relationship between open data and an innovative, knowledgeable, engaged, and just society. The Government of Canada defines open data as "structured data that is machine-readable, freely shared, used and built on without restrictions" and cites benefits as innovation, informed decision-making by consumers, the leveraging of public sector information, and increased government accountability (Treasury Board of Canada Secretariat, n.d.). The open data definition includes several key criteria: structured, machine readable, and freely available without restrictions—elements not typical for proprietary commercial data where data is typically not shared for reuse outside the bounds of an organization.

Libraries, archives, museums, galleries, cultural centres, and other organizations related to culture, such as music and historical projects, also have mandates to impact society within Canada positively. Because many of these organizations typically have needs related to the management of information and data, particularly around the management of material culture collections, they are also highly involved in creating and maintaining structured data. In



https://doi.org/10.15215/aupress/9781771994521.01

the case of libraries, archives, and sometimes other organizations, this data is machine readable and created, stored, and exchanged according to local, national, and international standards. These organizations also typically have an interest in the subjects of their organization, whether it be making collections accessible in the case of archives or increasing the impact of research in the case of a university.

Sharing, enhancing, or making available structured data is an excellent way to meet these organizational and broader societal goals. As with the case of open data from the government, sharing data openly also aids in making society more just and equitable and increases the impact of culture in Canada and beyond. A growing focus on artificial intelligence (AI) and new modes of search and discovery point to the need for large-scale datasets of highly structured data to support continued machine learning and processing innovation. Despite the benefits of creating and using structured data and making data available openly, many libraries, archives, and cultural organizations are frequently challenged by a lack of technical resources to make their data openly available (Allison-Cassin & Scott, 2018). Further, many systems used to manage collections within organizations are not available on the open web, restricting the ability to make information related to collections and culture creators visible and reusable. Cultural data is frequently not open data (Zhu et al., 2023).

My journey into Wikipedia and Wikidata began out of necessity. In 2015, I was appointed as the W. P. Scott Chair in E-Librarianship and began a research project focused on using linked data technologies to better describe music materials (*Chair for Research*, n.d.). The project's focus was materials related to the Mariposa Folk Festival (Allison-Cassin et al., 2015; Proffitt, 2018). The Mariposa Folk Festival is one of the longest-running folk festivals in North America, and the festival's archival holdings are kept at York University's Clara Thomas Archives and Special Collections (St. Onge, 2013). My original research plan was to mobilize the information in the festival programs, such as performers, venues, event dates, and festival organizers, as linked data, with the aim of testing enriched descriptions and exposing relationships within the festival. Standard methods of library description for music do not handle nonclassical music well, and focusing on folk music in Canada seemed to be an ideal case study to test concepts related to moving beyond a focus on musical works to the network of relationships within a cultural scene (Allison-Cassin, 2012).



https://doi.org/10.15215/aupress/9781771994521.01

I quickly ran into several challenges within the project, the first being that the creation and publishing of linked data without adequate access to resources is very difficult, and my plan to integrate content from Wikipedia into my project, such as biographies of musicians, fell short when I discovered that very few of the performers and venues connected to the Mariposa Folk Festival and the revival scene in Toronto were covered in Wikipedia. I began to write Wikipedia articles on figures and venues connected to the folk revival scene in Toronto and Canada, such as the Riverboat Coffee House ("Riverboat Coffee House," 2024). Attempting to fill these gaps led me to become more involved in supporting and using the platforms.

In this chapter, I discuss several projects, initiatives, and research involving Wikidata in Canada within the domain of libraries, archives, and other similar cultural data applications. Using these examples, I discuss opportunities and critical issues for the continued use and development of Wikidata as an openly available structured data knowledge base in Canada. As an active user and member of the Wikidata community, I will take a personal approach to this discussion, focusing on several themes and examples. For a more systematic approach to Wikidata, see Zhao (2023) and Tharani (2021) for research articles. I will begin the discussion with my beginnings within Wikidata with the Music in Canada @ 150 Project, and subsequent sections will move through groupings of kinds of activities.

Ultimately, while Wikidata can be used to support activities related to cultural data and as a mechanism for making structured data openly available, there are critical issues within the Wikidata platform and its relationship with the wider ecosystem that need to be considered when committing resources in this area. However, at the same time, there is growing recognition of the need for ethical practice within the structured data landscape to support a wide swath of concerns such as Indigenous data, privacy, and discrimination. Wikidata has provided an avenue for many libraries, archives, and other organizations to engage in projects to make information on topics related to Canada more visible, but there are critical issues such as lack of support for Indigenous data, weakness in the availability of data and information in Canada, and issues with the fundamental structure of the project itself that make it unlikely to be a dependable source for open data. However, given the role of structured data within the current and potential future AI landscape, Wikidata may be an important player in the space.


https://doi.org/10.15215/aupress/9781771994521.01

## Background

Wikidata, a Wikimedia project launched in 2012, is rapidly evolving into a significant source of online, free, openly licensed structured data. As its name suggests, Wikidata is a project that creates and supports the development of openly accessible, semantically structured data that anyone can contribute to. This collaborative effort, akin to Wikipedia, has a profound impact on the information and knowledge space, underscoring the significance of each contribution. The freely available, openly licensed, multilingual platform of Wikidata offers individuals and organizations the opportunity to shape structured data. In Canada, Wikidata has found a place within the cultural data space, aligning with organizational goals relating to structured open data (Allison-Cassin & Scott, 2018).

Wikidata, launched in 2012, is "a free, collaborative, multilingual, secondary database, collecting structured data to provide support for Wikipedia, Wikimedia Commons, the other wikis of the Wikimedia movement, and to anyone in the world" ("Wikidata: Introduction," 2022). Wikidata is a knowledge base, a repository of information structured to be read and processed by machines. Structured data is information (data) that is highly organized and easily used by machine processing—for example, addresses or birth dates that are regularized across a database. Each element within the dataset is defined and used the same way. Unstructured data is much harder for machine applications to use and is difficult to reuse. Wikidata is already integrated into applications and uses beyond the Wikimedia projects. Unlike Wikipedia, Wikidata is valued and important for these alternate "beyond the projects" uses, making it unique among the Wikimedia projects.

Wikidata was initially developed to solve a technical problem on Wikipedia. Wikipedia has numerous unique language versions, and "the first goal of Wikidata is to support Wikipedia with centralized language links and infobox data, thus reducing the workload for Wikipedia editors and at the same time increasing the content's consistency and quality" (Vrandečić, 2013, p. 90). Before the creation of Wikidata, there were no connections between Wikipedias in different languages, making it difficult, for example, to move from the article on musician and poet Leonard Cohen in French to the article on Leonard Cohen in English or German. Wikidata performs an intermediary role by acting as the connector, ensuring all the articles on Leonard Cohen, no matter the language, are linked. With Wikidata, instead "of the articles linking



https://doi.org/10.15215/aupress/9781771994521.01

to each other, Wikidata maintains lists of all articles about a certain topic in the different language editions. Whenever an article is rendered, the software queries Wikidata for that list and displays it. The Wikipedias are completely relieved of maintaining these lists" (Vrandečić, 2013, p. 91). The creation of these connections highly improved the functionality of Wikipedia.

While its original role was supporting important technical structures of Wikipedia, the platform quickly began to support other kinds of structured data work. Wikipedia is text based and focused on the narrative form, making it much more usable from a human perspective; Wikidata is friendly to machine processing. Like Wikipedia, "Wikidata is also based on a community-editing model; it harnesses the distributed efforts of a worldwide community of contributors, including domain experts and bot developers. Anyone can add new statements, ranging from individual facts to large-scale data imports" (Waagmeester et al., 2020). However, including larger-scale editing of Wikidata through uploading larger datasets is tightly connected to its utility within technical projects. This utility is further amplified by its licence structure. Unlike many commercial proprietary databases, "the data in Wikidata is published under the Creative Commons CC-0 license, allowing anyone to reuse and republish the data in any way. The data is free of charge and without any conditions and requirements" (Vrandečić, 2013, p. 90). These factors make the enormous datastore attractive to consume within other applications. The size and licence are critical to its success, including within the cultural heritage and education sectors.

## Wikidata and the GLAM Community

The cultural heritage and broader culture organizations, often using the shorthand GLAM (galleries, libraries, archives, and museums), have been active in Wikidata from an early period (Tharani, 2021). When Wikidata was launched, many GLAM organizations were already active in Wikipedia, Wikimedia Commons, and other Wikimedia projects, and they leveraged campaigns such as Art + Feminism as collaborative outreach initiatives (Proffitt, 2018). GLAM organizations in Canada have also been engaged in the use of Wikipedia for outreach activities. For example, the Art Gallery of Ontario hosted edit-a-thons as far back as 2015 (Art Gallery of Ontario, 2015). Because many GLAM organizations have strong connections to metadata through collection management needs and staff with high levels of knowledge and experience


https://doi.org/10.15215/aupress/9781771994521.01

with metadata practices, the barrier to entry for Wikidata was lower than for other communities with less technological aptitude and resources. Factors such as responsive vocabulary development, an easy-to-use interface, and the potential to integrate Wikidata into library catalogues are cited as some of the positive aspects of Wikidata (Allison-Cassin & Scott, 2018; ARL, 2019).

Wikidata has emerged as a tool for creating linked data and enhancing access to collections within the GLAM sector (Allison-Cassin & Scott, 2018; Ansovini et al., 2022; Tharani, 2021), but questions remain regarding its viability as a reliable part of the sector's ecosystem because of issues such as data quality, lack of cohesion in the data model, and underlying problems related to data and information availability in Canada. GLAM and other cultural sector domains within Canada use Wikidata as part of their efforts to manage and make their collections visible. Wikidata is a knowledge base used by larger machine-processing entities such as Google and the growing AI tools for enabling contextual search and connecting information regarding the location of related connections, which can aid in visibility (Gertner et al., 2023). While creating these links to authority files or other data sources serves a purpose in supporting data and information needs and processes on Wikipedia and across the Wikimedia projects, Wikidata also supports external-use cases in the GLAM sector such as authority control (Bianchini et al., 2021; van Veen, 2019), extending the impact of research and researcher profiles (Nielsen et al., 2017; Odell et al., 2022), and enriching collections information (Ansovini et al., 2022; Colla et al., 2021; Hawkins, 2022).

Wikidata plays a crucial role in connecting Canadian-focused content, thereby improving the visibility of cultural resources both within Canada and beyond. For instance, Wikidata connects the item for author, poet, and actress Pauline Johnson to her item in library catalogues and information resources, driving the creation of a Google knowledge card. Both Wikipedia and Wikidata feed Google's knowledge graph, but the structured, machine-readable data in Wikidata makes this process smoother and also allows Wikidata editors the opportunity to more effectively enhance the connections in the knowledge graph with links to other information sources. As a metadata platform offering opportunities to shift information back and forth between collection databases, the ability to amplify collections and materials and to provide strong interlinking between data sites leading to greater impacts and enhancements make it a more effective tool for the movement of data, thereby strengthening the connection to our cultural heritage.



https://doi.org/10.15215/aupress/9781771994521.01

## Wikidata and the Mariposa Folk Festival Data

As stated in the opening of this chapter, I began my work with Wikidata while working on a project related to my position as the W. P. Scott Chair in E-Librarianship. The Mariposa Folk Festival linked data project aimed to model a network of relationships among the festival performers based on a dataset developed from the programs. The process for the creation of the dataset began with the development of an ideal data model based on the categories of elements of the programs and then consideration of likely valuable attributes for each category. The initial data model included categories such as musical performers, artisans, dancers, venues, dates, and the festival administrative staff. A set of attributes was developed for each of these different categories of data; for performers, this included properties such as name, music genre, medium the artist worked in, and the date they played at the festival. Google Sheets was used to record data in a separate spreadsheet for each grouping. For example, a sheet for musicians is separate from a sheet for bands. Because this was a linked data project and a desire to enable interlinking between the Mariposa dataset and open data available on the web, sources of persistent identifiers (PIDs) were included.

Sources of PIDs included the Virtual International Authority File (VIAF), a source of linked information related to library authority data, and music-specific identifier sources such as MusicBrainz and Discogs. Through their Linked Digital Futures initiative, the Canadian Association for the Performing Arts (CAPACOA) has focused on creating structured data to support the visibility of artists, venues, and performing arts events in Canada. They have pointed to the necessity of increasing the availability of linked data and, by extension, uniform resource identifiers (URIs) as part of an urgent need for sound digital strategies for Canadian performing arts. A lack of a strong, open, stable metadata culture within Canada weakens the overall health of sectors like culture and the performing arts, as "performing arts metadata has emerged as one of the most pressing issues for the performing arts sectors in Canada" (CAPACOA, n.d.).

Along with developing the data related to the Mariposa Folk Festival, I investigated available methods to transform the spreadsheet data into linked data and make the data available on the web. Lacking accessible software, tools, and resources, including access to software developers to create linked data, I turned to Wikidata as a method to create and publish data related to


https://doi.org/10.15215/aupress/9781771994521.01

the festival. The excellent integration between software tools such as Google Sheets, OpenRefine, and Wikidata made uploading data to Wikidata relatively simple. Furthermore, I could do all the work myself because Wikidata is free to use, openly available, and supported by an active community. Publishing the data using Wikidata provided a helpful solution and surfaced several issues of interest to using Wikidata in the Canadian context.

## Lack of Coverage of Canadian Music

The Mariposa Folk Festival is one of Canada's primary folk music festivals, and performing at the festival indicates recognition of notability. However, finding information to fill out the data on individual performers, bands, and other figures proved to be challenging. While there is Wikipedia, encyclopedias, and other notable sources available for well-known performers such as Gordon Lightfoot and Joni Mitchell, the same could not be said for the bulk of the performers, particularly as Mariposa moved away from more prominent mainstream performers. Biographical information in the programs was insufficient to fill in the desired data. Typical reference sources such as the *Encyclopedia of Music in Canada*, which was folded into the *Encyclopedia of Canada* in 2003, continue to be weighted toward classical music established by jazz musicians. While there has been an attempt to expand coverage to popular music, the encyclopedia has much catching up to do to cast off the prejudices of its original editors of the music-specific publication over what constituted noteworthy and important music.

National and local newspapers with a regular music review column or section can often be a good source for coverage of performers. However, searching through newspaper sources did not yield good results. For example, although the Toronto-based band the Dirty Shames appears in the Mariposa programs and photographs in the Toronto Telegram (McFadden et al., 1966), little information about the band leaves the Wikidata item bare ("The Dirty Shames," n.d.). Some information on musicians in the Mariposa dataset was found in digital newspaper archives; however, a subscription is required for newspapers such as the *Toronto Star* or *The Globe and Mail*. The "locked down" nature of these resources means they are unavailable to the public, and the data is not open to the web. Other information was not available in digital form or could not be found. Smaller newspapers and publications can be critical for coverage of local music scenes, but the archives of these



https://doi.org/10.15215/aupress/9781771994521.01

publications may not be available, and many of these publications are no longer in operation, with their websites no longer available.

Local media, such as newspapers and local newscasts, is rapidly shrinking in Canada, with CBC News reporting the closure of 70 community newspapers across Ontario in 2023. The lack of community coverage and smaller local media doesn't just mean that local people will not be informed of issues; it means a critical loss of the Canadian cultural record and serious consequences on the ability of those working on the Wikimedia projects to document Canadian culture (CBC, 2023). At the same time, digital independent community newspapers, such as Toronto's *West End Phoenix*, are reconfiguring news coverage (*About West End Phoenix*, n.d.).

The lack of information about Canadian music in digital form not only is a challenge when researching individual items but also makes it challenging to ensure that data is reliable and trustworthy. References for statements were a good practice but not necessarily needed to ensure inclusion in the Wikidata knowledge base in the early years of Wikidata. The data for the Mariposa Folk Festival added to Wikidata as part of the project outlined here should have taken the inclusion of references, therefore, as an essential aspect of the work. However, since this original project, including references has become essential to demonstrating notability. Provenance metadata is a crucial aspect of the Wikidata platform. It acts like a citation: "References are used to point to specific sources that back up the data provided in a statement" ("Wikidata: Introduction," 2022). A statement suggesting Céline Dion's birthplace is Charlemagne is supported by references to reliable sources where the information can be verified. Preferred sources, as with Wikipedia, are secondary and tertiary "high-quality" and reliable sources. Including references in Wikidata statements increases the quality of linked data for many use cases and can also assist with providing higher-quality data for AI applications, and a lack of references can have implications for the usefulness of the data (Beghaeiraveri et al., 2024).

The lack of readily available references for statements about musicians in the Mariposa Folk Festival has a compounding impact and points to a broader problem with the availability of information on music and the arts more generally. In subsequent projects I have deployed using Wikidata, such as documenting literature, built heritage, and film, the inability to provide references and provenance for Canadian data on Wikidata presents a serious challenge. As has been noted for Wikipedia, reliable and established sources


https://doi.org/10.15215/aupress/9781771994521.01

are required to prove notability, and lack of documentation is a known issue (McCracken, 2018). This can make it difficult not only to write articles but also to ensure that such articles are not deleted. The connection between the need to reference statements and notability makes it harder for data related to Canadian items to be created, or they may be at risk of deletion. Broadly, the lack of documentation related to music activities in Canada makes it difficult to build comprehensive, well-referenced datasets related to music within Wikidata.

## Wikidata and the Music in Canada @ 150 Project

In response to the lack of information in Wikipedia and Wikidata on Canadian music, I began the Music in Canada @ 150 Wikimedia campaign, an effort to increase content and recognition of the lack of information about Canadian music in a wider community. The campaign was developed to focus efforts across Canada on creating content on music in Canada (Allison-Cassin & Scott, 2018). As a member of the Canadian Association of Music Libraries and Documentation Centre, it made sense to tap into the music library community in Canada, as there was a ready network of individuals interested and knowledgeable about both music and information.

The effort was initially conceived to tie into activities related to the 150th anniversary of the Confederation of Canada; to align with the conference theme of the annual meeting of the Canadian Traditional Music Society, the Canadian Association of Music Libraries, Archives and Documentation Centres, and the International Association for the Study of Popular Music, Canada Branch; and to receive funding from Wikimedia and a special event fund at York University. The campaign featured an in-person workshop at the University of Toronto, two virtual workshops, and numerous distributed events over a year. We used the Wikimedia events dashboard, and the effort included 11 programs and 124 editors and featured the creation of 24 new articles and 187 article edits. Events took place at the University of Prince Edward Island, Charlottetown; Memorial University, Newfoundland; Western University, London; Hamilton Public Library, Hamilton; York University, Toronto; University of Toronto, Toronto; Laurentian University, Sudbury; University of Manitoba, Winnipeg; University of Saskatchewan, Saskatoon; and MacEwan University, Edmonton. Organizers included Dan Scott (Laurentian), Caroline



https://doi.org/10.15215/aupress/9781771994521.01

Doi (University of Saskatchewan), and Monica Fazekas (Western). A logo was used throughout the campaign.

The campaign's kick-off was a full-day workshop at the University of Toronto. Because the workshop was colocated with the annual meeting of several music-related professional associations, the workshop was attended by music librarians and music researchers. The workshop was structured as a "train the trainers" session, with instruction on editing both Wikipedia and Wikidata and discussions on how to run an edit-a-thon. It was intended to assist those holding their edit-a-thons during the year of the campaign. The event featured a panel session with local individuals such as Amy Furness; Rosamond Ivey, special collections archivist and head of library and archives at the Art Gallery of Ontario; and John Dupuis, science librarian at York University, with experience holding campaign-associated edit-a-thon events such as Art + Feminism and Ada Lovelace Day. The workshop was a critical mechanism for participants to gain skills and experience and engage in discussions on some of the previous topics on the availability of information resources on Canadian music. A lively topic of conversation was the question of "why Wikipedia" versus an institutional website or other publication, which led to discussions on open access and open licensing.

During these events in 2016 and through 2017, Wikidata was still relatively unknown in Canadian GLAM organizations, and there were few available workshop materials. Dan Scott, a librarian at Laurentian University, was instrumental in creating and leading Wikidata workshops and creating instructional materials (2017a). Building on this work and concerning the Mariposa Folk Festival project, Scott and I worked on modelling a means of capturing music festival data using Wikidata. Scott focused on the Northern Light Festival Boréal, which focused on bilingual considerations (2017b). The collaborative work on Wikidata, starting with the Music in Canada @ 150 Project, provided an early engagement with the collaborative creation of music-related data and became a basis for future collaborative efforts between Allison and Scott, including a workshop at the Ontario Library Association conference on adding libraries to the Wikimedia projects and a workshop at the Semantic Web in Libraries conference on Wikibase. The campaign also spawned the creation of the GLAM-Wiki Toronto conference held in 2019, which attracted over 100 attendees from GLAM organizations across the Toronto area.


https://doi.org/10.15215/aupress/9781771994521.01

The Music in Canada @ 150 Project was formed out of gaps I perceived while working on the Mariposa Folk Festival data project. The campaign largely succeeded in mobilizing individuals in music libraries interested in critical issues related to music information. The work within the music library communities demonstrated a means of creating opportunities for Wikidata development within an already existing community of individuals interested in finding mechanisms for supporting the visibility of music in Canada. Scott succinctly summed up the efforts:

> Our central argument was that, rather than focusing on directly enhancing our own local data repository silos (for example, library catalogues, digital exhibits), libraries and archives should invest their limited resources in enriching Wikidata, a centralized data repository, to maximize the visibility of those entities and the reusability of that data in the world at large . . . and then pull that data back into our local repositories to enrich our displays and integration with the broader world of data. (Scott, 2017b)

However, assessing the long-term impacts of the campaigns on the usage of Wikipedia and Wikidata in relation to music-related content is challenging, and we have not followed up with participants. Certainly, running editing campaigns will not solve the issue of the lack of secondary and tertiary sources on Canadian music.

As with Wikipedia, Wikidata is a volunteer-organized project, and the properties available to be used are user-generated by consensus. While this can be a great advantage in providing flexibility to data creators, it does mean that data is less standardized than typical systems using ontologies such as the CIDOC Conceptual Reference Model (CRM) or standardized descriptive systems found in libraries. Wikidata has no strict hierarchical structure; for example, broader and narrower concepts are not strictly observed. An additional known problem is related to the multilingual nature of Wikidata. Language is never a one-to-one translation between languages, and concepts are complex to capture. Increasing the availability of online information on Canadian topics isn't a problem to be solved strictly by the Wikidata community. However, it is for government, libraries, archives, and music organizations to give more significant consideration to digitization and web archiving projects.



https://doi.org/10.15215/aupress/9781771994521.01

## Wikidata and Events

The Mariposa Folk Festival data project provided an opportunity to explore aspects of cultural data description that are not possible within traditional descriptive practices of libraries and archives. A major aspect of this work was the creation of Wikidata items related to events or items based on things that take place in time. Typically, materials in libraries and archives are related to collection objects—for example, books, manuscripts, or recordings. The focus on or need for a collection-based item creates limitations in how something can be described and limits access and understanding of time-based artworks like music. Wikidata opens up the possibility of describing events and providing great dimensions related to data. The ability to describe events was an appeal of using Wikidata for the Mariposa Folk Festival project.

Wikidata allows for the representation of items through time, and the creation of timeline generation tools has become a popular method of visualizing data. A number of tools have been developed to take advantage of Wikidata items with date-related data. However, the modelling of Mariposa Folk Festival events was handled in a more detailed way through the linking of participants to an event. Wikidata projects provide an effective avenue for learning the best way to model and create data in relation to specific topics and can often provide important help in understanding a topic. The Wikidata Project Cultural Venues and Culture are helpful in reviewing critical issues related to describing cultural events. For the Mariposa Folk Festival data, it became clear that creating event data manually was a painstaking effort, as each individual performer needed their own Wikidata item, a large output of time and effort. Thus, while the creation of granular data is a more inclusive representation of the festival itself, it became painstaking. And though several years have passed since the inception of the Mariposa Folk Festival data, because of the ongoing challenges of structuring it, some question whether the investment of labour outweighs its perceived positive impacts. The increasing availability of tools focused on natural language processing may reduce the need for manual data entry, but there remains a dependence on machine-readable data.

## Representing Archival Collections in Canada

Creating data related to archival collections and holdings on Wikidata can aid in giving a clearer picture of collections at Canadian institutions and help


https://doi.org/10.15215/aupress/9781771994521.01

make primary source documents more findable. The Mariposa Folk Festival data project focused on an archival collection held at the Clara Thomas Archives and Special Collections at York University. One of the ways to connect unique collections held by Canadian institutions and Wikidata is to use the property "archives at." This property on Wikidata "describes the data value of a statement and can be thought of as a category of data, for example 'color' for the data value 'blue.' Properties, when paired with values, form a statement in Wikidata" ("Help: Properties," 2023). Properties in large part define the structure of Wikidata and must be approved by a consensus community process. Wikidata is made up of statements that include a property and a value in the form of an item. The values in Wikidata can be an internal link to an item within Wikidata or a link to a unique and permanent address on the web called a URI. As of June 2024, Wikidata currently lists 140 properties related to archives and people or organizations (Wikidata Property Explorer, n.d.).

Most of the properties listed are for external identifiers, but there are a small number that are for use to describe collections. Property P485, "archives at" links the Wikidata item being described to the institutions holding the archives for that item. York University archivist Katrina Cohen-Palacios (2019) has presented and provided workshop materials on methods of adding archival holdings to Wikidata—for example, adding a link from "The *Toronto Evening Telegram*" newspaper Wikidata item to the Clara Thomas Archives and Special Collections. Cohen-Palacios (2019) suggested that creating information about finding aids in Wikidata can save archivists time by automating some tasks that might otherwise need to be maintained manually.

Following the examples laid out by Cohen-Palacios (2019), Ansovini et al. (2022) began an initiative to add select archival holdings information from the University of Toronto to Wikidata. The addition of archival information can help individuals (and machines) from around the world find and enhance collections and materials related to Canadian culture. Furthermore, the addition of archival holdings information links institutions into the Wikidata network, allowing for queries and data visualizations to support alternate views of a myriad of connections. Demonstrating their work on the Canadian author Margaret Atwood, Ansovini et al. (2022) found that

> the simple addition of one "archives at" triple, linking the records to an item in Wikidata, allows for the kinds of connections that may have been made in the reading room, where archivists use their contextual

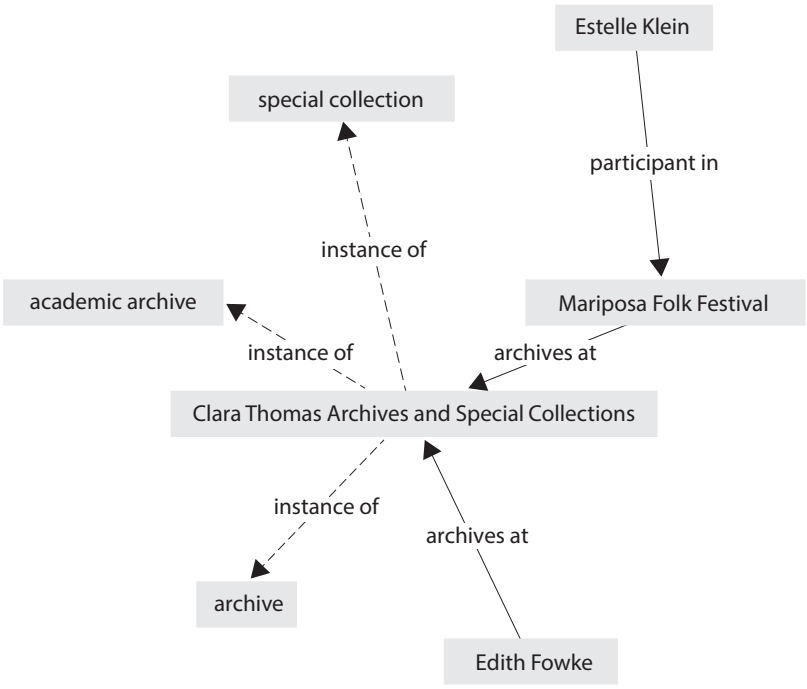


https://doi.org/10.15215/aupress/9781771994521.01

> knowledge to suggest to researchers records of related people, important dates, and notable publications. Wikidata can provide machine-actionable, community-generated access points that allow users to perform exploratory searches to visualize, search, and explore relationships to other entities. (p. 12)

The Mariposa Folk Festival data project also allowed for some experimentation in adding archival holdings information to Wikidata, and the Clara Thomas Archives and Special Collections also added archival holdings information to Wikidata in relation to the Mariposa Folk Festival. Figure 4.1, generated with a SPARQL query, makes these relationships visible in a network form. The Mariposa Folk Festival is now connected via a graph to the Clara Thomas Archives and Special Collections at York University. As an activity, adding a statement

**Figure 4.1.**
*Network representation of Wikidata properties related to the archival collection held at the Clara Thomas Archives and Special Collections, York University*





https://doi.org/10.15215/aupress/9781771994521.01

to Wikidata items relating to archival holdings is easier to accomplish than something like creating entries for festival events by being constrained in scope and not requiring a great deal of technical knowledge. While adding holdings could be impactful by creating structured data relating to archives in the wider internet, archival listings on Wikidata within Canada are uneven. Cohen-Palacios demonstrated this unevenness through a query (2019). The larger bubbles in this visualization represent between 300 and 550 holdings listed in Wikidata, with the bulk of institutions located in Ontario and Québec.

The creation of archival holdings entries could benefit from outreach and support to ensure greater equity in coverage; otherwise, the impact of these listings may remain relatively small. Other cultural heritage sites are also using Wikidata as a tool within their workflows. For example, the Canadian Heritage Information Network (CHIN) is using Wikidata as part of its ongoing Nomenclature project. The project has created a linked data vocabulary for use within the museum sector, and CHIN has been focused on the description of artworks at this stage of the project. While CHIN is creating alignment with CIDOC CRM as the content and schema standard, they are looking to Wikidata as a means of linking collections and making their content more visible.

## Wikidata and Indigenous Data in Canada

Another aspect of the Mariposa Folk Festival data project was considering the festival's "Native People's Area." The Native People's Area began in 1968 and was a physical location within the festival grounds. It included musicians, dancers, and storytellers and was included in its own section of the program. Alanis Obomsawin, an Abenaki filmmaker and performer, curated the Native People's Area from 1970 to 1976, and the influence of the rise in Indigenous rights movements within Canada and the United States is evident in the programming with the inclusion of groups such as the North American Indian Travelling College and activist poet Duke Redbird (Mariposa Folk Festival, 2014). The Native People's Area was handled as an additional dataset within the production of the Mariposa Folk Festival data. As with the other areas of the project, working with the festival data and Wikidata surfaced issues that are useful in a larger discussion of Wikidata within the Canadian context.

As stated in an earlier section of this chapter, finding information about individuals named in the program was frequently a challenge. However, it was



https://doi.org/10.15215/aupress/9781771994521.01

even more so with the Native People's Area. While some individuals, such as Shingoose, were named, many others only appeared in the program under a general name, such as "Six Nations Reserve Dancers" or "Metis Group." Research in the Mariposa Folk Festival archival holdings yielded no further information. The lack of specificity made it impossible to add these participants to Wikidata. The lack of information in relation to the Native People's Area could be an example of a lack of care or knowledge on the part of the festival organizers, and the lack of information generally on these groups points to a lack of information within the wider landscape of Canada. While it could be appropriate to include participants with vague or unclear names in an internal dataset, it is not appropriate to add such information to a global, open-knowledge base, as the meaning is too imprecise. The use of Wikidata versus an internally controlled and hosted repository or knowledge base should be carefully considered.

Additionally, a vital area of concern when looking at Wikidata within Canada is Wikidata's appropriateness for Indigenous data. Indigenous data can be defined as "data, information, and knowledge, in any format, that impact Indigenous Peoples, nations, and communities at the collective and individual levels; data about their resources and environments, data about them as Individuals, and data about them as collectives" (Carroll et al., 2021). This wide-ranging definition is in keeping with understandings of Indigenous peoples expressed in rights frameworks such as the United Nations Declaration on the Rights of Indigenous Peoples. Given this understanding, much of the data in Wikidata in relation to territories, environments, and Indigenous nations in the country now known as Canada is Indigenous data.

The Canadian federal government passed Bill C-15 on June 21, 2021, and the act respecting the United Nations Declaration on the Rights of Indigenous Peoples and its enactment "provides that the Government of Canada must take all measures necessary to ensure that the laws of Canada are consistent with the United Nations Declaration on the Rights of Indigenous Peoples (UNDRIP), and must prepare and implement an action plan to achieve the objectives of the Declaration" (Government of Canada, 2021). While both UNDRIP and Bill C-15 have been criticized by some Indigenous peoples, they provide a framework and means of implementing necessary measures to move to action regarding the recognition of Indigenous rights, including knowledge, cultures, data, and information. UNDRIP and Bill C-15 also provide impetus


https://doi.org/10.15215/aupress/9781771994521.01

for provincial and municipal governments, along with organizations, to consider measures to align policies and practices.

The growing global movement of Indigenous data sovereignty calls for the need for recognition, governance, and systems that ensure Indigenous peoples have control over their data, including data related to territories and cultures. While Wikidata, as an open-source and open-access project, adheres to the FAIR (findability, accessibility, interoperability, reusability) principles, making for positive impacts on the open-knowledge community (Odell et al., 2022), the FAIR principles ignore "power differentials and historical contexts" (Global Indigenous Data Alliance, or GIDA) and potentially uphold practices that are extractive and continue to harm Indigenous peoples (Hudson, 2020). Unnuanced and unmitigated approaches to open access are widely acknowledged to be problematic (Anderson & Christen, 2019). As of this writing, Wikidata has no mechanisms for identifying, safeguarding, or ensuring that Indigenous data is appropriately handled, including mechanisms to prevent data from being added in an unethical manner. Ideally, the Wikidata community and the Wikimedia Foundation would work toward implementing the CARE principles for Indigenous data: collective benefit, authority to control, responsibility, and ethics (Hudson, 2020). The GIDA calls on organizations to #BeFAIRandCARE. Lacking mechanisms for the ethical care and handling of Indigenous data, it is challenging to consider the ways in which Wikidata can be freely used for Indigenous data.

The data model also presents a problem for Indigenous data and respectfully representing Indigenous peoples, territories, and cultures on Wikidata. A key example is the lack of means to adequately identify Indigenous peoples. Indigenous peoples may identify themselves in a variety of ways, and their identity can be tied to their community and nation. Wikidata currently does not have the means to develop Indigenous nations as nations; rather, many Indigenous nations have been categorized as ethnicities. Indigenous identity is not an ethnicity, and recent works by Allison-Cassin (2023) and Johnston, Julian, and Singh (2022) suggest that the structural problems with the Wikidata data model for Indigenous identity are a serious barrier to considering Wikidata as an appropriate platform for information about Indigenous peoples. A further problem related to the identification of community is the lack of connection between Indigenous nations and Wikidata items.

In his *Elements of Indigenous Style*, Greg Younging (2018) points a number of times to the importance of recognizing Indigenous nationhood in referring



https://doi.org/10.15215/aupress/9781771994521.01

to Indigenous peoples. He suggests the term *nation* "has become widely accepted by Indigenous Peoples to describe separate Indigenous groups as political entities." Furthermore, Younging (2018) states that specific Indigenous peoples identify with a formal nation, such as the Métis nation of Alberta. However, Wikidata does not connect Indigenous peoples with nations, making the use of Wikidata inappropriate or questionable—especially in relation to understandings drawn from UNDRIP and various initiatives and practices related to self-determination.

Other structural problems with Wikidata in relation to Indigenous data are caused and amplified by structural problems with Wikipedia. While these issues might be unintentional, they represent an additional barrier to adequately supporting the appropriate handling of Indigenous data, making for a particular problem within the Canadian Wikidata landscape. A key example of this problem is related to the ways geographic areas are structured within Wikidata and is particularly striking in relation to Indigenous territories. Because of the Indian Act, many First Nations people were removed from the whole of their traditional territories to a system of smaller, restricted reserves of land. These reserves are geographic areas typically occupied by members of a single First Nation. Each reserve also has its own system of governance imposed by the federal government of Canada, which is known as a band government. A problem within Wikidata is confusion between "kinds" of entities in relation to First Nations communities. For example, a First Nation may be an "instance of" a First Nations band, a geographic location, or potentially another entity. The confusion in data structure oftentimes is because of the automated creation of a Wikidata item from a Wikipedia article where, in narrative form, there is little confusion. Brown states, "Description logics are hostile to ambiguity" (Brown, 2022, p. 3), and the ambiguity around what is meant by a First Nation is a structural problem making for a serious issue within data. See Allison-Cassin (2023) for a full discussion of this issue.

For the Mariposa Folk Festival data, attempting to document an individual's identity while recognizing the previous considerations was a challenge. While the consultation group was relatively small in CAPACOA's report, *Indigenous Artists and Wikidata: Explorations and Consultations Report* (Johnston et al., 2022), the overarching concerns expressed by participants about the appropriateness of Wikidata for their personal profile data are concerning. Wikidata, despite the appearance of being open and available for everyone without bias, is structured and modelled around Western understandings


https://doi.org/10.15215/aupress/9781771994521.01

of the world. While the data model can be changed and modified through community consensus, it remains that the community likely does not include many Indigenous peoples, and the scope of some of the issues is of such a large scale that a significant intervention may be required. Additionally, knowledge and understanding of different knowledge systems and the importance of understanding issues within intellectual property regimes are not widely known.

## Conclusion

The Mariposa Folk Festival data project was a way to experiment with linked data network production using Wikidata. The challenges of working with Wikidata from a Canadian perspective point to problems beyond Wikidata for cultural data in Canada. For example, significant issues remain about the place of Indigenous data within Wikidata that require timely and vital intervention. The continuing closure of community newspapers and local media will harm the ability of those working on the Wikimedia platforms to document Canadian-focused topics. As the Wikidata platform grows and becomes a notable hub for structured data, people in Canada will need to strategize on ways to advocate and provide pathways for adding more data to the platform—particularly in relation to communities not well represented in mainstream data. Wikidata may never play an obvious role in people's everyday lives. However, it plays a significant role within the open metadata movement in Canada, as evidenced by the number of different ways it is being used within Canada.

For institutions and individuals to participate more effectively, great attention must be paid to metadata and digital practices at individual organizations to improve the availability of stable identifiers, open data, and sources for referencing Wikidata items. Recognizing the need to support open data initiatives that will serve all people in Canada, several cultural agencies within the Canadian federal government have used Wikidata: "Such sharing reduces the work involved in many facets of digital collection management by drawing on expertise and updates provided by other teams with established authority" (Government of Canada, 2023).

It is worth remembering that "as the largest repository of knowledge ever collected, Wikipedia remains both an astounding human achievement as well as endless opportunity for improvement, both in content and in community"



https://doi.org/10.15215/aupress/9781771994521.01

(McDowell & Vetter, 2021). Wikidata will likely continue to grow within Canada by building and supporting the community. Canadian digital humanities researcher Susan Brown suggests, "Community building and cross-sectoral collaboration may be most important to address diversity effectively, given the substantial resources and infrastructure required to work with LOD" (Brown, 2022). A strong and diverse future for Wikidata in Canada will strengthen Canada's digital presence.

## References


*About West End Phoenix*. (n.d.). West End Phoenix. Retrieved July 15, 2024, from https://www.westendphoenix.com/about-us

Allison-Cassin, S. (2012). The possibility of the infinite library: Exploring the conceptual boundaries of works and texts of bibliographic description. *Journal of Library Metadata*, *12*(2–3), 294–309. https://doi.org/10.1080/19386389.2012.700606

Allison-Cassin, S. (2023). Métis nationhood: An examination of Indigenous nationhood, sovereignty and linked data through a Wikidata case study. In K. Burlingame, A. Provo, & B. M. Watson (Eds.), *Ethics in linked data*. Litwin Press. https://litwinbooks.com/books/ethics-in-linked-data/

Allison-Cassin, S., Ruest, N., St. Onge, A., & Suhonos, M. (2015). *Sounding it out: The Mariposa Folk Festival and a linked open data digital library*. Digital Humanities, Sydney, Australia. https://dh-abstracts.library.virginia.edu/works/2357

Allison-Cassin, S., & Scott, D. (2018). Wikidata: A platform for your library's linked open data. *code{4}Lib Journal*, *40*. https://journal.code4lib.org/articles/13424

Ansovini, D., Babcock, K., Franco, T., Jung, J. A., Suurtamm, K., & Wong, A. (2022). Knowledge lost, knowledge gained: The implications of migrating to online archival descriptive systems. *KULA: Knowledge Creation, Dissemination, and Preservation Studies*, *6*(3), Article 3. https://doi.org/10.18357/kula.234

ARL Task Force on Wikimedia and Linked Open. (2019). *ARL white paper on Wikidata: Opportunities and recommendations* [Report]. Association of Research Libraries. https://apo.org.au/node/254221

Art Gallery of Ontario. (2015). Art+feminism Wikipedia edit-a-thon at AGO, Toronto 15. In *Wikimedia Commons*. https://commons.wikimedia.org/wiki/File:Art%2BFeminism_Wikipedia_Edit-A-Thon_at_AGO,_Toronto_15.jpg

Beghaeiraveri, S. A. H., Gray, A. J. G., & McNeill, F. (2024). RQSS: Referencing quality scoring system for Wikidata. *Semantic Web Journal*. https://www.semantic-web-journal.net/content/rqss-referencing-quality-scoring-system-wikidata-1




https://doi.org/10.15215/aupress/9781771994521.01

Bianchini, C., Bargioni, S., & Pellizzari di San Girolamo, C. C. (2021). Beyond VIAF: Wikidata as a complementary tool for authority control in libraries. *Information Technology and Libraries*, *40*(2), Article 2. https://doi.org/10.6017/ital.v40i2.12959

Brown, S. (2022). Same difference: Identity and diversity in linked open cultural data. *International Journal of Humanities and Arts Computing*, *16*(1), 1–16. https://doi.org/10.3366/ijhac.2022.0273

Canadian Arts Presenting Association. (n.d.). *Linked digital future*. LDFI. Retrieved May 21, 2021, from https://linkeddigitalfuture.ca/

CAPACOA. (n.d.). *Linked digital future*. LDFI. Retrieved March 27, 2023, from https://linkeddigitalfuture.ca/

*Chair for research in e-librarianship 2015–2017*. (n.d.). York University Libraries. Retrieved July 1, 2024, from https://www.library.yorku.ca/web/about-us/wpscott-chair-e-librarianship/chair-for-research-in-e-librarianship-2015-2017/

Cohen-Palacios, K. (2019, October 24). *Wikidata and archivists*. Archives Association of Ontario (AAO) Institutional Issues Forum, Toronto. https://yorkspace.library.yorku.ca/xmlui/handle/10315/36898

Colla, D., Goy, A., Leontino, M., & Magro, D. (2021). Wikidata support in the creation of rich semantic metadata for historical archives. *Applied Sciences*, *11*(10), Article 10. https://doi.org/10.3390/app11104378

The Dirty Shames. (n.d.). In *Wikidata*. Retrieved July 15, 2024, from https://www.wikidata.org/wiki/Q84939500

Gertner, J., Hurst, A., Lozano, M., Woo, J., Ramirez, D., & Vancura, A. (2023, September 10). The Sunday read: "Wikipedia's moment of truth." *New York Times*. https://www.nytimes.com/2023/09/10/podcasts/the-daily/wikipedia-ai.html

Government of Canada. (2023). *Linked open data overview*. https://www.canada.ca/en/services/culture/history-heritage/museology-conservation/collections-management/linked-open-data/linked-open-data-overview.html

Government of Canada, Department of Justice. (2021, April 12). *Backgrounder: United Nations declaration on the Rights of Indigenous Peoples Act*. https://www.justice.gc.ca/eng/declaration/about-apropos.html

Hawkins, A. (2022). Archives, linked data and the digital humanities: Increasing access to digitised and born-digital archives via the semantic web. *Archival Science*, *22*(3), 319–44. https://doi.org/10.1007/s10502-021-09381-0

Help: Properties. (2023, January 22). In *Wikidata*. https://www.wikidata.org/w/index.php?title=Help:Properties&oldid=1817117774

Help: Sources. (2023, March 6). In *Wikidata*. https://www.wikidata.org/w/index.php?title=Help:Sources&oldid=1846065968

Hudson, M. (2020). *Indigenous data sovereignty: Towards an equitable and inclusive digital future*. A Digital New Deal: Visions of Justice in a Post-Covid World.



https://doi.org/10.15215/aupress/9781771994521.01

https://projects.itforchange.net/digital-new-deal/2020/11/01/indigenous-data-sovereignty-towards-an-equitable-and-inclusive-digital-future/

Johnston, B., Julien, F., & Singh, A. (2022). *Indigenous artists and Wikidata: Explorations and consultations report*. CAPACOA. https://linkeddigitalfuture.ca/research/indigenous-artists-and-wikidata/

McCracken, K. (2018, September 6). *Doing the work: Editing Wikipedia as an act of reconciliation*. Medium. https://medium.com/on-archivy/doing-the-work-editing-wikipedia-d82e927adb9f

McDowell, Z. J., & Vetter, M. A. (2021). *Wikipedia and the representation of reality*. Routledge.

McFadden, F., Raymond, J., & Toronto Telegram. (1966). *The Dirty Shames, in performance at The Riverboat [Carol Robinson and Roy Michaels sharing one microphone in foreground, Chick Roberts and Amos Garrett sharing another microphone in background]*. York University Digital Library. https://digital.library.yorku.ca/node/1078262

Music in Canada @ 150: A Wikipedia and Wikidata project. (2019, March 14). In *Wikimedia*. https://meta.wikimedia.org/w/index.php?title=Grants:Project/Smallison/Music_in_Canada_@_150:_A_Wikipedia_and_Wikidata_Project&oldid=18933864

Nielsen, F. Å., Mietchen, D., & Willighagen, E. (2017). Scholia, scientometrics and Wikidata. In E. Blomqvist, K. Hose, H. Paulheim, A. Ławrynowicz, F. Ciravegna, & O. Hartig (Eds.), *The Semantic Web: ESWC 2017 satellite events* (pp. 237–59). Springer International Publishing. https://doi.org/10.1007/978-3-319-70407-4_36

Odell, J., Lemus-Rojas, M., & Brys, L. (2022). *Selected tools for using Wikidata*. https://doi.org/10.7912/5mtv-h307

Proffitt, M. (2018). *Leveraging Wikipedia: Connecting communities of knowledge* (1st ed.). American Library Association.

Riverboat Coffee House. (2024, April 23). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Riverboat_Coffee_House&oldid=1220389506

Scott, D. (2017, May 28). Wikidata workshop for librarians. *Coffee|Code: Dan Scott's Blog*. https://coffeecode.net/wikidata-workshop-for-librarians.html

Scott, D. (2017, June 2). Wikidata, Canada 150, and music festival data. *Coffee|Code: Dan Scott's Blog*. https://coffeecode.net/wikidata-canada-150-and-music-festival-data.html

St. Onge, A. (2013). *Mariposa folk foundation fonds*. York University Libraries Clara Thomas Archives & Special Collections. https://atom.library.yorku.ca/index.php/mariposa-folk-foundation-fonds

Tharani, K. (2021). Much more than a mere technology: A systematic review of Wikidata in libraries. *Journal of Academic Librarianship*, *47*(2), 102326. https://doi.org/10.1016/j.acalib.2021.102326



https://doi.org/10.15215/aupress/9781771994521.01

Treasury Board of Canada Secretariat. (n.d.). *Open data 101*. Retrieved March 19, 2023, from http://open.canada.ca/en/open-data-principles

van Veen, T. (2019). Wikidata: From "an" identifier to "the" identifier. *Information Technology and Libraries*, *38*(2), Article 2. https://doi.org/10.6017/ital.v38i2.10886

Vrandečić, D. (2013). The rise of Wikidata. *IEEE Intelligent Systems*, *28*(4), 90–95. https://doi.org/10.1109/MIS.2013.119

Waagmeester, A. et al. (2020). Wikidata as a knowledge graph for the life sciences. *eLife*, *9*, e52614. https://doi.org/10.7554/eLife.52614

Wikidata Property Explorer. (n.d.). Retrieved March 26, 2023, from https://prop-explorer.toolforge.org/

Wikidata: Introduction. (2022, July 10). In *Wikidata*. https://www.wikidata.org/w/index.php?title=Wikidata:Introduction&oldid=1674555622

Younging, G. (2018). *Elements of Indigenous style: A guide for writing by and about Indigenous peoples*. Brush Education.

Zhao, F. (2023). A systematic review of Wikidata in digital humanities projects. *Digital Scholarship in the Humanities*, *38*(2), 852–74. https://doi.org/10.1093/llc/fqac083

Zhu, L., Xu, A., Deng, S., Heng, G., & Li, X. (2023). Entity management using Wikidata for cultural heritage information. *Cataloging & Classification Quarterly*, *61*(1), 20–46. https://doi.org/10.1080/01639374.2023.2188338



https://doi.org/10.15215/aupress/9781771994521.01

# 5 Wikimedia in a Québec Art Museum

## Exploring an Open Cultural Institution Model

Nathalie Thibault

The Musée national des beaux-arts du Québec (MNBAQ) is committed to expanding access to its content and data through three projects linked to the Wikimedia movement. Launched in 2018 and still underway, these initiatives are the Wikipedia project (with articles written primarily in French), the Wikidata project, and the Wikimedia Commons project. These initiatives seek mainly to promote and share on the internet information as well as royalty-free images held by the MNBAQ while enabling the museum to acquire new digital skills at an organizational level. These initiatives are in line with the various goals of the open-access movement, the open GLAM movement, and the participatory museum model. The notion of a "work in progress" underlines the museum's commitment to act as project manager for a set of interdependent works on which to base the vision and infrastructure of an open cultural institution.

In general, Wikimedia projects promote access to digital cultural content by means of more open licences. By "open" we mean "any content or data that is freely accessible, usable, modifiable, and shareable for any purpose, subject to requirements that preserve its origin and openness" (*Open Definition*, n.d.). This open access to digital cultural heritage as a commons now enables the museum to better respond to the needs and expectations of different audiences and to better support the work of researchers as well as that of the museum, artistic, and documentary communities. The European Commission



https://doi.org/10.15215/aupress/9781771994521.01

on open GLAM illustrated the diverse impacts of opening up content and data by highlighting the importance of sharing knowledge for research, innovation, and creativity. For instance,

- More openly licensed cultural content enables teachers across the world to reuse this work in the classroom.
- More open cultural data enables researchers to draw links between people, things, and events through the use of innovative techniques such as text mining and visualization.
- More open cultural content enables citizens from across the world to enjoy this material, understand their cultural heritage, and reuse this material to produce new works of art.

## Brief Review of the Literature and Conceptual Framework

In 2006, at a time when concern about the validity of Wikipedia content was dominating public and scientific debate, a research article proposed an argument and a method for increasing the visibility and promotion of memory institutions: It encouraged museums in detail to set up a Wikipedia entry (Bowen & Angus, 2006). It was a few years later, in 2008, that the open GLAM movement really emerged, following the Wikimania Conference held at the Bibliotheca Alexandrina: Holding the conference in this highly symbolic site provided a context where Liam Wyatt, an Australian Wikimedian, saw the potential for structuring long-term relations between the various cultural institutions and Wikipedia. In 2009, the GLAM-Wiki expression and community were forged together in Canberra, at a session organized by Wyatt and entitled "Finding the Common Ground," which brought together players from libraries, museums, and archives (Lih, 2018). This event helped generate a conversation aimed at mutualizing the capabilities of cultural institutions while converging with the commons movement. Shortly afterward, in 2010, Liam Wyatt became the first Wikipedian in residence at a museum, the British Museum. He set about building the relationship and sharing expertise between the museum world and Wikimedians. At the British Museum, the benefits of this collaboration were quickly realized, given that the article devoted to the museum's star attraction, the Rosetta Stone, was already being consulted five times more often on Wikipedia than on the museum's own website (Cohen, 2010).



https://doi.org/10.15215/aupress/9781771994521.01

In the past, members of the public and particularly researchers visited museums, archives, and libraries in person so they could consult documents and images for their work. In the digital environment, the strategies of museums, archives, and libraries now aim to broaden access: They offer new online services and increasingly open practices (Klein & Cardin, 2018, p. 101). These strategies aim to create a more participatory environment; cultural institutions are exploring the possibilities of collaboration and possibly cocreation by renewing the way in which they traditionally ensure user and visitor access to their collections. The concept of the "participatory museum" is becoming a reference point for transformation, with Wikimedia projects regularly serving as a springboard (Simon, 2010). As a result, the concept and open rationalization of access are shifting from a physical dimension to an online presence. They also take into account the capacities for collaboration and participation that form part of the participatory, inclusive, open museum movement and also characterize open GLAM (Evans, 2024; Salgado & Marttila, 2013).

For a cultural institution, the sharing and opening up of collections and information in the digital environment give rise to a number of issues and challenges, with effects going beyond the management of change within the organization. The tradition of museums and especially national museums is characterized by the dimensions of governance, power, authority, and control. These dimensions continue to play a predominant role in the management of digital collections, although this culture of cultural institutions is tending to change little by little despite resistance (Fouseki, 2013). Moreover, it seems that apprehensions about a loss of control over collections or a possible reduction in the number of face-to-face visits, which would depend on the opening up of content and data, are unfounded (Kapsalis, 2016).

The gradual exploration of different Wikimedia platforms therefore provides an opportunity to rethink and update the ethical positioning of museums, particularly with respect to the issues of ownership and copyright on which this redesigned and expanded access for the public depends. More specifically, the Wikipedia experience is leading to debates and negotiations on the ownership of objects, images, and digital data; on the place of their commercial or public uses; or even on their use as commons. Collaboration within Wikipedia projects encourages these emerging uses, which will reflect "the participatory nature, philosophy, and ideology of each museum institution" and the state or degree of each institution's transition (Fouseki, 2013). The three initiatives presented are part of this process and are indicative of


https://doi.org/10.15215/aupress/9781771994521.01

the museum's transitional identity in terms of openness and participation. The three initiatives serve as examples of strategies for opening up the museum's collections, information, and data and for adapting professional practices in terms of participation, social inclusion, and openness to new uses.

Other observations are worth making in terms of the changing needs and uses of museum information. Firstly, the use of Wikimedia platforms and the more open licences they recommend for the content of digital cultural heritage are expanding the possibilities for researchers from different fields in the arts, humanities, and social sciences for both accessing and studying cultural heritage (Terras, 2015). Secondly, digital cultural consumption is also likely to extend the range of uses "outside of the expected art context, enabling new forms of hedonic and utilitarian consumption" (Navarrete & Villaespesa, 2020). In other words, the Wikimedian use supported by museums such as the Metropolitan Museum of Art leads to a new conceptualization of authentic taste that takes into account digital consumption, which in turn enhances the possibilities of more informative uses of a collection (Navarrete & Villaespesa, 2020).

Moreover, the MNBAQ is exploring various collaborations in order to broaden access to the reuse of its data and content by other institutions, mainly of the GLAM type. These collaborations represent a significant challenge where an open and linked data strategy, through Wikidata, for example, appears to facilitate research by "enabling digital humanities researchers to establish links between (and make sense of) the multitude of digitized cultural heritages available on the web" (Alexiev, 2018). This exploration of the facets of Wikimedia as an open system has proved to be a timely, productive, and sustainable choice for the MNBAQ and one that is based on these various considerations, on the renewed missions of museums and their explicit democratization objectives, and on new individual, community, and institutional uses.

## The Museum's Transition to an Open Cultural Institution Model

The MNBAQ's three initiatives required the contribution of a varied multidisciplinary team drawn from several sectors within the institution: legal affairs, information technology, conservation, collections management, preventive conservation, restoration, photography, archives, and documentation.



https://doi.org/10.15215/aupress/9781771994521.01

As mentioned earlier, the initial aim was to make newly digitized cultural content accessible online to different audiences. It very quickly became apparent that the MNBAQ needed to disseminate synchronized, reliable, and verified data on the works and artists represented in its collections—namely, the permanent collection, the study collection, and the loan collection of works of art, not forgetting the vast section of private archives. The museum has a long-standing tradition of collaborating with its users. Each year, it responds to a growing number of requests for information from researchers, students, and the general public. All these people want access to images and descriptive data from national and public collections. Disseminating the entire heritage collection is a great way of bringing citizens closer to works that are less widely exhibited or have not necessarily been the subject of in-depth research. It is estimated that around 25% of works and objects have already been exhibited or reproduced in a publication. The online dissemination of content makes it possible to discover archived and often little-known elements.

In 2016, the inauguration of the Pierre Lassonde pavilion generated significant visitor numbers both on-site and online. At the same time that the MNBAQ got ready to unveil this pavilion, the museum rethought how its collections should be accessed online. As a result, the drive to broaden and open up access to our data and information coincided with the major architectural impetus represented by the new pavilion.

The MNBAQ is a Québec provincial museum that currently holds over 42,000 works of art and thousands of archival documents, objects, and rare books. Since it opened in June 1933, the museum collections have grown to the point where the catalogue now lists more than 4,500 artists, artists' collectives, craftspeople, authors, and others. As a publisher, it has published more than 1,000 exhibition catalogues and books related to its acquisitions, exhibitions, and other dissemination activities. These publishing activities have generated a significant amount of content, data, texts, photographs, and archival documents to be preserved and shared in the future in accordance with current copyright legislation. The museum's mission is to make known, promote, and preserve Québec art of all periods, from ancient art to contemporary art, and to ensure a place for international art through acquisitions, exhibitions, and other cultural activities.

The museum owns a number of works of art in the public domain and has a significant artistic and documentary heritage. This corpus is currently being digitized for future dissemination and direct online access. It seemed logical


https://doi.org/10.15215/aupress/9781771994521.01

for the MNBAQ to explore web dissemination projects whose orientations and objectives would be compatible with the approach favouring reuse by citizens as well as the museum, artistic, and documentary community. We were aiming for digital projects that would enable information to be updated and synchronized almost in real time, and that would give MNBAQ teams the flexibility to manage and monitor processes. For several years now, the museum has been going through a period of organizational and technological transformation, like other cultural organizations. It wants to deploy information technologies and sciences in order to achieve digital maturity and efficiency through increased access. The museum must take into account documentary norms, interoperability standards, the development of the semantic web, and ontologies in order to exchange and aggregate information through the creation of federating platforms currently under review (Canadian Heritage Information Network, Société des musées du Québec, Répertoire sur le patrimoine culturel du Québec, Artists in Canada directory, etc.). We are therefore preparing the ground by exploring the potential of Wikimedia tools.

Following the discussions initiated in 2018 to develop the MNBAQ's digital strategy and thanks to funding from Québec's Digital Cultural Plan, the museum decided to explore and experiment with open-access content and open data projects, considering that it was a museum, documentary, and research player within Québec's artistic and cultural ecosystem. The museum manages a library, a documentation centre, and private collections and archives, which it aims to disseminate and make available to as many users as possible. Before disseminating information on web platforms, the museum takes into account the legal and governmental framework in Québec and Canada regarding copyright, the respect for and protection of personal information, access to information, proactive disclosure, and so on. A major digitization project underway since 2018 has already resulted in thousands of images being made available on the MNBAQ's collections website.

According to Christelle Molinié of the Musée de Bretagne in Rennes, France, "By opening up to the collaborative practices of Web 2.0, the work of documentary mediation takes on a new dimension that encourages interaction, appropriation and re-use of works by the public, while continuing to participate in the fundamental missions of museums in terms of study, accessibility and dissemination of collections in the public domain and cultural democratization" (Fraysse, 2015).



https://doi.org/10.15215/aupress/9781771994521.01

We have therefore taken into account the importance of free access and the possible reuse of certain data, public information, or royalty-free images in the digital approach to all these projects at the MNBAQ.

The recent experiences and digital innovations of the Bibliothèque et Archives nationales du Québec (BAnQ) and the Cinémathèque québécoise have strongly influenced the MNBAQ's thinking about the documentary projects to be planned and tested within the organization and for the community on the web. This has in turn led to an approach of openness and the free sharing of information, which until now had only been accessible in the organization's databases. As Jessica de Bideran and Romain Wenz point out,

> For a cultural institution, a presence on Wikipedia is based on two movements that are a priori antagonistic: on the one hand, the desire to achieve tangible objectives (in terms of writing articles, enhancing the value of heritage holdings and raising the institution's profile), and on the other, taking into account the collaborative and evolving construction of knowledge resulting from the choices of a free and autonomous community. (De Bideran & Wenz, 2020)

The various initiatives have helped develop new digital skills and raise awareness of the challenges of open culture, thereby facilitating better links between traditional and participatory approaches to museums.

## First Project: Exploring Wikipedia Articles in French

In 2018, the MNBAQ explored the potential of Wikipedia in French by directly integrating existing articles on artists into the new version of its collections site. An overview of the collection sites of several art museums demonstrated the value of exploiting Wikipedia in different languages. The Tate Modern includes a link for each artist directly on its site, sending users to the relevant English-language Wikipedia article. The multilingual approach is a real advantage, as it allows us to display the names of artists in their original language—Innu, Japanese, and so on—in addition to French.

The trials and proof of concept involved testing a daily automated computer script. This script made it possible to update a selection of public data as contributors updated the French-language version of Wikipedia. In the summer of 2018, an initial experiment was carried out on a selection of


https://doi.org/10.15215/aupress/9781771994521.01

300 Québec and Canadian artists whose works of art are exhibited in the MNBAQ pavilions. We began by studying the number and quality of articles on Wikipedia in French. The follow-up, monitoring, and examination of these articles were conclusive in terms of access, script stability, monitoring for vandalism, and above all, reliability of the sources.

The project took off in autumn 2018. From the outset, we realized the importance of offering training to staff members about this collaborative project so we could find out about their apprehensions and perceptions and also pursue the project in line with certain priorities: the notoriety of artists, published sources, works exhibited in the museum's pavilions, and so on. Since 2019, the museum has been adding draft articles as well as articles on emerging Québec and Canadian artists, with the complicity and collaboration of the French-speaking Wikipedian community around the world. To date, more than 1,000 articles in French have been linked directly to our collections site. These articles are constantly evolving and are consulted by hundreds of internet users every day. There are as many articles in English as in French on the artists in the MNBAQ's collections. Some artists are represented in more than one language. In all, 128 languages are represented. This multilingual contribution is invaluable in promoting the visibility of Québec and Canadian artists.

We began organizing workshops on how to contribute to Wikipedia in 2019. As well as discussing the myths surrounding the encyclopedia, we wanted to make a range of reliable, published sources from the MNBAQ library accessible to the public, including books, periodicals, newspaper articles, and so on, as a way of documenting the subjects targeted by the participants. Unfortunately, the pandemic put a stop to this on-site component of the workshops. It may be repeated in the future with themes for each workshop. A project to write articles in French as part of a college course has been launched. As Marta Severo points out, "This trend is part of a phenomenon known as citizen science. This term has become particularly fashionable in recent years; it is intended to include all initiatives that facilitate the participation of non-professionals in scientific research and the construction of scientific knowledge" (Barbe & Severo, 2021, p. 88).



https://doi.org/10.15215/aupress/9781771994521.01

## Second Project: Exploring Wikidata

In June 2019, a second project linked to the opening up of content was added, with a view to combining the Wikipedia project with the revision of certain public biographical data held by the MNBAQ in its collections management system as well as in its integrated library management system. The comparison, standardization, scrubbing, and updating of the data in these two systems required the collaboration of several colleagues so we could harmonize the data and compare it with data already accessible in the free Wikidata knowledge base. This database is published collaboratively and hosted by the Wikimedia Foundation.

The Wikidata element for each emerging artist has been integrated directly into one of the fields in the MNBAQ's collection management system. This Wikidata identifier uses common reference systems so the data extracted and visible on the collections website can be linked directly and the artists can be identified efficiently. Each Wikidata element for an artist gathers and collates unique and perennial identifiers (Virtual International Authority File [VIAF], International Standard Name Identifier [ISNI]), perennial authorities from the Library of Congress, the Bibliothèque nationale de France, the Online Computer Library Center, the Union List of Artist Names, and so on, as well as direct links to the Répertoire du patrimoine culturel du Québec, the *Dictionary of Canadian Biography* identifier, each artist's official website, their social networks, and so on.

This open, structured data from Wikidata is used bidirectionally by the MNBAQ, which imports and adds new data as it acquires works of art. A visualization tool in the form of a dashboard helps track the progress of the project. This tool makes it easy to add new artists' authorities to be created as the weeks go by while keeping a history of the actions undertaken. This is a very important tool because it improves the management of the information disseminated, the traceability and logging of actions in terms of updating both external data contributors and MNBAQ staff. The museum has taken advantage of the recent creation of an ISNI agency within BAnQ to create ISNI identifiers for certain deceased artists, in collaboration with BAnQ. We are preparing the ground because this unique identifier with an international scope is proving to be extremely useful in interoperability platform projects to properly identify artists and communities and offer the possibility of a chain of monetary rewards in the not-so-distant future.


https://doi.org/10.15215/aupress/9781771994521.01

Information professionals, the registrar's team, and the museum's curators reviewed and approved the inclusion of thousands of items and properties in the Wikidata knowledge base under an open licence. This task was carried out with rigour and meticulousness and was needed so that the public data could be harmonized with published sources and up-to-date knowledge about the artists targeted by this Wikidata project. What's more, there is considerable potential for this data to be reused by other organizations or individuals in the long term. To date, the data entered in Wikidata links the artists in the MNBAQ's collections to more than 1,000 museums and collections around the world.

In addition, this project is exploring and clarifying the quality of data and the sharing of open-access content. The project has led to professional exchanges with other museums in Europe and the United States, and we have been able to confirm the authenticity of certain works of art by cross-referencing our data with that of international museums. The project provides daily monitoring of more than 7,000 items in Wikidata and makes it easier to update the data every day in our collection management and library systems. In addition, properties linked to the existence of archival fonds and documentary files on Québec and Canadian artists have been integrated into this Wikidata project, as well as mentions of Québec and Canadian museums, which hold works by artists represented at the MNBAQ. Finally, several elements have been created for art historians, museums, galleries, exhibitions, and so on in order to link the data described in Wikidata.

In June 2020, a Wikimedian contributor and the Wikidata community proposed and obtained the creation of a unique identifier for an MNBAQ artist. This property encompasses more than 4,000 elements and now enables the museum to effectively monitor data on the artists represented in its collections.

The MNBAQ thus joins a number of art museums interested in sharing content and proactively disclosing their data on Wikidata and Wikipedia, including the National Gallery of Canada, the Cleveland Museum of Art, the Art Institute of Chicago, the Smithsonian, the American Museum of Art, the Whitney Museum of American Art, the Nelson-Atkins Museum of Art, and the Saint Louis Museum of Art, to name but a few.

The fact that the Wikimedia community is also getting involved in the projects has been a motivating factor for several colleagues. After 3 years of work, we can see the extent of the collaborative work carried out on French-language Wikipedia articles and Wikidata data by the MNBAQ. Every day,



https://doi.org/10.15215/aupress/9781771994521.01

several additions, modifications, and adjustments are made to the data, properties, and texts.

The MNBAQ is also collaborating on the Wikidata project Sum of All Paintings, which lists more than 95,000 paintings from the collections of museums, public institutions, and private collections around the world. This project links data on artists to their paintings and sometimes constitutes the beginnings of a catalogue raisonné in which the data is sorted chronologically or by the theme or subject represented.

In December 2020, this project led to the use of a new identifier on Wikidata for the MNBAQ: the digital identifier of a work of art on the website of the Musée national des beaux-arts du Québec. This identifier lists descriptive data, images, and so on for each painting held by the MNBAQ. Once again, it was the community that created this identifier and made it possible to reuse the data. Our role is to make adjustments and modifications based on research carried out by the conservation team. This specific project has made it possible to discover the images contributed to Wikimedia Commons over the years. Contributors show continuous interest in royalty-free images of certain works in our collections. We regularly upload new images in line with user needs and current research topics.

## Third Project: Images in Wikimedia Commons

A third initiative relates to the potential for disseminating and reproducing images under a free licence. The museum has explored and targeted the possibility of placing certain images of works of art, archival documents, and so on in the public domain via the Wikimedia Commons media library. Allowing certain works to be downloaded, distributed, and reproduced without having to request permission from the MNBAQ is a major test for the institution. The selection of images for this project was based on the needs of the museum's internal and external users. Prior to this integration project, the team had already contributed images of works of art taken during their exhibition or presented in the museum's public spaces. The contributors to Wikimedia Commons are passionate photographers who devote their time to visiting public places and spaces and then sharing the fruits of their labour for the common good. As Simon Côté-Lapointe points out in his article "Les documents audiovisuels numériques d'archives," "Everyone is potentially an expert, a creator and a collaborator thanks to the democratization of digital tools,


https://doi.org/10.15215/aupress/9781771994521.01

as demonstrated by initiatives such as Wikipedia, YouTube and the Internet Archive. And this is particularly true of the audiovisual sector, thanks to the democratization of the means of manipulation" (2019, p. 46). In order to launch such information disclosure and sharing projects, we had to accept that the data shared would not be complete during the first iterations. The documentation and observation work has enabled us to acquire new skills in terms of data normalization and standardization. The museum team's multidisciplinary approach made it easier to meet deadlines, take ownership of the projects, and integrate a new way of working. Certain fields in our database and their concordance with properties in Wikidata are still being explored: gender identity, sexual orientation, nationality versus citizenship, and membership of an ethnocultural group. The reflex is to turn to the artists concerned and respect their choices as to whether or not to disclose certain information.

The impact of these three projects is tangible given that a percentage of referencing comes from digital actions carried out on Wikipedia, Wikidata, and Wikimedia Commons. Queries and analysis tools show that a growing number of visits to the collections website stem directly from these projects.

## Discussion and Conclusion

In short, these exploratory projects are putting in place a digital knowledge base among staff so they can learn how to edit content on the web. We need to plan for an alignment with the semantic web and the ontologies that will have to be implemented in the heritage and cultural sector over the next few years. Another notable benefit is that the conversation with internet users has been enriched since the collections were put online, as they participate directly in the documentation through their comments and exchanges with the museum. What's more, some of the data contributed by the MNBAQ to Wikidata have already been incorporated into other projects underway within the cultural community in Québec and Canada, focusing on the performing arts and the literary arts, including a number of collaborative projects in the regions of Québec. This effort by the communities of practice demonstrates the desire to ensure that everyone—on the basis of their databases, documentation, and published sources—gains greater visibility of Québec's cultural content while at the same time adopting the principles of standardization and future interoperability across different disciplines. Experiments such as the WikiProjects at the MNBAQ are gradually bringing users closer to a new



https://doi.org/10.15215/aupress/9781771994521.01

pooling of resources, digital humanities, and the potential for viewing and editing their data and metadata. This type of open-content project can only enhance the value of information in a museum institution through the digital learning that takes place. These projects pave the way for collective and collaborative intelligence while enhancing the value of the knowledge held by an organization and consolidating its commitment to the open-access and open GLAM movements.

## References


Alexiev, V. (2018). Museum linked open data: Ontologies, datasets, projects. *Digital Presentation and Preservation of Cultural and Scientific Heritage*, *8* (September), 19–50. https://doi.org/10.55630/dipp.2018.8.1

Barbe, L., & Severo, M. (Eds.). (2021). *Wikipédia, objet de médiation et de transmission des savoirs*. Presses universitaires de Paris Nanterre. https://doi.org/10.4000/books.pupo.15040

Bowen, J., & Angus, J. (2006). *Museums and Wikipedia* [Conference paper]. Museums and the Web 2006: The international conference for culture and heritage on-line, Albuquerque, NM. https://www.archimuse.com/mw2006/papers/bowen/bowen.html

Cohen, N. (2010, June 4). Venerable British Museum enlists in the Wikipedia revolution. *New York Times*. https://www.nytimes.com/2010/06/05/arts/design/05wiki.html

Côté-Lapointe, S. (2019). Les documents audiovisuels numériques d'archives. *Documentation et bibliothèques*, *65*(3), 39–57. https://doi.org/10.7202/1064748ar

De Bideran, J., & Wenz, R. (2020). *Enseigner la diffusion du patrimoine des bibliothèques sur Wikipédia: Retour sur une expérimentation*. Bulletin des bibliothèques de France (BBF). https://bbf.enssib.fr/matières-a-penser/enseigner-la-diffusion-du-patrimoine-des-bibliothèques-sur-wikipedia_69788

Evans, J. (2024). *Aligning and sharing data with Wikidata—a feasibility study*. Knowledge Commons. https://works.hcommons.org/records/5ym2e-bap78

Fouseki, K. (2013). Digital museum collections and social media: Ethical considerations of ownership and use. *Journal of Conservation and Museum Studies*, *11*(1), 5. https://doi.org/10.5334/jcms.1021209

Fraysse, P. (2015). La médiation numérique du patrimoine: Quels savoirs au musée? *Distances et médiations des savoirs*, *3*(12). https://doi.org/10.4000/dms.1219

Kapsalis, E. (2016). *The impact of open access on galleries, libraries, museums, & archives*. Smithsonian Emerging Leaders Development Program.





https://doi.org/10.15215/aupress/9781771994521.01

Klein, A., & Cardin, M. (2018). *Consommer l'information: De la gestion à la médiation documentaire*. Presses de l'Université Laval.

Lih, A. (2018). What are galleries, libraries, archives, and museums (GLAM) to the Wikimedia community? In M. Proffitt (Ed.), *Leveraging Wikipedia: Connecting communities of knowledge* (pp. 7–16). American Library Association.

Navarrete, T., & Villaespesa, E. (2020). Digital heritage consumption: The case of the Metropolitan Museum of Art. *Magazén*, *1*(2), 223–48. https://edizionicafoscari.unive.it/media/pdf/journals/magazen/2020/2/iss-1-2-2020_3AUKRGj.pdf

*The open definition: Defining open in open data, open content and open knowledge*. (n.d.). Open Knowledge Foundation. https://opendefinition.org/

Salgado, M., & Marttila, S. (2013). Discussions on inclusive, participative and open museums. In *NODEM 2013 conference proceedings*. Interactive Institute Swedish ICT, Sweden. http://repo.nodem.org/uploads/Discussions%20on%20inclusive,%20participative%20and%20open%20museums.pdf

Simon, N. (2010). *The participatory museum*. Museum 2.0.

Terras, M. (2015). Cultural heritage information: Artefacts and digitization technologies. In G. Chowdhury & I. Ruthven (Eds.), *Cultural heritage information: Access and management* (pp. 63–88). Facet.



https://doi.org/10.15215/aupress/9781771994521.01

# 6 Open Government
## A Wiki to Link Them All Together

Miguel Tremblay

Wikidata, the Wikimedia Foundation's free database, is an ideal convergence point for open government metadata from all over the world (Krabina & Polleres, 2021; Pellissier Tanon et al., 2016). Indeed, grouping together open data catalogues makes it possible to cross-reference the information in these datasets with each other but also with information from one hundred million elements already present on Wikidata (Hitz-Gamper et al., 2019; Krabina & Polleres, 2021). This turnkey solution offers an IT infrastructure as well as methods and software access for combining data; it can also count on the voluntary participation of an active community.[1] To our knowledge, this type of centralization, aimed at government open data catalogues, has not been explored in the scientific literature (Mora-Cantallops et al., 2019).

How might such centralization work in practice? We identify the first milestones for progress in this direction and also list a series of advantages and challenges for this solution. We take as an example a set of Government of Canada metadata that has already been uploaded to Wikidata. Inspired by this real example from the recent past, we explore the possibility and feasibility of adding other metadata from open data catalogues.

1 One of the special features of Wikidata is that it is based on a volunteer community. It was this community strength that led Google to migrate Freebase's data to Wikidata. See Pellissier et al. (2016).



https://doi.org/10.15215/aupress/9781771994521.01

## Open Government, Open Data

Moore's law states that the power of computer chips doubles every 2 years, and this also applies to computer data volumes. Faced with these challenges, governments are turning to cloud computing, taking the user to where the data are found rather than the other way round.[2]

At the same time, several public administrations are joining the open government movement[3] (Estermann, 2018). In 2011, the Government of Canada launched its first National Action Plan on Open Government. It was one of the first countries to join the Open Government Partnership (OGP), a multilateral partnership endorsed in 2011. As of June 2024, 75 countries and 150 local governments—representing more than two billion people—as well as thousands of civil society organizations are members of the OGP. This is a global movement.

The first area developed during the implementation of open government is access to data, commonly known as *open data*.[4]

## A Multitude of Data Portals in Public Administrations

Open data from public administrations is usually grouped together and made available on a single portal. These data are listed in a catalogue that can be searched using website-specific nomenclature and search tools (Tygel et al., 2016). The Government of Canada presents its catalogue of open data on its website, where approximately 40,000 metadata available can be searched.

Each government portal offers tools for identifying and sometimes visualizing datasets. These are often grouped by theme, such as health, culture, or technology (Tygel et al., 2016).

---

2 See, for example, the DIAS platforms, funded by the European Commission, and the Australian government's project Digital Earth Australia, which offers cloud computing space for the manipulation and posting of satellite data.

3 There are several definitions of *open government*. According to the Canadian government, open government is defined as "a governing culture that holds that the public has the right to access the documents and proceedings of government to allow for greater openness, accountability, and engagement" (Government of Canada, 2014).

4 Open data are "raw, non-nominative, royalty-free data produced or collected by a public or private organization, accessible to citizens via the Internet" (Office québécois de la langue française, 2013).



https://doi.org/10.15215/aupress/9781771994521.01

Government portals sometimes bring together descriptions of datasets from several administrations.[5] In Canada alone, there are at least 66 different open data portals.

There is a lack of uniformity in the interface, functionalities, and categories across these portals, many of which are only available in English, which limits both discoverability and access to data for people who do not speak this language.

To find a dataset specific to each government, users must therefore discover the portals, familiarize themselves with the search tools, and identify the dataset in question, and only if they are fluent in English. The opening up of data is a step in the right direction for open government, yet it has to be said that the existence of a multitude of portals creates silos restricting access to data.

## Centralizing Metadata on Wikidata

The Wikimedia Foundation was created in 2003, 2 years after the birth of Wikipedia. Its aim was to fund technical support for the free encyclopedia. Gradually, several projects were added, all with the aim of making knowledge accessible to everyone (Yoakim, 2020).

Wikidata was launched in 2012 by the German chapter of the Wikimedia Foundation. It is a free, multilingual, collaboratively edited knowledge base that provides answers to complex queries and is part of a context of linked data, the basis of the semantic web. Its content is licensed under the Creative Commons Zero (CC0) licence,[6] which encourages the free reuse of the knowledge and information it contains.

By June 2024, more than 10 years after its creation, Wikidata had more than 110 million items and more than 23,000 active users. Its data can be viewed using 50 external tools.[7] More than 500 articles have been published about Wikidata.

---

5 For example, the directory of the Government of Canada includes data from the Government of Alberta. The directory of the Government of Québec includes data from municipal administrations in Québec.

6 The CC0 license is a free Creative Commons license that allows copyright holders to waive as much of their rights as possible, within the limits of applicable law, in order to place their work as close as possible to the public domain.

7 The list of tools can be accessed on the following page: "Wikidata: Tools/Visualize Data."



https://doi.org/10.15215/aupress/9781771994521.01

Several open datasets (including Wikipedia's) have been integrated into Wikidata, in all languages. Wikidata is conducive to the centralization of government open data, not least because of its semantic web capabilities (Ajose-Ismail & Osanyin, 2019).

These capabilities are all the more interesting given that Wikidata contains information, for example, on administrative regions (1.9 million entries), chemical compounds (1.2 million), and communication channels (630,000 entries). Metadata on scientific publications account for over 30% of entries, with 22.6 million descriptive records ("Wikidata: Statistics," 2025).

Wikidata makes it possible to add structured data using scripts, and it offers a protocol (SPARQL) for searching for, adding, modifying, or deleting data. Wikidata constitutes a solid infrastructure capable of absorbing millions of items of metadata, as the example of scientific articles shows.

## Example of the Meteorological Service of Canada

Historically, the exchange of meteorological data has been an area where collaboration between countries and scientists has always predominated. Indeed, a weather forecast over more than a 48-hour period has to integrate data from all over the world. And as this is true everywhere on the planet, it encourages countries to exchange data since they all need data to make forecasts for their own territory. Even at the height of the Cold War, meteorological data circulated between Eastern Bloc and Western Bloc countries (Zillman, 2019). This is one of the reasons why this area of knowledge is particularly well suited to the Wikimedia Foundation's ideal of knowledge sharing.

In 2019, the Meteorological Service of Canada granted funding to Wikimedia Canada to upload weather station observation data and metadata to the Wikimedia Foundation ecosystem.[8] These data are uploaded to the Wikimedia Commons and Wikidata, respectively (see figure 6.1). The metadata includes a link to the corresponding weather station data in the commons. The metadata for Canada's weather stations are added to those for 12 other countries with more than 10 stations in Wikidata.

It is now possible to use the tools and exploit Wikidata's various calculation and query capabilities. For example, you can cross-reference station

8 A test was also undertaken using financial data from Austrian municipalities (Krabina & Polleres, 2021).



https://doi.org/10.15215/aupress/9781771994521.01

metadata and then create categories of stations by altitude. You can also cross-reference the metadata with other information already available in Wikidata and, for example, identify weather stations in Québec located within a radius of one kilometre of a bridge longer than 100 metres. The information on bridges already existed in Wikidata when the station metadata were imported. These few examples illustrate the power of Wikidata's search tools.

In connection with this project, Wikimedia Canada and Acfas welcomed two Institut de valorisation des données (IVADO) interns in the summer of 2021 from the Data Storytelling program.

Two articles were based on the collaborative work of Laurence Taschereau (UQAM, journalism) and Ali Akbar Sabzi Dizajyekan (Polytechnique Montréal, data visualization) on meteorological data. Their work is helping raise public awareness of the effects of climate change on the population. The first article deals with agricultural resilience (2021b) and the second with heat islands (2021a).

Wikimedia Canada and the World Meteorological Organization (WMO) are working together to find a way of uploading all the metadata from weather stations registered with the WMO to Wikidata. With a view to creating weather analysis products, exchanges are also taking place with users in Cameroon and Nigeria, countries where IT infrastructure capacity is less developed.

**Figure 6.1.**
*Map of existing or former Meteorological Service of Canada weather stations*

*Note:* Grouped in 500-metre altitude increments (a different colour for each increment).



https://doi.org/10.15215/aupress/9781771994521.01

## First Stage in Importing Data into Wikidata: Targeting Metadata and Standardization

Each open government follows a certain standardization to describe its datasets. The information often includes but is not limited to the licence, the administrative entity (government ministry) responsible for the dataset, the URL for downloading the data, and the date of the last update. The Open Data Charter is working to establish an international charter for open data, including interoperability, so that data can be compared and cross-referenced, regardless of the original source.

The aim is to convert catalogue metadata, which are standardized locally, into a form that corresponds to that metadata stored within the Wikidata database.[9] Some metadata lend themselves better than others to being imported into Wikidata, as they adopt an international standard as well as a local one. Converting them to a common format in Wikidata is therefore less complex.

Any international standard is supported by a community made up of users and contributors who are likely to take part in the project to adapt the catalogue for import into Wikidata. In addition to this meteorological data, which meet this criterion, georeferenced data serve as another example of standardized data. These metadata describe administrative areas (political and census) as well as routes for different means of transport, among others.

Taken together, these metadata could make it easier to deploy a host of applications on a global scale, which is currently impossible given the need for time-consuming manual processing of metadata.

## Benefits of Centralizing Open Data Catalogues

Open governments have centralized the data of their ministries and agencies under a single portal, imposing a standardization of dataset metadata, as

9 Some open software applications have been designed to host open data catalogues (Socrata, CKAN). More than 250 organizations publish their data using these open software applications using metadata standards such as DCAT or standards published by the W3C Government Linked Data (GLD) task force (Neumaier et al., 2016; Tygel et al., 2016). However, none of these data catalogues are retranscribed directly into Wikidata.



https://doi.org/10.15215/aupress/9781771994521.01

explained earlier.[10] The centralization of Wikidata[11] is a simple extension of this principle: creating a single place where not just one but all the catalogues of open governments can be found, making them accessible to citizens all over the world.

### Upgrading Existing Capacity

Copying metadata onto Wikidata is not intended to replace the various governmental open data portals. Rather, this is a way to add capacity to existing portals. Wikidata should not become a single point of failure[12] by being the official metadata repository. Governments should ideally retain an authoritative copy of all their metadata.

Governments remain responsible for any open data that remains on the original IT server. This makes it possible to carry out performance measurements and benefit from the usual control measures. Wikidata, like Wikipedia, requires that the sources of the information be cited ("Help: Sources," 2026). This feature preserves the link between the copy on Wikidata and the original government source.

Once the metadata have been copied into Wikidata, they have a broader reach because they are described in more than one place on the internet, where the metadata of other governments are colocated.

### Monitoring and Revising Metadata

The principle of the wiki is to preserve each version of the documents and make it possible to view and monitor changes. Wikidata applies this principle. The person responsible for metadata can therefore have all the entries corresponding to a catalogue in their watch list associated with their Wikidata account. They can be notified immediately by email of any changes and, if

10 Hitz-Gamper et al. (2019) have also drawn up a list of advantages and disadvantages but in a more general framework, focusing instead on government-linked data in the broadest sense. Here, we focus on the specific case of Wikidata.

11 Although the semantic web was designed to operate in a decentralized fashion, there remain challenges of nomenclature among databases (Ajose-Ismail & Osanyin, 2019), and harmonization could be undertaken in Wikidata.

12 A single point of failure (SPOF) is a point in a computer system on which the entire system depends and whose failure brings about the complete shutdown of the system.



https://doi.org/10.15215/aupress/9781771994521.01

necessary, revise them. Wikis therefore allow anyone to contribute, but all modifications can be checked.

The import of the Meteorological Service of Canada's weather observation data has not been subject to any unnecessary modifications or vandalism over the last 5 years. Instead, data custodians have been notified of errors in the data, which gives them the opportunity to make corrections.

### Multilingual and Standardized Information

The Wikidata interface is available in over 300 languages, each with a community that discusses the terms to be used to describe encyclopedic entries. This is more than a translation, as demonstrated by the deployment of the interface in the Atikamekw language, which involved the community and the custodians of the Atikamekw language in choosing or creating new words (by Casemajor et al., 2017). In addition to the interface, the elements themselves can be translated into more than 300 languages. Metadata for the label "weather stations" can, for example, be seen in Arabic (محطة رصد جوي) or Belarusian (метэаралагічная станцыя).

The use of alternative names (aliases) to designate items in each language is a feature that distinguishes Wikidata from Wikipedia. This feature allows different nomenclatures to be reconciled in the management of labels, avoiding a problem identified in other open databases (Tygel et al., 2016).

Wikidata's multilingualism is seen as an added value for organizations wishing to share their data (Evans, 2024). Wikidata's multilingual interface is a considerable advantage over open government portals, which are generally available in only one language (two languages in the case of Canada). The interface also encourages collaboration between people who do not share a common language. Speakers of the world's different languages can search for and use the datasets of their choice using an interface in their own language.

## Cloud Infrastructure and Community Already in Place

Wikidata is a database available on a reliable infrastructure, financed and managed by the Wikimedia Foundation. The platform is regularly monitored, and capacities are increased as required, which includes integrating new technologies. Governments wishing to contribute their metadata will benefit from existing servers, and users will have free access to the data.



https://doi.org/10.15215/aupress/9781771994521.01

Although it is difficult to estimate how much Wikidata would contain if governments added their metadata—consider, for example, that the Government of Canada has 40,000 entries in its open data catalogue—it is reasonable to think that the infrastructure would be able to tolerate these massive imports. By way of comparison, the number of entries in Wikidata for scientific articles is over 22 million. Following a massive import of elements, it is vitally important to maintain dialogue with the Wikidata community to avoid any spontaneous, sharp reaction. The incremental nature of the import process will leave room for negotiation in order to satisfy the community.

The results of user queries in Wikidata can be displayed in a number of ways: in tables, graphs, or maps. The results can be exported in a variety of formats, and the data can be reused in other systems.

Finally, there are a number of external applications for carrying out searches or calculations in Wikidata. A directory page lists these applications, which often work by thematic field ("Wikidata: Tools/Visualize Data," 2026). It is possible to develop specific tools for open government metadata.

## The Challenges of Centralization

The first and potentially most important step in copying metadata into Wikidata is to create a correspondence between the typology of open data catalogues and their representation in Wikidata. This requires knowledge and expertise in the content of both the catalogues and Wikidata. Such a task requires human resources and time, and it can extend to the creation of properties in Wikidata in collaboration with the community.

Once the matches have been created, the metadata is uploaded to Wikidata, ideally using automated scripts. This task requires programming skills, which are key when it comes to connecting a catalogue to Wikidata.

When the catalogue is updated on the government portal side, the information needs to be synchronized in Wikidata, which means uploading the metadata that have been added or modified. This can be done at regular intervals (every week, month, or year) or as and when required. Conversely, a feedback mechanism should be put in place to communicate the changes made in Wikidata to those responsible for government metadata. The community can detect errors or incongruencies in the metadata once they are being used (Krabina, 2023). The element tracking system, already present in Wikidata and mentioned earlier, can be used to provide this feedback.



https://doi.org/10.15215/aupress/9781771994521.01

The key to the massive import of government data into Wikidata is therefore regular communication between the community and its government counterparts. The ideal is to obtain the support of a community of practice with which to establish links of trust. The source code of the scripts should be free to share best practices.

### Support for Users

Greater sharing of metadata could lead to more users making use of open data. In fact, this is the ultimate goal. The downside of an increase in popularity is an increase in the demand for technical support for users, a demand that governments should, in theory, meet with appropriate tools.

### SPARQL or the Art of the Query

The query language in Wikidata, called SPARQL, requires an advanced level of expertise to be used to its full potential and to successfully identify the information being sought. Although efforts are being made to simplify the interface, notably through the online Wikidata Query Builder tool, specific computer programming skills are required to write complex queries. That said, it is possible to ask for help from the community of users who have this knowledge, but it may take several weeks to get a response.

In order to be able to cross-reference metadata in open data catalogues with other data in Wikidata, metadata managers need to develop documentation that includes typical examples of SPARQL queries. This documentation can be hosted directly in Wikidata.

Generative artificial intelligence tools such as ChatGPT have greatly helped democratize the writing of SPARQL queries. By asking a question in natural language, the tool translates this question into SPARQL code adapted for Wikidata. This allows users, even when they lack programming skills, to easily access the data they are looking for and take full advantage of Wikidata's capabilities (Lubiana, 2023).

To optimize the user experience, customized tools can be created within the Wikimedia Foundation's ecosystem or elsewhere on the web. Users then benefit from an intuitive interface that is easier to use than formulating a query in code.



https://doi.org/10.15215/aupress/9781771994521.01

### Metadata Licence

The content of Wikidata is licensed under the CC0 licence, under which the owners of copyright- or database-protected content waive those interests in their works and thereby place them as completely as possible in the public domain.

Governments uploading metadata to Wikidata must waive their copyright. This should not, in principle, be an issue, as we are talking about the description of datasets, not the data itself. Metadata represent a factual description of information rather than any added value related to the creation of the data. There is often little added value in this information. However, administrative pitfalls are to be expected depending on the relevant legislation in place.[13]

### No Standardized Categories

Most open data portals present datasets by theme. Yet there is no standardized categorization (Pellissier Tanon & Kaffee, 2018). This lack of conceptual uniformity is compounded by the problem of translating them into different languages (Farda-Sarbas & Müller-Birn, 2019; Pellissier Tanon & Kaffee, 2018).

Metadata must be classified according to properties that are specific to Wikidata. This classification effort raises challenges for nomenclature, logic, and ontology. We can work in collaboration with a community of fields of interest to try to answer these questions and to help our thinking evolve. There is, for example, a community interested in the performing arts that could answer questions relating to the cultural milieu ("Wikidata: WikiProject Performing Arts/fr," 2023).

## Granularity

When it comes to managing metadata, the question of granularity is omnipresent. At what level should datasets be represented? By grouping them together using the lowest possible level of abstraction or, on the contrary, by going to the highest possible level of granularity? Both solutions are equally valid.

To take the example of weather stations again, should we publish a single metadata entry containing all the 8,000 or so stations listed, or should we

13 In Canada, government productions are subject to Crown copyright and could be subject to a special copyright regime. For more information on the subject, see https://ised-isde.canada.ca/site/ised/en/terms-and-conditions/about-copyright.



https://doi.org/10.15215/aupress/9781771994521.01

publish one entry per station? Or both? The difference in the volume of metadata to be entered in this case is several orders of magnitude.

One of the advantages of the higher level of metadata granularity is that Wikidata's tools can be used to their full potential to identify data. This requires extensive work to establish the metadata correspondence between the government catalogue and Wikidata.

It would be surprising if a general rule could be identified. Metadata owners must make choices. As with categorization, this could be undertaken in consultation with the fields of interest community, where such a community exists.

## Conclusion

Initial interest in the meteorological metadata project was surprising. We were contacted by governmental and paragovernmental organizations as well as civil society institutions interested in open knowledge (libraries, NPOs, private companies). This growing interest has encouraged us to extend our thinking to other datasets, since the principles underlying the massive import and dissemination of meteorological data can be applied to them as well.

The notoriety of Wikipedia and other Wikimedia Foundation projects promotes understanding and acceptance by government officials. The attractiveness of using existing IT resources is an asset, where their methods of access, search, and dissemination are recognized by public administrations and open data users in general.

An infrastructure of this kind also facilitates collaboration between different levels of government or even between states because the issues of infrastructure costs and, above all, the policy of managing and funding this infrastructure are in the hands of a third party: the Wikimedia Foundation. All that remains for governments to do is invest in the human capital needed to use this infrastructure and, of course, to answer the ontological questions that cannot be ignored.

We have sketched out the advantages and challenges of centralizing metadata in Wikidata. The next step will be to identify the metadata whose sharing could benefit society as a whole. We will then need to invest in the longer term to explore and mark out the path leading to greater dissemination of open data.

Discoverability issues are the main obstacles to the dissemination of open data. Investing in new solutions and joining a community or existing projects



https://doi.org/10.15215/aupress/9781771994521.01

are actions in line with the values underpinning the open democracy movement. All that remains is to take the second step.

Generative AI tools, such as ChatGPT, will help mine the data to answer various questions. However, for these tools to be truly effective, it is essential that the basic information be accessible. It is crucial that these data are reliable and sourced to ensure that the answers provided are based not solely on statistical inference but on accurate and verified data. So accessibility and quality of data are essential prerequisites for taking full advantage of the capabilities of generative AI in the management and analysis of open data.

## References


Ajose-Ismail, B. M., & Osanyin, Q. A. (2019). A semantic approach for facilitating search for discovery of open government data on open data portals. In *Book of proceedings of the 4th National Development Conference of the School of Pure and Applied Science, The Federal Polytechnic Ilaro, Ogun State, December 2–5, 2019* (pp. 249–56). https://eprints.federalpolyilaro.edu.ng/1645/

Casemajor, N., Béland, J.-P., & Rochon, B. (2017). Comment dit-on page d'accueil? Incorporer le savoir autochtone dans Wikipedia. *Diff*. https://diff.wikimedia.org/fr/2017/05/15/creer-dune-encyclopedie-atikamekw-a-partir-de-zero/

Estermann, B. (2018). Development paths towards open government—an empirical analysis among heritage institutions. *Government Information Quarterly*, *35*(4), 599–612. https://doi.org/10.1016/j.giq.2018.10.005

Evans, J. (2024). *Aligning and sharing data with Wikidata—a feasibility study*. Knowledge Commons. https://works.hcommons.org/records/5ym2e-bap78

Farda-Sarbas, M., & Müller-Birn, C. (2019). *Wikidata from a research perspective—a systematic mapping study of Wikidata*. arXiv. https://doi.org/10.48550/arXiv.1908.11153

Help: Sources. (2026, January 19). In *Wikidata*. https://www.wikidata.org/w/index.php?title=Help:Sources&oldid=2456148705

Hitz-Gamper, B. S., Neumann, O., & Stürmer, M. (2019). Balancing control, usability and visibility of linked open government data to create public value. *International Journal of Public Sector Management*, *32*(5), 451–66. https://doi.org/10.1108/IJPSM-02-2018-0062

Krabina, B. (2023). Building a knowledge graph for the history of Vienna with semantic MediaWiki. *Journal of Web Semantics*, *76*. https://doi.org/10.1016/j.websem.2022.100771

Krabina, B., & Polleres, A. (2021). *Seeding Wikidata with municipal spending data*. ResearchGate. https://www.researchgate.net/publication/355205810_Seeding_Wikidata_with_Municipal_Spending_Data




https://doi.org/10.15215/aupress/9781771994521.01

Lubiana, T. (2023). *SPARQL queries, GPTs and large language models—where are we currently?* Wisecube. https://www.wisecube.ai/blog/sparql-queries-gpts-and-large-language-models-where-are-we-currently/

Mora-Cantallops, M., Sánchez-Alonso, S., & García-Barriocanal, E. (2019). A systematic literature review on Wikidata. *Data Technologies and Applications*, *53*(3), 250–68. https://doi.org/10.1108/DTA-12-2018-0110

Office québécois de la langue française. (2013). *Données ouvertes*. Vitrine linguistique. https://vitrinelinguistique.oqlf.gouv.qc.ca/fiche-gdt/fiche/26519745/donnees-ouvertes

Pellissier Tanon, T., & Kaffee, L.-A. (2018). Property label stability in Wikidata: Evolution and convergence of schemas in collaborative knowledge bases. In *Companion of the Web Conference 2018 on the Web Conference 2018 (WWW '18)* (pp. 1801–3). ACM Press. https://doi.org/10.1145/3184558.3191643

Pellissier Tanon, T., Vrandečić, D., Schaffert, S., Steiner, T., & Pintscher, L. (2016). From Freebase to Wikidata: The great migration. In *Proceedings of the 25th International Conference on World Wide Web* (pp. 1419–28). International World Wide Web Conferences Steering Committee. https://doi.org/10.1145/2872427.2874809

Taschereau, L., & Sabzi Dizajyekan, A. A. (2021a). Des inégalités climatiques à rendre insomniaque. *Acfas Magazine*. https://www.acfas.ca/publications/magazine/2021/11/inegalites-climatiques-rendre-insomniaque

Taschereau, L., & Sabzi Dizajyekan, A. A. (2021b). La résilience agricole se trouve à nos pieds . . . dans le sol. *Acfas Magazine*. https://www.acfas.ca/publications/magazine/2021/11/resilience-agricole-se-trouve-nos-pieds-sol

Tygel, A., Auer, S., Debattista, J., Orlandi, F., & Machado Campos, M. L. (2016). Towards cleaning up open data portals: A metadata reconciliation approach. In *2016 IEEE tenth international conference on semantic computing (ICSC)* (pp. 71–78). IEEE. https://doi.org/10.1109/ICSC.2016.54

Wikidata: Statistics. (2025, August 31). In *Wikidata*. https://www.wikidata.org/w/index.php?title=Wikidata:Statistics&oldid=2398684746

Wikidata: Tools/visualize data. (2026, January 19). In *Wikidata*. https://www.wikidata.org/w/index.php?title=Wikidata:Tools/Visualize_data&oldid=2455976888

Wikidata: WikiProject performing arts/fr. (2023, December 3). In *Wikidata*. https://www.wikidata.org/w/index.php?title=Wikidata:WikiProject_Performing_arts/fr&oldid=2023250515

Yoakim, W. (2020). Wikipédia, Wikimedia Commons et Wikisource, un Eldorado de visibilité. *Archives*, *48*(2), 41–81. https://doi.org/10.7202/1067524ar

Zillman, J. W. (2019, November 27). Origin, impact and aftermath of WMO resolution 40. *World Meteorological Organization*. https://wmo.int/media/magazine-article/origin-impact-and-aftermath-of-wmo-resolution-40



https://doi.org/10.15215/aupress/9781771994521.01

Focus III

# Literacies

*This page intentionally left blank.*

# 7 Public Knowledge During the COVID-19 Infodemic

## Health Literacy and the Effect of Wikipedia

Denise Smith

The COVID-19 pandemic emerged at a time when information is produced with a frequency and volume never seen before. Simultaneously, the 24-hour news cycle and myriad social media platforms effectively monopolized our attention to focus on the pandemic. These two elements of the information landscape worked in conjunction to provide the public with an opportunity to witness the knowledge of a new infectious disease evolve in real time. Public exposure to typical processes in the development of scientific or medical knowledge—namely, the ups and downs of learning in science—has sown significant seeds of distrust among the public.

Early in the COVID-19 pandemic, politicians and public health officials openly blamed misinformation (CBC Ideas, 2021; Patel, 2021) to the extent that the term *infodemic* has experienced a resurgence since its initial use in 2003 during the SARS epidemic (*Words We're Watching*, n.d.).

Although the ubiquity of misinformation about COVID-19 has presented its challenges, this chapter will highlight that, beyond misinformation, the pandemic has amplified an overlooked but important gap in the general populace: health literacy. While a COVID-19 misinformation crisis might have posed a threat to public health (Islam et al., 2020), it is unreasonable to expect to erase misinformation from the landscape. A more appropriate response is to equip the public with the skills and tools necessary to identify and detect misinformation. However, a mass public education initiative



https://doi.org/10.15215/aupress/9781771994521.01

for health literacy that aims to educate the populace on how to critically appraise health information in various contexts predictably presents its own unique set of challenges. This is where Wikipedia enters the spotlight.

Wikipedia emerged as a leader in health information communication and has situated itself as a key resource for COVID-19-related information (Benjakob, 2020; Benjakob et al., 2022; Cohen, 2020). In addition to its potential to provide relief from the misinformation crisis, it has also been positioned as a documentarian of the pandemic itself (Harrison, 2020; Yang & Tanaka, 2023). It has also been explored as a resource for epidemiological surveillance and measuring public interest in the disease (Chrzanowski et al., 2021; Gozzi et al., 2020).

We know the public relies heavily on the internet for health information (Heilman et al., 2011) and that health literacy among the public is varied and influenced by many variables. Wikipedia's potential to balance health literacy inequities among the general populace must not be overlooked. Wikipedia is frequently updated, can be used to easily understand an evolving situation, and provides information that can either make sense of the situation or simply inform one's decision-making. The reader can reasonably expect objective, evidence-based information in easy-to-read plain language. We also know that Wikipedia and its community-monitored editing guidelines work to create a source of medical evidence that summarizes high-quality medical evidence from reliable and reputable sources. As such, the following four arguments will guide this chapter:

1. Health literacy is a mediating determinant of health.
2. The public relies on the internet for health information, but they need advanced health literacy skills to appraise and make appropriate use of the information they find.
3. Health literacy skills vary, and inequities are influenced by socioeconomic variables.
4. Wikipedia can mitigate the requirement for critical and interactive health literacy to discern reliable evidence from misinformation.

This chapter will introduce the concept of health literacy and discuss why Wikipedia has a role in addressing health literacy inequities highlighted by the COVID-19 pandemic. The author will survey Wikipedia's position as a key pandemic information resource and explore whether this appraisal applies in the Canadian context. The chapter will conclude with a summary of



https://doi.org/10.15215/aupress/9781771994521.01

what is currently unknown about Wikipedia as a health information resource and describe forthcoming research into the unexplored relationship between health information consumers and Wikipedia.

## Health Literacy

Health literacy is not a new concept, but with the emergence of COVID-19, it has been thrust into the spotlight under the guise of conversations about misinformation. Health literacy is frequently reduced to what is considered *functional* health literacy: the ability of a patient to apply basic literacy skills to be able to engage in rudimentary health behaviours, such as understanding appointment cards or prescription directions. However, this chapter will focus on *interactive* and *critical* health literacy (Nutbeam, 2000, 2008). Interactive health literacy refers to the ability to "extract information and derive meaning from different forms of communication, and to apply new information to changing circumstances" (Nutbeam, 2000). Critical health literacy applies more advanced cognitive skills to analyze information and then use it to "exert greater control over life events and situations" (Nutbeam, 2000). Beyond the ability to follow directions, the more advanced one's health literacy is, the more autonomy they have in finding, understanding, evaluating, and applying health information in their daily life.

People with advanced health literacy skills have the tools necessary to identify what kind of health information they need, locate and retrieve information that meets their needs, and then discern whether that information is reliable and credible. They may also have the skills to triangulate the information they encounter against other sources of information and choose the most reliable evidence. In the context of COVID-19, this means an individual with interactive and critical health literacy skills can clearly identify what questions they have, then take an intentional path toward finding information that answers their questions. Alternatively, many individuals simply encounter COVID-19 information without looking for it. Whatever the path to information discovery, the individual must be able to verify its credibility and reliability, apply this new information to what they already know, and be able to extrapolate meaning from it or understand its potential impact.

Health literacy sits at the core of the COVID-19 infodemic. The crisis, I argue, is not only the prevalence of misinformation but also that the public, broadly speaking, is not equitably equipped with adequate health literacy skills


https://doi.org/10.15215/aupress/9781771994521.01

to be able to identify and use appropriate health information. Further, health literacy has been established as a mediating social determinant of health, wherein recent evidence has demonstrated that strengthening health literacy, alongside other initiatives, could reduce health disparities and promote health equity (Nutbeam & Lloyd, 2021). A key finding from a literature review suggests that one's health literacy is reflective of their experience with the use of information, their comfort with technology, their educational background, and their socioeconomic status (Martzoukou & Sayyad Abdi, 2017). This review finds that "the internet is an important source of health information across different demographics" and the need for health literacy increases as public access to online health information also increases (Martzoukou & Sayyad Abdi, 2017). These findings align with what is currently known about the digital divide and how existing social structures, alongside socioeconomic status, can influence not only an individual's access to adequate health information but also how they perceive their ability to access such information (Bodie & Dutta, 2008; Smith, 2021).

## Wikipedia and COVID-19: Filling the Health Literacy Gap

When it comes to levelling out the varied health literacies of individuals, Wikipedia has promise. In interactive and critical health literacy, advanced cognitive skills are required so the information seeker or user can adequately evaluate the quality, relevance, and utility of the information at hand. In the context of COVID-19, anyone with an internet connection has probably experienced information overload. New knowledge related to the pandemic is generated and shared at a velocity that outpaces a person's ability to keep up. Add to this the noise of misinformation that can drown out methodologically sound scientific knowledge, and it is easy to picture an overwhelming experience that makes it difficult for anyone, including those with adequate or advanced health literacy, to make sense of the pandemic, the virus, or the disease.

Despite best efforts from well-meaning organizations and individuals to create high-quality and credible consumer health websites (e.g., Health Canada, MedlinePlus, Mayo Clinic, NHS), Wikipedia is one of the most frequently accessed health information resources on the internet (Birkun, 2024; Heilman et al., 2011; Heilman & West, 2015; Shafee et al., 2017). It has been situated as a well-used and valuable health information resource in various contexts for



https://doi.org/10.15215/aupress/9781771994521.01

the public, physicians, teachers, students, and researchers alike (Smith, 2020). Early research into understanding the use of Wikipedia for health information indicates that the general familiarity of Wikipedia—internet users know it and recognize it when scanning search engine results—lends itself to the broad reach and readership of Wikipedia's health and medical content (Huisman et al., 2020; Smith, 2023a).

The quality of Wikipedia's medical content cannot be assessed in whole because "Wikipedia's articles are individual pieces of a larger whole . . . a mosaic . . . where some pieces are more complete than others" (Smith, 2020), but its potential to contribute to the closure of health literacy gaps is based on its self-governed community guidelines for content creation and maintenance. Volunteer contributors must comply with stringent community guidelines to which they are held accountable by other contributors. Among these guidelines are the requirement to maintain a neutral point of view ("Wikipedia: Neutral Point of View," 2026) within the article. NPOV can minimize the requirement for readers to assess the article for bias or seek out counterpoints. Ideally, all available information is in the article, giving readers a clear picture of what knowledge has been generated and is presented in a tone that does not require the reader to appraise the article for potential bias.

The requirement to verify information added to the article with high-quality, reliable sources with even more specific guidelines for reliable sources of medical information ("Wikipedia: Reliable Sources," 2026; "Wikipedia: Identifying Reliable Sources," 2025) limits the reader's need to evaluate the quality of sources used to generate the Wikipedia article. Evaluation of news media and academic sources cited in an article is presumed to be undertaken by the contributor adding them. Further, the credibility or reliability of sources is negotiated by the editing community, creating an information environment for the reader wherein the reader consumes information that has already undergone a certain degree of scrutiny. Wikipedia is also not the place to create new knowledge ("Wikipedia: No Original Research," 2026). As such, it situates itself as a valuable resource to learn about existing knowledge without requiring the reader to rigorously appraise its content in ways primary or secondary studies must be appraised: for the possibility of bias or methodological rigour.

For COVID-19, Wikipedia is a single resource that can be consulted to get a sense of what is known about the virus, the vaccine, the pandemic, and more. Early in the pandemic, it garnered media attention for how quickly new


https://doi.org/10.15215/aupress/9781771994521.01

and evolving information was added to relevant articles, the expanse of its coverage of the global pandemic, the dedication of its editors to updating relevant articles (Bedirian, 2021), and the importance of protecting its approach to content creation "in service of the public interest" (Keton & Humborg, 2021). In June 2021, a Canadian news media source released an article refuting Wikipedia's decades-long reputation as unreliable, suggesting that all information sources should be subjected to equal scrutiny (Ibrahim, 2021). Existing knowledge about COVID-19 that is summarized in Wikipedia is monitored by other editors, updated as our knowledge of the virus evolves, and supported with vetted citations that must meet reliability guidelines. Evidence emerged in 2021 that the resource experienced an increase in volunteer editors and in productivity in 2020, after the COVID-19 virus was declared a global pandemic (Ruprechter et al., 2021).

Wikipedia can mitigate inequities in health literacy with respect to one's ability to "discriminate between different sources of information" and appraise health information about risks to personal health (Nutbeam & Lloyd, 2021), but it cannot resolve health literacy inequities. It is not, nor does it claim to be, a resolution for inequities in health literacy among the general populace. However, its editorial model situates it as a promising, accessible, consumer health resource that, in its best form, stands to provide impartial fact-based medical information that has been scrutinized by an internal peer-review system to ensure verifiability and reliability—removing much of that mental labour for the consumer (Smith, 2023b). Although it remains the responsibility of the consumer to identify that each page in Wikipedia must be evaluated individually (Epstein, 2022), the work within the scope of WikiProject COVID-19 eases this responsibility.

One may argue that other consumer health websites about COVID-19 are also frequently updated, cite high-quality evidence, and use plain, accessible language. While this may be true, the simple act of the reader navigating to Wikipedia over and above other resources must not be ignored. The fact is, people commonly access Wikipedia even if it is not necessarily their preferred information source (Sebelefsky et al., 2015; Smith, 2023a). It is often selected from search engine results in health information searches because it is familiar, easy to read, and comprehensive (Huisman et al., 2020; Smith, 2023b). This finding aligns with theories of human information behaviour research, specifically, the principle of least effort, which is based on the finding that "people invest little in seeking information, preferring easy to use, accessible



https://doi.org/10.15215/aupress/9781771994521.01

sources to sources of known high quality that are less easy to use and/or less accessible" (Fisher et al., 2005, p. 4).

Finally, with the great strides Wikipedia has made in documenting the evolution of a global pandemic and summarizing the most currently available knowledge as the situation evolves, inequities within Wikipedia's content must not be ignored. Wikipedia's coverage of COVID-19 has benefited a broad range of people, but much like the real world, it is built by a group of privileged individuals with similar life experiences. It is widely acknowledged that editors of English Wikipedia are, generally, a homogenous group with limited diversity, which means there are important voices missing from the editing community.

One important example is the poor documentation of the unique experiences of Canada's Indigenous populations in the pandemic article for Canada. Many of the pandemic-related articles for Canada's provinces or territories have limited or no information about First Nations, Inuit, or Métis populations, even though these populations were prioritized for early vaccination, continue to face unique challenges with controlling transmission, and in some northern or remote communities, struggle with limited capacity to manage an influx of new cases. There are several contributing factors to this shortcoming that, although important for discussion, are beyond the scope of this chapter. Examples include a lack of Indigenous representation among Canada's Wikipedia contributors or incompatible perceptions of what it means for a source to be reliable between "Wikipedia: Reliable Sources" / "Wikipedia: Identifying Reliable Sources (Medicine)" and Indigenous ways of knowing and Indigenous ways of knowing.

## Wikipedia and COVID-19 in the Canadian Context

In a time of heightened need for high-quality, reliable, unbiased health information that is easily accessible by all people, including individuals with limited health literacy skills, Wikipedia can be a useful tool for public health communication. But in practice, has this held true for Canada? Let's investigate two Wikipedia articles relevant to COVID-19 in the Canadian context.

By December 2020, 6,950 articles were created about COVID-19 in 188 languages with contributions from more than 97,000 editors. These pages received close to 580 million pageviews, with a record of more than 10 million pageviews on a single day—March 23, 2020 (*Wikipedia and COVID-19*, 2020).


https://doi.org/10.15215/aupress/9781771994521.01

As of June 2024, more than 2,400 English-language articles in Wikipedia were related to COVID-19 or within the scope of "Wikipedia: WikiProject COVID-19" (2021). Relevant articles include the major topics, hereafter referred to as the "big four": "COVID-19" (the disease), "SARS-CoV-2" (the virus), "COVID-19 Vaccine," and the "COVID-19 Pandemic." Cumulatively, these four articles were viewed over 60 million times between January 2020 and December 2022. They averaged more than 86,000 views per day in the first year of the pandemic. These pageview counts do not include the various iterations of these articles in languages other than English. For example, French- and Spanish-language articles about the same topics have cumulatively received more than 7.7 million and 21.2 million pageviews from January 2020 through December 2022, respectively.

The "COVID-19 Pandemic in Canada" article is semiprotected, meaning only registered editors who have made at least 10 edits are preauthorized to contribute. It has been assigned a quality score of B-class and is rated as mid-importance by WikiProject COVID-19. As of June 2024, the article, which received the bulk of its contributions near the beginning of the pandemic, was flagged because of a need for more recent updates. It was created on February 26, 2020, before the World Health Organization (WHO) declared the outbreak of the novel coronavirus a pandemic. As such, it was originally titled "2020 Coronavirus Outbreak in Canada." According to the page statistics as of June 2024, 604 editors contributed to the article more than 4,500 times. It has been viewed more than 1.8 million times; most views were between article creation and December 2022 (1.7 million). Early in the pandemic (May 2020), its pageviews peaked with an average of 10,000 views per day. The article is nearly 6,000 words long and cites 295 unique references. This page was also a source of data for Google. A search for "COVID-19 pandemic in Canada" in November 2021 yielded statistics boxes about new and active cases generated with data from Wikipedia in conjunction with other sources.

Ontario is Canada's most populated province or territory. For some Canadian provincial context, the "COVID-19 Pandemic in Ontario" article is also semiprotected. It has also been assigned a quality score of B-class and rated as low importance by WikiProject COVID-19. This article was created on March 13, 2020, 2 days after the WHO declared the outbreak a global pandemic. Based on data from the page statistics, 344 editors have contributed to the article more than 6,200 times. It has been viewed more than 970,000 times, and by May 2020, it had averaged more than 4,800 views per day. As



https://doi.org/10.15215/aupress/9781771994521.01

of June 2024, the article is 4,500 words long and has citations to 196 unique references. Like "COVID-19 Pandemic in Canada," a November 2021 Google search for "COVID-19 pandemic in Ontario" yielded statistics boxes generated with data from Wikipedia, in conjunction with other sources, such as Our World in Data.

It seems safe to say that Wikipedia was frequently used for information about the pandemic in the Canadian context as well as the more specific provincial context of Ontario. This claim can be supported by the active pageviews recorded for the relevant Wikipedia articles but also by more passive user engagement with Wikipedia, wherein Google search results draw data from Wikipedia. While these two Wikipedia articles indicate some level of engagement within the Canadian context, these articles are limited to summarizing knowledge about the pandemic. It is probable that Canadians are also among the users frequenting the "big four" articles. Some evidence to suggest this can be found in the pageview statistics for "COVID-19 Vaccine." Although not solely responsible for an uptake in pageviews, in late May 2021, the Ontario government announced it would shorten the interval between first and second doses from 3 months to 8 weeks (CBC News, 2021). Internationally, the Johnson and Johnson vaccine was approved in the UK (Taylor, 2021), and the US media introduced the idea of employer vaccine mandates (Diaz, 2021). While the "COVID-19 Pandemic in Canada" and "COVID-19 Pandemic in Ontario" articles remained relatively stable in their pageviews at this time, the COVID-19 vaccine article experienced a significant spike in pageviews, jumping from 20,033 on May 28, 2021, to 47,282 on May 29, 2021. Any or all of these changes to the vaccine landscape, covered by the media, could have contributed to the spike in pageviews, but at best, this would indicate a positive correlation. A causal relationship cannot be determined because the pageview data lacks context.

Wikipedia's data and statistics, while helpful, are limited in that they only provide insight into engagement insofar as that engagement can be measured by pageview activity or contribution activity. It's easy to argue that since Wikipedia was accessed frequently, it has a role to play in relieving some of the inequities in health literacy skills that can perpetuate the infodemic. However, if we don't know how users engage with Wikipedia's content, such as what they do with the information they access, if anything, our insight into Wikipedia's potential to provide relief to the infodemic is limited to drawing


https://doi.org/10.15215/aupress/9781771994521.01

connections between what is known about health literacy and data that can be difficult to contextualize.

## What We Don't Know

There is limited research investigating why or how people use Wikipedia's health and medical content. Theories in information behaviour or information practices might be helpful in the development of our understanding of what motivates an individual to seek out or use health information encountered online. In 2012, Johnson and Case described "white, middle-aged women who are members of high socio-economic status (SES) groups and also highly educated" as the typical profile of a high information seeker (Johnson & Case, 2012). This profile is an important consideration when we consider Nutbeam and Lloyd's (2021) correlation between health literacy and social status. As important as it is to develop a richer understanding of how people use Wikipedia's health and medical content, it is equally important to acknowledge that any exploration into how people use Wikipedia is limited to the segment of the population who have the privilege of access to this technology in addition to adequate literacy and reading comprehension skills.

Health information behaviour (HIB) can be defined as "the *key variables* that influence the seeking, receipt, avoidance, sharing, management or use of health information and the outcomes that result from the behaviour" (Smith, 2021, emphasis added). One study serendipitously encountered rich data about Wikipedia in the context of using it for health information (Huisman et al., 2020), but at the time of writing, the only intentional investigation of this nature was conducted by the author of this chapter (Smith, 2023a). This study conducted interviews with individuals who had used Wikipedia's health and medical information. Through the voices of Wikipedia users, it aimed to add some qualitative context to Wikipedia's usage. The results do not contribute new knowledge to our current understanding of health literacy. Nor do they position Wikipedia as a possible source of relief for some of the health disparities that inequitable health literacy levels can contribute to. However, it does provide one of the first steps in understanding how Wikipedia's health content is used and why it is used, and it has the potential to possibly shed more light on what is already known about health literacy and health information behaviour.



https://doi.org/10.15215/aupress/9781771994521.01

## Conclusion

Wikipedia is not the solution to health literacy inequities, nor is it, by extension, a solution for disparities in health and health care that are driven by socioeconomic factors. It does not generate health literacy skills, but it can reduce some barriers to health information access created by inadequate health literacy.

It is already known that health literacy is a determinant of health and that one's health literacy is often influenced by various social or economic variables. It is proposed here that, given what we know about the factors that contribute to individuals' advanced health literacy skills, Wikipedia's availability and accessibility to the public, its editorial processes, and its emergence as a health information leader in the COVID-19 pandemic, Wikipedia has an important role to play in ensuring equitable access to important health information during a time when socioeconomic disparities have been amplified. Wikipedia's "big four" articles related to COVID-19, in conjunction with its frequently accessed content and dutiful coverage of the pandemic in the Canadian context, provide some evidence that Wikipedia has positioned itself highly as a reliable source of plain language health information supported with high-quality evidence from reputable sources.

The very nature of how Wikipedia content is created and curated situates it as a key player in the battle against misinformation, wherein our biggest challenge is not necessarily the ubiquitous existence and creation of misinformation but rather a public audience that is imbalanced with respect to its capacity to critically appraise, engage with, and make decisions about health information. Research into how the general public uses Wikipedia is needed. It has the potential to provide deeper insight into Wikipedia's potential to address health literacy and information access inequities and, consequently, to contribute in part to addressing health literacy gaps in a time when equitable access is more important than ever.

## References

Bedirian, R. (2021, August 16). Arab medical professional Alaa Najjar honoured by Wikipedia for COVID-19 coverage. *The National*. https://www.thenationalnews.com/arts-culture/2021/08/16/arab-medical-professional-alaa-najjar-honoured-by-wikipedia-for-covid-19-coverage/



https://doi.org/10.15215/aupress/9781771994521.01

Benjakob, O. (2020, April 8). Why Wikipedia is immune to coronavirus. *Haaretz*. https://www.haaretz.com/us-news/.premium.MAGAZINE-why-wikipedia-is-immune-to-coronavirus-1.8751147

Benjakob, O., Aviram, R., & Sobel, J. A. (2022). Citation needed? Wikipedia bibliometrics during the first wave of the COVID-19 pandemic. *GigaScience*, *11*(1), giab095. https://doi.org/10.1093/gigascience/giab095

Birkun, A. (2024). Wikipedia is a highly accessed although unreliable source of knowledge on cardiopulmonary resuscitation. *Clinical and Experimental Emergency Medicine*, *11*(4), 392–95. https://doi.org/10.15441/ceem.24.223

Bodie, G. D., & Dutta, M. J. (2008). Understanding health literacy for strategic health marketing: eHealth literacy, health disparities, and the digital divide. *Health Marketing Quarterly*, *25*(1–2), 175–203. https://doi.org/10.1080/07359680802126301

CBC Ideas. (2021, October 22). The danger of COVID-19 misinformation is "mind-boggling," says Dr. Anthony Fauci. *CBC Radio*. https://www.cbc.ca/radio/ideas/the-danger-of-covid-19-misinformation-is-mind-boggling-says-dr-anthony-fauci-1.6218070

CBC News. (2021, May 28). Ontario speeding up 2nd dose eligibility for COVID-19 vaccines. *CBC News*. https://www.cbc.ca/news/canada/toronto/covid-19-vaccine-rollout-second-doses-may-28-2021-1.6043862

Chrzanowski, J., Sołek, J., Fendler, W., & Jemielniak, D. (2021). Assessing public interest based on Wikipedia's most visited medical articles during the SARS-CoV-2 outbreak: Search trends analysis. *Journal of Medical Internet Research*, *23*(4), e26331. https://doi.org/10.2196/26331

Cohen, N. (2020, March 15). How Wikipedia prevents the spread of coronavirus misinformation. *Wired*. https://www.wired.com/story/how-wikipedia-prevents-spread-coronavirus-misinformation/

COVID-19. (2026, January 29). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=COVID-19&oldid=1335541051

COVID-19 pandemic. (2026, January 28). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=COVID-19_pandemic&oldid=1335387577

COVID-19 vaccine. (2026, January 26). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=COVID-19_vaccine&oldid=1334874655

Diaz, J. (2021, May 28). Employers can (mostly) require vaccines for workers returning to the office. *NPR*. https://www.npr.org/2021/05/28/1001116485/for-employers-the-law-is-mostly-on-their-side-when-it-comes-to-vaccines

Epstein, H.-A. B. (2022). Health information on Wikipedia: Can you trust it? *Journal of Consumer Health on the Internet*, *26*(2), 218–24. https://doi.org/10.1080/15398285.2022.2069391



https://doi.org/10.15215/aupress/9781771994521.01

Fisher, K. E., Erdelez, S., & McKechnie, L. E. F. (2005). *Theories of information behavior*. Information Today.

Gozzi, N., Tizzani, M., Starnini, M., Ciulla, F., Paolotti, D., Panisson, A., & Perra, N. (2020). Collective response to media coverage of the COVID-19 pandemic on Reddit and Wikipedia: Mixed-methods analysis. *Journal of Medical Internet Research*, *22*(10), e21597. https://doi.org/10.2196/21597

Harrison, S. (2020, May 27). Future historians will rely on Wikipedia's COVID-19 coverage. *Slate*. https://slate.com/technology/2020/05/wikipedia-coronavirus-information-future-historians.html

Heilman, J. M., Kemmann, E., Bonert, M., Chatterjee, A., Ragar, B., Beards, G. M., Iberri, D. J., Harvey, M., Thomas, B., Stomp, W., Martone, M. F., Lodge, D. J., Vondracek, A., Wolff, J. F. de, Liber, C., Grover, S. C., Vickers, T. J., Meskó, B., & Laurent, M. R. (2011). Wikipedia: A key tool for global public health promotion. *Journal of Medical Internet Research*, *13*(1), e14. https://doi.org/10.2196/jmir.1589

Heilman, J. M., & West, A. G. (2015). Wikipedia and medicine: Quantifying readership, editors, and the significance of natural language. *Journal of Medical Internet Research*, *17*(3), e62. https://doi.org/10.2196/jmir.4069

Huisman, M., Joye, S., & Biltereyst, D. (2020). Health on Wikipedia: A qualitative study of the attitudes, perceptions, and use of Wikipedia as a source of health information by middle-aged and older adults. *Information, Communication & Society*, *24*(12), 1797–813. https://doi.org/10.1080/1369118X.2020.1736125

Ibrahim, N. (2021, June 10). Is Wikipedia as "unreliable" as you've been told? Experts suggest the opposite may be true. *Global News*. https://globalnews.ca/news/7921230/wikipedia-reliablity/

Islam, M. S., Sarkar, T., Khan, S. H., Kamal, A.-H. M., Hasan, S. M. M., Kabir, A., Yeasmin, D., Islam, M. A., Chowdhury, K. I. A., Anwar, K. S., Chughtai, A. A., & Seale, H. (2020). COVID-19–related infodemic and its impact on public health: A global social media analysis. *The American Journal of Tropical Medicine and Hygiene*, *103*(4), 1621–29. https://doi.org/10.4269/ajtmh.20-0812

Johnson, J. D., & Case, D. O. (2012). Socio-psychological factors in health. In G. L. Kreps (Ed.), *Health information seeking* (Vol. 4, pp. 39–61). Peter Lang.

Keton, A., & Humborg, C. (2021, November 29). Digital regulation must empower people to make the internet better. *TechCrunch*. https://techcrunch.com/2021/11/28/digital-regulation-must-empower-people-to-make-the-internet-better/

Martzoukou, K., & Sayyad Abdi, E. (2017). Towards an everyday life information literacy mind-set: A review of literature. *Journal of Documentation*, *73*(4), 634–65. https://doi.org/10.1108/JD-07-2016-0094

Nutbeam, D. (2000). Health literacy as a public health goal: A challenge for contemporary health education and communication strategies into the 21st



https://doi.org/10.15215/aupress/9781771994521.01

century. *Health Promotion International*, *15*(3), 259–27. https://doi.org/10.1093/heapro/15.3.259

Nutbeam, D. (2008). The evolving concept of health literacy. *Social Science & Medicine*, *67*(12), 2072–78. https://doi.org/10.1016/j.socscimed.2008.09.050

Nutbeam, D., & Lloyd, J. E. (2021). Understanding and responding to health literacy as a social determinant of health. *Annual Review of Public Health*, *42*, 159–73. https://doi.org/10.1146/annurev-publhealth-090419-102529

Patel, R. (2021, February 14). Tam takes aim at COVID-19 "infodemic," urges vigilance over misleading online content. *CBC News*. https://www.cbc.ca/news/politics/theresa-tam-covid-19-infodemic-1.5914178

Ruprechter, T., Horta Ribeiro, M., Santos, T., Lemmerich, F., Strohmaier, M., West, R., & Helic, D. (2021). Volunteer contributions to Wikipedia increased during COVID-19 mobility restrictions. *Scientific Reports*, *11*(1), 21505. https://doi.org/10.1038/s41598-021-00789-3

SARS-CoV-2. (2026, January 16). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=SARS-CoV-2&oldid=1333159463

Sebelefsky, C., Karner, D., Voitl, J., Klein, F., Voitl, P., & Böck, A. (2015). Internet health seeking behaviour of parents attending a general paediatric outpatient clinic: A cross-sectional observational study. *Journal of Telemedicine and Telecare*, *21*(7), 400–407. https://doi.org/10.1177/1357633X15583431

Shafee, T., Masukume, G., Kipersztok, L., Das, D., Häggström, M., & Heilman, J. (2017). Evolution of Wikipedia's medical content: Past, present and future. *Journal of Epidemiology and Community Health*, *71*(11), 1122–29. https://doi.org/10.1136/jech-2016-208601

Smith, D. A. (2020). Situating Wikipedia as a health information resource in various contexts: A scoping review. *PLOS ONE*, *15*(2). https://doi.org/10.1371/journal.pone.0228786

Smith, D. A. (2021). Wikipedia: An unexplored resource for understanding consumer health information behaviour in library and information science scholarship. *Journal of Documentation*, *78*(3). https://doi.org/10.1108/JD-03-2021-0049

Smith, D. A. (2023a). "I'm comfortable with it": User stories of health information on Wikipedia. *First Monday*, *28*(8). https://doi.org/10.5210/fm.v28i8.12897

Smith, D. A. (2023b). It's time to recognize Wikipedia as a health information resource. *Journal of Consumer Health on the Internet*, *27*(2), 210–20. https://doi.org/10.1080/15398285.2023.2211498

Taylor, C. (2021, May 28). UK approves Johnson & Johnson's single-dose Covid vaccine for use. *CNBC*. https://www.cnbc.com/2021/05/28/uk-approves-janssen-single-dose-covid-vaccine-for-use.html



https://doi.org/10.15215/aupress/9781771994521.01

*Wikipedia and COVID-19*. (2020, April 13). Wikimedia Foundation. https://wikimediafoundation.org/covid19/data/

Wikipedia: Identifying reliable sources (medicine). (2025, December 23). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Wikipedia:Identifying_reliable_sources_(medicine)&oldid=1329047369

Wikipedia: Neutral point of view. (2026, January 15). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Wikipedia:Neutral_point_of_view&oldid=1333125479

Wikipedia: No original research. (2026, January 27). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Wikipedia:No_original_research&oldid=1335052383

Wikipedia: Reliable sources. (2026, January 29). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Wikipedia:Reliable_sources&oldid=1335477844

Wikipedia: WikiProject COVID-19. (2021, September 8). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Wikipedia:WikiProject_COVID-19&oldid=1043202989

*Words we're watching: "Infodemic."* (n.d.). Merriam-Webster. Retrieved December 7, 2021, from https://www.merriam-webster.com/words-at-play/words-were-watching-infodemic-meaning

Yang, K., & Tanaka, M. (2023). Crowdsourcing knowledge production of COVID-19 information on Japanese Wikipedia in the face of uncertainty: Empirical analysis. *Journal of Medical Internet Research*, *25*, e45024. https://doi.org/10.2196/45024



https://doi.org/10.15215/aupress/9781771994521.01

*This page intentionally left blank.*

# Afterword

## The Value of Verified Knowledge in the Age of Generative AI

Pierre Lévy

Most people, myself included, who use Wikipedia on a nearly daily basis are generally unaware of the importance, both theoretical and practical, of the debates taking place internally among the contributors and touching on essential social, cultural, and epistemological issues. After reading this book on open culture and the Wikimedia movement in Canada, the reader will have come across some fascinating analyses of the complex issues facing Wikipedia's authors and editors. The reader will also have gained a glimpse of the growing mine of information that is Wikidata, a semantic metadata repository that today serves as the skeleton for numerous symbolic artificial intelligence (AI) systems.[1]

I would like to start this afterword with a quote from Denny Vrandečić, one of the initiators of Wikidata who had worked on Google's knowledge graph as an ontologist and is now the leader of the Abstract Wikipedia project, which is aimed at making Wikipedia articles data language independent (i.e., translatable into all languages). Speaking at the Knowledge Graph Conference in May 2023, Denny said, "In a world of infinite content, knowledge becomes valuable" (Vrandečić, 2023). Clearly, this world of potentially infinite content results from the now massive use of generative artificial intelligence. This new situation poses a number of problems. Let's mention two that are particularly

1 Symbolic AI uses explicit logical rules and semantic networks (rather than statistical methods).



https://doi.org/10.15215/aupress/9781771994521.01

significant from the point of view of access to knowledge. First, despite the uses made of generative models to obtain direct answers quickly, it should be remembered that today's statistical AI (also known as neural AI) is unlike the classic symbolic AI of the twentieth century: Statistical AI is limited only by its capabilities and offers no guarantee of truth. GPT-4 and other similar models *do not constitute knowledge bases*. Statistical AI leads to many errors of both fact and reasoning, and we only need to be a specialist in a single field of endeavour to notice the weaknesses of this kind of AI, as we usually do when we read an article involving one of our areas of expertise when it is written by a journalist in a hurry who is content to restate the common beliefs in which they are immersed. ChatGPT's probabilistic answers are only *plausible*. Secondly, as generative AI models are trained on web data and this data is increasingly written and illustrated by the models in question, we find ourselves in the presence of a dangerous vicious circle, all the more so as the low-paid workers responsible for aligning the models and correcting their biases or errors are themselves using generative AI to accomplish their task!

This presents us with epistemological quicksand, and the best way to extricate ourselves is to invest more than ever in building reliable sources of information. In other words, the explosion in the use of generative AI doesn't dispense us from building up and using Wikipedia, Wikidata, and other verified knowledge bases; on the contrary, it makes it even more necessary that we contribute to them and have the pleasure of consulting them! That said, we have seen in the rest of this collective work that neural AI is nonetheless destined to play a positive role in the sharing of knowledge that is so dear to Wikipedians.

I am not a researcher specializing in Wikimedia studies. Instead, I am a philosopher committed to thinking about the digital world. I will confine myself to a few thoughts by way of coda on the triangular relationship between collective intelligence, artificial intelligence, and the uplifting goal of making knowledge available to all.

## Collective Intelligence

Collective intelligence processes are narrowly defined as a means of solving problems, and they take many forms, the most widely studied of which are statistical, deliberative, and stigmergic (Baltzersen, 2022; Lévy, 1994).



https://doi.org/10.15215/aupress/9781771994521.01

At the beginning of the twentieth century, the English scientist Francis Galton visited an agricultural fair. A competition was organized with 800 participants, most of them farmers, and they were asked to guess the weight of an ox. But none of them had found the exact weight. Galton took the average of all the estimates and found that it was much closer to the real weight than any of the individual estimates. The "wisdom" of the crowd was thus superior to each of the individual intelligences (Galton, 1907). This form of collective statistical—or accounting—intelligence assumes that individuals do not communicate with each other and do not coordinate in any way. It works all the better if the distribution of choices or predictions is spread over a wide spectrum so that individual errors and biases compensate for each other. Paradoxically, this type of collective intelligence presupposes mutual ignorance. It is expressed in opinion polls or elections, where it is forbidden to communicate partial results before everyone has voted. This approach to statistical collective intelligence, with no connection between its members, was popularized by James Surowiecki in his book *The Wisdom of Crowds*.

A second form of collective intelligence is deliberative intelligence, and it is based on direct communication between the members of a community. Deliberative intelligence results from the exchange of arguments and points of view. Faced with a common problem, deliberative intelligence can converge on a consensus, or it can be divided between a number of solutions whose pros and cons have been weighed up together. Provided there is a willingness to listen, everyone brings their own point of view and their own particular expertise, and this in turn enriches the general debate (Mulgan, 2017; Zara, 2021). This type of collective intelligence is apparently ideal because it is open and reflexive, yet it is all the more difficult to implement when the community is large. Forms of hierarchy and delegation then need to be established, and while these are certainly essential, they also disrupt the transparency of the collective intelligence itself.

I would now like to introduce a lesser-known form of collective intelligence that is nonetheless at work in many animal societies and has found its highest degree of completion in humanity: stigmergic communication. The Greek etymology explains the meaning of the word *stigmergy* quite well: Marks (*stigma*) are left in the environment by the action or work (*ergon*) of members of a community, and these marks in turn—and recursively—guide their actions (Heylighen, 2016). The classic case of stigmergy is when ants leave a trail of pheromones in their wake as they bring food back to the anthill.



https://doi.org/10.15215/aupress/9781771994521.01

The smell of the pheromones encourages other ants to follow in their footsteps to discover the booty and bring food back to their underground city, leaving a scented message on the ground in their wake. Language gives humanity a high degree of collective intelligence, superior to that of other mammals and comparable to that of bees or ants. Like other eusocial species, we communicate largely in a stigmergic way. But instead of marking a physical territory with pheromones or other types of visual, auditory, or olfactory signals, we leave symbolic traces. As culture evolves, signifiers accumulate in increasingly sophisticated *external memories*: standing stones, totem poles, sculpted landscapes, monuments, architecture, writing signs, archives, libraries, and databases. It could be argued that any form of writing that is not precisely addressed is a form of stigmergic communication: Traces are deposited for future reading and act as the external memory of a community.

The various processes of collective intelligence that have just been mentioned—statistical (without communication), deliberative (with direct communication), and stigmergic (with indirect communication)—are obviously not mutually exclusive. Indeed, they may well succeed one another or combine. For example, Wikipedians coordinate, deliberate, and vote via shared databases.

On the scale of the species as a whole, human collective intelligence is a continuation of animal collective intelligence, but in humans, this collective intelligence is more sophisticated than in other species because of the language, techniques, and political, economic, legal, and other institutions that characterize us. The main difference between animal and human collective intelligence lies in culture. In a diachronic dimension, our species is driven by a speed of learning greater than that of biological evolution. Our know-how is accumulated and passed on from one generation to the next by means of our external memories, sign systems, social conventions, and tools. No individual would be "intelligent" if they did not inherit the knowledge created by their ancestors. In a synchronic dimension, we participate in a coordinated collective intelligence in which the conceptual architecture of our shared memories and the social organization of our communities resonate and revive each other. The reciprocal definition of identities and the recognition of problems are decided at this metalevel of culture. As a result, beyond the useful procedures (stigmergic, statistical, and deliberative) for solving problems, there is a more holistic collective intelligence that circumscribes the cognitive capacities of a society.



https://doi.org/10.15215/aupress/9781771994521.01

Cultural evolution has already crossed several thresholds of collective intelligence. The inventions of writing, printing, and electronic media (music recording, telephone, radio, television) have already irreversibly increased our capacity for memory and social communication. Probably the greatest social change we have experienced in the last 25 years is the emergence of global communication via digital memory. This new form of distributed read-write communication in a collective digital memory represents a far-reaching anthropological transformation. We are immersed in the new digital environment, and we interact through the oceanic mass of data that brings us together. Wikipedia's encyclopedists and GitHub's programmers collaborate via the same database. Unbeknown to us, every link we create, every tag or keyword we affix to a piece of information, every act of evaluation or of approval, every "like," every request, every purchase, every comment, every share we post—all these operations subtly modify the shared memory—in other words, the inextricable magma—of relationships between data. Our online behaviour emits a continuous flow of messages and cues that transform the structure of memory and help direct the attention and activity of our contemporaries. We deposit electronic pheromones in the virtual environment that determine the actions of other internet users in a loop and that train the formal neurons of AIs as well.

## Artificial Intelligence as an Enhancement of Collective Intelligence

Let's now turn to artificial intelligence, but from the angle—which may strike some readers as unusual—of collective intelligence. Journalists and the general public tend to classify applications that are considered advanced at the time they appear as "artificial intelligence." But a few years after being introduced, these same applications will have become commonplace, everyday occurrences and will more often than not be reinterpreted as belonging to ordinary computing. Since the mid-twentieth century, despite the apocalyptic headlines and images of young women with chromium-plated brains that are supposed to embody artificial intelligence, we have been witnessing a process of formal commodification and externalization of cognitive functions. Increased power and lower hardware costs are distributing these objectified cognitive functions throughout society. Interconnected machines record and retrieve information, perform arithmetic or algebraic calculations, simulate



https://doi.org/10.15215/aupress/9781771994521.01

complex phenomena, reason logically, conform to syntax and systems of rules, and extract shapes from entangled statistical distributions. Computers automate and socialize our ability to communicate, our capacity for memory, perception, learning, analysis, and synthesis.

Artificial intelligence, by virtue of its very name, naturally conjures up the idea of an autonomous machine intelligence that sits opposite human intelligence, simulating or surpassing it. But if we look at how AI devices are actually used, we have to admit that, most of the time, they augment, assist, or accompany the operations of human intelligence. Back in the days of expert systems—in the 1980s and 1990s—I observed that once the critical knowledge of specialists within an organization was codified in the form of rules to drive knowledge bases, it could then be made available to the members who needed it most, responding precisely to current situations while remaining continuously available. Rather than supposedly autonomous artificial intelligences, these were media for disseminating practical know-how, the main effect of which was to increase the collective intelligence of the user communities.

In the current phase of AI development, the role of the expert is played by the crowds who produce the data. The role of the cognitive engineer who codifies the knowledge is played by neural networks. Instead of asking linguists how to translate something or authors how to produce a text, statistical models unknowingly interrogate the multitudes of anonymized writers on the web to automatically extract standard structures that no human programmer would have been able to work out. Algorithms are conditioned by their training and can then recognize and reproduce data corresponding to the forms they have learned. But because they have abstracted structures instead of recording everything, they are now capable of correctly conceptualizing forms (of images, text, music, code, etc.) that they have never encountered and of producing an infinite number of new symbolic arrangements. This is why we speak of *generative* artificial intelligence. Neural AI synthesizes and mobilizes shared memory. Far from being autonomous, it extends and amplifies collective intelligence. Millions of users contribute to perfecting the models by asking them questions and commenting on the answers they receive. Take Midjourney, for example, where users exchange instructions (prompts) and constantly improve their AI skills. Midjourney's Discord servers are now the most widely distributed on the planet, with over one million users. A new collective stigmergic intelligence is emerging from the fusion of social media, AI, and creative communities.



https://doi.org/10.15215/aupress/9781771994521.01

Contemporary AI thus serves as the conduit for a feedback loop between the shared digital memory and the individual productions that exploit it and accumulate in turn in data centres. We need to get behind the machine and glimpse the collective intelligence that it reifies and mobilizes.

## Sharing Knowledge: Toward Neurosymbolic Collective Intelligence

The collective intelligence currently supported by artificial intelligence is still only partial. In fact, the use of internet data to train models mobilizes collective stigmergic intelligence (the feedback loop between individual behaviour and shared memory) and statistical intelligence (neural learning). In the early 2020s, the connection between and mutual reinforcement of these two forms of collective intelligence by the new AI devices provoked an intellectual shock—and strong emotions—in people who gained a glimpse of how powerful they could be. But a deliberative and reflexive collective intelligence was still missing. At the scale at which we are operating, this deliberative collective intelligence must relate to the organization of data—that is, the conceptual structure of memory, inevitably coupled with the practices of communities. How can we ensure that the networks of concepts that inform digital memory are subjected to an open, transparent conversation that is attentive to the consequences of the choices digital memory makes? The semantic web and its panoply of standards (XML, RDF, OWL, SPARQL) have certainly established format interoperability, but they do not provide the proper semantic interoperability—that of concept architectures—that we need. The web giants have their knowledge graphs, but unfortunately, these are private and secret. Wikidata offers an example of an open-knowledge graph, but it is still very difficult for the general public to explore and use this on a daily basis. What's more, it is presented as a stand-alone ontology, that of the Wikipedia encyclopedia, whereas we need to harmonize and dialogue with the multitude of ontologies emerging from practices as diverse as we like.

I invented IEML (Information Economy MetaLanguage) to solve this problem of the emergence of a deliberative (or reflexive) collective intelligence in digital form. IEML is an artificial language with a regular algebraic structure, whose semantics can be calculated and that can say anything and translate any concept network (Lévy, 2023). IEML is an open-source language designed with a view to increasing the knowledge commons, and its



https://doi.org/10.15215/aupress/9781771994521.01

development must be subject to decentralized governance. However heterogeneous or diverse it may be, IEML projects ontologies, knowledge graphs, collections of labels, and data models onto the same semantic coordinate system: a virtually infinite universe of conceptual differences that algorithms can use. IEML can serve as a *pivotal language* between natural languages, between humans and machines, and between AI models.

It goes without saying that most of IEML's users will not have to learn this metalanguage, since the application interfaces, including the editor themselves, will be in natural languages or in iconic form. The "code" side of IEML is intended for computers only. We can therefore envisage a multitude of knowledge bases with singular conceptual architectures being able to exchange ontological modules and information thanks to the semantic interoperability provided by this *common metadata language*.

Let's return now to what unites all Wikipedians—namely, the authors, editors, and readers of the movement. Not the "encyclopedia" object, which is ultimately no more than a particular means adapted to the technical and cultural possibilities of an era, but the much broader aim, which resonates into an as-yet-unimaginable future that will consist of making knowledge available to everyone. This concept can be broken down into two exercises: (1) allowing all knowledge to be expressed, accumulated, and communicated and (2) facilitating the exploration and appropriation of knowledge based on the wide range of practical situations, learning paths, and cognitive styles. We can see the affinity of the Wikipedia ideal with that of collective intelligence, which is diametrically opposed to "groupthink" and aims to maximize creative freedom and collaborative efficiency simultaneously.

The philosopher in me will be forgiven for evoking a concrete utopian vision that is no doubt technically feasible but in the short term aims first and foremost to get people to think. So let's imagine a system for sharing knowledge that makes the most of contemporary technical possibilities. At the heart of this system is an open *ecosystem of knowledge bases* categorized as IEMLs, which emerge from a multitude of communities of research and practice. Between this core of interoperable knowledge bases and living human users lies a "no-code" neural interface (an ecosystem of models) that provides access to data control, supply, exploration, and analysis. Everything happens intuitively and directly, in line with the sensory-motor modalities selected. It is also through this gigaperceptron—an immersive, social, and generative metaverse—that groups exchange and discuss the data models and semantic



https://doi.org/10.15215/aupress/9781771994521.01

networks that organize their memories. This new knowledge-sharing system is a good knowledge management tool that encourages the recording of creations, supports learning paths, and presents useful information to those involved in their practices.

On the shared side, each knowledge base, whether personal or collective, displays its universe of discourse, its data, and its statistics, which is as transparent for algorithms as it is for human eyes. On the private side, however, our knowledge-sharing system ensures that individuals and groups have practical and legal sovereignty over the data they produce, which they disclose only to select players.

The decisive increase in the deliberative dimension of collective intelligence through the use of a common metadata language has multiplier effects on the statistical and stigmergic collective intelligences already at work in our day. A new neurosymbolic infrastructure will plunge the collective intelligence of the future into the explorable universe emanating from its own cognitive activities. However, we need to distinguish between the collective intelligence that drives living human individuals and communities and the mechanical extensions and media representations that augment it. Let's not turn artificial intelligence into an idol.

Quoting Ibn Rushd (the Averroes of the Latins), Dante writes in chapter I, 3 of the *Monarchy*,

> The highest potentiality of mankind is his intellectual potentiality or faculty. And since that potentiality cannot be fully actualized all at once in any one individual or in any one of the particular social groupings enumerated above, there must needs be a vast number of individual people in the human race, through whom the whole of this potentiality can be actualized. (Alighieri, 1996)

Let this vast number of individual people become transparent to itself in the new algorithmic medium, and we will have moved from the anthill to the city.

## References

Alighieri, D. (1996). *Monarchy* (P. Shaw, Trans.). Cambridge University Press.

Baltzersen, R. K. (2022). *Cultural-historical perspectives on collective intelligence*. Cambridge University Press.



https://doi.org/10.15215/aupress/9781771994521.01

Galton, F. (1907). Vox populi. *Nature*, *75*, 450–51. https://www.romolocapuano.com/wp-content/uploads/2023/10/Francis-Galton_Vox_populi.pdf

Heylighen, F. (2016). Stigmergy as a universal coordination mechanism II: Varieties and evolution. *Cognitive Systems Research*, *38*, 50–59. https://doi.org/10.1016/j.cogsys.2015.12.007

Lévy, P. (1994). *L'intelligence collective: Pour une anthropologie du cyberspace*. La Découverte.

Lévy, P. (2023). Calculer la sémantique avec le langage IEML. *Humanités numériques*, *8*. https://doi.org/10.4000/revuehn.3836

Mulgan, G. (2017). *Big mind: How collective intelligence can change our world*. Princeton University Press.

Vrandečić, D. (2023, July 19). *The future of knowledge graphs in a world of LLMs* [Video]. YouTube. https://www.youtube.com/watch?v=ww99npDh4cg

Zara, O. (2021). *Le chef parle toujours en dernier: Manifeste de l'intelligence collective*. Axiopole.



https://doi.org/10.15215/aupress/9781771994521.01

# Contributors

**Stacy Allison-Cassin** is assistant professor in the Department of Information Science in the Faculty of Management at Dalhousie University where she researches and teaches on topics related to knowledge organisation and knowledge equity and justice. She started editing Wikipedia in 2015 and Wikidata in 2016 and has been an active member of the open community, including the Creative Commons and open access communities, for more than twenty years.

**Gabriel Arsenault** is associate professor of political science at the Université de Moncton and associate researcher at the Donald J. Institute. He is the editor of *The Higgs Years. Leading and Dividing New Brunswick* (2025).

**Nathalie Casemajor** is a professor at the Urbanization, Culture and Society Research Centre of the Institut national de la recherche scientifique (INRS). Her research examines the intersections of culture, technology, and territory. She holds the INRS Chair in Digital Commons and directs the Fernand-Dumont Chair on Culture. She has been a Wikipedian since 2012.

**Jean-Michel Lapointe** is digital education project manager at Université du Québec à Montréal. A Wikipedian since 2014 and a former academic librarian and teacher, he promotes the educational use of Wikimedia projects in higher education settings.

**Pierre Lévy** is a philosopher, a member of the Royal Society of Canada, and former holder of the Canada Research Chair in Collective Intelligence at the University of Ottawa. He has authored thirteen books translated into more than a dozen languages, including *Collective Intelligence*, *Becoming Virtual*, and *The Semantic Sphere*. He is the CEO and co-founder of INTLEKT Metadata Inc., and the inventor of the Information Economy Meta Language.



https://doi.org/10.15215/aupress/9781771994521.01

**Marie D. Martel** is an associate professor at the École de bibliothéconomie et des sciences de l'information at the Université de Montréal. A specialist in libraries and open culture, she is also involved in research and outreach projects related to Wikimedia platforms. She has been a Wikipedian since 2010.

**Denise Smith** is the Head of the Research Lifecycle Department at Brock University Library. She has practiced academic librarianship since 2013. She has explored Wikipedia and contributed to the Wikipedia movement since 2017, hosting edit-a-thons, teaching undergraduate and graduate students to edit Wikipedia, and researching Wikipedia as a health information resource. She is currently a PhD student at Western University's Faculty of Information and Media Studies.

**Nathalie Thibault** is archives curator at the Musée national des beaux-arts du Québec (MNBAQ) since 1987. Starting in 2016, she has been collaborating on various digital projects in the cultural field, like data standardization, metadata, Wikimedia projects, digitization, and dissemination of art and archive collections.

**Miguel Tremblay** is a physicist and has been a contributor to Wikipedia since 2003. With more than 16,000 contributions to Wikipedia, 3000 contributions to Wikimedia Commons, he is on Wikidata as the founder and animator of the Wikiproject on weather observations.

**Simon Villeneuve** is a physics and astronomy teacher at Cégep de Chicoutimi (Québec). He has fifteen years of experience as a trainer for wikis hosted by the Wikimedia foundation (Wikipedia, Wikidata, Wikimedia Commons but also Wikisources and Wikinews). A Wikipedian since 2007, he has contributed 10% of the French content on Wikipedia and is recognized as the sixth largest contributor in the world on Wikidata.

**Mathieu Wade** is a professor in the Department of Sociology and Criminology at Université de Moncton. His research focuses on Acadian nationalism and the history of land use planning in Acadia.



https://doi.org/10.15215/aupress/9781771994521.01