Quantifying the Gap: The Gender Gap in French Writers' Wikidata

Melanie Conroy[1]

[1] University of Memphis

Journal of Cultural Analytics

Vol. 8, Issue 2, 2023

One of the recurring questions of world literary history is how to ensure that marginalized writers are represented. The advent of a data-driven literary history has made this question even more pressing, as collaborative and distributed projects like Wikidata have been shown to exhibit large gaps between groups, despite the diversity of topics and contributors represented. In order to get an idea of how entrenched the gender gap is within literary Wikidata, I will examine the representation of male writers versus writers who are women or other genders using Wikidata. Since the data are vast and complex, I will particularly focus on the subset that is related to French and Francophone writers in Wikidata with an eye to how the gender gap evolves across nations, geography, and time. I will show that the gender gap is less significant in recent periods and in smaller Wikidata communities and that the largest Wikidata communities with the longest histories have larger gender gaps. As in other subject fields, literary topics in Wikidata are disproportionately linked to male authors. Finally, I consider some ways that the gender gap intersects with linguistic justice movements and how the gender gap can be reduced in literary Wikidata. The patterns in the data and procedure may be generalizable to literary Wikidata as a whole, especially larger Wikidata communities, because the gender gap in both the French and the Francophone subsets of the data is close to the global average; there is also a higher-than-average representation of writers of other genders that resembles other large Wikidata communities.

Gender bias in literary history is a fascinating problem because there are so many potential confounding variables: gaps in the publication record, differences in how many people of different genders wrote and published, and cultural differences, to name but a few.[1] Feminists have long challenged the masculinist literary history that would claim that women's texts have often been excluded from literary canons due to poor quality. Far from a retrospective activist intervention by modern feminists, this critique of literary history's focus on male writers is present in literary history, as well as in women's writing from the medieval to early modern period. In their introduction to *A History of Feminist Literary Criticism*, editors Gill Plain and Susan Sellers affirm that their subject's "eventual self-conscious expression was the culmination of centuries of women's writing, of women writing about women writing, and of women—and men—writing about women's minds, bodies, art and ideas" (2). Within French literary history specifically, deep biases against women as literary producers have long been expressed in both literary works and literary history. Especially before the twentieth century, French

women who claimed to be professional writers or spoke out against institutional limits on their self-expression were often referred to as "un homme manqué" or otherwise not feminine. "Once a woman writer decided to publish any criticism of the patriarchal status quo in early modern France," Anne L. Schroder observes, "she risked a humiliating backlash intended to force her to retreat into silence" (376). But sometimes even women writers were famous in their own era yet were forgotten soon after. For instance, writers such as Mme de Genlis and Marie Jeanne Riccoboni were celebrated in the eighteenth century, but they were, nevertheless, sidelined in later literary histories. In Riccoboni's case, this bias was amplified by the world of publishing, such as the widespread plagiarism of her works; nevertheless, she achieved "a successful commercial literary venture in Enlightenment France, where such an undertaking was commonly doomed to fail due to the deeply ingrained gender bias against women writers," argues Marijn S. Kaplan, as well as "the limited legal protection for both authors and publishers from literary piracy and in defense of textual ownership" (187). Even in the case of individual known writers, it can, thus, be difficult to untangle gender bias from other forms of competition and hardship that authors—both male and female—experienced.

The feminist re-examination of early modern and modern literary history has often turned up writers like Genlis and Riccoboni who were highly praised at the time they published but became less so when canons were formed for use in schools and universities. Women writers of the early modern period, including Christine de Pisan, one of the earliest French women writers to appear in many literary histories and the author of *The City of Women* (*La Cité des Dames*, 1405), or Mme de La Fayette, author of *The Princess of Clèves* (*La Princesse de Clèves*, 1678), were hardly unknown when they initially published, but they have seen their reputations rise in recent decades. But some French women authors were not "published authors" in the same way that authors of books are. Many more women writers produced handwritten manuscripts, occasional poetry, or correspondence, rather than braving the publishing industry to produce books. Despite the many cultural constraints they faced, "[e]arly modern French women were prolific writers," as Colette H. Winn stresses, but "[t]hose who engaged in authorship would simply circulate their manuscripts. Relatively few of them went so far as to publish their works themselves" (1). We can, therefore, wonder how many such unpublished, or under-published, works by women writers exist in the archives. Digital platforms like Wikipedia present an opportunity to make such under-published writers known.

Indeed, we can ask whether women's literary history should be written as a separate history at all, or whether it should be integrated into masculinist literary history. Christine Planté has explored both options in a recent essay where she lays out how women's literary history can be either written into or written outside of traditional literary histories (Planté 657). Planté insists on the distinction between history as a series of events—in this case,

publications—and history as a story about those events that is open-ended and constantly retold by future generations. Open-ended digital projects have much to offer both open-ended histories and more closed versions of literary history. In fact, I would argue that data-driven literary history is uniquely good at toggling between histories written at different scales, since data can be re-organized and curated more easily than textual histories. That said, it should be noted that the debate around the importance of gender as a category is particularly contentious in French culture, which has often posed and denied the centrality of gender as a construct, as Riot-Sarcey, Planté, and Fougeyrollas argued in their 2003 book *Le genre comme catégorie d'analyse: sociologie, histoire, littérature.*

Despite decades of feminist activism to discover and reference more works by women and people of other genders, the gender gap continues to mark projects like Wikidata and Wikipedia that rely on established reference works like *Encyclopedia Britannica*, the *Oxford Dictionary of National Biography*, or *La Biographie Michaud*. Indeed, while far more research has been done on the gender gap in online encyclopedias like Wikipedia, the gender gap in projects like Wikidata is remarkably similar to that in more traditional encyclopedias. There is little doubt that women are under-represented as writers and topics in French language encyclopedias like Diderot and d'Alembert's *Encyclopédie Méthodique*, where zero out of 140 encyclopédistes were women and topics tended towards the practical mechanical arts, mathematical, geographical, botanical, and other highly technical subjects, and controversial topics like philosophy. That said, less work has been done on encyclopedias that completely exclude women like the *Encyclopédie*. There have been, however, many encyclopedias and encyclopedic projects that included women, at least to some degree, and even Diderot and d'Alembert's *Encyclopédie* participated in a culture in which women were active. Adeline Gargam notes that a small number of eighteenth-century French women were permitted to teach science publicly, one of whom, Marie-Marguerite Biheron, turned her cabinet de curiosités into "une véritable école au service de l'instruction publique" that gave lessons in human anatomy to the likes of encyclopédistes Diderot, Grimm, and d'Alembert (paragraph 11). The lack of written participation in these early encyclopedic projects is, however, striking.

It is not so easy to say why the gender gap persists in current encyclopedias. Is it primarily because women have contributed fewer canonical works to world literature or because their books are *now* perceived as "less significant" by the predominantly male Wikidata community? Encyclopedias that "have recently been subject to programmes of extensive revision and republication" like the *Oxford Dictionary of National Biography* and the *Oxford English Dictionary* have managed to add more notable women, suggesting that there are still important stories to be told about women's publications in early eras (Baigent

et al. 13). As we shall see, online and contemporary encyclopedias like Wikipedia have made some progress in closing the gender gap, but much work remains to be done.

The gender gap in Wikipedia has been documented by Wikipedians, the Wikimedia Foundation, and researchers on Wikipedia. For example, the English-language article "Gender bias on Wikipedia" cites as key evidence for the gender imbalance within Wikipedia the 2018 Wikimedia Foundation survey showing that 90% of Wikipedia contributors who responded to the survey identify as male, as well as the fact that "Wikipedia's articles about women are less likely to be included, expanded, neutral, and detailed." A similar French-language article, "Biais de genre sur Wikipédia," raises many of the same questions, focusing on women's representation in French Wikipedia and the 2008 survey that began the Wikimedia inquiry into the demographics of its users. Emma Paling details how "[s]ome female editors have been the target of harassment from their male colleagues," driving them away from the Wikipedia community.

The gaps in Wikidata have been less explored than those in Wikipedia and traditional encyclopedias. Given the substantial bias in both traditional literary history and the Wikidata communities, how can we increase the visibility of women and gender minorities in Wikidata? Well before the creation of Wikipedia, there have been attempts to quantify the gender gap and weigh different possible causes (Saint Martin 52), but Wikipedia and Wikidata provide us with new data sources to answer these longstanding questions. While I cannot hope to solve this thorny problem across the entirety of Wikidata, I would like to propose a method for approximately quantifying the gap between women's writing and the representation of women in literary history through one case study: the representation of French and Francophone women writers in Wikidata. Hopefully, the ideas presented here can be of use to literary historians seeking to integrate more high-quality data about other marginalized groups such as writers who use less common languages, indigenous writers, or writers from smaller or marginalized countries that have been inadequately represented in online spaces. In particular, I would like to examine the ways that women writers are integrated, or not, into Wikidata's knowledge graph in ways that contribute to world-historical narratives like national literatures, periodization, and spatial influence. I will also present some ways to quickly and efficiently increase literary women's representation within Wikidata and, by extension, Wikipedia as a whole, drawing on the methods of larger projects to monitor the gender gap (Klein et al.). In particular, I will consider the ways that the linguistic justice movement and feminist activism within Wikidata interrelate and how they influence the representation of literary history in Wikipedia.

Wikipedia has become an increasingly data-heavy resource. "Originally conceived in 2001 as a mainly text-based resource, Wikipedia has collected increasing amounts of structured data, including numbers, dates, coordinates, and many types of relationships, from family trees to the taxonomy of species" (Vrandečić and Krötzsch 78). Wikidata itself is a large knowledge repository of links and statements describing an item, often ascribing a property to it. Wikidata is accessible through the Wikidata website (https://www.wikidata.org/) or the Wikidata query service (https://query.wikidata.org/). Wikidata contains a lot of information about the lives of people that can be used for historical research, with enough knowledge of how the historicity of topics was arrived at. Wikidata is not a part of Wikipedia; rather, as explained on its main page, Wikidata "acts as central storage for the structured data of its Wikimedia sister projects including Wikipedia, Wikivoyage, Wiktionary, Wikisource, and others" ("Main Page"). Not all of the items in Wikidata have an associated Wikipedia or other Wikimedia page, but they almost always appear on one (for example as a relative or colleague of a more famous individual). The Wikidata stats as of August 2022 are 9,670,744 "humans," 3,954,304 of which have an associated page, 674,283 of which are labelled as "female," 3,113,838 are labelled as "male," and 1,604 as "other." 164,579 "humans" with an associated Wikipedia link do not have a gender assigned. Of the articles or topics for unknown or non-binary "humans" it is hard to draw many conclusions. The small number of articles on "other" or non-binary subjects (1,604) are, on average, more recent, longer, and higher quality than the larger number of articles on subjects of "unknown" gender (164,579), but that may be a correlation based on poor documentation that could be resolved through further research.

What constitutes the gender gap in Wikidata? Regarding Wikipedia, the gender gap in the online encyclopedia's authorship, topics, and readership has been well documented in the sciences, the arts, and general culture since the site's origin in 2001 (Konieczny and Klein 4608–10). For the most part, these disparities persist in Wikidata, despite the fact that the Wikimedia Foundation has documented the problem in surveys and organized hack-a-thons and other events aimed at reducing the gender gap in Wikipedia since 2008. These disparities exist in every major language and across a range of levels from 1) the gender of Wikipedians worldwide (around 90% of whom identify as male), 2) participation in editing and leadership roles, 3) the gap in number of entries (both of biographical articles and Wikidata topics labelled "human"), 4) length of biographical articles which are longer and more likely to be high quality for male subjects than for women, and 5) within the topics of non-biographical articles where "male-coded" topics are more often explored than "feminine-coded" topics. Gender gaps have been documented in the creation of Wikipedia biography pages (Graells-Garrido et al. 165–67); researchers have shown that women's biographies are more likely to be deleted due to perceived lesser notability (Tripodi), and the women's biographies that remain in Wikipedia are actually more notable due to deletions (Wagner et al. 2–3). There

is some debate about whether the lack of representation of women among Wikipedia's editors is a cause of excess deletions of women's biographies, or whether the bias is found in underlying works and databases that serve as sources. Whatever the causes, the lack of biographical articles about women reduces the data that can be extracted to Wikidata either automatically or by users.

At the most basic level, the gender gap that interests me here is the difference between the number of women, men, and people of other genders who appear in the Wikidata knowledge graph as item with the statement "instance of 'human.'" In other words, I am only going to deal with these inequalities at the level of prosopography, or the quantification of individuals and their traits. Computational approaches, such as attempts to distinguish between men's and women's styles in a large number of texts and to define an "écriture feminine" (Olsen 147–48), have been fruitful. Here I am concerned with a rather more mundane question: "how many women writers appear in Wikidata?". Currently, there are more than nine million items purported to be "human" across all languages in Wikidata. Of these, almost four million have an associated sitelink; more than three million of them are "male" and around 675,000 are "female." This is the main gender gap that I would like to explore. Wikidata items that have a sitelink are a good proxy for people who are considered notable enough to have their own Wikipedia page or Wikisource link since the sitelink is an internal link that often corresponds to a Wikipedia page; for example, the Wikidata item "Lou Andreas-Salomé (Q38873)" has 61 sitelinks, of which 54 are sitelinks to Wikipedia pages in various languages, four are to Wikiquote pages, and two are to Wikisource pages. I will also refer to people of "other genders," as Wikipedia classifies trans, non-binary, and other gender minorities, but that data is far more spotty and my conclusions are limited. Fewer than 2,000 people in Wikidata with a sitelink are currently assigned the gender "other," although this group is growing and no doubt under-researched and under-documented, appearing mostly in the most recent decades.

1. Wikidata as a source for literary history

Wikipedia users regularly encounter the sort of information contained in Wikidata in the form of the infobox on many Wikipedia pages. Whether generated from Wikidata or created independently, the infobox often displays a condensed version of the information in Wikidata itself, notably the person's name (called a "label"), a short description of the person, normally providing the person's nationality and occupation, and sometimes aliases. The rest of the box displays the statements about the person in a tabular format. For people, this box often contains the birth name, the birthday, birthplace, day of death and place of death. It also frequently contains a list of family members, such as spouse(s), children, parents, and other relatives. For writers, it often contains an occupation, such as "writer," "poet," "playwright" or "novelist." The infobox can also display statements about political affiliations, religion,

other jobs and roles, associates, influences, works, a signature, and tens of other possible topics across all of Wikipedia. It is important to note that the arrangement and display of information are highly variable across languages as well as among pages in the same language; the infobox is not automatically generated from Wikipedia, but rather an act of curation that each Wikipedia project has undertaken.

What may be less obvious to the general reader is that Wikipedia pages themselves vary in quality and depth across languages. Users who only interact with pages within one language community may not realize which types of information are present or could be added, either to Wikipedia pages or to Wikidata, because that community does not commonly use that information category. And, of course, Wikidata pages display far more information than the infobox, sometimes even information that is missing or incorrect in the infobox. That said, the extent of complete information in Wikidata often provides clues to the importance of a person to posterity for particular groups, whether they be linguistic communities, national ones, or somewhere in between. This is because individuals who have one fairly complete Wikipedia page—in French, for instance—may have stubs or very low quality pages in other languages. The "missing" data is much more apparent in Wikidata because the entire knowledge structure is displayed on a single page.

As Wikidata's introduction page explains, an item in Wikidata, including a person, is given a unique identifier, which is usually linked to a name; for example, the nineteenth-century French novelist George Sand (1804-1876) has the unique identifier "Q3816," which is also linked to other versions of the writer's name—"George Sand," "Lucile Aurore Dupin," and 14 other versions of her name or pseudonym, including common misspellings and versions of her pseudonym in non-Roman alphabets. By connecting these versions of a name via a unique identifier, Wikidata links not only the different versions of a person's name but also the various Wikipedia and other Wikisource pages that use those names. Via Wikidata, it is possible to see that the "Жорж Санд" [George Sand] in Russian Wikipedia is the same as "乔治·桑" [George Sand] or "阿曼蒂娜-露西-奥萝尔·杜班" [Amantine-Aurore-Lucile Dupin] in Chinese-language Wikipedias; all of these refer to the same person as the "Baroness Dudevant" (Sand's married name) in English or the French "Amantine-Aurore-Lucile Dupin" (her birth name). George Sand's name appears in Arabic, Belarusian, Central Kurdish, Persian, Russian, multiple versions of Chinese, and other languages with non-Roman scripts. Sand also has a large number of names under English, French, and other languages in the "also known as" values. Indeed, Wikidata records 74 Wikipedia pages in various languages that are associated with her, as well as many unique identifiers for library databases around the world that are linked to her profile, Wikiquote pages in 32 languages, and Wikisource pages in twelve. The disparity between

the number of languages in which Sand has a biographical page and the number of languages in which Sand's works are available is striking and perhaps indicative of a writer whose works are more talked about than read today.

For less internationally famous writers, such as the French socialite and writer Delphine de Girardin (1804-1855, Q437094), who wrote at a similar time as Sand and likewise had many pseudonyms, we find less coverage across different language Wikipedias and, therefore, fewer versions of the name or sitelinks. Girardin has a birth name ("Delphine Gay"), a married name ("Mme Émile de Girardin"), and several pseudonyms ("Vicomte Delaunay," "Charles de Launay," etc.), just as Sand does. Nevertheless, she has far fewer aliases in different languages, due to the fact that she is not present in nearly as many Wikipedia projects as George Sand is. She does have a label in many languages that do not use the Roman alphabet, such as Arabic ("كاتبة فرنسية "), Bulgarian ("Делфин дьо Жирарден"), Japanese ("デルフィーヌ・ド・ジラルダン"), Russian ("Дельфина де Жирарден"), Ukrainian ("Дельфіна де Жирарден"), and several others that transliterate Girardin's name directly. But there are fifteen languages in which she does not even have a label. And Girardin's pseudonyms only appear in the French and the English descriptions, suggesting a lack of familiarity with her writing, much of which was published pseudonymously, in other Wikipedia communities. Wikidata records thirteen Wikipedia pages in various languages that are associated with Girardin, mostly in Romance languages, but also in Hebrew, Russian, Arabic, and Ukrainian, as well as many unique identifiers in library databases, Wikisource pages in four languages, and Wikiquote pages in four—reflecting, perhaps, less integration into world literary history, or a lack of translations of her works, or a lack of interest among digital communities.

Other evidence we can find in Wikidata for the wider distribution of George Sand's works than Girardin's comes in the form of the descriptions that are attached to the unique identifier. All of the descriptions for Girardin contain the same phrase translated into various languages: "écrivaine française," "French writer," "französische Dichterin," etc. This suggests that the description is a direct translation from a common source and not created independently by various Wikipedia communities to represent a familiar subject. The descriptions of George Sand, on the other hand, while not more numerous, are more varied, from the English "French novelist and memoirist; pseudonym of Lucile Aurore Dupin," to the French "romancière et dramaturge française" ["French novelist and playwright"], to the Finnish "ranskalainen vallankumouksellinen ja feministinen kirjailija" ["a French revolutionary and feminist writer"], to the Chinese "法国作家，阿芒蒂娜-露西尔-奥萝尔·迪潘的化名" ["the pseudonym of French writer Amantine-Aurore-Lucile Dupin"]. The descriptions of George Sand refer to different genres (novel, memoir, theater). The English and the Chinese refer to her use of a pseudonym, although they use different versions of her real name. A few of the descriptions erroneously include dates or other information which is

Table 1. Sample Wikidata metrics

Metric	George Sand	D. de Girardin	Clarisse Tremblay
Wikidata sitelinks	121	26	2
Wikiquote links	32	9	0
Wikipedia links	76	13	2
Wikidata statements	336	127	26

* Three examples of Wikidata profiles for Francophone women writers

not supposed to appear in the description. Finally, the Finnish text refers to her as "a French revolutionary and feminist writer" and several others refer to her politics. The variation in these descriptions suggests that George Sand's biography, if not her works, is of enough interest for users of Wikipedia to be invested in describing her work and her politics, rather than using direct translations.

Both Sand and Girardin have attracted enough attention from Wikipedia and Wikidata editors to garner impressive documentation of their lives and works across languages. An example of a less complete Wikidata profile describes the Québécoise poet and journalist Clarisse Tremblay (1951-1999); Tremblay appears only in the French and English language editions of Wikipedia and has only four descriptions in Wikidata (in English, French, German, and Dutch) and no non-Roman script transliterations or translations. Her biography is less complete and her data appear across fewer Wikidata projects. This pattern of less complete biographies appearing in fewer Wikidata communities is repeated in other time periods and for other women writers.

Table 1 shows various metrics that can be used to measure the footprint of these three writers within the Wikidata ecosystem and how those metrics correlate to status: the number of sitelinks, links to Wikiquote pages for different languages, to Wikipedia pages, and the total number of statements. These numbers reflect the three writers' disparate statuses: George Sand as a representative of "world literature" with a global presence, Girardin as a representative of transnational literature, and Clarisse Tremblay as an example of a national literary figure (table 1). There are tens of French women writers with a footprint similar to George Sand's, hundreds who appear in many national Wikidata projects like Girardin, and thousands of writers who appear primarily in one Wikipedia community like Tremblay; indeed, these less represented writers are the majority and form the "long tail" of the data. Since they have few Wikipedia pages or sitelinks, writers with fewer links do not appear in many Wikidata subsets and may not get added to lists and other parts of Wikipedia that might bring more attention to writers who are less well known internationally. They have fewer connections and, therefore, do not appear in as many queries.

We can, therefore, use Wikipedia and Wikidata as an indicator of the prestige and influence of an author on contemporary world culture, but with the caveat that their very centrality to Wikidata may be a factor in how famous they are. Earlier studies have looked at which authors are the most central to the page network of Wikipedia and explored how this network centrality might correspond with international literary prestige for authors like Voltaire, Victor Hugo, and other prominent figures in world literary history (Hube et al. 28). Any attempt to transform an author with few connections into a super-connected author like George Sand will encounter numerous impediments in Wikipedia. These impediments exist to stop Wikidata from being overrun with hoaxes, spammers, or self-promoters. It may, however, be worth thinking about minor interventions that we can make to bring more attention to lesser known writers, using Wikidata as a tool.

2. Searching for patterns: nationality, place, and time

As we have seen, items from Wikidata can be viewed directly on the Wikidata site as well as in the infobox on some Wikipedia pages in languages like French and English. They can also be queried via SPARQL or the Wikidata query tool. The underlying data model for much of Wikidata is triples. Statements assign values, such as "female," to items, such as "George Sand." This simple language allows for unlimited connections between items.

I will be using Wikidata queries in order to get rough estimates for the number of women writers in various parts of Wikidata's knowledge graph. These numbers will change as Wikidata is updated and may be sensitive to small changes in the script or terms queried. The scripts have been published to my "Literary Wikidata" Github repository [https://github.com/mrconroy/literary-wikidata] so that readers can check my work and see the current results from the queries. The Wikidata identifiers are "P106: Q36180" for "occupation: writer" and "P21: Q6581072" for "gender: female." In order to specify French as a written language, the property for written language is "P6886," while French is "Q150."

Let's begin with the broadest picture of women writing in the French language. Querying Wikidata for "human" with the occupation of "writer" and the gender "female," we find that there are 90,352 women writers with a Wikipedia page. 4,977 of these women writers in Wikipedia are of French nationality. If we focus on the Wikipedia projects that are most likely to feature French women writers (French, English, German, Arabic), we find there are 1,626 women in Wikidata who are listed as using French as a written language and have the occupation "writer." This compares to 10,970 entries for humans with the gender "male," the profession "writer," and the written language "French" in Wikidata. Table 2 shows how "George Sand is female" is represented in Wikidata.

Table 2. Sample Wikidata statement

Item	Property	Value
Q3816	P21	Q6581072
George Sand	sex or gender	Female

* Wikidata statement: George Sand has the gender female

If we want to display the items labelled "human," "female," and "writer" on a map, we must also filter for items with an associated birth place and then change the default view to "map."[2] The script for this query (https://w.wiki/6Laa) appears as follows:

```
#Map of birth places of French-language women writers
("écrivaines d'expression française")
SELECT DISTINCT ?item ?itemLabel ?placeofbirth ?coord ?dob
WHERE {
?item wdt:P31 wd:Q5.
?item wdt:P21 wd:Q6581072.
?item wdt:P106/wdt:P279* wd:Q36180.
?item wdt:P6886 wd:Q150.
OPTIONAL {?item wdt:P19 ?placeofbirth.
?placeofbirth wdt:P625 ?coord.}
OPTIONAL {?item wdt:P569 ?dob.}
SERVICE wikibase:label
{ bd:serviceParam wikibase:language "fr, ar, en, de".}
 }
#defaultView:Map
```

Figure 1 shows the birth places of all women who are said to have the written language "French" and the occupation "writer." There are some duplicates, notably because some writers have multiple geographical coordinates for their birth places; many coordinates also represent more than one writer. There are also many writers missing, due to the fact that many writers do not have a written language explicitly marked, or even a code in Wikidata. We can see that Europe, North Africa, and Quebec, where French is a dominant written language in many countries and regions, are well represented, as are Francophone countries like Haiti. So, even though we know that this is a fraction of the total number of women writers of French expression, there is a reasonable geographical diversity to those who are in the database.

The resulting dataset includes 1,626 women who write or wrote in French, including a large number of international women writers, such as the American-born Parisian socialite Natalie Clifford Barney, English-Canadian Nancy Houston, who writes primarily in French, and Hispanic women who write in French like Silvina Ocampo. There are also the many Francophone writers of the former French empire (Magie Faure-Vidot) and former Russian

---

2 For a step-by-step tutorial on how to create these queries, see Alex Stinson, "Writing a Wikidata Query: Discovering Women Writers from North Africa" from the WikiIndaba 2018 Conference. The Wikidata identifiers are "wdt:P106 wd:Q36180" for "occupation: writer" and "ps:P31 wdt:P279" for "gender: female." In order to specify French as a written language, the property for written language is "P6886." French is "Q150."
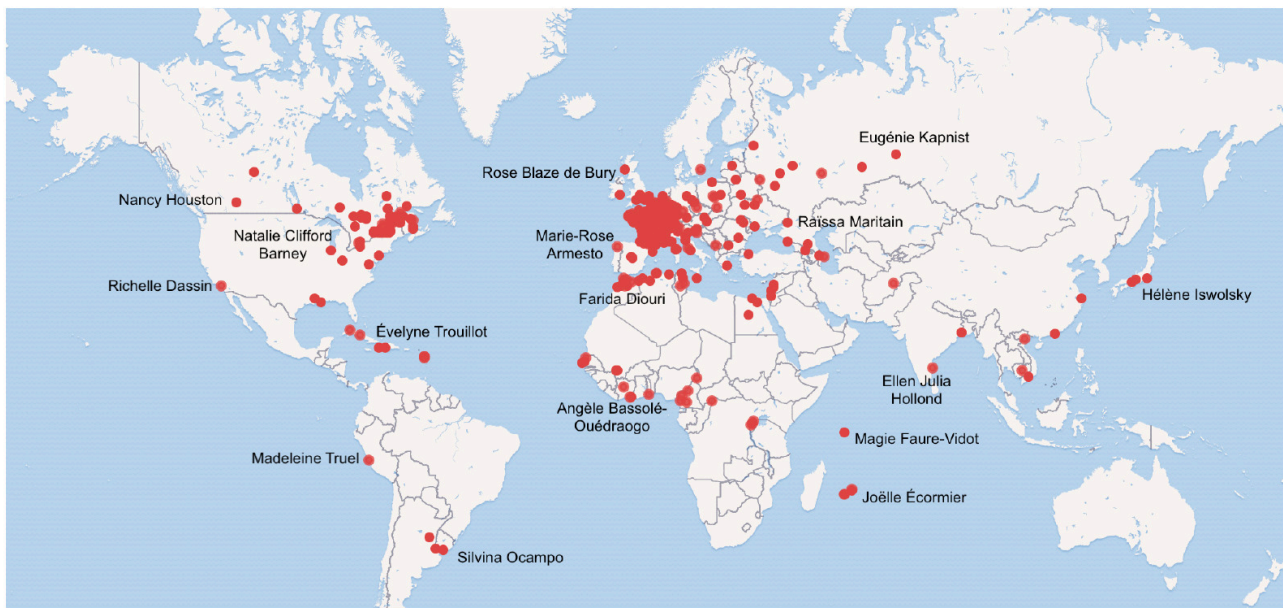
Figure 1. Map of birth places of women writers who write in French, Data: Wikidata (French, English, Arabic, German).

empire (Eugénie Kapnist). At the same time, we can see that French-speaking Africa, the home of the majority of Francophones in the world, is extremely under-represented compared to the number of speakers of the language in that region, and the regions of North Africa and the Antilles are not particularly well represented. Of course, people in these regions often speak and write in other languages, but seeking out individuals born in these regions who write in French is very likely to uncover many more new writers of French expression.

How representative of French women writers is Wikidata by time period? Placing these writers on a timeline based on their birth year, we see that the majority of these women writers with a birth year assigned in Wikipedia were born after 1800, although estimation and the small number of women writers born in most years appears to have made the data quite noisy (figure 2).

From this broad sample of women writing in French, we can see that Wikidata contains a representative sample of Francophone women writers insofar as there are representatives of all major Francophone countries and even of women from countries that are not traditionally or officially Francophone. It is also representative insofar as there are women writers from various periods, especially those born after 1800. The coverage is far spottier before 1800 and it is worth asking whether there are significant writers missed during those earlier periods, especially between 1500 and 1800 when French is fully established as a language of written work.

Feminist literary histories have given us a number of periods to examine for evidence of a rise in the number of women writers: the seventeenth and eighteenth centuries in conjunction with the rise of the French novel and new opportunities for promotion in the salons or private theaters; the early nineteenth-century sentimentalist wave; the late nineteenth-century rise of
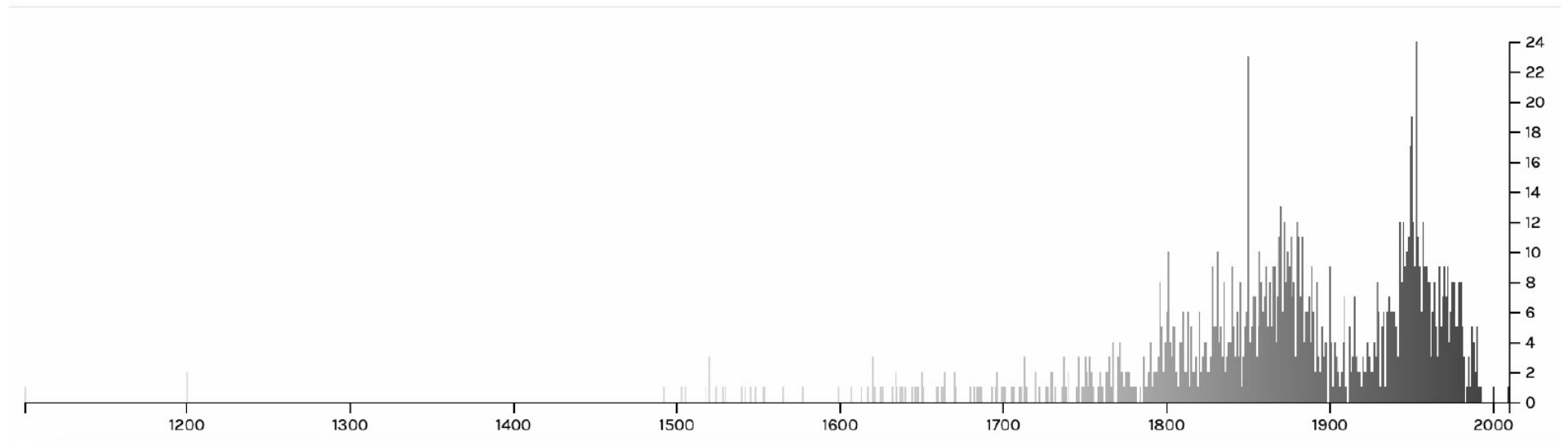
Figure 2. Timeline of birth years of women writers who write in French, Data: Wikidata (French, English, Arabic, German).

women's magazines; the avant-garde movements during and between the two world wars. Indeed, we can see in [Figure 2](#) a bump at each of these times, particularly at the beginning of the nineteenth century and again in the twentieth century. Keeping in mind that many of the writers who were born after 1980 may not have become active yet or may not yet have experienced wide acclaim, the broad picture remains a slow rise in women writing as a profession. It is important to remember that this rise in both the number and proportion of women writers occurs as people outside the gender binary are also rising. The timeline suggests that the literary community has gotten better at integrating people of diverse genders over time, or that Wikidata becomes more inclusive during later periods, or perhaps some combination of the two.

3. The French gender gap compared to other nationalities'

How does France's gender gap compare to other countries'? As of 2022, the global gender gap in Wikipedia sitelinks was 82% men / 18% women with 0.042% people of other genders. The gender gap in French Wikipedia sitelinks was slightly higher than average at 85% men / 15% women. Being close to the average, France sits around the middle of the distribution of larger countries in terms of its gender gap and the distribution of the gender gap over time. Canada has the smallest total gap at 35% women / 65% men, 0.2% other gender, among writers of that nationality. The gender gap for articles with the profession of "writer" and French nationality is virtually identical to the global average for the website (82% / 18%). Countries that have a small gender gap in Wikipedia links include Finland, Norway, the United States, Sweden, the United Kingdom, and the Netherlands. These countries are disproportionately wealthy countries of the global North, whether English-speaking or not. We cannot, therefore, entirely dismiss the possibility that wealth is correlated with the capacity to pursue gender equity.

Nevertheless, we can see patterns other than the impact of national wealth. Three other patterns arise in the data when looking at the proportion of women writers by nationality. One is that smaller Wikipedia communities, like those for Norwegian, Welsh, and Haitian Creole, often have a smaller gender gap than communities like the English or French Wikipedia. Haitian Creole Wikipedia has a particularly high percentage of women writers represented at 24%. The second pattern is that some countries with long histories of writing, notably Italy, often have the largest gender gaps; indeed, of all major Wikipedia communities Italian Wikipedia has the largest gender gap (at 10% women, 90% men, 0.03% other), as well as some of the best coverage of earlier centuries. For this reason, it is important to bear in mind change or improvement in gender balance over time. A Wikipedia or Wikidata community might have a very large gender gap when viewed as a whole, but much less so with regard to the slice representing the present day. Finally, the number and proportion of women and non-binary writers rise in all major communities over time. This suggests that Wikipedia as a whole is becoming more inclusive of gender identities, and with it, Wikidata is becoming more inclusive as well.
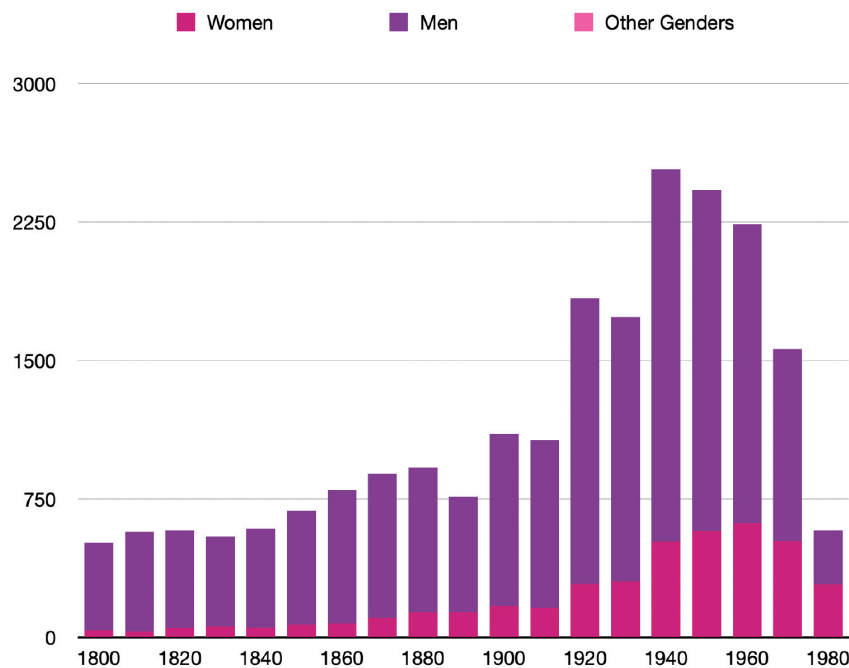
Figure 3. Bar chart of French (citizen) writers by gender (women, men, and people of other genders) by decade. Data: Wikidata.

4. The French gender gap over time

If we visualize the raw number of biographies for writers of various genders (women, men, and other genders) by decade (figure 3), we see that the number of French women writers has mostly increased, albeit unevenly, up until the 1980s, when it approaches parity (49% women / 51% men).

While gender parity in articles has been reached for those born in the 1980s (later generations are too small for statistical analysis), the rise of women writers as subjects of Wikipedia biographies across history has been slow and uneven. From the birth cohort of the 1800s to that of the 1880s, the percentage of Wikipedia biographies dedicated to women varies between five percent and eleven percent. The average percentage climbs substantially after that but stays below 20% until the birth cohort of the 1940s. The rise to near parity in the birth cohort of the 1980s is substantial, but we should remember that fewer people born in the 1980s have assigned occupations in Wikipedia. This gender gap is consistent with the global pattern within Wikipedia by birth cohort. We should note, as well, that people of other genders are featured more as time goes on but remain a small proportion of biographies of French writers. That said, French Wikipedia has one of the highest numbers of people of other genders (487), 113 of whom are writers, and 9 of whom are French citizens.

Aside from the gender gap, there are significant cultural and linguistic gaps in how much various demographic groups contribute to Wikipedia, with the English-language edition dominating non-English languages in the number and completeness of articles, a gap which means that the English-language edition often serves as a model for articles in other languages and strongly

influences how topics are covered or even deemed worthy of inclusion at all. This gap is not strictly related to the number of speakers of a language, since, for example, Chinese has a relatively small Wikipedia community and more than a billion worldwide speakers. This language bias matters, even if many native speakers of other languages use Wikipedia in English. Just as in the authorship and editing process, there are gender and cultural differences in who reads Wikipedia and how they read it: female readers and minor language groups are less represented, leading to further alienation from the Wikipedia community. Similarly, people and works with their own Wikipedia pages are perceived as more prominent or historically significant, despite the disparities in gender, culture, and language that have been documented through the encyclopedia and among Wikipedians. These documented inequities have led researchers like Julie McDonough Dolmaya to call for "linguistic justice" by closing some of the gaps in representation, rather than merely documenting them.

One aspect of Wikipedia that has held back progress in gender and linguistic justice is the concept of "notability." In order for a person to be declared "notable," he or she needs to be referenced in a third-party document, usually a paper encyclopedia or external verified database, such as a national or international dictionary of biography. Otherwise, a person can be considered "notable" by winning a prize or achieving "a widely recognized contribution." Here are the criteria for notability from external sources:

1. The person has received a well-known and significant award or honor, or has been nominated for such an award several times; or

2. The person has made a widely recognized contribution that is part of the enduring historical record in a specific field; or

3. The person has an entry in a country's standard national biographical dictionary (e.g. the *Dictionary of National Biography*).

The notability rule is a good one insofar as it prevents commercial or government interests from passing off press releases as documents verifying the importance of people who may not be of genuine interest to either the Wikipedia community or casual users of the site. Likewise, the rule has benefits in that it asks for external confirmation of the importance of a person to a field, effectively outsourcing the verification to external arbiters. But if these arbiters are themselves biased, as prize competitions and dictionaries of national biography have been in the past, such a rule risks replicating ingrained gender inequities. For this reason, considering option 3, it may be that the creation of external biographies and the linking of that data to Wikidata is the best method for quickly and cheaply removing bias from Wikidata.

5. Previous attempts to reduce the gender gap

"[O]nly 12.64% of contributors are female," according to the 2010 Wikipedia user survey, the first to thoroughly document the predominance of men among contributors and editors (Glott et al. 1171). Most previous research on gender and Wikipedia has focused on Wikipedia articles, rather than Wikidata items, but many of the same quantitative patterns exist across both projects, not least because a large number of Wikidata sitelinks are to Wikipedia articles. Later research on the gender gap in Wikipedia and Wikidata has more often analyzed data from Wikipedia or Wikidata, rather than conducting surveys of site users; for example, a 2011 presentation by Zhang and Terveen found that gender disparities persist in Wikidata (Zhang and Terveen). Wikipedia and Wikidata have both made efforts to reduce the gender gap that is persistent when Wikipedians create and structure content without combating bias. Wikipedia's active contributors are younger, more male, wealthier, and more interested in technology and adjacent topics than is the general public. In general culture, this has meant that contemporary topics, such as popular video games, consumer technology and devices, or science fiction television shows that are currently on air, get more coverage than older cultural artefacts or subjects that are less linked to computer technology, such as knitting or cooking. Within Wikidata, topics popular with young tech-positive men in the global north receive far more attention than topics that are popular with older populations, populations with less access to computers, or that are feminized. In order to overcome these common biases of internet projects, editors must make an effort to include topics related to people who do not resemble the Wikipedia contributors in gender, race, age, and cultural background.

Within literary topics, this preference for the new and technological has meant that popular literature and science fiction are over-represented, compared to classical literature. There is also some evidence that novels and graphic novels that have popular Hollywood movies derived from their intellectual property or have been adapted for mass consumption feature more prominently than books that have not. "[F]ans of pop culture are among the most enthusiastic of Wikipedia's editors," Paul Thomas asserts, because Wikipedia allows editors to create their own paratext whose usage by the general public "is a form of implicit approval that affirms the editors' knowledge and encourages them to make more edits" (1). Older works and literary works with a smaller popular culture footprint tend to have shorter, less detailed articles, and fewer sitelinks. While not authoritative, Wikipedia metrics can be used to gauge the popularity or canonicity of literary authors and works, notably within the same school or tradition (Blakesley 433–35). It is important to keep in mind the way that these metrics differ according to historical period, genre, and language. For example, medieval poetry may be less represented within Wikidata than contemporary novels when authors are considered. Similarly, poets may appear to have fewer

works to their names if the poems are listed as parts of collections, rather than as individual texts. Nevertheless, such metrics can be useful within a well-constructed set of authors or texts where such issues are understood.

Historical genres of literature that are not currently dominant in popular culture get less treatment in online encyclopedias than they would in top-down encyclopedias. A top-down editor-driven approach to encyclopedia creation and editing is not, however, the only way to deal with gender gaps. Another approach is to educate the contributors and editors about the gaps that exist and to encourage them to create solutions. This is the approach that the Wikipedia community and Wikipedia have taken. By organizing conferences, discussions, and online forums, members of the Wikipedia community, including the founder Jimmy Wales, have drawn attention to the presentist nature of Wikipedia, the tendency to focus on so-called "cult" topics with strong online audiences, and the concomitant tendency to ignore traditional cultural topics that top-down encyclopedias cover more completely, including women's literature.

Equity-driven projects have succeeded in reducing the gender gap, especially in smaller Wikipedia communities such as Haitian Creole and the Scandinavian countries. The fact that smaller and non-Western Wikipedia and Wikidata communities have often reduced the gender gap more than communities of the global north or majority white countries is an interesting phenomenon that deserves further study. In Wikidata, not only are men over-represented compared to other genders but there is also an over-representation of white individuals and citizens of countries in Europe and North America; people of other races and nationalities are under-represented in comparison to their share of the world's population (Shaik et al. 6). Nevertheless, many smaller and majority non-white communities are reducing the gender gap more quickly than larger ones. China and South Korea (both 30% female) have also reduced their gender gap. Larger Wikipedia projects like French have no doubt made progress but not as dramatically as these others have. The United States and France still have a high gender gap, as do other large countries like India, the United Kingdom, and Spain, all with large Wikipedia projects. As we have seen, there is a non-linear relationship between gender equity and linguistic justice. Increasing the presence of under-represented languages does not automatically reduce the gender gap. Indeed, it seems to be in the articles on the most recent topics that gender parity is being approached in both larger and smaller Wikipedia communities like Israel's or Norway's. The smaller gender gap in these communities is a testament to their efforts to represent women and gender minorities either in Wikipedia or in their online culture more generally. Increasingly, this approach is taken within Wikidata too.

One of the most active Wikipedia gender equity projects is the WikiProject Women in Red (Q23875215). Women in Red has been working for a decade on creating higher quality Wikipedia articles about women, using the

Wikipedia quality ranking system to identify biographical articles about women of different grades: high quality, stubs, A, B, C quality, etc. They also track the number of sitelinks that are being produced in specific categories so that users can identify categories with many "stubs," short or low quality articles that could be expanded or for which the quality could be improved. Hack-a-thons and group projects have been created to address the lack of content about historical figures who are gender minorities, women in STEM, women authors, women artists, and others. One such project is Mairely Lemus-Rojas's to increase the representation of women artists, specifically in modern and contemporary art (Lemus-Rojas). Saundra Fauconnier has also published an online tutorial for this purpose through the Wikimedia Foundation called "Making women more visible online—with Wikidata tools!" (Fauconnier).

Women in Red and related groups have also been working to reduce the gender gap in Wikidata. The WikiProject Women is spearheading a large Wikidata effort "to get every item about a woman described properly on Wikidata." Their interest is in the quality as well as the quantity of the items related to women, such as sitelinks to articles on books, art works, and historical events. There has also been work done to improve data related to the visual representation of women through Wikimedia links to images related to biographical articles, which tend to be less numerous for biographies of women; these Wikimedia and WikiCommons projects impact the number of sitelinks and other data related to women available in Wikidata. Images provide another interesting example of the gender gap with fewer images of higher average quality attached to women's biographies, compared with men's biographies (Beytía et al. 11), suggesting that more lower-quality images may be available to link to women's biographical pages or that some potential sitelinks may be missing.

One model for how to integrate data related to literary women into Wikidata comes from the Women Writers Project at the Huygens Institute for Dutch History in the Netherlands, led by Suzan van Dijk. The Women Writers Project has an online database and a large amount of documentation of their previous activities available online (van Dijk). Their Wikidata project page (P2533) shows data that they have made available. By adding their data to Wikidata, they have made it visible in the Wikidata query service, making it instantly available to millions of Wikipedia users, while maintaining control over their unique identifiers. You can see the list of French women writers in French-language Wikidata with a Women Writers ID and birth and death years using this query (https://w.wiki/6LaZ.):

```
SELECT DISTINCT ?author ?authorLabel {

    ?author wdt:P2533 ?wid;
            wdt:P21 wd:Q6581072;
            wdt:P27 wd:Q142;
            wdt:P569 ?birth;
```

```
            wdt:P570 ?death;

      SERVICE wikibase:label { bd:serviceParam
      wikibase:language "fr".}}
```

By using such a system, academic authors can increase the gender representation in Wikidata, by adding curated datasets. Further, if the data linked to from Wikidata are genuinely third-party, are properly sourced, and do not bear conflicts of interest, then the existence of a third-party, properly sourced database with links to proof can actually help meet the Wikipedia criteria for notability.[3]

Since producing and integrating data into the knowledge graph is time-consuming, many in the Wikidata community have been turning to bots and other ways to automate data extraction and curation. Indeed, Wikidata is increasingly edited by bots rather than humans and data points may never be checked over by a human user. As early as 2014, Wikipedia and Wikidata were edited by "about 50% bots and by about 23% anonymous users," meaning that little more than a quarter of the content may be checked by editors, or even humans (Steiner 1). In 2014, these bots were far more likely to edit large Wikipedias like English and French, and most bots were active in 5 or fewer languages; many smaller languages had fewer than 10 active bots, so the use of bots is highly variable across languages (Steiner 4–5). The use of bots has only increased since that time and bots now produce so much Wikidata content that, according to a more recent pre-print study, "most of this content is likely to be never seen" nor "checked by any human user. With more than 45M entities in the graph, large swathes of it may be never consulted by anyone" (Piscopo 3).

The use of bots to create data that may or may not be consulted by humans is only one of the aspects of automated editing that has been ethically controversial. A related issue is that bots and mass edits often rely upon strong assumptions about gender that can be made based on partial biographical information, or even on the basis of names alone, with no cited sources, to generate a value for "P21" (Lindsey et al. 5–7). That said, the use of bots is essential to complete routine tasks that humans are reluctant or unwilling to do, such as extracting data from Wikipedia for use in the Wikidata knowledge graph; indeed, using supervised bots is one way to solve the lack of human-readable labels in "multilingual labels in particular" (Kaffee et al. 1).

---

3 Another large project that has a presence on Wikidata but is not yet fully integrated into Wikidata is the Women Writers Project led by Julia Flanders at Northeastern University (Q8031351). The Women Writers Project has an online database and a large amount of documentation of their previous activities available online; for more information, see their website [https://www.wwp.northeastern.edu].

6. Recommendations for reducing the gender gap

So how can we reduce the gender gap in larger Wikidata projects that shape our online access to the data of world literary history? Here are some easy, low-cost ways for digital humanists to reduce the gender gap in Wikidata:

- Advocating for the "notability" of under-documented writers, such as women and gender minorities, in both articles and data projects. Without the inclusion of more women in databases and biographical dictionaries, fewer articles will be created and the notability problem will persist in Wikimedia and Wikisource projects.

- If you have a dataset with a large number of women or nonbinary writers, consider adding those individuals directly to Wikidata, whether or not your project has a separate web repository, so that others can query Wikidata to retrieve information.

- We can use machine translation to do "first drafts" of Wikidata, such as transliteration into non-Roman scripts and the creation of short statements based on identity categories, in order to make marginalized writers appear as writers across Wikidata languages and communities. In particular, it is possible to create—computationally and/or in bulk—short biographies of writers who appear in Wikidata from statements about writers. Once these are visible to users and editors, they can be corrected or modified by native or near-native speakers.

- For those whose work focuses on writers who are women or gender minorities, consider adding attributes to the items associated with them in Wikidata so that they appear in lists or can be queried. For example, adding Francophone writers from non-Francophone countries or countries where French is no longer an official language, as well as other languages that the person speaks or writes, can make that person visible in queries from researchers interested in all of those topics. Writers like George Sand from dominant countries and languages benefit from this effect already, but writers who are principally known in "peripheral" literary communities that are denied exposure within the global context need an extra push to make them more broadly discoverable.

This article is, thus, in part a call for literary historians who work on marginalized groups to consider adding their data to Wikidata in the interests of greater equity. Scholars have often described Wikipedia and Wikidata as unequitable but have less often intervened to improve inclusivity and the notability of marginalized writers. By completing these (sometimes laborious) data tasks, we can improve both the quality and the representativeness of Wikidata. Working within Wikidata, or adding curated and sourced data from another project, in many ways goes against academic notions of research credit. Yet if the data collected is tracked as the Women Writers project/database has

done, then academic reputations can be built and quality can be monitored in ways that are consistent with academic research expectations. As an international project that includes people from all regions and all identity groups to some degree, Wikipedia has brought readers and critics of all kinds into conversation. Although that conversation has not always been totally equitable, the community has the tools and the proven creativity to resolve issues of equity.

Data Repository: https://doi.org/10.7910/DVN/5PERE3
Peer reviewer: Sandra Folie (University of Jena)

REFERENCES

Baigent, Elizabeth, et al. "Gender in the Archive: Women in the Oxford Dictionary of National Biography and the Oxford English Dictionary." *Archives: The Journal of the British Records Association,* vol. 30, 2005, pp. 13–35, doi:10.3828/archives.2005.2.

Beytía, Pablo, et al. "Visual Gender Biases in Wikipedia: A Systematic Evaluation Across the Ten Most Spoken Languages." *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 16, 2022, doi:10.31235/osf.io/59rey.

Blakesley, Jacob. "World Literature According to Wikipedia Popularity and Book Translations: The Case of Modern Italian Poets." *Comparative Critical Studies*, vol. 17, no. 3, Oct. 2020, pp. 433–58, doi:10.3366/ccs.2020.0373.

"Clarisse Tremblay." *Wikidata*, http://www.wikidata.org/wiki/Q2975638.

Conroy, Melanie. "Literary Wikidata." *Github Repository*, http://github.com/mrconroy/literary-wikidata. Accessed 21 Aug. 2022.

Fauconnier, Saundra. "Making Women More Visible Online—With Wikidata Tools!" *Wikimedia Foundation*, 2018, https://diff.wikimedia.org/2018/03/29/increasing-visibility-women-with-wikidata/.

Flanders, Julia, et al. *Women Writers Project*. http://www.wwp.northeastern.edu.

Gargam, Adeline. "Savoirs mondains, savoirs savants : les femmes et leurs cabinets de curiosités au siècle des Lumières." *Genre histoire*, vol. 5, no. 5, Nov. 2009, doi:10.4000/genrehistoire.899.

"Gender Bias on Wikipedia." *Wikipedia*, http://en.wikipedia.org/wiki/Gender_bias_on_Wikipedia. Accessed 21 Aug. 2022.

"George Sand." *Wikidata*, http://www.wikidata.org/wiki/Q3816. Accessed 21 Aug. 2022.

Glott, Ruediger, et al. "Wikipedia Survey– Overview of Result." *United Nations University: Collaborative Creativity Group*, vol. 8, 2010, pp. 1158-1178, https://tinyurl.com/WikipediaUNStudy.

Graells-Garrido, Eduardo, et al. "First Women, Second Sex: Gender Bias in Wikipedia." *Proceedings of the 26th ACM Conference on Hypertext Social Media*, 2015, doi:10.1145/2700171.2791036.

Hube, Christoph, et al. "World Literature According to Wikipedia: Introduction to a DBpedia-Based Framework." *arXiv Preprint*, 2017, https://arxiv.org/pdf/1701.00991.pdf.

"Introduction." *Wikidata*, http://www.wikidata.org/wiki/Wikidata:Introduction.

Kaffee, Lucie-Aimée, et al. "When Humans and Machines Collaborate: Cross-Lingual Label Editing in Wikidata." *Proceedings of the 15th International Symposium on Open Collaboration*, Aug. 2019, doi:10.1145/3306446.3340826.

Kaplan, Marijn S. "Publication, Authorship, and Ownership in Marie Jeanne Riccoboni." *The French Review*, vol. 88, no. 1, 2014, pp. 179–91, doi:10.1353/tfr.2014.0115.

Klein, Maximilian, et al. "Monitoring the Gender Gap with Wikidata Human Gender Indicators." *Proceedings of the 12th International Symposium on Open Collaboration*, 2016, doi:10.1145/2957792.2957798.

Konieczny, Piotr, and Maximilian Klein. "Gender Gap through Time and Space: A Journey through Wikipedia Biographies via the Wikidata Human Gender Indicator." *New Media & Society*, vol. 20, no. 12, June 2018, pp. 4608–33, doi:10.1177/1461444818779080.

Lemus-Rojas, Mairely. "Contributing to the Representation of Women Artists in Wikidata." *Museum Computer Network Conference*, 2019, https://hdl.handle.net/1805/21402.

Lindsey, Nerissa, et al. "Ethical Considerations of Including Gender Information in Open Knowledge Platforms." *KULA: Knowledge Creation, Dissemination, and Preservation Studies*, vol. 6, no. 3, July 2022, pp. 1–15, doi:10.18357/kula.228.

"Main Page." *Wikidata*, http://www.wikidata.org/wiki/Wikidata:Main_Page.

Olsen, M. "Écriture féminine: Searching for an Indefinable Practice?" *Literary and Linguistic Computing*, vol. 20, no. Suppl 1, Jan. 2005, pp. 147–64, doi:10.1093/llc/fqi020.

Paling, Emma. "Wikipedia's Hostility to Women." *The Atlantic*, Oct. 2015, https://www.theatlantic.com/technology/archive/2015/10/how-wikipedia-is-hostile-to-women/411619/.

Piscopo, Alessandro. "Wikidata: A New Paradigm of Human-Bot Collaboration?" *arXiv Preprint*, 2018, doi:10.48550/arXiv.1810.00931.

Plain, Gill, and Susan Sellers, editors. *A History of Feminist Literary Criticism*. Cambridge University Press, 2007.

Planté, Christine. "La place des femmes dans l'histoire littéraire: annexe, ou point de départ d'une relecture critique?" *Revue d'histoire littéraire de la France*, vol. 103, no. 3, 2003, pp. 655–68, http://hal.archives-ouvertes.fr/hal-00986581/.

Riot-Sarcey, Michèle, et al. *Le genre comme catégorie d'analyse. Sociologie, histoire, littérature*. L'Harmattan, 2003, doi:10.2307/3654973.

Saint Martin, Monique de. "Les 'femmes écrivains' et le champ littéraire." *Actes de la recherche en sciences sociales*, vol. 83, no. 1, 1990, pp. 52–56, http://www.persee.fr/doc/arss_0335-5322_1990_num_83_1_2936.

Schroder, Anne L. "Going Public against the Academy in 1784: Mme de Genlis Speaks Out on Gender Bias." *Eighteenth-Century Studies*, vol. 32, no. 3, 1999, pp. 376–82, doi:10.1353/ecs.1999.0023.

Shaik, Zaina, et al. *Analyzing Race and Country of Citizenship Bias in Wikidata*. arXiv, 2021, doi:10.48550/arXiv.2108.05412.

Steiner, Thomas. "Bots vs. Wikipedians, Anons vs. Logged-Ins (Redux) a Global Study of Edit Activity on Wikipedia and Wikidata." *Proceedings of The International Symposium on Open Collaboration*, Aug. 2014, pp. 1–7, doi:10.1145/2641580.2641613.

Stinson, Alex. "Writing a Wikidata Query: Discovering Women Writers from North Africa." *WikiIndaba 2018 Conference, Medium.Com*, 2018, https://tinyurl.com/northafricanwomen.

Thomas, Paul. "Wikipedia and Participatory Culture: Why Fans Edit." *Transformative Works and Cultures*, vol. 22, Sept. 2016, doi:10.3983/twc.2016.0902.

Tripodi, Francesca. "Ms. Categorized: Gender, Notability, and Inequality on Wikipedia." *New Media & Society*, June 2021, p. 146144482110237, doi:10.1177/14614448211023772.

Vrandečić, Denny, and Markus Krötzsch. "Wikidata: A Free Collaborative Knowledgebase." *Communications of the ACM*, vol. 57, no. 10, Sept. 2014, pp. 78–85, doi:10.1145/2629489.

Wagner, Claudia, et al. "Women through the Glass Ceiling: Gender Asymmetries in Wikipedia." *EPJ Data Science*, vol. 5, no. 1, Mar. 2016, pp. 1–24, doi:10.1140/epjds/s13688-016-0066-4.

Wikidata Query Service. https://query.wikidata.org. Accessed 15 Mar. 2022.

"WikiProject Women." *Wikidata*, http://www.wikidata.org/wiki/Wikidata:WikiProject_Women. Accessed 17 July 2022.

"WikiProject Women in Red." *Wikipedia*, http://en.wikipedia.org/wiki/Wikipedia:WikiProject_Women_in_Red. Accessed 17 July 2022.

Winn, Colette H. "Early Modern French Women Writers." *The Palgrave Encyclopedia of Early Modern Women's Writing*, 2020, pp. 1–9, doi:10.1007/978-3-030-01537-4_1-1.

Zhang, Charles Chuankai, and Loren Terveen. "Quantifying the Gap: A Case Study of Wikidata Gender Disparities." *17th International Symposium on Open Collaboration*, 2021, doi:10.1145/3479986.3479992.