

The Realienation of the Commons: Wikidata and the Ethics of “Free” Data

ZACHARY J. MCDOWELL
University of Illinois at Chicago, USA

MATTHEW A. VETTER
Indiana University of Pennsylvania, USA

Wikipedia’s founding in 2001, accompanied by the techno-optimism of Web 2.0 and the ambitious agenda for free knowledge, inspired countless volunteers to contribute. The success of the encyclopedia both inspired and provided evidence of the power of “wikinomics,” “crowd-sourcing,” and “commons-based peer production.” In many ways, Wikipedia, and its parent company Wikimedia, can be viewed as the standard-bearers of Web 2.0’s early promises for a free and open Web. However, the introduction of Wikipedia’s sister project Wikidata and its movement away from “share alike” licensing has dramatically shifted the relationship between editors and complicated Wikimedia’s ethics as it relates to the digital commons. This article investigates concerns surrounding what we term the “re-alienation of the commons,” especially as it relates to Google and other search engine companies’ reliance on data emerging from free/libre and open-source (FOSS/FLOSS) Web movements of the late 20th and early 21st centuries. Taking a Marxist approach, this article explores the labor relationship of editors to Wikimedia projects and how this “realienation” threatens this relationship, as well as the future of the community.

Keywords: digital commons, labor, alienation, Wikidata, Wikimedia, Wikipedia, knowledge panels, digital ethics

Now over 22 years old, Wikipedia has slowly but carefully matured into the Web’s “grown-up.” A top hit for nearly every Web search and one of the most heavily trafficked sites on the Internet, Wikipedia has been favorably compared in numerous studies to the reliability of Encyclopedia Britannica and other traditionally “vetted” reference sources (Brown, 2011; Casebourne, Davies, Fernandes, & Norman, 2012; Giles, 2005; Hwang, Bourgeois, & Seeger, 2014). Recently, Wikipedia was even enlisted by social media sites to counteract fake news, an indicator of its continuing relevance in the global knowledge economy (Cohen, 2018). Wikipedia’s long history remains a testament to its community’s labor as well as its “prime objective”—“Imagine a world in which every single person on the planet is given free access to the sum of all human knowledge” (“Prime Objective,” 2022, para. 1). Although he first coined the term commons-based

Zachary J. McDowell: zmcdowell@gmail.com

Matthew A. Vetter: mvetter@iup.edu

Date submitted: 11-9-2022

Copyright © 2024 (Zachary J. McDowell and Matthew A. Vetter). Licensed under the Creative Commons Attribution Non-commercial No Derivatives (by-nc-nd). Available at <http://ijoc.org>.

peer production more than 20 years ago (Benkler, 2002), Yochai Benkler (2020) more recently noted that “Wikipedia and commons-based peer production (CBPP) more generally continue to offer an existence proof that there can be another way”—an alternative to the more prevailing forces of market exchange and surveillance that characterize the current Web (p. 43). Indeed, much has changed in the last 20 years, and the sharing economy of the early Web has been almost wholly corporatized. Cooperation, sharing, and peer production continue as prevalent practices in digital communities, but very little of that collaboration is truly “commons-based,” that is, nonprofit and for the good of a community or society at large (Jemielniak & Przegalińska, 2020). Benkler’s (2020) take encapsulates the increasingly rare value in the ethos of the digital commons that continues to inspire the labor that creates Wikipedia, as well as other digital projects and communities in the free/libre and open-source movements (FOSS/FLOSS).

When Benkler (2020) highlights the value of CBPP as “another way,” it is not just against these market forces on the Internet and within modern capitalist societies but also in the way in which most information and knowledge is held, transmitted, and put to use. All mediums of writing can be used for both capture and enclosure, as well as transmission of information. Mediums can interface and contort, and different mediums have different “ethics” insofar as they “do things” (Galloway, 2013, p. 23) that can then conserve and control information’s movement and distribution. In contrast to CBPP—where information is shared freely and distributed openly—much of the history of written information has been one of conservation and control, limiting access to information to a select few, whether because of cost, location, or other factors. Those who control the medium of capture and storage can choose how to employ the medium’s manner of control, and whether stone tablets, books, or paywalled news organizations, methods of capturing and storing information have long kept knowledge from the larger populace, including subjects of that knowledge (or even creators of it). In this manner, knowledge has often been alienated from knowledge-making communities, and the ideals of CBPP remain quite radical.¹ The commons has often faced legal action; in particular, Project Gutenberg and The Internet Archive have fought for preservation (Bailey, 2022). However, recent developments in the Wikimedia ecosystem complicate (and even threaten) the ethos and sustainability of the digital commons.

In this article, we investigate ethical concerns surrounding what we term the “realienation of the commons,” especially as it relates to the extraction and commodification of datafied knowledge emerging from FOSS/FLOSS Web movements of the late 20th and early 21st centuries. Wikimedia projects especially are susceptible to large-scale reuse for commercial intent, whether it be by search engines such as Google’s Knowledge Graph, virtual assistants (VAs) such as Siri and Alexa, or more recently large language model generative artificial intelligence (AI) applications such as ChatGPT or Google’s Bard. To accomplish this, we first establish context for the ethics of CBPP as it relates to Wikipedia. We then examine the case of Wikipedia’s sister project, Wikidata, the structured data peer production (SDPP) platform that has emerged as the largest open knowledge database in the Web’s history. After contrasting and contextualizing the ethical implications of Wikidata using Benkler’s (2020)

¹ Although original copyright laws in both the United States and the United Kingdom both intended to secure rights “for a limited time,” and the expiration of these would ensure that knowledge became “public domain,” changes to these laws have increased this from 14 years to 70 years after the death of the author, ostensibly creating an indefinite system of ownership.

framework, we call for more critical scrutiny of the project in terms of (1) its usage of the Creative Commons CC0 “No Rights Reserved” license (equivalent to licensing as public domain and waiving all rights),² and (2) how knowledge from Wikidata is extracted, reappropriated, and commodified beyond the intent of its original creators (or Wikidata contributors), which we term “realienation.” Other scholars have problematized the ways that Wikidata’s licensing results in a loss of data provenance via automated extraction (Ford, 2022; Iliadis, 2023), but this article attends to the motivations of Wikimedia contributors as they lose confidence in the promise of the digital commons. Finally, centering this motivational concern, we explore trust issues with Wikimedia, suggesting the need to rethink big tech’s approach to data ethics, and how this affects a variety of data uses, particularly AI systems. We conclude with a discussion of what’s at stake in the future of Wikimedia as an open knowledge movement, and we share suggestions about how Wikidata should define its relationships with other platforms going forward.

Licensing “Another Way”

Benkler’s (2020) promise that “there can be another way” centers around the type of openness that creates (and can ensure) additional openness. Such an ideological direction not only is a practice but also is backed up by legal language through a Creative Commons license. The Creative Commons “Attribution—ShareAlike” (CC-BY-SA), the license that Wikipedia (and much of Wikimedia’s projects) is licensed under, is a critical legal tool because it enables a type of openness that ensures perpetual openness. A CC-BY-SA license allows for individuals or businesses to share, or “copy and redistribute the material in any medium or format,” adapt, or “remix, transform, and build upon the material for any purpose, even commercially,” and that the license is irrevocable as long as the license is followed (Creative Commons, 2023a, para. 1). The terms of the license are fairly simple, requiring that attribution (“appropriate credit, provide a link to the license, and indicate changes were made”) and that any content built upon or remixed must also be distributed with the same license as the original (hence “share alike”) (Creative Commons, 2023a, para. 3). This is one of the licenses that Creative Commons deems a “free culture” license, precisely because it encourages a CBPP-informed ethos. In fact, without Creative Commons, Wikipedia (and Wikimedia in general) would never have become the standard-bearers of Web 2.0’s early promises for a free and open Web. Even from its early days, Wikipedia and its parent organization both embodied this ethos in the free knowledge work they made possible as well as in the systems they created to encourage (and demand) future “freedom.”

Importantly, CC-BY-SA is open in a way that is also closed—it shuts off information for particular types of usages to preserve the openness. This way ensures that what is common creates more “common” and creates a new type of oikos (home) in the oikonomos (economy), one that focuses on sharing and giving away for the common good and walling off that which is “common” from those who wish to extract it. This type of restructuring, which encourages reuse while demanding both attribution

² Creative Commons defines CC0 as a lack of copyright, because the work is deeded to the public domain, and provides the following explanation: “The person who associated a work with this deed has dedicated the work to the public domain. . . . You can copy, modify, distribute and perform the work, even for commercial purposes, all without asking permission” (Creative Commons, 2023b, para. 1).

and license adoption, restructures the potential for economies of information, because it ensures that the information is free and that anything done with that information must be free. This is a radical departure from many practices of big tech companies, despite their reliance on much “open” software and information.

For quite some time, the relationship between Google and Wikipedia had been (somewhat) mutually beneficial. Google utilized Wikipedia’s information in its knowledge graphs and information boxes. In exchange, Google donated on multiple occasions to Wikipedia and the Wikimedia Foundation (WMF) (Litman-Navarro, 2019), and the *Guardian* has labeled this philanthropy an “investment” (Johnson & Francisco, 2010, para. 10). A \$2 million grant in 2010 was one of the earliest well-publicized grants from Google. In 2019, another major gift went to both the WMF, in the amount of \$1.1 million, and to the WMF Endowment, in the amount of \$2 million (Dickey, 2019). In addition, the initial start-up money for Wikidata, \$1.7 million, was partially funded by Google, alongside Microsoft’s Paul Allen Foundation and the Gordon and Betty Moore Foundation (Terdiman, 2012). Although one could argue that this relationship had been one where Google extracted significantly more revenue than it offered to Wikipedia, it still represented a two-sided exchange.

With the rise of the Semantic Web, however, requirements started to change when systems started “doing things” with data—and CC-BY-SA created a barrier to streamlining the process from data to product (particularly in AI-powered apps and systems). Essentially, the sharing requirement for Wikipedia’s information curbs the ability to monetize products effectively, presenting significant challenges to the big tech companies who wish to use the information and data. The turn toward CC0 licensing creates a space for the selling off of and disconnection of labor-product with the creators of that information.

Wikipedia’s Radicalism and Wikidata’s Turn

Wikipedia “frees” knowledge not only because it displays the (arduous and meticulous) labors of thousands of authors for free (both cost and usage) but also because it acts as a (verifiable) representation of countless knowledge sources that are not “free” (as in usage and accessibility but also often cost) and otherwise unavailable for public viewing and consumption (e.g., paywalled or subscription-only publications). CC-BY-SA is key here because it creates the type of space that can act as a reconnection with knowledge for the general public—it recaptures that which has been sold off or otherwise alienated from the larger knowledge commons.

The emergence of Wikipedia and Creative Commons in 2001 came about alongside the techno-optimistic zeitgeist of the 1990s and early 2000s, characterized by a belief in the power of Internet technology to add value to everyday users’ lives through sharing, collaboration, and the seemingly limitless democratic potential of the Web. Wikipedia volunteer labor remains unique, so much so that numerous books and research studies have attempted to understand what made it and other related projects successful (Brabham, 2013; Shirky, 2009; Tapscott & Williams, 2008). One way to understand the economics behind Wikipedia (but certainly not the only way) is to consider how volunteer editors, living under late capitalism and performing other wage labor as day jobs, became attracted to the techno-optimistic narrative of Wikipedia as well as its economic mode of production as something strictly beyond or outside their typical

relationship to work. In Wikipedia, editors volunteer to create new knowledge that is then made freely available—not for the benefit of a wage, but because doing so allows them to participate, access, and share the outcome of that labor. This is remarkable for many reasons, but mainly because such motivational factors often operate outside traditional market forces.

In this way, Wikipedia editing allowed the average editor to subvert the capitalist status quo. The Wikipedia community was created around this new economic model—CBPP, which connected editors with their labor and connected other editors to each other through that labor. As it manifested in Wikipedia, CBPP enabled a community of volunteers who sought to connect with the world through sharing knowledge. Karl Marx (2012) defined alienation as “appropriation as estrangement” and stated that “realisation of labour appears as loss of realisation for the workers” (p. 69). Marx’s concept here refers to the relationship between the product of the labor and how it is both used and disconnected from the laborer. This relationship with labor (and the community around it) marks the important distinction that helps illustrate our use of the term “realienation” with regard to Marx’s usage of “alienation.” This distinction arises due to the circumstances that editors find themselves in 20 years later as AI, large language models, and structured data systems and applications bring about new ways to capture, extract, and commodify CBPP labor, of which Wikimedia projects are perhaps one the most valuable and susceptible targets.

In contrast, Wikidata serves as a radically open, structured knowledge database that appears, at first glance, to continue the openness of the commons, but shifts the way in which the data are used. As an enormous database with nearly 100 million data items that can be (and are) used freely in machine learning networks to create new products, Wikidata is the largest Semantic Web platform humanity has ever created and the largest system for “doing things” with data, whether simple calculation or training machine learning systems (Iliadis, 2023; “Statistics,” 2023). If Wikipedia was the promise of Web 2.0, Wikidata represents Web 3.0, connected data in a Semantic Web.

Launched in 2012, Wikidata is a collaboratively edited knowledge base that is the structured data repository for all Wikimedia projects, particularly Wikipedia. Wikidata employs user-generated metadata standards that reflect consensus in the scientific community as well as already-established standards. As the backbone for Wikipedia, it is the repository of information that is utilized by a variety of projects that represent information on a variety of platforms (e.g., Facebook fact-checking and Google search results). The data model for entities in Wikidata utilizes specific subfields’ data structures so that data can be passed back and forth between databases. Thus, Wikidata uses application programming interfaces (APIs) to “talk to” entities in other databases developed and maintained by National Institutes of Health, National Center for Biotechnology Information, Ensembl, and Homologene, among others, helping to reinforce core concepts around scientific, structured data. Wikidata is “free, collaborative, and multilingual” (“Introduction,” 2022, para. 1) and, as such, offers numerous possibilities for the advancement of Wikimedia’s overall mission of bringing “free educational content to the world” (Wikimedia, 2022, para. 1). And yet the project’s use of a “No Rights Reserved” CC0 license also exposes the database to corporate actors that might not share Wikimedia’s altruistic values and goals. We connect Marxist theory to the extraction, appropriation, commodification, and alienation that occurs because of this exposure as a way to explore the relationship

with the community labor that creates and maintains Wikipedia and that of larger corporate interests utilizing the information for profit.

“No Rights Reserved” and Wikidata’s Usage of CC0

Instead of Wikipedia’s CC-BY-SA (“share alike”) license (a license that requires derivatives and other uses of the licensed material to retain the same license), Wikidata utilizes a license that has no requirements.³ This might sound ideal for “freedom,” but in reality, Wikidata seems to appropriate that particular FOSS imaginary of sharing while instead delicensing information into data by assigning it a CC0 license—allowing companies to extract, commodify, and otherwise use these data in ways to create systems without requirements to honor the license or reference the works that were utilized.

Wikidata sounds like a huge step forward in providing “free” data. However, the ways in which large tech companies (in particular those training machine learning systems) use these data and where the data originated highlight ethical concerns about data provenance and the exploitation of volunteer labor. What Wikidata shifts is the way in which the commons are no longer preserved and utilized, creating concerns for continued and increased extraction and the abandonment of the CBPP model in favor of a model that drops the “C” for “Commons” from its acronym: namely, from commons-based peer production (CDPP to structured data peer production (SDPP).

At least some Wikimedians recognized these concerns early on in Wikidata’s inception, as Andreas Kolbe (2015) writes for the Wikipedia Signpost: The “no-attribution CC0 license means that third parties can use the data on their sites without indicating their provenance, obscuring the fact that the data came from a crowdsourced project subject to the customary disclaimers” (n.p.). Perhaps even more significant, Wikidata’s usage of CC0 enables the reappropriation of content originally licensed under a CC-BY-SA license and created by a volunteer community in Wikipedia. Wikidata is composed (at least partially) of scraped Wikipedia metadata to populate its system with data, skirting the copyright issue. Although metadata is not copyrightable, the original information that created this metadata was created by countless volunteer hours under the guise of this “share alike” license. However, much of the community, including those directly contributing to Wikidata, often remains in the dark about how this information is used (Zhang, Houtti, Smith, Kong, & Terveen, 2022).

Wikidata and its system and licensing operate in such a way as to maintain nearly complete opacity in terms of the end-user due to the way end-users use other software intermediaries for mining the knowledge base. Despite actively populating, developing, and contributing to it, it is of deep concern that the Wikimedia community and Wikidata volunteers know very little with regard to how third-party consumers use Wikidata. These “invisible machine intermediaries make it difficult for contributors to access information about how their contributions are used by others” (Zhang et al., 2022, p. 14), which is not the case in sister projects like Wikipedia. An additional finding of Zhang and colleagues (2022) is the report from Wikidata contributors that some editors (certainly this may be a minority given how little is known about end-users) experience feelings of

³ It is important to note that the CC-BY-SA license does not prohibit commercial uses, and Wikipedia itself can be copied, used, and sold by anyone and any organization, as long as the license requirements are met.

“aversion . . . towards the idea of volunteering their time in a way that creates profit for large businesses, e.g. Google” (p. 15). This circumstance presents a catch-22: volunteer editors might be demotivated because they do not know who or what they are contributing to, but they would be equally demotivated to know that their contributions are serving big tech’s profits. These findings are especially interesting considering the prevalence of the WMF’s connection with Google, suggesting that editors are either not making the connection between the two or simply ignoring the connection in their altruistic pursuits.

The Wikipedia Detour: How Wikidata’s CC0 License Can Break the Reference Trail

One of the reasons that Wikipedia has become so popular as a source of information is because of the WMF’s interdependence and relationship with Google and other tech corporations (McMahon, Johnson, & Hecht, 2017; Vincent & Hecht, 2021; Vincent, Johnson, & Hecht, 2018). McMahon and colleagues (2017) have demonstrated a mutual and “extensive interdependence” between the two tech giants, acknowledging how without Wikipedia included in its search results, Google is “simply a worse search engine for many queries” (p. 148). In contrast, because Wikipedia articles appear so prevalently in Google search results, the search engine is the source of the majority of Wikipedia’s page views (readership), from which it both raises funds via individual donations and recruits editors (McMahon et al., 2017). Although such interdependence was never completely mutual given Wikimedia’s status as a nonprofit, in recent years, the relationship has become much more one-sided, because the introduction of Google Knowledge Graph has significantly reduced traffic to Wikipedia as well as average Web users’ understanding of where information comes from when sourced from Wikipedia (McMahon et al., 2017).

VAs such as Siri, Google Home, and Alexa often provide extractions of information from Wikipedia without attribution, particularly without the context that that information is socially and historically constructed by volunteers in the Wikipedia community via community-mediated policies and practices (McDowell & Vetter, 2021). However, at the same time, the same users of those VAs are often unwilling to trust Wikipedia, often remembering information they were told a decade ago about the encyclopedia’s unreliability (boyd, 2017, p. 84). This concern is not new—much of the early criticism surrounding the issue revolves around the evolution of Google’s Knowledge Graph, aka “Knowledge Panels” (Ford, 2020). Seemingly benign, these “infoboxes” repackaged Wikipedia’s information for over a decade, displaying Wikipedia’s information alongside search results. Although a small blue link often allows the user to navigate to Wikipedia for the full article, the infobox remains part of the search results, furthering the “oracular” usage of Google search in response to user queries (McDowell & Vetter, 2022). Wikidata, however, allows for more complex and radical ways to repackage and reuse information and data.

In an interview quoted by Kolbe (2015) and originally conducted by Heather Ford (2014), Max Klein notes:

Wikidata being CC0 at first seemed very radical to me. But one thing I noticed was that increasingly this will mean where the Google Knowledge Graph now credits their “info-cards” to Wikipedia, the attribution will just start disappearing. This seems mostly innocent until you consider that Google is a funder of the Wikidata project. So in some way, it could seem like they are just paying to remove a blemish on their perceived omniscience. (Ford, 2014, n.p.)

More recently, Heather Ford (2020) describes the evolution of Google Knowledge Panels and their treatment of attribution:

As panels evolved, blue links to Wikipedia articles started shrinking in size. Over time, the underscore was removed so that the links weren't clickable, and then the links were lightened to a barely visible grey tone. Now, facts under the opening paragraph tend not to be cited at all, and hyperlinked statements refer back to other Google pages. (p. 196)

According to Ford and Graham (2016), Google "refuses to answer questions about how its results are garnered" (p. 962). Although more recent studies have shown that Wikipedia is referenced in a majority of infoboxes across devices and browsers (Vincent & Hecht, 2021), there is no guarantee that this will continue because Google is under no legal obligation to make a reference to where the data are actually coming from.

What Klein, Ford, and Graham have all acknowledged is that this issue is symptomatic of a larger problem regarding data licensing—it is not just the case that Google is able to, as Klein noted earlier, "remove a blemish on their perceived omniscience," but instead that CC0 allows them to display data unattributed. The data become part of their larger data ecosystem and is able to be used in any way they deem fit—whether simply stating facts in the knowledge panels (which, of course, also power VAs and train newer AI systems) or utilizing them in machine learning systems. This disassociation of data sourcing breaks the reference trail that is so important to Wikipedia's accountability (verifiability of sources remains key to how it operates) but also breaks the promise of Wikipedia's licensing—that information contributed to Wikipedia is legally secured as "share alike." The implications for user behavior are significant: without attribution, and often even with a proper reference, most casual users detour Wikipedia altogether, relying on the trust they place in Google instead of critically interacting with the information and its sources on Wikipedia.

When Google uses Wikidata's data derived from Wikipedia, it now has free access to do with it what it will, because it has been transformed from CC-BY-SA into CC0. Not only does this violate the citational ethos of the commons, but also this extracted data can now be used to create new products, train algorithmic systems, and create copyrightable, wholly owned systems and ideas—all from information that was created, sourced, and labored on as a gift to the commons. This extraction violates the ethics of CBPP and the promise of the commons, and it should concern anyone involved in the countless hours of labor volunteered to support this dream. Our analysis of policy and data utilization illustrates how we should be concerned about the extraction of the commons and how technology giants continue to enclose, extract, repackage, and resell our own labor back to us. This is not new, of course, just a shift in how it is done—even with clear attribution or reference back to Wikipedia, there is clear evidence that the labor of Wikipedia volunteers subsidizes Google's search engine results (Vincent & Hecht, 2021).

Recent investigations of Google's appropriation of Wikipedia content into its systems have shown that even before the radical transformation through AI machine learning systems, such appropriation can have devastating effects on Wikipedia's ecosystem, because not only do "users rarely know that the answers to their questions come from Wikipedia" (Jemielniak & Przegalińska, 2020, pp. 20–21), but also this precludes users from accessing and experiencing Wikipedia, which can reduce its

ability to find new editors, users, and potential donors. New users are absolutely imperative because “only a typically small fraction of Wikipedia readers volunteer to write and edit its entries, and there is a natural burnout; and because volunteer engagement in crowdsourced movements escalates slowly, through gradually increased participation” (Jemielniak & Przegalińska, 2020, pp. 20–21). More to the point, Wikipedia has a distinct problem recruiting new editors, particularly in diversifying editorship. The redirection away from Wikipedia only compounds existing problems, with editor recruitment becoming even more problematic when seen along the shift in the labor relationship that editors can experience (Jemielniak & Przegalińska, 2020).

Realienation of the Commons

Due to the WMF’s ties to tech giant donors such as Google and how Google actively utilizes Wikimedia project data to train AIs (“Announcing WIT,” n.d.), there remain serious concerns about the ethics of utilizing the commons to create new “products” that are then sold back to those who toiled tirelessly to create what now feeds tech giants’ machines. Although Google has gifted donations to Wikimedia on multiple occasions, Vincent and Hecht (2021) argue that these “donations” from Google “represent only a tiny fraction of the value the Wikipedia community’s labor has created for these companies” (pp. 11–12). Not only have the contributions been undervalued, but Vincent and Hecht (2021) also argue the community has lost control over the product of their labor (p. 12).

Because Wikipedia and other FLOSS movements (supposedly) function so differently from traditional movements of capital, it is important to understand some of the framings with regard to CBPP. Discussed by Benkler and Nissenbaum (2006) and more recently in relation to Wikipedia by Benkler (2020), these scholars sought to distinguish Wikipedia and other FLOSS movements as existing outside capitalism’s distinguishing features of “private property, commodified exchange, and wage labor” (Benkler, 2020, p. 44). That is, in its conception and following the theory to its realization, Wikipedia and other early FLOSS projects allowed its producers a “freedom from markets” that brought them closer to a meaningful exchange (and relationship) with their labor. We might understand the mystifying success of some of these projects as due, in some small part, to the ways that the economies of Wikipedia and other CBPP, in opposition to the typical capitalist alienation, actually fosters a relationship between the worker and their labor. At its core, the Wikipedia volunteer/worker is able to choose their writing subject (thus, the object they produce) as well as benefit from their labor (and understand the benefits of their labors to others) via a digital commons for free knowledge. Such a relationship, and its value and meaning, stands in direct opposition to Marx’s notion of “alienation” as applied to wage labor in capitalist societies. Furthermore, because Wikipedia does not sell advertisements, it does not surveil its users or capture data for use in the same way as most of the rest of the Web. Benkler (2020) describes this relationship by thinking broadly about what CBPP offers: “the ability to live our lives under the constraints of social relations other than those dictated by the need to buy and sell labor to obtain the basic necessities of life” (p. 44).

Marx’s concept of alienation is appropriate because of the recent turn toward transforming (“objectified,” via metadata and into Wikidata) the products of socially donated labor, which become “dispossessed” (literally depriving the information of its licensing under CC-BY-SA) and then extracted

into capital (by companies using the data to make new products) (Marx, 1993, p. 832). In other words, Wikipedians donate labor to create content on Wikipedia and Wikidata under the guise of gathering and distributing knowledge for free, under a copyright license that specifically requires any use of that content must carry a license ensuring not only its continued freedom (share alike) but also that it can be traced back (attribution), and that content is then extracted from that site of sharing using a legal loophole that strips it of its licensing for use in the creation of wholly owned and licensed products for multibillion-dollar corporations with no attribution. As Sartre (1984) notes, alienation “begins with exploitation” (p. 227), and the methods of extraction and usage of labor donated under the guise of sharing seem fairly exploitative.

This creates what we refer to as “realienation of the commons,” in which the fundamental agreement to donate labor under the guise of the commons and sharing, to what Benkler (2020) refers to as “another way,” has become broken and the product of that labor utilized, transformed, monetized, and sold back to the community (a community that worked so hard to create something shared). The way of CBPP attempted to subvert this massive alienating system by creating a community of sharing that reunites people with meaningful work, creating relationships between the laborer and the product of that labor. However, the erosion and privatization of the commons results in new methods of extraction and estrangement, and these new methods could not occur without exploitation of the licensing schemes.

Estranged labor, according to Marx, is the value exchange between the worker and their production of wealth: “The worker becomes an ever cheaper commodity the more commodities he creates. . . . Labour produces not only commodities: it produces itself and the worker as a commodity—and this at the same rate at which it produces commodities in general” (Marx & Engels, 1978, p. 71). We can think here not just of the Wikipedian as a content creator but also of the labor for that content, which becomes less about content and more about raw material, grist to be used by corporate interests, distancing the Wikipedian from the content they helped to create. Marx’s conceptualization of “estranged labor” describes a form of alienation in which the worker’s relationship to their labor, as well as the object they produce, exists in a system of value exchange apart from the worker. Although in some ways all labor is somewhat alienating, what we want to differentiate here is the continued ownership and connection with the labor and its product, which is unique to CBBP. For content creators on Wikipedia, this is simple—the value exchange happens in the form of extracted labor that is then used in a way so foreign to the original intention of the laborer that it becomes unrecognizable. Of course, the products of this labor then become “something alien, as a power independent of the producer” (Marx, 2012, p. 69). Marx (2012) continues: “Labour realisation is its objectification. Under these economic conditions this realisation of labour appears as loss of realisation for the workers; objectification as loss of the object and bondage to it; appropriation as estrangement, as alienation” (Marx, 2012, p. 69). Marx characterizes this objectification as the worker’s loss of any meaningful relationship to their own labor and to what they might produce. This runs completely counter to the way of CBPP.

What is perhaps most problematic as it concerns this turn toward big tech is that the average Wikimedian contributes because they are motivated by altruism, the promise of the “digital commons,” and they are also largely unaware of the project’s commodification by tech giants like Google (Li et al.,

2022; Xu & Li, 2015). Zhang and colleagues (2022) have identified two prevalent archetypes of Wikidata editors more specifically: the architect and the mason. Architects are those editors who are mainly interested in developing the "ontological infrastructure of Wikidata"; masons are those filling in the gaps, "building the database through data entry and editing" (Zhang et al., 2022, p. 10). For both archetypes, common motivations for volunteering their time to work on Wikidata were consistent with previous research on other Wikimedia projects and open-source initiatives and included entertainment (Dholakia, Bagozzi, & Pearo, 2004), altruism (Kuznetsov, 2006; Oreg & Nov, 2008) and the desire to spread knowledge (Farič & Potts, 2014; Kuznetsov, 2006), as well as the added incentive of being able to work on structured activities.

In contrast to the motivations of Wikipedians, and especially divergent when thinking about other instances of CBPP projects, a majority of the participants in Zhang and colleagues' (2022) study "did not have a concrete understanding of how their contributions are actually used, or by whom" (p. 13). This "usage opacity," as Zhang and colleagues (2022) label the phenomenon, represents not only a realiation of the volunteer contributor but also a unique problem for Wikidata's continued sustainability. On the one hand, Wikidata editors could be further motivated to contribute if they had more information about the use and usefulness of their work. On the other hand, the opacity also serves to keep the "invisible machines" (i.e., for-profit entities like Google that commodify volunteer labor) concealed. "Revealing the machines and increasing usage transparency," although an immediate solution to the problem of opacity, could also "function as a demotivating factor for some editors, if the machines are run by for-profit corporations" (Zhang et al., 2022, p. 18).

In some ways, Wikidata actually starts to appear as what Fuchs (2020) refers to as a "pseudo-commons" (p. 123) because it obscures its participation in digital capitalism. Fuchs originally used this term to differentiate between projects like Wikipedia on the one hand and Google on the other, but the term here becomes a useful heuristic in understanding Wikipedia's participation in a type of shadow-capitalism alongside its FOSS/FLOSS rhetoric. Pseudo-commons are those that "provide free access to certain digital resources, crowdsource human labor so that it is performed online and unpaid, and accumulate capital in ways that prevent humans from experiencing their exploitation in an immediate manner" (Fuchs, 2020, p. 123).

In many ways, Wikipedia's "free and open" promise has enabled the potential for a 20-year-long "bait and switch" in which volunteer labor has been transformed into corporate subsidies, and the donations of Google to Wikimedia pale in comparison to the ways that the search engine profits from the labor of volunteer Wikipedia editors (Vincent & Hecht, 2021). The more recent formation of Wikimedia Enterprise enabled the two companies to enter into a more formal business relationship (Roth, 2022).⁴ Wikidata, and its use of the CC0 license, allows for the extraction and commodification of the volunteer labor that helped make Wikipedia valuable in the first place. As machine learning, AI, and other large-scale uses of Wikidata become more prevalent, there remains a critical need to critique big tech's appropriation and

⁴ WMF launched Enterprise, their first for-profit project, in October 2021, and it later announced its first (paying) customer, Google, in June 2022. Enterprise is a service that offers "Enterprise-grade APIs Built for Search, Social, and Voice Assistants" ("Wikimedia Enterprise," 2023, para. 1) to data and information in Wikimedia's products (Wikipedia, Wikidata, Wiki Commons, etc.).

commodification of volunteer labor, as well as serious concerns around data equity, data provenance, and the original promise of CBPP.

What's at Stake?—The Future of Wikipedia

If Wikipedia and other similar projects offered a unification (as opposed to alienation) for workers to their labor (through freedom from markets and the promise of a digital commons), then the realienation of the worker to that labor occurs when that promise of CBPP is broken. For CBPP and Wikipedia in general, this should be unthinkable—the cocreation of Wikipedia is as much of a relationship to that labor as it is to the dedication to the movement. Accordingly, we ask the question: What might happen if Wikipedians are estranged from Wikipedia? Asking this question helps to think not only about the individuals currently editing Wikipedia but also about its potential future editors—something that the WMF has been aware of and concerned about for more than a decade (Hill, 2012). Although more than 44 million usernames are currently registered in the English language version, only about 110,000 of these actively contribute. Furthermore, an even smaller number of these contributors participate in community discussions (Wikipedians, 2022). Equally problematic is the fact that this editorial demographic is severely lacking in diversity, with one study reporting that 90% of editors identified as male (“Community Insights,” 2018). When Wikipedia is effectively bypassed by AI apps such as Google Knowledge Graph, VAs (e.g., Siri, Alexa) and instances of large language model Chatbots, the encyclopedia receives less traffic, fewer individual donations, and, more importantly, less engagement with the production of knowledge overall.

In their discussion of digital assistants, or “voice interfaces” such as Siri and Alexa, Jemielniak and Przegalińska (2020) note that “users rarely know that the answers to their questions come from Wikipedia” (p. 20). Because of this, users may not feel the need to investigate a question further, especially when and if the “Google sidebar supplied sufficient information” (Jemielniak & Przegalińska, 2020, p. 20). Although Google’s knowledge graphs have evolved over the years and currently do provide a link back to Wikipedia in some instances, there is no guarantee that this will persist, because there is no legal obligation to do so. As Jemielniak and Przegalińska (2020) argue, the circumvention of Web traffic away from Wikipedia “significantly limits not only Wikipedia’s ability to raise funds but also to attract new users” (pp. 20–21). In essence, the threat of Google’s commodification of Wikidata is at least twofold, when it comes to what is at stake for Wikipedia’s future. Not only are fewer readers directed to Wikipedia through their search for information, resulting in lost opportunities for critical engagement with Wikipedia’s knowledge-making processes and policies, but also even fewer readers are introduced to Wikipedia as a possible site for community engagement and interaction. The encyclopedia’s community of editors peaked in 2007 and has yet to reach the same level of participation (Simonite, 2013). Yet problems related to diversity, bias, and content gaps remain quite pressing (McDowell & Vetter, 2021, p. 71).

These are the most tangible and perhaps immediate effects of Wikidata’s CC0 compromise. Yet there is something bigger at stake, something fundamental to Wikipedia. The community’s success remains largely due to the techno-optimistic and even Enlightenment-era ideology that accompanied Web 2.0, and the narratives around crowd-sourcing and CBPP drove the altruistic motivations of the editors. Wikipedia’s vision for “a world in which every single person on the planet is given free access to the sum of all human knowledge” (“Prime Objective,” 2022, para. 1), despite its idealism and perhaps even impossible promise

(McDowell & Vetter, 2021; Vetter, 2020), remains a compelling idea that captivates the imagination of thousands of contributors. When those same volunteers become estranged from this “labor of love,” the narrative, the big idea encapsulated in Wikipedia’s enlightenment rhetoric, begins to fall apart.

Google also changed its tune over the years, dropping “don’t be evil” from the beginning of its code of ethics in 2018 (Conger, 2018). This quippy phrase was located in their code of ethics since 2000, reflecting an earlier techno-utopian and techno-optimistic orientation that was deeply embedded in the formation of many early Internet firms. Despite keeping “do the right thing,” their shift in language might also be an indicator of a shift in ethos, one that eschews earlier ethical orientations. Dropping (or at least moving away from) “don’t be evil” might have been a move to corporatize Google’s image and streamline the company to keep it in line with typical corporate language, but it seems to also represent a departure from the earlier “hacker” mentality that launched much of the early Web. There remain numerous concerns regarding techno-utopian ideals, but we believe there is ample opportunity to harken back to a (modified) version of their foundational ethos—particularly those associated with the “free culture” and “hacker” mentality of early Internet technology firms. Considering the ethical concerns with extracting the commons, perhaps it might behoove these firms (Google in particular) to return to some of these roots and honor the commons through participation rather than extraction. They already do this when it comes to software (Android itself is predominantly open source, Google Chrome’s binaries are licensed as freeware, and the majority of the Web runs on FOSS/FLOSS-based projects), so they are not unfamiliar with CBPP practices.

Focusing on participation in the commons, particularly participation that continues to hold up a more “share alike,” CBPP approach could be a win-win for big tech firms, because increasing transparency around participation and use helps to increase trust not only in the companies but also in the information, increasing transparency and making the “machines” less invisible. This helps to alleviate trust issues around information sourcing, in the firms providing that information, and increases trust within the communities that create the information used by those firms. Of course, this goes for all “big tech” firms, including the WMF.

It is understandable that the WMF wishes to get ahead of this extraction and get involved because some control is better than none (as well as finding ways to financially support the nonprofit that supports the community). We have tried to explore what is at stake and the power that Wikipedia (and now Wikidata) has in shaping the future of how the commons are utilized. More research is needed, and we would particularly call for more work on understanding the motivations of Wikidata contributors, but Wikipedia’s greatest asset remains the community and its relationship to its labor (that being Wikipedia itself and now Wikidata). This should underscore the importance of that community retaining this relationship and retaining more control over its community-produced knowledge and data. Transparency and openness remain imperative here because, much like Vincent and Hecht (2021) argued, the community “should have agency in how its content is used” (p. 12). As it stands, not only has the community lost control over the product of its labor, but also it faces alienation and estrangement, breaking the relationship with its product (Wikipedia) as well as the bonds that tie the community together. Considering the deep relationship between Google and the WMF, there seems to be an opportunity here to encourage stronger participation in the commons, which might in turn alleviate concerns for both parties.

However, considering that WMF's assets have risen from \$57,000 in 2004 to \$231 million (plus another \$100 million endowment) in 2021 (the year it launched Enterprise), it might be the case that WMF does not actually need Google's financial assistance at all (WMF, 2023). In this case, it might only be Google that requires WMF, Wikipedia, and Wikidata, which puts WMF in a prime spot to shape the way that Google accesses the commons. Although Enterprise might be their way of addressing this, it still seems problematic because the donated labor is sold without the original license. With this in mind, it might behoove WMF and Google to agree to a more CBPP-like way of addressing attribution and share alike licensing for future uses of the information and data utilized.

If the prime directive of Wikipedia remains to "Imagine a world in which every single person on the planet is given free access to the sum of all human knowledge" ("Prime Objective," 2022, para. 1), we must also recognize that with CC-BY-SA, the notion of "free" here must represent not only "freedom" but also the promise of continued freedom for the information. Instead of CC0, where big tech firms can extract and manipulate data into a new, controlled product or system, what might it look like if the commons persisted? The Wikipedia and Wikimedia communities not only can shape the future of what the commons currently are (how it can be used and what is included) but also might then shape the future of what the commons can be (reunification of products into the commons when they emerged from that commons-based labor). Instead of turning our backs on the "other way," the community here has an opportunity to double down and ensure the perpetuation of the ethics of the commons (against the "pseudo-commons") by requiring their donated labor to serve the common good, instead of the "invisible machines" of profiteering multinational corporations. On the other side, WMF and Google have a responsibility to the community, and ensuring transparency and reciprocity remains imperative for this relationship to continue.

In the end, the deepest concerns should not be about the product (the information and data) but instead the community of volunteers—a community of people—that built (and continue to build and maintain) the free knowledge movement. It is this community that characterizes Wikimedia and that is itself central to the project's survival. The foundation and connection of that community is a uniquely valuable model for labor relationship, which this "realienation" now threatens. This community that already dwindles in numbers and struggles to diversify and increase its ranks now faces even more challenges. As mentioned elsewhere, "Wikipedia's greatest potential is change" (McDowell & Vetter, 2021, p. 105), and, most important, that change emerges directly from the community.

References

- Announcing WIT: A Wikipedia-based image-text dataset. (n.d.). *Google AI blog*. Retrieved from <http://ai.googleblog.com/2021/09/announcing-wit-wikipedia-based-image.html>
- Bailey, L. (2022, October 8). *Internet archive files final reply brief in lawsuit defending controlled digital lending*. Internet Archive Blogs. Retrieved from <https://blog.archive.org/2022/10/08/internet-archive-files-final-reply-brief-in-lawsuit-defending-controlled-digital-lending/>

- Benkler, Y. (2002). Coase's penguin, or, Linux and the nature of the firm. *Yale Law Journal*, 112, 369–446. doi:10.2307/1562247
- Benkler, Y. (2020). From utopia to practice and back. In J. Reagle & J. Koerner (Eds.), *Wikipedia @ 20: Stories of an incomplete revolution* (pp. 43–54). Cambridge, MA: MIT Press.
- Benkler, Y., & Nissenbaum, H. (2006). Commons-based peer production and virtue. *Journal of Political Philosophy*, 14(4), 394–419. doi:10.1111/j.1467-9760.2006.00235.x
- boyd, d. (2017). Did media literacy backfire? *Journal of Applied Youth Studies*, 1(4), 83–89. Retrieved from <https://search.informit.org/doi/10.3316/informit.607936397466888>
- Brabham, D. C. (2013). *Crowdsourcing*. Cambridge, MA: MIT Press.
- Brown, A. R. (2011). Wikipedia as a data source for political scientists: Accuracy and completeness of coverage. *PS: Political Science and Politics*, 44(2), 339–343. Retrieved from <https://www.jstor.org/stable/41319920>
- Casebourne, I., Davies, C., Fernandes, M., & Norman, N. (2012). *Assessing the accuracy and quality of Wikipedia entries compared to popular online encyclopaedias: A comparative preliminary study across disciplines in English, Spanish and Arabic*. Retrieved from http://commons.wikimedia.org/wiki/File:EPIC_Oxford_report.pdf
- Cohen, N. (2018, April 6). Perspective: Conspiracy videos? Fake news? Enter Wikipedia, the 'good cop' of the Internet. *Washington Post*. Retrieved from https://www.washingtonpost.com/outlook/conspiracy-videos-fake-news-enter-wikipedia-the-good-cop-of-the-internet/2018/04/06/ad1f018a-3835-11e8-8fd2-49fe3c675a89_story.html
- Community Insights/2018 Report/Contributors—Meta. (2018). Retrieved from https://meta.wikimedia.org/wiki/Community_Insights/2018_Report/Contributors
- Conger, K. (2018, May 18). *Google removes "don't be evil" clause from its code of conduct*. Gizmodo. Retrieved from <https://gizmodo.com/google-removes-nearly-all-mentions-of-dont-be-evil-from-1826153393>
- Creative Commons. (2023a). *Attribution—ShareAlike 4.0 International—CC-BY-SA 4.0*. Retrieved from <https://creativecommons.org/licenses/by-sa/4.0/>
- Creative Commons. (2023b). *CC0 1.0 universal*. Retrieved from <https://creativecommons.org/publicdomain/zero/1.0/>
- Dholakia, U. M., Bagozzi, R. P., & Pearo, L. K. (2004). A social influence model of consumer participation in network- and small-group-based virtual communities. *International Journal of Research in Marketing*, 21(3), 241–263. doi:10.1016/j.ijresmar.2003.12.004

- Dickey, M. R. (2019, January 22). *Google.org donates \$2 million to Wikipedia's parent org*. TechCrunch. Retrieved from <https://techcrunch.com/2019/01/22/google-org-donates-2-million-to-wikipedias-parent-org/>
- Farič, N., & Potts, H. W. (2014). Motivations for contributing to health-related articles on Wikipedia: An interview study. *Journal of Medical Internet Research*, *16*(12), e260. doi:10.2196/jmir.3569
- Ford, H. (2020). Rise of the underdog. In J. Reagle & J. Koerner (Eds.), *Wikipedia @ 20: stories of an incomplete revolution* (pp. 43–54). Cambridge, MA: MIT Press.
- Ford, H. (2022). *Writing the revolution: Wikipedia and the survival of facts in the digital age*. Cambridge, MA: MIT Press.
- Ford, H., & Graham, M. (2016). Provenance, power and place: Linked data and opaque digital geographies. *Environment and Planning D: Society and Space*, *34*(6), 957–970. doi:10.1177/0263775816668857
- Fuchs, C. (2020). The ethics of the digital commons. *Journal of Media Ethics*, *35*(2), 112–126. doi:10.1080/23736992.2020.1736077
- Galloway, A. R. (2013). *The interface effect* (1st ed.). Malden, MA: Polity.
- Giles, J. (2005). Internet encyclopaedias go head to head. *Nature*, *438*(7070), 900–901. doi:10.1038/438900a
- Hill, K. (2012). Sue Gardner on “herding cats” and halting Wikipedia’s editor erosion. *Forbes*. Retrieved from <https://www.forbes.com/sites/kashmirhill/2012/08/22/sue-gardner-wikipedias-editor-erosion/?sh=125ff8bf3a4c>
- Hwang, T. J., Bourgeois, F. T., & Seeger, J. D. (2014). Drug safety in the digital age. *New England Journal of Medicine*, *370*(26), 2460–2462. doi:10.1056/NEJMp1401767
- Iliadis, A. (2023). *Semantic media: Mapping meaning on the Internet*. New York, NY: Polity Press.
- Introduction. (2022). *Wikidata*. Retrieved from <https://www.wikidata.org/wiki/Wikidata:Introduction>
- Jemielniak, D., & Przegalińska, A. (2020). *Collaborative society*. Cambridge, MA: MIT Press.
- Johnson, B., & Francisco, S. (2010, February 18). Wikipedia wins the Google lottery – but why? *The Guardian*. Retrieved from <https://www.theguardian.com/technology/blog/2010/feb/18/wikipedia-google>
- Kolbe, A. (2015). Whither wikidata? *The Signpost*. Retrieved from https://en.wikipedia.org/wiki/Wikipedia:Wikipedia_Signpost/2015-12-02/Op-ed

- Ford, H. (2014, September 8). Max Klein on Wikidata, "botpedia" and gender classification. *Hblog.Org*. Retrieved from <https://hblog.org/2014/09/08/max-klein-on-wikidata-botpedia-and-gender-classification/>
- Kuznetsov, S. (2006). Motivations of contributors to Wikipedia. *ACM SIGCAS Computers and Society*, 36(2), 1–10. doi:10.1145/1215942.1215943
- Li, H., Ajmani, L., Zhou, M., Vincent, N., Hwang, S., Piccardi, T., . . . Veselovsky, V. (2022). Ethical tensions, norms, and directions in the extraction of online volunteer work. *Companion Computer Supported Cooperative Work and Social Computing*, November, 273–277. doi:10.1145/3500868.3560923
- Litman-Navarro, K. (2019, January 23). *Google will conquer the world one charitable donation at a time*. The Outline. Retrieved from <https://theoutline.com/post/6999/google-will-conquer-the-world-one-charitable-donation-at-a-time>
- Marx, K. (1993). *Grundrisse: Foundations of the critique of political economy (rough draft)*. London, UK: Penguin Classics.
- Marx, K. (2012). *Economic and philosophic manuscripts of 1844*. Mineola, NY: Dover.
- Marx, K., & Engels, F. (1978). *The Marx-Engels reader* (R. C. Tucker, Ed.). New York, NY: W. W. Norton & Company.
- McDowell, Z. J., & Vetter, M. A. (2021). *Wikipedia and the representation of reality*. New York, NY: Routledge.
- McDowell, Z. J., & Vetter, M. A. (2022). Fast truths and slow knowledge: Oracular answers and Wikipedia's epistemology. *Fast Capitalism*, 19(1), 104–112. doi:10.32855/fcapital.202201.009
- McMahon, C., Johnson, I., & Hecht, B. (2017). The substantial interdependence of Wikipedia and Google: A case study on the relationship between peer production communities and information technologies. *Proceedings of the International AAAI Conference on Web and Social Media*, 11(1), 142–151. doi:10.1609/icwsm.v11i1.14883
- Oreg, S., & Nov, O. (2008). Exploring motivations for contributing to open source initiatives: The roles of contribution context and personal values. *Computers in Human Behavior*, 24(5), 2055–2073. doi:10.1016/j.chb.2007.09.007
- Prime Objective. (2022). *Wikipedia*. Retrieved from https://en.wikipedia.org/w/index.php?title=Wikipedia:Prime_objective&oldid=1070972014

- Roth, E. (2022, June 22). *Google is paying the Wikimedia Foundation for better access to information*. The Verge. Retrieved from <https://www.theverge.com/2022/6/22/23178245/google-paying-wikimedia-foundation-information>
- Sartre, J. P. (1984). *Critique of dialectical reason, Vol. 1: Theory of practical ensembles* (1st ed.). New York, NY: Verso.
- Shirky, C. (2009). *Here comes everybody: The power of organizing without organizations*. London, UK: Penguin Books.
- Simonite, T. (2013). *The decline of Wikipedia*. MIT Technology Review. Retrieved from <https://www.technologyreview.com/2013/10/22/175674/the-decline-of-wikipedia/>
- Statistics. (2023). *Wikidata*. Retrieved from <https://www.wikidata.org/w/index.php?title=Wikidata:Statistics&oldid=1876166148>
- Tapscott, D., & Williams, A. D. (2008). *Wikinomics: How mass collaboration changes everything* (Expanded ed.). Alberta, Canada: Portfolio.
- Terdiman, D. (2012). *Wikidata to provide structured data for all Wikipedia versions*. CNET. Retrieved from <https://www.cnet.com/culture/wikidata-to-provide-structured-data-for-all-wikipedia-versions/>
- Vetter, M. A. (2020). Possible enlightenments: Wikipedia's encyclopedic promise and epistemological failure. In J. Reagle & J. Koerner (Eds.), *Wikipedia @ 20* (pp. 285–296). Cambridge, MA: MIT Press. doi:10.7551/mitpress/12366.003.0024
- Vincent, N., & Hecht, B. (2021). A deeper investigation of the importance of Wikipedia links to search engine results. *Proceedings of the ACM on Human-Computer Interaction, 5*(CSCW1), 1–15. doi:10.1145/3449078
- Vincent, N., Johnson, I., & Hecht, B. (2018). Examining Wikipedia with a broader lens: Quantifying the value of Wikipedia's relationships with other large-scale online communities. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, April*, 1–13. doi:10.1145/3173574.3174140
- Wikimedia. (2022). Retrieved from <https://www.wikimedia.org/>
- Wikimedia Enterprise—API for Search Engines & Social Networks. (2023). *Wikimedia Enterprise*. Retrieved from <https://enterprise.wikimedia.com/>
- Wikimedia Foundation. (2023). *Wikipedia*. Retrieved from https://en.wikipedia.org/w/index.php?title=Wikimedia_Foundation&oldid=1150408631
- Wikipedians. (2022). *Wikipedia*. Retrieved from <https://en.wikipedia.org/w/index.php?title=Wikipedia:Wikipedians&oldid=1095561306>

Xu, B., & Li, D. (2015). An empirical study of the motivations for content contribution and community participation in Wikipedia. *Information & Management, 52*(3), 275–286.
doi:10.1016/j.im.2014.12.003

Zhang, C. C., Houtti, M., Smith, C. E., Kong, R., & Terveen, L. (2022). Working for the invisible machines or pumping information into an empty void? An exploration of Wikidata contributors' motivations. *Proceedings of the ACM on Human-Computer Interaction, 6*(CSCW1), 1–21.
doi:10.1145/3512982