This is a repository copy of *The search for Wikipedia’s edges*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/95526/

Version: Accepted Version

**Book Section:**

(c) 2017 - Routledge. This is an Accepted Manuscript of a book chapter published by Routledge in *The Routledge Companion to Digital Ethnography*. on 15th Dec 2016 available online: http://www.routledge.com/9781138940918

**Reuse**
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

**Takedown**
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.
The search for Wikipedia's edges
Heather Ford

Abstract:

Wikipedia has become an authoritative source for facts about the world. It promises to be working towards reflecting “the sum of all human knowledge” in 288 languages. Every day, however, thousands of contributions to Wikipedia are rejected. A large proportion of those whose contributions are rejected remain silent on Wikipedia’s platform. In this chapter, I describe my attempt to give voice to some of those silenced. In 2011, I met and interviewed a small group of Wikipedians in Nairobi in order to investigate the apparent success of the Kenyan Wikipedia community. There were many surprises as people spoke about their hopes and disappointments regarding the community and the platform into which they had recently been inducted. This experienced highlighted what I believe to be a vastly understudied area. Instead of more ethnographic work that considers only those who are highly active and visible on Wikipedia, there is a need to also study the lacunae, the people who remain on the edges of Wikipedia because of a feeling of disappointment, rejection or invisibility. I highlight the gaps in ethnographic research on Wikipedia to date and suggest the principle of multiplicity within ethnography as a way of articulating a new research agenda that focuses on Wikipedia as a significant global, political space.

Fanfare around the Internet’s potential as a platform for speech by all still echoes around us. Heralded as the site in which geographic, social and political barriers were being overturned, early rhetoric around the impact of networked technologies on the media and education posited the rise of new forms of communitarian platforms. Such rhetoric positioned platforms as the means by which users were able to take back the means of production and publish information goods without the need for media power brokers and intermediaries of the past. The 2006 Time Magazine Person of the Year Award perhaps best illustrates this zeitgeist. In December 2006, Time Magazine

---

1 Heather Ford is a Fellow in Digital Methods at the School of Media and Communications at Leeds University and has recently completed her DPhil in Information, Communication and the Social Sciences at the Oxford Internet Institute (University of Oxford) where she studied Wikipedia as a “Fact Factory” that is reconfiguring the ways in which claims become facts in the network society. An ethnographer and former Internet rights activist, Heather studies the construction of knowledge and the performance of authority in online, collaborative systems and the realignments of power in new media ecosystems using digital methods.
declared that its person of the year was “you”. The cover of the magazine featured a computer screen with the words, “You. Yes, you. You control the Information Age. Welcome to your world” (Grossman 2006). The editorial argued that ordinary people now controlled the means of producing information and media because they dissolved the power of the gatekeepers who had previously controlled the public’s access to information.

[2006 is] a story about community and collaboration on a scale never seen before. It’s about the cosmic compendium of knowledge Wikipedia and the million-channel people’s network YouTube and the online metropolis MySpace. It’s about the many wresting power from the few and helping one another for nothing and how that will not only change the world, but also change the way the world changes. (Grossman, 2006)

The idea that the “many” were “wresting power from the few” was shared by a host of commentators and scholars at the time. This colourful metaphor was inspired by the evidence that many more people were now doing work that had previously been done by credentialed individuals within large institutions. Now, non-academics were writing encyclopaedia articles, laypeople producing films, and concerned citizens producing news articles. With the decrease in the costs of the means of production and the connection of millions of ordinary people to a network of potential audience members, co-producers, employees and publishers, this logic extended to the idea that anyone could be a journalist, an engineer or a scientist (Gillmor, 2008; Leadbeater & Miller, 2004; Shirky, 2009); now anyone could be an expert in something (Weinberger, 2011). We had moved
from a “read only” culture to a “read write” culture (Lessig, 2009) that was characterised by active cultures of participation (Jenkins, 2006) instead of passive consumption of culture that was presented to the public ready-made.

Cracks have, however, begun to appear in this revolutionary picture in which the user reigns triumphant. In this chapter, I outline the current debates and research relevant to this more complicated picture in the area of Wikipedia studies and argue that what is missing from Wikipedia research is a deeper analysis of what is missing from Wikipedia. I argue that the ethnographic tradition is vital for studying Wikipedia’s edges, its silences, its lacunae and that instead of more ethnographic work that considers only those who are highly active and visible on Wikipedia, there is a need to also study those who remain on the edges of Wikipedia because of disappointment, rejection or invisibility. By highlighting Wikipedia’s widening multiplicity, ethnography has the potential to focus attention on the platform as a significant global, political arena in which there are emerging power configurations and new landscapes of inequality. Such research is necessary in order to continue struggling towards greater global democratic decision-making within the world’s critical representative platforms.

**Wikipedia as a Positioned Platform**

The dominant position of platforms including Facebook, Google and YouTube in the Western world have led to some to question the opportunities afforded by Web 2.0 sites to enable democratic debate. Commentators have highlighted the role of such platforms in the surveillance of users – either in the interests of capitalism (Lovink 2012) or autocratic regimes (Morozov 2011). Wikipedia has,
however, remained largely unsullied in such debates. Christian Fuchs (2013) in his critique of social media, for example, holds up Wikipedia as an example of info-communism as opposed to the likes of capitalist Facebook and Google. A closer analysis of the politics of Wikipedia, however, starts to complicate the notion of Wikipedia as the model for global democratic production. Although Wikipedia’s structure and values may be the best opportunity that we have for envisioning a global democratic online public sphere at the present moment, it is important to understand that it, too, suffers from tensions relating to its position as the dominant reference work on the Internet.

Wikipedia promises to be working towards reflecting “the sum of all human knowledge” (Wikimedia Foundation 2015) in 288 languages. Every day, however, thousands of contributions to Wikipedia are rejected. In a study that I conducted with Stuart Geiger (2011) of Wikipedia articles for deletion, for example, we learned that the majority of articles in English Wikipedia are deleted under the category of ‘no indication of importance’. Understanding the power dynamics behind such decision-making is a critical task for scholars of knowledge and the media. Wikipedia has become the site of significant power to represent. This power stems from its identity not as a social media site in which thousands of editors participate in the construction of articles, but as an authoritative source for facts about the world.

Understanding who (or what) governs Wikipedia and which users and uses such configurations empower is important for two key reasons. Firstly, Wikipedia is often used as a model for the participatory development of public information
goods. Wikipedia has become an important model for answering questions about whether the Internet can and should be regulated, who should ideally regulate it, as well as how regulation should occur.

Secondly, Wikipedia has become an authoritative source for facts about the world, used not only by the millions of individual users who visit the site everyday, but by other platforms (especially search engines such as Google, Yahoo! and Bing) as a source of public information displayed as fact to the user. With the advent of Google’s Knowledge Graph, a search for “What is the capital of Israel?” on Google, for example, will result not only in a list of possible sites that may provide answers to our query, but also a fact box presented as the answer. Wikipedia is one of the core sources from which search engines are drawing data for these answers. Understanding Wikipedia as a site of power and politics is therefore critical to analyzing the ways in which the power to represent is being reconfigured in today’s mediated environment (see, for example, Ford and Graham, 2015).

**Researching Wikipedia**

Wikipedia has been the subject of a rich and varied array of research over its fifteen-year history. Because the site makes data about every edit available to the public, the availability of openly accessible data has meant that Wikipedia research has been largely dominated by quantitative studies in computer science and information systems (Okoli et al. 2012). Some of the leading views on Wikipedia are, however, derived from qualitative or mixed methods research
approaches. Studies in this arena benefit from the rich open data that Wikipedia provides about editing behavior, but complement such research with interviews and participant observation.

As the rhetoric about the revolutionary potential of the Internet and peer production in particular to enable particular freedoms and values has waxed and waned, so too have responses to the stock of evidence that has been built up around how platforms like Wikipedia are both unique but also extend historical trends. Taken as a whole, Wikipedia research can be summarized as attempting to answer three core questions over time: a) Does Wikipedia work? b) How does it work? c) For whom does it work?

a) Does Wikipedia work?

Wikipedia reached its growth peak in early 2007 (Suh et al. 2009). In six years, Wikipedia’s editor numbers had reached over 300,000 and English Wikipedia moved over the 1 million-article milestone. In this first phase of Wikipedia’s development the key question driving research and commentary was whether Wikipedia was as good as its commercial counterparts. A breakthrough study by Nature in 2005 started off this slew of research, comparing the accuracy of Wikipedia with that of Encyclopedia Britannica and finding that Britannica contained only marginally fewer errors of fact (Giles 2005). Since the Nature study, others have investigated Wikipedia’s quality according to topics that include famous Americans (Rosenzweig 2006) and Western philosophers (Bragues 2007).
Theoretically, Yochai Benkler’s *Wealth of Networks* (2006) made an important contribution to debates about the validity of Wikipedia’s production model. According to Benkler, there have been significant changes in the organization of information production catalysed by the Internet that have resulted in the rise of nonmarket and non-proprietary production where individuals are able to take a more active role than was previously possible in the production of information goods. Benkler argued that the rise of individual agency leads to an inevitable clash with previous hierarchical market-driven industries which will inevitably decide the fate of each of these models. Wikipedia is used as a key example of nonmarket, non-proprietary and non-hierarchical peer production where incentives to create cultural goods are based not on price signals but by pro-social goals. According to Benkler, Wikipedia works because it is a natural outcome of a network in which the costs of connecting and producing information products together has seen a radical decline.

**b) How does Wikipedia work?**

The question of Wikipedia’s comparative quality was followed by empirical research into *how* the site works, with a number of research projects laying the foundations for understanding the social and processional dynamics behind Wikipedia’s everyday functioning. One of the most popular research questions in this topic area concerns the motivations for contribution to Wikipedia. Why, the question was asked, would anyone write for Wikipedia for free?
In response to this question, Forte and Bruckman (2005) found that Wikipedia resembles the scientific community in their cycles of credit and that Wikipedia editors are motivated to 'collaboratively identify and publish true facts about the world' (Forte and Bruckman 2005, 1). Despite the potential for anarchy, researchers have found that there is a strong focus on group coordination mechanisms within the architecture and norms of Wikipedia that enable users to rapidly respond to vandalism (Viegas et al. 2007). Wikipedia policies, also developed by community members, are considered an important element of social interaction on Wikipedia because they provide a common resource for new users to learn about editing and behavioural conventions (Viegas et al., 2007). One such policy or principle, that of “assum(ing) good faith” is the subject of Joseph Reagle's (2010) ethnographic study of Wikipedia. Reagle considers the history and rhetoric employed by the Wikipedia community on mailing lists, talk pages and a host of other forums in order to argue that the principle of good faith shaped the success of Wikipedia in its early years.

Another strand to research around questions of how Wikipedia works is to pay attention to the significant role of automated software agents in the everyday functioning of Wikipedia. Geiger and Ribes (2010), for example, demonstrate the role of non-human actors in the process of vandal banning, arguing that the decentralized activity enabled by automated and semi-automated tools on Wikipedia is a type of ‘distributed cognition’ (2010, 117). Similarly, Niederer and van Dijck (2010) investigate the increasingly important role of bots in the rise of Wikipedia, arguing that it is impossible to understand Wikipedia’s response to vandalism without an appreciation of the encyclopaedia as a sociotechnical
system driven by collaboration between users and bots. The authors provide a schematic overview of Wikipedia users according to their technical permission levels, and note that the permission level of bots is below that of administrators but well above the authority of registered users (2010, 1373). Wikipedia’s content management system, argue Niederer and van Dijck, allows for “protocological control, a mode of control that is at once social and technological – one cannot exist without the other” (2010, 1373).

In my own research, I have discovered that addiction to responses from other editors as well as other negative social and psychological influences can catalyse increased participation (Ford 2013). Following an editor who began his work with Wikipedia as a model editor and ended it by being banned from English Wikipedia enabled me to discover the dark side of Wikipedia editing and the driving forces behind it. More work is needed, however, in order to fully understand motivations (both positive and negative, automated and human) for contributing to Wikipedia if we are to gain a fuller perspective of how Wikipedia works.

c) For whom does Wikipedia work?

In response to the early democratic rhetoric espoused by researchers and pundits of Wikipedia and the Internet in general, a number of researchers have started to expose inequalities in Wikipedia’s representation of the world and to ask questions about the source of these inequalities. Such questions are an inevitable step from the recognition that, whether we like it or not, Wikipedia is
an important feature of the global information environment for many countries around the world and, as such, Wikipedia needs now to be seen not as an anomaly but as a feature of our public information infrastructure.

A number of valuable studies into Wikipedia’s representational scope and coverage have highlighted significant sources of inequality on the platform. The first group of studies investigates Wikipedia’s representation along the lines of social stratification including gender and geography. Demographic studies in this area indicate that between 84 and 90% of Wikipedia editors are men (Glott and Ghosh 2010; Hill and Shaw 2013; Lam et al. 2011), the majority have tertiary education and a significant number of editors across language versions speak English (Wikimedia Foundation, 2011). Lam et al. (2011) find that the low proportion of females participating in English Wikipedia have resulted in measurable imbalances relating to content quality so that articles relevant to women are significantly shorter and have lower assessment ratings than those interesting to men (Lam et al., 2011, p. 6).

Similarly, mapping geotagged articles in English Wikipedia, Mark Graham finds that “almost all of Africa is poorly represented in Wikipedia” (Graham 2011, 275) thus mirroring existing inequalities online. Analysing accounts of Singaporean and Philippine history on Wikipedia, Luyt (2011) argues that, despite the potential of new media for making visible previously marginalized voices, a more likely outcome is a reproduction of the status quo in historical representation. Graham et al. (2014) argue that it isn’t only issues of connectivity that prevent people in developing countries from contributing to Wikipedia and that
representation is a vicious cycle for those with strong editing cultures in local languages, while those on the peripheries of these countries fail to reach critical mass.

Other studies have indicated that Wikipedia's representations do not only mirror existing asymmetries but can actually exacerbate those asymmetries. Reagle and Rhue (2011) find that, although Wikipedia biographies on women are longer and more numerous than Encyclopaedia Britannica in absolute terms, “Wikipedia articles on women are more likely to be missing than are articles on men relative to Britannica” (p. 1138). Reagle (2013) extended these findings with a qualitative study in which he argues that low female participation in free culture communities, particularly within Wikipedia, is the product of a culture that is alienating towards women and that gender disparities are even worse in Wikipedia than in the computing culture from which it arose. Reasons for this include that geek stereotypes can be alienating, open communities are especially susceptible to difficult people, and the ideas of freedom and openness can be used to dismiss concerns and rationalize the gender gap as a matter of preference and choice.

Although the aforementioned studies are useful in highlighting Wikipedia as a source of inequality in opposition to the rhetoric being espoused, they tend not to fully consider the systematic sources of bias inherent in Wikipedia's processes, policies and technologies. Wikipedia's environment is characterised by a highly software mediated and rife with hundreds of policies that stand in for editorial hierarchy. Instead of a small group of editors making site-wide
decisions on what to include or exclude, there is instead an array of policies, templates, guidelines and tools that aim towards standardisation. The growing stratification of users along the lines of their experience deploying such objects has begun.

A handful of researchers have started to investigate the peculiarities of the Wikipedia system and how it is affecting power configurations on the site. Wikipedia policy is an important feature in defining the rules by which editors “delimit the sayable, define legitimate perspectives and fix the norms for the elaboration of concepts” (Pentzold and Seidenglanz 2006, 65). Although policy provides a common language for action, their ambiguity leads to power play among participants who may interpret policies differently according to their editorial interests (Kriplean et al. 2007).

Apart from complex policy debates applied to editorial decision-making, effective Wikipedia editors also demonstrate a sophisticated “trace literacy” (Ford and Geiger 2012) that results in significant power imbalances among users. The ability to both read and write according to the rules of Wikipedia's complex site-specific vocabulary has a significant impact on who is able to have their edits sustained. Along similar lines, Hargittai and Shaw (2015) note that inequalities in representation of women on Wikipedia are exacerbated by a similarly significant Internet skills gap. They find that the most likely contributors are “high-skilled males” and that, actually, among low-skilled Internet users no gender gap exists.
Such research is beginning to shed light on questions relating not only to how peer production platforms such as Wikipedia may serve to exacerbate certain inequalities, but also how Wikipedia is resulting in the reconfiguration of the power to represent knowledge. Those who are able to master Wikipedia’s technocratic system of representation with an emphasis on facts and other modular pieces of verifiable information are consequently able to emerge as power brokers within this environment. Those who fail to master this system, either because their knowledge of the world does not fit with what Wikipedia recognises as knowledge (see, for example, Gallert and Van der Velden 2013) or because negative social interactions on the platform have led to their leaving the platform (Halfaker et al. 2013), mean that many will remain on Wikipedia’s edges, unable to contribute and have their knowledge represented.

**Situating ethnography**

Although the aforementioned studies are a useful starting point for future research into Wikipedia’s shadows, there is a significant gap in research that investigates Wikipedia’s edges according to the perspectives of those whose knowledge remains unrepresented. Ethnography can play a significant role in highlighting such knowledges, but ethnographers need to find innovative ways of seeking out actors who are largely invisible to Wikipedia’s system – actors whose traces are feint or isolated if they exist in digital form at all.

There have been at least three previous ethnographic studies of Wikipedia including *Good Faith Collaboration: The Culture of Wikipedia* by Joseph Reagle.
(2010), *Janitors of knowledge: Constructing knowledge in the everyday life of Wikipedia editors* by Olof Sundin (2011) and *Common Knowledge? An Ethnography of Wikipedia* by Dariusz Jemielniak (2014). Each of these ethnographies provides a rich source of understanding about Wikipedia laws, norms and dynamics but all of them are filtered through the perspective of Wikipedia's policy makers and administrators – that is, through the lens of those who already have significant power and experience on Wikipedia.

Instead of more ethnographic work that considers only those who are highly active and visible on Wikipedia, there is a need to study the lacunae, the people who remain on the edges of Wikipedia because of disappointment, rejection or invisibility. What is missing from Wikipedia research are ethnographies of the dispossessed, the invisible, the subalterns. Ethnography is a useful tool in starting to shed light on the edges of Wikipedia’s network for three core reasons. These include ethnography’s focus on multiplicity that suits the study of diverse and multifarious communities of practice within and outside of Wikipedia’s domain, the flexibility in ethnographic methods that enables the straddling of both the quantitative and qualitative traditions, as well as the ethnographic tradition of highlighting those in the shadows of dominant systems of power.

**a) Multiplicity**

Wikipedia started as a small, well-defined community of individuals working towards a relatively simple set of goals, but as it has grown in numbers and in authority, Wikipedia’s networks of individuals, technologies and policies have
grown in numerous diverse directions. Wikipedia is now constituted by multiple, striated communities and identities. In addition to discussions about Wikipedia articles taking place on talk pages, Wikimedia mailing lists and community pages, Wikipedia articles are the subject of discussion on social media sites, in the mainstream press and in sites like Wikipediocracy that criticize Wikipedia’s governance structures, its community dynamics and what its authors believe are the hidden power dynamics behind its functioning.

According to anthropologist, George Marcus (1995, 25) “any cultural identity or activity is constructed by multiple agents in varying contexts, or places, and that ethnography must be strategically conceived to represent this sort of multiplicity, and to specify both intended and unintended consequences in the network of complex connections within a system of places”. Wikipedia’s complexity, the fact that it means many different things to many different people in multiple assemblages of humans and technological artefacts, makes it a particularly ripe field of study for ethnography.

b) Flexibility

Ethnographic research, with its focus on principles rather than specific methods, enables the researcher to employ a flexible suite of methods in order to study people and the meanings they ascribe to phenomena. Using the technique of “following the thing” suggested by Marcus (1995), Jenna Burrell (2009) suggests that we construct a network of human actors, non-human actants or objects, as
well as the processes and practices that become enrolled as we follow the thing across numerous places, platforms or phenomena.

Applied to Wikipedia, such flexible approaches enable ethnographers to collect diverse sources of data, including trace data about who (or what, in the case of automated bot editors) has edited articles over which time periods, rhetoric employed by editors in arguing for particular editorial or policy decisions, as well as to follow representations of phenomena to the subjects of such representation through interviews and participant observation. Mixed methods that employ the benefits of openly available trace data, but do not stop at such data, are useful to the ethnographer in order to fill in the gaps that remain invisible to the system. Such methods enable researchers to overcome challenges in being held to the study of a single platform when users and representations move dynamically between those platforms and, indeed, outside of those platforms in ways that the digital does not reach.

c) Shadows

There is a rich tradition in ethnographic research of engaging fully with the morality and ethics of particular systems. According to Nancy Schepier-Hughes (1992), ethnography calls on us to reveal our sympathies rather than trying to disguise them. By offering a lens for filtering the experiences of the dispossessed, ethnographers are in an important position to call for change to make systems more equitable in the light of their growing authority and power to represent.
In addition, participant observation enables the ethnographer to experience the position of the uninitiated, the newcomer, the powerless within the network. When one begins editing on Wikipedia, for example, one’s edit count is low and so is one’s authority within the network. I used such moments as opportunities to reflect on the challenges and problems faced by new editors in my own ethnographic study of Wikipedia (Ford, 2015). There are individuals and communities who will always remain on the outside of the Wikipedia network. The ethnographic experience of the new editor puts researchers into a good position for documenting these narratives if ethnographers are sufficiently reflexive.

**Conclusion**

Although the Internet’s liberatory potential is no longer a feature of commentary and scholarship in the same way that it was a decade ago, the excitement that has accompanied the vast availability of digital trace data has meant that those who have fallen on the dark side of the digital remain largely unheard. George Marcus writes that “the heart of contemporary ethnographic analysis is not in the reclamation of some previous cultural state or its subtle preservation despite changes, but rather in the new cultural forms to which changes in colonial subaltern situations have given rise” (1998, 79). Wikipedia is an example of a new cultural form that has arisen out of the global spread of the Internet and the dominance of Western epistemology and the logic of verifiable facticity. Understanding the political dynamics of those who remain outside of this logic is a critical job for ethnographers.
Rather than another final word on Wikipedia, ethnographic studies should seek to explore the edges of the Wikipedia network, to provide a myriad of alternative perspectives that shed light on the multiplicity of Wikipedia and its position as critical global representative infrastructure. Ethnographic studies should aim at achieving yet an alternative perspective, another frame with which to view what is produced out of Wikipedia’s socio-technical structures and what its place in our world is. In order to construct this frame, researchers must seek out the actors and phenomena that are in the shadow of the encyclopedia: the banned, the deleted, the disillusioned. In doing so, Wikipedia research will begin to take on the shape and voices of those who have a critical stake in its representations.

Bibliography


Kriplean, Travis, Ivan Beschastnikh, David W. McDonald, and Scott A. Golder. 2007. “Community, Consensus, Coercion, Control: Cs*W or How Policy


---

i Encyclopedia Britannica criticized the report’s methodology and refuted its conclusions but Nature refused to retract it (Nature 2006).