Ethnographies of the Digitally Dispossessed
Heather Ford

Abstract: Defining a fieldsite is one of the most challenging aspects of media anthropology work in the context of global, digital platforms. In this chapter, I argue that, in addition to the ethnographic studies of platforms' pro-users, its dominant actors, its most popular or active groups, we need ethnographies of the digitally dispossessed: those who are routinely ignored, omitted, denigrated and denounced by platforms. I do so in the context of an internet that is increasingly subject to the logic of automation and AI, and where voice is enabled through one's mastering of data centric representational forms. Using the case of English Wikipedia and an article that faced significant challenges in its first year of development, I propose three strategies for exploring digital knowledge platforms from the perspective of the digital subaltern including: entering via flagged erasures, following sources and their characterization and analysing the networks that traverse articles. I explain how I applied these strategies in the context of my own fieldwork, demonstrating the ways in which digital and datafied knowledge representations increasingly gain purchase or are stopped in their tracks via digital knowledge infrastructures.

In 2011 the "Oral Citations Project" was funded by the Wikimedia Foundation, the non-profit organisation based in the United States that hosts Wikipedia. The project was first documented on "Meta", a site used for coordinating and planning Wikimedia projects. The project was introduced on the site as follows:

> Imagine a world in which every single person on the planet is given free access to the sum of all human knowledge. To many within the Wikimedia movement, this idea is the guiding ambition that drives us. The problem with the sum of human knowledge, however, is that it is far greater than the sum of printed knowledge.

The Oral Citations Project was a response to Wikipedia's dearth of published material about topics relevant to communities in places outside North America and Western Europe. To the project director, Wikimedia Foundation Advisory Board member, Achal Prabhala, the source of this problem was Wikipedia's policy that defines "reliable sources" as published texts.

According to Prabhala, Wikipedia policies suggest that only printed knowledge can be used as a basis of articles. But books and printed material are a luxury of only the "rich economies (of) Europe, North America, and a small section of Asia." In the introduction to the project, he noted that there was very little scholarly publishing in languages other than English in India and that most South African languages other than English and Afrikaans have had a "primarily oral existence".

Because of this disparity, the knowledge of some communities is privileged on Wikipedia, while for others, it remains hidden. The lack of written, printed and published material in countries and languages outside the Global North was not only problematic for Wikipedia editions written in languages other than English and other major European languages. It also resulted in a less rich Wikipedia, one that couldn't possibly fulfil its goal to represent "the

sum of all human knowledge" (Wikimedia Foundation, n.d.).

> As a result of this disparity, everyday, common knowledge - things that are known, observed and performed by millions of people - cannot enter Wikipedia as units of fact because they haven't been written down in a reliably published source. This means that not only do small-language Wikipedias in countries like India and South Africa lose out on opportunities for growth, so also does the Wikimedia movement as a whole lose out on the potential expansion of scope in every language.

Prabhala suggested that a new type of source, an "oral citation", be employed in cases where no published information about a topic exists on Wikipedia. The Oral Citations Project team visited communities where oral tradition had facilitated the transfer of knowledge about local customs and culture. Wikipedia editors involved in the project conducted in-person interviews with community members in the rural village of Ga-Sebotlane in Limpopo province, South Africa and over the phone with interviewees from Kannur, a city in North Kerala, India. They uploaded recordings and transcripts of those interviews onto Wikimedia Commons, a sister site to Wikipedia where multimedia files are stored. They then wrote Wikipedia articles using the oral citations as references.

In response to the project, many other Wikipedia editors attempted to discredit, delete and vandalise the articles, while debates about the validity of oral citations for the encyclopedia raged on mailing lists and working groups. One of the articles based on oral citations was about "surr", a game that was once popular among children living in the villages of northern India. Surr is a game played by two teams of four players each. A rectangular playing field is divided into four equal quadrants. One team gathers in the first quadrant, while the other team gathers along the lines of defence at the borders of the adjoining quadrant. The objective of the game is for a team to enter the other three quadrants without being touched by a player from the opposing team. If all members of a single team survive and reach the final quadrant, then they will win the game. Once all the surviving players gather in a new quadrant, they shout, "Bol Den Goivan Surr!"

Surr is no longer commonly played in India. Some say that it is because children prefer video games or that they now live in high-rise apartments and aren't allowed to run and play outside by their increasingly vigilant parents. Despite this, surr is a phenomenon etched in the memories of those who once played it. It animates the memories of their childhood. It reminds them of a culture unique to where they come from. A man born in the neighbourhood of Ayodhya in Uttar Pradesh, who played the game in his childhood, talked about how surr was played by *Bahujan* people. *Bahujan* is the Hindi word for Dalits and "Other Backward Classes" (OBC), a collective term used by the Government of India to classify castes which are educationally or socially disadvantaged.

> "(G)ames like Surr have a strong connection with soil and agriculture. Surr used to be played at the barren or emptied fields. And in most cases, the people who used to inhabit near to these fields were those, who work in the fields, minimum waged workers. Their kids used to play Surr. In rare cases, when the kids from metropolitan cities, like Mumbai, visited the village and happened to see kids playing Surr might join them." (Interview, 17 June, 2021)

On Wikipedia, the article explained where the game was played and its rules. Statements were referenced to transcribed audio interviews with two Indian public servants about their

experience of the game. But not everyone agreed with the validity of the knowledge represented there. Many Wikipedians opposed the use of the interviews as sources in Wikimedia mailing lists and working groups dedicated to Wikipedia policy discussion. The hope that at least on Wikipedia, a space would be offered to share the knowledge of those already suffering from significant hardship born by class and caste, was dashed.

One of those who weighed in on oral citations via the Wikimedia listserv was the the long-time Wikimedia volunteer, Ziko van Dijk, who has been a Wikimedian since 2003, has a doctorate in history and has played a significant role in Wikimedia in the Netherlands and Germany. He argued that only academic experts ("historian, ethnologist etc.") could legitimately record knowledge held by "illiterates".

> There are good reasons for this way. One is, that it is not very practical to cite from audiotapes/audiofiles. Another, that what this individual is describing may be true for his personal environment but cannot be generalized to others. For that, one needs the scholar. Remember: witnesses are the most unreliable source ever. People tell you plain nonsense - not because they want to (lie) or are stupid but because the human brain is simply not created to be a historian. It has the greatest difficulties to store information truthfully. So you need to record, and compare the different assertions from different people.
>
> It is a possibility to record oral and visual expressions from illiterates, and only later to do something with it scholarly. But all this has nothing to do with Wikipedia. (Ziko van Dijk, Wikimedia-l mailing list, 25 February 2012)

Former Wikimedia Foundation Board Chairperson, Ting Chen responded that van Dijk's was the traditional view of how encyclopaedia should be produced, but that Wikipedia could produce the encyclopaedia differently; Wikipedia, *was* in fact, already doing things differently:

> Yes, it is the way (that) classic encyclopedia(s) worked. But Wikipedia is not a classic encyclopedia, and I don't see the sense to bound ourselves… just to please some old traditional rules… Scholars have limited capacities… (s)cholars cannot pay attention to everything. (If we) give everyone the possibility to pay attention to what they think is interesting and important in their life, we can free a lot of potential… (Ting Chen, Wikimedia-l mailing list, 25 February 2012)

The story of surr is difficult to reconcile with the dominant paradigm of Wikipedia. By most accounts, Wikipedia is a platform that offers the only viable, virtuous alternative to Big Tech. Wikipedia is among the fifteen most popular websites in the world, used daily by nearly 500 million people. Featuring more than 40 million articles, it works through what Yochai Benkler (2006) calls 'commons based peer production', in which large numbers of people work cooperatively to produce collective, public goods.

Wikipedia is open to contributions by anyone who has an internet connection and its content can be freely used, reused and modified for commercial and non-commercial purposes. It is operated by the Wikimedia Foundation, a non-profit organisation that is headquartered in San Francisco, with chapters in 39 countries. Early in its development, Wikipedia's co-founder, Jimmy Wales pronounced that the site's goal was to give freely "the sum of all human knowledge". The project's logo is a globe represented as a puzzle that is yet to be completed.

This dream of global accord in a universal encyclopedia continues a tradition from ancient encyclopedic efforts to H.G. Wells' "World Brain" that would bring forth "a common understanding" (Reagle, 2010: 25).

Wikipedia's apparent pluralism was what attracted me to the project. When I co-founded Creative Commons in South Africa in 2005, I recognised that Wikipedia was important not only because it provided free *access* to knowledge but because it provided a platform for local *participation* on a global platform. When I was appointed as the Executive Director of iCommons, a non-profit organisation set up by Creative Commons to lead international efforts around Creative Commons, I recognised that Wikipedia was the world's greatest example of the power of free and open source software and open content.

I became a passionate activist for Wikipedia in Africa. On the 10th of November, 2007, I co-organised the first Wikipedia Academy (as edit-a-thons used to be called) in Africa (Wikimedia Foundation, 2007). The event was hosted by CIDA City Campus, a university that served a majority of disadvantaged students at their campus in Johannesburg, offering students a full scholarship for tuition and living expenses. Jimmy Wales and Swahili Wikipedian, Ndesanjo Macha came to the event to talk about the importance of Wikipedia for preserving local knowledges. Students from a local tertiary education institution enthusiastically started articles in their home languages. On South Africa's annual Heritage Day, the iCommons team set up a stall in our local mall to accept donations of old photographs and artefacts to Wikimedia Commons, the site where most of the photographs used on Wikipedia are housed. We enthusiastically accepted a dribble of donations and proselytized to the already converted about the value of "free knowledge". I believed, at that time, that Wikipedia was truly open to the knowledge of all, and that it only required our acceptance of the invitation to participate in building "the sum of all human knowledge".

But a few years on and I had become increasingly sceptical of the global promise of the movement from my vantage point in South Africa. Even Wikipedia was being unmasked as a project riddled with bias and ignorance about knowledge from outside the West. When the oral citations project was launched in 2011, I was in graduate school, trying to understand how social norms and organisational rules could create barriers to information and knowledge sharing, despite the existence of open licenses. I started researching Wikipedia for a small project during my Master's degree at UCBerkeley and started my ethnographic study of Wikipedia in 2012, when I worked as an ethnographer for the Kenyan non-profit technology company, Ushahidi.

Media anthropology inspired me to discover Wikipedia from an alternative perspective to the one that dominated the field known as Wikipedia Studies at the time. In this chapter, I argue that, in addition to the ethnographic studies of platforms' pro-users, its dominant actors, its most popular or active groups, we need ethnographies of the digitally dispossessed: those who are routinely ignored, omitted, denigrated and denounced by platforms. In the case of Wikipedia, ethnographies of the digitally dispossessed enable us to understand how platforms shape the world, how we see one another, and how we might improve the ways in which they work for the benefit of underserved groups.

Media anthropology and the sources of power/knowledge

Media anthropology's principles for understanding the *multiplicity* of peoples' experiences of global platforms is key to its value in studying global digital cultures. At the heart of media

anthropology is the idea that media practices are not universal. Media anthropologists study the ways in which media are designed or adapted for use by specific communities or groups. Borrowing a phrase coined by postcolonial theorist, Dipesh Chakrabarty (2000), the media anthropologist, E. Gabriella Coleman (2010) writes that ethnography is a useful methodology to study digital media platforms that are supposedly global because of how they can "provincialize" digital media. This enables ethnographers to "push back against peculiarly narrow presumptions about the universality of digital experience" (Coleman, 2010: 489).

For some ethnographers, these details are used to reveal "the splendour of sociocultural life" (Coleman, 2010: 497). In this tradition, ethnographies of Wikipedia have generally answered the question: how can Wikipedia exist? Wikipedia seems like an impossibility. Why would people give their labour for free to build an encyclopedia? Answering this question, Wikipedia ethnographers have illustrated the importance of the principle of "good faith collaboration" for how Wikipedia operates (Reagle, 2010) and the unique organisational form that constitutes Wikipedia work (Jemelniak, 2014).

For Joseph Reagle (2010), participant observation meant "observing – and occasionally participating in – the Wikipedia online community by "follow[ing] them 'in real time' via a number of venues" including Wikipedia pages and edits to them, talk pages, mailing lists, newsletters and Wikipedia meetups (Reagle, 2010: 9). For Dariusz Jemelniak, it resulted in significant Wikipedia labour which saw him "moving through all the ranks of the Wikipedia parahierarchy" (Jemelniak, 2014: 197). Jemelniak eventually became a steward (the role with the widest access to technical privileges across all Wikimedia projects) and the Chair of the Funds Dissemination Committee, a global advisory body to the Wikimedia Foundation.

For other ethnographers, details are "ethically deployed to push against faulty and narrow presumptions about the universality and uniformity of human experience" (Coleman, 2010: 497). It is this orientation that aligns most strongly with my own ethnographic approach. Like other ethnographies, my own ethnographic study of Wikipedia involved with editors and participant observation of practice both in the work of the encyclopedia and in the context of the face to face events that Wikipedians gather at. But my choice of articles to follow, people to interview and practice to participate in is shaped by a commitment to the subaltern perspective. By subaltern, I mean those who are "removed from all lines of social mobility" (Spivak, 2005: 475). The subaltern, in this sense, is not just Other, minority, or disadvantaged. They are essentially unable to speak for themselves within existing power structures that constitute Wikipedia as a knowledge project.

Previous ethnographies of Wikipedia have tended to explore Wikipedia from the perspective of its most active contributors, that is, through the lens of those who already have significant power on Wikipedia. Other ethnographies have pointed to the multiple alternative ways of seeing and experiencing Wikipedia. My own ethnographic study of Wikipedia is situated from the perspective of those that exist in the shadow of the encyclopedia – those who edit Wikipedia far from its cultural center in Silicon Valley, those who have been banned, whose articles have been deleted and who are opposed to what they believe is the hegemony of the mainly-Western, white, male Wikimedia community (Ford, 2011; Ford and Geiger, 2012; Ford, 2016; Ford, 2017). This orientation has influenced how I have situated myself within the field, who I have chosen to interview, which articles I have chosen to select as cases, what kinds of practice I have observed and participated in.

Indeed, when faced with the opportunity to study a platform ethnographically, there are a myriad of choices that need to be made. Wikipedia is made up of millions of articles, hundreds of working parties (or Wikiprojects) and article types, a myriad policy pages and "meta" discussions taking place on wiki and on mailing lists and other social media platforms. There are over 300 language versions of Wikipedia and Wikipedia is only one of a number of Wikimedia projects interconnected with one another. Wikimedia Commons stores media files, for example; Wikidata stores infobox data, interwiki links and other data. Projects are often distinguished by their own rules, norms and cultures. I found myself asking: How can one claim to capture Wikipedia's culture when one can only experience a tiny slice of it? More importantly: which slice to capture? Which projects to follow? Which practice to participate in?

In the sections that follow, I present three strategies that I used to explore Wikipedia from the perspective of the digital subaltern. First, I focus on how the surr article raised questions about how factual claims are obstructed, refused or demeaned by others. Second, I explore how different forms of authority were used to support (or oppose) these claims and what kinds of expertise are ignored or downplayed. Finally, I trace how claims about the authority of the surr article were (or were not) constituted and animated by data and what this means for how knowledge becomes authoritative (or not). I explain how I applied these strategies in the context of the article about surr on English Wikipedia, demonstrating the ways in which digital and increasingly datafied knowledge representations gain purchase or stop in their tracks via online platforms like Wikipedia.

Entering via flagged erasures

The ethnographer, Jenna Burrell suggests seeking entry points to networks rather than identifying sites in order to locate one's fieldwork. This requires first establishing "what position(s) to take within the network" (Burrell, 2009: 190). Wikipedia's articles that had been flagged for deletion were useful entry points in line with my focus on the subaltern perspective and enabled me to learn more about the practices by which some knowledges are excluded from Wikipedia. In order to facilitate the representation of phenomena in Wikipedia, a space needs to be created for the recording of facts surrounding the phenomenon. The ideal form of space creation in Wikipedia is the creation of an article. The new article contains no baggage of previous articles, with their embedded authors, narrative frames, and teeming desires.

Anyone can create a new Wikipedia article if a topic doesn't yet exist. But those articles can be just as quickly deleted by other editors. Wikipedia's editors are patrolling new articles with the assistance of automated tools that mark up the articles according to some of the traces of their construction. Those traces include the edit experience of the authoring editor, the existence of categories in the article, and links to other Wikipedia articles.

One of the key features of the software and the practice of page review is to be able to rapidly remove what Wikipedia calls "bad-faith contributions," such as "attack pages" or copyright violations. The feed enables editors to view metadata related to the article, such as who "patrolled" (reviewed) the article or whether it has been nominated for deletion, as well as data about the article itself (its size, a preview of the text, whether it has been categorized, and how many users have contributed to it). Page patrollers' first view of the article is through its data.

In the case of surr, I followed edits made to the Hindi and English version of the article (including attempts to delete and vandalise it), discussions about the article on the Reliable Sources Noticeboard on English Wikipedia and the global Wikimedia mailing list (Wikimedia-l). I interviewed editors involved in creating articles and sources, and one editor leading efforts to discredit them. The article flagged for deletion generated a wide map of people, discourses and practices and technologies while centring on the topic of exclusion.

Following sources and their characterization

Articles can be accepted or deleted on Wikipedia, but another significant way in which knowledge is included or excluded is via the sources that are cited in support of facts in articles. Wikipedia is built on the principle of verifiability, where all statements must be able to be verified by a "reliable source".

Reliability on Wikipedia has specifically been equated with "secondary" sources published according to institutional standards, rather than sources as people themselves (as in the tradition of journalism or social science). English Wikipedia policy forbids the publishing of "original research". This means that articles "may not contain any new analysis or synthesis of published material that serves to reach or imply a conclusion not clearly stated by the sources themselves" (Wikipedia: No Original Research, n.d.).

The ways in which Wikipedians decide which sources are reliable, as well as the ways in which they summarise those sources, all involve Wikipedians actively creating knowledge by deciding what is excluded. Wikipedia practice involves discourse and actions that serve to classify some sources as reliable and other sources as unreliable. Following the ways in which local sources of knowledge on Wikipedia are accepted, rejected, debated or not, is central to understanding the practices of exclusion on the platform.

In the case of surr, the article became a site of controversy because it challenged the idea that Wikipedians do not actively curate knowledge. The oral citations model enabled activist Wikipedians to record the knowledge of ordinary people who they recognised as authorities on local culture. But many Wikipedians fought back, re-asserting the expertise of institutional academics (in this case, historians and anthropologists) as the proper authority on knowledge and reinforcing the idea that they, the editors, were merely passive curators.

Following discussions about citations for the Wikipedia surr article makes visible multiple layers of sources for the facts stated in Wikipedia articles. These include the editors themselves, who make decisions about what facts to select, which sources are cited and how facts are summarized. It includes secondary sources, which may reflect a topic alternatively or in opposition. In the background are the knowledge holders, who are not always represented accurately (or at all) by sources. They include the witnesses to events, or those identified by information.

I followed discussions about the oral citations used in surr in at least two key sites outside the article to understand how decisions are made to exclude knowledge on Wikipedia. The Reliable Sources Noticeboard on English Wikipedia, for example, is a place where editors can go to query and discuss the appropriateness of particular sources being used in articles. In the case of surr, a number of editors weighed in on the reliability of oral citations on the surr article by pointing to the identities of those involved in producing the citations (interviewees,

interviewers and Wikipedia editors). In one example, the local expert interviewed in the audio clip was described as "a layperson".

> (T)he person interviewed has no academic authority in the field. He is simply a layperson, who has played a game of unknown notability. (NativeForeigner, Reliable Sources Noticeboard discussion, 13 February 2012)[1]

In another statement, the publisher of the source on Wikimedia Commons (Aprabhala) is labeled as a mere 'commons content creator'. Reliability, according to these editors, is determined not by the quality of the source but by the identity of its producers.

> I'm sorry, but commons user Aprabhala is not a professional or academic ethnographer; they're a commons content creator. (Fifelfoo, Reliable Sources Noticeboard discussion, 13 February 2012)

A week after the debate on the Reliable Sources Noticeboard, the Oral Citations Project was discussed on the Wikimedia-l mailing list. Wikimedia-l, is one of about 400 mailing lists administered by the Wikimedia Foundation but it occupies a special position in the Wikimedia network because of its role in providing a forum for issues relating to Wikimedia Foundation projects. The Wikimedia-l mailing list is used to discuss issues relating to new projects, new chapters, polling and fundraising and is where the majority of high-level, strategic and long-time discussions about movement-wide issues take place between the Wikimedia Foundation and non-Foundation members of the Wikimedia community.

The discussion about oral citations began when the project lead, Achal Prabhala sought advice from the Wikimedia community about the Oral Citations Project.[2] Fourteen editors weighed in on the question, many were significant members of the Wikimedia movement. Disagreements centred around the role of Wikipedians in terms of the knowledge that they represent, and about the subjects of knowledge and the role of those subjects in the representation process.

The discussion about oral citations ended after less than a week of replies with no clear consensus on whether allowing oral citations as defined by Prabhala and his colleagues was conducive to Wikipedia's goals or identity. None of the editors participating in the thread on Wikimedia-l edited the surr article during or after this discussion, but work continued on the article by a variety of actors engaged in the daily practice of article maintenance and construction.

Opposing editors attempted to label the article as unreliable and incomplete. They added warning tags to the article in order to cast doubt on the content of the article, they deleted text, thereby discrediting claims within the article and they removed the links to the oral citations hosted on Wikimedia Commons. They also threatened to delete the citations themselves by adding a request for copyright permissions in the format specified by policy

---

[1] The RS/N discussion about oral citations is archived at
https://en.wikipedia.org/wiki/Wikipedia:Reliable_sources/Noticeboard/Archive_115#Oral_Citations.

[2] Archived discussion available at https://lists.wikimedia.org/pipermail/wikimedia-l/2012-February/thread.html.

and denied the authority of the Oral Citations Project page on Meta (Wikimedia's coordination site) by arguing that Meta was irrelevant to the work of English Wikipedia.

Analysing the networks that traverse articles

Wikipedia is increasingly a data project rather than a traditional encyclopedia project. Its facts are extracted to populate vast knowledge bases that present single answers to our questions about the world on search engines and digital assistants. Asking Google or Siri who won the Russian election, what happened in Egypt in 2011 or what is the capital city of Israel results in facts often gleaned from Wikipedia.

In order for facts to travel to more popular sites like Google, they must be structured as data on Wikipedia. This primarily happens in the infobox, the box on the right-hand side of an article that lists its key facts. Starting in 2013, Wikipedians began to develop infoboxes as structured data. Structured data differs from digital data in that it is tightly controlled by a data model that dictates its structure and order. It is well-defined in this way so that it can be easily accessed by computer programs.

The structuring of claims, it turns out, is determined by a particular kind of representational politics. The heterogeneity of these networks creates power dynamics because only certain users are able to control representation in multiple spaces, each of which requires familiarity and specialised literacy. Tracing how facts travel across networks and how they are structured as data, then, is a useful way of understanding how power is inscribed according to the ways in which claims are able to circulate.

Some knowledge doesn't seem fit the data models provided, some claims haven't yet been structured because of a lack of interest or understanding by Wikipedians who care about them. Whereas the sport of cricket on Wikipedia has an infobox that is richly detailed, including facts about when and where the sport was first played, for example, surr has no infobox. It lists just one other "interwiki" link to the Hindi version of the article and it links to only two categories on Wikipedia. It is thinly described on Wikidata, the Wikimedia project where structured data from infoboxes is stored.

The arrival of Wikidata has shifted power relations among knowledge communities in Wikimedia. Even though Wikidata's entities are open for public scrutiny and editing, many Wikipedia editors lose their ability to make meaningful changes to the facts that they constructed once they move to Wikidata. Editing a semantic database turns out to require very different expertise and the ways in which editors think about the topics that they edit are different still. If there is conflict in the ways that items are edited, discussions about those items need to take place in English because Wikidata is a centralised database for Wikipedia's more than 300 language projects.

Analysing the networks that traverse articles, I uncovered a universe of new relations that were being discounted as "merely technical" but had real, material effects on the ways in which people and their knowledge was represented. The structured data projects in which Wikimedia is currently involved will have an impact on how minority language groups are able to control representations on Wikipedia and outwards to the wider Web. Understanding the metadata that drives those representations along silent transport routes are critical to engaging with knowledge platforms like Wikipedia. The question, as the Wikipedia ethnographer, Stuart Geiger (2017) writes, is "for whom are algorithmic systems (and the

organizations that rely on them) formal, rigid, and consistent, and for whom are they in flux, revisable, and negotiable?"

Structured data matters in the context of Wikipedia because it is a new means of making knowledge visible – not only on Wikipedia but on the Web, since structured data claims travel more readily to powerful platforms like Google, Siri and Alexa. In order to understand how knowledge is being excluded, we have to trace the circulation of data in addition to focusing on where it lands. It is not enough for knowledge claims to be represented on platforms like Wikipedia. If they aren't discoverable, then they will be buried and ignored, unmaintained and unrecognised. The ways in which communities are represented by factual claims on platforms like Wikipedia (and consequently, Google) matters to people who are so often unseen and whose knowledge and experience are unacknowledged in global systems. Data makes knowledge visible and it is in data's circulation that new centres of power/knowledge arise.

New centres of power/knowledge

The arrival of the Internet was accompanied by a new idealism about global solidarity. The Internet enabled people to come together on a common platform to share knowledge and understanding. As John Perry Barlow (1996) famously stated about the Internet in his "Declaration of the Independence of Cyberspace," "We are creating a world that all may enter without privilege or prejudice accorded by race, economic power, military force, or station of birth." More than 15 years on and we now scoff at Barlow's idealism. As Science and Technology Studies scholar, Sheila Jasanoff (2004: 36) writes that power is continually reinscribing itself in institutions, practices, discourses, claims and products of science and technology. No platform exists outside of this constant process.

How can we best understand how power is reinscribing itself in the context of digital systems, beyond the simplistic claim that platforms or algorithms are the only subjugating force? How can we conduct research that will actually help to change the conditions in which so many have to endure when attempting to present their knowledge as legitimate? The dominant framing of research in the context of digital inequalities is in terms of platform "bias". The majority of studies exploring platform bias use statistical models to compare how certain subjects or sources are represented compared to others. In the context of Wikipedia, much work has been done to demonstrate how Wikipedia over-represents current history rather than events from the distant past (Graham, Hale and Stephens, 2011), men rather than women (Reagle and Rhue, 2011), places in the global North rather than the global South (Graham et al., 2014; Sen et al., 2015).

Although this work has been instrumental in determining what gaps exist, they don't get us any closer to understanding the sources of bias or what might be done to solve such problems. Furthermore, knowledge claims are increasingly made accessible to the world not via the platform itself but by its data doppelgangers, copied automatically across to digital assistants and smart search engines as answers to users' queries. The types of content, topics and language versions of content that are highlighted in these systems also demonstrates biases. Some content is more likely to be represented in more popular engines than others, but the source of that content is often obscured. What does platform bias mean in the context of a Web that is increasingly interconnected and fragmented at the same time?

Ethnographies that follow disputed claims, their incumbent authorities on both sides of the debates and the networks that traverse them are a useful alternative to the quantitative methodologies exploring bias. Starting from deleted or disputed knowledge claims as entry points to heterogeneous networks, ethnography can surface the conversations and actions that work to disable and destabilise certain types of knowledge. The example of surr demonstrates how the power to represent knowledge is lodged in discourses concerning the sources of reliable knowledge, representations that determine who the expert and subject are, and identities that authorise certain knowledge holders as legitimate and others as illegitimate.

Achal Prabhala completed the Oral Citations Project but it left him bitterly disappointed at Wikipedia's response. In my interviews with him, he talked about opportunities lost in favour of what he saw as the Western conservatism of Wikipedia. Although surr remains on English Wikipedia, ten years after it was first published, it is thinly described. The potential for including many more local sports and aspects of cultural life was abandoned as proponents of the oral citations project were left dismayed at the attitudes of Wikipedia's dominant groups. Wikipedia rejected the Oral Citations Project because they thought it would move them out of the passive curator role they thought they inhabited. But Wikipedians are active curators of knowledge, as discussions like this confirms.

Advocates of the project chose players of surr, rather than historians or anthropologists to be interviewed for the surr article, effectively upending Wikipedia's notions of expertise as located within institutions, particularly academia. The project advocates were effectively suggesting that the source of knowledge is in the embodied experience of phenomena rather than through institutional certification. This idea represented a threat, not only to the series of articles using oral citations, but to Wikipedia's conceptions about its own role in the system of expertise more broadly. By choosing to interview individuals in the community who had first-hand knowledge of the game, the editors representing surr in the original Hindi had effectively transformed those community members into experts. Similarly, the editors who conducted the interviews also displayed expertise because of their knowledge of the phenomenon existed and knowledge of how to identify and contact the experts in the villages.

The response by some English Wikipedia editors was to reaffirm the expertise of academics and institutions, thereby denigrating the embodied experience of the villagers. When the article entered English Wikipedia's socio-technical system, the voices of those who experience the game were no longer available since the link to the audio file had been removed. Instead of the community members being recognised as experts, they were relegated to being the subjects of the newspaper article about surr that constitutes the only linked citation within the article when this study was conducted. Wikipedia editors' success lay in their ability to define the roles and identities of Wikipedians and the experts that they rely on.

It seems that the call from the "Critical Point of View: A Wikipedia Reader" is still as relevant as it was in 2011. "The task is to create new encounters and point to new modes of inquiry, to connect the new with the old, and to give voice to different, 'subjugated' histories." (Lovinck and Tkacz, 2011: 10) What platforms include and exclude is not just a technical data point on a map about coverage. In the case of surr, for example, the acceptance or rejection of an article about cultural activity in rural India is symbolic of whether colonial attitudes towards former colonies have been abandoned or whether those attitudes have merely taken on a new form.

The results of platform ethnographies are constituted according to decisions made at every step in forming a fieldsite within these massive territories. I advocate for exploring how subaltern populations try to use supposedly global platforms in order to be heard, how those attempts often fail and what this says about how they can be more inclusively designed. In the context of an Internet that is increasingly subject to the logic of automation and AI, voice is enabled through one's mastering of practice within data centric representational forms.

My research on the surr Wikipedia article is intended to push back against the idea that Wikipedia can only be one thing to all people. This orientation aligns strongly with my desire to study systems not only to understand them but in order to improve their workings for those who can most benefit from them. Ethnographies of the digitally dispossessed are one key component of this ongoing task. Connecting knowledge platforms to the stories of the subjugated, dispossessed and disenchanted is a valid goal as we remain under the influence of their representations.

References:

Barlow, J.P. (1996). A Declaration of the Independence of Cyberspace. Retrieved 1 August from https://www.eff.org/cyberspace-independence.

Burrell, J. (2009). The fieldsite as a network: A strategy for locating ethnographic research. Field Methods, 21(2), 181–199. doi: 10.1177/1525822X08329699

Chakrabarty D. (2000). Provincializing Europe: Postcolonial Thought and Historical Difference. Princeton, NJ: Princeton Univ. Press

Coleman, E. G. (2010). Ethnographic approaches to digital media. Annual review of anthropology, 39, 487-505.

Ford, H. (2011). The Missing Wikipedians. In Geert Lovink and Nathaniel Tkacz (eds), Critical Point of View: A Wikipedia Reader, Amsterdam: Institute of Network Cultures, 2011. ISBN: 978-90-78146-13-1

Ford, H. and Geiger, S. (2012). Writing up rather than writing down: Becoming Wikipedia literate'. ACM WikiSym '12, Linz, Austria

Ford, H. (2016) Wikipedia and the sum of all human information. Nordisk Tidsskrift for Informationsvidenskab og Kulturformidling. 5(1), 9-13.

Ford, H. (2017) The search for Wikipedia's edges. In L. Hjorth, H. Horst, A. Galloway, G. Bell (Eds), Routledge Companion to Digital Ethnography. Routledge.

Ford, H. and Wajcman, J. (2017) 'Anyone can edit'. Not everyone does. Wikipedia and the gender gap. Social Studies of Science Journal.

Geiger, R. S. (2017). Beyond opening up the black box: Investigating the role of algorithmic systems in Wikipedian organizational culture. Big Data & Society. https://doi.org/10.1177/2053951717730735

Graham, M. (2009, December 2). Wikipedia's known unknowns. *The Guardian.* https://www.theguardian.com/technology/2009/dec/02/wikipedia-known-unknowns-geotagging-knowledge

Graham, M., Hale, S. and Stephens, M. (2011). Geographies of the World's Knowledge. Technical report. Convoco, London.

Lovink, G. & Tkacz, N. (Eds.). (2011). Critical Point of View: A Wikipedia Reader, Amsterdam: Institute of Network Cultures, 2011. ISBN: 978-90-78146-13-1.

Graham, M., Hogan, B., Straumann, R.K. and Medhat, A. (2014). Uneven geographies of user-generated information: Patterns of increasing informational poverty. Annals of the Association of American Geographers, 104, 4, 746–764.

Jasanoff, S. (Ed.). (2004). States of knowledge: the co-production of science and the social order. Routledge.

Jemielniak, D. (2014). *Common knowledge?: An ethnography of Wikipedia*. Stanford University Press.

Raymond, E. (1999). The cathedral and the bazaar. *Knowledge, Technology & Policy*, *12*(3), 23-49.

Reagle, J. (2010). Good Faith Collaboration: The Culture of Wikipedia. MIT Press.

Reagle, J., & Rhue, L. (2011). Gender bias in Wikipedia and Britannica. *International Journal of Communication*, *5*, 21.

Sen, S. W., Ford, H., Musicant, D. R., Graham, M., Keyes, O. S., & Hecht, B. (2015). Barriers to the localness of volunteered geographic information. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 197-206).

Spivak, Gayatri. (2005). Scattered Speculations on the subaltern and the popular. *Postcolonial Studies, 8(4).* https://doi.org/10.1080/13688790500375132

Wikimedia Foundation. (2007, October 31). Wikipedia founder Jimmy Wales leads South African Wikipedia Academies in Johannesburg. Wikimedia Foundation Website. https://foundation.wikimedia.org/wiki/Press_releases/Wikipedia_Academies

Wikimedia Foundation. (n.d.) Our Work. Retrieved August 1, 2021, from https://wikimediafoundation.org/our-work/

Wikipedia: No original research policy. (n.d.) Retrieved August 1, 2021, from https://en.wikipedia.org/w/index.php?title=Wikipedia:No_original_research&oldid=1012596 131