World Heritage sites on Wikipedia: Cultural heritage activism in a context of constrained agency

Ben Marwick ⓘ **and Prema Smith** ⓘ

Abstract

UNESCO World Heritage sites are places of outstanding significance and often key sources of information that influence how people interact with the past today. The process of inscription on the UNESCO list is complicated and intersects with political and commercial controversies. But how well are these controversies known to the public? Wikipedia pages on these sites offer a unique dataset for insights into public understanding of heritage controversies. The unique technicity of Wikipedia, with its bot ecosystem and editing mechanics, shapes how knowledge about cultural heritage is constructed and how controversies are negotiated and communicated. In this article, we investigate the patterns of production, consumption, and spatial and temporal distributions of Wikipedia pages for World Heritage cultural sites. We find that Wikipedia provides a distinctive context for investigating how people experience and relate to the past in the present. The agency of participants is highly constrained, but distinctive, behind-the-scenes expressions of cultural heritage activism are evident. Concerns about state-like actors, violence and destruction, deal-making, etc. in the World Heritage inscription process are present, but rare on Wikipedia's World Heritage pages. Instead, hyper-local and process issues dominate controversies on Wikipedia. We describe how this kind of research, drawing on Big Data and data science methods, contributes to digital heritage studies and also reveals its limitations.

Keywords

Wikipedia, UNESCO, digital heritage, World Heritage, social archaeology, data science

This article is a part of special theme on Heritage in a World of Big Data. To see a full list of all articles in this special theme, please click here: https://journals.sagepub.com/page/bds/collections/heritageinworldbigdata

Introduction

Heritage is the processes and outcomes of people engaging with elements of the past – material and immaterial – and attributing social and cultural meanings to them in the present (Harrison 2013; Smith 2006). These are important to understand because they shape peoples' identities and influence how they think and behave toward other people. Digital heritage are engagement with elements of the past that are enabled by the Internet (Bonacchi and Krzyzanska, 2019), leaving traces that can be identified and quantified using data science methods. Digital heritage studies represent a major turn from traditional heritage studies, characterized by post-modernism (Kristiansen, 2014), critical theory, and qualitative methods, toward novel ontologies, data-intensive ethnographies, and a new role for heritage scholars as data

scientists. Bonacchi et al. (2018, 2019) have sketched out the new digital heritage research program with their combination of data-intensive and qualitative investigations of 1.4m Facebook posts in Brexit-related community groups. They found recurring parallels – both pro- and anti-Brexit – made by Facebook users between the European Union, the Roman Empire, and "barbarians" as they use heritage to support their political activism. They demonstrate the

Department of Anthropology, University of Washington, Seattle, WA, USA

Corresponding author:
Ben Marwick, Department of Anthropology, University of Washington, Denny Hall 117, Box 353100, Seattle, WA 98195-3100, USA.
Email: bmarwick@uw.edu

potential for understanding public perceptions and experiences of the past in contemporary society using Big Data obtained from social media. In this paper, we extend the digital heritage research program in two substantial new directions. First, we introduce Wikipedia as an example of an online peer production community where people engage with elements of the past in measurable ways. Second, we present a case study using data science methods to investigate the ways people create and consume English-language Wikipedia articles on cultural sites inscribed on the UNESCO World Heritage List (hereafter CS-WHL).

While social media, such as Facebook and Twitter, is a vast and diverse online space in which we are only just beginning to explore how people use to engage with the past, there are other contexts of online interactions where heritage is practiced in distinctive, if poorly understood, ways. We can contrast social media, with its fundamental elements of identity, conversations, sharing, presence, relationships, reputation, and groups (Kietzmann et al., 2011), with online peer production communities, where users participate in the collaborative, asynchronous, creating, sharing, promoting, and classifying of content in highly structured and goal-directed ways (Wilkinson, 2008). Online peer production communities are comparable to more traditional kinds of voluntary associations where groups set and execute goals, with explicitly democratic organizational ideals. While the ideals of many online peer production communities emphasize non-hierarchical and non-bureaucratic organization, analysis of large amounts of user activity indicates that most of these communities are actually undemocratic and noninclusive, functioning as entrenched oligarchies (Shaw and Hill, 2014). This emphasis on governance and management of collective action is a key detail that distinguishes social media from online peer production. It follows that user interactions in the process of generating content in online peer production communities include technological and social mechanisms that enact the community's written or unwritten governance policies. These may include, for example, limiting a user's activity according to their status in the community's hierarchy or managing conflict with highly structured procedures. Here, we show how the distinctive organizational and technical qualities of online peer production communities make them a unique context of heritage production to study digital traces of human activity resulting from engagement with the past.

Wikipedia as a context of heritage production

We present a study of how people engage with elements of the past in one of the largest and long-lived online peer production communities, the English-language Wikipedia. Originating in 2001, this is a highly influential and well-known online encyclopedia, that anyone can edit, with nine billion page views per month as of September 2020 (https://stats.wikimedia.org/#/en.wikipedia.org). Although anyone can edit, most internet users do not. Factors that strongly predict if a user has ever edited Wikipedia include their gender (male), age (younger), education level (has BA), Internet use frequency (higher), and Internet use skills (higher) (Adams and Brückner, 2015; Ford and Wajcman, 2017; Hill and Shaw, 2013; Shaw and Hargittai, 2018). There are also geographical disparities. Articles about rural areas have systematically lower quality, are less likely to have been produced by contributors who focus on the local area, and are more likely to have been generated by bots (automated software agents) (Johnson et al., 2016). These studies indicate that participation in online peer production communities often follow existing patterns of social exclusion. Graham et al. (2014) examined the global distribution of 3.4 million geotagged Wikipedia articles and find a pattern of places in the Global North being represented in local languages, while articles about places in the Global South are largely being written by others.

An additional consideration for understanding participation in online peer production communities are the technical schemas of MediaWiki, the software that Wikipedia runs on. This is a complex toolkit that enables participation in Wikipedia in highly structured ways. On one hand, these structured behaviors produce structured datasets that are well suited to data science methods for efficient computational analysis of large numbers Wikipedia articles. On the other hand, they constrain and limit the agency of the user, canalizing their behavior into a small number of possible actions and acceptable modes of discourse and engagement with other users (Iba et al., 2010). While Wikipedia has elements that are ubiquitous on the Internet, such as links that take the user to other articles or pages on the Internet, it also has several less common elements that contribute to its unique technicity, resulting in specific types of relationships between human users and the technical elements of the Wikipedia project (Niederer and Van Dijck, 2010; Weltevrede and Borra, 2016). For example, every edit to an article is tracked in a publicly accessible version control system associated with that article. This exposes the article creation process in highly granular detail; for any given article, we can see how many editors contributed, the size of their edits and their distribution over time, among other things (Priedhorsky et al., 2007). Wikipedia has a special category of edit called the "revert" which allows a user to restore an article to an earlier state to remove recent vandalism (such as the addition of irrelevant or offensive

material). This special revert action, combined with a "talk" page attached to each encyclopedia article for threaded discussion among editors, allows us to detect and study the social dynamics arising from the creation and editing of articles, for example, the controversiality of an article (Suh et al., 2007; Yasseri et al., 2012). While articles themselves must be written to conform to the fundamental Wikipedia policy of Neutral Point Of View (NPOV), the talk page is where different views are expressed and negotiated among editors.

In addition to the human users and the technological system that enables and constrains their activities on Wikipedia, there is an important third element of the ecosystem that contributes to Wikipedia's uniqueness: the bots. Wikipedia bots are computer scripts that automatically handle repetitive and mundane tasks to develop, improve, and maintain the encyclopedia (Zheng et al., 2019). While bots are not unique to Wikipedia, they are important contributors and are responsible for a large proportion of edits (Geiger, 2009, 2014; Niederer and Van Dijck, 2010). They also evolve and autonomously engage in complex interactions with other bots to modify the encyclopedia (Geiger and Halfaker, 2017; Tsvetkova et al., 2017).

Contentious UNESCO World Heritage cultural sites

We investigate how the unique technicity of Wikipedia shapes interactions between people and the past with a case study of cultural sites inscribed on the WHL. We chose the CS-WHL as a bounded set of cultural heritage elements with several characteristics that make it

of general interest. It has a global geographic distribution; broad public interest at local and international scales, in both online and face-to-face communities; a wide temporal distribution in both the age of the cultural sites, the ages of inscription on the WHL, and the ages of their appearance on Wikipedia; and finally, many CS-WHL have a high intensity of cultural and political discussions that surround events affecting these sites, such as their inscription on the WHL. These qualities make it an ideal data set as an entry point for case studies of digital heritage in online peer production communities, where activities are typically goal-driven (e.g. "write quality articles") compared to social media activity where user activities are more often event-driven (e.g. "share reactions to Brexit").

UNESCO was established in 1945, shortly after the end of the Second World War, for the purpose of helping to rebuild after the war and preserve peace by promoting the international exchange of ideas. In 1975, the UNESCO-drafted "Convention Concerning the Protection of the World Cultural and Natural Heritage" came into force and established the WHL to protect natural and cultural sites and landscapes around the world that have outstanding universal value. As of September 2020, there are 869 cultural properties on the UNESCO WHL, with the first sites inscribed in 1978. On average, most countries have two to three sites, with most sites located in Italy and Western Europe, and several countries having no sites at all, for example, several central African countries, Taiwan, and New Zealand (Figure 1).
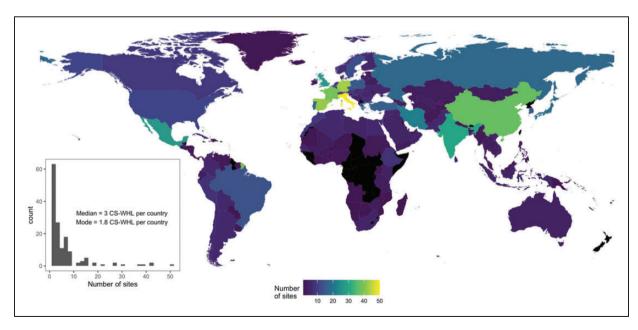

Figure 1. Cultural sites on the UNESCO WHL as of September 2020. Countries colored black have no listed cultural sites at the time of writing. Inset shows the distribution of sites per country. Map data from naturalearthdata.com.

Several CS-WHL sites are notable for the conflicts and tensions that have surrounded their inscription (Meskell, 2018). The 1992 inscription of Angkor (an ancient city and empire in Cambodia, prominent during the 9th to the 15th centuries AD) was encouraged by exiled supporters of the genocidal Khmer Rouge regime, hoping to strengthen territorial claims (Locard, 2015). They appropriated Western discourse on national cultural heritage to argue for the safeguarding of Angkor as part of their quest for national independence and international recognition. Early in the Khmer Rouge regime, Angkor was declared a symbol of enslavement by a primitive culture, but when the Khmer Rouge adopted a new rhetoric of a supposedly civilizing mission, they presented it as the site one of the great world civilizations (Falser, 2015). The 2003 inscription of Mapungubwe (the site of the first indigenous kingdom in Southern Africa, 900–1300 AD) was preceded by a recommendation from ICOMOS (International Council on Monuments and Sites, a professional association that is a key advisory body to the World Heritage Committee) not to inscribe because of the farming and mining activity in highly sensitive areas near the site and the unclear ownership of the mining rights at the time (Meskell, 2011). Despite this negative recommendation, geopolitical machinations within the Committee, especially by the Indian and Russian delegates, led to Mapungubwe being inscribed on the list, although without the typical prerequisites of a management plan or complete buffer zone (Meskell, 2012). These examples of Angkor and Mapungubwe demonstrate the attention that the WHL inscription process can generate due to political activism, conflicts, and intrigue.

Physical conflicts at or near CS-WHL are major events that also galvanize public interest in these locations. World Heritage sites in Palestine, Mali, Syria, Congo, and Cambodia have recently been sites of violence, in many cases, specifically linked to their potential WHL nomination, listing, or management. In 1998, anti-government and mostly Hindu Tamil groups bombed the holy Buddhist site of the Temple of the Tooth at the WHL site of Kandy (the last capital of the ancient kings of Sri Lanka), killing 17 people and substantially damaging the temple (Coningham and Lewer, 1999). In Mali, during 2012, fighting between government and rebel groups lead to the damage and destruction of tombs at the CS-WHL sites of Gao and Timbuktu (Brioschi, 2017). The World Heritage Committee found itself powerless to intervene because of political gridlock (Meskell, 2015), and these Mali sites are among the 53 cultural sites on the List of World Heritage in Danger, as of March 2021 (https://whc.unesco.org/en/danger/). In 2015, ISIS militants destroyed the Temple of Bel in Palmyra, Syria (a CS-WHL site of monumental ruins, once great city at the crossroads between east and west in the ancient world) (Gornik, 2015). Preah Vihear, inscribed in 2008, is a CS-WHL located on a long-disputed section of the Thai–Cambodia border that has been a site of both violent military clashes and international political intrigue. Although both Thailand and Cambodia supported the nomination of the site to the WHL, the Thai government objected to maps in the nomination package that showed Cambodia as the owner of disputed land next to the temple, leading to protests and military clashes (Sothirak, 2013). US diplomatic cables released by WikiLeaks reveal that settlement of disputes over Preah Vihear were intricately tied to broader issues of foreign policy and US and Chinese investment, especially access to natural gas reserves in the Gulf of Thailand (Meskell, 2016).

Methods

Our brief review of contentious cultural sites on the WHL shows the intensity and diversity of conflicts and tensions that surround these sites. Many CS-WHL are symbols of national, cultural, political, and religious identity, and the extent of political involvement in negotiations of WHL inscriptions indicates that they are of great public interest among local and diasporic communities. Our goal in this study is to answer the question of how this interest is expressed within the socio-technical constraints of the English-language Wikipedia. We surveyed the basic characteristics of content (article length, number of Wikilinks out to other pages, number of citations to non-Wikipedia items), consumption (page view counts, Wikilinks in from other Wikipedia pages), and production (edit counts, edit densities, edit sizes, number of unique editors per article, talk page length, talk page topics). By comparing these basic characteristics of English-language Wikipedia articles about CS-WHL to 10,000 random English-language Wikipedia articles, we can approach the question: can metrics of content, consumption, and production indicate engagement with the past via CS-WHL on Wikipedia? Can we detect conflict in the edit histories, bot activity, and talk pages for Wikipedia articles about CS-WHL sites, and how does this conflict relate to the types of controversies noted above? Random articles were obtained by sending GET requests to the "random" module in the Wikimedia REST API (https://en.wikipedia.org/api/rest_v1/).

The highly detailed edit histories that Wikipedia keeps for every article allow us to further investigate spatial and temporal questions relating to engagement with the past and conflicts surrounding CS-WHL sites. When an article is anonymously edited, for example, by

a user who does not have a Wikipedia user account (or is not logged into their account), their edit is identified by that person's IP address. An IP address can be used to geolocate the user to the country they were in when the made the edit. Edits made by people who are logged in to their Wikipedia user account do not include the user's IP address, only their Wikipedia user account name. This means that edits from registered Wikipedia users cannot be used for tracing the geographic origin of an edit, but anonymous edits can. We used the rgeolocate package for R (Keyes et al., 2020) to geolocate all edits with IP addresses for all English-language Wikipedia articles CS-WHL sites to determine the country of origin of those edits. This helps us to answer the question: are the editors of articles about CS-WHL located near the sites they edit, indicating local community interest in the online representation of their heritage? The time and date stamps attached to every edit on every article allow us to investigate temporal patterns of activity on CS-WHL Wikipedia articles. Analyses of these temporal data help us to answer the question: is Wikipedia editing activity correlated with events outside of Wikipedia relating to the CS-WHL sites, such as conflict events, or their inscription on the WHL?

We obtained data about Wikipedia articles by scraping the HTML pages with the rvest package for R (Wickham, 2019). We used the SelectorGadget (Cantino and Maxwell, 2017) extension for the Chrome web browser to identify specific page elements of interest, or nodes, on the HTML pages and wrote custom R functions to extract data from these nodes. Our entry points were the Wikipedia articles that are lists of World Heritage sites in major geographical regions of the globe. We found 15 of these and scraped the CS-WHL site names from the tables on these pages and followed the links to scrape the article text, edit history, and talk page text for each CS-WHL site included on those tables. A small number of CS-WHL sites have Wikipedia articles that are not included on these tables, but we did not include these in our sample. Starting at these regional lists of sites was a pragmatic choice because the individual Wikipedia article titles for CS-WHL sites very frequently differ from the official site name on the UNESCO list. A limitation of this approach is that it excludes "orphan" pages for CS-WHL that, while present in Wikipedia, have not been curated by editors into a table listing all the sites in a region. Thus, our sample is not the complete set of articles about CS-WHL, but only those that have been curated into regional groups. This approach ensures that our all sites in our sample are meaningful by sharing the essential quality of a taxonomic status of being categorized by Wikipedia editors as a CS-WHL in a certain region.

Reproducibility and open source materials

We collected data during May 2019, and due to the highly dynamic nature of Wikipedia, it is likely that articles in our study have subtly changed since our data collection, or that new ones have appeared. Our original code may no longer work on the most current version of Wikipedia without modification as the tables on Wikipedia articles continue to be modified by editors. Although we recognize that the fragility and temporally specific nature of our methods limits the reproducibility of our results, we include the entire R code (R Core Team, 2020) used for all the analysis and visualizations contained in this article in our compendium at http://doi.org/10.17605/OSF.IO/AY27G to enable reuse of our materials and improve reproducibility and transparency (Marwick, 2017). Also in this version-controlled compendium are the raw data for all the results reported here. All of the figures and quantitative results presented here can be independently reproduced with the code and data in this repository. In our compendium, our code is released under the MIT license, our data as CC-0, and our figures as CC-BY to enable maximum reuse (for more details, see Marwick et al. (2018)).

Results

Article content

Of the 869 cultural sites on the WHL at the time of writing, we found Wikipedia articles for 582. As a group, the basic details of content for CS-WHL Wikipedia articles differ little from a sample of 10,000 random Wikipedia articles (Figure 2). The scholarly nature of the articles, measured by the number of sources cited in the reference list per thousand words in the article body, has similar distributions for CS-WHL articles and random articles. The number of Wikilinks out from the target article to other Wikipedia articles are also similarly distributed for CS-WHL articles and random articles. The total number of words in a CS-WHL article is typically much higher than a random article, indicating that they receive more generative effort from editors than other articles.

Article production

Although details of content of CS-WHL articles are similar to our random sample, variables related to the production of Wikipedia articles on CS-WHL differ in important ways from other articles (Figure 3). The number of edits per thousand words, or edit density, and the number of unique editors per thousand words, or editor density, are substantially higher for CS-WHL
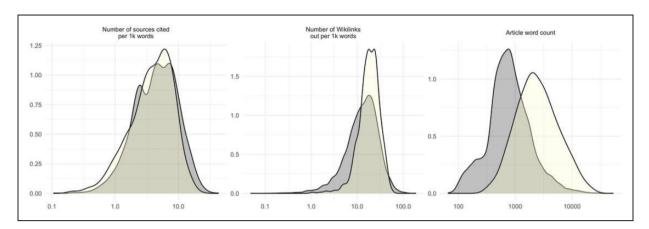
Figure 2. Content of Wikipedia articles about CS-WHL. The density plots show the distributions of basic content characteristics of Wikipedia articles about CS-WHL (yellow) compared to 10,000 random Wikipedia articles (grey).
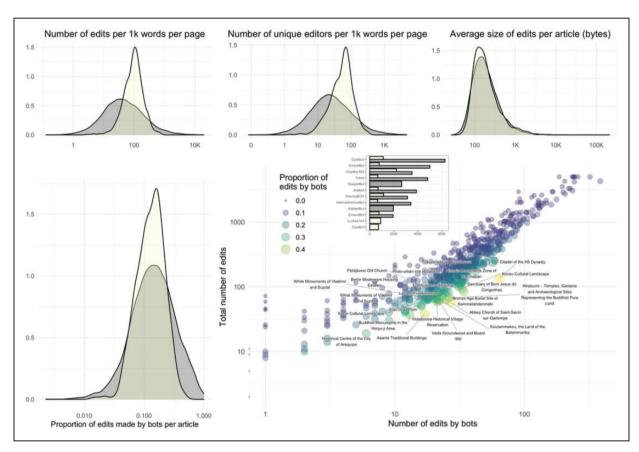

Figure 3. Production of Wikipedia articles about CS-WHL. The top row of density plots show the distributions of basic article production characteristics of Wikipedia articles about CS-WHL (yellow) compared to 10,000 random Wikipedia articles (grey). The density plots on the lower left show the distribution of edits made by bots. The lower right shows a scatterplot of production-by-bot metrics for Wikipedia articles about CS-WHL and includes labels on the articles where bots were responsible for >30% of edits. Inset on the scatterplot shows the number of edits for the top ten bots in our sample.

articles. This tells us that CS-WHL articles are intensively word-smithed by a more diverse community of editors than for other articles. The absolute size of edits (i.e. additions or removals of text) is about the same for CS-WHL articles as other articles. The involvement of

bots in producing CS-WHL articles is also about the same as for other articles. Bot activity is most intense on shorter, low-profile CS-WHL articles; in Figure 3, the labeled points are sites where bots have done >30% of edits. The most active bot on CS-WHL articles is
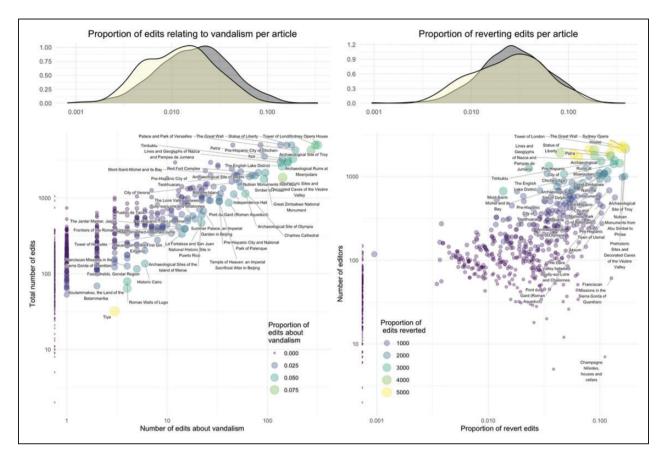
Figure 4. Reverted edits and edits about vandalism in Wikipedia articles about CS-WHL. The top row of density plots show the distributions of proportions of edits relating to vandalism, and the proportion of revert edits in Wikipedia articles about CS-WHL (yellow) compared to 10,000 random Wikipedia articles (grey). The scatterplots below show reverted edits and edits about vandalism metrics for Wikipedia articles about CS-WHL and include labels on the articles with high proportions of these types of edits.

Cluebot NG (vandalism detection and reverting) compared to Cydebot (automatic implementation of category deletions) for the random articles. The AnomieBOT, which performs clerical duties in an article's reference list, is highly active on CS-WHL articles compared to random articles. Most bot edits on CS-WHL articles are in the fixer, tagger, connector, and clerk roles (Zheng et al., 2019). None of these articles with intensive bot activity are CS-WHL sites of conflict or on the List of World Heritage in Danger, indicating that these sites receive little or no vandalism.

For the special "revert" edit type, we see that the proportion of all edits per CS-WHL articles is similar to other articles, but has a left-skewed distribution indicating a higher number of articles that have few revert edits (Figure 4). We also identified edits with the string "vandal" in the edit summary as a similar type of edit to the revert edit, e.g., "Edits by 72.49.241.71 identified as vandalism." CS-WHL articles generally have fewer edits about vandalism than our random sample. The shape of the distribution of edits about vandalism has a

smaller second mode to the left of the peak, indicating that a large number of CS-WHL articles have few edits about vandalism (Figure 4). Among the CS-WHL articles that have high proportions of reverts and edits about vandalism are highly iconic sites in the Western canon of culture history, e.g., the Sydney Opera House, the Tower of London, and the Statue of Liberty (cf. Harrison, 2013). In reviewing a sample of several hundred reverted edits for each of these, we found that nearly all of them are undoing the addition of short strings of text (e.g. profanities, spam, and nonsense). Much of this vandalism is playful, in the spirit of "'I am', a statement that one is present and alive", as Baker (2003) described historical graffiti on the Reichstag in Germany by Russian soldiers in the Second World War. Once again, of the CS-WHL sites with a history of conflict or on the in-danger list, only Timbuktu appears here as having high proportions of revert and vandalism-reversing edits.

Talk pages are an important locus of article production activity where we expect to see conflicts and debates unfold on Wikipedia. Wikipedia talk pages
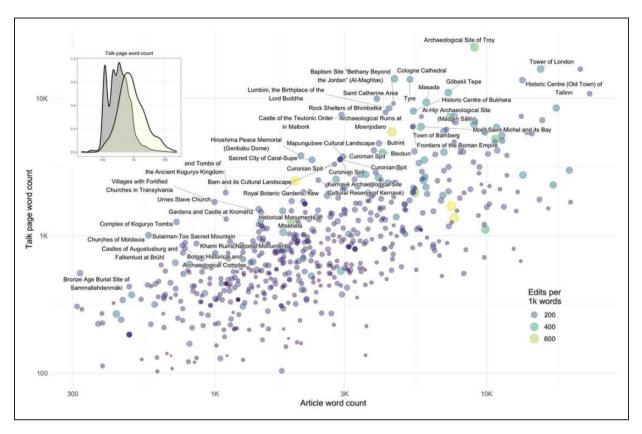
Figure 5. Scatterplot showing the length of each CS-WHL article and the length of each article's talk page. Labeled points are articles where the talk page is longer than the article. Inset shows the distribution of talk page lengths for CS-WHL articles and 10,000 random articles.

are a popular subject of investigations to understand the collaborative generation of knowledge, and online conflict management (Ho-Dac et al., 2016, 2017; Kittur et al., 2007; Schneider, Passant and Breslin, 2012). Yasseri et al. (2012) has shown that the length of an article's talk page is correlated with the controversality of the article and thus an effective simple proxy for conflict. We counted the words on all talk pages of the CS-WHL articles to identify conflict (Figure 5). Talk pages for CS-WHL articles tend to be much longer than other articles, which we expect due to the CS-WHL articles themselves being generally longer than other articles. However, the distribution of talk page lengths for CS-WHL articles has a long right tail, indicating that a higher number of articles have long talk pages compared to other articles. Some of these articles with long talk pages, such as Cologne Cathedral and Troy, have clear evidence of conflict among the editors in the contents of the text. However, close reading of the discussions on these talk pages reveals that these debates are dominated by technical issues of article production rather than conflicts and tensions at the CS-WHL or surrounding their inscription. For example, the Cologne Cathedral

talk page includes some debate about the correct calculation of the interior volume of the structure, and the Troy talk page includes heated comments by one editor about the removal of unsourced claims and unencyclopedic prose.

The talk page for Mapungubwe, which is longer than the article itself, is lengthy with expressions of concern about the accuracy of information in the article. In particular, editors were concerned about the depiction of Indigenous people, especially the degree to which early European colonists were aware of Indigenous communities and the correct cultural affiliation of the Indigenous groups that originally occupied the site. In this remarkable case, we see the Wikipedia editors engaging in conflicts that go beyond the typical technical details of article production. What is especially notable about these discussions on the talk page is how closely they echo the debates of land ownership that complicated the WHC process (Meskell, 2011). The specific editors involved in this debate did not make any edits to the article itself, which raises questions about their motivations for engaging in debate on the talk page, since they do not seem personally invested in the content of the

article. Without directly interviewing these editors, we cannot be sure of their motives for participating in the talk page.

A second important example of conflict on a talk page is the article about Preah Vihear Temple. Although not notable for the length of its talk page, the discussions among editors of the article about Preah Vihear Temple include a strongly worded disagreement among seven editors about the legitimacy of the ownership claims by Thailand and Cambodia of the territory that includes the temple. As for Mapungubwe, the debate among Wikipedia editors here mirrors closely the political tensions surrounding the nomination of Preah Vihear. The talk page also has a lengthy argument from July to October 2008, involving several editors, about whether or not the Thai name of the site should be included in the text of the article. The editors appear to be people of Thai and Khmer heritage, with comments that personalize the national tensions such as "we invaded you," and references to contemporary political tensions such as the 2008 street demonstrations in Thailand that saw conflict between the royalist People's Alliance for Democracy and the populist People's Power Party. One editor indicates their likely Thai ethnicity here through their use of Thai language phrases (using the Thai alphabet). This debate also recapitulates broader substantive issues of whether Thailand or Cambodia has the stronger claim for ownership of the temple. Amid accusations of nationalism among the editors, one of them asks "Can we put aside politics and be more collaborative on the encyclopedia project?" and appeals to Wikipedia's policies of "Third opinion" and "Dispute resolution" to diffuse the tensions and refocus attention on the common goal. Unlike the Mapungubwe talk page, many of the editors involved in the Preah Vihear talk page debates have made contributions directly to the Preah Vihear article and are deeply invested in how the article represents the site.

Article consumption

The basic metrics of consumption of CS-WHL articles show substantial differences from our sample of random articles (Figure 6). We measure consumption by counting the total number of views of the article over the 100 days prior to our data collection date and the number of Wikilinks from other articles into the target article. Wikipedia article view counts are popular widely used measures of cultural interest or salience (Cao et al., 2020; McIver and Brownstein, 2014; Roll et al., 2016). Wikilinks from other articles are a measure of the centrality of an article, if many other articles link to it, then the article is well-integrated into the encyclopedia and viewed as important for supporting information presented in other articles. CS-WHL articles are typically viewed far more frequently than other Wikipedia articles, reflecting high consumption by internet users generally. They are also much more often linked to by other Wikipedia articles than our random sample of other articles, indicating consumption by other Wikipedia articles and Wikipedia users in their editing work (Figure 6). This indicates that consumption of CS-WHL articles is generally high, relative to other articles, and confirms our assertion of Wikipedia as an important source of heritage information. But how is attention distributed across all CS-WHL articles, and how does it relate to sites with conflicts and tensions?

We can see an answer to this question in the scatterplot on the lower part of Figure 6, which shows the values of inward Wikilinks, article views, and article word count for all CS-WHL articles. The labeled points in the upper right quadrant are the articles that receive most of the attention. As noted above for sites with high proportions of reverts and edits, highly iconic sites in the Western canon of culture history are also the majority of sites that are highly consumed. Of the CS-WHL notable for conflict and tension, only Timbuktu is visible among these highly popular articles. It is also the only site on the List of World Heritage in Danger that is among these highly popular articles. We reviewed the talk page for the Timbuktu article to determine if the attention received by the article might relate to the issues leading to armed conflict at the site. Of the 6162 words on the talk page, only 168 are on the topic of conflict and destruction, with the editors discussing how to describe the scale of the damage to the temples. The majority of talk page content for Timbuktu is about filling in missing detail, suggestions for, or notifications of minor corrections. This is also the case for the talk page for Auschwitz, another popular CS-WHL article with six archived talk pages including over 100,000 words. Although Wikipedia articles in other languages on Auschwitz have different points of emphasis (Wolniewicz-Slomka, 2016), the majority of the discussion among editors on the talk pages is technical and precise. Provocative comments on the English language Auschwitz talk pages, for example, by editors denying the Holocaust, usually end after brief exchanges with other editors requesting credible sources. The rarity of conflict on the Auschwitz talk page can be contrasted with the talk page for the Holocaust, where Pfanzelter (2015) found abundant evidence of conflict among editors. Generally, our results show that conflict and tension at a CS-WHL are not highly correlated with how much attention an article receives.
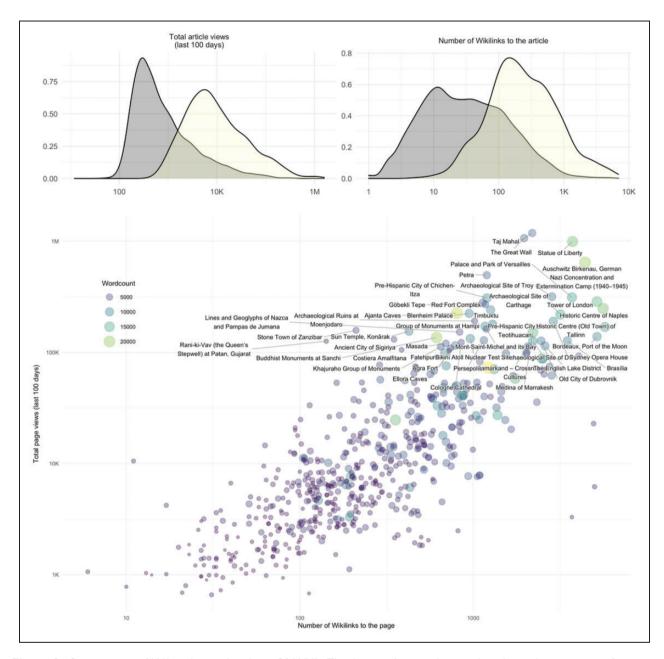
Figure 6. Consumption of Wikipedia articles about CS-WHL. The density plots on the top show basic characteristics of consumption for Wikipedia articles about CS-WHL (yellow) compared to 10,000 random Wikipedia articles (grey). The scatterplot at the bottom shows consumption metrics for Wikipedia articles about CS-WHL, with labels on the most intensively consumed articles.

Spatial patterns

Spatial patterns in the coverage of CS-WHL sites on Wikipedia closely resemble the physical distribution of sites. Figure 7 shows that most countries have between 2 and 3 CS-WHL on Wikipedia with Germany, India, China, France, and Spain being the countries with the most articles about their CS-WHL. This pattern is consistent with Graham et al. (2014)'s observations that places in the Global North tend to be over-represented on Wikipedia. We also see that some Asian countries outside of the Global North are highly represented on Wikipedia (e.g. India and China). The lower panel of Figure 7 further demonstrates this with South American countries having low proportions of their CS-WHL sites represented on Wikipedia. The pattern for African countries is that they either have a very small number of sites that are all on Wikipedia (indicated by yellow on the figure) or they do not have any CS-WHL.
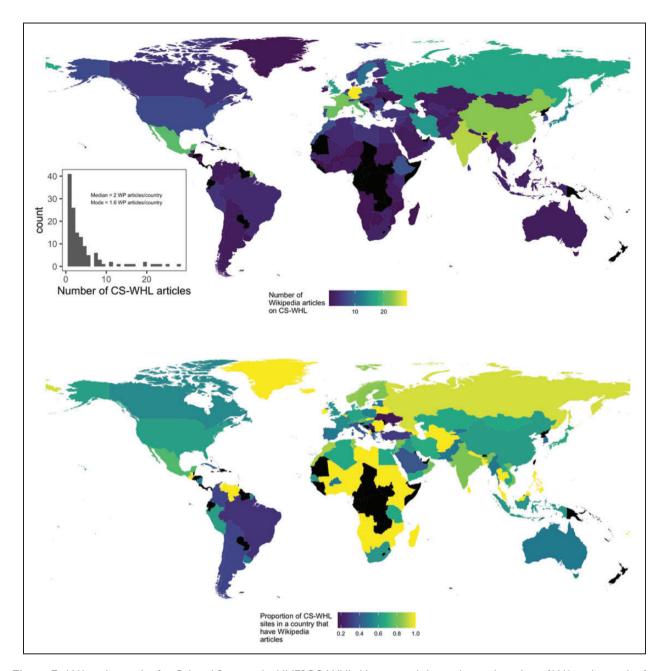
Figure 7. Wikipedia articles for Cultural Sites on the UNESCO WHL. Upper panel shows the total number of Wikipedia articles for CS-WHL sites per country. Lower panel is the proportion of all CS-WHL sites that have Wikipedia articles per country. Inset shows the distribution of numbers of CS-WHL articles per country. Countries colored black have no CS-WHL with Wikipedia articles at the time of writing. Map data from naturalearthdata.com.

The distribution of editor locations shown in Figure 8 was determined by geolocating the IP addresses attached to each anonymous edit on all CS-WHL articles (only anonymous edits contain IP addresses, edits made by registered Wikipedia users contain no information suitable for geolocation). This figure shows the flow of edits, with the arrows starting at the country where the editor was located when they made their edit, and ending at the country where the

CS-WHL site that they are editing is located. The inset plot on Figure 8 shows that this visualization accounts for a relatively small proportion of all edits to CS-WHL articles. Around 20% of edits of most CS-WHL articles are anonymous, a higher proportion than our random sample. Nevertheless, there are 79,077 anonymous edits, and so, this sample likely has some informational value in characterizing the spatial distribution of edits. The most striking detail in
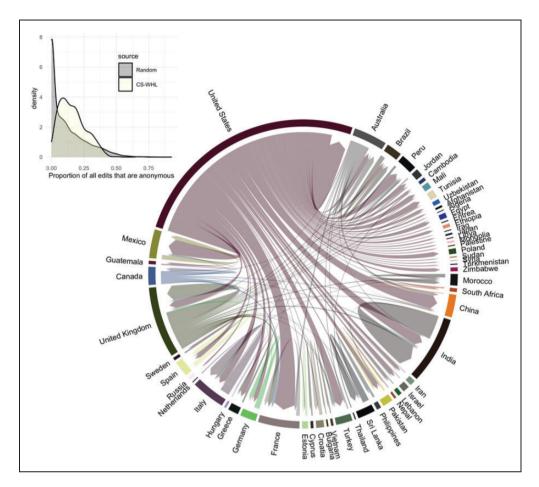
Figure 8. Flow of edits by location on Wikipedia for articles about Cultural Sites on the UNESCO WHL. The arrows indicate the country where the editor is location (arrow nock or start) and the location of the CS-WHL site that they are editing the article of (arrow point or end). Editor locations were determined by geolocating the IP addresses associated with anonymous edits. Inset shows the distribution of proportions of edits per article that are anonymous. Non-anonymous edits are identified by Wikipedia usernames, rather than IP addresses, and cannot be geolocated.

Figure 8 is the large proportion of edits that originate in the United States and that the vast majority of these are edits of articles about CS-WHL sites located elsewhere in the world. The country with the next largest proportion, the United Kingdom, is less than half of the United States, and a much greater proportion of edits originating in the UK are on articles about CS-WHL sites located in the UK compared to the US. After the United Kingdom is India, and nearly half of those edits are on CS-WHL sites located in the same country. Generally, articles about CS-WHL sites receive edits from editors located elsewhere, mostly, the US and other English-speaking countries.

Temporal patterns

Time series analysis of edits on articles about CS-WHL sites helps us identify relationships between activity on Wikipedia articles and external events related to the sites. Figure 9 shows that most articles about CS-

WHL sites are much older than articles in our random sample, with most CS-WHL articles created 5–10 years after Wikipedia first appeared in 2001. This rapid addition of articles early in the life of Wikipedia supports our earlier observation that articles about CS-WHL sites are more central to the encyclopedia and considered more worthy of inclusion than many other types of articles. The distribution of years between the date of a site's inscription on the WHL and the date of the creation of that site's Wikipedia article has a distinctive trimodal shape. The first two modes reflect the relatively large number of sites inscribed on the WHL in 1983, and the early 1990s, well before Wikipedia was created in 2001. The third mode at the zero point on the horizontal axis corresponds to CS-WHL sites that were inscribed after the creation of Wikipedia and has Wikipedia articles that were created in or around the same year the CS-WHL site was inscribed on the list. Only a relatively small proportion of CS-WHL sites had Wikipedia articles created about
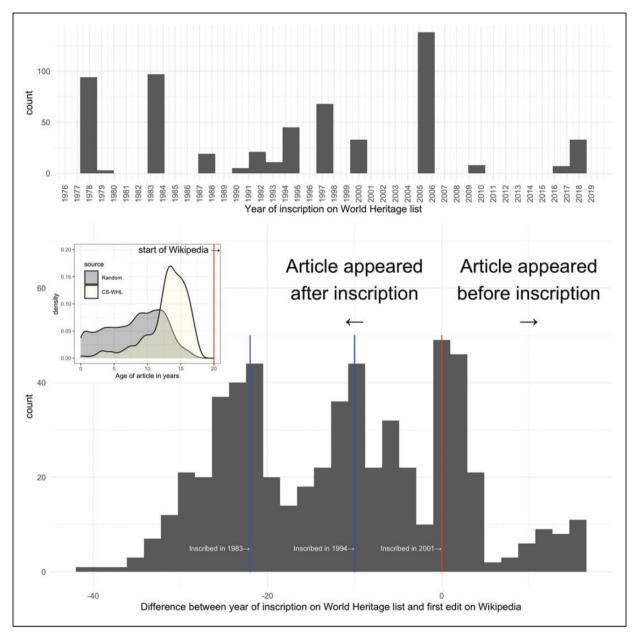
Figure 9. Article editing relative to WHL inscription date. Upper panel is a histogram of WHL inscription years. Lower panel is a histogram of duration between inscription and the appearance of the Wikipedia article. Inset shows the distribution of CS-WHL article ages and ages of random sites.

them before they were inscribed on the WHL because the majority of sites were inscribed before Wikipedia was created.

The time series of editing activity at individual sites are highly variable and display complex patterns (Figure 10). Timbuktu shows spikes of editing activity right after the rebel attacks in mid June 2012, with descriptions of the attacks added to the text of the article at that time. However, the tallest peak in editing activity on the Timbuktu occurs slightly earlier, in April 2012, with numerous edits concentrated on

expanding the sections on local prehistory. A detailed analysis of all the article-specific time series is beyond the scope of this paper, but we can see several patterns indicating important relationships between editorial activities and WHL listing. For example, one pattern is of editing activity spiking sharply at the time of WHL inscription (e.g. Tyre, Masada), a second pattern is editorial activity peaking at the time of WHL inscription and slowly decaying to a baseline level (e.g. Carthage), a third pattern is editorial activity that is highest shortly after the time of WHL inscription
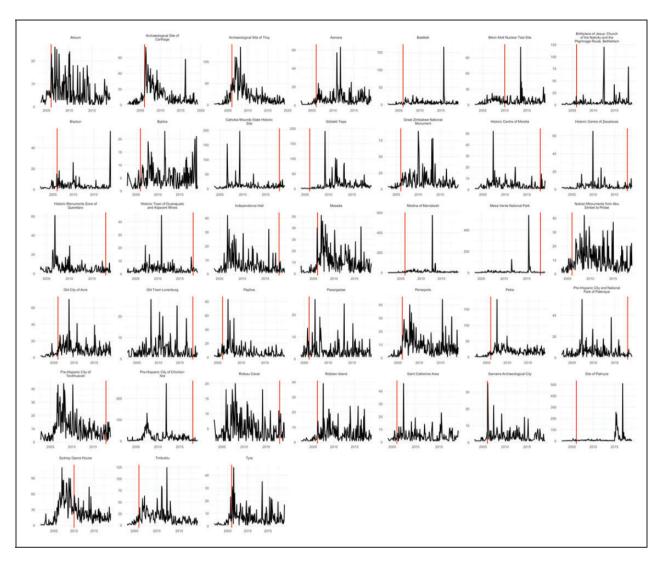
Figure 10. Time series of editing activity for CS-WHL articles for all sites inscribed after Wikipedia began in 2001 and with a minimum of 150 edits in total. Red vertical line indicates the time of inscription for each site.

(e.g. Troy, Masada), and finally many sites where editorial activity that shows no signal at all at the time of WHL inscription (Figure 10).

Discussion

Our results establish the English-language Wikipedia as an online community actively engaging with cultural heritage (Pentzold et al., 2017), especially cultural sites on the UNESCO WHL. We have defined Wikipedia as an online peer production community with technical and cultural qualities that make it distinct from social media services such as Twitter and Facebook. On Wikipedia, the editing mechanics, bots, reverts, and talk pages work together to enable and limit discourse and debate about CS-WHL. The goal-driven culture of Wikipedia editors and the policies that guide the production of articles strongly constrain heritage discourse, such that exchanges of different perspectives are rarely found in articles themselves. However, we have identified conflict and tension on the article talk pages, sometimes in technical and highly specific discussions, and sometimes in overtly hostile exchanges between editors. Our results show how the technicity of Wikipedia shapes how people engage with elements of the past and attribute social and cultural meanings on Wikipedia. What we have found resembles "contingent collaborations" and "productive frictions" (Tsing, 2011) where local, individual actions of knowledge creation are circumscribed by forces and structures that encompass their expression but also give meaning to their interactions.

Predictably, many of the Wikipedia articles about CS-WHL are central to the encyclopedia project. Many of them were among the first articles written for the encyclopedia. They are more frequently

viewed, are more frequently linked to by other articles, are longer, and more intensively edited by a greater variety of editors than the average article. These results indicate the high value that Wikipedia users place on CS-WHL, consistent with the goal of the UNESCO World Heritage Committee, to maintain a list that reflects the world's cultural diversity of outstanding universal value. Similarly, we found that the effort and attention sites receive on Wikipedia is highly unequal, reflecting the fairly narrow demographic traits of the majority of Wikipedia users and editors, which is further exacerbated by the Eurocentric value system of the World Heritage Committee (Cleere, 2003; Smith, 2006: 98). For example, CS-WHL in countries in the Global South are generally under-represented on Wikipedia, a geographical tension that is also evident in the WHL (Brumann, 2014). Articles about CS-WHL in the Global South are also shorter and have received less attention in the form of page views and inward Wikilinks.

The most striking spatial pattern in our results is the high proportion of anonymous edits (as the only type of edit that can be geolocated) that come from the United States on articles about CS-WHL sites, regardless of where in the world, the CS-WHL site is located. These anonymous editors of articles about CS-WHL are generally not located in the same country as the site they are editing. Finding the United States at the first ranked location for anonymous edits is perhaps not surprising, as it has the world's largest number of English speakers. We might similarly expect that the French language Wikipedia (2.3m articles, 173m edits, 3.9m users) would have more edits originating in France or francophone countries on articles about CS-WHL sites in any part of the world. We have limited the scope of our initial foray into this topic to the English language Wikipedia because it is the largest by a considerable margin (6.2m articles, 998m edits, 40m users, Wikipedia contributors, 2020). The narrow linguistic focus of our study is an important limitation, and question of whether or not our findings generalize to non-English-language Wikipedias should be a priority for future research.

Although the dominance of the United States in the spatial patterns may be unsurprising, our data show some patterns that cannot be fully explained by language dominance alone. After the United States (297m speakers), the top five countries for numbers of English speakers are India (238m), Nigeria (104m), United Kingdom (60m), Philippines (50m), and Germany (45m) (David M. Eberhard and Fennig, 2020). However, our spatial data show that the countries that originate the largest numbers of edits to articles about foreign CS-WHL sites (i.e. sites not located in the country where the edits where made)

are the United States, United Kingdom, Australia, and Canada. Although many edits originate from India, they are almost entirely of articles about CS-WHL sites also in India.

The United States, United Kingdom, Australia, and Canada are an important group because they are collectively known as the core Anglosphere countries, sharing a common heritage as former colonies of the British Empire (Richards, 2019). Vucetic (2011) defines the Anglosphere as a "distinct international, transnational, civilisational, and imperial entity within the global society." The dominance of editors of CS-WHL articles located in the Anglosphere indicates an extension of this imperial project into the organization of World Heritage information on the English-language Wikipedia. The prominence of Anglosphere editors on CS-WHL articles represents a digital colonization of World Heritage information on Wikipedia, where heritage is interpreted and communicated by people who are not part of the descendant communities traditionally associated with the site. Although we could only geolocate about 20% of edits, our spatial patterns resemble those found by Mandiberg (2020) who examined 884m edits to the English-language Wikipedia and found that, at the country level, the "five largest contributors were part of what once was the British Empire, and account for nearly 75 percent of all editors."

This is an important observation because it is these editors that control the production of knowledge about CS-WHL articles, which has several implications for understanding how information about heritage is curated. First is that the choices of these editors define community identity, that is, the community of Wikipedia editors, and also the broader community of Internet users that read Wikipedia, by indicating which sites are important and visible, and which are not, by contributing to a digital "authorized heritage discourse" (Smith and Waterton, 2012) that prioritizes their own self-interests (cf. the poor representation of CS-WHL sites located in the Global South). Second, traditional communities associated with CS-WHL have little control over how their own heritage is represented on the English-language Wikipedia. We see this spatial pattern on Wikipedia as part of a trend of globalization of heritage, where places that had traditionally contributed to local or regional identities are turned into sanitized playgrounds for rich tourists (Bernbeck and Pollock, 2004). Wikipedia may be seen as virtual space for Anglosphere citizens to construct a universal heritage that reflects their neo-liberal cultural values and validates their identities, perhaps reflecting a cultural rhetoric of "heritage populism" (Reynié, 2016). In this way, the representation of heritage on Wikipedia recapitulates a long-standing tradition of colonial

subordination of one group's heritage and identity by a more powerful and dominant group.

This colonial logic may help to explain our findings of minimal conflicts surrounding the production of CS-WHL articles. We found that bot activity, numbers of revert edits, and edits relating to vandalism for CS-WHL articles were comparable to or less than other sites. Most of this activity relates to playful vandalism and occurred on uncontroversial CS-WHL sites such as the Sydney Opera House and the Tower of London. Controversial sites, identified by the presence of military or diplomatic conflicts, were not prominent in these metrics. Our analysis of the talk pages of CS-WHL articles shows that they are often longer than a typical Wikipedia article and are the location where conflict is most evident. While much of the conflict is highly technical and local about the policies and processes of editing Wikipedia, we did observe signs of the political and military conflict at CS-WHL played out on some of the talk pages, such as Mapungubwe and Preah Vihear. We interpret this as editors exploiting the talk pages for behind-the-scenes heritage activism. The technical and cultural constraints of writing directly on the article page mean that the talk page is the only suitable location for negotiating disputes whose eventual trace in the edit history of the article might be negligible.

Overall, we find that editors of CS-WHL articles appear conflict-averse, probably because many of them come from similar cultural, socio-economic, and educational backgrounds due to their Anglosphere origins. This means they share similar values about heritage that align with their shared goal of making an encyclopedia. The low levels of conflict show continuity between the printed encyclopedia project as a compendium of universal knowledge and the WHL as universal heritage. These projects share origins in the cultural elites of Western Europe, and their distinctive sense of what is universal that is linked to 19th-century nationalism and liberal modernity, vestiges of which are still evident in the modern political philosophy of the Anglosphere (Smith, 2006).

The results of our time series analyses reveal complex patterns in how heritage concerns are realized on Wikipedia. Some sites show spikes in Wikipedia editing activity that are directly related to the inscription of the site on the WHL or conflict events at the site. However, for other locations, the rhythms of editing activity seem unrelated to events directly related to the site. Like the talk pages, which have potential for close reading of heritage activism and disputes, close study of individual edit histories of CS-WHL articles may reveal further patterns of how people engage with heritage on Wikipedia.

Conclusion

The agency of people participating in heritage-related activities on Wikipedia is highly constrained by the technical characteristics of editing, bots, and consensus-valuing culture. Although there are some niches where editors can debate, concerns about state-like actors, violence and destruction, deal-making, etc. are minor. Typical conflicts about English-language Wikipedia CS-WHL articles are hyper-local and process-centered. There are also cases of editorial conflicts about substantive issues of specific CS-WHL, such as Mapungubwe and Preah Vihear, that include editors with a personal investment in the debate, if not the article. Spatial patterns in the edits to Wikipedia articles about CS-WHL show a kind of digital colonialism of World Heritage information, with most edits emanating from the Anglosphere, and generally poor coverage of CS-WHL in the Global South. Because of these technical and cultural characteristics of Wikipedia, its articles present a discourse that actively obscures the power relations that give rise to them and makes opaque debates about how heritage is involved in the production of identity and power (cf. Harvey, 2001).

That said, Wikipedia is valuable and important as a Big Data source of social information on heritage for several reasons. Although it lacks the velocity of other social media, such as Twitter, it preserves a distinctively complex record of human interactions with the past at simultaneously high scales, high resolution, and in highly structured ways, afforded by no other platform. As a non-profit entity, Wikipedia's data are freely available to researchers, unlike Twitter and Facebook. This has important ethical implications because it eliminates the need for academic collaboration with commercial platforms, whose business interests may not align with the researchers' ethics (Schroeder, 2014). Although free of these commercial interests, a different set of ethics is implied by the substantial administrative and technical complexity of Wikipedia. These structures serve to normalize and historize inequalities and manage threats to models of control and expertise about heritage (Harrison, 2015). As a context of heritage production, a resource of information on heritage, and itself an artifact and archive of contemporary digital heritage, Wikipedia is remarkable and holds substantial research potential. While the NPOV is a central policy for Wikipedia articles, representations and definitions of cultural heritage are never neutral in any discourse because they always come from some perspective (Lähdesmäki, 2019). Given the tremendous popularity and visibility of Wikipedia, it is important that future work on digital heritage in this context continues to critically

examine what is absent from, and uncontested in, Wikipedia, and who benefits and suffers from these absences.

Acknowledgements

Declaration of conflicting interests

Funding

ORCID iDs

Ben Marwick ⓘ https://orcid.org/0000-0001-7879-4531
Prema Smith ⓘ https://orcid.org/0000-0002-5485-359X

References

Adams J and Brückner H (2015) Wikipedia, sociology, and the promise and pitfalls of Big Data. *Big Data & Society*. Epub ahead of print 1 December 2015. https://doi.org/10.1177/2053951715614332

Baker F (2003) The red army graffiti in the Reichstag, Berlin: Politics of rock-art in a contemporary European urban landscape. In: Christopher C and George N (eds) *European Landscapes of Rock-Art*. London, UK: Routledge, pp.38–56.

Bernbeck R and Pollock S (2004) The political economy of archaeological practice and the production of heritage in the Middle East. In: Meskell L and Preucel RW (eds) *A Companion to Social Archaeology*, pp.335–352.

Bonacchi C, Altaweel M and Krzyzanska M (2018) The heritage of Brexit: Roles of the past in the construction of political identities through social media. *Journal of Social Archaeology* 18(2): 174–192.

Bonacchi C and Krzyzanska M (2019) Digital heritage research re-theorised: Ontologies and epistemologies in a world of Big Data. *International Journal of Heritage Studies* 25(12): 1235–1247.

Brioschi G (2017) *Discourses of heritage: UNESCO and the local community in Timbuktu*. PhD Thesis. Brussels School of International Studies, Belgium.

Brumann C (2014) Shifting tides of world-making in the UNESCO World Heritage convention: Cosmopolitanisms colliding. *Ethnic and Racial Studies* 37(12): 2176–2192.

Cantino A and Maxwell K (2017) SelectorGadget: Point and click CSS selectors. Available at: https://selectorgadget.com/ (accessed 12 May 2021).

Cao Y, et al. (2020) Analysis of Wikipedia pageviews to identify popular chemicals. In: Samuel A and Ramesh R (eds) *Reporters, Markers, Dyes, Nanoparticles, and Molecular Probes for Biomedical Applications XII*. Bellingham, WA: SPIE, p.112560I.

Cleere H (2003) The uneasy bedfellows: Universality and cultural heritage. In: Layton R, Stone P and Thomas J (eds) *Destruction and Conservation of Cultural Property*. London, UK: Routledge, pp.38–45.

Coningham R and Lewer N (1999) Paradise lost: The bombing of the temple of the tooth – A UNESCO World Heritage site in Sri Lanka. *Antiquity* 73(282): 857–866.

Eberhard DM, Simmons GF and Fennig CD (2020) *Ethnologue: Languages of the World*. 23rd ed. Dallas, TX: SIL International.

Falser M (2015) Representing heritage without territory – The Khmer rouge at the UNESCO in Paris during the 1980s and their political strategy for Angkor. In: *Cultural Heritage as Civilizing Mission*. Berlin, Germany: Springer, pp.225–249.

Ford H and Wajcman J (2017) "Anyone can edit", not everyone does: Wikipedia's infrastructure and the gender Gap. *Social Studies of Science* 47(4): 511–527.

Geiger RS (2009) The social roles of bots and assisted editing programs. In: Riehle D and Bruckman A (eds) *Proceedings of the 5th international symposium on wikis and open collaboration*, New York, NY: Association for Computing Machinery, pp.1–2.

Geiger RS (2014) Bots, bespoke, code and the materiality of software platforms. *Information, Communication & Society* 17(3): 342–356.

Geiger RS and Halfaker A (2017) Operationalizing conflict and cooperation between automated software agents in Wikipedia: A replication and expansion of 'even good bots fight. In: Nichols J (ed) *Proceedings of the ACM on human-computer interaction*, New York, NY: Association for Computing Machinery, pp.1–33.

Gornik V (2015) Can UNESCO do more to counter terrorists' destruction of World Heritage? *Present Pasts* 6(1): p.Art. 5. http://doi.org/10.5334/pp.61

Graham M, et al. (2014) Uneven geographies of user-generated information: Patterns of increasing informational poverty. *Annals of the Association of American Geographers* 104(4): 746–764.

Harrison R (2013) Forgetting to remember, remembering to forget: Late modern heritage practices, sustainability and the "crisis" of accumulation of the past. *International Journal of Heritage Studies* 19(6): 579–595.

Harrison R (2015) Beyond "natural" and "cultural" heritage: Toward an ontological politics of heritage in the age of Anthropocene. *Heritage & Society* 8(1): 24–42.

Harvey DC (2001) Heritage pasts and heritage presents: Temporality, meaning and the scope of heritage studies. *International Journal of Heritage Studies* 7(4): 319–338.

Hill B and Shaw A (2013) The Wikipedia gender gap revisited: Characterizing survey response bias with propensity

score estimation. *PLoS ONE* 8(6): e65782. https://doi.org/10.1371/journal.pone.0065782

Ho-Dac L-M, et al. (2016) French Wikipedia talk pages: Profiling and conflict detection. In: *4th conference on CMC and social media corpora for the humanities, September 2016, Ljubljana, Slovenia*.

Ho-Dac L-M, et al. (2017) Exploring Wikipedia talk pages for conflict detection. In: Fišer D and Beißwenger M (eds) *Investigating Computer-Mediated Communication: Corpus-Based Approaches to Language in the Digital World*, Ljubljana: Ljubljana University Press, pp.146–168.

Iba T, et al. (2010) Analyzing the creative editing behavior of Wikipedia editors: Through dynamic social network analysis. *Procedia-Social and Behavioral Sciences* 2(4): 6441–6456.

Johnson IL, et al. (2016) Not at home on the range: Peer production and the urban/rural divide. In: Kaye J, Druin A, Lampe C, Morris D and Hourcade JP (eds) *Proceedings of the 2016 CHI conference on human factors in computing systems*, New York, NY: Association for Computing Machinery, pp.13–25.

Keyes O, et al. (2020) Rgeolocate: IP address geolocation. Available at: https://CRAN.R-project.org/package=rgeolocate (accessed 12 May 2021).

Kietzmann JH, et al. (2011) Social media? Get serious! Understanding the functional building blocks of social media. *Business Horizons* 54(3): 241–251.

Kittur A, et al. (2007) He says, she says: Conflict and coordination in Wikipedia. In: Rosson MB and Gilmore D (eds) *Proceedings of the SIGCHI conference on human factors in computing systems*, New York, NY: Association for Computing Machinery, pp.453–462.

Kristiansen K (2014) Towards a new paradigm. The third science revolution and its possible consequences in archaeology. *Current Swedish Archaeology* 22(4): 11–71.

Lähdesmäki T (2019) Conflicts and reconciliation in the post-millennial heritage-policy discourses of the council of Europe and the European Union. In: Lähdesmäki T, Passerini L, Kaasik-Krogerus S and van Huis I (eds) *Dissonant Heritages and Memories in Contemporary Europe*. Berlin, Germany: Springer, pp.25–50.

Locard H (2015) The myth of Angkor as an essential component of the Khmer rouge Utopia. In: Falser M (ed) *Cultural Heritage as Civilizing Mission*. Berlin, Germany: Springer, pp.201–222.

Mandiberg M (2020) Mapping Wikipedia. *The Atlantic*. Available at: https://www.theatlantic.com/technology/archive/2020/02/where-wikipedias-editors-are-where-they-arent-and-why/605023/ (accessed 12 May 2021).

Marwick B (2017) Computational reproducibility in archaeological research: Basic principles and a case study of their Implementation. *Journal of Archaeological Method and Theory* 24(2): 424–450.

Marwick B, Boettiger C and Mullen L (2018) Packaging data analytical work reproducibly using r (and friends). *The American Statistician* 72(1): 80–88.

McIver DJ and Brownstein JS (2014) Wikipedia usage estimates prevalence of influenza-like illness in the United States in near real-time. *PLoS Computational Biology* 10(4): e1003581. https://doi.org/10.1371/journal.pcbi.1003581

Meskell L (2011) From Paris to Pontdrift: UNESCO meetings, Mapungubwe and mining. *The South African Archaeological Bulletin* 66(194): 149–156.

Meskell L (2012) The rush to inscribe: Reflections on the 35th session of the World Heritage committee, UNESCO Paris, 2011. *Journal of Field Archaeology* 37(2): 145–151.

Meskell L (2015) Gridlock: UNESCO, global conflict and failed ambitions. *World Archaeology* 47(2): 225–238.

Meskell L (2016) World Heritage and WikiLeaks: Territory, trade and temples on the Thai-Cambodian border. *Current Anthropology* 1: 57.

Meskell L (2018) *A Future in Ruins: UNESCO, World Heritage, and the Dream of Peace*. Oxford, UK: Oxford University Press.

Niederer S and Van Dijck J (2010) Wisdom of the crowd or technicity of content? Wikipedia as a sociotechnical system. *New Media & Society* 12(8): 1368–1387.

Pentzold C, et al. (2017) Digging Wikipedia: The online encyclopedia as a digital cultural heritage gateway and site. *Journal on Computing and Cultural Heritage (JOCCH)* 10(1): 1–19.

Pfanzelter E (2015) At the crossroads with public history: Mediating the holocaust on the Internet. *Holocaust Studies* 21(4): 250–271.

Priedhorsky R, et al. (2007) Creating, destroying, and restoring value in Wikipedia. In: Gross T and Inkpen K (eds) *Proceedings of the 2007 international ACM conference on supporting group work*, New York, NY: Association for Computing Machinery, pp.259–268.

R Core Team (2020) R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Available at: https://www.R-project.org/.

Reynié D (2016) The specter haunting Europe: "Heritage populism" and France's national front. *Journal of Democracy* 27(4): 47–57.

Richards E (2019) British emigrants and the making of the Anglosphere: Some observations and a case study. In: Payton P and Varnava A (eds) Australia, Migration and Empire: Immigrants in a Globalised World. Berlin, Germany: Springer, pp.13–43.

Roll U, et al. (2016) Using Wikipedia page views to explore the cultural importance of global reptiles. *Biological Conservation* 204: 42–50.

Schneider J, Passant A and Breslin JG (2012) A content analysis: How Wikipedia talk pages are used. In: *WebSci2010, web science conference, Raleigh, NC*.

Schroeder R (2014) Big Data and the brave new world of social media Research. *Big Data & Society* 1(2): 1–11. https://doi.org/10.1177/2053951714563194

Shaw A and Hargittai E (2018) The pipeline of online participation inequalities: The case of Wikipedia editing. *Journal of Communication* 68(1): 143–168.

Shaw A and Hill BM (2014) Laboratories of oligarchy? How the iron law extends to peer Production. *Journal of Communication* 64(2): 215–238.

Smith L (2006) *Uses of Heritage*. London, UK: Routledge.

Smith L and Waterton E (2012) Constrained by commonsense: The authorized heritage discourse in contemporary debates. In: Skeates R, McDavid C and Carman J (eds) *The Oxford Handbook of Public Archaeology*, pp.153–171.

Sothirak P (2013) Cambodia's border conflict with Thailand. *Southeast Asian Affairs* 1: 87–100.

Suh B, et al. (2007) Us vs. them: Understanding social dynamics in Wikipedia with revert graph visualizations. In: *2007 IEEE symposium on visual analytics science and technology*, NW Washington, DC: IEEE Computer Society, 1730 Massachusetts Ave., pp.163–170.

Tsing AL (2011) *Friction: An Ethnography of Global Connection*. Princeton, NJ: Princeton University Press.

Tsvetkova M, et al. (2017) Even good bots fight: The case of Wikipedia'. *PloS One* 2: 12.

Vucetic S (2011) *The Anglosphere: A Genealogy of a Racialized Identity in International Relations*. Redwood City, CA: Stanford University Press.

Weltevrede E and Borra E (2016) Platform affordances and data practices: The value of dispute on Wikipedia. *Big Data & Society*. Epub ahead of print 1 June 2016. https://doi.org/10.1177/2053951716653418

Wickham H (2019) *Rvest: Easily harvest (scrape) web pages*. Available at: https://CRAN.R-project.org/package = rvest (accessed 12 May 2021).

Wikipedia contributors (2020) List of Wikipedias – Meta. Available at: https://meta.wikimedia.org/wiki/List_of_Wikipedias (accessed 12 May 2021).

Wilkinson DM (2008) Strong regularities in online peer production. In: Fortnow L, Riedl J and Sandholm T (eds) *Proceedings of the 9th ACM conference on electronic commerce*, New York, NY: Association for Computing Machinery, pp.302–309.

Wolniewicz-Slomka D (2016) Framing the holocaust in popular knowledge: 3 articles about the holocaust in English, Hebrew and Polish Wikipedia. *Adeptus* (8): 29–49.

Yasseri T, et al. (2012) Dynamics of conflicts in Wikipedia. *PloS One* 7(6): e38869.

Zheng L, et al. (2019) The roles bots play in Wikipedia. *Proceedings of the ACM on Human-Computer Interaction* 3: 1–20.