

Perspectives

This collection of essays was developed by the Audiences team as a multi-month collaborative exercise. These point of view pieces were developed through conversations with subject matter experts inside the Movement and the Foundation, and incorporated as much secondary research and reference points as possible. These perspectives reflect the collective opinion of this team as we embark on medium-term planning process.



Trust

Experience

Scale

Augmentation

Culture

Tools

Trust

This collection of essays explores the role of Trust in relation to the Movement, the content and the Foundation. These positions are intended to stimulate discussion and inform the development of new capabilities.



Published
January 2019

Authors

Trevor Bolliger
Alex Hollender
Jon Katz
Margeigh Novotny

Contributors

David A.M. Goldberg
Danny Horn
Josh Minor
Jonathan Morgan
Toby Negrin
Jacob Orlowitz
Nirzar Pangarkar
Alex Stinson
Ben Vershbow



Trust is a fundamental building block of community and culture, and is therefore critical to consider as we define next stage strategy toward 2030 goals. In commercial contexts, mutual interest and mutual concern mandate some level of trust between individuals or entities engaged in trade. In civic contexts, trust is often a by-product or outcome of the process by which collective good is determined—and a baseline criteria in selecting who can represent or advocate for that good. In interpersonal contexts, trust is the social lubricant that allows us to progress, step by step, from strangers to acquaintances to intimates. In each of these contexts (all of which have parallels in the open knowledge ecosystem) the rituals of trustworthy engagement share the same basic characteristics.

Parties in trustworthy relationships...

Take an interest in or are curious about one another

Act in good faith and assume good faith

Share relevant information voluntarily

Take information that is shared at face value

Behave respectfully toward one another

Strive for equitable value exchange

Are reliable and honor their agreements

Sometimes share mutually beneficial objectives

These characteristics, and the relative weight placed on each, are manifest at cultural, social and transactional levels in ways that have the potential to significantly impact our Movement and its goals. And, as a global organization, it's important to understand how our predominantly northern/western model of trust-building may be reflected at each level.

In the papers that follow we explore three

vectors of trust: reliability of content, loci of accountability and transparency of the Foundation and the Movement.

Investment in Trust means the Movement, the Foundation and every community must cultivate awareness and facility with the cultural, social and transactional aspects of trust-building interaction. That is, in order to empower and activate all communities to take part in the creation of knowledge we must develop an actionable model for building trust and strengthening alliances - one that is effective across projects and between cultures. This will require awareness of our own trust-building paradigms, and a fluency with reading and sending trust-signals in our governance practices, our community norms and policies, and via our content. We will need to consciously define and embody trustworthy practices at every touchpoint of the Movement.

04

Notes

1 Even in illicit markets, trust is a factor in determining who will do business with whom.

2 Reaching our 2030 objectives will require effective and dynamic partnership with sister-organizations and for-profit partners (trade). The “civic” dimension of the movement must include support for a multiplicity of interest-groups within the Movement, and the balancing of leaders’ voices. It will require these civic leaders to exhibit trustworthy practices and exemplary judgement (organizers, content curators and moderators fall into this “civic” dimension). Their objectives must be understood and communicated transparently, and leaders of interest groups within the Movement must have the ears of other leaders. On the interpersonal level, individual users must be able to move from anonymous consumer to engaged contributor through a set of mutually agreeable and understandable gates; and once there, experience only respectful interactions with others.

3 At a minimum, they “trust but verify”

4 But isn’t it necessary for parties in a trusted relationship to have some “mutually beneficial objective”? No, it may not be necessary. It’s possible for one party to take an interest in another, and to demonstrate every other characteristic of a trusting relationship, but in the end see no mutually beneficial objective to work toward together, at a particular time. This distinction marks the difference between allies and partners. When Trust is present, both allies and partners are able to work constructively to support one another, whether or not they have an immediate mutual objective (and while cultivating allies is a reward in itself, allies can become partners later).

01

Reliability

Wikipedia is one of the most trusted sources of knowledge in the world.[1] In one telling example, a 2013 report determined it to be the top source of healthcare information for both doctors and patients in the US. [2] The importance of this trust has intensified in the last few years, as the world has polarized further; new, niche media sites have proliferated; blatant lies are told by politicians on television; and “fake news” spreads quickly. It is increasingly hard to tell what is true and what is not. At this time, Wikipedia is increasingly seen as a uniquely trustworthy source, a reputation demonstrated by the reliance on Wikipedia by large media platforms to fact-check assertions made on their platforms. [3]

At the same time, there is room for improvement and reason to invest in reader trust. While Wikipedia is trusted by many, it is also famous for misinformation.[4] Wikipedia has historically been regarded with unease and contempt by traditional sources of information (teachers, journalists) [5] and it is easy to imagine how a high profile scandal might undermine public faith in Wikipedia for years to come. A recent poll of Wikipedia users found that the number one request of respondents was “More trustworthy

content.”[6] Stories of vandalism frequently make headlines, and teachers around the world tell their students not to trust Wikipedia. Some of the many reasons not to trust Wikipedia content are outlined in a very thorough Quora answer by Andreas Kolbe [7] (a longtime community member and Wikipedia documenter) for anyone who is interested.

Challenges

One of the challenges readers have with Wikipedia is determining what to trust and what not to trust. Multiple studies have shown that readers tend to use proxies like article length, pictures and grammar along with citations as proxies for trustworthiness.[8] [9] [10] When we asked Donors what we should improve, we received many responses like these:

“A way to measure/judge how reliable the information I’m reading is. A truth-meter or fact-meter. I’m told not to trust the information I read on Wikipedia so many times I don’t consider it factual until I read it published somewhere else.”

“As a reader it would be hard to know what is fake information. But i trust Wikipedia, never even checking the references unless I need them.”[11]

Given the wide range in quality of articles on Wikipedia,[12] this lack of signal within Wikipedia represents unappreciated value (in the case of good articles) and avoidable liability (in the case of bad ones).

Admittedly, it is not clear how distrust impacts Wikipedia: a recent study found that as students move through the educational system, they are more likely to be told not to use Wikipedia and yet remarkably use Wikipedia more and more. [13] However, distrust does impact how our content is leveraged in everyday life. While doctors may use Wikipedia extensively, they don’t rely on it. This UK doctor sums it up nicely: “I use Wikipedia to gain a quick overview of a subject/topic that I am unfamiliar with or to jolt my memory of a subject. I would never base management or treatment of a patient I find there – for that I use my own knowledge, hospital protocols/guidelines, textbooks and advice from colleagues.” [14]

The impact of greater trust in this case, might only be seen in offline behavior.

Room for improvement

Marginalized Communities

As we consider the next 2 billion people who are coming online, we need to consider the trust they will have in Wikipedia. If someone goes to Wikipedia and doesn’t see their own experience reflected there, how can they truly trust it?

In many ways, the citations tools that create trust among the majority by relying on and mirroring traditional structures of power and authority, undermine the trust by the minority. Along these lines, Wikipedia is facing legitimate criticism from groups who feel that Wikipedia does not represent their truth. Examples include: WhoseKnowledge, WomenInRed, WikiMujeres, Wikigap, Afrocrowd, Art & Feminism.

The efforts to improve trust globally and among marginalized communities will likely overlap in some areas (showing number of editors) and deviate in others (loosening reliable sources criteria).

Investing in Trust

Wikipedia is currently trusted enough for broad, global usage and increasing trust further has diminishing returns. Yet, it is also self-evident that maintaining some minimum degree of trust is essential for meaningful usage. It is also true that giving users greater clarity over what they can trust on Wikipedia would allow people to use Wikipedia more effectively. In this respect, we hope to increase trust in Wikipedia overall, by, at times, lowering reader trust in a particular piece of content.

For this reason, we think that “Reader Trust” should be invested in as an asset that increases our impact and requires defensive support.

Specific Trust Concerns

Citations need support

We need to invest in trust, by improving our citation infrastructure. Verification via citation is one of the cornerstones of our trust and there are significant opportunities to improve our citation experience. However, citations are neither sufficient nor accessible signals of trustworthiness and there is more room for signalling to readers in other ways how much to trust a particular piece of content.

Citations are a valuable measure for accountability that we have on Wikipedia. However, they currently suffer from several limitations:

- User awareness and understanding of citations
- Unable to serve function of verifiability [15]
 - Incomplete coverage
 - Incomplete archiving
 - Paywalls
 - Language barriers
- Once verified, hard to evaluate
 - A proliferation of new media outlets to vet
 - An proliferation of real-looking, fake media outlets
 - Cross-article analysis. (Wikicite, in particular, looks to open this up)
- Dishonest or incorrect use of citations by editors
- Existing reliable source definitions limit information

Some of these challenges are harder to solve than others, but we should invest in solving or accounting for them. It is worth noting that two other popular mechanisms are currently employed signal content trustworthiness:

article quality grades, and hatnotes (page issue templates). Both of these are manually derived, manually updated and subject to both gaps in either definition or usage. Hatnotes, are also subject to over-use. In fact, the term “tag:bombing” [16] was created to describe the destructive power of tag bombing. Here [17] is a famous example.

Specific solutions are recommended in the last section.

Verifiability is Insufficient

However, citations are not the only tool at our disposal, and, even at their best, represent an incomplete picture of content trustworthiness. For instance, one of the most common forms of deception is selective truth. If the article on Barack Obama only had only one sentence that said: Barack Obama is the author of the bestselling book “Dreams of my father”, the article would be tantamount to a lie. Similarly, the following statement, while factually accurate would be highly misleading:

While Barak Hussein Obama claims to have been born in the United States, there are many prominent figures who dispute this claim. The current president of the United States publicly stated his belief that Mr. Obama founded ISIS.

As another example, found on the Wikipedia article on Wikipedia’s reliability, a 2014 exploration found that FDA drug warnings were not sufficiently updated on Wikipedia in sometimes stark circumstances.[18]

There is also the concern that most users do not actually open the citations and are therefore apt to falsehoods that are planted there either erroneously or intentionally. Planted falsehoods are the sort of behavior that users (anecdotally) seem most concerned about. Its okay if a 7th grader learns that the Eiffel tower was built in 1830 (maybe Encyclopedia Britannica has a typo), but not okay if they learn that it was built by a troupe of runaway mimes.

As is often the case, the deficiencies here are maximized in the case of promoting knowledge equity. All too often, content, well-cited or not, takes the perspective of the traditionally powerful, Western white world. In doing so, it undermines the trust of marginalized communities.

Siko Bouterse, of the group Whose Knowledge,[19] references the following example from Wikipedia to show how existing power structures undermine trust in Wikipedia. During an editathon at Wikicon San Diego in 2016, with individuals from the Kumeyaay, a tribe indigenous to California, the California Gold Rush page was edited. While the goldrush was many things, it was considered a genocide by California's indigenous people. One of the first changes they wanted to make was to change the picture on the page about indigenous people, from one of an indigenous person attacking a settlement, to one in which the Western settlers are firing their arms.



BEFORE User:NorCalHistory CC BY-SA 3.0 [20]



AFTER Illustration by John Ross Browne Public Domain [21]

As you can see dramatically here, the truth of our content goes well beyond whether or not something is cited, but the same solutions as above apply.

Policies Can Negatively Impact Trust

In 2012, the celebrated author, Philip Roth, noticed that an article about his 2000 novel “The Human Stain” had a verifiable fact that was nonetheless misleading. It said the novel was “allegedly inspired by the life of the writer Anatole Broyard.” This fact was cited. When Roth contacted administrators at Wikipedia to tell them this was incorrect and what really inspired the novel, his correction was rejected. His interlocutor was told: *I understand your point that the author is the greatest authority on their own work, but we require secondary sources.*[22]

Roth wrote about this experience in the New Yorker and thereby publicly illustrated how the focus on reputable sources can actually decrease trust in our content.

Again, these drawbacks have bigger impact on marginalized communities. While Philip Roth, being famous, was able to publish this piece in the New Yorker, that option is not available to most people. Indeed the problem Philip Roth illustrated, of overly restrictive “reliable source” policy is greatly exacerbated when we consider regions and peoples whose access to and coverage by traditional media forms is limited.

Priorities and Approaches

Given the value of investing in trust for our users and the danger to Wikipedia’s credibility if we don’t, the following areas of intervention are suggested in order of suggested priority.

Signals of Trustworthiness

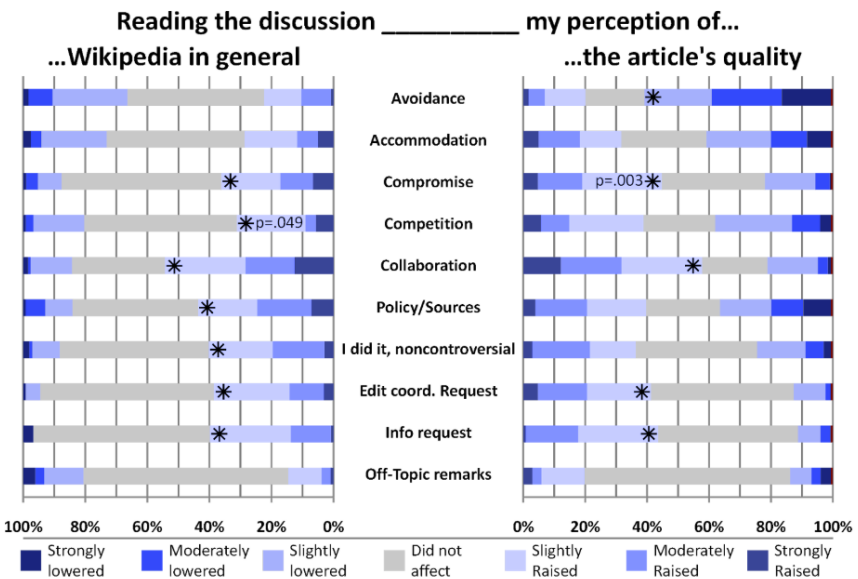
In addition to citations, we need to bring to the forefront other tools to give users the signals they need to make educated assessments of how much to trust individual pieces of Wikipedia content and thereby, trust Wikipedia more. This will

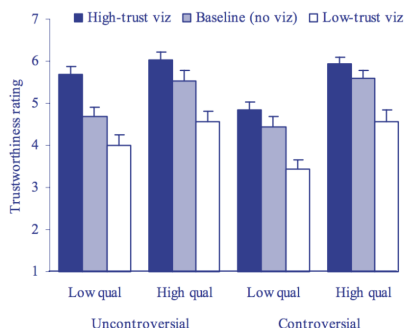
serve all readers and will not further reinforce traditional power structures. Right now we only use two tools: citations and page issue templates, and as described above, these are incomplete.

We can do this by exposing information about the article’s development and current status in easy-to-understand ways. On one end of the spectrum is access to a history page and talk page--this requires a great deal of time and head-scratching. On the other end of the spectrum is a trust score-- this is troubling because it is bound to be inaccurate and undermines a readers’ critical thinking skills. The recommendation is for visualizations and statistics that readers can use to form their own opinions.

Educating readers about how Wikipedia works is going to require great skill. One study showed that learning more about our process (seeing an article talk page) lead to a drop in trust, independent of article quality.[23]

Another study showed that trust in the same





Wikipedia article could go up or down, depending on how the material was presented. [24]

Here is an example of a visualization that led to higher trust:



Another small study showed that having students make edits to Wikipedia had a dramatic increase in their qualitative assessment of Wikipedia's value and accuracy.[25] A final study of Wikipedia found that a particular set of visualizations helped readers better distinguish between low credibility and high-credibility search results, but had little result on web-pages.[26] Interestingly, in the first study, users reported that seeing the discussion had changed their opinion. So it is clear that care will have to be taken when employing and testing these visualizations.[27]

Biased or Incomplete information

Priority 1: Show readers indicators accessible information (visualizations) of how biased or incomplete an article might be.

Solutions:

- Showing Wikipedia readers how many authors contributed to content
- A ratio of article depth to article importance (links in, centrality)
- A signal of how active the discussion is
- A signal of editor diversity (based on editor behavior patterns or geo)

Vandalism or Misuse of citations

Give readers digestible information about how likely a piece of content is to be valid.

Solutions:

- Showing readers how many "watchlists" the article appears on
- How recently it was edited, the ORES score of the last edit
- A blame map of which there are already more than a dozen [28]
- How many times the page was viewed since the last edit
- Providing users with annotated screenshots of the content from the cited source material

Priority is given to this set of solutions, because there is rich data that is currently invisible or inaccessible to most users.

Priority 2. Include the perspective of marginalized communities

In order to reflect the “Truth” of marginalized communities, we cannot continue to rely solely on the signals from traditional sources of power and authority. More specifically, we need to very carefully move away from a sole-reliance on published secondary sources by a handful of institutions.

Narrow Reliable Sources Policy

No clear product interventions exist for broadening reliable source policies. The heavy lifting might need to come through community activism and supported by product interventions (such as color-coding citations based on what kind of source they were). Dangers to this project include undermining our system of trust by opening up massive loopholes in the fact-checking process. However, steps can be made to mitigate these concerns.

Priority 3. Promote media literacy

We should then invest in teaching users critical media literacy: about how articles are created, the protections in place and how they can be flawed.

Solutions:

- UX features like those mentioned above
- Making citations more visible
- Online resources for journalists, students and educators (teachers, librarians, other professionals)
- Outreach programs for journalists, students and educators (teachers, librarians, other professionals) teachers

Priority 4. Address gaps in citation verification infrastructure

We also should continue to invest in making citations easier to create and verify. As mentioned

above, the current verification system is plagued by several issues including:

- Unable to serve function of verifiability:
 - Incomplete archiving
 - Paywalls
 - Language barriers
- Once verified, hard to evaluate
 - A proliferation of new media outlets to vet
 - An proliferation of real-looking, fake media outlets
 - Cross-article analysis. (Wikicite, in particular, looks to open this up)

Citation Gaps

Solutions here are around making the source material easier to find and digest

Solutions

- Leveraging wikidata to multiply fact-checking across languages
- Anti-paywall, free academic knowledge advocacy
- Reliable source database--allowing for validation of sources and quick cleanup when/if a source is deemed unreliable
- Reference material translation
- AI fact streams (primer.ai),
- Providing users with annotated screenshots of the content from the cited source material

02

Accountability

A key factor in building and maintaining trust is holding ourselves accountable for the human dynamics in our Movement and on our platform. That is, in the coming years, we must not only recognize the various ways that bias undermines the growth of our Community, we must also actively cultivate trust-building practices across cultures, within communities and between individuals. This is a strategic imperative because only by examining our assumptions, behaviors and norms (as a mostly northern, mainly western, largely male movement) will we be capable of recognizing all possible forms of

knowledge, expertise and notability. To this end we must develop facility with navigating between the cultural, social and transactional factors that foster or undermine trust. Accountability must become a pillar of this socio-technical platform, built by the Movement, modeled by the Foundation, and maintained by each Community and contributor.

Accountable for Cultural Literacy

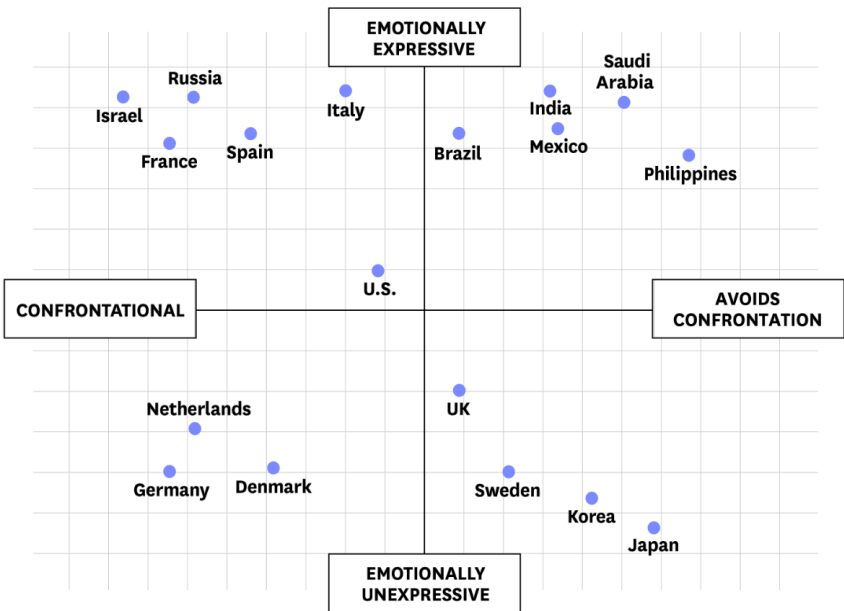
Paths to developing trust differ in different parts of the world, with some cultures placing early emphasis on the more cognitive aspects [1] (evidenced by a willingness to depend on the other based on demonstrations of competence, reliability and expertise), and others on the affective aspects (the degree to which either party feels comfortable sharing problems and aspirations; the willingness to be vulnerable to the other).[2] These tendencies are driven by cultural values and directly influence our ability to establish trust/find common ground as allies operating across cultures.

While direct debate is a paradigm we take for granted in a western/northern intellectual tradition, in a cultural perspective that values harmonious interaction, a passionate debater might be read as less trustworthy.[3] So when we base trust judgements on the signals we bring from our own cultural contexts, it's easy

to misread the intentions and motives of others (bias).

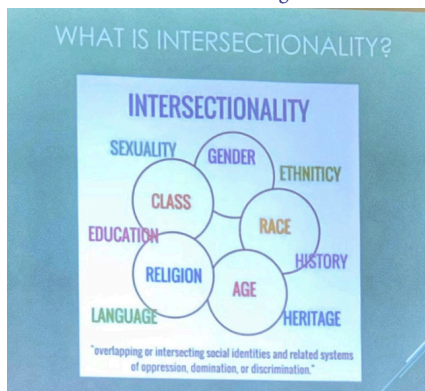
As it relates to the open knowledge movement, this phenomena may play out in governance practices and organizational principles. It could be a factor, for example, in determining or recognizing progress in new chapters, or in the criteria used to decide which projects to fund. It could show up as a preference for "having things in writing", or as a disadvantage for those in leadership roles who use cultural signals that can't be read at a distance (e.g. body language). It may also play out in terms of expectations of how decisions should be made,[4] or in the authenticity we assign to a particular types or sources of information.

And, to complicate things further, we must also recognize the intersectional [5] dimensions of trust in any cultural context - where who we are constrains what we can express and to whom. To



SOURCE ERIN MEYER FROM "GETTING TO SI, JA, OUI, HAI, AND DA," DECEMBER 2015

its credit, the Foundation has evolved a range of workplace standards and practices intended to promote more cultural and intersectional awareness within teams and between geographies. But the same kind of intentionality must be applied at the Movement level in order to build and maintain trust at a global scale.



From WMF All Hands 2018 presentation by Russell Robinson [6]

Accountability Within Social Groups

In addition to cultural norms around trust signals and trust-building, trust is also manifest at a social level, via the personalities, [7] norms [8] and perceptions of any particular social group. The distribution and prevalence of certain tendencies, values and attributes across a social group [9] within a given cultural context [10] impacts who will be able to dominate or influence and who will be perceived as more or less competent or trustworthy. [11] For example, in a social group that is risk averse, more daring suggestions will be considered less trustworthy. Or, in a group that engages mostly to share perfunctory information, emotional or subjective messaging (e.g. demonstrative praise, gratitude, etc.) could be regarded as unusual or suspicious.

It is important to recognize the nested relationship between cultural and social factors - how the cultural frame, and aggregate of

individual personality traits, facets [12] and sub-facets, [13] set the prevailing tone, tendencies, and qualities of a social group. And hence, determine whose judgements and decisions are to be valued and trusted.

Communities take on the personalities and tendencies of their most influential characters, and must be challenged directly when these tendencies become exclusionary or dysfunctional. But challenged by whom? Should it be the responsibility of the Foundation to sanction communities? No. But the Foundation can model zero tolerance for negative social dynamics within its own practices, and develop capabilities for identifying and supporting communities struggling with community health issues. Ultimately, and ideally, it must be the community which holds itself accountable for fostering dynamics which encourage growth, and sustain involvement.

Accountability for Interactions

Manners

“Cosmopolitanism”, a term dating back to fourth century BC, meant “citizen of the cosmos” – the controversial and paradoxical notion that every person belongs both to the particular place they live and to the universe at the same time. The concept was a rejection of the then prevalent idea that a person’s fealty and citizenship was defined entirely by their polis, and its stature among other city states. Voltaire later referred to cosmopolitanism as “an obligation to understand those with whom we share the planet.” [14]

“So there are two strands that intertwine in the notion of cosmopolitanism. One is the idea that we have obligations to others... beyond those to whom we’re related... or even [by] the more formal ties of a shared citizenship. The other is that we take seriously the value not just of human life but of particular human lives. Which means taking an interest in the practices and beliefs that lend them significance.” [14]

This “valuing of the practices and beliefs that

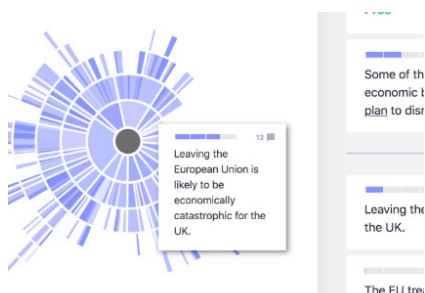
lend particular lives significance” typically shows up in the rituals of respect, or manners, we employ when engaging with strangers. Some form of etiquette exists in every culture, but what is considered polite in one context may not be in another. Yet we can demonstrate our cosmopolitanism, our obligation to and curiosity about others, whenever we act according to any form of politeness, since polite interactions are the means by which we demonstrate respect while exploring our differences.

In the late 1970’s, Singapore instituted the National Courtesy Campaign [15] “as a means of ensuring a smooth transition to a new Singapore which would be densely populated, where people lived and worked in high rise towns, offices and factories, while travelling in crowded buses and lifts.” The goal of the campaign was to encourage behavior that would make life better for everyone, and to improve the “self-esteem” of the Singapore resident. The decades long public relations campaign included a mascot, Singa the Courtesy Lion, who became a beloved cultural icon by promoting norms of behavior that would signal Singapore as a cosmopolitan (world) culture. While it’s worth noting that Singa retired in 2013 with an open letter saying it was “just too tired to continue facing an increasingly angry and disagreeable society” asking real people to step up as the mascot stepped aside. “The letter also stressed that “it’s not that we aren’t a gracious society, or that kindness is not innate in all of us. But some days it feels like not very many of us believe in or care about expressing kindness.”

The Movement is at a similar inflection point. In anticipation of the many, many “others” we must engage in order to meet our 2030 goals, the onus is squarely on each community member to step up, to demonstrate kindness, tolerance and curiosity in every interaction they have on the platform. That is, to be cosmopolitan... by honoring our obligation to one another and by choosing to take an interest in the belief systems that lend significance to the individual lives of our fellow community members.

Civil Debate

Kialo, the Esperanto word meaning “reasoning for”, is also the name of a web-based forum [16] designed to encourage and develop parliamentary debate as a counterpoint to the “Internet Shouting Factory”. [17] Kialo provides an interface for the crowdsourcing of ground truth answers to any question – from “Is Morality Objective?” to “Should the UK Remain a Member of the European Union?” to “Are we Living in a Simulated Reality?”. [18] Similar to Quora [19] and Aardvark [20], a Kialo debate starts from a user-contributed question. Unlike Quora and Aardvark, where questions are answered by “experts” whose status as such are determined by up-voting by their user-peers, Kialo questions are answered in structured debate via pro and con assertions, with each pro and con pair forming new branches of the “tree”. Each assertion, pro or con, can be commented upon, justified, and up/down voted, with the aggregate effect being a more nuanced consideration of every aspect of a contentious question, rather than the more typical conflation of aspects which often leads to anti-social comments.



Kialo Interface: Should The UK Remain a Member of the European Union? [21]

The Movement and its communities must not only advocate for civil debate, they must also model civil debate within the product, through interfaces that promote civilized discourse. This imperative will have two-fold returns: a more overt system for encouraging positive behavior,

and a structured database of dialog on every topic. The latter is a potential asset in and of itself.

03

Transparency

The general level of trust in digital platforms, both in terms of the accuracy of their content, and of their operating practices, has hit a new low.[1] Most content sites have optimized around the popularity of their content and the speed at which they pump it out, often to the detriment of quality and trustworthiness. As a culture, we can now see the social costs of moving fast and breaking things.[2] And while it's easy to bemoan the experience gap between wiki projects (slow and old) and other people-powered platforms there's a hidden upside. Our projects adapt at a plodding (human) pace, change is slow and painfully incremental. But slowness is an advantage when it comes to trust because trust is built on consistency and predictability. The open knowledge model has an innate stability and is inherently more reliable since very little of this output changes at the pace of the world around us.

In the current climate of distrust, being perceived as trustworthy presents an opportune moment for Wikipedia. With so much positive social capital built up over so many years, it's now time to take a risk: we must openly critique the flaws in this edifice in order to retain trust in the long run. Being more intentionally transparent about the messy process by which knowledge is created would almost certainly invite criticism but it's also the only way to begin to address gaps and bias at a system level.

How does Wikipedia work?

At the heart of a lot of the distrust of digital content, as well as the companies behind that content, is a lack of understanding about what's going on behind the scenes. How does Facebook's revenue model compete with the quality of information in your feed? How does Quora make money? What kinds of policies have worked to make Twitter more or less safe? How does a Wikipedia article come to be? While technology in general, and specifically content/knowledge/information has become more accessible, the processes behind it have become more and more obscure/opaque. These days information simply arrives at your doorstep, and perhaps we don't spend enough time thinking about how or why it got there. Google provides a decent example of what it looks like to explain the inner-workings of "Search" to people.[3] While Wikipedia is certainly complex, we need to do a better job of explaining what's going on, how it works, what hasn't worked in the past, where the money comes from, how it gets used, what a citation is, why it might not make sense for there to be a page about your grandfather, etc. What are some of the central tensions that we grapple with, and how might an understanding of those tensions make people feel more aware and engaged in the platform?

What this could look like in practice:

- Create a visual, ELI5 type of experience where we explain the central concepts of Wikipedia. This could exist as a standalone "micro-site"
- Create case-studies about certain articles (either directly on the platform, or as a supplemental thing) — narrative storytelling around how this article came to be, and what the future might hold for it
- Invest in exploring the general question of "what information can I trust?" (think about Snopes as a brand) — we can be facilitators of the general discussion of what trust means, and who deserves it

- Invest in contextual, in-product explanations of things like references, verifiability, and notability

Create a Wikipedia steward certification program where people can become certified in understanding how Wikipedia works, and explaining it to others.

Insider Knowledge

A big part of what has eroded trust in other social digital platforms is the degree to which they enable and engender echo-chambers - one now has to be very deliberate in seeking information that doesn't reinforce one's existing points of view. While Wikipedia doesn't share many of the negative attributes of other platforms, it does in this one regard: our contributor community places high value on neutrality but is not itself neutral. It attracts the like-minded, rewards the technically adept, and celebrates insider knowledge. "High standards" can no longer be used to justify lack of diversity in content or among contributors. Only by being more transparent about these shortcomings will the community be challenged to evolve, and make it possible for more diverse perspectives to be represented.

What this could look like in practice:

- Call attention to the demographic makeup of an article's authors/contributors
 - Highlight articles in need of more diverse perspectives and reward contribution from such folks
 - Publicly communicate what we know about content gaps and invite new community members to come help us address them
 - Be more aggressive about highlighting page issues, and find other opportunities to be proactive about where our content falls short
- Policies are set and upheld by the community, so by what mechanisms could the community be motivated to change its norms? What leverage is

there at the product-level to drive this change? What can be done to make invisible power structures explicit and more easy to navigate as newcomers? How can verifiability and accessibility be reconciled?

Calling Attention to Flaws

On the consumption side of the platform, readers must be able to see not only where content is under debate but also where content has been removed or altered by someone standing to gain from the change. By tracing edits to IP addresses at a particular locations,[4] or user names to corporate interests, [5] readers with sufficient motivation can ferret out abuses of the system and decide for themselves what to believe. But is there more the platform could do to identify conflict of interest? Is there more that can be done to surface tampering?

An objection might be that to call attention to vulnerabilities will undermine the credibility of the corpus overall, however transparency is simply preventive medicine.

“What sets us apart, I believe, has been our willingness to be transparent in our journey. We own our failures, we learn from them, and we share them publicly so that others can learn from our failings as well, which has helped us to bounce back higher than before when we fall.” [6]

When issues are known but go unaddressed, it invites greater scrutiny down the road. Being called out on a known issue is far more damaging to the organization than tackling the issue head on. As we see with Facebook, an organization's resources are far more strained by damage control than by self-initiated remedies.[7] Transparency is just good business.

Coming to terms with platform weaknesses - in terms of who is allowed to contribute and where abuses of the system occur - will call the integrity of the system into question. But only by

leveraging positive social capital, and tackling these challenges head-on will we retain trust over time. What this could look like in practice:

- Investing in raising public awareness around abuse, harassment, and bias on the platform

- Case study around a Wikipedia-article-gone-wrong

- Take even more responsibility around what qualifies as a good/bad source — develop a model for quantifying reliability of a source and share it openly so that others can understand how it works and help us improve it

What does this have to do with reaching our 2030 goals? Getting there will require scaling up participation in and consumption from the platform. Emerging communities will only invest to the extent that they feel their contributions will be welcome and secure- for this reason it will be critical for the existing community to trust them in order to be trusted in return.

Notes

1. Edelman Trust Barometer Global Report 2018 <https://www.edelman.com/trust-barometer>
2. Feezell and Krupnikov (2018) What's True and Fake About The Facebook Effect <http://behavioralscientist.org/whats-true-and-fake-about-the-facebook-effect/>
3. How Search Works <https://www.google.com/search/howsearchworks/>
4. https://en.wikipedia.org/wiki/United_States_Congressional_staff_edits_to_Wikipedia
5. The Atlantic “Wikipedia Editors For Pay” <https://www.theatlantic.com/business/archive/2015/08/wikipedia-editors-for-pay/393926/>
6. Forbes “Trust, the Most Valuable Business Commodity” <https://www.forbes.com/sites/davidkwilliams/2013/06/20/the-most-valuable-business-commodity-trust/#5c7539da6500>
7. <https://www.ou.edu/deptcomm/dodjcc/>

- Appiah, (2006) "Cosmopolitanism: Ethics in a World of Strangers", Norton
- Anderson (2009) "Why Do Dominant Personalities Attain Influence in Face-to-Face Groups? The Competence-Signaling Effects of Trait Dominance" https://pdfs.semanticscholar.org/d0d8/13eec040dce91df0012767ac45d9a7694b07.pdf?_ga=2.130622479.1083641897.1546406941-103297660.1546406941
- The Atlantic "Wikipedia Editors For Pay" <https://www.theatlantic.com/business/archive/2015/08/wikipedia-editors-for-pay/393926/>
- Big 5 Personality Traits https://en.wikipedia.org/wiki/Big_Five_personality_traits
- Business Insider (2015) "Personality Types That Clash at Work" <https://www.businessinsider.com/personality-types-that-clash-at-work-2015-5>
- De Young, Quilty and Peterson (2007) "Between Facets and Domains: 10 Aspects of the Big 5" <https://experts.umn.edu/en/publications/between-facets-and-domains-10-aspects-of-the-big-five>
- Edelman Trust Barometer Global Report 2018 <https://www.edelman.com/trust-barometer>
- EPIC Oxford Report "Assessing the accuracy and quality of Wikipedia entries" https://upload.wikimedia.org/wikipedia/commons/2/29/EPIC_Oxford_report.pdf
- Feezell and Krupnikov (2018) What's True and Fake About The Facebook Effect <http://behavioralscientist.org/whats-true-and-fake-about-the-facebook-effect/>
- Fleeson (2001) "Toward a Structure and Process-oriented View of Personality: Traits as Density Distributions of States." <https://personality-project.org/revelle/syllabi/classreadings/fleeson.2001.pdf>
- Forbes "Trust, the Most Valuable Business Commodity" <https://www.forbes.com/sites/davidkwilliams/2013/06/20/the-most-valuable-business-commodity-trust/#5c7539da6500>
- Harder, Velasco, Evans and Rockmore (2015) "Measuring Verifiability in Online Information" <https://arxiv.org/pdf/1509.05631v2.pdf>
- Hwang et. al. "(2014) Drug Safety in the Digital Age" <https://www.nejm.org/doi/10.1056/NEJMp1401767>
- Human Computation Blog (2013) CSCW 2013: Your process is showing... <https://humancomputation.com/blog/?p=4214>
- IIHI social media report "Doctors' #1 Source for Healthcare Information: Wikipedia" <https://www.theatlantic.com/health/archive/2014/03/doctors-1-source-for-healthcare-information-wikipedia/284206/>
- Intersectionality <https://en.wikipedia.org/wiki/Intersectionality>
- Lucassen, Shraagen (2010) "Trust in Wikipedia: how users trust information from an unknown source" <http://www.ra.ethz.ch/CDstore/www2010/wicow/p19.pdf>
- Lucassen, Muilwijk, Noordzij, Shraagen (2012) "Topic familiarity and information skills in online credibility evaluation" <https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.22743>
- Meyer (2015) "Getting to Si, Ja, Oui, Hai, and Da, Harvard Business Review" <https://hbr.org/2015/12/getting-to-si-ja-oui-hai-and-da>
- Meyer (2017) "How Cultures Across the World

Approach Leadership”, Harvard Business Review, <https://hbr.org/video/5476393165001/how-cultures-across-the-world-approach-leadership>

Mothe and Sahut (2108) “How trust in Wikipedia evolves” <http://www.informationr.net/ir/23-1/paper783.html>

Motherboard “Why People Trust Wikipedia More Than the News” https://motherboard.vice.com/en_us/article/ae37ee/in-defense-of-wikipedia

The New Yorker, “An Open Letter to Wikipedia” <https://www.newyorker.com/books/page-turner/an-open-letter-to-wikipedia>

PC World https://www.pcworld.com/article/170874/The_15_Biggest_Wikipedia_Blunders.html

Quora “Is Wikipedia Trustworthy?” <https://www.quora.com/Is-Wikipedia-trustworthy/answer/Andreas-Kolbe>

Schmitt (2007) “Geographic Distribution of Big 5 Personality Traits” <http://www.toddshackelford.com/downloads/Schmitt-JCCP-2007.pdf>

Schwartz and Morris (2011) “Augmenting Web Pages and Search Results to Support Credibility Assessment” https://www-cs.stanford.edu/~merrie/papers/WebCredibility_CHI2011.pdf

Times Online “Wisdom? More like dumbness of the crowds” https://web.archive.org/web/20110814104256/http://www.timesonline.co.uk/tol/comment/columnists/guest_contributors/article2267665.ece

The Utopian Fantasy of Rational Debate On the Web”, Nov 2017 <https://www.urbandaddy.com/articles/40999/kialo-is-an-internet-unicorn>

Washington Post “Conspiracy videos? Fake news? Enter Wikipedia, the ‘good cop’ of the Internet” https://www.washingtonpost.com/outlook/conspiracy-videos-fake-news-enter-wikipedia-the-good-cop-of-the-internet/2018/04/06/ad1f018a-3835-11e8-8fd2-49fe3c675a89_story.html?noredirect=on&utm_term=.3abf96bc9826

Whoseknowledge.org

Wikimedia Brand Awareness, Attitudes, and Usage, Executive_Summary https://meta.wikimedia.org/wiki/Strategy/Wikimedia_movement/2017/Sources/Brand_awareness,_attitudes,_and_usage_-_Executive_Summary

Wikimedia Blog, “Doctors use, but don’t rely totally on, Wikipedia” <https://blog.wikimedia.org.uk/2012/04/doctors-use-but-dont-rely-totally-on-wikipedia/>

Wikimedia Donor Survey: “What can we do to make Wikipedia better?” <https://docs.google.com/spreadsheets/d/1VaRS7PjOgdtB7HWz5QKue6PdFiNGTNZvStbFk7VRjPw/edit#gid=0>

Urbandaddy.com “Kialo Is An Internet Unicorn



Experience

This collection of perspectives on the various aspects of user experience are based on insights from the Communities and WMF staff, as well as synthesis of secondary research.



01
Form Factor

02
Rich Content

03
Contributors

04
Customization

05
Discovery

Published
January 2019

Authors

Ramsey Isler
Nirzar Pangarkar
Anne Gomez

Contributors

Adam Baso
Volker Eckl
Charlotte Gauthier
David A.M. Goldberg
Rita Ho
Alex Hollender
Danny Horn
Jess Klein
Jon Katz
Josh Minor
Toby Negrin
Margeigh Novotny
Olga Vasileva



Wikimedia projects have become important fixtures in the infrastructure of knowledge sharing on the internet. However, our share of media interaction and consumption is shrinking as new populations come online, new platforms arise and the internet becomes multi-modal, more interactive, and more social.[1][2] These changes challenge both our consumption and contribution models, and to date, we have struggled to adapt to the opportunities that have arisen and the changing expectations of our users.

A user's expectations of quality are shaped by the totality of their digital experiences. To match these expectations, we have to match the quality of other experiences that users are exposed to. The simplest way to do that is to re-use and refine patterns, methods, and mental models of popular platforms.

These expectations extend to the media types they engage with. Our platform does support diverse media types including video, audio, images, and animated GIFs, and allows them to be mixed in in a single document. However, the experience of uploading and consuming this media does not match the use of media on other major informational platforms. Moreover, our content is saved and presented as a single blob of mostly text, and for some forms of knowledge, such as explanatory knowledge, we do not provide a space or tools to generate, curate or engage with that form of knowledge.[3]

On the contributors' side, Wikimedia projects compete with modern platforms that provide gentler on-boarding and guidance to new users. The competing platforms provide rich, multimedia editing tools and emotionally reward their users with explicit gratitude, meaning, and status.

Additionally, user expectations are rapidly growing with regard to tailored experiences. Software is becoming more aware of individual user needs. The likes, dislikes, and personal preferences of users are vital considerations for modern software design. We distinguish between personalization, in which a system uses what it knows about the user to determine a person's experience, and customization, which empowers users to control their experience. Through customization we can provide tailored experiences, without sacrificing our values or principles.

Data adaptability and content structure are required for the creation of modern user experiences across form factors. Users should be able to engage with Wikimedia, as consumers and creators, in the diverse variety of form factors that are the contemporary internet. And the platform must provide the flexibility to build new experiences for emerging form factors.

Finally, discovery models will be key, as will having captivating content people want to discover. But that discovery process must be proactive on our part. A large and growing body of research supports a key product theory about today's media - content finds the consumer.[4][5]

Our current products severely lack user awareness and interface customization for the vast majority of our potential audience. We can remedy this by following some basic modernizing principles in our user experience and development processes:

- Embrace a "factoid" paradigm; a lot of people still want to read long-form content, but a lot of people don't.
- Go where the people already are and utilize platforms they already like.[6]

- Provide suitable content format alternatives for subjects that are not well-served by long text or require advanced levels of prior knowledge.[3]
- Purposely become a tool that empowers others to create, promote, and remix knowledge-based content in many formats.[7]
- Separate the advanced editing experience from the reading and basic editing experiences.
- Provide easier customization of information and interface to match individual needs.
- Use user-centered design to meet consumer expectation.[8]

By understanding our users needs and expectations we can modernize our products, and provide a user experience that informs and delights.

Examples

Apps Explore feed
Page previews
Popcorn video editor
Content APIs

Areas of Impact

Wikidata [9]
Commons [10]
iOS and Android apps [11]
Wikipedia [12]

Key External Factors

Social Platform dynamics
Demographics
Technological waves [13]
Google [14]

Notes

[1] 2018 Adobe Consumer Content survey contains multiple data points that describe what people expect from digital content experiences <https://www.slideshare.net/adobe/2018-adobe-consumer-content-survey>

[2] Research indicates that millions of users say they use Facebook, but not the Internet. “In their minds, the Internet does not exist; only Facebook.” <https://qz.com/333313/millions-of-facebook-users-have-no-idea-theyre-using-the-internet/>

[3] Learning Styles: Concepts and Evidence”
- Research shows that there’s little evidence supporting the popular idea of catering to “learning styles”, however; “differences in educational backgrounds can be a critical consideration in the optimization of instruction. [...] student’s prior knowledge is bound to determine what level and type of instructional activities are optimal for that student” and “the optimal instructional method is likely to vary across disciplines. For instance, the optimal curriculum for a writing course probably includes a heavy verbal emphasis, whereas the most efficient and effective method of teaching geometry obviously requires visual-spatial materials.” https://www.psychologicalscience.org/journals/pspi/PSPI_9_3.pdf

[4] “I Just Google it”: Folk Theories of Distributed Discovery, is fantastic and very recent (June 2018) research on how people find content on the Internet. <https://academic.oup.com/joc/article/68/3/636/4972617> “

[5] Effects of the News-Finds-Me Perception in Communication: Social Media Use Implications for News Seeking and Learning About Politics: “The news-finds-me effect stems from individuals’ perceptions that a) they are

well informed about current events despite not purposely following the news, because b) the important information 'finds them' anyway, through their general media use, peers, and social connections." <https://onlinelibrary.wiley.com/doi/full/10.1111/jcc4.12185>

[6] Jonathan Morgan's 2015 research on free online learning resources used by students https://meta.wikimedia.org/wiki/Research:Student_use_of_free_online_information_resources/Results

[7] "Sharing small pieces of the world": Increasing and broadening participation in Wikimedia Commons - recent research and interviews with people who use other image sharing platforms but aren't heavy Commons users http://www.opensym.org/wp-content/uploads/2018/07/OpenSym2018_paper_30-1.pdf

[8] The Aesthetic-Usability Effect: "Users are more likely to want to try a visually appealing site, and they're more patient with minor issues" <https://www.nngroup.com/articles/aesthetic-usability-effect/>

[9] Wikidata can potentially provide the common vocabulary for many organizations/sites to use for referencing their content so it is discoverable by both humans and algorithms

[10] Commons has to change (or be used differently) in significant ways, primarily because its user experience is far outdated and not at all what users of modern commercial multimedia tools want.

[11] The apps are already utilizing many modern experiences, but further integration of social media, discovery, and multimedia will be needed.

[12] Wikipedia's article consumption experience

is adequate, but there is room for improvement. Additionally, an improved editing/contributing user experience is an opportunity for attracting more people to enrich existing pages and become regular contributors.

[13] The future is hard to predict, but there are definite tech trends in place today that aren't going anywhere in the next 5 years. According to the 2018 Adobe Consumer Content survey, the top devices consumers expect to use most in the next 5 years include: Home Entertainment Streaming Devices (82%), Smart speakers/home assistants (64%), connected home appliances (44%), and wearable devices (42%)

[14] Google controls many of the most powerful content discovery tools on the planet.

01

Form Factor

A significant challenge in making all the world's knowledge accessible to all the world's people will be to ensure that it is optimized and future-proofed for a rapidly-evolving digital consumption environment. The term Form Factor generally refers to the various entry points, devices, channels and formats that define a digital product offering. In the context of Wikimedia, form factor will include (a) the variety of devices that Wikimedia content can show up on now and in the future; (b) the size, and flexibility of the content itself. Such explicit form factor considerations are the main focus of this paper, however, other implicit aspects of form factor must be considered as well.

Partnerships, for example, may require Wikimedia content to be adapted to, or deeply integrated into, third party products with assistance and guidance from the makers of those products; several of the explicit considerations suggest this type of partnership. Likewise, ideas such as making Wikimedia content available as a utility, or layer on top of the Internet, is another implicit example of form factor, and explored in many of the concepts described in the paper.

Devices

The future will bring new devices and screen sizes. Wikimedia content should be easy to access and easy to use on every device

In the last five years alone we've seen a remarkable rise in the number of devices people use to consume internet content. In just three years, the Apple Watch has become the best-selling wearable device in the world and a common sight. In ten years, smart appliances have become a fixture in millions of homes. Just recently, Amazon announced a new line of Echo devices boasting its Alexa technology. Most of these devices only have audio capabilities, but some include small screens as well.

Amazon's Echo is the most popular smart speaker system and has already reached 10% of US respondents in the Digital News Report. [1] Amazon's Alexa (which powers Echo) can already read Wikipedia articles by using text-to-speech technology, but its ability to hone in on specific facts within that article is limited.

In the near future, we'll see other competing devices come onto the market. In addition to text content, these devices will need audio content to play. If I ask Alexa to play Franklin D. Roosevelt's famous "Day of Infamy" speech, I'm prompted to buy it from Amazon Music (even though it's in the public domain). Other devices, without their own massive digital storefronts, will want to be

able to use that kind of content from an openly licensed source. We should ensure that we are that source. We should also put focus on optimizing all Wikimedia content for a range of devices with screens. Home entertainment devices are a major new platform for consumption of media. 82% of consumers expect such devices to be their most used devices in the next 5 years.

Desktop experiences (and even mobile experiences) are becoming less relevant. Even though mobile usage is still high, 2018 saw the first global decline in mobile sales and app installs and opens are in a downward trend [2] too. Wikimedia properties will be left behind without a viable platform for atomic content that can adapt to these devices..

One exception, however, may be wearable devices. The current top wearable is the Apple Watch, but with its limited capabilities and tiny screen size there may not be much we can do there.

Takeaways and things to do:

Tactical To Do's

- Decide that it is in our best interest to influence, perhaps even control, the user experience that Echo and other device users have with Wikimedia content. It's not "our" content, but we are the stewards of the systems used to create and disseminate that

PROPORTION THAT USE FACEBOOK MESSENGER AND WHATSAPP FOR NEWS

Selected markets

Messaging Apps	Greece	Norway	United States	Australia	Finland	Argentina	Hong Kong	Malaysia
FB Messenger for news	22%	11%	7%	11%	5%	9%	8%	12%
WhatsApp use for news	4%	2%	4%	10%	10%	37%	38%	54%

content, and therefore it's our responsibility to ensure those systems are used optimally and appropriately

- Encyclopedic content by itself, while useful and what we're best known for, isn't all we have to offer. We have a wealth of imagery, video, audio, and instructional content that better fits the audio and visual strengths of new devices.

Technical To Do's

- Create tools and APIs that are purpose-built for the audio-based smart speaker/home assistant experience
- Improve our backend tech and documentation for thumbnailing uploaded images. We recently had an issue with the Amazon team regarding this (they were trying to download size-appropriate images linked to Wikipedia articles for the Echo Show, but were requesting files that were too big, too often)
- Improve our support for common multimedia formats, especially audio. We now have MP3 on Commons, but we should take initiative in making sure existing files are available in that format.
- Utilize Structured Data to make multimedia easier to find and easier to associate with content from a number of sources

Content

Wikimedia Content will have to adapt to accommodate different user needs

Currently, our flagship project, Wikipedia, specializes in long-form in-depth content. This should definitely be considered a strength, and one that serves most of our current user base well. Wikipedia's brand is currently (and probably solely) centered around accurate, informative, long form information. But having only long-form information can be a problem..

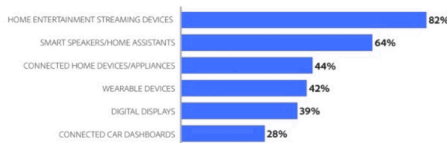
Moving forward, it'll be critical for our content to adapt to shifting habits and the expectations of new audiences we begin to reach. As form factors change, the long form and complex nature of Wikimedia content may start to undermine the value of its accuracy and extensive coverage.

According to the 2018 Adobe Content Consumer Survey (US only), [3] when content is too long, 47% of consumers stop reading, and 23% switch devices. Our content needs to adapt to different contexts that reflect how people actually use social media and messaging.

The social messaging use case is a very important one to focus on because direct communication

Devices to be Used Most in the Next Five Years

- The devices that are most expected to be used in the next 5 years are connected home entertainment streaming devices, smart speakers/home assistants, and connected home devices/appliances.
 - Females are more likely to expect to use wearable devices (46%) while males are more likely to expect to use connected car dashboards (32%).
 - Millennials are more likely to expect to use smart speakers/home assistants (70%) and wearable devices (47%). Baby boomers are more likely to expect to use digital displays (45%).



TOP DEVICES CONSUMERS EXPECT TO USE MOST IN THE NEXT 5 YEARS

tools like Whatsapp and Messenger are on the rise for news, [4] particularly in emerging markets, Asia, and South America. This phenomenon barely exists in the US, but it is a huge content consumption driver in other countries.

Wikimedia projects need to become stronger here, and that means having content and systems that naturally fit with messaging usage patterns. This could include possible future features like:

- Chatbots - Imagine a Wikipedia bot on WhatsApp, Facebook Messenger, or Telegram that actually answers questions and links to citations when you talk to it. It could even be a form of interactive instruction that guides you through a topic based on your prior knowledge, available time, etc.
- Media bots - “Hey Wikipedia, show me video of World War II.” Instead of taking you to a link, the bot can put customized/curated video clips right into the chat
- Have link previews that show the fact you’re interested in—automated page links with <title> tags that have the text of the specific fact you want to share; so instead of seeing “Patrick Stewart - Wikipedia” as the link preview, users can see “Patrick Stewart was knighted on June 2, 2010”

Adaptive Learning [5] can help readers customize their path through content. Imagine a world where we could ask a reader what level of knowledge they have for a Wikipedia page topic, and then automatically reconfigure the facts and citations to fit the reader. For Readers with advanced knowledge, the page can automatically skip the basics, while readers with basic knowledge can be presented with an “explain like I’m five” version. This approach also provides an on-demand customization experience which helps us avoid privacy concerns.

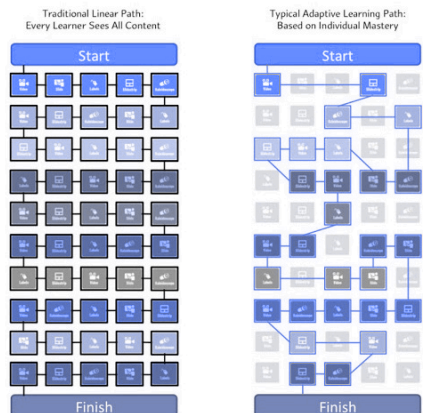
The Article vs. The Fact

All of the scenarios above are problematic for us right now, mostly because of one thing - our core, fundamental element is the article, not the fact.

As mentioned in the Discovery document, major Internet players like Google and Facebook are already grabbing our text content, chopping it up, and presenting it in factoid-sized chunks. We currently don’t have influence over this process and the resulting user experiences, largely because we don’t have any facility that we can point to and say, “do it this way, it’s better and we’ve already done the work for you.”

There are several ways we might achieve this “atomizing” of articles:

- Automatically break up the entire article into elemental parts (sentences/passages perhaps)
- Take the top 5 most important elements/sections from each article and atomize that (although it’s unclear at the moment how we would identify those top 5 elements).
- Have the community decide which elements should be atomized for each article (essentially a new editor function/workflow)



Of these three options, the first is probably the most flexible, most likely to scale, and the most likely to fit every possible need we may have in the future.

The Reasonator [6] project has made an attempt at “prettifying” Wikidata facts into human-readable form with mixed results. With better technology and techniques, we may find a better automated fact-generated system in the near future.

Takeaways and things to do:

Tactical To Do’s

- Embrace a “factoid” paradigm; a lot of people still want to read long-form content, but a lot of people don’t.
- Encourage and enable quick answers to discrete questions. ensure those systems are used optimally and appropriately
- Explore how we can optimally serve content in short-form environments like social messaging

Technical To Do’s

- Architect a methodology for breaking up, storing, and serving our text content into individual, atomic elements that can be paired with citations
- Explore content adaptation architecture so that pages can change their form based on context and/or reader needs
- Build our own social media/messaging APIs and improve our integration with others
- Explore automated video/audio file creation (combining multiple clips or images into one and sending it off to the user’s touchpoint)
- Utilize Structured Data to help put it all together. This could potentially be Wikidata (or a new feature on Wikidata), or an entirely new tool altogether since some fact formats just don’t easily fit into Wikidata right now.
- Improve our on-wiki search capabilities to enable “factoid” searches on our sites just

as we would on other platforms. This could include integrating structured data into search to ensure semantic matches, improve accuracy, and enable highly focused searches (see structured data search on Commons as an early example).

Notes

[1] Reuters (2018) “Digital News Report” <http://www.digitalnewsreport.org/>

[2] Connelly (2017), The Drum “Future of mobile apps looking bleak” <https://www.thedrum.com/news/2017/02/27/future-mobile-apps-looking-bleak>

[3] 64% of respondents in Adobe’s Content Consumer survey said that smart speakers/home assistants are devices they expect to use most in the next 5 years. <https://www.slideshare.net/adobe/2018-adobe-consumer-content-survey>

[4] Kalogeropoulos, Reuters Digital News Report (2018) “The Rise of Messaging Apps for News” <http://www.digitalnewsreport.org/survey/2018/the-rise-of-messaging-apps-for-news/>

[5] Adaptive Content Learning provides a possible framework for the future

[6] Manske’s Reasonator <https://meta.wikimedia.org/wiki/Reasonator>

02

Rich Content

The **Open Knowledge Movement** [1] encompasses much more than Wikipedia articles. Reading long-form text is not currently the only, or optimal, way people choose to gain knowledge. In fact, a recent study [2] shows that only 20% of Wikipedia consumers are in-depth readers, no matter what language you consider.

And while Wikipedia is mainly associated with long-form, informative, encyclopedic text content, several Wikimedia projects [3] already offer much more than that. Commons, [4] while flawed, is an established source for freely licensed multimedia files. WikiVoyage, [5] though largely unknown, is full of rich and useful crowdsourced travel content. Wikisource [6] has a small but dedicated community of transcribers, translators, and archivists who combine imagery and text into useful digital reproductions of old publications.

Our existing projects already offer rich opportunities to expand beyond encyclopedic content and give our users useful and fulfilling experiences, so it won't be a stretch for us to continue to explore all types of media and formats to accomplish our goals.

Roadmap for the future?

The National Geographic Society is one of the most well-known and successful global non-profits. They began with a magazine, which stood as their only media platform from 1888 until 1964 when they aired their first television content on CBS.

Today, in partnership with 21st Century Fox, the Society still operates the magazine that got them started, but they have also branched into other forms of media including TV channels, films, a website that features extra content, worldwide events, and other media operations. After decades of exclusively being a magazine brand, today “NatGeo” is truly a successful omni-channel presence. We should seriously consider using this approach as well.

An Omni-channel approach

Disruption. It is perhaps the one word that best describes what happened to the print industry in the past 20 years. Technology didn't kill print, but it certainly gave it a mortal wound. Disruption wrecks companies, and the best defense against it is diversification.

What would National Geographic be today if they had not ventured into other media? What if they'd remained strictly a magazine company? In the early 1990s the company's flagship publication (National Geographic Magazine) boasted 15 million subscribers. That number was estimated to be closer to 3.5 million in 2015. [7] Although it is possible that National Geographic would still be around if they'd stuck with magazines, they wouldn't have been able to do so without a massive restructuring, and there's no question that their current setup increases their outreach to millions more people than they could ever hope to reach with just a print magazine.

How does all this apply to us?

It is doubtful that we'll see another online encyclopedia rise to compete with Wikipedia, but that's mainly because the encyclopedia business isn't exactly a growth industry these days. Sooner or later, something will disrupt our model. It might be that our donations dry up, or larger companies grab all our data and start their own thing (think Wikitravel but in reverse), or Artificial intelligence algorithms perfect the automatic creation of articles based on news. Something's going to happen, and it's in our best interest to diversify so we have defenses against the inevitable and the unforeseeable.

Three types of knowledge

Before we focus on the content formats we should consider for the future, let's talk about “types” of knowledge. Our vision statement mentions “a world in which every single human being can freely share in the sum of all knowledge”, but what does that really mean? There are least 3 types for us to target:

Factual

This is already a strength of Wikipedia and its straightforward, neutral, “citations needed” format. However, the same things that are Wikipedia's strengths are also its weaknesses. They make it hard to find new editors who want to work in the intimidating and often conflict-laden processes of the Wikipedia world. Few people feel like factual experts, and even fewer feel like vigorously defending their claims.

Instructional

This is a weakness for Wikimedia projects. Wikipedia is very good at describing things but very bad at telling you how things work. You can see how bad we are at this with a simple experiment. Search for “rocket” on Google and the Wikipedia article for rocket shows up very early in the search results. Now try “how do

rockets work” and see what happens (hint: you’re gonna have to go to the dreaded page 2 of search results). This isn’t just an artifact of poor SEO—the rocket page really does not do a good job of simply explaining how rockets work.

Experiential

This is an important area where we again are lacking useful content. People want to know what it’s like to be X or do Y or visit Z. Knowledge of someone’s else’s experience is just as valuable as discrete facts, and a key element of getting humans to understand each other. Similar to the shortcomings of explaining how things work, Wikimedia projects are not very good at conveying experiential knowledge.

Now, let’s talk about the content formats we can use to convey knowledge - Video, Audio, Images, Text, and Interactive.

Video : The Elephant In the Room

“What are we going to do with video?” is a question often asked not only within the Foundation, but within our community as well. Video has become the most preferred learning method for the majority of Millennials and GenZ. In the next 3-5 years it will be crucial for us to expand the Foundation’s video capabilities. [8]

Factual Video

“Explainer” videos are a popular genre. We should give users the ability to create their own videos explaining certain topics or giving highlights of their favorite Wikipedia articles.

Instructional Video

Even the most talented writer in the world couldn’t write an article to teach you sign language. But it would only take a decent teacher to make a useful sign language course with video. That, in a nutshell, is the advantageous power of video when it comes to instruction.

YouTube, solely through the power of its user community, has become a prime hub for learning how to do things. Although much of YouTube’s content consists of non-informational vlogs and commentary, there is a large body of knowledge-based instructional content, from life skills like cooking to professional skills. It is even used by surgery trainees. [9]

The Wikimedia Foundation has the unique opportunity to learn from the YouTube model and improve it with openly licensed instructional video content that is translated, vetted, and highly curated by our communities for joyful consumption by all.

Experiential Video

Imagine free, openly licensed video content that provides detailed and compelling stories of the experiences of a wide variety of people, cultures, events, and walks of life. Think “Wikipedia originals”

Documentaries are a time-tested form of information sharing that can raise awareness and enter the social consciousness. It makes natural sense for Wikimedia to explore this medium, but not in the typical way. Other organizations already do documentary content; the BBC, Discovery Channel, and even National Geographic and Smithsonian. But all of these organizations focus on telling stories from a limited number of perspectives. Our strength is the Wiki way - allowing multiple perspectives and contributors.



Fortunately, we don't have to guess or theorize about what this might look like. There's a perfect real-world example in the form of *Winter on Fire*, [10] an Oscar nominated, Emmy nominated, feature length, Netflix original documentary that chronicles the deadly anti-government protests in Kiev, Ukraine that took place in 2013.

Winter on Fire had 28 credited cinematographers, using video footage captured by ordinary people who were on the ground during the conflict. In many ways it was a crowd-sourced film. [10]

From the director of the film:

“We got footage from people’s phones, from GoPro cameras, from TV crews, from wherever we could. Without these volunteer cinematographers and the variety of technology available, it would have been impossible to document the movement.”

The parallels to the Wikimedia processes and movement are clear.

Tactical & Technical Takeaways

Tactical To Do's

- Embrace video as a cornerstone of our media strategy going forward
- Invest in architecture, policies, and community members that support online editing tools for video
- But also support static, immutable content that is created by collecting disparate pieces of Wiki content and combining it into a finished “work” that is greater than the sum of its parts (ala *Winter on Fire*) [10]
- Answer an important question: Do we have to limit ourselves to the Internet as our only distribution method?

Technical To Do's

- Figure out how and where we're going to serve and store all this video
- Explore and invest in collaborative online video editing solutions like Popcorn.js
- Devise tools that enable admins to vet and moderate video content

Audio - the forgotten format

When most people think of audio content they think of music, but there's so much more to hear. Audiobooks, podcasts and storytelling shows are more popular than ever. Meanwhile, text-to-speech technology, in the form of personal assistants like Siri and home smart speakers like Amazon's Echo, are creating a renaissance for audio content.

In their US only 2018 Consumer Sales Survey [11], the Association of Audio Publishers found the following:

- audiobook sales in 2017 totaled more than \$2.5 billion, up 22.7% over 2016, and with a corresponding 21.5% increase in units
- this continues the six-year audiobook trend of double-digit growth year over year.
- 54% of audiobook listeners are under the age of 45 (in other words, it's not just for the olds)
- 24% of listeners said they have listened on a smart speaker and 5% said they listen most often on a smart speaker
- the top three activities while listening to audiobooks are: driving (65%), relaxing before going to sleep (52%), and doing housework/chores (45%)
- 52% of people said borrowing from a library/library website was important or very important for discovering new audiobooks. 43% of listeners said they downloaded an audiobook from a library

If audiobooks are a growing form of media as a “port” of a type of long-form text content (books), perhaps we can adopt that model as well, but for Wiki articles. And perhaps we can

serve as the library source for audiobooks in areas that simply don't have a lot of libraries.

Factual Audio

Factoids in audio form, and beyond

What does the fox say? We can answer that question with recordings of fox sounds that are openly licensed and freely available as part of a Wiki database of animal sounds.

"Alexa, in what year was Nelson Mandela released from prison?" Audio facts like that are already being provided by products using our platforms, however, we can enhance that work with "links" to audio files. For example, with structured data we can link topics to available media on that topic, so when someone asks about Mandela getting out of prison we can programmatically suggest the user listen to the public domain speech Mandela gave after he was released.

Instructional Audio

Audio can teach you more than just French:

Imagine a young person in India learning to speak the indigenous language of their region, or an art student in Canada listening to a public domain podcast that is produced by a group of women in Quebec and focuses on traditions of Inuit art.

And, yes, we could also have language courses so you can learn French.

Experiential Audio

Listen to someone tell their story

Oral histories have been a hot topic within the foundation for some time, and they are just the kind of experiential knowledge that we've largely neglected. However, oral tellings don't just expose us to mythology and history of esoteric cultures in far off lands. They can be used to give us deep connections to topics most of us only

scratch the surface of.

Imagine a CC0 collection of audio recordings from Holocaust survivors recounting their personal stories.

Tactical & Technical Takeaways

Tactical To Do's

- Don't forget audio! It is a flexible and easy to use file format
- Invest in obtaining/acquiring rich histories in audio format
- Explore what it would take to become a repository of all the sounds in the world (machine, animal, and other)

Technical To Do's

- Consider ways to use data to link topics/facts to available audio related to that fact
- The audio playback experience for our files on the web is terrible now. Commit to fixing it.
- File formats. We have MP3 support now (patent expired). What else might we need to do to provide good streaming audio quality?

Images: a long way to go

Commons falls far short of expectations for a modern image sharing platform. There. Now that we have that out of the way, let's talk about the kind of image content tools we should have going forward.

Factual Imagery

Photos are a type of fact that ordinary people feel comfortable adding. What exactly does the Bambino cat breed look like? It's a simple question with a simple answer (just show me a picture of one), but you won't find it on Wikipedia [12] or Commons because we don't have an image for it.

Commons has numerous and varied knowledge gaps, and doesn't get close to being a great tool for even illustrating Wikipedia articles, let alone being a source for visual reference for all knowledge.

A large part of this problem is that Commons itself provides a poor experience as a destination. Its UI and design are outdated, it lacks features people have come to expect on platforms like Flickr and Instagram, it has almost no social component, and its processes are even more obscure than Wikipedia's.

Put simply - we can't get great, comprehensive image coverage because people don't want to use the site.

So the first step is to change that. We've enlisted the help of gurus like George Oates to give us outsider perspective on what's wrong with Commons, [13] but changing Commons is an expensive and time consuming task (as anyone on the SDC project can tell you). The more prudent approach seems to be to forget the idea of Commons as a destination for the masses and instead use it as a piece of infrastructure maintained by people who like that sort of thing. Imagine a world where we do image campaigns like Wiki Loves X every week instead of every quarter. Imagine a massive social media campaign like #WikiLovesWednesday, where every Wednesday we ask the whole world to donate photos of that week's topics.

But we can't do that now because directing millions of people to Commons would make them hate us and ask why we made them go to that terrible, terrible place and they're never coming back.

We propose a New Commons, which would include the creation of a purpose-built, user-

friendly new presentation layer on top of Commons. This is still just in the concept phase, but has gotten some support so far.

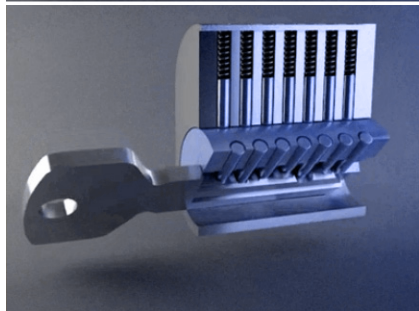
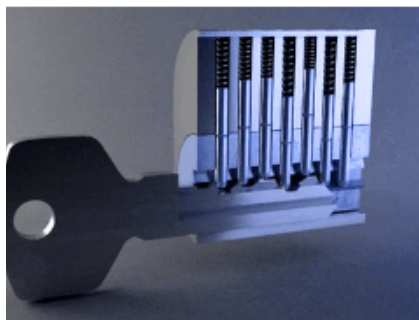
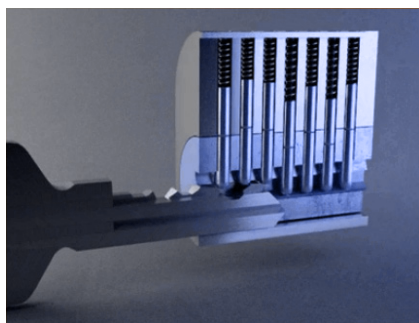
Instructional Imagery

Images are often much more efficient at explanation than text (and usually don't need translation). Let's say you're trying to learn how ordinary door locks work.

Which experience do you find more elucidating? This text from the Wikipedia entry....

"The pin tumbler lock uses a set of pins to prevent the lock from opening unless the correct key is inserted. The key has a series of grooves on either side of the key's blade that limit the type of lock the key can slide into. As the key slides into the lock, the horizontal grooves on the blade align with the wards in the keyway allowing or denying entry to the cylinder. A series of pointed teeth and notches on the blade, called bittings, then allow pins to move up and down until they are in line with the shear line of the inner and outer cylinder, allowing the cylinder or cam to rotate freely and the lock to open."

Or this GIF explaining How a lock and key work?



Experiential Imagery

Everyone can be a photojournalist. Social photo sharing is a common activity now, with people sharing their personal experiences of travel, dining, and events both mundane and fantastical. The good news is that the Wiki way - where everyone and anyone is invited to contribute - nicely meshes with broader internet usage patterns, and other hugely popular platforms have already trained the masses to always have their smartphone camera at the ready.

Additionally, we can take a page out of the National Geographic book and encourage people to capture ordinary life in extraordinary places. We've seen some good photo essays come out of Wiki Loves Africa, but they're wasted on Commons, which simply doesn't have the reach or format to really showcase this content.

With the right tools and design and a motivated community, we can do what many photojournalistic outlets do, but at a scale they can't achieve.

Tactical & Technical Takeaways

Tactical To Do's

- Invest in at least one (possibly more!) welcoming, useful, and usable place for people to share/donate their images
- Animated GIFs have made a strong comeback. They are also fantastic for informational and instructional content. Support and explore the idea of static images having less prominence in the future.
- Encourage experiential storytelling through imagery

Technical To Do's

- Make frictionless mobile image contribution a priority

Text : Fix it

Wikipedia isn't perfect, but it works. Its success is undeniable, and it will stand in history as a world-changing project. If you're the kind of person who loves deep dives into complex topics, and you don't mind spending time with text that can be challenging, Wikipedia's current format totally works for you. There are certainly parts of its formula that work and should be left alone.

But there are other Wikimedia projects that are heavily text-based and far more accessible than Wikipedia. They are in drastic need of some attention and fixes.

Factual Text

Facts matter, but we don't have all of them yet. Wikisource is a fantastic...well...source, for all sorts of information that just wouldn't work on Wikipedia. Want to read a biology text for kids that was published in 1875? It's there! Want to read the One Thousand and One Nights story in the original Arabic? It's there!

There are countless rich and engaging pieces of public domain or freely licensed text content out there. Some of them are digitized already, but many are not. A lot of stuff is locked away in archived books sitting in vaults (Charlotte's Culture Outline touches on this). There's a large opportunity for us to facilitate the process of freeing this text and bringing it to the people.

Sometimes, however, the content we want won't be under a license we like. Recent out-of-print content can still be under copyright, but perhaps there's room for us to fund the process of purchasing the rights for valuable content and then releasing it to the public domain or CC0 license.

Instructional Text

Video is king, but text is still like...a duke or

something. Video is still the most engaging and powerful medium if you want to tell someone how to do something, but video is not always easy to create, much harder to edit, and not as portable as well-written text instructions.

As we've discussed earlier, Wikipedia is terrible at instructions. But Wikisource and Wikivoyage are pretty good at it! You can find an entire book on how to teach yourself Chinese on Wikisource, and Wikivoyage has a wealth of content like how to buy a kimono in Japan.[14]

Experiential Text

Good writers can make text descriptions come alive. This is another area where Wikisource and Wikivoyage can shine if we let them. Wikipedia's neutral point of view rule makes the content fair and less prone to bias (although not impervious to it), but that rule also makes the content bland and no fun to write.

Investing in our other projects with less stringent content rules will help attract new readers looking for something less dry, and contributors who can really write and want to use that skill for a good cause.

Tactical & Technical Takeaways**Tactical To Do's**

- Spend money on Wikisource
- Spend money on Wikivoyage
- Spend money on acquiring rights to books, articles, and other text content we want (then make it free)

Tactical To Do's

- Much of Wikisource is held together with string, gum, and hope. Volunteer developers have kept it working with many disparate hacks over the years. We'll have to either commit to fixing it or, perhaps even better, create a new platform that is purpose-built for this use case.

Interactive

This is a complex topic, best described in slide deck form. See Future of Commons Deck, Path #3. [15]

Priorities

This is our recommendation for how to prioritize investment in the rich media types described in this paper:

- Video support. Brion Vibber has put together an excellent to-do list for improving video on our project. [16] There are many legal and engineering problems involved here, and they will take time to sort out, which is why we need to start on this list ASAP OR decide we're going to build a separate, purpose-built, video platform from ground up.
- We desperately need to improve our image situation. Commons does not match modern user expectations for what an image sharing site should be, and we're really missing lots of photo opportunities where simple, common sense changes could make massive differences in scale (more images and more diverse audiences uploading) and ubiquity (more people reusing and discovering the content).
- Define a clear strategy for audio. This is the "open sky" area where there aren't many competitors and we could potentially do groundbreaking things very quickly.
- Develop and set in motion plans to nurture Wikivoyage and Wikisource.

Notes

- [1] https://en.wikipedia.org/wiki/Open_knowledge
- [2] "Research:Characterizing Wikipedia Reader Behaviour/Human development index and Wikipedia use cases" https://meta.wikimedia.org/wiki/Research:Characterizing_Wikipedia_Reader_Behaviour/Human_development_index_and_Wikipedia_use_cases
- [3] https://meta.wikimedia.org/wiki/Our_projects
- [4] https://commons.wikimedia.org/wiki/Main_Page
- [5] <https://www.wikivoyage.org>
- [6] https://en.wikisource.org/wiki/Main_Page
- [7] Thielman (2015) The Guardian "How Fox ate National Geographic" <https://www.theguardian.com/media/2015/nov/14/how-fox-ate-national-geographic>
- [8] Pearson (2018) "Beyond Millennials: The Next Generation of Learners" https://www.pearson.com/content/dam/one-dot-com/one-dot-com/global/Files/news/news-announcements/2018/The-Next-Generation-of-Learners_final.pdf
- [9] Al-Khatib "Surgical Education on YouTube" <https://pdfs.semanticscholar.org/fbc3/96b2d3f4fcd3a3844e2e02866992204c6032.pdf>
- [10] Winter on Fire utilized footage from dozens of sources, including cell phones used by protestors during the conflict <https://www.youtube.com/watch?v=RibAQHeDia8>

[11] 2018 Consumer Sales Survey, US only
<https://www.audiopub.org/uploads/pdf/2018-Consumer-Sales-Survey-Final-PR.pdf>

[12] Bambino Cat https://en.wikipedia.org/wiki/Bambino_cat

[13] <https://docs.google.com/presentation/d/1yMApru-GP1iY4NRUL6a5ZJj4z1DwZVbawpftUviMi0/edit#slide=id.p>

[14] Wikivoyage “How to Buy a Kimono in Japan” https://en.wikivoyage.org/wiki/Purchasing_a_kimono

[15] Isler (2018) “The Future of Commons”
https://docs.google.com/presentation/d/15HC6lxwd3mCXXGe0fwzaFH2Wotp_ZQhrgOnWbaCOGek/edit?pli=1#slide=id.g32e28c599f_0_87

[16] Vibber “To-do list for improving video on our project” <https://docs.google.com/document/d/1MDE2j69b0FQwWK-kWPdblQsqB3dT056DEhJu54G0HJw/edit?usp=sharing>

03

Contributors

In order to reach our 2030 knowledge equity goal [1] it will be necessary to diversify the pool of people contributing to the projects, to expand the modalities of engagement and to open the door to new types of content that can be contributed. That is, it must be possible for the next wave of contributors to provide raw text, speech, images, video and other multimedia formats; to contribute new data in a structured way; and to perform discrete editing tasks to improve existing content. This next wave of editors must also be empowered with the tools necessary to storify (assemble and add context to) the raw content uploaded by others. The next wave of contributors must be able to create short form content (e.g. parts of articles) as well as rich articles. We're going to need to make our contributing experiences both cheaper and easier, as well as richer and more complex.

Wikipedia is currently a reading and editing tool, but in the next three to five years these experiences must be treated separately and optimized for their purpose. In this near term future Wikipedia will become a reading and recruiting entry point (driven by search traffic,

and reinforced by citations across the internet) while another more purpose-built experience will be created to optimize the contributors' experience. This new experience must focus on better supporting the myriad of tasks related to uploading, labeling, editing and monitoring contributions.

Finally, we're going to need to make sure that the moderation experiences around the new contribution modalities and types are going to work for existing and new contributors (eg. satisfying "Risks's checklist") [2] ensuring that content creators are happy with the new content, and it meets reader needs for quality and trust

Modalities of Content

To diversify the pool of contributors and perspectives we must diversify the modalities of content we accept. This means that we must “open the aperture” and start managing contributions that are both richer and more complex than existing models, but also cheaper and easier.

With diverse modalities comes the new responsibilities and possibilities for collating and storifying these contributions, thus creating new types of editing tasks, even as other types of content creation (see augmentation et al) reduce traditional editing tasks.

Speech

Speech represents a frontier for developing new and more inclusive modes of access. There is a clear need for tools that support voice consumption and navigation of Wiki content, or, at the very least, augment existing content with audio. New methods like speech-to-text are vital for many use cases and audiences, particularly older users: the typical user of voice assistant applications is a 52 year old woman. [4] Also, input for local languages are not always supported by the small screens and keyboards of contemporary technologies. [5] Finally, oral histories and culturally influential evolving slang lexicons (such as the “swardspeak” of Filipino gay men) represent a frontier on the frontier. These types of aural and oral knowledge introduce new cultural and potentially political complications to our efforts.

Video

Our own research has shown there is demand and need for rich content on Wikipedia. Our commitment to open media formats has held us back for years and we may need to find a way to find a compromise between the open source values of the community and the modern expectations of web users.

Youtube is a surging platform for procedural learning. Research shows using TAM (Technology Acceptance Framework) [3] framework, the user acceptance of this behaviour is sufficiently high enough to call it a leading place for learning. [6] Therefore, Wikimedia should seek partnership opportunities with Youtube or Youtube like service to serve as a potential compliment to open source limited video distribution on Wikipedia projects.

We use Wikipedia as reading and editing tool, but because reading and editing have divergent intentions, processes and target audiences, in the next 3 to 5 years these two tasks need to be more cleanly separated.

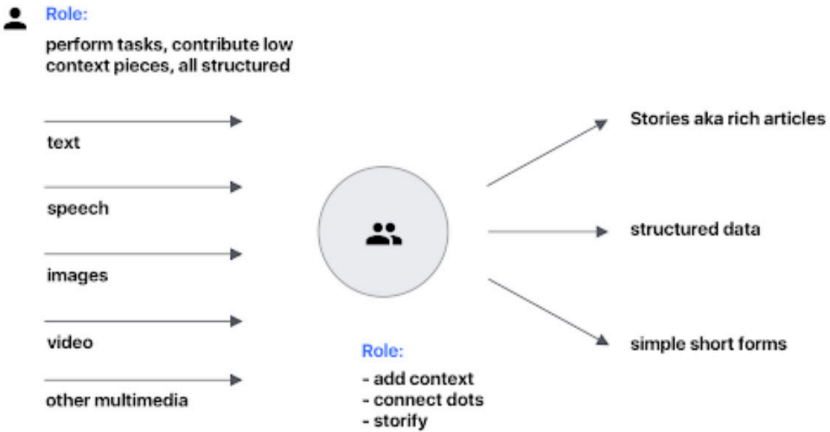
The contribution experience

We will need to design editing (and moderation) tools for the rich media experiences described in the rich content experience section. It seems likely that a new type of contributor as well as tools will be necessary to produce the type of content necessary for an omnichannel experience.

Potential framework

Driven by search traffic and reinforced by robust citations across the Internet, Wikipedia will become an entry point for reading and for recruiting contributors. The necessarily diverse processes employed by these groups will require separate, purpose-built experiences that are supported by machine learning facilitating tasks and allowing user customization.

Not just desktop or mobile options but modes of contribution that are customized to their context. These modes need to provide a mental map of the contribution process to make the knowledge creation process transparent and navigable to all users, new and experienced.



Bret Victor, [7] noted design strategist, observes that current digital software is a medium where you do not manipulate your environment to match your needs. We imagine two possible modes: the Workstation as the experience for veteran editors and the Launchpad for new ones. The Workstation is similar to an integrated development environment (IDE) that doubles down on the community's use of add-ons and specialized tools. The Launchpad is a software environment for newcomers to get help and see what impact they are making.

Both would be built on top of a foundation of structured data (see also the Ubiquity paper) and an associated API, to integrate rich media seamlessly.

The Workstation

The Workstation focuses on productivity, discovery based on customization, and detailed records of interactions with others (interpersonally and in relation to content from a managerial perspective). It should facilitate the creation of pipeline (If... Then... That...) workflows that make use of machine learning to streamline repetitive tasks, build out simple workflows, and aid in discovery

(to highlight issues such as potential bias and suggest citations). A necessary part of increasing the number and diversity of contributors is increasing the tasks and contexts where people can contribute. Therefore the Workstation is about task management, from automatic identification of abusive contributions, vandalism and spam, to guiding the complex process of project development.

The Launchpad

The Launchpad focuses on the immediate presentation of a new editor's impacts, and facilitates discovery centering the editor amid records of their (as opposed to global or community) interaction, attribution and causation. While the Workstation's interface is tailored for co-designed workflows, The Launchpad focuses on workflows that are assigned. Authors should always be able to contribute to discussions, review other edits quickly and easily so they feel connected to the community and the project all the time.

We will need to provide tools that connect other knowledge-gathering activities outside Wikipedia (such as reading, browsing, researching, discussing, taking photographs,

recording sounds, downloading sensor data, etc) and make them available to the process of content creation. For example a “publish to Wikipedia” button on browsers that carries a fact and citation information straight to the article.

The contribution reward

Social networks have succeeded because they variably distribute a deeply evolutionarily compelling reward: approval [8]. Our ecosystem does not give consistent reward for participation in the knowledge creation process nor is there any signposting for how to progress in skill and responsibility.

The community has created some mechanisms to do this but it is not particularly accessible and the look and feel is not in keeping with modern reward systems on other social platforms. We need to retain community ownership but clearly support a first-class model of editing progression and provide mechanisms and rewards that make editing and rewarding a sticky experience.

Notes

- [1] Knowledge equity goals https://meta.wikimedia.org/wiki/Strategy/Wikimedia_movement/2017/Direction#Knowledge_equity:_Knowledge_and_communities_that_have_been_left_out_by_structures_of_power_and_privilege
- [2] user:Riskier “Checklist for content creation extensions” https://en.wikipedia.org/wiki/User:Riskier/Riskier%27s_checklist_for_content_creation_extensions
- [3] Technology Acceptance Model https://en.wikipedia.org/wiki/Technology_acceptance_model
- [4] Boyd (2018) “The Past, Present, and Future of Speech Recognition Technology” <https://medium.com/swlh/the-past-present-and-future-of-speech-recognition-technology-cf13c179aaf>
- [5] Knight, (2016) “Powerful speech technology from China’s leading Internet company makes it much easier to use a smartphone.” <https://www.technologyreview.com/s/600766/10-breakthrough-technologies-2016-conversational-interfaces/>
- [6] Lee and Lehto (2013), “User acceptance of YouTube for procedural learning: An extension of the Technology Acceptance Model” <https://www.sciencedirect.com/science/article/pii/S0360131512002229>
- [7] <http://worrydream.com/MagicInk/>
- [8] Nethercutt (2018), “We’re Primed to Be Addicted to Social Media” <https://zandercutt.com/2018/09/18/were-primed-to-be-addicted-to-social-media/>

04

Customization

Customized and Personalized user experiences are the new norm in consumer products, but are they appropriate for Wikipedia?

A Personalization-focused strategy would conflict with Wikipedia privacy policy, product principles [1] and Movement equity goals, [2] but Customization could contribute to greater usability for readers, communities and editors and for this reason must be considered as part of product modernization overall. In terms of the reading experience, the platform should support a set of user-modifiable customization options and a set of community-modifiable customization options (to allow for language-specific and culture-specific preferences). In terms of the editing experience, the platform should support customization based on common usage patterns and contributor activities at scale, and across whole groups of wiki projects. By making customization options part of the platform, it is possible to serve a more diverse set of needs and preferences without forking the main product. This approach will make scaling much more achievable, and the process of integrating new customization features less dependent upon the technical resources of regional communities.

Customization & Personalization are sometimes used interchangeably but are fundamentally different in terms of user experience.

Personalization is the automatic adaptation of a system to the behaviors and preferences of a user. Because Personalization requires much less explicit input from the user, it is appealing to non-expert users - their experience is automatically tailored without additional effort on their part. But a truly satisfying bespoke experience can only be delivered with a more technically sophisticated system, and requires sign-in, or another form of user identification, and the ongoing collection of user behavior data over time. This approach would be unfeasible and incompatible with the principle of intentional transparency.

Customization refers to the degree to which a user can tailor their use of a product through overt controls and settings. Customization features allow a user to optimize their experience through preferences, defaults and conditionals. But unlike Personalization, Customization requires the user to take action in order to have a more optimal experience and hence is most appropriate in products where users are highly motivated to make such adaptations. Customization tends to result in greater satisfaction among power users, and lower satisfaction among non-power users. [4]

The Reading Experience

Customization For Individual Users

The Customized aspect of consumption experiences may not be differentiating or particularly memorable, but multiple studies [3] [4] suggest customization features are a factor in user satisfaction, enjoyment, and perception of usefulness.

In the context of Wikipedia, it is tempting to treat customization for individuals as a matter of agency (i.e. that it is self-evident that a user should have control over what they are being shown, how information is presented, and how the context they're in affects the modes of presentation). However while basic levels of customization in reading experiences have been shown to result in quick wins, [3] the considerable effort required to implement them may not result in a commensurate impact. So while the product landscape may have set an expectation of agency over aesthetics, accessibility, data control, screen-oriented view modes for comfort, short form/long form reading, implicit topical interests, control of notifications and abuse filters, such settings will not necessarily result in sustained or significantly greater satisfaction with the product.

Customization For Communities

Reading habits and aesthetic preferences are drastically different in different cultures. [6] Communities should be able to customize the reading experiences for their language wikis drastically as well. There is a small subset of contributors who work towards reading

audience, (main page designers, maintainers, template designers) but we need to give more choices and agency over how they present content to their communities.

So while Customization can be a good thing for readers and communities, in an ecosystem with limited resources, the return on investment (in terms of change in user satisfaction) must be considered against the potential hit to scalability.

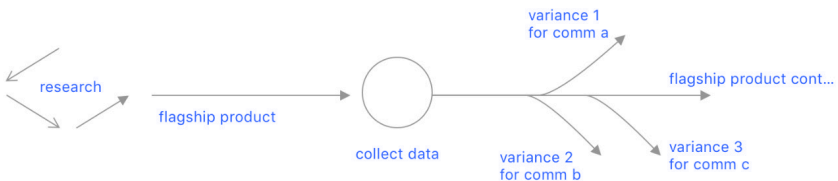
The Editing Experience

Customization For Individual Editors

In spite of the fact that customizing how a system behaves is a superficial change in the medium, it has been shown to lead to greater retention and satisfaction. [4]

Customization for Communities

Similarly, community members should be able to see reflections of their needs, histories and perspectives. In our context, this would require starting with research to understand and drive products that optimize for the “lowest common denominator” solutions. This core solution would continue to develop over time but, in response to community needs and an understanding of its use, support variations and forks. This strategy is summarized in the diagram. The task before us is to identify specific efforts and opportunities for intervention for target communities to see a particular change in that community.



Notes

[1] Product Principles https://www.mediawiki.org/wiki/Product_Principles

[2] Equity Goals https://meta.wikimedia.org/wiki/Strategy/Wikimedia_movement/2018-20

[3] Because Personalization requires a high degree of technical sophistication it would likely only be implemented in communities with a high degree of technical acumen and the developer resources to enable and support it. Sweet and Wirth (2017) “One-to-One Personalization in the Age of Machine Learning” https://books.google.com/books/about/One_to_One_Personalization_in_the_Age_of.html?id=zyM_DwAAQBAJ&printsec=frontcover&source=kp_read_button#v=onepage&q&f=false

[4] Customization leads to perceived ease of use and perceived ease of use leads to perceived usefulness. A user’s perception of control positively affects their attitudes toward the product and the creator of the product. H. Lee and E. Chang

[5] This pattern holds in “low privacy” contexts, but the opposite is true in “high privacy” contexts. Sundar and Marathe (2010) “Personalization versus Customization: the Importance of Agency, Privacy, and Power Usage” <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-2958.2010.01377.x>

[6] Liu, Lee and Lee (2013), “Exploring the Relationship between Reading Habits and Aesthetic Preferences in Different Cultural Contexts and Design Practices” <http://design-cu.jp/iasdr2013/papers/2066-1b.pdf>

05

Discovery

In today's world, content finds the consumer. Over the past decade, social media have had a profound impact on the way people discover content on the internet: time and attention are scarce resources and users have become increasingly accustomed to consuming information surfaced and filtered by friends and family via social feeds. In the current paradigm, information automatically flows toward readers – it simply shows up as part of whatever journey they're on. [1]

“paying active attention to the news was unimportant, because such information was “in the air” as an ambient part of daily life.” [2]

In this way, social media channels have (re)defined the expectations and habits of users all over the world. Users now expect relevant information to find them as a result of their preferences, feed settings and serendipitous browsing. Facebook

[3] and WhatsApp have become primary entry points for new users accessing the internet, and are, for many emerging communities, simply conflated with “the Internet”. For these communities, the page-based mental model [4] of the Internet will effectively never have existed. So to meet our newest users where they are we must recognize the interdependence of form factor and discoverability in a consumption model not driven by search.

Social Media Facilitates Discovery

In a global user study by SDL, [5] when asked “How do you typically discover new and interesting things online?” respondents indicated they turn first to social networks for content discovery, and then to online and customizable newsfeeds. More traditional means, like email and search engines, ranked far behind.

That same study found, on average, millennials share six pieces of content via social media a day, which has overtaken email with five shares a day as the de facto channel for sharing content.

Content discovery and sharing is driven by social media, and if Wikimedia content isn't there, we're missing an opportunity to increase our reach and be part of the global conversation.

This is particularly true in emerging markets, where social media is sometimes the entirety of a user's online experience. This is especially true in emerging markets where Facebook has reached low-income mobile users by partnering with cell networks to provide Facebook access for free (Facebook Zero inspired the ill-fated Wikipedia Zero).

Researchers evaluating how Facebook Zero shapes information and communication technology use in the developing world found that 11% of Indonesians who said they used Facebook also said they did not use the Internet. Essentially, for them, “Facebook is the Internet”

Statistics show the number of shares on social media are down for many types of content, but Wikimedia has an advantage - “evergreen” content

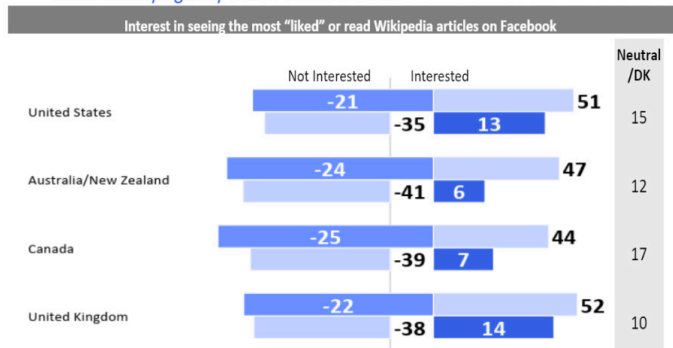
In his Content Trends 2018 report, Steve Rayson, found data showing that “evergreen” posts have resisted shifts in user behavior, tastes, and changes in Facebook algorithms. Despite Facebook share traffic being down overall, evergreen content remains.

According to Rayson, articles that qualify as evergreen have at least one of the following characteristics. They are:

- Research-based
- Reference-style
- Topics that are relevant over time
- Updated regularly

These attributes are perfect, natural matches

Readers are more interested in seeing the most “liked” or read Wikipedia articles on Facebook, with approximately half of readers in all five countries saying they would be interested.

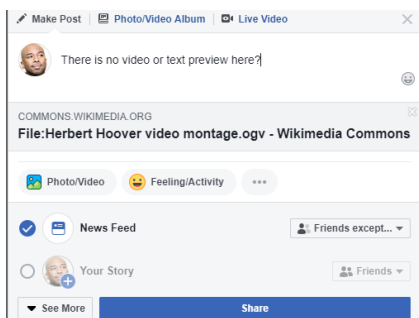


for Wikimedia content and show a clear opportunity for us to gain reach with a strategy for encouraging wider sharing of our content.

“In this new world of content saturation and falling social shares, the big winners are sites that have built a strong reputation for original, authoritative content.

“The majority of content gets zero backlinks but authoritative research and reference content continues to gain links. In particular, authoritative evergreen content consistently gains shares and links over time.”

- Steve Rayson, Content Trends 2018 [6]



Improve discoverability of content

Feeds, top articles, and relevance

On the mobile Wikipedia apps, top stories have become a promising avenue. Jonathan Morgan’s 2017 research [7] on the Top Articles feature in our apps found that on average, raters reported that they would be more interested in reading the articles in the ‘top read’ list than the ‘trending’ list. The results were consistent across groups, and (marginally) significant for India-based raters.

Additionally, recent (though perhaps skewed) research about Wikipedia readers consistently found that around half of respondents were interested in seeing most liked or read Wikipedia

articles on Facebook.

These data points indicate that we’re missing an opportunity to promote popular content at key places, either on-Wiki or on social media. Not everyone is interested in Top Articles, but a lot of people are and we can enhance their experience with functionality that is optional and unobtrusive but very useful for users who want it.

Discovery of Multimedia

Discovery of multimedia content should also be a key component of our future strategy. With a focus on:

- Helping editors find useful multimedia to add to articles
- Helping readers find multimedia related to what they’ve read

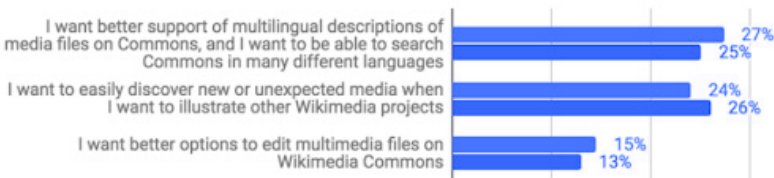
As Ubiquitous as Internet Advertising

Advertising is an omnipresent force on the Internet today. It’s not just tolerated, but expected, and as more people in developed nations become “cord cutters” and move away from traditional TV, internet advertising’s omnipresence increasingly becomes a primary way that people discover new content, products, and news. However, growing trends in consumer trust and content saturation suggest that there is room for a new omnipresent force on the Internet: facts.

Major players like Facebook, Amazon, and Google have already integrated Wikimedia content into their platforms. The Wikimedia experience is slowly being integrated into the user experience of other major platforms, but without our input.

Still, the ultimate goal companies like Facebook and Google are trying to achieve is a good one. Misinformation has become a top consumer

PR30: Which features do you want for Wikimedia Commons?



concern on the Internet. The Reuters/Oxford Digital News Report for 2018 [8] found:

“Over half of those polled (54%) say they are very or extremely concerned about what is real and ‘fake’ on the internet. This is highest in countries like Brazil (85%), Spain (69%), and the United States (64%) where polarised political situations combine with high social media use.”

Modern Internet users aren’t sure what to believe. This is becoming a defining element of the Internet usage experience in modern times. In 2018, Facebook and Youtube began using content from Wikipedia to help combat this problem. [9]

Notes

[1] Like the Burma Shave signs motorists encountered on US roadways in the 1950’s and 60’s, modern day internet users tend to consume whatever information shows up along whatever route they’ve chosen. https://en.wikipedia.org/wiki/Burma-Shave#Roadside_billboards

[2] Toff and Nielsen, (2018) “I Just Google It”: Folk Theories of Distributed Discovery <https://academic.oup.com/joc/article/68/3/636/4972617>

[3] Daily Active Facebook Users <https://www.statista.com/statistics/346167/facebook-global-dau/>

[4] Definition of Mental Model [https://en.wikipedia.org/wiki/Mental_model mental model]

[5] SDL Study “Global Study Finds Social Media Drives Content Discovery with Millennials” <https://www.itbusinessedge.com/slideshows/global-study-finds-social-media-drives-content-discovery-with-millennials-05.html>

[6] Rayson (2018) “Content Trends 2018” <https://buzzsumo.com/blog/content-trends-2018/>

[7] Morgan (2017), “Research:Comparing most read and trending edits for Top Articles feature” https://meta.wikimedia.org/wiki/Research:Comparing_most_read_and_trending_edits_for_Top_Articles_feature

[8] Reuters/Oxford Digital News Report for 2018 <http://www.digitalnewsreport.org/survey/2018/misinformation-and-disinformation-unpacked/>

[9] By “2020, experts estimate 2.95 billion people to access social networks regularly. The majority of this growth is projected to come from mobile devices, as emerging markets catch up on online connectivity.” <https://www.statista.com/statistics/346167/facebook-global-dau/>

Adobe Content Survey (2018) <https://www.slideshare.net/adobe/2018-adobe-consumer-content-survey>

Al-Khatib “Surgical Education on YouTube” <https://pdfs.semanticscholar.org/fbc3/96b2d3f4fcd3a3844e2e02866992204c6032.pdf>

Boyd (2018) “The Past, Present, and Future of Speech Recognition Technology” <https://medium.com/swlh/the-past-present-and-future-of-speech-recognition-technology-cf-13c179aaf>

Burma Shave Signs https://en.wikipedia.org/wiki/Burma-Shave#Roadside_billboards

Connelly (2017), The Drum “Future of mobile apps looking bleak” <https://www.thedrum.com/news/2017/02/27/future-mobile-apps-looking-bleak>

Daily Active Facebook Users <https://www.statista.com/statistics/346167/facebook-global-dau/>

Equity Goals https://meta.wikimedia.org/wiki/Strategy/Wikimedia_movement/2018-20

Gotter (2017) “Desktop vs. Mobile Video Consumption: What You Need to Know” <https://blog.shakr.com/desktop-vs-mobile-video-consumption-what-you-need-to-know/>

Hempel, Wired, 2018 What Happened To Facebook’s Grand Plan To Wire The World? <https://www.wired.com/story/what-happened-to-facebooks-grand-plan-to-wire-the-world>

Huffington Post (2017) “The Media Habits Of Millennials, Generation Z, And The Rest Of Us: In Five Key Charts” https://www.huffingtonpost.co.uk/entry/the-media-habits-of-millennials-generation-z-and-the-rest-of-us-in-five-key-charts_uk_5a149436e4b0815d3ce65ac5?guc-counter=1&guc_referrer_us=aHR0cHM6Ly-93d3cuZ29vZ2xLmNvbS8&guc_referrer_cs=hkCzYDU0_kr_tHiPSqxiAg

Isler (2018) “The Future of Commons” https://docs.google.com/presentation/d/15HC6lxwd-3mCXXGe0fwzaFH2W0tp_ZQhrgOnWba-COGek/edit?pli=1#slide=id.g32e28c599f_0_87

Kalogeropoulos, Reuters Digital News Report (2018) “The Rise of Messaging Apps for News” <http://www.digitalnewsreport.org/survey/2018/the-rise-of-messaging-apps-for-news/>

Knight, (2016) “Powerful speech technology from China’s leading Internet company makes it much easier to use a smartphone.” <https://www.technologyreview.com/s/600766/10-breakthrough-technologies-2016-conversational-interfaces/>

Knowledge equity goals https://meta.wikimedia.org/wiki/Strategy/Wikimedia_movement/2017/Direction#Knowledge_equity:_Knowledge_and_communities_that_have_been_left_out_by_structures_of_power_and_privilege

Lee and Lehto (2013), “User acceptance of YouTube for procedural learning: An extension of the Technology Acceptance Model” <https://www.sciencedirect.com/science/article/pii/S0360131512002229>

Liu, Lee and Lee (2013), “Exploring the Relationship between Reading Habits and Aesthetic Preferences in Different Cultural Contexts and Design Practices” <http://design-cu.jp/iasdr2013/papers/2066-1b.pdf>

Manske, Reasonator <https://meta.wikimedia.org/wiki/Reasonator>

Mirani (2015) “Millions of Facebook users have no idea they’re using the internet” <https://qz.com/333313/millions-of-facebook-users-have-no-idea-theyre-using-the-internet/>

Morgan (2017), “Research:Comparing most read and trending edits for Top Articles feature” https://meta.wikimedia.org/wiki/Research:Comparing_most_read_and_trending_edits_for_Top_Articles_feature

Nethercutt (2018), “We’re Primed to Be Addicted to Social Media” <https://zandercutt.com/2018/09/18/were-primed-to-be-addicted-to-social-media/>

Oates (2018) “Commons: ‘tell us what’s wrong and how to fix it’” https://docs.google.com/presentation/d/1yMApru_GP1i-Y4NRUL6a5Zj4z1DwZVbawpftUviMi0/edit#slide=id.p

Pearson (2018) “Beyond Millennials: The Next Generation of Learners” https://www.pearson.com/content/dam/one-dot-com/one-dot-com/global/Files/news/news-announcements/2018/The-Next-Generation-of-Learners_final.pdf

Product Principles https://www.mediawiki.org/wiki/Product_Principles

Radcliffe (2017) “Five things you need to know about millennial media habits” <https://digitalcontentnext.org/blog/2017/09/22/five-things-need-know-millennial-media-habits/>

Rayson (2018) “Content Trends 2018” <https://buzzsumo.com/blog/content-trends-2018/>

Reuters (2018) “Digital News Report” <http://www.digitalnewsreport.org/>

SDL Study “Global Study Finds Social Media Drives Content Discovery with Millennials” <https://www.itbusinessedge.com/slideshows/global-study-finds-social-media-drives-content-discovery-with-millennials-05.html>

Sundar and Marathe (2010) “Personalization versus Customization: the Importance of Agency, Privacy, and Power Usage” <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-2958.2010.01377.x>

Sweet and Wirth (2017) “One-to-One Personalization in the Age of Machine Learning” https://books.google.com/books/about/One_to_One_Personalization_in_the_Age_of.html?id=zyM_DwAAQBAJ&printsec=frontcover&source=kp_read_button#v=onepage&q&f=false

Technology Acceptance Model https://en.wikipedia.org/wiki/Technology_acceptance_model
Toff and Nielsen, (2018) “I Just Google It”: Folk Theories of Distributed Discovery <https://academic.oup.com/joc/article/68/3/636/4972617>

Thielman (2015) The Guardian “How Fox at National Geographic” <https://www.theguardian.com/media/2015/nov/14/how-fox-ate-national-geographic>

user:Riskier “Checklist for content creation extensions” https://en.wikipedia.org/wiki/User:Riskier/Riskier%27s_checklist_for_content-creation_extensions

US Consumer Sales Survey 2018 <https://www.audiopub.org/uploads/pdf/2018-Consumer-Sales-Survey-Final-PR.pdf>

Vibber, To-do list for improving video on our project <https://docs.google.com/document/d/1MDE2j69b0FQwWK-kWPdblQSQb-3dT056DEHJu54G0HJw/edit?usp=sharing>

Brett Victor <http://worrydream.com/MagicInk/>

Wikimedia Foundation Research: “Characterizing Wikipedia Reader Behaviour/Human development index and Wikipedia use cases” https://meta.wikimedia.org/wiki/Research:Characterizing_Wikipedia_Reader_Behaviour/Human_development_index_and_Wikipedia_use_cases

Wikivoyage “How to Buy a Kimono in Japan” https://en.wikivoyage.org/wiki/Purchasing_a_kimono



Scale

We want the sum of all knowledge to be available to everyone in the world. We also want the process to assemble that knowledge to be inclusive, balanced, and safe for all participants.



01
Community

02
Content

03
Resilience

04
Ubiquity

Published
January 2019

Authors

Olga Vasileva

Adam Baso

Danny Horn

Josh Minor

Contributors

Contributors

Amir Aharoni

C. Scott Ananian

Asaf Bartov

John Bennett

Tilman Bayer

Brandon Black

Arlo Breault

Brian Davis

Dan Foy

Charlotte Gauthier

David A.M. Goldberg

Aaron Halfaker

Ramsey Isler

Jon Katz

SJ Klein

PPS Naryan

Toby Negrin

Jon Robson

Sam Smith



By 2030, up to 90% of the world's population will be using the internet¹ They will bring new languages, new customs, and ways of communicating - and the technologies we have today will have to evolve to account for their needs. The internet will change and the patterns and interactions of today will become less relevant. Some will grow, transform and reinvent themselves. Others will fade into obscurity.

It is difficult to envision a future where Wikimedia projects, in their current form, continue to be essential to the needs of new internet users exposed mainly to social media, short-form text, and multimedia across a variety of platforms.^{2,3,4,5} It is equally difficult to envision Wikipedia restructured as a social network or atomized into a database providing knowledge throughout every corner of the internet while retaining its active readers, communities, and donors.^{6,7}

For years we have established a baseline for quality content for the world's internet population. Yet as the identity of this population changes and our content gap widens, we are found increasingly wanting. If our goal is to increase readership in new markets, or even to provide the content readers are interested in within existing markets, we must focus on not only the size, but also the relevance of our content. An increase in locally-relevant content can not only bring in new readers, but provide them with an opportunity for representation that has so far been sparse not only within Wikimedia projects, but within all media. In addition, allowing communities to create different types of content can make accessibility for a variety of different audiences much easier.

Yet growing relevant content works under the assumption that the supply of content is equal to the demand from readers and, unfortunately, this does not apply to our current structure. Along with our readers, we must grow our communities by focusing on decreasing the barriers between readers and editors⁸ and ensuring new editors have the support they need to begin providing quality content to projects they are interested in.⁹ Additional focus must be placed on the content

itself. While we are not capable of predicting the needs of all of our future users, we can ensure that our content is adaptable to any technical trends that may occur and support our communities by providing them with the tools necessary to create, curate, and moderate such content. We can focus on building relationships between projects and communities so that people looking to find, or contribute to, different types of information can do so with ease¹⁰

One change that may seem inevitable is syndication across other platforms - providing the ability for partners to use our content and for others to access it. We must note that such a future, if implemented without proactive management, can put the sustainability of our communities at risk. Without a steady rate of visits to the site, fewer readers become editors and, over time, the quality of our content will suffer. To account for this imbalance, we can explore the relationship between content creation and syndication and focus on building tools that will allow content creation to continue in an increasingly dispersed network. We can expand our presence on other platforms while continuing to navigate users back to our projects. Success in the aforementioned areas will provide equitable growth to our projects and communities and ensure ubiquity of our content throughout the fabric of daily technology. Yet it does not address our vulnerability to external threats or offer us protection. Protection from such threats must also be treated as a priority. We can explore options such as making censorship and security threats more expensive for those who wish us harm, exploring different ways for accessing our content, and supporting other organizations that stand against censorship.

If we hope to become the "essential infrastructure of the ecosystem of free knowledge" and to allow "anyone who shares our vision [to] be able to join us", we must focus on providing knowledge-seekers with content relevant to their needs and interests, sustainably growing healthy and diverse communities, and ensuring our continued presence throughout the fabric of the internet.

We must also focus on protecting ourselves and ensuring we are resilient to internal and external threats.

Examples

- Structured Data
- Global tools
- External contribution models
- Identifying content gaps
- Platform-agnostic content

Areas of Impact

- All wiki projects
- Community Relations
- Community developers
- Partnerships
- Infrastructure
- Research

01

Community

The Foundation is currently working on features designed to bring more people into our communities. ¹ What will all of these new residents need in order to find their place, and what does the existing community need, to deal with this influx of new strangers? The population of English Wikipedia has famously dropped since hitting a peak ten years ago, going from a high of around 28,000 active editors in 2007 down to about 15,000 by 2013. ² Because this figure has remained more or less consistent since then, it may suggest the number of people the current structure of the site can support. But if the goal is to grow the active population of Wikipedia by attracting and assimilating a large number of new strangers, then a new conceptual model is required- this paper posits a model based on urban planning theory, conceptualizing each contributor community as if it were a city. Such a model would require tools and support to help established residents and newcomers be more visible to one another, and interact. That is, each community must be reconsidered in terms of scale - like a city is organized by neighborhood - in order to ensure that newcomers land in a place they can identify with, among others with similar interests and motivations. Such a

model would ensure greater cohesion within contributor communities, transfer of knowledge between related contributor communities, and the opportunity for more visibility and awareness of the activities of others (both in terms of recognizing positive contributions and policing negative behavior). This new model begins to suggest role structures that are more flexible and nuanced (official, ceremonial, interest-based, activity-based, time-based etc.).

The Problem of Strangers

Growing from a population of 15,000 active editors to 150,000 is like moving from a small town to a busy city. New York City isn't just a small town that got bigger; the scale creates new levels of complexity. In a small town, you see the same people every day, and you can keep track of all the important happenings around town.

But you could walk around New York for a week, and not see the same person twice. In a big city, you're constantly surrounded by strangers, and there's far too much going on to keep track of, which means that there are different expectations around the way that people behave.

For established residents of the community, Wikipedia still feels like a town -- they see a lot of the same people, and they know where all the important meeting points are. But it's a town that's overrun by strangers -- there are 15,000 active editors per month, and around 350,000 people making 1-4 edits.

And for visitors who enter the community by making an edit, it's like stepping into a dark, empty street, where their first interaction might be a stranger jumping from the shadows and bullying them. The newcomer doesn't know if anyone is around to stick up for them, or help them.

Eyes on the street

In a city, both the residents and the newcomers want to feel safe among all of these strangers, and that feeling arises from the natural, active use of the city's streets and sidewalks. In a safe neighborhood, people are sitting on the steps, looking out of the windows, and hanging out in front of the stores. People are less likely to commit crimes or bully people, if there are other people watching.

In *The Death and Life of Great American Cities*, Jane Jacobs identifies this as the key to public safety:

"There must be eyes upon the street, eyes belonging to those we might call the natural proprietors of the street. The buildings on a street equipped to handle strangers and to ensure the safety of both residents and strangers, must be oriented to the street. They cannot turn their backs or blank sides on it and leave it blind." Jacobs, ch 2

This amateur surveillance doesn't need to be organized and explicit; it happens naturally, because people enjoy the sight of other people.

"Nobody enjoys sitting on a stoop or looking out a window at an empty street. Almost nobody does such a thing. Large numbers of people entertain themselves, off and on, by watching street activity." Jacobs, ch 2

But the street needs to be active, in order to be safe:

"The sidewalk must have users on it fairly continuously, both to add to the number of effective eyes on the street and to induce the people in buildings along the street to watch the sidewalks in sufficient numbers." Jacobs, ch 2

In *City: Rediscovering the Center*, William H. Whyte says that the same strategy applies to other public spaces:

"The best way to handle the problem of undesirables is to make a place attractive to everyone else. The record is overwhelmingly positive on this score. With few exceptions, center city plazas and small parks are safe places." Whyte, ch 10

But this strategy works on a local level, street by street. What Jacobs refers to as "the natural proprietors of the street" are the people who

feel some ownership and responsibility -- the people who live on that block, the people who own and work at the local businesses, and the regular visitors who have a connection to the neighborhood.

You can't keep an eye on the whole city at the same time, and nobody really wants to. People like to watch places that are busy, and places that they care about.

Wiki Neighborhoods

The scale of a big city is comprehensible because there are neighborhoods, smaller areas with their own characters and specialties. There's a big difference between Greenwich Village, a bohemian artist's neighborhood, and Wall Street, an international financial center. In these neighborhoods there are different schedules and different expectations about how people talk, dress and behave. The kind of people who feel welcome there will be different. On Wall Street, everyone wears a business suit, and they all go home at 6pm. In Greenwich Village, the neighborhood doesn't really get started until the middle of the afternoon, and things are open all night.

Currently on Wikipedia, the closest analogue to a neighborhood is the WikiProject, topic-based project pages where editors coordinate editing work around a shared interest. Editors join the project by adding their username to a list of members, and active projects organize communal events and work toward shared goals. These should be places where the "natural proprietors" of a topic area can watch people walk by, and perform both functions of the city street -- protecting the neighborhood from bad-faith strangers, and making sure good-faith strangers are treated well.

Unfortunately, for the most part, WikiProjects don't perform those helpful functions, because the project pages are static -- explaining what

the project is and how to get involved, but not providing any signs of activity that would encourage visitors to come back. These are buildings that turn a blank wall to the street, creating empty plazas that don't inspire people to take action.

Dynamic Environments

The active WikiProjects on English Wikipedia go out of their way to make sure that there's new activity to look at, often near the top of the project page:

WikiProject Women in Red[3] has a tally at the top of the page that shows the percentage of biographies about women which is updated each week, and has recent announcements and events listed right under that.

On WikiProject Military history, [4] there's a monthly newsletter [5] that comes out more often (and appears to be better-loved) than the general-interest Wikipedia Signpost.

WikiProject Medicine [6] has a regularly-updated Discussions module, which automatically lists current talk page conversations about pages in the WikiProject's subject area.

Topic areas could become functional neighborhoods that help to solve the problem of strangers, but the people who work in that topic area need a street to watch. There should be topic-specific Recent changes feeds, where people who care about that subject can see the activity happening on the relevant pages. The feed could be based on the existing WikiProject categorization, with new articles added automatically, based on a proposed expansion [7] of ORES machine learning.

Making people more visible

Seeing other people on city streets also helps visitors and new residents to fit into the existing neighborhood. Watching other people helps people understand how to behave in a new environment. If we want thousands of

new strangers to assimilate to the Wikipedia communities, then they need to see other people interacting successfully.

Currently, Wikipedia articles are designed as if the existence of editors was a closely-guarded secret. Besides the button marked “View history”, there’s no way for a reader to understand what editors do on Wikipedia, or how they behave.

The received wisdom is that well-written article pages should look professional and encyclopedic, with all visible signs of human activity tucked away on the talk page and history page. However, there are clues about editor activity on pages with issues -- warning templates explain that “The neutrality of this article is disputed,” or “This article has no lead section.” This means that visitors are only made aware of editing activity when something has gone wrong.

This same urge to make things look clean and professional occurs to city planners as well, as Jacobs points out:

“This last point, that the sight of people attracts still other people, is something that city planners and city architectural designers seem to find incomprehensible. They operate on the premise that city people seek the sight of emptiness, obvious order and quiet. Nothing could be less true. People’s love of watching activity and other people is constantly evident in cities anywhere.”

If we want the active population of Wikipedia to grow by attracting and assimilating a large number of new strangers, then we need to design as if Wikipedia was a city, rather than a small town. We need to help established residents and newcomers to see each other, and interact.

Notes

- 1 Mobile editing tools, more onboarding features, and an easier-to-use communication system.
- 2 The estimates of active editor participation comes from stats.wikimedia, using 25+ edits/mo as the definition of “active”
- 3 https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Women_in_Red
- 4 https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Military_history
- 5 https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Military_history/News/October_2018/Interview
- 6 https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Medicine
- 7 https://meta.wikimedia.org/wiki/Research:Automatic_new_article_topics_suggestion

02

Content

Wikimedia is many things: a software platform, a global movement, a collaborative community. But for the vast majority of our daily users [1] Wikimedia means one thing: informational content. Readers come to Wikimedia (and largely Wikipedia) for many reasons[2], for example to satisfy an intrinsic curiosity, or to become more informed about something they see in other media. But no matter the motivation their satisfaction rests, finally, on one thing: relevant content.[3] Satisfying this need for new users in new markets will be the key to encouraging growth in readership, just as it did in the early growth phase of Wikipedia.[4]

This core user need also aligns with our strategic direction. That is, locally relevant content is not only a potential engine of growth in new markets, but filling these gaps in the content is core to combating the larger inequities in the knowledge that historically has been stored and shared on Wikimedia. By encouraging and enabling new content and topic growth in previously excluded areas, Wikimedia can drive not just growth for its own sake but equitable growth: growing specific audiences and content which have previously not been able or allowed to participate in global knowledge production and distribution.

“Topics about the global south are not as strong in English Wikipedia... [Getting more content in these gaps...] that has an important effect for us as a movement, broadly. the more content there is in Wikipedia that is relevant to people in a certain part of the world, the more likely they are to use it and engage with it. It’s sort of a self feeding cycle.”
[5]

Content and Participation

One significant barrier to wider participation and filling of content gaps with content that is relevant to new readers is the asymmetry between the experience and tools of our current editing community and the reading community whose needs they fill. Currently only about 5% of edits are made on mobile devices. However nearly 60% of our total device access comes from mobile devices. This means there is a disconnect between the way people are writing and curating the content and the way people are consuming it. Although this affects issues like presentation and content form very directly, it also means that the people writing Wikipedia do not reflect the reader population, its context or experience. It is key that we enable participation on the devices and in the contexts where content is consumed. Although our ultimate goal is to satisfy the information needs of users around the world, Wikimedia also relies on a dedicated community to create and most importantly for this discussion, provide quality control for the information. In order to preserve trust and reliability we must also balance the pressure between content growth and quality, and moderation systems. Merely bringing in new eyeballs with click-bait for fake news might create growth, but it undermines the value of that same content. [6] This means that as we encourage new content contributions and the growth of new topics, we must monitor and support the curators and administrators. However, we also must overcome the significant bias and inherent exclusionary nature of certain policies and current community attitudes. Qualitative research [7] and user reports [8] suggest that policies, particularly around notability and reliable sourcing are especially problematic.

Regional Relevance

One way to provide relevant content for many users is to replicate the existing content in their language. This resolves one barrier for users (ie. the content is at least in a language they read and write). However, many many topics of local importance and interest may not exist on any

Wikipedia. Additionally for many users English (or other large colonial language) remains the primary language of the internet and of education more generally, and users expect to search and read about their topics of interest in this global language. This means that we cannot fill these information needs and expectations purely by translating content from large to small languages. It means that we will need to ensure large global wikis like English accommodate and, indeed, encourage a multi-cultural tolerance of difference and variation, and get support for curation tools that enable this tolerance and cooperation.

For a plurality of users, [9] coming to Wikipedia to have your information needs met starts not on Wikipedia but on Google. Their journey begins by searching for keywords. If these keywords are found on Wikipedia, there is a good chance [10] they will see that result and come to us to satisfy their information need. This results in increased readership, which in turn, should result in additional contributors and content growth. Encouraging this virtuous cycle between search, content and knowledge generation applies energy to the flywheel that is at the heart of Wikimedia's content engine. By identifying and filling content gaps, in English and across languages, we add more search keywords that help readers find us. Some of those readers share and contribute, expanding the movement. These new contributors keep our collaborative community diverse and active, adding and improving content. And all this comes back to help satisfy readers needs for the sum of all knowledge.

There are many potential ways to improve and encourage this cycle, including some already underway.

Software interventions:

directing interest with recommendations, improving inclusive interactions

Programmatic interventions:

interest groups, Project Tiger, content campaigns

Notes

1 We count approx 200,000 contributors a month, and roughly 1B devices. The means 99.98% of use is non-contributory consumption. This undercounts as it doesn't account for the readers reached through massive re-publication channels, such as the Knowledge Panel.

2 <https://blog.wikimedia.org/2018/03/15/why-the-world-reads-wikipedia/>

3 “Interestingly, one of the barriers to adoption that this report cites is a lack of local content. “In trying to connect the unconnected to the internet, content has for many years been the forgotten ingredient, with efforts prioritised in expanding coverage and lowering the cost of ownership. These are, of course, fundamental, but so too is the question: is the internet relevant for me?” https://meta.wikimedia.org/wiki/Strategy/Wikimedia_movement/2017/Sources/Considering_2030:_Future_technology_trends_that_will_impact_the_Wikimedia_movement#cite_note-8 pointing to <https://www.gsmaintelligence.com/>

4 https://www.mediawiki.org/wiki/Wikimedia_Audiences/2018_Product_points_of_view/Scale/Content#cite_note-4

5 D.Scott, Lead Organizer of Wikimania 2018 <https://www.youtube.com/watch?v=TTtb4dEypQk> at roughly 22 minutes in

6 There are a number of papers and books which examine the network effect and symbiotic growth between Google and English Wikipedia in the early years of the project. Andrew Lih's is probably the most narrative. Academic version: The Substantial Interdependence of Wikipedia and Google: A Case Study on the Relationship Between Peer Production Communities and Information Technologies

7 New Editors and New Readers research both make the case that learning and understanding policies and the suitability of those policies for other cultures or underserved topics may present significant barriers.

8 In the commentary that follows the quote that opens this document, for example, notability and reliable sources policies are cited as barriers for African participants in English Wikipedia for example. Interestingly Asaf Bartov recently claimed in a related discussion that notability is not the core problem faced by emerging communities, but rather reliable sourcing. In either case these are community policy issues.

9 Search referral traffic is 35% of of daily traffic, per <https://discovery.wmflabs.org/external/>

10 For English Wikipedia our average search position in November of 2018 was 5th (on the first page of results). For Hindi it was 4.3, even higher on the first page of results. Note: numbers are unweighted average across desktop and mobile.

03

Resilience

In order to reach its 2030 goals, the Wikimedia product platform must prepare for rapid scaling of development efforts, contributor participation, and content. In the process it will be critical to design for resilience: the ability to engender sustainable growth and fend off threats. For example, it will be necessary to define countermeasures against external threats such as censorship, misinformation [1], climate and policy related threats, as well as attacks on security or privacy by state actors. It will also be necessary to anticipate and countermand threats that could undermine the projects from within: communities or affiliates turning against one another, communities turning against themselves [2] and communities turning against the Foundation. And finally, perhaps the most critical existential threat is relevance; what barriers to entry can be erected to prevent loss of mind share? What pre-emptive measures

must be taken to guarantee mind share as new communities come online? This paper explores each type of threat and offers a set of economic, cultural, and technical countermeasures. As the incumbent nonprofit internet presence defending a neutral point of view and access for all, it is critical that Wikimedia maintain and strengthen itself to preserve a future with truly free knowledge.

Intro

This paper recommends a number of countermeasures to be supported by the Wikimedia Foundation's Audiences department in order to bolster Wikimedia's resilience. A synthesized version on Officewiki is forthcoming as part of the Audiences department's 3-5 year planning FY 2018-2019.

External Threats

There are four major external threats to Wikimedia:

Censorship

Misinformation, principally from state actors or sophisticated PR firms

Climate- and policy-related disasters

Attacks on security or privacy by state actors

Internal Threats

There are also several ever present internal threats:

Communities turning against each other

Communities turning against newcomers

Communities turning against themselves

Communities turning against the Foundation

The WMF turning against communities

Wikimedia becoming irrelevant

Countermeasures

The following alternatives seek to address a number of the threats listed above. There are not one-size-fits-all countermeasures for the threats, and thus a set of the alternatives would likely need to be applied for a robust defense.

Domain Name Consolidation

Consolidate Wikimedia production traffic under one domain name. This will discourage DNS poisoning and make DNS poisoning and TLS negotiation-based blocking more evident when it does occur.

Give Huggle Hug

Support growth and diversity of the editor ecosystem through targeted product enhancement: adapt (possibly mainstream) tools like Huggle with low BRD (Bold, Revert, Discuss) reciprocation rates. Make these tools

run on additional contemporary platforms, adding features to streamline guidance to good faith editors, with integrated follow-up discussion, and promoting praise of edits going through this BRD cycle. Shepherd appropriately sized coalitions of users focused on the new platform tools and updated approaches.

Decentralized Internet Distribution

Work with key experts and OS and browser vendors to build a secure protocol stack for decentralized distribution that

Ensures availability

Maintains content integrity and recognizable URLs (e.g., Signed HTTP Exchanges)

Shields reading habits from intermediaries (e.g., inbuilt browser tunneling or use of trusted peer nodes)

Shields metrics logging from intermediaries (e.g., opaque out-of-band logging)

Reasonably accommodates protecting readers from outdated reverted material for the common consumption case. [3]

This is in addition to other resilient Wikimedia hosted solutions. Forthcoming enhancements to core protocols (e.g., DNS over HTTPS and ESNI coupled with proxying through critical hosting intermediaries) present additional opportunities to raise the costs of eavesdropping and denial of service.

Cloud Storage and Mirrors

More proactively place Wikimedia dumps on BitTorrent, Github, Gitlab, BitBucket, AWS/S3/Cloudfront, Azure, GCP, Rackspace, Akamai, and Cloudflare. Also foster more mirroring relationships with a global network of universities. Consider coordinating with Google, Cloudflare, and Bing to serve as hosts for AMP as a fallback of last resort in case of widespread system outage or blockade. Apply cryptographic signatures to these distributions.

This would provide redundancy and would create obstacles to censorship while allowing experts to better verify edit histories.

Two Factor Authentication

Add support for two-factor authentication for

all interested users. Holding all other factors constant, this is one of the surest ways to confound a broad class of attacks on security and privacy.

Invest in AI

Consider further investment in AI resources for: Liar, outlier, and bias detection

Machine vision and speech-to-text

Labeling and model tuning

This will be necessary for combatting bad faith state actors and PR firms. It will also be necessary to support a probable influx of multimedia content that needs moderation (and tagging and translation). Product opportunities for high value micro-contributions abound here as well.

Wikipedia All Up

Begin streaming of algorithmic or volunteer curator (or both) selected content via one or more of the following means. Consider a consolidated global Wikipedia brand. Offer language content in one to thirty languages, depending on the format.

Internet radio

Global radio frequencies

YouTube (with permissive syndication)

Multicast for broadcast and cable television

Satellite TV

In addition to reinforcing Wikipedia as a global brand presence and an information utility, this strategy would open the door to further future investment. It would also create an outlet for Foundation and Movement thought leaders to explain how Wikipedia works and why. Channelization raises the costs of censorship at comparatively lower costs of support. Finally, it is an opportunity for forging collaborative user groups for durability and a global brand.

It should be noted this concept could easily be applied in native fashion on various consumer appliances as well, although that is a separate product question.

Structured Markup

Embrace distribution on syndicating and interactive agent platforms, utilizing partnership conversations for bespoke treatments where appropriate. A broader presence not only keeps Wikimedia relevant, it makes suppression harder - for two reasons: (1) when Wikimedia is part of the fabric of life people won't take kindly to it disappearing, and (2) when Wikimedia is everywhere it's technologically harder to suppress. Employ five principles:

Use of structured markup. As specific next steps, (i.) add Schema.org support to TemplateWizard and (ii.) conduct a consultation with the Wikidata and major wiki communities about the Wikidata community modeling templates using Schema.org and weaving that modeling into the non-Wikidata projects (by mainstreaming of Parsoid markup). This is an opportunity to build trust between communities and help define some functional roles for the future.

Ability to measure impact. It's important to know if and to what extent distribution is helping the cause.

Overt branding. This is important for brand presence and enforcement.

Attribution. This is important for compliance and staying true to Wikimedia's values.

Positive contribution feedback loop. Not all distribution platforms will have this capability, but contribution should always be intentionally encouraged, and ideally, co-designed.

Node.js and Python Support

Add support for Node.js, and possibly Python, to Scribunto. Scribunto supports the Lua language, which is not widely used. It should support Node.js, and possibly Python, which has a huge developer following.

Steer volunteer engineering toward:

template (Scribunto) scripting, gadgets, and bots improving MediaWiki Core

This places a higher emphasis on growing content and workflows for the wikis in a more sandboxed fashion while simultaneously making basic MediaWiki more excellent software for collaborative knowledge production (a global

ecosystem form of resilience). Further investment in first class global templates, ideally with a mechanism to fuse data with Wikidata, is complementary. These new technologies are an opportunity to consider more contemporary code contribution workflows.

Fund Anti-surveillance and Anti-censorship Research.

Provide funding to 1-2 reputable anti-censorship / anti-surveillance firms (or fund incrementally internally). This lets more sophisticated forms of distribution and protection be developed.

Summary

The following table is a guide to the countermeasures, how they address the major threats, their relative cost, and how the countermeasures might complement other efforts

Countermeasure	Threats addressed	Cost and horizon	Complements
<p>1. Consolidate to one domain name</p> <p>Content and APIs are all served from wikipedia.org.</p> <p>Censorship of one language is censorship of all, which is costly for censors.</p>	<ul style="list-style-type: none"> → Censorship → Attacks on security or privacy by state actors 	Medium, two year project with cross-functional team at 50%	Brand unification under Wikipedia
<p>2. Give Huggle a hug</p> <p>Get Huggle on Android and iOS. Improve its UX. Invest in productive in-app feedback loops.</p> <p>People work nicely with each other, more editors stay around, the ecosystem flourishes.</p>	<ul style="list-style-type: none"> → Communities turning against newcomers → Communities turning against the Foundation 	Medium, three year project for one team	Making wiki projects bustling neighborhoods
<p>3. Decentralized internet distribution</p> <p>Host Wikipedia in a decentralized fashion with secure tunneling and digital signing.</p> <p>Wikimedia is accessible even when servers are down or blocked. This is in addition to other resilient Wikimedia hosted solutions.</p>	<ul style="list-style-type: none"> → Censorship → Climate- and policy-related disasters → Attacks on security or privacy by state actors → Wikimedia becoming irrelevant 	Big, five year project for one small team with support from several other teams. Incremental milestones.	Eventually, offline editing
<p>4. Database copies to more cloud storage providers and mirrors</p> <p>Digitally verifiable database dumps become more pervasive. It's even harder to erase Wikimedia and its chain of edits.</p>	<ul style="list-style-type: none"> → Censorship → Misinformation, principally from state actors or sophisticated PR firms 	Small, one year project with one additional dedicated FTE	Academic research outreach

Countermeasure	Threats addressed	Cost and horizon	Complements
<p>5. Two factor authentication</p> <p>Anyone who wants it gets the option of two factor authentication. Account compromise becomes much harder.</p>	<ul style="list-style-type: none"> → Attacks on security or privacy by state actors → The Foundation turning against communities 	<p>Medium, 18 month project with three dedicated FTEs and recurring SMS fees</p>	<p>Potentially, scoring and brand positioning</p>
<p>6. AI</p> <p>Investment in liar, outlier, and bias detection; machine vision and speech to text; labeling and model tuning.</p> <p>It's easier to spot the bad guys. It's also easier and more fun for users to interact with and moderate content</p>	<ul style="list-style-type: none"> → Misinformation, principally from state actors or sophisticated PR firms → Communities turning against each other → Communities turning against newcomers → Wikimedia becoming irrelevant 	<p>Large, 5 year project with paradigm shift for Audiences and Technology - varying levels of commitment by team.</p>	<p>Translation, scoring, mobile contribution and AI training, multimedia contribution, oral history</p>
<p>7. Wikipedia All Up</p> <p>Content is streamed online, over the airwaves, and by satellite.</p> <p>People can catch Wikimedia anywhere. Wikimedia is a trusted brand everyone knows will always be there, even for those without computers or smartphones. It's harder to block an omnichannel presence.</p>	<ul style="list-style-type: none"> → Censorship → Misinformation, principally from state actors or sophisticated PR firms → Wikimedia becoming irrelevant 	<p>Medium, 3 year project with small team with escalating brand penetration</p>	<p>Brand unification under Wikipedia</p>
<p>8. Structured markup</p> <p>Using structured markup and partner management, Wikimedia content is further embedded online, with impact measurement and Wikimedia values. Interactive agents automatically rely on the structured markup.</p>	<ul style="list-style-type: none"> → Censorship → Wikimedia becoming irrelevant 	<p>Medium, 3 year project with small team with escalating brand penetration</p>	<p>Granularization of the article, translation</p>
<p>9. Add Node.js and Python support to</p>	<ul style="list-style-type: none"> → Wikimedia becoming irrelevant 	<p>Medium, 2 year project with one dedicated FTE and one code review/tester</p>	<p>Global templates</p>
<p>Templates</p> <p>Would-be template editors no longer need to use Lua for scripting (Scribunto), they can also use programming languages they know and love. This allows a key piece of the ecosystem to grow and thrive.</p>			
<p>10. Fund anti-surveillance and anti-censorship research</p> <p>The next round of privacy tools gets researched and built while we pursue efforts on the current tools.</p>	<ul style="list-style-type: none"> → Censorship → Attacks on security or privacy by state actors 	<p>150K-300K annual investment</p>	

Other Considerations

The following items are efforts the Product team can tackle alone - they represent potential opportunities for collaboration with other departments:

Cooperate with Technology on a continuity plan in case both primary data centers go down for an extended period due to climate or policy disaster.

Explore international governing body action on censorship on the basis of anti-competition (e.g., most blocks have corresponded with unfairly positioned state-supported alternatives) or adverse health and safety externalities (medical information and other critical information has become unavailable). This is a longshot, and the consequences for scrutiny on the content and the positioning as an NGO would need to be considered, but it may provide a defense.

Conclusion & Other Opportunities

The countermeasures preempt future, and in some cases squash current, threats. You'll notice that they are also oriented around the space where the Wikimedia Foundation is uniquely positioned to take action, as these are large and difficult efforts requiring personnel. These recommendations do not yet fully capture the range of discussions or feedback received during late September and early October 2018 as part of the 3-5 year planning process.

Many other potential community or feature interventions have been outlined in other papers in this series, but the following (heavily informed by recent conversations) are examples of how to bolster resilience in various other ways. They principally speak to creating the content and ecosystems that can activate and sustain growth, which is germane to the general theme of SCALE, as well as several other themes.

Abuse filters

Creating spaces to inform editors where there is surging demand or probable surging demand

(based on algorithms) for topics and those topics do not yet meet a particular quality bar.

Encourage multilingualism. Exploring with professors the concept of translation proofreading as coursework.

Ensuring inflows of translations into English Wikipedia and other major wikis.

Investing in generalized work backlog solutions, catered for various personae and form factors.

Emphasizing product experiences for mobile that are catered principally for AI training.

Supporting federated SSO with major social identity providers, and flowing contribution activity back to user social channels.

Scaling analysis of interventions by further integrating with academics in our data analysis.

Partnering with a provider such as Coursera on a free course such as Programming Wikimedia: APIs, Bots, Gadgets, and Template Scripting.

Supporting content snapshots (i.e., branded, perma-linkable, countable, attributed hypermedia fragments) for embeddable content.

This would be a complement to the summary endpoint and context cards.

Principally from state actors or sophisticated PR firms

For example, veteran contributors working against newcomers.

Note: risk concerning potentially infringing content, perhaps avoided by simply obfuscating discovery, needs analysis.

04

Ubiquity

If our goal is to make it possible for the content from Wikimedia projects to be a presence everywhere in the broad ecosystem of new internet platforms and modes of usage—for example voice-driven search—then this content will need to be adaptable, structured, empowered by a unified set of tools, enriched by multiple media and federated for continuous connection to one another. In short, we need to develop a content vision rooted in platform agnosticism. While some of this effort will involve direct internal development, a significant portion depends upon developing a syndication model - where the uses of and dependencies on our content happen through means other than direct access on our website. While Wikipedia content supports the efficacy, trustworthiness and reliability of entities such as Google search and the Dictionary application in Mac OS X, providing access to this content does not always

feed into our own community growth or the financial sustainability of Wikipedia itself. The Wiki projects are tremendously heterogeneous in terms of content, size, usage, rate of growth and degrees of engagement, which presents unique challenges in reconciling regional use patterns and reader demographics.

The goals of platform agnosticism and syndication introduce new challenges in terms of sustainability of contribution, regional relevance and access. These challenges must be met with a combination of adaptive tools and partnerships to provide flexible access to the entire range of content that our current and future readers will require.

Platform agnosticism

“Stroll through Sanlitun, a bustling neighborhood in Beijing filled with tourists, karaoke bars, and luxury shops, and you’ll see plenty of people using the latest smartphones from Apple, Samsung, or Xiaomi. Look closely, however, and you might notice some of them ignoring the touch screens on these devices in favor of something much more efficient and intuitive: their voice.” [1]

The Chinese language, as many other languages, was not built for typing tiny letters on a small screen. But that’s okay because technology, as it usually does given large-enough demand, is making its way around such initial difficulties. In this particular case, the answer might be voice search, alongside AI and new messaging paradigms.

This is just one example of how growing populations are coming to the internet with new needs, new languages, and new modes of expression. What’s certain is that their arrival will change the fabric of and forms in which knowledge is created, shared, and used. As internet access and usage rises in growing economies, the internet will become a more diverse place and platforms will be required to adapt to the needs and motivations of these new users.

If Wikimedia projects want to be participate in this growth and to “break down the social, political, and technical barriers preventing people from accessing and contributing to free knowledge”, we must ensure adaptability to any platform or mode of usage.

Yet predicting trends can be tricky and the risks that have prevented us from being at the forefront of technical innovation so far still apply. Unlike Google, who have the resources to try to do everything-everywhere-all-the-time. We don’t have the luxury or expertise to take large risks, especially if they’re not initiated from

the the ground-up (i.e. from our communities).

For us, ubiquity means adaptation - skipping the guessing game of what will be big in the future, investing in the needs of our current and potential communities to make sure our content is available for use in any future trend and for presentation on any device.

To support the goal of ubiquity we must focus on re-structuring our content so that it can easily be repurposed, remixed and repackaged by us, our communities, or other platforms. For us, structured Wikipedia content could significantly content porosity between our projects over time - facilitating use cases like a reader’s smooth transition from the Wikipedia article on Istanbul, to the Wikivoyage guide, and onward to related media about the city from Commons. Structured content would also support the establishment of global and customizable modular templates for articles, portals and projects. Standardized formats for the subcomponents of these experiences (such as sections, ideas or themes) via well-documented Wikipedia APIs—ie. the means of retrieving knowledge in whole or part—would support both non-Wikipedia platforms and future Wikimedia uses. A more structured content API platform would also make it easier for our diverse communities to generate the tools they feel they need - while maintaining consistent and reliable standards, and that work smoothly across the entire Wikimedia platform (see also Tools: For Developers).

Finally, as we consider this issue of ubiquity at the intersection of Wikipedia and its consumers, structured content would relieve us from the requirement to anticipate, monitor or otherwise be directly aware of how all populations in all emerging economies are developing their own unique relationships to the internet.

Syndication

“Wikipedia content appears to play a substantially more important role in the Internet ecosystem than anticipated, with other websites having critical dependencies on Wikipedia content.”

“Google becomes a worse search engine for many queries when it cannot surface Wikipedia content (e.g. click-through rates on results pages drop significantly) and the importance of Wikipedia content is likely greater than many improvements to search algorithms. Our results also highlight Google’s critical role in providing readership to Wikipedia. However, we also found evidence that this mutually beneficial relationship is in jeopardy: changes Google has made to its search results that involve directly surfacing Wikipedia content are significantly reducing traffic to Wikipedia.” [2]

So far, Wikipedia’s relationship with Google has been fairly symbiotic. We provide a trusted source they can show at the top of the page; they provide an increase in pageviews and, in turn, an increase in donations, in new editors, and in the continued creation of quality content they can then show to users. Everybody wins and information is distributed freely.

Yet exposing more information outside of the site, such as in Google’s knowledge panels, has decreased pageviews to Wikipedia. It is unfortunate that this is an issue. While we still met our goal of providing the information a reader sought without the direct traffic to our sites, we face not only a decrease in funds, but eventually a decrease in quality. Potential editors never see the site, let alone have a way to contribute, and current editors have less motivation to continue writing. Over time, we’re in trouble.

But, so is Google. The study quoted above clearly shows that Google is a worse search engine in a world without Wikipedia. Wikipedia’s importance is so large that the “mere presence of Wikipedia links may have an effect

approximately 80 times larger than the difference between a good ranker algorithm and bad one (for many queries)”. Similar patterns have been found for other online websites such as Reddit and StackOverflow, where Wikipedia content is widely shared.

Thus we find ourselves in an odd paradox where our current level of ubiquity is also a potential threat. One option would be to take a purely defensive stance and work towards preventing any information from usage outside of the site. Needless to say that that goes directly against the free-culture underpinnings of our movement, as well as our licensing. The other option would be to take syndication for granted - to imagine our content spread throughout the fabric of the internet, and shift our content creation and revenue model to such a future.

We need to open or deepen conversations with our partners, to provide them with insights into our side of the relationship. Being able to present them with our perspectives, such as those outlined in the previous sections will make it easier for them to respond to this more nuanced recognition of our interdependency. Larger institutions in particular must be made aware of the financial, legal, trust and cognitive dimensions of a relationship where they are getting a tremendous amount of value for no cost while putting the sustainability of that resource at risk. For example, our partners need to be more aware of information attribution issues: sampling our content without attribution that links to its full context not only negatively affects Wikipedia pageviews but potentially diminishes its functional value (i.e. outside the context of the community that can vouch for/dispute its veracity).

This recognition of the mutual downside is a potential opportunity for deepening our relationship with these high traffic drivers. Just as we now have a process for reviewing and adopting potential affiliate chapters, so too could we institute a model of official corporate affiliation with Wikipedia (e.g. “Google, an official partner of Wikipedia”), that makes that partner an official sponsor of the Open Knowledge Movement, according to some

mutually agreeable terms. A similar concern was voiced in recent research with regard to GLAM partners - that we have no way of bringing them into the fold in an official manner... “we can’t even provide them with a logo to use on their website”.[3] In this way we have the potential to amplify the “building a better world” missions of, for example, Apple and Google... and even to elevate the “don’t shoot the messenger” vibe of Reddit. Providing current and future partners with access to structured content, including contribution actions, via an API would support more symbiotic relationships, and open the door to creating workflows of contribution from other places where our content is used.

Content Relevance

“In the English Wikipedia, articles of strong insufficient quality alone receive close to half of the pageviews, and in the Russian Wikipedia, they receive more than half.” [2]

For our projects to be ubiquitous, we must provide relevant content to all of our users. Not all wikis are the same, nor do they grow in similar fashions and users of different projects have widely varied motivations for reading. For example, our research shows that readers in Western-language Wikipedias are more likely to focus on quick-fact information whereas the speakers of languages in growing economies are more likely to use Wikipedia for deeper learning and for work or school purposes. To be able to cater to the needs of individual wikis or groups of wikis, we must be able to distinguish their needs. Features that might work great on one target audience, might not work for another. Similarly, content that might be notable for a particular community, might not be for another. Focusing on targeting our work to match our unique audiences as well as providing them with the tools to build according to their needs will help us cover the entire range of content that our current and future readers will require. Only by analyzing the needs of readers, editors, and moderators can we address imbalances

in projects which constrain their growth. For example, knowing which Wiki projects may have quality content but low readership, or a high volume of low quality content, would targeted interventions toward more sustainable approaches to growth. Achieving that sustainability will mean assisting projects according to their specific needs. Being able to model the extent of a wiki’s content gaps along with nuances of its editor retention history would allow us to more effectively focus on the factors that limit that project’s ability to scale. We will need new tools to do so, based on a foundation of structured content and communication.

Notes

1. Knight, MIT Technology Review “Conversational Interfaces: Powerful speech technology from China’s leading Internet company makes it much easier to use a smartphone.” <https://www.technologyreview.com/s/600766/10-breakthrough-technologies-2016-conversational-interfaces/>
2. McMahon, Johnson and Hecht (2017), “The Substantial Interdependence of Wikipedia and Google: A Case Study on the Relationship Between Peer Production Communities and Information Technologies” <https://aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15623>
3. Paraphrasing a comment from an AFFCom board member in a Jan 2019 interview for Movement Strategy.

Ban, Perc and Levnajic (2017) “Robust clustering of languages across Wikipedia growth”

<http://rsos.royalsocietypublishing.org/content/4/10/171217>

“Can NC-licensed content be used in the Wikipedia project? – No. Wikipedia contents are being used commercially.” https://meta.wikimedia.org/wiki/Free_knowledge_based_on_Creative_Commons_licenses#8._Can_NC-licensed_content_be_used_in_the_Wikipedia_project?_%E2%80%93_No._Wikipedia_contents_are_being_used_commercially

Creative commons non-commercial license https://meta.wikimedia.org/wiki/Free_knowledge_based_on_Creative_Commons_licenses#8

De Zuniga, Weeks and Ardevol-Abreu (2016) “Effects of the News-Finds-Me Perception in Communication: Social Media Use Implications for News Seeking and Learning About Politics” <https://onlinelibrary.wiley.com/doi/full/10.1111/jcc4.12185>

Dixon (2018) “Why Decentralization Matters”, <https://medium.com/s/story/why-decentralization-matters-5e3f79f7638e>
Econsultancy (2018) “Why we need to stop repeating the “50% by 2020 voice search prediction” <https://econsultancy.com/why-we-need-to-stop-repeating-the-50-by-2020-voice-search-prediction/>

Guazzini, Vilone, Donati, Nardi and Levnajic (2015) “Modeling crowdsourcing as collective problem solving” <https://www.nature.com/articles/srep16557>

Hagedom et. al (2011) “Creative Commons licenses and the non-commercial condition:

Implications for the re-use of biodiversity information” <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3234435/>

Halfaker et. al. (2018)”The Rise and Decline of an Open Collaboration System: How Wikipedia’s reaction to popularity is causing its decline” https://www-users.cs.umn.edu/~halfaker/publications/The_Rise_and_Decline/halfaker13rise-preprint.pdf

Jacobs (1961) “The Death and Life of Great American Cities”, Random House

Jun, Lee and Jeong (2016) “Early onset of structural inequality in the formation of collaborative knowledge, Wikipedia” <https://arxiv.org/abs/1610.06006>

Jun, Lee and Jeong (2015) “Intellectual interchanges in the history of the massive online open-editing encyclopedia, Wikipedia” <https://arxiv.org/abs/1510.06092>

Kittur, Pendleton and Kraut, (2009) “Herding the Cats: The Influence of Groups In Coordinating Peer Production” <http://citeseerx.ist.psu.edu/viewdoc/>



Augmentation

We want the sum of all knowledge to be available to everyone in the world. We also want the process to assemble that knowledge to be inclusive, balanced, and safe for all participants.



01
Content Generation

02
Content Curation

03
Governance

04
Machine Translation

Published
January 2019

Authors
Marshall Miller
Jess Klein

Contributors
Dan Garry
David Goldberg
Aaron Halfaker
Danny Horn
Jon Katz
Josh Minor
Jonathan Morgan
Toby Negrin
Margeigh Novotny
Nirzar Pangarkar



The Wikimedia movement wants the sum of all knowledge to be available to everyone in the world. We also want the process to assemble that knowledge to be inclusive, balanced, and safe for all participants. But there is too much knowledge needed, in too many languages, for humans to do this alone. As an example, if we assume that a Wikipedia that covers a substantial amount of knowledge has 2 million articles (likely a low estimate), and we believe that 300 languages should have access to that knowledge, we should expect there to be 600 million articles. There are currently only 48 million articles¹, which is 8% of the way there. There are simply not enough potential human editors, especially in smaller languages, to get there. Whether or not we believe that long-form articles will be the medium of the future, this illustrates the problem we face.

Augmentation for contribution activities is our path to closing these gaps. Augmentation refers to any technology that helps humans do their work, and wikis have been using augmentation almost since their beginnings: Rambot created 34,000 articles from Census data in 2002², Twinkle has been automating repetitive tasks since 2007,^[3] ClueBot has been reverting vandalism since 2011,^[4] and the Content Translation tool has employed machine translation to generate content since 2016. Over the next three to five years, human editors will need to increasingly wield augmentation tools, especially those that incorporate artificial intelligence, to create content, curate content, and maintain a safe environment on the wikis. Artificial intelligence will not replace human editors -- it will allow human editors to focus on the most impactful and fulfilling work, and, if used correctly, will open up more avenues for more contributors.^[5]

But although artificial intelligence is a powerful editing aide, it also has the potential to powerfully magnify the problems of bias and unfairness^{[6][7]} that already exist in the wikis, and has the potential to discourage new editors.^[8] Therefore, the role of human editors

will change in the future to focus on wielding these tools safely to guard the wiki values that only humans understand.^[5] In pursuing any augmentation technology, we should stick to the principles we apply to code and content: transparency and the ability for anyone to contribute. We should build closed-loop systems that essentially make augmentation “editable” by community-members, even non-technical ones. By making it possible for members of all communities to audit augmentation tools, contribute training data, flag errors, and tailor tools to their wikis, we will ensure that wikis are not unduly influenced by the smaller set of people who build the tools, while also opening up a new avenue of contribution.

In terms of capabilities we need to build, the Wikimedia Foundation should do two main things:

Build an infrastructure platform for many people to contribute augmentation tools.

Provide interfaces that make it possible for non-technical editors to apply, adjust and contribute to those tools.

The former would likely be pursued by the Technology department, while the latter would be pursued by the Audiences department. The Audiences work will create on-wiki tools that allow non-technical editors to record training data, identify errors in existing algorithms, and tune algorithms to fit their wiki’s culture; surfacing those tasks as first-class wiki work that other editors can see. Through these interfaces, the shepherding of augmentation tools will become a new, major way of contributing that will ensure that machines are fair and healthy contributors to every wiki.

Assembling the platform and the interfaces that allow a feedback loop are the most important parts of this strategy -- more important than the particular applications of augmentation. That said, particular augmentation tools will generally fall into three aspects: content generation, content curation, and community conduct. We will need to develop design principles in each

of these aspects that ensure augmentation tools are transparent and editable; and that ensure augmentation respects the boundaries between human work and machine work. These principles should also govern the ways we incorporate augmentation resources from third parties not controlled by the Wikimedia movement, such as machine translation services.

And finally, in order to be successful with this strategy, we will need to continuously recognize and embrace augmentation as a major way to contribute to the wikis. We can do this through community capacity building, holding events, providing training, and encouraging discussion in the community.

Examples

Rambot (content generation)
Twinkle (content curation)
ClueBot (content curation)
SuggestBot (content generation)
HostBot (governance)
Bot approval process (governance)
ORES models in RecentChanges and Watchlist (content curation)
Content Translation tool (content generation)
Article Placeholder (content generation)

Areas of Impact

Wikidata[9]
ORES[10]
Experienced editors[11]
Volunteer developers[12]

Key External Factors

The rate of improvement to artificial intelligence, especially machine translation.[13]

Efforts by other tech companies to automatically translate English Wikipedia, or to otherwise make massive amounts of information available. [14]

The movement's ability to get top talent to work on these issues as staff or volunteers.[15]

01

Content Curation

Content curation is the aspect of wiki activity related to editing, refining, and cleaning up content that has been generated. The Wikimedia movement's ambitious aspiration to make the sum of all knowledge available to everyone in the world means that the movement has a tremendous amount of work to do with respect to making judgments about what information belongs, and how to organize, phrase, and cite it. Most of the hundreds of languages in the world have Wikipedias with less than 10% the number of articles that English Wikipedia has, and even the largest Wikipedias have serious gaps in terms of the depth of their articles, and the subject matter covered by their articles. As all that content gets added, the curation workload will increase beyond what humans are capable of doing.

Augmentation is a potential pathway to curating the massive amount of information needed in the Wikimedia projects. By applying algorithms and artificial intelligence in the right ways, human editors can be aided in making the most important judgments about the content in the wikis, allowing the content to be well-organized and reliable. This kind of human-machine partnership is not new in the wikis. Tools like

Twinkle and Huggle have been helping to automate the tasks of reviewing recent changes and patrolling for vandalism since 2007 and 2008. ClueBot has been independently reverting vandalism since 2011. And in more recent history, ORES machine learning models have begun to surface the edits and pages most in need of attention.

As humans and machines work together to curate content, we can think about that interaction on a spectrum of how much work the human editor does and how much work the machine does. In some scenarios, the machine may just direct human attention to important curation needs. In other scenarios, the human may review tasks completed (e.g. edits reverted) by an algorithm. This paper explores some specific examples of content curation activities that can exist in the future, drawing from all along the spectrum of the human-machine partnership.

Augmentation strategy summary

To meet our movement's goal of making all the world's information available to everyone, we have more work to do than human editors can do alone. We need help in the form of augmentation, which is when humans and algorithms work together. Though augmentation in the wikis is not new, it will be a growing part of their future. To ensure that the contributions made by algorithms are productive, unbiased, and fair, we will need to stick to our movement's principles of openness, transparency, and the ability for anyone to contribute. We should build closed-loop infrastructure and interfaces that allow anyone to contribute new algorithms, and participate in training and tuning them. These principles would apply to augmentation as it is applied to content generation, content curation, and governing interactions between people.

Definition of content curation

Content curation is the activity of editing, refining, and cleaning up wiki content. This is in contrast with content generation, which is about adding new content, and with the interactions and communications between wiki editors (governance).

Aspiration

Making the sum of all knowledge available to everyone in the world is a tremendously ambitious goal. Today even the largest Wikipedias—such as the English wiki—have serious gaps in terms of the depth of their articles, and the subject matter covered by their articles. Assembling knowledge is about more than compilation—it means curation: making judgments about what information belongs, and how to organize, phrase, and cite it. Effective curation makes the sum of all knowledge more accessible, and also makes it more trustworthy. There is going to be too much curation work in

the future for humans to do it unassisted.

Augmented content curation

Augmentation is a potential pathway to curating the massive amount of information needed in the Wikimedia projects. The effective application of algorithms and artificial intelligence can help human editors make important judgments about the organization and reliability of wiki content, and human-machine partnerships are not new. Tools like Twinkle (2007) and Huggle (2008) have helped automate the tasks of reviewing recent changes and patrolling for vandalism. ClueBot has been independently reverting vandalism since 2011. More recently, ORES machine learning models have begun to surface the edits and pages most in need of attention.

As humans and machines work together to curate content, we can think about that interaction on a spectrum of how much work the human editor does and how much work the machine does. In some scenarios, the machine may direct human attention to important curation needs, while in others the human reviews tasks completed by an algorithm (e.g. edits reverted). This table provides some specific examples of content curation activities that can exist in the future, drawing from all along the spectrum of the human-machine partnership.

Content curation strategy

The technical effort behind developing and deploying the human-machine partnership scenarios described above is only part of the challenge. The more important challenges involve defining technical frameworks and design principles that ensure that algorithms are implemented as forces for unbiased and fair curation.

Human-driven content curation necessarily reflects human biases. Because bias and

unfairness already exists in Wikimedia projects, algorithms that learn from human work have the potential to magnify and exacerbate those problems. For instance, one human editor's preference for writing with a certain style might accidentally exclude edits done by members of other demographic backgrounds. The Wikimedia movement should confront this with the same principles that have led to our success in the past: transparency and the ability for anyone to contribute.

Activity	Algorithm role	Human role
Identifying vandalism	Flag edits that are likely vandalism	Review the flagged edits
Identifying unsourced content	Flag parts of articles that make claims that do not appear to be sourced	Review the flagged content and correct or delete
Checking sources	Check sources for the facts cited from them	Review any flagged citations found by the algorithm
Identifying copyright violations	Check edits for whether their contents appear in other sources	Review flagged edits and correct or revert
Grouping tasks	Assemble related individual curation tasks into a prioritized queue	Use the queue to work more efficiently
Routing tasks	Route individual curation tasks to the editors who are most likely to be interested or capable	Receive and take action on routed tasks
Improving article composition	Make automatic improvements to the tone, style, grammar, and spelling of written content	Review automated improvements

Concretely, for algorithms to successfully participate in content curation, the following standards should be adopted:

Algorithms should be able to be created and deployed by anyone. For example, if the creators of algorithms for identifying vandalism are all from the English-speaking world, the algorithms might do a poor job at identifying vandalism in other languages.

The provenance of Algorithmically-produced

work should be clearly indicated. For example, the user who made an edit should be notified if an algorithm reverts it.

Algorithmically-produced works should always include a human in a “closed loop” of editing, improvement and auditing. For example, in the Recent Changes feed, ORES models suggest edits that need attention, but do not automatically take action. Further, there is no way for humans to flag ORES suggestions that are incorrect so that ORES can be improved.

Shepherding, tuning, and training algorithms should be an important wiki role that non-technical editors can take on. Any editor should be able to wield, adjust, and provide data for improving augmentation. This work should “count” as wiki work, as actual edits, and editors should find their way to this augmentation niche. For example, flagging ORES judgments as incorrect should count as an edit.

As described in the overall augmentation theme strategy, the Wikimedia Foundation should do two concrete things to make the above possible:

One.

Build an infrastructure platform for many people to contribute augmentation tools, coupled with Wikidata (or something like it) to serve as a repository of facts.

Two.

Provide interfaces that make it possible for non-technical editors to adjust and contribute to those tools.

02

Content Generation

Content generation is the aspect of wiki activity related to adding new facts, writing, translations, or images to the wikis. The Wikimedia movement's ambitious aspiration to make the sum of all knowledge available to everyone in the world means that the movement has a tremendous amount of work to do with regard to content parity across all wiki projects. Most of the hundreds of languages in the world have Wikipedias with less than 10% the number of articles that English Wikipedia has, and even the largest Wikipedias have serious gaps in terms of the depth of their articles, and the subject matter covered by their articles.

Augmentation is a potential pathway to closing the gaps described above. By applying algorithms and artificial intelligence in the right ways, human editors can be assisted in generating the most important content for the wikis, allowing us to close the most important gaps fastest. This kind of human-machine partnership is not new in the wikis.

Activity	Algorithm role	Human role
Suggesting articles	List which articles should exist but do not	Create the articles listed
Suggesting information	Find new sources, distill facts, and surface the ones not yet in the wikis	Integrate the missing facts into the wikis
Suggesting updates	Identify and flag when information may be out of date	Correct the flagged information that needs to be updated
Initial translation	Create an initial translation of content from one language to another	Improve translations and remove bias
Starting articles	Create the beginnings of articles from sources	Correct and expand machine-generated stubs
Multimedia	Reassemble information from one medium (e.g. long-form article) into another (e.g. visual slideshow)	Correct machine-generated content

Definition of content generation

Content generation is the activity of adding new facts, writing, translation, or media to a wiki. This is in contrast with content curation, which is about vetting, editing and organizing content, and with the interactions and communications between wiki editors (governance).

Augmented content generation

Augmentation is a potential pathway to generating the massive amount of information needed to close gaps in the Wikimedia projects. This kind of human-machine partnership is not new in the wikis. Tools like Rambot (2002) generated 32,000 stub articles in English Wikipedia using Census data, and in 2018, algorithms automatically translate thousands

of articles. On the horizon are technologies like Quicksilver, which detects fact-related entities in news articles and collates them for human editors to turn into articles.

As humans and machines work together to generate content, we can think about that interaction on a spectrum of how much work the human editor does and how much work the machine does. In some scenarios, the machine may suggest a task that the human editor would execute. In other scenarios, the human may edit and improve on algorithmically-generated. The table provides specific examples of possible content generation activities that drawing from the full spectrum of human-machine partnership.

Content generation strategy

The technical effort behind developing and deploying the human-machine partnership scenarios described above is only part of the challenge. The more important challenges involve defining technical frameworks and design principles that ensure that algorithms generate high-quality and unbiased content. are implemented as forces for unbiased and fair curation.

Human-driven content curation necessarily reflects human biases. Because bias and unfairness already exists in Wikimedia projects, algorithms that learn from human work have the potential to magnify and exacerbate those problems.

Concretely, for algorithms to successfully participate in content generation, the following standards should be adopted:

Algorithms should be able to be created and deployed by anyone. For example, if the creators of algorithms for suggesting notable female scientists are all from the English-speaking world, it is possible that the algorithm neglects notable female scientists from outside the English-speaking world.

The provenance of Algorithmically-produced work should be clearly indicated. For example, if an article is primarily generated through machine translation (such as through the Content Translation tool), that characteristic should be clear to readers and editors. This will increase transparency, and potentially encourage human editors to improve the result.

Algorithmically-produced works should always include a human in a “closed loop” of editing, improvement and auditing. For example, in the Content Translation tool, users are required to review and correct the automated translation done by the algorithm.

Shepherding, tuning, and training algorithms should be an important wiki role that non-technical editors can take on. Any editor should be able to wield, adjust, and provide data for

improving augmentation. This work should “count” as wiki work, as actual edits, and editors should find their way to this augmentation niche. For example, if a topic suggestion algorithm existed, editors should have a way to assess the notability of those topics in such a way that the algorithm could improve.

As described in the overall augmentation theme strategy, the Wikimedia Foundation should do two concrete things to make the above possible: Build an infrastructure platform for many people to contribute augmentation tools, coupled with Wikidata (or something like it) to serve as a repository of facts.

Provide interfaces that make it possible for non-technical editors to adjust and contribute to those tools.

03

Governance

In order to meet our movement's goal of making all the world's information available to everyone, we have more work to do than human editors can do alone. We need help in the form of augmentation, which is when humans and algorithms work together. Though augmentation in the wikis is not new, it will be a growing part of their future. To ensure that the contributions made by algorithms are productive, unbiased, and fair, we will need to stick to our movement's principles of openness, transparency, and the ability for anyone to contribute. We should build closed-loop infrastructure and interfaces that allow anyone to contribute new algorithms, and for even non-technical editors to participate in training and tuning those algorithms. These principles would apply to all types of augmentation, whether it is in the aspect of content generation, content curation, or governing interactions between people.

Governance is a word meant to capture a broader scope than "Code of Conduct". It refers to all the ways that people interact with each other on wiki projects, in both constructive and unconstructive situations. Current newcomers rarely contribute past their initial edits because of bad reactions to quality control mechanisms, algorithmic tools

(bots) or policy. We see augmented governance practices as the vehicle that will safeguard, and simultaneously empower the Wikimedia community to become that desired safe haven for knowledge discourse through a set of human-centered principles.

Aspiration

The Wikimedia movement wants the assembly of all the world's knowledge to be inclusive, balanced, and safe for all participants. Current newcomers rarely contribute more than initial edits because of bad reactions to quality control mechanisms, algorithmic tools (bots) or policy. We see augmented governance practices as the vehicle that will safeguard, and simultaneously empower the Wikimedia community to become that desired safe haven for knowledge discourse through a set of human - centered principles .

Augmented governance

Wikipedia wants to attract and retain a person who edits in good faith and has a relatively high quality of edits. Currently the interplay between augmentation and governance has been explored through bots such as HostBot, which welcome new contributors to Wikipedia, and through processes such as the Bot Approval Process. The reception of these machine-generated greetings and process guidance has been initially cold due to impersonal and sometimes ineffective guidance. However, the way forward is leveraging the technology to create a partnership spectrum in which bots and algorithmic tools support and enhance the role of humans within the Wikimedia communities.

This table provides some specific examples of content curation activities that can exist in the future, drawing from all along the spectrum of the human-machine partnership.

Governance Strategy

In the 1971 film *RoboCop*, a picture is painted of a dystopian city on the decline that is dealt with by employing an army of robots to brutally police the city. This is not the future of governance that we want at Wikimedia. The majority of AI systems and related tools that are being created in the world today are being put in place with minimal oversight, few accountability mechanisms and little research into their broader implications. As the table indicates, artificial intelligence is potentially a powerful editing aide, but each advantage is also has the potential to powerfully magnify the problems of

bias, unfairness, and hostility to new editors that already exist in the wikis. Currently, there are no internally agreed-upon methods to measure and assess the social implications of governance augmentation. Therefore, we will need to develop ways to measure, audit, analyze and improve them. There are two ways that we can concretely do this: First, the Wikimedia community can generate a set of guiding principles to ensure that the wikis are authentic and sincere representations of the worlds' knowledge. These principles should focus on human-centered AI, learnability as a core metric and transparency. The Human-centered AI nudges and opens the door to human connection instead of closing it. In so doing, Wikimedia augmentation champions you—your security, privacy, and the quality of your online activity within our tools. Learnability as a core metric means optimizing Wikimedia algorithms for learning and assessing their capacity to do so in terms of how well it assisted in the learning experience. Transparency refers to a commitment to presenting the provenance of any augmentation tool's introduction and use.

Second, developers and administrators will create and utilize open governance algorithms. If anyone can contribute to these tools, and if Wikimedia makes it possible for non-technical editors to watchguard them so that they aren't allocation or representation biased, we have the opportunity to define the role of human computer interaction within the context of governance within the communities that we create and run.

Guiding Questions

What do WE mean by fairness?

How might augmentation help us detect 'fake news' and misinformation campaigns from powerful malicious actors, like nation states?

How can we ensure that individual editors (new, experienced) feel that their contributions are still valued in a wiki where so much is created and curated by machines?

How can machines enhance the governance work done by editors?

03

Machine translation

Machine translation can contribute to the Movement's goal of making the sum of all knowledge available to everyone in the world by making the work done in one language available in others. Such an effort would dramatically accelerate the transmission of information, making more knowledge available and freeing contributors to work on original research. However, third party machine translation by the likes of Google and Facebook represents a threat to our model, potentially promotes English dominance, siphons traffic from our wikis, offers a poor experience, and discourages global contribution. We cannot ignore this change in the Internet's climate, but must adapt to it.

No matter how fast humans create or translate content, we will not be able to create or, crucially, update content faster than a competitor who uses machine translation. We need to adopt a proactive, long-term approach to the widespread use of machine translation that serves our users, aligns with our values, and supports our ecosystem. If we do this correctly, machine translation could represent the ingredient that allows us to realize our vision of global knowledge sharing.

Machine translation is here

Imagine logging onto the internet, conducting a search, and seeing no results in your language. It is hard for English speakers to imagine, but this is the experience of many Internet users today.

Lack of relevant content was cited as one of the top reasons people don't read online by our New Reader's research. However, automated translation is changing how people read, write and relate to each other. In late 2016, Google announced that it was now using neural networks to power its translation, as these had quickly surpassed traditional, algorithmic models. As a result, content translations for over 100 languages are now just 1-click away. As of 2017, Google Translate served 200M users a day and, as of 2018, translates 143 billion words a day. Meanwhile, the Facebook platform now performs 6 billion translations a day.

As impressive as this is, machine translation as it is deployed by third parties such as Google and Facebook is a threat to our model, our philosophy and our traffic. For some time, we have expected Google or someone else to provide translated versions of Wikipedia if we didn't. It is now clear that Google will be implementing a pilot of this in Bahasa Indonesian very soon, serving Wikipedia pages translated into Bahasa Indonesian from their English versions if a native language version doesn't exist.

We cannot ignore this trend, but must adapt to it. We are currently working with Google to ensure that translated pages provide an option to modify the translation and save it as a page on Bahasa Indonesian Wikipedia. However, this short-term solution may lead to static forks of pages that do not update over time because there are no human editors to attend to them or readers that know about them.

We need to adopt a proactive, long-term approach to machine translation that serves our users, aligns with our values and supports our ecosystem.

Short Term Necessity

If Google's pilot is successful, we can expect it to roll out in other languages. Google search users will be offered native language Wikipedia articles that have been machine-translated from the English language version. Internet wide deployment of this strategy would promote the epistemological dominance of English, divert traffic away from our wikis, offer an inconsistent experience and handicap our greatest strength: the invitation to contribute. Ideally Google would let us control this experience, but that is not an option and our licensing does not allow us to enforce it.

Instead, we are working with Google to address the issues of experience and community health. We are asking them to ensure that wikipedia users know that the content is not written by humans and offer them a way to modify the translation for addition to the wiki in their language. In the long-term, we hope they will choose to use articles from other languages when appropriate, but this is not currently on the table.

A Path to Leveraging Machine Translation

Google's efforts will change how people read Wikipedia globally. To take advantage of the opportunity, we will have to adapt our current approach from one in which content translation is a single fork of content, to one in which content flows readily from one wiki to another.

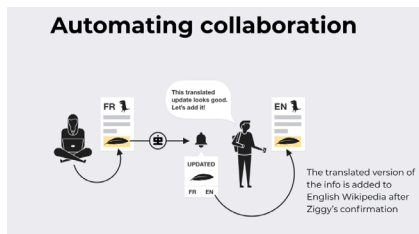
For example, today, the article about Genetically Modified Organisms (GMOs) might be brought from English to Hebrew. However, from this point on, the development of the two articles is forked. In English, several paragraphs might be added about the history of GMOs. In Hebrew, someone might add a paragraph about the economics of GMOs. At this point, neither wiki is benefiting from the scholarship of the other.

As articles change, their counterparts in other languages should have the option to import the new material. For example, editors would be informed about changes to articles in other languages (sections added, facts, etc).

Machine Translation

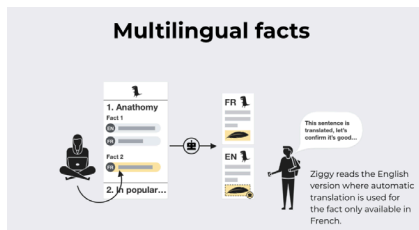
23

Here is a symbolic flow of what that might look like:



(From Pau Giner's "A Multilingual Wikipedia")

Another approach would be to focus on generating facts which can be migrated from language to language via machine translation. Whether or not Wikidata, already a global repository, is used as a semantic storage and mediation platform between wikis is somewhat contested.



(From Pau Giner's "A Multilingual Wikipedia")

In either approach, we are taking the best of augmentation: using machines to replicate existing efforts and bringing in humans to confirm. This will require a shift in the kinds of work that needs to be done and the kinds of people who work on the encyclopedia. For every 1 writer who has a book in their hand and cites it, there need to be 100 other editors whose job is to import the new content into the appropriate place in the destination wikis.

Machine translation will impact different wikis in different ways at different times. An automated translation approach would focus mostly on the wikis for which machine translation is good enough— as measured by our users. Other

segments won't be as immediately impacted by machine translation.

The Glorious Future

The future is already here — it's just not very evenly distributed.

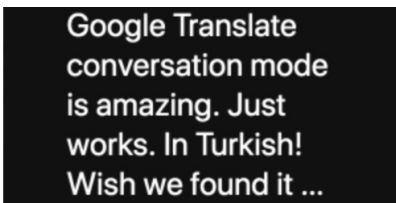
— William Gibson

Machine translation will make possible real-time collaboration across languages in any given digital scenario. Facebook and Google have already built early versions of such technology into their services. Here is an example where a woman was able to share the news of her father's death one-time in Hebrew and have her English-speaking and Hebrew-speaking friends discuss on the same thread:



Screenshot from Facebook, pulled November 1st, 2018

Facebook's next step would be to automatically translate other languages into the language of the reader. This began earlier this year in select places. Google implemented real-time translation of speech in 2015 and made this functionality a key feature of the earbuds they introduced in 2017.[13] A friend travelling around the world just posted this from Turkey:



Screenshot from Facebook, pulled November 9th, 2018

It is possible that we will eventually arrive at a single knowledge corpus, offering a variety of perspectives and written and consumed in many languages. In fact, such a system was recently proposed in a paper by Denny Vrandečić and shared on wikimedia-l.

The idea of a more global Wikipedia was also promoted by a community member in response to our announcement of a new machine translation service being available:

“I believe that if in the future it is possible to write the wikipedia articles in a universal language (I am not referring to English, Latin or something like that, but something for computer), and any change made in any language was visible to all...” This would be a more efficient way to amass the sum of human knowledge, but whether or not this idea is actually feasible will depend in large part on the interests of our communities. The time for that conversation is far away.

Happening Now

Today, Amisha, a biology student in Indonesia, looks up *Cystoseira baccata*, a species of brown seaweed, on Google. There is no Wikipedia article in Bahasa Indonesia for this topic, so Google provides the Wikipedia article in English in the results page.

There is an offer to translate it. If she instead clicks on the link, she is taken to an article in English and her browser may or may not offer to translate it.

There is a Google effort to improve this experience. In the near future, if Amisha conducts the same search, Google will by default offer up a translated version of the English page, hosted by Google with a Google header, yet with of our site’s appearance and branding.

Here is one possible view:

This offering has been tested, and Google has suggested that readers reacted very positively to it. They plan on doing this for any page in which there isn’t a suitable Bahasa Indonesian article. Bahasa Indonesian is the only language they are

currently applying this to, because it is a fairly well-structured and easy-to-translate language. However, it is obvious that Google’s ambitions do not stop at Indonesian or with Wikipedia. Their goal is to make the world’s knowledge available in every language on the web.

Guiding principles

To serve a more adaptive, fluid strategy for our engagement with widespread machine translation, here is a set of principles we stand by:

More knowledge availability is good

Contributions from people all over the world are necessary for capturing valuable perspectives and because a single region cannot capture all the world’s knowledge

Machines are neither good nor bad, but they require oversight.

All wikis should have the chance to export content other languages. Default translations should be derived from the most compelling content from whichever (translatable) language provides it.

Wikimedia Foundation decisions regarding MT need to be informed by what readers and community members want with respect to machine translation usage. Deciding based only on our values or perception has the potential to exacerbate unhealthy global power dynamics.

Risks and Open Questions

We still are many questions about machine translation that we will have to address over the coming months and years. The following section presents a set of risks and mitigations that are significant but over the horizon.

Risk: Disintermediation. Even if Google agrees to provide the users options to circle back to edit articles on their destination wiki, it is possible the mechanism will not be sufficient and local wiki communities will stagnate. Similarly, if users never reach Wikimedia-hosted site, we will not be able to fundraise.

Defense: Work with partners to establish the necessary entry points into our system As

Machine Translation

25

rapidly as possible fill out wiki content using machine translation to augment human effort.

Risk: Machine translation enforces bias. If we continue to rely on third party machine translation services, we are subject to the unknown biases built into those tools. In the case of language, some specific examples include using the male form of “prime minister” by default in gendered languages.

Defense: There aren’t great options here. Work with tool providers to ensure that feedback is registered and to promote transparency. Using Wikidata’s label system might help here. Another response might be to create our own, open machine translation tool.

Risk: Dependence on a third party tool makes us vulnerable. If a fact is created on Wikipedia and then is replicated 100x using machine translation, most of the work is now being done by a third party tool. This dependency means we risk losing the tool at any time and would be vulnerable to the demands of a few key players, such as Google or Yandex.

Defense: There aren’t great options here. Work with tool providers to establish terms up front. Push towards storing content as structured data. Another response might be to create our own, open machine translation tool.

Risk: Low quality machine translation. If machine translation is pushed on users by a third party and creates incomprehensible text or even promulgates falsehoods, we risk harming our users and Wikipedia’s reputation.

Defense: Work with Google to push back where necessary. Research tools and listen to user feedback. We are currently planning an investigation into how machine translation is perceived.

Summary

We need to multiply our efforts on the translation front in three areas:

Real-time translation: Google is using machine translation to provide meaningful default articles where the users’ primary language version is non-existent (or maybe even a short stub), using the English article. Eventually move to some notion of article quality and translation quality to choose the source language. We use default articles as funnel to contribution. “Improve this translation”. A key component of this is working with Google.

Syncing articles: Harness machine translation to improve cross-wiki collaboration. Move from forking articles to syncing articles, Expand the use cases so that contributors do not need to know more than one language to propagate changes from one wiki to another.

Potential Global Corpus: Eventually, we approach a state where there is a global corpus and when someone does the research to expand an existing Wikipedia entry, that research and writing doesn’t necessarily need to be manually distributed in 200 separate places. We are able to examine whether we want different perspectives to be reflected along ethnic or philosophical lines, rather than along language lines (which have, by necessity, served as a convenient, but imperfect proxy for “perspective”).

1 <https://en.wikipedia.org/wiki/Wikipedia:Twinkle>

2 <https://en.wikipedia.org/wiki/Wikipedia:Huggle>

3 <https://www.mediawiki.org/wiki/ORES>

4 <https://en.wikipedia.org/wiki/Wikipedia:Twinkle>

5 <https://en.wikipedia.org/wiki/Wikipedia:Huggle>

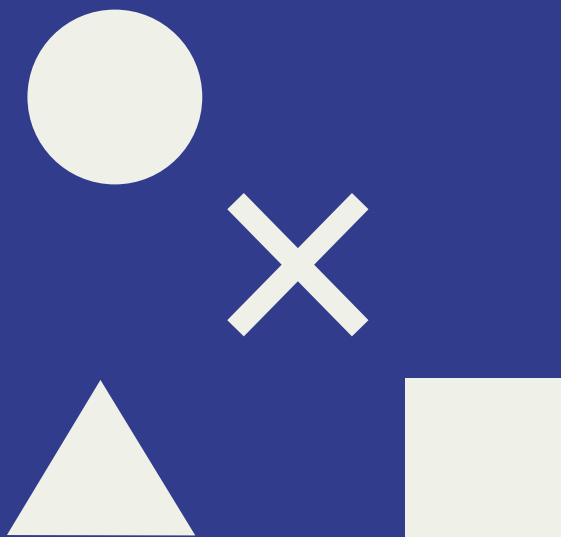
6 https://en.wikipedia.org/wiki/User:ClueBot_NG

7 <https://www.mediawiki.org/wiki/ORES>



Culture

This collection of essays on language, content and inclusive culture are based on insights from the Communities and WMF staff, as well as synthesis of secondary research. These positions will be used to inform product strategy over the next three to five years.



Published
January 2019

Authors

Charlotte Gauthier

Rita Ho

Runa Bhattacharjee

Josh Minor

Contributors

Amir Aharoni

Adam Baso

Volker Eckl

David A.M. Goldberg

Danny Horn

Ramsey Isler

Jon Katz

Toby Negrin

Margeigh Novotny

Nirzar Pangarkar

Santhosh Thottingal



Abstract

In many ways, the theme of Culture synthesizes the movement's strategic directions of Service and Equity. Knowledge as a service means enlisting allies and partners to make knowledge available outside the confines of Wikimedia - and that must include institutions and form factors in which the majority of the world's knowledge still resides, un-digitized and unavailable to a public who might not even know it exists. This means we have to think outside the existing projects and begin acting as a platform for knowledge organization, dissemination and exchange, and as a catalyst for other organizations and institutions engaged in knowledge dissemination. This will require new cultural as well as technical competencies.

Knowledge equity means removing the barriers preventing anyone from sharing what they know, and encouraging "respectful collaboration" between people. However, it is not only historical "structures of power and privilege" that have prevented people from sharing knowledge via Wikimedia. Focusing on specific forms of diversity that come from mainly an American context is likely to hinder rather than help the movement's global inclusiveness. Rather, we need to meet people where they are - in both the literal and figurative senses. A coherent culture strategy must start with awareness, for both consumers and contributors. Research has indicated that awareness of Wikipedia is low in emerging markets,[1] and that the fact that it is editable might actually work against some of our other priorities in the short-term, such as engendering trust.[2] We need, therefore, to find an approach in underserved communities that works - not assuming that people in these communities share our values or that the movement, as it is, can smoothly be applied to new cultural contexts. It is not certain that new users from emerging markets will necessarily share in the movement's goals right away, if ever. If we really mean all the world's knowledge, that means stepping outside our comfort zone.

In addition to fulfilling our mission of making all the world's knowledge available to everyone on our own projects, we also have the opportunity

to make a lasting positive impact on the world's free culture by pursuing (or enabling others to pursue) other societal goals, such as the preservation of endangered languages.[3] Where we can identify areas our movement goals overlap with other free culture efforts, we should partner with the people and institutions pursuing them.

Examples

Content Translation
GLAM outreach and strategy
Structured Data on Commons

Areas of Impact

Most Wiki projects (but especially Wikipedia, Commons, Wiktionary, and Wikisource)
Community Relations
Communications
Mobile (both apps and mobile web)
Community policies and guidelines
Research
Disabled or Disenfranchised Communities

Key External Factors

Between 53% [4] and 71% [5] of the world's population will be online by 2030, with the growth mainly driven by emerging markets. The overall growth rate is slowing [6] due to saturation in developed markets.

Notes

- 1 Movement Strategy Executive Summary of Brand Awareness https://meta.wikimedia.org/wiki/Strategy/Wikimedia_movement/2017/Sources/Brand_awareness,_attitudes,_and_usage_-_Executive_Summary
- 2 Finding from New Readers research in Nigeria, India and Mexico “Trust in Wikipedia is shaken when people find out anyone can edit pages.”
- 3 https://meta.wikimedia.org/wiki/New_Readers/Findings/India#Theme:_Using_Wikipedia
- 4 <https://wikitongues.org/>
- 5 <https://blog.euromonitor.com/2015/04/half-the-worlds-population-will-be-online-by-2030.html>
- 6 The Mobile Economy Global <https://www.gsma.com/mobileeconomy/wp-content/uploads/2018/02/The-Mobile-Economy-Global-2018.pdf>
- 7 USA Today: Smartphone Sales Have Hit a Wall

01

Inclusion

What do “diversity” and “inclusion” mean in the context of software products? Diversity in this context is in reference to having a variety of social stratifications within a collective, namely class, race, sexual orientation, age, (dis)ability and gender. An important aspect of diversity to consider is the variety of ideological viewpoints that inform the types of knowledge considered for inclusion. Inclusion is recognizing the need for greater diversity in Wikimedia addressing issues of content contribution, consumption and policy and decision-making. The web is fundamentally designed to work for all people, whatever their hardware, software, language, location, or ability.

[1] Ultimately, the drive for inclusion of different contributors, beneficiaries and movement leaders must be to ensure diversity of knowledge. That is, broadening perspectives of the various forms of content that we consider as valid and valuable part of the knowledge space, while at the same time ensuring this broad range of content is accessible to all.

Why diversity matters

Besides being one of the core values of the Wikimedia Foundation,[2] diversity is critical to the goal of making all knowledge available to all people, and is an important factor in ensuring high quality content. Direct examples of diversity directly improving knowledge are demonstrated in two studies from 2017. The first study showed that Political, Social Issues, and Science articles whose editors were comprised of more diverse political viewpoints (across Democratic & Conservative) were of higher quality than those with politically homogeneous editor groups. [3] Similarly, a separate second study noted that language-specific topics on Wikipedia are generally better quality on relevant language-editions of Wikipedia, and could be leveraged to improve the knowledge base across language editions.[4]

What's holding us back?

Conceptual barriers/issues

Currently, we talk about diversity and inclusion as an ongoing pursuit, without a, clear end goal. Resource constraints and biases as identified below mean that there is continuing work needed to help recognize and include marginalized groups. But it is hard to reach consensus as to which groups are of higher or lower importance/ need and what goals should be universal (for example, the number of articles of a particular language). Without clear targets for manifesting diversity and inclusion it will be difficult to progress effectively.

Individual biases (conscious and unconscious/ implicit) and systemic bias challenge our capacity to address diversity and inclusion. An example of individual bias would be the notion of notability of BLP [5] articles being assessed based on criteria that often leads to underrepresentation

of certain groups. Systemic biases describe processes and organizational structures that exist within social-cultural groups and lead to decisions that under-represent other groups. And finally, there's a lack of awareness between communities, in the sense of of "peer" communities that use or contribute to Wikipedia (i.e. Group X doesn't know, trust or feel welcomed by Group Y), and those defined by the border between the Foundation and the communities it serves (i.e. Community X doesn't understand how the Foundation itself, or something the Foundation is proposing, is a benefit to them).

Internal organizational/resource constraints

There are gaps in making content accessible to everyone: at a high level, web content is considered accessible when it is perceivable, operable, usable, and robust. [6] Therefore there should be resources that help providers meet Web Content Accessibility Guidelines [7] and tools for readers and contributors with accessibility needs.[8]

There are also gaps in content across languages and many other factors of diversity. This is an area where the Foundation has been reluctant to intervene directly, out of respect for volunteer editorial control. But it is clear that a totally community driven approach, supported only by grants, is insufficient for filling content gaps effectively.

Finally there are constraints associated with human capital. People and their volunteer time are also a limited resource. There are trade-offs to be considered if we are to truly prioritize greater diversity and inclusivity without diminishing or diluting the overall impact of the movement.

External factors

Internet and technology access is inequitable depending on geography, socioeconomic and other demographic factors. This is consistently reported in multiple studies, including a recent

report from the UN noting that less than 50% of the world's population are online as of 2018, and there is a trending decline in internet growth rates.[9]

There are cultural factors as well. Certain communities may not want to participate, share, or be “served” for their own individual reasons and interests. A couple of Wikimania 2018 sessions touched upon some of the reasons Communities fail to engage or disengage - from differences in communication norms (e.g., communities which have a stronger oral tradition,[10][11] distrust due to a history of being marginalized,[12] to fundamental differences in cultural conceptions of what is an appropriate channel for knowledge-sharing.[13]

Reducing barriers to inclusion

Correcting for bias within the Foundation or Community

We must prioritize the closing of gaps in representation at higher leadership and decision-making levels within the Movement. This will help reduce systemic and unconscious bias, but also encourage greater participation from the top down.

We should establish some commonly agreed upon definitions or baseline for measurement. For example, assessing our success in making knowledge accessible to every person is dependent on how we define knowledge, diversity, and what it means to be accessible. [14] Broadening measurement and research means both including currently identified underrepresented voices [15] and understanding what ways they feel excluded from participation, but also continuing to research where there are still existing unidentified gaps in representation.

Improving lines of communication

Besides representation, one of the ways to reduce the conceptual biases of Communities is for the Wikimedia Foundation to facilitate better lines of communication, in general, across groups. Part of this involves continuing research to show benefits of diversity in advancing knowledge quality as well as equity (see above “Why diversity matters” section) and showing these are not conflicting, but mutually reinforcing aims. Another parallel effort is the continuous investment in tools and programs that encourage civil discussions and foster more positive relationships in our communities. This relates to our work in anti-harassment tools,[16] UI standardization's focus on accessibility and initiatives that strive to provide better help and support for new members (e.g. New Readers [17] program, and the Growth Team).[18]

Broadening representation of the movement is not only important in reducing barriers to inclusion but is also a clear signal to newcomers that the Movement is a welcoming place.

Recognizing that Wikimedia has limited resources to provide, and some factors are outside of our control, we should explore strategic partnerships that optimize our reach. Partners might include local experts and embedded members of a particular community with more access to content, people and support; Also, GLAM institutions who may be interested in preserving collections without taking on the burden of maintenance; distribution platforms including search (e.g. Google,) news media organizations (e.g. NYTimes,) and social media sites (e.g. Twitter and Facebook); and finally education groups such as MOOCs and other institutions to adapt Wikimedia content to newer learning courses tailored to helping improve digital literacy in marginalized communities.

Technology and tools

We can utilize technology to both attract more voices as well as help to fill content gaps. Content tailored to newer methods and behaviors of consumption, sharing and learning will enable more people in low reach areas to become aware of and begin accessing our knowledge repository. As noted in “Brand awareness, attitudes, and usage”, [19] awareness and familiarity with Wikipedia is a main area of concern. Some areas where we can invest to increase awareness of the overall Wikipedia brand to broaden reach in low awareness groups (eg. those with limited internet access, younger users) include: optimization and improvement for mobile content in terms of design and delivery and improving sharing via social media platforms such as Facebook and messaging platforms such as Whatsapp, WeChat, and Telegram. Such recommendations address countries like Cambodia where 30% of users access their information via Facebook, [20] and the greater diversity of mobile messaging apps outside of the United States where Internet and technology access are restricted.

Improving contribution tools

It is important to recognize and remember that Wikipedia is a participatory tool. We should invest in new technology and tools that promote and attract new contributing voices. Another recent paper, “The Pipeline of Online Participation Inequalities: The Case of Wikipedia Editing” [21] highlights participation is limited earlier in the ‘pipeline’ of possible editors based on income and racial biases, then by technical knowledge, and finally a gender gap in awareness of its participatory nature. [22]

With that in mind, some broad recommendations for improving contribution tools address issues of accessibility, diversity in technical skills, and the ever-increasing emphasis on the mobile space. Other content consumption platforms

such as voice-interfaces also benefit from simpler interfaces and underlying APIs.

We should ensure that our products conform to accessibility standards and guidelines to work for all people (so, for example, vision impaired users are not excluded from reading and contributing). There are similar overlaps in the needs of in developers, moderators and organizers when we consider mobile compatible tools, and easy-to-navigate systems for recommending content needing contribution or directing new editors to key areas, via, for example “micro-contributions” tools. [23][24]

Programmatic Initiatives

These potential software improvements are one way we can continue to increase contribution, and specifically targeting marginalized or underrepresented communities. But this work must be paired with outreach efforts such as in-person, off-wiki awareness events to recruit newcomers, edit-a-thons to expand content in identified underserved topic areas, and robust community organizing and capacity building programs. [25]

Notes

1 <https://www.w3.org/WAI/fundamentals/accessibility-intro/>

2 https://meta.wikimedia.org/wiki/Values/2008#Wikimedia_Foundation_values

3 <https://arxiv.org/abs/1712.06414>

4 <https://www.mdpi.com/2227-9709/4/4/43>

5 https://en.wikipedia.org/wiki/Wikipedia:Biographies_of_living_persons

6 WCAG “POUR” Accessibility principles - <https://www.w3.org/WAI/fundamentals/accessibility-principles/>

7 <https://www.w3.org/TR/WCAG21/>

8 We are striving to add reader tools as well, e.g. dark mode for reading Wikipedia.

9 Sample, I. (2018-10-18). “Exclusive: dramatic slowdown in global growth of internet access”. the Guardian. Retrieved 2018-10-21. <https://www.theguardian.com/technology/2018/oct/18/exclusive-dramatic-slowdown-in-global-growth-of-the-internet>

10 Wikimedia 2018 presentation: “Wikipedia and Bhutan can learn from each other” <https://www.theguardian.com/technology/2018/oct/18/exclusive-dramatic-slowdown-in-global-growth-of-the-internet>

11 Wikimedia 2018: presentation: “The quotation of oral sources in a decolonization context” https://wikimania2018.wikimedia.org/wiki/Program/The_quotation_of_oral_sources_in_a_decolonization_context

12 Wikimedia 2018 Panel: “Centering Knowledge from the Margins: A Whose Knowledge? discussion” https://wikimania2018.wikimedia.org/wiki/Program/The_quotation_of_oral_sources_in_a_decolonization_context

13 Wikimedia 2018 presentation: “Wikipedia for Indigenous Communities” https://wikimania2018.wikimedia.org/wiki/Program/Wikipedia_for_Indigenous_Communities

14 UNESCO’s Fostering inclusive knowledge societies report particularly focuses on the importance of definitions and measurements for what it means to be inclusive, have access to knowledge, etc; and stresses the need for establishing common definitions so their members states can aim for common goals. http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/internet_draft_study.pdf

15 See Brand Awareness, Attitudes, And Usage - Executive Summary https://meta.wikimedia.org/wiki/Strategy/Wikimedia_movement/2017/Sources/Brand_awareness,_attitudes,_and_usage_-_Executive_Summary

16 https://www.mediawiki.org/wiki/Anti-Harassment_Tools

17 https://meta.wikimedia.org/wiki/New_Readers

18 https://meta.wikimedia.org/wiki/New_Readers

19 See Brand Awareness, Attitudes, And Usage - Executive Summary https://meta.wikimedia.org/wiki/Strategy/Wikimedia_movement/2017/Sources/Brand_awareness,_attitudes,_and_usage_-_Executive_Summary

20 “In 2016 Internet/Facebook became the most important channel through which Cambodians access information (30%) — surpassing TV (29%) and almost doubling radio (15%)” – from “Mobile Phones and Internet Use in Cambodia 2016” http://www.open.org.kh/research/phones_2016.pdf

21 Shaw, Aaron; Hargittai, Eszter (2018-02-01). “The Pipeline of Online Participation Inequalities: The Case of Wikipedia Editing”. *Journal of Communication*. 68 (1): 143–168. doi:10.1093/joc/jqx003. ISSN 0021-9916. (but still available via archive.org)

22 Already identified underrepresented groups are those discussed in the references, including but not limited to: groups in other Western-European regions (per “Geographies of the world’s knowledge”), Women and Non-binary (per “Gender equity report 2018”, and those in areas with limited internet access (per “UNESCO’s Fostering inclusive knowledge societies report”). https://www.oii.ox.ac.uk/archive/downloads/publications/convoco_geographies_en.pdf, https://meta.wikimedia.org/wiki/Gender_equity_report_2018, http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/internet_draft_study.pdf

23 https://www.mediawiki.org/wiki/Mobile_design/Micro_contributions

24 https://www.mediawiki.org/wiki/Reading/Readers_contributions_via_Android

25 Hargittai, E., & Shaw, A. (2015). “Mind the skills gap: The role of Internet know-how and gender in differentiated contributions to Wikipedia.” *Information, Communication & Society*, 18(4), 424–442. doi: 10.1080/1369118X.2014.957711

02

Language

This document is an exclusionary act. Conceived and written in English by highly literate speakers of English, the world's most dominant and well-represented system of knowledge. But Wikimedia's vision aims to be for all people, in their language, calling for us to break down the exclusionary systems of knowledge creation and access. To do this, our systems, software and governance will have to evolve away from an English first, others maybe model, to a system that not only supports all languages but empowers them. For 60% of the world [1] using multiple languages, for different purposes, is the normal way to live. Pidgins and inter-languages produce knowledge too, so how must the current one-project-one-language model adapt to support evolving languages such as Hinglish and Spanglish? And while capturing and growing

knowledge in all languages is an uncontroversial goal, it is at these fuzzy edges that we can see, for example, the unintended exclusionary potential of technologies such as machine translation. As isolated cultural "space" collapses, a tension arises between support for cultural uniqueness and support for the experience of living between languages. And finally, language acquisition and cultural adaptation will be a huge factor in the foundation's success – what new capabilities does this demand of our organization and of the movement as a whole?

Privilege

Recognizing and Addressing our English first model

How does culture (both Foundation/Wiki community culture and the cultures of the underserved language communities) play into what languages and/or topics are “marginalised”? As a first step toward answering this question, the Movement must develop a pragmatic model for mapping similarities and differences between language cultures. Only with such a model can there be understanding of where the common ground exists between language cultures. Languages, and the facility with which one moves between them, may not be the connection in all cases, but can be a catalyst.

It is understood that there will be predominance of some languages due to various factors like demographic majority, regional prevalence, social prestige, or academic needs. However, isolated cultural spaces are collapsing thus more languages are going to co-exist in the same space. This will need more case based exploration.

Where to Focus?

Scoping Our Approach

Language is a vast source of information and influence. Small and large impacts arising from languages are happening around us continuously (from misuse of words that give rise to misunderstanding, to re-purposing or invention of words to describe new cultural phenomena, to policy decisions about what languages will be used in education). There are many variables that factor into how language use changes over time, and tracking all of them would be a heavy investment without obvious return. We must therefore identify the key factors that have the strongest relationship to knowledge gathering (e.g. language population migration) and develop mechanisms for monitoring these factors over time.

For the purposes of our our strategic planning it will be necessary to identify a few areas of high

potential impact, and use these to center our technical and social interventions.

Priorities

How do language interventions get prioritized? What should drive this?

Language interventions must be prioritized in response to social factors. Tracking cultural shifts due to economic migration and other global factors will be key to understanding the adaptive linguistic uses that will impact our platform. Forced migrations from conflict and climate change impact zones, and economic migrations for subsistence and specialized jobs. Related to migrations are efforts at cultural preservation triggered by cultural overlaps, friction and intermingling.

Improved access to education, increased digital literacy, economic improvements and the expanding reach of technological and digital communication services and devices also contribute to linguistic transformations and pressures.

Interoperability

Multilingual use patterns

The Wikimedia platform must begin to support context-switching between languages (according to context of use). Code switch(es), pidgins and inter-languages produce knowledge too, so how should our model of one project/one language adapt to a world of Hinglish and Spanglish?

Language and culture are intimately related. In multilingual societies languages come to serve specialized purposes, and some types of knowledge are associated with a particular language. For example it is not uncommon for there to be a language for science, a language for government and a day-to-day spoken language, in which mass media and popular culture are conducted. The idea that each language will have a complete vocabulary and source material to build an encyclopedia that looks like English's doesn't stand scrutiny.

The model of language around which the projects are built, with parallel encyclopedias and dictionaries for each language, and some set of shared content that is available across them all, understates the complex way people use languages. For 60% of the world [2] using multiple languages, for different purposes, is a the normal way to live.

(How) Can we create content that is consumable in multiple languages, or doesn't require language at all to enjoy and use? Perhaps new formats are needed.

Support

How do we provide not just content, but also support in marginalised or underserved languages? What happens when a contributor who only speaks e.g., Yoruba or Tibetan has a policy question? Participation demands support, especially on the long tail.

In parallel, we need to continue pushing forward Unicode font support and basic text entry (reference-- Minh, Odia, etc.), to ensure the most basic forms of access for all languages.

Opportunities & Threats

Machine translation is making the boundaries between languages softer, as are Wikidata and Commons. So how do we preserve the cultural values and self-determination of each speaker community in a world where these communities and languages are becoming more porous and mixing more than ever? How can we avoid reinforcing language inequalities, or worse, promote language extinction, when we use technology (machine translation, for example) to fill our content gaps?

Content translation is crucial, but insufficient. Language and culture are deeply bound to categorization, notability and content expectations. Simply translating all articles between all pairs of languages would not make a relevant or even comprehensible experience for most humans.

Advocacy

Societal Goals

How can the Wiki movement contribute to larger societal goals? Should it? Is there a way that the

Foundation can facilitate language teaching and learning?

Otherness

Should the Foundation take on the role of extracting and cataloguing knowledge from languages that don't have a writing system? Can our intervention could be at the level of the platform - for example a DIY kit for capture?

"Since the beginning of time, every culture has developed means of passing on important information to its people. For Hawaiians, there was no written language per se until the 1820s. The missionaries introduced the alphabet which made it possible to represent Hawaiian language in the written form. Until then, all information was passed orally through the use of songs, chants, and poems.

Hawaiians devised various methods of recording information for the purpose of passing it on from one generation to the next. The oli was one such method. Elaborate chants were composed to record important information, e.g. births, deaths, triumphs, losses, good times and bad."[3]

Preservation

To what extent should the Foundation advocate for or orchestrate the preservation of endangered languages? One possible route is to pursue partnerships, another is to develop educational materials in non-endangered languages that are spoken by many people, but are rarely used for education and business.

Notes

1 Categories of Multilingual <http://ilanguages.org/bilingual.php>

2 <http://ilanguages.org/bilingual.php>

3 Hawaiian Oli <https://apps.ksbe.edu/olelo/learning-place/performance-indicators/chant/foundational/significance-oli-chants-in-hawaiian-society>

03

Content Gaps

Are we there yet? Making the sum of the world's knowledge accessible to all the world's people presumes that the extent of knowledge can be described, [1] measured, [2] and tracked. [3] It also implies that content gaps can be identified and filled with context-relevant takes on the missing knowledge—this may be the most difficult part of the problem. For our purposes, gap knowledge can be said to fall into one of these general categories: potentially portable knowledge, [4] contextually nuanced knowledge [5] or contextually anchored knowledge. [6] The three types require different intervention strategies, and may have different relative priority for a given a regional community. For this reason, intervention strategies (product features, global tools, community mobilization) must be developed in parallel.

Solutions for filling content gaps also surface new challenges around notability and

verifiability[7]—how must current standards for authority evolve in order to encourage the growth of each type of gap knowledge? And finally, assuming that it is possible to close the content gap through various means, what support mechanisms will be required to maintain this body of knowledge? It will be necessary to have tools that monitor and signal the freshness of content to the community best suited to maintain it.

Making all knowledge available to all the world's people will demand the organization and participation of all the world's peoples, and interventions deployed at pace with the rate of growth of the overall pool of world knowledge.

Identifying Gaps

How Do We Identify Content Gaps And The Appropriate Ways To Fill Them?

What is “the sum of all knowledge?” Is it a thing? Does it concretely exist? Is the sum what results from everyone contributing or is it a matter of circumscribing an area? And in either case, how does one measure progress?

Content gaps may exist for a number of reasons that go beyond the well-known factors of institutional bias or lack of internet access. Knowledge may be held by a group or groups of people who are entirely unaware of Wikipedia’s existence; and if they are, they might not know that it is editable, or understand its relevance to them. People may have specific cultural, religious or other taboos against sharing specific knowledge (such as what they know about religious rites). They may also have ingrained cultural or personal beliefs that community-generated knowledge is less valuable than institutional-generated knowledge.

Filling-in content gaps goes beyond identification of the gap and the creation of initial content. It also needs to include finding and building audiences for contributed materials, and identifying and supporting contributors who want to maintain and expand the initial content. In both cases the risk is that content will stagnate and result in a loss of audiences and creators.

It will not be possible to fill all gaps immediately, and there are some gaps that may never be able to be filled. Prioritisation is therefore necessary based on the characteristics of potential audiences (including size, language, internet penetration and awareness), on the size of the potential contributor community, and on the notability of the topic in that particular language. Care should be taken with finding automated ways to fill the gaps, so that we do not inadvertently reinforce inequalities (see Language section, for example).

How Do We Bring Knowledge Out Of Legacy Media And Institutions And Onto The Web?

There is a vast amount of knowledge locked away

in legacy media that has never been digitised. These heritage materials are not available online for multiple reasons, including ignorance, copyright restrictions, and the degree to which the information is considered “specialized” or “esoteric.” Such cultural constraints are combined with technical ones, including the difficulty or impossibility of digitising with current technologies due to intangibility, deteriorated media or anachronistic communication styles (e.g. pre-19th century typefaces and manuscripts).

In many of these cases it may be a better strategy to forego rights acquisition efforts and digitising the content ourselves, and instead collaborate with or catalyze entities that are already pursuing digitization efforts such as libraries, archives, museums and dedicated online projects such as Gutenberg and Internet Archive.

Institutions (universities, governmental bodies, etc) also have huge amounts of knowledge that is not available on the web.

We might also partner with academics to bring their research to the public (such as by adding citations of published material to articles,) advocate for digitisation of government materials where governments are not already doing this, and advocate for meta-knowledge: knowledge of how to acquire knowledge.

Evolving Notability Standards

(How) Do We Manage Types Of Knowledge That Don’t Conform To Our Current Standards Of Notability Or Published Authority?

In order to make Wikimedia the repository of ALL the world’s knowledge, it would be necessary to consider how various alternative forms of knowledge can fit in with existing or new Wiki projects beyond Wikipedia. There are types of knowledge (such as oral history) that by their very nature do not conform to Wikipedia’s standards of notability or published authority. In such cases, Wikimedia should consider acting as a catalyst for organisations that are gathering and digitising these forms of knowledge, rather than attempting to reinvent the wheel—especially

when those organisations already have an existing body of work and useful contacts that would take us a long time to build up.

Incorporating Other Formats

(How) Do We Move Beyond Primarily Text-based Content And Into Other Formats Such As Video, Audio, Images, Etc?

Answering this question begins with defining the purpose of this type of content: is it to add to existing Wikipedia articles, encourage social media sharing, or serve as a destination unto itself? As we look to other sites that have embraced video we find that short-form video content is popular on social media and newspaper sites for example, but rarely repays its costs. On the contribution side, Instagram and YouTube demonstrate that people want to provide and share video, but the question remains as to whether this interest will extend to providing Wikipedia with such content. The approval and review practices that are a part of Wikipedia's culture, along with those of maintenance and updating, are different from the culture and performance of social media. Issues of consumption and contribution are both measured against tooling changes and process changes

Supporting New Platforms

How Can We Make The Tools For Filling In Content Gaps Available On The Devices People Are Likely To Need Them On?

Many of the most urgent content gaps are in languages and about topics relevant to places outside Europe and North America. Internet penetration is growing in these countries, though the majority of users access the web on mobile devices. Addressing content gaps therefore means thinking outside the desktop editing paradigm. Therefore we will have to consider several factors as we imagine expanding Wiki into these areas: Mobile-friendly tools for content creation, translation tools, and less data-intensive tools for rich media creation.

Notes

1 T. Negrin: "The Static Knowledge Pie": using the sum of all topics currently covered on all wikis as a measure of completeness, the content gap is measured as topics existing on some projects and absent on others. Requires an exhaustive but extensible topic model.

2 T. Negrin: "The Potential Knowledge Pie": The Static Knowledge Pie with the addition of known gaps across all wikis (e.g.estimated number of articles there should be about female scientists).

3 T. Negrin: "The Dynamic Knowledge Pie": the diameter of the pie increases over time making the content gap a dynamic function of the relationship between rate of knowledge created in the world and knowledge captured on the system.

4 Potential Portability reflects the degree to which certain classes of knowledge have the potential to be "ported" or translated with low loss in fidelity. Scientific terminology and topics are an example of PPK. Certain types of legacy media (e.g. illuminated manuscripts) that can be ported from one medium to another could also be considered PPK. Transferring Potentially Portable topics and formats is challenging, but methods for tackling them are known unknowns.

5 Contextually Nuanced knowledge is a category of knowledge that requires a higher degree of human input to transfer (e.g. subject matter expertise, topical knowledge and interpretation). The history of the Korean War written from a Japanese, or American or Korean point of view would be an example of CNK.

6 Contextually Anchored knowledge is a category of knowledge that is utilized by a specific community in a specific way. This knowledge can be described in other contexts, but not utilized in a comparable way. The Hawaiian Oli chant is an example of CAK. <https://apps.ksbe.edu/olelo/learning-place/performance-indicators/chant/foundational/significance-oli-chants-in-hawaiian-society>

7 Decline of editor retention in mature language communities on Wiki being (partially) caused by lack of welcome for newcomers <https://drive.google.com/file/d/1FzMCptTMOvvRRGQPyh2AsBrGneOIG3F/view>

- Article coverage by language
 “Increasing article coverage” https://meta.wikimedia.org/wiki/Research:Increasing_article_coverage
 “Growing Wikipedia across languages” <https://drive.google.com/open?id=1Fp4ktS3tzZDbfXuxraELk8EuCRb4W4dh>
 Content translation recommendation API https://meta.wikimedia.org/wiki/Recommendation_API
 Expanding Articles Cross-language” https://meta.wikimedia.org/wiki/Research:Expanding_Wikipedia_articles_across_languages
 “Brand awareness, attitudes, and usage - Executive Summary” https://meta.wikimedia.org/wiki/Strategy/Wikimedia_movement/2017/Sources/Brand_awareness,_attitudes,_and_usage_-_Executive_Summary
- Cisco Visual Networking Index: Forecast and Trends, 2017–2022 <https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.html>
- “Cultural Issues Hindering Content Creation” from New Voices synthesis: [https://meta.wikimedia.org/wiki/Strategy/Wikimedia_movement/2017/Sources/New_Voices_Synthesis_report_\(July_2017\)#Challenges_for_Wikimedia](https://meta.wikimedia.org/wiki/Strategy/Wikimedia_movement/2017/Sources/New_Voices_Synthesis_report_(July_2017)#Challenges_for_Wikimedia)
- Euromonitor <https://blog.euromonitor.com/2015/04/half-the-worlds-population-will-be-online-by-2030.html>
- Gapfinder <https://www.mediawiki.org/wiki/GapFinder>
- “Geographies of the world’s knowledge” https://www.oii.ox.ac.uk/archive/downloads/publications/convoce_geographies_en.pdf
- “Gender Equity Report” http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/internet_draft_study.pdf
- Google-KPMG, ‘Indian Languages - Defining India’s internet’ <https://assets.kpmg.com/content/dam/kpmg/in/pdf/2017/04/Indian-languages-Defining-Indias-Internet.pdf>
 Halfaker , Geiger, Morgan and Riedl (2018) “The Rise and Decline of an Open Collaboration System: How Wikipedia’s reaction to popularity is causing its decline” https://drive.google.com/file/d/1FzMCp_tTMOvwwRRGQPyh2AsBrGneOIG3F/view
- Hargittai, E., & Shaw, A. (2018). “The Pipeline of Online Participation Inequalities: The Case of Wikipedia Editing”. *Journal of Communication*, Vol. 68, Iss. 1, 1-Feb-2018, pp143–168, doi. [org/10.1093/joc/jqx003](https://doi.org/10.1093/joc/jqx003)
- Hargittai, E., & Shaw, A. (2015). “Mind the skills gap: The role of Internet know-how and gender in differentiated contributions to Wikipedia”. *Information, Communication & Society*, 18(4), 424–442. doi: [10.1080/1369118X.2014.957711](https://doi.org/10.1080/1369118X.2014.957711)
- Hawaiian Oli Chants <https://apps.ksbe.edu/olelo/learning-place/performance-indicators/chant/foundational/significance-oli-chants-in-hawaiian-society>
- “Knowledge Gaps” <https://research.wikimedia.org/knowledge-gaps.html>
- Language Extinction:
 Wikitongues <https://wikitongues.org/>
 Foundation for Endangered Languages <http://www.ogmios.org/manifesto/index.php>
 Rosetta Project <http://rosettaproject.org/>
 Living Tongues Institute <https://livingtongues.org/>
 UNESCO linguistic diversity on the internet project <http://www.unesco.org/new/en/communication-and-information/access-to-knowledge/linguistic-diversity-and-multilingualism-on-internet/>
 BBC: “Why We Must Save Dying Languages” <http://www.bbc.com/future/story/20140606-why-we-must-save-dying-languages>
 T.Skutnabb-Kangas, (2013) “Linguistic Genocide in Education” https://books.google.co.uk/books/about/Linguistic_Genocide_in_Education_or_Worl.html?id=sGtObjnBydQC&redir_esc=y
 Lewoniewski, Węcel and Abramowicz (2017), “Relative Quality and Popularity Evaluation of Multilingual Wikipedia Articles” <https://www.mdpi.com/1099/13/1/1>

All Sources

com/2227-9709/4/4/43

“Microcontributions” https://www.mediawiki.org/wiki/Mobile_design/Micro_contributions

“Mobile Phones and Internet Use in Cambodia 2016” http://www.open.org.kh/research/phones_2016.pdf

New Voices Synthesis: Attracting New Users [https://meta.wikimedia.org/wiki/Strategy/Wikimedia_movement/2017/Sources/New_Voices_Synthesis_report_\(July_2017\)#How_to_attract_new_users](https://meta.wikimedia.org/wiki/Strategy/Wikimedia_movement/2017/Sources/New_Voices_Synthesis_report_(July_2017)#How_to_attract_new_users)

“Readers Contributions Via Android” https://www.mediawiki.org/wiki/Reading/Readers_contributions_via_Android

“The Rise and Rise of Hinglish in India” <https://theconversation.com/the-rise-and-rise-of-hinglish-in-india-53476>

Sample, I. (2018-10-18). “Exclusive: dramatic slowdown in global growth of internet access”. *the Guardian*. Retrieved 2018-10-21. <https://www.theguardian.com/technology/2018/oct/18/exclusive-dramatic-slowdown-in-global-growth-of-the-internet>

Shi, Teplitskiy, Duede, and Evans (2017) “The Wisdom of Polarized Crowds” <https://arxiv.org/abs/1712.06414>

“The State Of The Article Expansion Recommendation System” (Research showcase) https://www.mediawiki.org/wiki/Wikimedia_Research/Showcase#December_2017

UNESCO’s “Fostering Inclusive Knowledge Societies Report” http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/internet_draft_study.pdf

WCAG “POUR” Accessibility principles - <https://www.w3.org/WAI/fundamentals/accessibility-principles/>

Wikimedia Foundation Values [https://meta.](https://meta.wikimedia.org/wiki/Values/2008#Wikimedia_Foundation_values)

[wikimedia.org/wiki/Values/2008#Wikimedia_Foundation_values](https://meta.wikimedia.org/wiki/Values/2008#Wikimedia_Foundation_values)

Wikimania 2018 presentation: “Wikipedia And Bhutan Can Learn From Each Other” https://wikimania2018.wikimedia.org/wiki/Program/Bridging_knowledge_gaps:_Wikipedia_and_Bhutan_can_learn_from_each_other

Wikimania 2018: presentation: “The Quotation Of Oral Sources In A Decolonization Context” https://wikimania2018.wikimedia.org/wiki/Program/The_quotation_of_oral_sources_in_a_decolonization_context

Wikimania 2018 Panel: “Centering Knowledge from the Margins: A Whose Knowledge? discussion” https://wikimania2018.wikimedia.org/wiki/Program/Centering_Knowledge_from_the_Margins:_A_Whose_Knowledge%3F_discussion

Wikimania 2018 presentation: “Wikipedia for Indigenous Communities” https://wikimania2018.wikimedia.org/wiki/Program/Wikipedia_for_Indigenous_Communities

Wikitongues <https://wikitongues.org/>

Tools

The essays that follow explore the need for tool development to better support developers, organizers and moderators. These position papers were synthesized out of input from staff and community experts as well as prior independent research.



01
For Developers

02
For Organizers

03
For Moderators

Published
January 2019

Authors

Niharika Kohli
Ryan Kaldari
Joe Matazzoni

Contributors

Amir Aharoni
Asaf Bartov
Adam Baso
Trevor Bolliger
Maria Cruz
Bryan Davis
David A.M. Goldberg
Aaron Halfaker
James Hare
Danny Horn
Ramsey Isler
Jon Katz
Josh Minor
Toby Negrin
Margeigh Novotny
Nirzar Pangarkar
Joseph Seddon
Alex Stinson
Nick Wilson
Leon Ziemba
Leila Zia



In early 2006, a large-scale vandalism attack wiped out thousands of articles on the English Wikipedia. The editors were defenseless and the site was vulnerable. In response, four Wikipedians came together to write TawkerBot, the first anti-vandalism bot [1] for Wikipedia. This bot proved to be a life-saver for the site, and today more than 300 bots work round the clock on English Wikipedia to ensure the smooth functioning of the site. Tools like Twinkle [2] (tool library that helps editors perform wiki maintenance), Huggle[3] (a diff browser for rapidly reverting vandalism), HotCat[4] (allows a user to easily add and remove categories to pages), AutoWikiBrowser [5] (semi-automated MediaWiki editor), etc. drive many of the tasks power editors do on English Wikipedia every day.[6]

At the same time, smaller language wikis like Hindi Wikipedia[7] have problems coping with vandalism and keeping up with content moderation needs. Unlike English Wikipedia, they don't have the corps of volunteer developers able to write tools to defend and curate the site's content. It's much harder for those communities to grow their content or their editor base, because active contributors are stuck doing manual drudge work that bigger wikis automated years ago.[8][9]

Tools for Developers

Wikis need code contributors as much as they need content contributors. Templates, gadgets, and bots act as superpowers in making editors more efficient at their tasks. Experienced editors use these tools to create and maintain backlogs, track incoming edits and their quality, perform mass actions on pages, ward off vandalism and more. However this superpower is limited to wikis which have contributors who are able to write such code for the site. Bringing these important resources to all wikis is fundamental to bridging the equity gap across all language wikis, and is going to be a key factor in helping us realize the goal of amassing the sum of all knowledge.

Tools for Organizers

The Foundation's 2018-19 annual plan recognizes organizers as "fundamental implementers" and a "core asset" of the free-knowledge movement. Unfortunately tools that support organizers' efforts are frequently ad-hoc, poorly documented and available only on certain wikis. Access problems can be particularly acute in smaller communities, where the technical skills required to set up and run bots, scripts and other technologies are often scarce.

Organizers' needs fall into four main areas. "Community-building" tools are required to help organizers inform, engage and manage the work of their communities. "Outreach and promotion" tools will help organizers advertise their activities and recruit new members. "Event-management" tools are necessary to more efficiently carry out tasks like event signup and conference scheduling. Emotional labor, the fourth category of need, demands better support infrastructure to nourish and sustain the passionate organizers on the ground, over time. Finally, two overarching meta-problems are key areas of interest among organizers. One is the need for better guidance about best practices and the tools that do exist. As one organizer put it, "There are a lot of tools we don't know about or know what they can do for us. We need someone to help us understand what we are missing, and what to do and how to do it." The second is the need for a mechanism that can replace or augment categories, so that organizers will be able to classify content effectively and more efficiently tap volunteers' subject interests—a primary motivator, especially of new editors.

Tools for Moderators

A critical, but often overlooked, aspect of the workflows that make our projects successful are the tools and processes used to review and moderate our content. For the most part, the Wikimedia Foundation has taken a hands-off approach to content moderation tools and allows the community develop their own solutions (with a few exceptions such as Recent Changes filtering). As one would expect, the systems

built by the community utilize the building blocks at their disposal: templates, gadgets, user scripts, and wikitext. These tools and processes suffer from several significant problems, many of which have already been mentioned above: lack of portability, limited capabilities, lack of automated testing and code review, lack of localization support, etc.

Another major problem with these tools and processes, especially those created for content moderation, is their high learning curve. For example, on English Wikipedia there is a system for submitting, reviewing, and publishing draft articles called Articles for Creation (AfC). In order to participate as a reviewer in AfC, you have to install a special user script, be granted a special right though a unique vetting process, and use several obscure templates and categories. The complexity of this process limits the number of people who are able and willing to participate, which in turn leads to a less diverse pool of reviewers. This lack of diversity may contribute to problems of systemic bias in our content.[10] [11] The small number of reviewers also makes the review process slow, often taking a week or longer to review a submitted draft. This likely contributes to the overall inefficiency, resulting in decreased newcomer productivity. [12] Unless we make moderation tools work for less technical users, it is unlikely the pool of moderators will grow or diversify.

Similar examples can be found throughout the moderation processes for our projects, including workflows for article assessment, deletion, and problem triaging; workflows for reviewing edits; workflows for reviewing and organizing multimedia contributions; workflows for proofreading Wikisource transcriptions; and more. While the Wikimedia Foundation has historically focused on building software for reading and editing content, the other critical pieces of the wiki ecosystem have been neglected, leading to volunteers feeling overwhelmed and unsupported. In a 2015 survey of experienced editors across 10 projects, only 35% said that the Foundation was mostly or completely meeting the community's technical needs around identifying and surfacing content problems.

[13] Unfortunately, the Foundation's hands-off approach has resulted in a lack of credibility in this area. To build our credibility, we should first focus on the areas where there is a clear need for better tools, such as fighting vandalism and sock-puppetry. We should also investigate how editors transition into becoming moderators so that we can better facilitate that transition. Once we've proven our capacity to understand their motivations and work with moderators to build effective tools, we will then have the mutual trust needed to tackle more difficult workflows such as article deletion and conflict mediation.

Clearly, there's a lot of work for us to do in this area as we have only scratched the surface thus far. If we want to increase the capacity of our communities to efficiently and effectively moderate content, it is time for the Foundation to begin investing seriously in this area.

Examples

- HotCat
- Huggle
- Twinkle
- AutoWikiBrowser
- Programs and Events Dashboard
- Wikimedia Cloud Services
- CentralNotice
- GeoNotice

Areas of Impact

- Template
- Gadgets
- Bots
- Editing and Administration APIs
- Discussion systems
- Messaging systems
- Contributor Analytics
- Developer Advocacy and Outreach
- Translation and Localization Infrastructure
- API and Tool Documentation

01

Tools for Developers

Empowering our volunteer developers to write better code that can work across wikis is going to be a key factor in helping us gather the sum of all knowledge. Wikis need code contributors as much as they need content contributors. Templates, gadgets, and bots act as superpowers in making editors more efficient at their tasks. Experienced editors use these tools to create and maintain backlogs, keep track of quality of incoming edits, perform mass actions on pages, ward off vandalism and more. However this superpower is limited to wikis which have contributors able to write code for the site. This creates disparity in the resources available to

wikis. Bringing these important resources to all wikis is fundamental to bridging the equity gap across all language wikis.

Pan-wiki Tools Platform

Empowering our volunteer developers to write better code that can work across wikis is going to be a key factor in helping us gather the sum of all knowledge. Wikis need code contributors as much as they need content contributors. Templates, gadgets, and bots act as superpowers in making editors more efficient at their tasks. Experienced editors use these tools to create and maintain backlogs, keep track of quality of incoming edits, perform mass actions on pages, ward off vandalism and more. However this superpower is limited to wikis which have contributors able to write code for the site. This creates disparity in the resources available to wikis. Bringing these important resources to all wikis is fundamental to bridging the equity gap across all language wikis.

We must evolve our platform so that support tools can work on all our wikis seamlessly. Right now a lot of developer code lives on specific wikis (gadgets, Lua modules, templates) where it really isn't possible to do any type of testing, code reviews or debugging; nor is there any straightforward way to add localization or RTL support. This often leads to issues like security vulnerabilities, [1] conflicts with MediaWiki deployed extensions, [2] and bugs due to lack of maintenance. Also, in its current state, having code hosted on the wikis (in a per-project fashion) makes it hard to get in the mindset of having the code work across wikis. It's easy to get sucked into customization and forget to think about things like RTL rendering or localization. This evolution depends on core services being available to developers across communities which they can use for building tools. Examples of these services include APIs can be used to do copyright violation detection, vandalism detection, and image recognition, provide access to better statistics, and so on. Part of growing these services involves better partnerships with

companies like Google, Turnitin and others already providing such functionality. Finally, for this to succeed, Wikimedia staff will need to collaborate with our volunteer developer communities to come up with documentation and best practices for creating new tools. Tools that facilitate communication among engineers and volunteer developer communities will be key to achieving this goal.

An example of these might be tutorials and guidelines like these documenting how gadgets can make use of OOU [3] to standardize our interfaces and make them more accessible for everyone.

Bots : Potential and Risks

Bots automatically perform repetitive tasks, but like any type of automation the advantages come with risks. The way that bots are wielded by editors can shape the character of a wiki's community. An editor's ability to rapidly apply changes that impact contributions that may have taken hours of individual work is a form of power that can alter contributors' attitudes about their willingness to participate. Bots are also complex pieces of programming, typically developed by a single individual trying to solve a specific problem in a specific context. There is nothing inherently wrong with this, as volunteer development is a core component of our culture. However, there is little standardization in the development of bots, or guidelines for creating them to serve the wider community of projects. We must begin to evolve our thinking about the future of Bots in the context of the discussions on structured data outlined in the Ubiquity paper, and in our Developer, Moderator and Organizer Tool positions.

Gadgets & Gadget Usage

Unlike bots which may make use of external resources or specialized programming (e.g. neural net programming for ClueBot),[4]

gadgets are built for and within the browser ecosystem in JavaScript or CSS. The number of gadgets enabled by default for English Wikipedia, sixteen, [5] far exceeds the number for other wikis such as German [6], Hebrew, [7] Tamil, [8] and Italian [9], while eight other gadgets are manually enabled by over 30,000 active users. We have seen some migration of gadgets between projects, based on volunteer efforts by those who are familiar with JavaScript and CSS. Examples include HotCat, [10] Navigation popups, [11] UTC:LiveClock,[12] and WikEd. [13]

Gadgets are comparable to apps and extensions on other platforms, and we are faced with the same challenges of testing, reliability, developer standards, and review. Gadgets can easily break things for users due to bad programming practices, poor testing, or conflicts with other MediaWiki extensions, [14] [15] [16] and there is no easy way to identify them as the source of the problem. Most importantly, gadgets do not go through any sort of staff review process before being deployed which means they are a potential vector for abuse.

In addition to the issues of segregated usage, many gadgets are prone to browser performance issues due to their large JavaScript code consuming bandwidth and processor resources [14].

Conclusions & Implications

Though gadgets are heavily used on most Wikipedias their application is uneven due to three factors: overspecialization, concentration (in terms of development and use), and lack of instrumentation.

Larger projects with larger language communities (e.g. English) dominate the development of gadgets. And while correlation between programming expertise and English usage does not dictate their development, it does reflect the concentration of support based on privilege seen in other contexts.

The functionality of gadgets tends to be overspecialized and adapted to the wiki for which they were developed, thereby limiting internationalization, adaptation between left-to-right/right-to-left writing systems, and the development of more general and migratable solutions.

Finally, we lack instrumentation to track gadget requests, development, distribution and usage. Without the means to measure and analyze the gadget space we will continue to suffer the inefficiencies of redundant development efforts, overspecialization without awareness of the costs or consequences of these issues.

Wishlist & Technical Requests

In addressing the annual Wishlist, [18] the Community Tech team has recognized that many of the frequently requested tools, bots and gadgets already have existing solutions created by the community. Some of these tools have alternate versions adapted to specific languages (such as the TemplateWizard [19] developed for Hebrew [20] and German [21] wikis, while others, like the Blame tool, have been created multiple times [22][23]. Tools like Crosswatch [24] and a pageview stats tool were on the wishlist in 2015 but had already been developed by volunteers. Similarly, user Legoktm developed the GlobalPreferences extension in 2013, well before it was nominated for the wishlist [25]. Looking forward, all of these examples demonstrate the need to improve the tracking of requests and implementations across our various communities.

Conclusions & Implications

Clearly we must continue to rely upon a robust and dedicated community of volunteer developers to deliver the means to interact with Wikipedia's complex features. Much of their work has become indispensable to their specific communities and on more global scales as well, and in this sense they have the potential to have a

multiplying effect on the efforts of the Foundation. However, in the context of tools, we once again confront the challenges that come from our uniquely heterogeneous character: knowledge and tools concentrated in certain communities do not cross over to those that might make use of them. Often there are communication gaps between different wikis, which mostly act as individual ecosystems of users and tools. We lack a place to surface the tools used by various wikis, which would not only address issues of redundancy in both development and requests, but encourage the type of cross-community tool portability that is envisioned here. This is a challenge that mixes technical and cultural issues, some—like multilingualism and different community densities across different projects—are inherent to Wikipedia, while others—like widely divergent practices in documentation and standardization—are associated with any volunteer development community. By prioritizing the development of a shared platform for tool identification, user requests, and pre-deployment assessment we can begin to address both of these issues simultaneously.

Notes

- 1 “Persian Wikimedia cryptocurrency mining incident” <https://lists.wikimedia.org/pipermail/wikitech-l/2018-March/089636.html>
- 2 Using Hotcat after saving with visual editor tries to edit old version <https://phabricator.wikimedia.org/T170896>
- 3 OOUI is a UI component library created by WMF, aimed at providing a consistent UI experience that works well for all languages. <https://www.mediawiki.org/wiki/OOUI>
- 4 ClueBot https://en.wikipedia.org/wiki/User:ClueBot_NG
- 5 Default gadget usage on English
- 6 Wikipedia [Special:GadgetUsage](https://en.wikipedia.org/wiki/Special:GadgetUsage)
- 7 <https://de.wikipedia.org/wiki/Spezial:GadgetUsage>
- 8 <https://he.wikipedia.org/wiki/מגוון:GadgetUsage>
- 9 <https://ta.wikipedia.org/wiki/சிறப்பு:GadgetUsage>
- 10 <https://it.wikipedia.org/wiki/Speziale:GadgetUsage>
- 11 Hotcat <https://en.wikipedia.org/wiki/MediaWiki:Gadget-HotCat>
- 12 Navigation Popups https://en.wikipedia.org/wiki/Wikipedia:Tools/Navigation_popups
- 13 UTC LiveClock <https://en.wikipedia.org/wiki/MediaWiki:Gadget-UTCLiveClock>
- 14 WikEd <https://en.wikipedia.org/wiki/MediaWiki:Gadget-wikEd>
- 15 <https://phabricator.wikimedia.org/T178348>
- 16 <https://phabricator.wikimedia.org/T170896>
- 17 <https://phabricator.wikimedia.org/T22134>
- 18 <https://phabricator.wikimedia.org/T142461>
- 19 Community Wishlist https://en.wikipedia.org/wiki/Wikipedia:Bot_requests
- 20 Template Wizard <https://www.mediawiki.org/wiki/Help:Extension:TemplateWizard>
- 21 Hebrew Template Wizard https://he.wikipedia.org/wiki/מגוון:קני_הידום:Gadget-TemplateParamWizard.js
- 22 German Template Wizard <https://de.wikipedia.org/wiki/Wikipedia:Technik/Skin/Gadgets/Vorlagenmeister>
- 23 <https://github.com/wikiwho/WhoColor>
<http://wikipedia.ramselehof.de/wikiblame.php>
- 24 Crosswatch <https://tools.wmflabs.org/crosswatch/welcome>
- 25 Global Preferences <https://www.mediawiki.org/w/index>

02

Tools for Organizers

The Foundation's 2018-19 annual plan recognizes organizers as "fundamental implementers" and a "core asset" [1] of the free-knowledge movement. But tools that support organizers' efforts are frequently ad-hoc, poorly documented and not universally available—particularly to smaller communities. As the movement puts an increasing emphasis on knowledge equity, the need to understand and support movement organizers is more vital than ever. This white paper is an early effort to analyze and document organizers' main areas of need. This examination

will be followed and deepened soon by the annual plan-mandated Movement Organizer Study. [1]

Who are organizers and what do they contribute?

For the purposes of this analysis, a “movement organizer” is anyone who purposely seeks to motivate, attract and/or direct volunteer attention. Organizers’ efforts generally fall into the following categories: content creation (e.g. content drives, editathons, photo expedition and writing competitions,) knowledge dissemination (e.g. training, conferences and campaigns,) process improvement (e.g. standardizing sources, developing procedures, style guides and templates,) and outreach (e.g. lobbying and partnership development, especially in relation to GLAMs, governments, NGOs, etc.) and emotional labor.

Some organizers work independently, possibly having been trained by more formal groups. But most operate within the framework of various structures that support their efforts – they range from legally incorporated nonprofit entities (Chapters, of which there are about 40 that work to improve laws and negotiate content donations for example,) to Independent Organizations operating on the edges of recognized groups within a given movement. Unincorporated User Groups, with regional or thematic focuses (e.g., Community User Group of Greece, Wiki Medicine,) of which there are about 100 provide the other structures of support.

Many groups coalesce around shared interests in organizing or improving efforts around subject areas or types of tasks (e.g. copyediting). These “WikiProjects,” “Portals” or just “Projects” appear across many wikis, with about 300 concentrated on en.wiki. Finally there are Campaigns: unaffiliated working groups, usually in part supported by organizers at existing affiliates or user groups. These are usually time limited efforts run with support from the more permanent Affiliates and User groups,, such as Art+Feminism, Wikipedia Asian Month, or Wiki Loves Monuments.

Wiki “functionaries” such as stewards, admins, and bureaucrats, whose activities generally don’t focus on “attracting and directing volunteer attention,” are not considered organizers.

Problems and Needs

As one organizer put it, “There are a lot of tools we don’t know about or know what they can do for us. We need someone to help us understand what we are missing, and what to do and how to do it.”

As one staffer put it,

“We need to give people the scaffolding of how to be successful.”

We are some distance from addressing these two points because Wikipedia organizer tools remain hard to discover, largely restricted to use by people with technical skills, and limited by documentation that is non-standardized and/or inconsistently maintained. Organizer tools broadly fall into four groups: community-building, outreach and promotion, event-management, and conference tools.

Common Needs

These have many current issues in common. There is no step-by-step process that organizers can follow to create a new project or campaign and make sure that it will be successful. A lack of end-user documentation (as opposed to technical) often makes the tools organizers do locate unusable. Most organizers do not have technical backgrounds, so the bots, scripts, Wikidata tools and other technologies that benefit some groups enormously are not available to all. This problem can be particularly acute in smaller communities, where it is less easy to find people with the required skills. Organizers want an easy way to announce an event and find out who is going to come, for which they generally must turn to off site tools. Setting out work for a community to accomplish is another common organizer need, yet the wikis lack even the most basic functions of task-management software.

Community-Building Tools

“Community-building” refers to a nexus of functions that organizers require to inform, engage and motivate their communities — to

build “a movement not an event”, as one organizer put it. Performing these functions currently requires a patchwork of tools, a high degree of technical sophistication and considerable manual effort.

The key community-building function is conversation: between group members and between organizers and the group. Organizers need easy ways to make announcements and invite large numbers of people to participate in a discussion, while group members need better ways to share and discuss information among themselves.

Outreach and Promotion Tools

Community-Building tools address a need for more effective conversation within a group, but organizers also need to reach editors, readers and others who are not yet in their circle of contacts. There is currently no simple or effective way for organizers to promote their groups or events to wiki readers or editors who are likely to be interested (e.g. based on geography or demonstrated subject interest). A founder of one of the largest and most active user groups recently named their inability to reach out beyond a circle of existing, very active users as one of the biggest issues with current tools. Finding new members is slow and often accomplished through personal networks.

Tools like CentralNotice [2] and related tools like Sitenotice exist and do reach readers, but they lack important features such as targeting. Whereas Geonotice, [3] for example, can target by location but reaches only registered users and only on their Watchlists. Such tools are also subject to many restrictions and layers of approval. [4]

Event-Management Tools

Organizers of edit-a-thons, training sessions, photo walks and similar in-person events have urgent needs related to event management. The Event Metrics [5] project, currently in development, will aid event organizers with better data about their contributions, but event management itself is out of scope for that project. An opportunity exists for synergy between Event

Metrics and future event-management tools, since both make use of similar input data about the event and its participants. Event organizers need specific tools that support in person participant signup and sign-in, “day of” Wiki account creation, and conference tools.

Generally, event management is complicated by privacy issues and technical restrictions designed to prevent exploitation of the wiki platform. In the first case, the management of event participants requires providing an email address, or the use of third-party management tools such as Eventbrite. In the second, limitations on the number of wiki accounts that can be created from a single IP complicate the process by which participants engage the event. There are workarounds for some needs, but they are not universal or well known.

Conference Tools

At a higher level of organization, conferences are large annual or semi-annual events with their own needs. Each conference must rebuild or remix the infrastructure for conference calendaring, signup, payment, scheduling etc for each event. This is true even for recurring annual events. More direct links to on-wiki activities, such as the pages related to sessions, including proposal submission, scoring of submissions, scheduling, and presenting scheduled programs that happen in isolation on wiki.

Recommendations

The Foundation is currently undertaking a Movement Organizer Study which will provide more data on how organizers do their work, and their needs and challenges. This broad survey of organizer workflows be useful for community interventions, product teams and others wanting to better support organizers’ functional and emotional needs. Addressing these opportunities will be non-trivial, but the benefits will be felt more broadly than by organizers alone. “Subject interest” and “Group conversations”, in particular, are fundamental tools whose potential applications are widespread.

Improve The Category System

Research and experience indicate that an abstract interest in the movement, per se, is not what motivates volunteers at the beginning of their wiki journey: they come to Wikipedia to share knowledge and passion for some subject. The Wikiproject directory [6] demonstrates the wide range of interests organizers and volunteers pursue formally, from folklore to pharmacology, football to firearms. However, such projects are hampered by the fundamental weaknesses of the wiki category system [7].

Fundamental improvements to categorization will produce benefits across all communities, but serve organizers particularly well. Better categorization would introduce a foundation for better traversal and interconnection of subject areas, and support better recognition and promotion of interests for participants and organizers, respectively. Organizers need automated ways to classify (and therefore assemble and search for) articles and tasks by subject, and to reach out to potential participants based on demonstrated subject interest.

Facilitate Group Conversations and Task Management

The consultation about “fixing talk pages” [9] in our 2018-2019 annual plan will initiate a conversation about the shortcomings of this core wiki communication platform. However changes to the talk page system may not be the only way to address organizers’ communication needs.

Tracking, annotating and debating changes are fundamental “gestures” for all engaged wiki editors and contributors, but organizers bring an additional set of requirements such as task assignment and management that could potentially transform communication for everyone. Organizers strive to keep groups engaged and informed, and their efforts cross the border between the Wikipedia universe and the rest of the world in synergistic ways.

Organizers and group members need tools to work with what may become a new core wiki concept: the task. Different from the implicit notion of the edit which drives the development of an article, a task would have

a fixed identity, a lifespan, the means of being classified (e.g. by type, subject area, or degree of difficulty,) ownership privileges, and associated discussion threads. Tasks might not necessarily be circumscribed by or destined for single wikis (e.g. getting decentralized support on a communications plan, solicit support for event organization, or other “jobs” that don’t aren’t limited to single, on-wiki activity,) thereby introducing a new mode of cross-community connection.

Email is the only message-delivery mechanism outside talk pages that our system currently supports. But the world we live in now requires that organizers broadcast and stay in touch with members on multiple social-media platforms at the same time (see also Experience). Feeding these multiple platforms manually is labor-intensive. An important area will be determining how we should best should incorporate social-media channels into our notifications and communications system.

Improve Outreach and Promotion Tools

Given the enormous traffic the wikis command, failing to tap the communication potential of our platforms would be an enormous opportunity missed. Better outreach and promotion tools would enable organizers to more effectively reach desired audiences. Overcoming the problems associated with using the wikis for mission-focused promotion will require efforts in both the technical and social arenas.

Meeting organizers’ promotional needs may require the reconsideration of some longstanding ideas and prohibitions, e.g. experimenting with limited, noncommercial, movement-focused advertising to wiki readers. Given the wikis’ enormous traffic, even narrowly focused banners might prove effective, but this would require new oversight to make sure the level of promotion is not excessive. There may be ways to respect editors’ privacy while still targeting them with messages based on geography or demonstrated subject interest (for example editors might be requested to voluntarily submit such information).

As highlighted in earlier sections, subject interest

is a key motivator of wiki activities. Providing simpler and better ways for organizers to perform outreach based on subject will pay dividends in increased efficiency and effectiveness across a wide range of activities. Rather than focus on revising and reforming the MediaWiki Category system itself is unlikely to be the fix to this problem. Future solutions may come from experiments currently ongoing with structured data, AI projects like ORES Draft topic model, [10] or from a system based on link analysis, like the one that powers Recommendation API. [11]

Improve Tool Discovery, Documentation, Ease of Use

Wiki tools are hard to discover, use and install, and they are poorly documented. It is time to dedicate more resources that are responsible for documentation and organizing and standardizing community writing about best practices and model workflows. To address discoverability, the existing Tools Directory [12] and the Toolhub [13] project demonstrate a concrete effort. The Directory provides a list of tools by title, author and categorization tags, while Toolhub documents the underlying structure [14] for cataloging them. However, there is much room for improvement in terms of how these tools are explained and promoted to less-experienced users.

Immediately Fix “Day-of” Account Creation

There is something we can do address a core need in the very near term. The limits and controls around account creation on shared IPs are the most immediate block on effective events. This ticket [15] includes a good discussion of solutions. Current solutions share many of the common problems identified across organizers: it is not available on all wikis, not well known or understood and thinly documented.

Notes

- 1 https://meta.wikimedia.org/wiki/Wikimedia_Foundation_Annual_Plan/2018-2019/Audiences#Outcome_1:_Progressive_Onboarding
- 2 Central Notice <https://meta.wikimedia.org/wiki/Special:CentralNotice>
- 3 GeoNotice <https://en.wikipedia.org/wiki/Wikipedia:Geonotice>
- 4 Usage Guidelines for Central Notice https://meta.wikimedia.org/wiki/CentralNotice/Usage_guidelines#Goals
- 5 Event Metrics [16] https://meta.wikimedia.org/wiki/Community_Tech/Tools_for_program_and_event_organizers
- 6 Wikiproject Directory https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Directory
- 7 A few of these weaknesses: Categories are monolingual, ad-hoc and completely nonstandard, so developing universal tools based on them is challenging. Items in sub-categories don't inherit from parents, which is why searching broad categories typically yields few article results, contrary to user expectations. This non-inheritance also leads to hyper-specificity and bloat, to the point of absurdity. [8] Technically speaking, the category system is not a “taxonomy”—a logical structure where all child categories are more specific classifications of the parents and wholly contained by them (e.g., Musical Instruments > String Instruments). It is, instead, a “category network,” where children have some relationship with parents, but the relationship is unpredictable and therefore less useful (Musical Instruments > Orchestras). Categories can even be circular, with one category being both a parent and a child of itself.
- 8 Twitter feed dedicated to surfacing absurd categories related to Cats on Commons <https://>

twitter.com/CommonsCat

9 Talk pages Consultation [17] https://meta.wikimedia.org/wiki/Wikimedia_Foundation_Annual_Plan/2018-2019/Audiences#Outcome_2:_Communication

10 ORES Draft topic model https://meta.wikimedia.org/wiki/Research:Automatic_new_article_topics_suggestion

11 Recommendation API https://meta.wikimedia.org/wiki/Recommendation_API

12 Tools Directory <https://tools.wmflabs.org/hay/directory/>

13 Toolhub <https://meta.wikimedia.org/wiki/Toolhub>

14 Toolhub data model https://meta.wikimedia.org/wiki/Toolhub/Data_model

15 <https://phabricator.wikimedia.org/T202759>

16 Eight organizer interviews: As part of the research into the ongoing Event Metrics project, I conducted about 12 hours of interviews with seven event organizers suggested by program staff. (I also received one written interview.) These interviews, for which I have extensive notes, cover the full workflow of event creation, management and reporting. I did not ask subjects for the right to publish but could request if desirable. https://meta.wikimedia.org/wiki/Community_Tech/Event_Metrics

17 Extensive talk page discussion: Also as part of Event Metrics, J. Matazzoni engaged extensively with organizer on the project talk page. (The discussion is organized by subject. E.g., here on the problems associated with Account Creation.) https://meta.wikimedia.org/wiki/Talk:Community_Tech/Tools_for_program_and_event_organizers, https://meta.wikimedia.org/wiki/Talk:Community_Tech/Tools_for_program_and_event_organizers#Step_3:_Wiki_account_creation

03

Tools for Moderators

The tools and processes used for reviewing and moderating content are critical to the growth of the projects, but have historically been overlooked. Communities have, for the most part, been left to develop their own moderation workflows, tools and solutions – which means that community-built systems tend only to utilize the building blocks at their disposal: templates, gadgets, user scripts, and wikitext.[1] This approach results in solutions with limited capabilities, that tend not to be portable, lack automated testing and code review, and don't support localization.

Many of these tools and processes also have steep learning curves and require considerable technical acumen to use. [2] The more complex the process is, the more it limits the number of people who are able or willing to participate. This, in turn, leads to a less diverse pool of reviewers and contributes to problems of systemic bias in

our content.[3][4] And, fewer moderators makes the review process slow, impacting newcomer productivity.[5] In order to grow and diversify the pool of moderators, moderation tools must be made to work for less technical users, and the Foundation must drive investment that enables community developers to create more portable, localizable tools, and in resources to support better practices in tool development.[6]

The Example of English Articles for Creation

“New article creation is a battlezone where socializing newcomers seems to take a back seat to ensuring quality control.”

Since Wikipedia’s dramatic rise in popularity it has faced an endless stream of spammers, vandals, PR firms, and “clueless” newbies. In response, around 2009, English Wikipedia began directing new users to create new articles in the Draft namespace and seek review before publication. This new process, Articles for Creation (AfC), was meant to ensure that new articles measure up to Wikipedia standards for notability and verifiability, [2] but it can be a frustrating and discouraging process for newcomers. This is indicated by a low percentage of drafts that are reviewed (29.4% of AfC drafts are never submitted for review. 11% of reviews take longer than a week) and a falloff in new article longevity. Since AfC was introduced (2007), the proportion of articles of each new page creator that survive at least 30 days has declined. One reason for this is that draft articles are hidden from potential collaborators. It also suggests that the review process is often slow and confusing.

Implications

Overall, we should create better tools for the community to review new articles with, increase the percentage of drafts that are reviewed and decrease the number of reviews that take a week or longer to happen. By improving these metrics, we will improve newcomer productivity.

What Experts Say

Wikipedia has a very strong need for better tools to deal with vandalism and sock-puppeting. Though the means to block such malicious actors exist, they are easy to evade, especially for dedicated vandals who know how to use proxies. [7] Current proxy blocking tools aren’t very effective, especially in combating zombie proxies. [8]

Community-built tools are important for

moderators, but have lots of problems including lack of consistent or ongoing developer support, being tailored to one wiki and with generally steep learning curves for installation and use.

Discussion-based workflows (see Articles for Deletion, [9] Bot Request for Approval, [10] Article Peer Review, [11] for example, are often overly complicated and would benefit from dedicated interfaces to streamline their processes..

Generally, this points to the need for clearer guidelines for collaborating with the community on tool development to make sure we aren’t building things in isolation. Further, they must be built in close collaboration with the community in order to be successful. One key requirement here is consistent communication: project managers should provide weekly updates for the community, post screenshots, ask questions, and respond to feedback.

Implications

The top priority for moderators’ tools should be counter-vandalism tools, particularly automated proxy-blocking tools.

Later, we can focus on the tools used to facilitate discussion-based workflows like Articles for Deletion.

Decision Quality [12]

Because disagreements and conflicts on Wikipedia are frequent and inevitable, effective decision-making and conflict resolution processes are essential to community health.

The Articles for Deletion (AfD) process is one of the core moderation processes on English Wikipedia. AfD works through a consensus process rather than voting, and thus far we have learned that many factors affect the quality of AfD decisions, for example the likelihood that they will be reversed later. On a broader level it has been demonstrated that larger groups of participants make better decisions, but that the degree of participant experience makes a difference. Lacking understanding of Wikipedia norms and processes tends to yield worse decisions. Bias also plays a decisive role in decision quality: biased admins make worse decisions than impartial ones, and diverse

groups tend to reduce bias. At the same time, orchestrated bias does affect decision quality, especially when users with established positions are specifically recruited.

Implications

More tools should be built to facilitate more inclusive participation in moderation processes such as Articles for Deletion since this will likely improve the quality of the decisions. These tools should provide onboarding support and education for newcomers so that they can participate constructively.

Notes

1 Wikitext https://meta.wikimedia.org/wiki/Help:HTML_in_wikitext

2 Articles for Creation (AfC), just one example from English Wikipedia, is a system for submitting, reviewing, and publishing draft articles. In order to participate as a reviewer in AfC, you have to install a special user script, be granted a special right through a unique vetting process, and use several obscure templates and categories. https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Articles_for_creation

3 Lam, Shyong K., et al. (2011). "WP:Clubhouse? An Exploration of Wikipedia's Gender Imbalance", WikiSym '11.

4 Purtil, Corinne; Schlanger, Zoë (October 2, 2018). "Wikipedia rejected an entry on a Nobel Prize winner because she wasn't famous enough", Quartz.

5 Schneider, Jodi, et al. (2014). "Accept, decline, postpone: How newcomer productivity is reduced in English Wikipedia by pre-publication review", OpenSym '14.

6 Code review procedures, documentation standards, etc.

7 Open Proxies <https://en.wikipedia.org/wiki/>

Wikipedia:Open_proxies

8 A zombie proxy is a proxy server being run on a computer that has been compromised by hackers.

9 Articles for Deletion https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Articles_for_creation

10 Bot Request for Approval https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval

11 Article Peer Review https://en.wikipedia.org/wiki/Wikipedia:Peer_review/guidelines

12 Lam, Karim and Riedl (2010) "The Effects of Group Composition on Decision Quality in a Social Production Community" <https://dl.acm.org/citation.cfm?id=1880083>

Platform agnosticism

“Stroll through Sanlitun, a bustling neighborhood in Beijing filled with tourists, karaoke bars, and luxury shops, and you’ll see plenty of people using the latest smartphones from Apple, Samsung, or Xiaomi. Look closely, however, and you might notice some of them ignoring the touch screens on these devices in favor of something much more efficient and intuitive: their voice.” [1]

The Chinese language, as many other languages, was not built for typing tiny letters on a small screen. But that’s okay because technology, as it usually does given large-enough demand, is making its way around such initial difficulties. In this particular case, the answer might be voice search, alongside AI and new messaging paradigms.

This is just one example of how growing populations are coming to the internet with new needs, new languages, and new modes of expression. What’s certain is that their arrival will change the fabric of and forms in which knowledge is created, shared, and used. As internet access and usage rises in growing economies, the internet will become a more diverse place and platforms will be required to adapt to the needs and motivations of these new users.

If Wikimedia projects want to be participate in this growth and to “break down the social, political, and technical barriers preventing people from accessing and contributing to free knowledge”, we must ensure adaptability to any platform or mode of usage.

Yet predicting trends can be tricky and the risks that have prevented us from being at the forefront of technical innovation so far still apply. Unlike Google, who have the resources to try to do everything-everywhere-all-the-time. We don’t have the luxury or expertise to take large risks, especially if they’re not initiated from

the the ground-up (i.e. from our communities).

For us, ubiquity means adaptation - skipping the guessing game of what will be big in the future, investing in the needs of our current and potential communities to make sure our content is available for use in any future trend and for presentation on any device.

To support the goal of ubiquity we must focus on re-structuring our content so that it can easily be repurposed, remixed and repackaged by us, our communities, or other platforms. For us, structured Wikipedia content could significantly content porosity between our projects over time - facilitating use cases like a reader’s smooth transition from the Wikipedia article on Istanbul, to the Wikivoyage guide, and onward to related media about the city from Commons. Structured content would also support the establishment of global and customizable modular templates for articles, portals and projects. Standardized formats for the subcomponents of these experiences (such as sections, ideas or themes) via well-documented Wikipedia APIs—ie. the means of retrieving knowledge in whole or part—would support both non-Wikipedia platforms and future Wikimedia uses. A more structured content API platform would also make it easier for our diverse communities to generate the tools they feel they need - while maintaining consistent and reliable standards, and that work smoothly across the entire Wikimedia platform (see also Tools: For Developers).

Finally, as we consider this issue of ubiquity at the intersection of Wikipedia and its consumers, structured content would relieve us from the requirement to anticipate, monitor or otherwise be directly aware of how all populations in all emerging economies are developing their own unique relationships to the internet.

Syndication

“Wikipedia content appears to play a substantially more important role in the Internet ecosystem than anticipated, with other websites having critical dependencies on Wikipedia content.”

“Google becomes a worse search engine for many queries when it cannot surface Wikipedia content (e.g. click-through rates on results pages drop significantly) and the importance of Wikipedia content is likely greater than many improvements to search algorithms. Our results also highlight Google’s critical role in providing readership to Wikipedia. However, we also found evidence that this mutually beneficial relationship is in jeopardy: changes Google has made to its search results that involve directly surfacing Wikipedia content are significantly reducing traffic to Wikipedia.” [2]

So far, Wikipedia’s relationship with Google has been fairly symbiotic. We provide a trusted source they can show at the top of the page; they provide an increase in pageviews and, in turn, an increase in donations, in new editors, and in the continued creation of quality content they can then show to users. Everybody wins and information is distributed freely.

Yet exposing more information outside of the site, such as in Google’s knowledge panels, has decreased pageviews to Wikipedia. It is unfortunate that this is an issue. While we still met our goal of providing the information a reader sought without the direct traffic to our sites, we face not only a decrease in funds, but eventually a decrease in quality. Potential editors never see the site, let alone have a way to contribute, and current editors have less motivation to continue writing. Over time, we’re in trouble.

But, so is Google. The study quoted above clearly shows that Google is a worse search engine in a world without Wikipedia. Wikipedia’s importance is so large that the “mere presence of Wikipedia links may have an effect

approximately 80 times larger than the difference between a good ranker algorithm and bad one (for many queries)”. Similar patterns have been found for other online websites such as Reddit and StackOverflow, where Wikipedia content is widely shared.

Thus we find ourselves in an odd paradox where our current level of ubiquity is also a potential threat. One option would be to take a purely defensive stance and work towards preventing any information from usage outside of the site. Needless to say that that goes directly against the free-culture underpinnings of our movement, as well as our licensing. The other option would be to take syndication for granted - to imagine our content spread throughout the fabric of the internet, and shift our content creation and revenue model to such a future.

We need to open or deepen conversations with our partners, to provide them with insights into our side of the relationship. Being able to present them with our perspectives, such as those outlined in the previous sections will make it easier for them to respond to this more nuanced recognition of our interdependency. Larger institutions in particular must be made aware of the financial, legal, trust and cognitive dimensions of a relationship where they are getting a tremendous amount of value for no cost while putting the sustainability of that resource at risk. For example, our partners need to be more aware of information attribution issues: sampling our content without attribution that links to its full context not only negatively affects Wikipedia pageviews but potentially diminishes its functional value (i.e. outside the context of the community that can vouch for/dispute its veracity).

This recognition of the mutual downside is a potential opportunity for deepening our relationship with these high traffic drivers. Just as we now have a process for reviewing and adopting potential affiliate chapters, so too could we institute a model of official corporate affiliation with Wikipedia (e.g. “Google, an official partner of Wikipedia”), that makes that partner an official sponsor of the Open Knowledge Movement, according to some

mutually agreeable terms. A similar concern was voiced in recent research with regard to GLAM partners - that we have no way of bringing them into the fold in an official manner... “we can’t even provide them with a logo to use on their website”.[3] In this way we have the potential to amplify the “building a better world” missions of, for example, Apple and Google... and even to elevate the “don’t shoot the messenger” vibe of Reddit. Providing current and future partners with access to structured content, including contribution actions, via an API would support more symbiotic relationships, and open the door to creating workflows of contribution from other places where our content is used.

Content Relevance

“In the English Wikipedia, articles of strong insufficient quality alone receive close to half of the pageviews, and in the Russian Wikipedia, they receive more than half.” [2]

For our projects to be ubiquitous, we must provide relevant content to all of our users. Not all wikis are the same, nor do they grow in similar fashions and users of different projects have widely varied motivations for reading. For example, our research shows that readers in Western-language Wikipedias are more likely to focus on quick-fact information whereas the speakers of languages in growing economies are more likely to use Wikipedia for deeper learning and for work or school purposes. To be able to cater to the needs of individual wikis or groups of wikis, we must be able to distinguish their needs. Features that might work great on one target audience, might not work for another. Similarly, content that might be notable for a particular community, might not be for another. Focusing on targeting our work to match our unique audiences as well as providing them with the tools to build according to their needs will help us cover the entire range of content that our current and future readers will require. Only by analyzing the needs of readers, editors, and moderators can we address imbalances

in projects which constrain their growth. For example, knowing which Wiki projects may have quality content but low readership, or a high volume of low quality content, would targeted interventions toward more sustainable approaches to growth. Achieving that sustainability will mean assisting projects according to their specific needs. Being able to model the extent of a wiki’s content gaps along with nuances of its editor retention history would allow us to more effectively focus on the factors that limit that project’s ability to scale. We will need new tools to do so, based on a foundation of structured content and communication.

1. Wikitext https://meta.wikimedia.org/wiki/Help:HTML_in_wikitext
2. Articles for Creation (AfC), just one example from English Wikipedia, is a system for submitting, reviewing, and publishing draft articles. In order to participate as a reviewer in AfC, you have to install a special user script, be granted a special right though a unique vetting process, and use several obscure templates and categories. https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Articles_for_creation
3. Lam, Shyong K., et al. (2011). “WP:Clubhouse? An Exploration of Wikipedia’s Gender Imbalance”, WikiSym ’11.
4. Purtil, Corinne; Schlanger, Zoë (October 2, 2018). “Wikipedia rejected an entry on a Nobel Prize winner because she wasn’t famous enough”, Quartz.
5. Schneider, Jodi, et al. (2014). “Accept, decline, postpone: How newcomer productivity is reduced in English Wikipedia by pre-publication review”, OpenSym ’14.
6. Code review procedures, documentation standards, etc.
7. Open Proxies https://en.wikipedia.org/wiki/Wikipedia:Open_proxies
8. A zombie proxy is a proxy server being run on a computer that has been compromised by hackers.

Ban, Perc and Levnajic (2017) “Robust clustering of languages across Wikipedia growth”

<http://rsos.royalsocietypublishing.org/content/4/10/171217>

“Can NC-licensed content be used in the Wikipedia project? – No. Wikipedia contents are being used commercially.” https://meta.wikimedia.org/wiki/Free_knowledge_based_on_Creative_Commons_licenses#8._Can_NC-licensed_content_be_used_in_the_Wikipedia_project?_%E2%80%93_No._Wikipedia_contents_are_being_used_commercially

Creative commons non-commercial license https://meta.wikimedia.org/wiki/Free_knowledge_based_on_Creative_Commons_licenses#8

De Zuniga, Weeks and Ardevol-Abreu (2016) “Effects of the News-Finds-Me Perception in Communication: Social Media Use Implications for News Seeking and Learning About Politics” <https://onlinelibrary.wiley.com/doi/full/10.1111/jcc4.12185>

Dixon (2018) “Why Decentralization Matters”, <https://medium.com/s/story/why-decentralization-matters-5e3f79f7638e>

Econsultancy (2018) “Why we need to stop repeating the “50% by 2020 voice search prediction” <https://econsultancy.com/why-we-need-to-stop-repeating-the-50-by-2020-voice-search-prediction/>

Guazzini, Vilone, Donati, Nardi and Levnajic (2015) “Modeling crowdsourcing as collective problem solving” <https://www.nature.com/articles/srep16557>

Hagedom et. al (2011) “Creative Commons licenses and the non-commercial condition: Implications for the re-use of biodiversity

information” <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3234435/>

Halfaker et. al. (2018) “The Rise and Decline of an Open Collaboration System: How Wikipedia’s reaction to popularity is causing its decline” https://www-users.cs.umn.edu/~halfaker/publications/The_Rise_and_Decline/halfaker13rise-preprint.pdf

Jacobs (1961) “The Death and Life of Great American Cities”, Random House

Jun, Lee and Jeong (2016) “Early onset of structural inequality in the formation of collaborative knowledge, Wikipedia” <https://arxiv.org/abs/1610.06006>

Jun, Lee and Jeong (2015) “Intellectual interchanges in the history of the massive online open-editing encyclopedia, Wikipedia” <https://arxiv.org/abs/1510.06092>

Kittur, Pendleton and Kraut, (2009) “Herding the Cats: The Influence of Groups In Coordinating Peer Production” <http://citeseerx.ist.psu.edu/viewdoc/>

