

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 in⁹ 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/>

