WIKIMEDIA
FOUNDATION

Address Knowledge Gaps, Three Years On

An updated roadmap for knowledge gaps research at the Wikimedia Foundation.

April 19, 2022. Prepared by Miriam Redi (mredi@wikimedia.org), Isaac Johnson (ijohnson@wikimedia.org), Martin Gerlach (mgerlach@wikimedia.org), and Leila Zia (lzia@wikimedia.org). DOI 10.6084/m9.figshare.19589662

The Wikimedia Foundation Research Team shares an update to its 2019 Knowledge Gaps White Paper. We reflect on what we have done and learned so far towards this strategic direction, revisit the research questions in the original white paper, and look to the future of research on this topic.

In February 2019, the Wikimedia Foundation Research Team published three white papers outlining the plans and priorities for the next five years: Address Knowledge Gaps (the focus of this update), Knowledge Integrity, and Foundational Research. Since then, the team has learned about existing research, and developed new methods and tools that can foster effective work towards addressing knowledge gaps. Three years later, we came together to reflect on the past, present, and future of our research around Knowledge Gaps. We present an update to the original white paper, with a summary of our findings and contributions, and a revised structure and plan for our future research and interests in this direction.

Executive Summary

In fall 2021, we came together to reflect on our roadmap towards addressing knowledge gaps, and did a deep dive in the research conducted over the past few years. Thinking about our work five years down the road, we identified three main areas of developments and improvements:

1. **Guiding Principles**. A set of principles guide our research in knowledge gaps and beyond: our role in Knowledge Equity, the projects we focus on, and a methodology focusing on communities, inclusivity, privacy, and openness.

2. **Consolidated Research Areas**. The original white paper identified five main areas of research: Identify, Measure, Bridge, Multimedia Gaps, and the Knowledge Gaps Index. Our updated version retains the three main directions: **Identify, Measure, and Bridge knowledge gaps**, and incorporates the remaining two into these three strategic directions. In this update, we will cover our initial and current vision for each direction, as well as past and future work towards that vision.

3. **Ideas for Future Research**. The lessons learned, and the large network of collaborators, encouraged us to think about big research questions, spanning a 5 to 10-year horizon.

Wikimedia Research, three years on.

Three years ago, our first white paper searched for an answer to one important question about our movement's future: How can we help Wikimedia projects thrive in a world that is becoming increasingly different from the one we are building for today? In 2019, we began advancing knowledge equity with a research program to address knowledge gaps. The first white paper shared ways in how the program and the Foundation itself can best tackle these gaps. Although greatly guiding our plans and priorities for the next five years and helping further the 2030 Wikimedia Strategic Direction, the world has changed and the Wikimedia research team has too. So must our 2019 report.

We have grown and **developed new capacities that are key for addressing knowledge gaps**. New members of the team have brought expertise in complex networks, data engineering, and community building. New **formal collaborations** have grown our expertise in human-computer interactions and disinformation. New **teams at the Foundation have become part of a broader network** of mutual support and collaboration in the context of addressing knowledge gaps:

- The Abstract Wikipedia team who started a new project where structured data can be transformed into Wikipedia-like articles through community-written functions;
- The Structured Data Across Wikimedia project who are building tools to add machine-readable structured metadata to wikitext pages;
- The Global Data & Insights team who gathers and analyzes data about the movement
- The Machine Learning Platform team who develops and maintains Wikimedia's machine learning infrastructure for productizing ML models.

Finally, the **Movement Strategy recommendations** are driving and shaping our research projects. Taken together, these developments have brought new knowledge and clarity around our strategy towards addressing knowledge gaps in Wikimedia projects.

1. Principles Guiding Knowledge Gaps Research

Knowledge Equity.

The Knowledge Equity strategic direction is at the heart of the knowledge gaps research. Knowledge Equity promotes the inclusion into Wikimedia projects of those communities and forms of knowledge that have been left out by structures of power and privilege. Through the knowledge gaps research, we aim to operationalize knowledge equity by focusing on atomic and more measurable knowledge gaps. The gaps we study reflect the disparity of coverage for different types of knowledge or communities. For example, the gender gap reflects a disparity in the representation of different gender groups in Wikimedia projects. As researchers, our contribution towards knowledge equity is to provide tools and research to identify, measure, and address those knowledge gaps that prevent us from reaching knowledge equity.

Beyond one project.

We are open and committed to conducting research on and for Wikimedia projects including but not limited to Wikipedia. While a significant part of our research in the past has focused on Wikipedia, we see great research opportunities when we start looking beyond English-language communities or Wikipedia. In particular, we intend to conduct research on projects such as Wikidata, Wikimedia Commons, Wikisource, and Wiktionary as opportunities arise. These projects represent different forms of knowledge than Wikipedia but all relate directly to efforts to address knowledge gaps on Wikipedia.

Community-driven research.

As researchers at the Wikimedia Foundation, we are inspired and influenced by Wikimedia communities. Our aim is to develop research and technologies that work in harmony with community principles and mechanisms. In our workflow, understanding the context in which our research and tools live and develop is key: qualitative research to systematically understand community practices, guidelines, and processes should be at the heart and the start of all our projects. An example of this approach is our research to create a taxonomy of reasons for citation need, as the first step towards a model detecting sentences needing citations.

Inclusive research methods.

Knowledge equity starts from the technologies we develop. Our outputs are designed to support a diverse community. Our research and tools should be inclusive by design, and focus on methods that can be scaled across platforms, content types, languages, and cultures. For example, unlike most natural language processing or image classification tools, which are modeled after main cultures/languages, we aim at producing models that are language-agnostic and inclusive of different cultures.

Machine-in-the-loop and human agency over automated systems.

The development of our machine learning models is centered in a machine-in-the-loop framework. In contrast to human-in-the-loop approaches — which aim to include humans in the process of training machine learning models by, for example, providing labels for training data — our models are designed to play a supporting role for editors, empowering them and improving their capacity. Researchers should focus on model interpretability, simplicity, and flexibility. Editors and other users should keep full control of whether to adopt or reject the model's suggestions. Examples of this approach include models for language-agnostic topic classification and the add-a-link task for newcomers. Moreover, while we build systems to support editors and readers with machine-assisted solutions, users keep full agency over automated systems. Content is not automatically pushed or adapted to users, who can fully choose what tasks and items they want to interact with.

Privacy.

Our research is guided by the respect for privacy that is core to all Wikimedia projects. For example, our respect for reader privacy means that we do not have ready-made, exact datasets of how all individuals navigate through articles or other fine-grained or longitudinal aspects of reader behavior (see Arora et al.). We are careful to avoid overly intrusive profiling of editors when supporting patrollers in their work to detect sockpuppets or identify vandalism. This constraint also helps us be creative in designing research models that respect editors and readers and grants them the freedom to interact in a flexible and open manner with the Wikimedia projects.

Openness.

Freedom and Open Source is a guiding principle of the Wikimedia Foundation and our research. The technologies, tools, code, and scientific papers that we produce or support other researchers to produce need to embrace and promote this principle following the Wikimedia Foundation's Open Access policy.

2. Three Research Directions: Identify, Measure, and Bridge Knowledge Gaps

We reflect here on the directions in the original white paper, summarize goals achieved, and reformulate the plans for future work. The structure of the 2019 manuscript included three broad directions — namely (1) identify knowledge gaps, (2) measure knowledge gaps, and (3) bridge knowledge gaps — and two cross-thematic directions that became the paper's de facto fourth and fifth overall directions : (4) multimedia knowledge gaps, and (5) the knowledge gaps index. For our updated version, we kept the three main directions: (1) identify, (2) measure, and (3) bridge. We then assigned individual ideas and projects of directions (4) and (5) to one of the corresponding three main directions, thus consolidating the original five strategic directions into three, more complete areas of research.

2.1. Identify Knowledge Gaps

This direction focuses on developing systematic definitions of knowledge gaps and their context as the first step towards operationalizing knowledge equity.

Our initial vision.

In "Section 1: Identify Wikimedia knowledge gaps" of the 2019 white paper, we identified the need for a systematic definition of knowledge gaps as a required first step for identifying knowledge gaps. We laid out a research plan for the development of a taxonomy of knowledge gaps in Wikimedia projects, which would describe and classify the content gaps of Wikimedia projects, gaps about the readers and contributors of the projects, as well as the usage gaps, namely the accessibility barriers. Finally, our plan included a taxonomy of causes for Wikimedia knowledge gaps.

Our current vision.

Our current vision for this direction is largely aligned with the original white paper. We are committing to defining gaps for readers, contributors and content, and creating a separate taxonomy for barriers. While identifying causes for all knowledge gaps is not feasible with our current expertise and resources, we will develop a framework to study causes of knowledge gaps using one example gap as a use-case. As part of these efforts, we also commit to in-depth investigation of how readers and contributors interact with Wikimedia sites, to uncover and prioritize new and existing knowledge gaps.

Our contributions so far.

A Taxonomy of Knowledge Gaps of Wikimedia Readers, Contributors and Content.
Following our initial vision of a multi-dimensional definition of knowledge gaps, we completed the first taxonomy of Wikimedia knowledge gaps. We defined a knowledge gap as a *disparity in representation of a specific category of content, readers or contributors.* We reviewed 200+ references to find evidence of potential gaps and inequalities across Wikimedia projects. We released an initial version in August 2020, collected feedback from the community, and incorporated it into a second version that was released in Jan 2021. The taxonomy is a structured list of the majority of gaps we can find in Wikimedia projects, across three main pillars of the Wikimedia Movement: readers, contributors, and content. Defining an exhaustive list of gaps' potential causes and barriers (*usage* gaps) required a separate effort. However, while doing this research, we found evidence of elements that could potentially amplify or cause inequalities in readers, contributors, and content, and added them to the final taxonomy manuscript as barriers.

Understanding Wikimedia readers and contributors.
To uncover and prioritize knowledge gaps in contributorship, we are working on a large-scale analysis of editors' edits on Wikipedia based on Wikimedia's publicly available information. We are building a taxonomy and classifier of edit types to systematically analyze the most/least common types of edits that contributors make on Wikipedia. Moreover, working towards removing barriers to readership requires a deep understanding of how readers use our platforms. In the past few years, we worked on foundational research about readers' behavior. We investigated what brings readers around the world to Wikipedia, and how this changes across different demographics, readers' navigation patterns, how readers interact with images on English Wikipedia, and the role of external links and citations on Wikipedia navigation.

Our plan for future work.

Including more types of knowledge and contributors.
The taxonomy developed is a simplified version of the ever-evolving, complex landscape of knowledge gaps in Wikimedia readership, contributorship, and content. Our intent is to expand the taxonomy on two fronts: types of knowledge, and types of contributors.

First, the currently available version of the taxonomy largely focuses on Wikipedia as the main field of study. Our aim is to expand the coverage, when possible, to the four Wikimedia projects that are central to our research: Wikipedia, Wikidata, Wikimedia Commons, and Wikisource.

Second, the current notion of contributorship is limited to people who are editing the content of Wikimedia projects. However, we want the taxonomy to cover other types of Wikimedia contributors. For example, movement organizers are central to the health and diversity of the community; developers build tools and contribute to MediaWiki that are essential to bridge Wikimedia knowledge gaps; donors sustain the projects through financial support; and Wikimedia Foundation partners support innovation and the growth of participation globally.

Understanding the resources gap.

As part of our studies around the inequalities and barriers preventing people from accessing and contributing to free knowledge, it became clear that different communities have access to vastly different tools and resources. New research is needed to understand and map the technical resources available to our communities – e.g., vandalism detection bots, templates for Wikidata-driven infoboxes, and fine-grained data about readership in their geographic region. A deeper analysis of the data/tools available to editors in different communities will shed light on how resource availability affects their productivity and the creation of high-quality, well-maintained content.

A taxonomy of barriers preventing people from accessing free knowledge.

Advancing knowledge equity not only means understanding, measuring, and bridging knowledge gaps, but also requires an understanding of the underlying causes to design effective interventions. While some of our recent research has attempted to identify hypotheses for gender gaps, causal evidence on what drives gaps is scarce and can require large experimental settings and theoretical frameworks. While studying causes for all possible gaps is outside the capacity of the Wikimedia Research team, our aim is to support the broader research community towards this direction. First, leveraging literature and existing efforts from the Wikimedia Foundation, we plan to devise a taxonomy *of barriers* that could contribute to knowledge gaps across readers, contributors, and content (for example, censorship or connectivity). Specifically for readers, research based on large-scale website traffic data analysis can be helpful to understand how different regions engage with Wikimedia content and uncover barriers to access free knowledge. Second, based on the taxonomy of barriers, we will develop a framework to study causes of knowledge gaps, focusing on one or a few example gaps.

Understanding Readers' Curiosity.

A new area of work in readership studies is around readers' curiosity. From past research we know that on average 20% of Wikipedia readers come to Wikipedia motivated because they are bored or are exploring Wikipedia for fun. While curiosity is generally regarded as beneficial to learning, there are now indications that there is a "dark side of curiosity," which is associated with errors in discerning the novelty and quality of information. We are interested in how readers are curious when seeking information in Wikipedia and in quantifying the degree to which readers are critically engaging with information on Wikipedia.

Research on Multimedia data.

While most of our efforts have focused on understanding the usage of visual content, multimedia is beyond images. Audio, video, and structured data are becoming increasingly popular on the web and on the Wikimedia platforms. We aim to open up new research directions around the usage of multimedia content beyond images, studying the role of multimedia content in our platforms, the extent of its presence, and its impact on navigation.

Disseminating the taxonomy.

Building the taxonomy was a large qualitative research effort. The output can be reused and adopted by Wikimedia communities and researchers. The next challenge is to *disseminate* the taxonomy, namely to elevate the awareness of the existence of the taxonomy and to increase interaction points with the taxonomy as part of the existing workflows in the Wikimedia Movement.

2.2. Measure Knowledge Gaps

Once knowledge gaps are systematically defined, the next challenge is to develop ways to measure them. How do we quantify the knowledge gaps in the taxonomy? What is the extent of knowledge gaps across different Wikimedia projects? Which tools do we need to expose quantifiable evidence about knowledge gaps? In this section, we revisit the directions in the original white paper, further scope what we mean by knowledge gaps measurements, and discuss past and future work on this front.

Our initial vision.

In the 2019 paper, the notion of “measurements” is widely present. In "Section 2: Measure and prioritize knowledge gaps,” we define ideas to measure the content dimension of knowledge gaps and propose two research directions to discover missing content in Wikimedia projects. The first aims at mapping missing content across projects based on internal and external knowledge repositories. The second proposes to identify missing knowledge based on readers’ and contributors’ needs, and to build a working model of knowledge equity based on which we can prioritize which gaps to focus on. Similar ideas around the measurement of multimedia knowledge are also present in "Section 4: Multimedia Knowledge Gaps,” and in "Section 5: A Knowledge Equity Index,” where we envision an index providing detailed metrics about knowledge equity.

Our current vision: measuring the state of knowledge gaps.

Our initial understanding of measurement has evolved in the past three years. The original vision conflated two separate ideas under this term: the idea of “quantifying” gaps, and the ideas of “discovering and prioritizing missing content.”

- The “Measuring knowledge Gaps” direction focuses on providing quantifiable evidence about the state of knowledge gaps. We will continue to report research on this front under the current direction.
- Our research on content discovery and prioritization is now captured under the bridging knowledge gaps direction.

Our goal for this direction is to provide metrics and measurements for many of the knowledge gaps in the taxonomy, create snapshots of the state of knowledge gaps over time, and provide tools to explore this data. Setting targets, namely defining what “good” values for knowledge gaps are, is outside the scope of our work and expertise. Through our tools and measurements, we hope to enable the communities closest to the work to set their own targets and monitor the progress towards those targets and goals.

Our contributions so far.

Defining Metrics and Measuring Individual Knowledge Gaps.
Following the above vision, in the past three years, we defined, productionized, and surfaced metrics to quantify readers, contributors, and content gaps. We identified a set of metrics, namely the tools, data, and logic needed to generate measurements about most of the gaps in the taxonomy. We also worked on testing and implementing these measurements on a few fronts.

- ***Readers***: We piloted multilingual Reader Demographics Surveys to capture the distribution of Wikipedia readers by gender, age, and motivations — discovering, for example, that women and men exhibit different topical preferences.
- ***Contributors***: We conducted Editor Gender Surveys across different languages.
- ***Content***: We started implementing content gap metrics for five demographic content gaps: gender, geography, cultural context, time, and sexual orientation. And we started researching metrics for interaction gaps: the multimedia gap and the readability gap.

A Framework for Content Gaps Metrics.
Through the case studies listed above, we defined a multi-faceted framework for developing content gap metrics. The framework requires every content gap metric to capture at least the three following dimensions of coverage:

- ***Selection:*** whether the content is present or not
- ***Extent:*** how much and how good the content is
- ***Framing***: whose priorities and perspectives are reflected in the content

On top of the five demographic content gaps that we have made progress on identifying (Selection), we have also begun prototyping language-agnostic quality models to assist in measuring Extent for the various content gaps. As detailed below, Framing remains largely unexplored but is a direction for future research.

Our plan for future work.

Measuring all knowledge gaps.
With the end goal of quantifying all of the knowledge gaps in the Wikimedia projects, more research and tools are required to systematically measure and monitor disparities in content, contributors, and readers. For example, measuring readership and contributorship gaps requires the infrastructure to periodically survey readers and contributors. We will encourage the creation of such tools within Wikimedia, leveraging existing initiatives around reader and editor surveys. On the content front, while most of our efforts have focused on quantifying the textual content of Wikimedia projects, more measurements are needed to analyze gaps in multimedia content, such as images, videos, and audio files.

The Knowledge Gap Index Tool.
Measuring our progress towards knowledge equity is key. Inspired by our initial ideas of a knowledge equity working model and index, we are committing to building a knowledge gap index. The index will expose our measurements for content, readers, and contributors using an accessible and user-friendly interface. Wikimedians, the Wikimedia Foundation staff, and researchers will be able to use the index to

understand the distribution of readers/contributors/content across different knowledge gap categories or by Wikimedia project, and monitor its evolution over time. With the index, we aim to empower those that we serve directly (the Wikimedia Foundation, affiliates, developer community, and Wikimedia researchers) to set targets, prioritize gaps, and take actions to reach those targets. Today, the knowledge gap index is part of a broader set of efforts from the community and the Foundation around measuring equity in Wikimedia projects. Among these, the Movement Equity Data Landscape, which is gathering data about the movement presence and engagement at country level, was partly inspired by our original knowledge gaps white paper.

Defining framing and content bias.
Content *framing* measurements rely on a body of research that understands how and why biases in Wikimedia content develop. Since literature on this front is still relatively scarce, a first required step is to scope the problem and develop a systematic understanding of content framing. Further research will build on this required layer to then define metrics to quantify the space of content framing.

2.3. Bridge Knowledge Gaps

This direction discusses ways to address Wikimedia knowledge gaps, informed by the systematic definitions and measurements described so far. From research about content tagging, to the deployment of systems for content recommendation, this section describes our efforts to provide insights and tools to bridge knowledge gaps.

Our initial vision.

In "Section 3: Bridge knowledge gaps” of the 2019 white paper, we identified programmatic and technological levers to address known gaps in Wikimedia projects. Programmatic levers refer to the community practices that can help address knowledge gaps. Technological levers include all tools and systems that support the discovery and prioritization of content to address knowledge gaps.

Our current vision.

Most of our research focuses on designing technologies to bridge knowledge gaps, along the lines of what we originally defined as “technological levers.” Operating in harmony with community practices, we are using machine learning, recommender systems, and large-scale data analysis to discover content, patterns, and insights that can help communities address gaps and lower barriers to access and contribute to free knowledge. Our tools and models are designed to be integrated in a machine-in-the-loop system, namely to be used as support tools to enhance and simplify already existing community processes.

Our contributions so far.

Discovering Wikimedia content.
We developed a series of tools and prototypes to **discover** content that can be added to Wikipedia articles. We worked on list-building tools that, given an article title in any language, use language-agnostic approaches to automatically retrieve similar articles that can be created or edited. We collaborated with the Growth and Language teams at the Wikimedia Foundation to develop algorithms to support structured tasks, namely micro-tasks that can help (mainly newcomer) editors easily add content to Wikipedia. We developed an entity-linking system for the add-a-link task, and an image-retrieval framework for unillustrated articles for the add-an-image task. By aligning content in different Wikipedia language editions, we developed a section recommendation algorithm that can suggest new sections to be added to Wikipedia articles in any language. Finally, we launched the Wikipedia image/caption matching competition to accelerate the generation of possible solutions for the problem of automated image captioning and recommendation.

Prioritizing Wikimedia Content
We also worked on studying content **prioritization.** We are working on a systematic understanding of what articles are more important, and which pages need contributors' attention, as there is a misalignment between article quality and readers' demand.

Tagging Wikimedia Content
Finally, we are developing content **tagging** models that can help filter the content to be recommended or prioritized across languages. We built a set of language-agnostic topic filters that can categorize articles and Wikidata items according to their topic, quality, and geographic location.

Tools and systems to address knowledge gaps.
As part of our initiatives to support editors in bridging knowledge gaps, we developed a set of tools that facilitate the visualization and exploration of inequalities in Wikimedia projects. For example, we built the WikiNav tool to navigate the clickstream dataset and understand readers' preferences and scripts for visualizing link gender/quality/clicks *in-situ* on Wikipedia articles.

Our plan for future work.

We will continue to build technologies that lower barriers to access and contribute to the Wikimedia projects.

Bridging Content gaps.
The content of Wikimedia projects can take many forms and languages. Our aim is to continue developing and fostering research on three fronts: Content tagging, with new ways to automatically categorize articles, items, and their sub-components; content prioritization, with novel algorithms to quantify article importance based on reader and editor needs; and content discovery, providing product and campaign teams within the Wikimedia Foundation with models recommending content to be added or improved in Wikipedia and its sister projects. On the multimedia front, we will continue promoting tools research on advanced multimodal systems for image captioning and image to text retrieval.

Tools to Understand Readership and Contributors Gaps.

We aim to continue our collaborations with teams inside of the Wikimedia Foundation, the Wikimedia affiliates, and the developer community to build products that address knowledge gaps. For example, we are currently building tools for better understanding edit diffs and exploring reader data, which build a more systematic understanding of our reader and contributor processes and practices and help communities prioritize and address knowledge gaps.

Equity in Recommender systems.

As our contributions towards machine-in-the-loop systems grow, it is our responsibility to think about how potential biases of automated systems can amplify existing inequalities in the Wikimedia projects. As a starting point, we are studying ways to promote equity and reduce biases through our recommender systems.

3. Beyond the 5-Year Horizon: Ideas for the Future

The aim of this paper is to provide an update about our research directions in the next 3-5 years. However, during our conversations, several ideas for research projects with a longer-term horizon emerged. By sharing the following directions, we hope to generate interest and awareness around research questions that are crucial for the future of Wikimedia projects.

Learning.

One of the main motivations bringing people to Wikimedia sites is intrinsic learning. But are our content, tools, and recommender systems designed to foster and maximize learning? In this context, being able to operationalize and measure learning becomes fundamental. With a few works around understanding the role of images in learning, we have just started scratching the surface, and we will be encouraging and promoting more on the fundamental problem of quantifying learning.

A Model of Wikipedia's Complexity.

Researchers have long studied ways to model complex systems such as climate change, biological networks, or spread of diseases in social networks. While often simplistic in nature, these models have been shown to be extremely useful to understanding the underlying mechanisms and providing decision-makers with quantitative forecasts for different scenarios. The Wikimedia ecosystem and its inner mechanisms represent a complex phenomenon that can be studied using similar methodologies (similar to a recent agent-based model for other peer-production communities). Having a model that reflects Wikipedia's processes could be useful to answer questions about knowledge gaps, design interventions, and imagining scenarios of exceptional growth or decay of readership or contributorship.

Named Entity Recognition in images.

Named Entity Recognition is a well-established task for NLP models that, given a piece of natural language text, such as a Wikipedia article, extracts semantic entities, and potentially links them to a node in a knowledge graph such as Wikidata. Adding machine-readable information to unstructured text is critical for knowledge search, retrieval, and translation. Being able to add semantic tags to images with the same level of granularity would greatly improve how we search and recommend images in free knowledge spaces. The recognition of named entities in text has been widely explored in the natural language processing field. However, the classification of highly granular semantics in images is still at its early stages, except from a few attempts for instance-level recognition at scale, mainly focusing on landmarks and artworks. To foster research on this front, we need more large-scale, richly annotated datasets, connecting images with Wikipedia articles and Wikidata items, as well as more complex image recognition models.

New and External forms of knowledge.

While most of our work focuses on understanding gaps in Wikimedia spaces as of today, Wikipedia and its sister projects are in continuous evolution. In the next few years, we expect projects such as Wikifunctions and Abstract Wikipedia to become prominent in this ecosystem. We will be closely monitoring these spaces and will investigate how knowledge gap research should be adapted and expanded to include these and other developing projects. Moreover, Wikimedia projects do not live in isolation: they are an essential part of the larger web ecosystem. Future research should explore knowledge gap research questions from a macro perspective, understanding how Wikipedia addresses the knowledge gaps of the broader web, and what forms of knowledge Wikimedia needs to acquire in order to fulfill its role in the web.

Acknowledgements

The research described in this update is the result of three years of collaborative work between the current and past members of the Research team, the Wikimedia Foundation's Research Fellow Bob West, formal collaborators (Akhil Arora, Michele Catasta, Giovanni Colavizza, Djellel Difallah, Mo Houtti, Florian Lemmerich, Tiziano Piccardi, Daniele Rama, Rossano Schifanella, Markus Strohmaier, Loren Terveen) and contractors (Jesse Amamgbu, Muniza Aslam, Aiko Chou, Bahodir Mansurov, Marc Miquel).

The Research team would like to thank everyone who has supported us throughout the years with this research: the Site Reliability Engineering, Data Engineering and Release Engineering teams for helping with all the infrastructure aspects of the research, from server access to deployments to hardware supplies, Platform Engineering and Machine Learning Platform for deploying the research code to production, Global Data and Insights for their data and wisdom about our movement, Technical Engagement for the tools and outreach programs, Security, for the support in every single data release. Thanks to the Structured Data, Android, Campaigns, and Growth teams in Product for welcoming our research into their products and to Product Analytics for the continued support, suggestions and data insights. Thanks to Product Design Strategy for helping us with the design of the Knowledge Gaps Index tool. Thanks to the Legal team, for providing us with the frameworks and support for formal collaborations, as well as data collections, competitions, and releases, and to the Communications Team, for believing in our research and helping us disseminate our work. Big thanks also to our Community Relations Specialists, for making sure that our research is in harmony with the community practices and culture. And everyone in the Research and Wikimedia Community who has given love and feedback to the Knowledge Gap projects: thank you all.

This document is the result of brainstorming sessions with the Wikimedia Research team, our Research Fellow Bob West, our contractors and collaborators, as well as the inputs of our Legal, Machine Learning Platform, Traffic, and Major Gifts teams. Thank you all!

Funds for this research are provided by donors who give to the Wikimedia Foundation, by grants from the Siegel Family Endowment and the Argosy Foundation. Thank you!