Title: Funding Knowledge Equity: Enhancing the Visibility, Verifiability, and Sustainability of African Knowledge on Wikipedia

Project Duration: July 1, 2025 – June 30, 2026

Team:

- **Person 1 (Lead)** – Dr Lawrence Ejike Ugwu

- **Person 2** – Candy Tricia Khohliwe

Abstract

Wikipedia aspires to be the "sum of all human knowledge," yet African perspectives remain underrepresented. This one-year project addresses critical knowledge equity gaps by investigating the visibility, verifiability, and sustainability of African content on Wikipedia. We will analyse content coverage across languages, source diversity, and contributor experiences informed by global studies. Rather than piloting interventions, we will consolidate findings from content mapping, surveys, and interviews into policy and practice recommendations. Outputs include peer-reviewed publications, community-informed policy proposals, multilingual training materials, and open datasets, supporting Wikimedia's equity, diversity, and knowledge integrity goals.

Introduction

Despite Wikipedia's global reach, there remains a significant disparity in how African knowledge is represented. For instance, more English Wikipedia articles exist in Paris than in the entire African continent (Sanneh, 2021). African content is disproportionately sparse, with gaps in visibility, representation, and verifiability (Graham, 2009; Prabhala, 2011).

This inequity is rooted in structural, infrastructural, and epistemological barriers. Wikipedia's sourcing guidelines prioritise traditional academic and media publications, which often exclude or devalue African oral traditions, community-held histories, and local sources (Ford et al., 2016). Only 37% of Africans had internet access in 2023 (Ecofin Agency, 2024), and power outages and high data costs continue to impede editor participation and content sustainability.

Beyond technical and infrastructural barriers lies a deeper issue of epistemic injustice (Fricker, 2007). Marginalised knowledge systems, such as oral traditions or indigenous interpretations of history, are often excluded because they do not align with dominant epistemologies shaped by the Global North. This reinforces a Western-centric model of reliability, undermining the representation of alternative knowledge.

In addition to these barriers, Wikipedia editing policies and perceptions of reliable knowledge exclude alternative forms of African epistemology. Limited publishing infrastructure in Africa contributes to a lower prevalence of sources deemed 'reliable' by Wikipedia's standards (Ford et al., 2016), further exacerbating underrepresentation. The growth in internet penetration across Africa (Ecofin Agency, 2024) presents a timely opportunity to support emerging contributor communities before the gap widens further.

Research Questions:

1. What is the current state of African knowledge representation across multiple Wikipedia language editions?

2. What challenges exist in citing African knowledge under existing standards, and how does this affect reliability?

3. What social, technical, and institutional factors influence African editor participation and the sustainability of Africa-related content?

This study responds to Wikimedia's 2030 strategy on Knowledge Equity by undertaking an integrated, Africa-led research effort. We draw on comparative insights from research in Asia, Latin America, and Europe to provide a roadmap for Wikipedia's knowledge equity agenda in Africa.

Related Work

Global research has consistently highlighted the underrepresentation of the Global South in digital knowledge production. Graham et al. (2015) demonstrated that content about Africa remains sparse and is largely authored by non-Africans, revealing significant geographical biases in Wikipedia's content distribution. This disparity underlines systemic biases in digital knowledge production and the need for more inclusive participation.

The Oral Citations Project (Prabhala, 2011) was a pivotal initiative to validate non-textual knowledge, particularly oral histories. Despite its initial promise, the project's impact was curtailed by policy constraints, community scepticism, and implementation challenges (Ford et al., 2016). These challenges highlight the complexities of integrating alternative knowledge forms into established digital platforms.

Further emphasising the challenges faced by contributors from low-resourced language communities, Nigatu et al. (2024) conducted a needs-finding study focusing on Ethiopian languages such as Afan Oromo, Amharic, and Tigrinya. The study revealed that contributors struggle with inadequate language technology support and difficulty sourcing reliable references, which are crucial for content creation on Wikipedia.

In the South African context, Botha (2016) explored the notion of digital equity and social justice, discussing how systemic inequalities affect access to digital resources and participation in knowledge creation. The paper provides a framework for understanding the

social justice implications of digital divides, emphasising the need for contextualised and equitable approaches to digital inclusion.

Addressing the challenges faced by students with disabilities, Zongozzi (2023) reviewed the literature on digital higher education in South Africa. The study identified significant gaps in institutional policies related to digital access, capacity development, and disability inclusion, offering insights into broader digital inclusion issues and equity relevant to knowledge representation on platforms like Wikipedia.

Laniado and Miquel-Ribé (2016) studied identity-based motivation among editors to examine the influence of cultural identities on Wikipedia content. Their research found that cultural identity significantly influences editing patterns, leading to content that reflects contributors' cultural backgrounds. This emphasises the importance of cultural context in knowledge representation and the potential biases arising from dominant cultural narratives.

Collectively, these studies provide a multifaceted understanding of the barriers to equitable knowledge representation on Wikipedia, which informs the methodologies and objectives of this proposal.

However, no prior study has systematically assessed African Wikipedia content using multilingual datasets, contributor interviews, and citation analytics in a comparative and regionalised way. This proposal fills that gap by examining representation, reliability, and sustainability continent-wide.

Methods

We adopt a mixed-methods design consisting of five key phases that are iterative and mutually reinforcing:

1. Content Mapping

We will gather and analyse data dumps and live API queries from four language editions of Wikipedia (Swahili, Hausa, French, Setswana) to identify Africa-related articles. Articles will be tagged using WikiProject Africa templates, geotags, and relevant category trees. Variables of interest include article length, edit frequency, content ratings (e.g., stub vs. good article), and thematic coverage (e.g., historical, biographical, cultural). Tools such as ORES (Objective Revision Evaluation Service) and WikiData queries will be deployed to assess the depth and breadth of content (Wikimedia Research, 2024).

2. Citation Analysis

Citation patterns provide insight into content verifiability. We will perform a source provenance analysis on the references in Africa-related articles to determine the proportion of citations from African publishers, institutions, or authors. Tools for domain parsing and manual tagging will classify citation origin. Metrics will assess citation diversity (e.g., Herfindahl Index) and local-global sourcing imbalances. This step draws on methods from Ford et al. (2016) and earlier citation studies by the Wikimedia Foundation.

3. Surveys

Two multilingual surveys will be developed and distributed:

  i.   **Editor Survey**: This survey targets contributors from African countries and diaspora communities. Items include motivations, source access, reliability perceptions, and community support. It will follow a purposive sampling approach through Wikimedia affiliate mailing lists and user groups.

ii.   **Reader/Knowledge User Survey**: This survey reaches educators, librarians, and students via institutional partners and professional networks. It investigates trust in Wikipedia, the perceived relevance of African content, and editing intentions.

Both surveys will be translated into Swahili, Hausa, French, Setswana. Ethical protocols, including informed consent and voluntary participation, will be developed. We anticipate 100–150 completed responses across at least four countries.

4. Interviews and Focus Groups

We will conduct semi-structured interviews and focus groups to deepen survey insights. Participants will be drawn from diverse backgrounds, including:

i.   Veteran editors and affiliate leaders

ii.   GLAM (Galleries, Libraries, Archives, Museums) professionals

iii.   Students, first-time editors, and digital knowledge activists

Interviews (n=20) and focus groups (n=4) will explore lived experiences, systemic editing barriers, content retention, and source availability. Interviews will be recorded, transcribed, and analysed using NVivo software.

5. Data Analysis

Data integration will be conducted through a concurrent triangulation strategy:

i.   Quantitative survey and content data will be analysed using descriptive statistics, logistic regressions, and content scoring metrics in SPSS and R.

ii.   Qualitative interview/focus group transcripts and open-ended responses will be thematically analysed using NVivo.

iii.   The mixed-method synthesis will inform training recommendations, policy implications, and community strategies.

This approach ensures actionable outputs, community relevance, and empirical rigour. It also builds on earlier Wikimedia-funded studies and methodology from the "Open the Knowledge" research series (Wikimedia Foundation, 2021).

Expected Outputs

By the end of the one-year project, we anticipate delivering the following in line with open knowledge principles:

We will produce two peer-reviewed research articles published in open-access journals or conference proceedings, contributing to a scholarly understanding of knowledge equity and African representation in digital knowledge platforms. Open datasets and analytic tools will be released to enable ongoing analysis and replication, including structured data on citation sources and content coverage. A multilingual training toolkit will be created to support editors in contributing African knowledge to Wikipedia, including guides, videos, and community-curated source lists. We will also deliver a white paper detailing policy recommendations for Wikimedia affiliates and the Foundation, highlighting source policy reforms and strategies for editor retention. Community engagement outputs such as blog posts, a public webinar, and Wikimania/WikiIndaba presentations will ensure broad visibility and community feedback. All resources will be openly licensed.

Risks and Mitigation Strategies

We anticipate the following risks and outline our mitigation strategies:

Low community engagement may pose a challenge, especially in surveys or interviews. We will mitigate this through the early involvement of Wikimedia affiliates, the use of trusted community liaisons, and modest incentives where appropriate. Resistance to including non-traditional sources may arise within Wikimedia policy circles; we will address this by

presenting empirically grounded evidence and involving community leaders in framing recommendations. Technical risks related to citation analysis will be addressed through piloting and collaboration with experienced Wikimedia researchers. Language barriers will be mitigated by hiring professional translators and involving native-speaking Wikimedians for validation. Data protection will be ensured through institutional ethics clearance, anonymised datasets, and secure data storage practices.

Community Impact Plan

This project is deeply rooted in community empowerment. We will collaborate with at least four Wikimedia affiliate groups across West, East, and Southern Africa. Community advisors will co-design tools, assist in validation, and co-lead dissemination. We will develop a "Train-the-Trainer" initiative to foster long-term skill transfer. Our findings will be localised through workshops, roundtables, and multilingual summaries to ensure relevance. Source lists and editing guides will be co-developed and hosted on Meta-Wiki, ensuring ongoing reuse. Community validation and co-authorship will be encouraged in publications and outputs, fostering ownership and long-term sustainability. The project also aims to build networks between editors, researchers, and cultural institutions, laying the groundwork for future collaborations.

Evaluation

Project evaluation will adopt a utilisation-focused framework. We will assess the achievement of research goals using quantitative and qualitative milestones. Research activities will be monitored through a project dashboard. Participation indicators (e.g., survey responses, training attendees, and interview count) will be tracked monthly. Knowledge equity indicators will include article improvement metrics, local source usage, and diversity

of content. Community impact will be evaluated through pre/post-training feedback, focus

group reflections and follow-up engagement. Uptake of policy recommendations by

Wikimedia stakeholders will be monitored through citations, endorsements, and

incorporation into strategy documents. A comprehensive final report will summarise

outcomes, lessons learned, and pathways for future work.

Budget

- **Research Lead** (12 months, part-time) – $9,000
- **Research Co-lead** (12 months, part-time) – $6,000
- **Translation** (Swahili, Hausa, French, Setswana $200 per language) – $800
- **Community Liaisons** (3) – $900
- **Public Webinar** – $500
- **Software Licenses** – $500
- **Data Storage & Security** – $500
- **Publication Fees** (2 open-access articles) – $1,500
- **Contingency Fund** – $500
- **Conference Funds** – $1,800
- **Fiscal Sponsor Fee (15%)** – $3,300
- **Total** : $25,300

References

Botha, J. (2016). *Digital equity and social justice: Whose reality? Reflections from South

   Africa*. In Show Me The Learning: Proceedings ASCILITE 2016 Adelaide (pp. 76–

   85). Retrieved from https://2016conference.ascilite.org/wp-

   content/uploads/ascilite2016_botha_full.pdf2016conference.ascilite.org

Ecofin Agency. (2024, February 27). *Africa's internet use doubles in a decade despite high

   costs – ITU report*. Retrieved from https://www.ecofinagency.com/

Ford, H., Sen, S., Musicant, D. R., & Miller, N. (2016). *Getting to the source: Where does Wikipedia get its information from?* In Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (pp. 836–847). https://doi.org/10.1145/2818048.2819937

Fricker, M. (2007). *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press.

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

Graham, M., Straumann, R. K., & Hogan, B. (2015). *Digital Divisions of Labor and Informational Magnetism: Mapping Participation in Wikipedia*. Annals of the Association of American Geographers, 105(6), 1158–1178. https://doi.org/10.1080/00045608.2015.1072791

Laniado, D., & Miquel-Ribé, M. (2016). *Cultural Identities in Wikipedias*. In Proceedings of the 12th International Symposium on Open Collaboration (pp. 1–10). https://doi.org/10.1145/2957792.2957795

Nigatu, H. H., Canny, J., & Chasins, S. E. (2024). *Low-resourced Languages and Online Knowledge Repositories: A Need-Finding Study*. arXiv preprint arXiv:2405.16669. https://arxiv.org/abs/2405.16669​:contentReference{index=16}

Prabhala, A. (2011). *Oral citations project*. Meta-Wikimedia. Retrieved from https://meta.wikimedia.org/wiki/Research:Oral_Citations​:contentReference{index=17}

Sanneh, A. (2021). *Wikipedia's cultural imbalance*. Statement via WikiAfrica. Retrieved from

https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Africa​:contentReference{index=18}

Wikipedia. (n.d.). *WikiProject Africa: Africa Sources List*. Retrieved from https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Africa/Africa_Sources_List​:contentReference{index=19}

Wikimedia Foundation. (2021). *Open the Knowledge: Research Series – Knowledge Equity: How far do we have to go?* Retrieved from https://wikimediafoundation.org/​:contentReference{index=20}

Wikimedia Foundation. (2022a). *Knowledge Equity Research – South Africa*. Retrieved from https://wikimediafoundation.org/​:contentReference{index=21}

Wikimedia Foundation. (2022b). *Knowledge Equity Research – Nigeria*. Retrieved from https://wikimediafoundation.org/​:contentReference{index=22}

Wikimedia Research. (2024). *Address Knowledge Gaps*. Wikimedia Research Programs. Retrieved from https://research.wikimedia.org/​:contentReference{index=23}

Zongozzi, J. N., & Ngubane, S. A. (2025). *Equitable access to digital higher education for students with disabilities in South Africa*. African Journal of Disability, 14, a1525. https://doi.org/10.4102/ajod.v14i0.1525