Tracking the Reuse and Public Domain Status of Works on Wikimedia Commons

Isaac Oloruntimilehin
Free Knowledge Africa

Alaafiabami Oladipupo
Free Knowledge Africa

Bolaji Alabi
Free Knowledge Africa

Abstract

This research proposes to investigate how public domain-tagged works such as books, music, sound recordings, newspapers, artworks, and monuments uploaded on Wikimedia Commons are being shared and reused globally. It will also identify the geographic distribution and thematic areas with the highest concentration of these works, providing critical insight into gaps and opportunities for African cultural heritage contributions. As a core component of this project, we will develop a suite of custom Wikidata SPARQL queries to track public domain-tagged works on Wikimedia Commons. These queries will retrieve metadata such as creator death dates, publication dates, geographic origins, and reuse instances, enabling the identification of works nearing or entering the public domain based on country-specific copyright laws.

Introduction

Despite growing digitization efforts in African countries, there is limited data on how public domain cultural works uploaded to Wikimedia Commons are reused or where these works are geographically concentrated. This hinders advocacy for open access and the strategic digitization of African cultural assets. By analyzing reuse patterns and tracking public domain status and tagging based on country-specific copyright laws, this project will close a crucial gap in Wikimedia's knowledge infrastructure and enhance the visibility of African content within the Commons and Wikidata ecosystems.

Research Questions:

- How are public domain works on Wikimedia Commons being reused across Wikimedia projects and beyond?
- What are the geographic distributions of public domain works currently uploaded, particularly focusing on Africa?
- Which categories (books, sound recordings, music, monuments, etc.) are most represented?
- How can a Wikidata-powered query support the calculation of public domain eligibility dates using metadata such as author death dates and copyright laws?

Date: July 1, 2025 - June 30, 2026

Related work

There is a growing body of research on open access, digital preservation, and public domain

content within Wikimedia projects and open GLAM initiatives, both in Nigeria and internationally. Some of the most relevant work includes:

- **Reuse of Wikimedia Commons Cultural Heritage Images on the Wider Web (Kelly,2019):**
  The study analysed the reuse of cultural heritage images from Wikimedia Commons, finding that they were most frequently repurposed within Wikimedia projects (e.g., wikis, aggregators, and mirror sites). Outside Wikimedia, common reuse platforms included news websites (particularly for images of notable people and historical events), social media (e.g., Pinterest), media galleries, and educational sites. However, many instances lacked proper attribution or contextual information. The study used Reverse Image Lookup (RIL) via Google Images to track 171 images from 51 institutions, documenting reuse patterns, licensing accuracy, and contextual trends. The findings highlight both the broad dissemination of open-access cultural heritage images and the ongoing challenges in ensuring correct attribution, providing valuable insights for institutions managing digital collections.

- **Case Study of Open Access Practices: Limitations and Opportunities in Public Libraries in Nigeria (Oloruntimilehin, 2023):**
  This case study examined factors influencing the adoption of open access practices in a Nigerian public library, offering insight into metadata management and digital access challenges relevant to this project's

focus on Wikimedia Commons.

- **Collaborate, Communicate, and Navigate Ethical Considerations with the Ethical Sharing Card Game (Oloruntimilehin et al., 2023):**
  This resource provided an innovative, gamified approach to teaching ethical considerations around open licensing, a topic closely connected to the responsible tagging and reuse of public domain materials online.

- **The Public Domain: Enclosing the Commons of the Mind (Bollier, 2002):**
  The study emphasises the critical role of the public domain in fostering creativity and cultural growth. He argues that it provides a foundational resource for new artistic and intellectual works, enabling creators to build upon existing ideas without copyright restrictions. By offering access to diverse cultural and scientific materials, the public domain encourages innovation, collaboration, and cross-cultural exchange. It also challenges conventional notions of ownership, promoting creativity as a shared, collective practice rather than a profit-driven endeavour. Additionally, the public domain lowers barriers for emerging artists, allowing them to participate in creative economies and contribute fresh perspectives.

Despite this growing research interest, significant gaps remain in understanding how public domain works uploaded to Wikimedia Commons are being reused, categorized, and distributed geographically, especially for African content. Additionally, the development of tools for tracking public domain status based on copyright laws and metadata has not yet

been fully addressed within the Wikimedia or OpenGLAM communities — this project seeks to fill those gaps.

Methods

This research project will combine quantitative data analysis, metadata mapping, and tool development approaches to achieve its objectives:

- **Data Collection:**
  We will extract data from Wikimedia Commons, focusing on public domain files across different media types (books, sound recordings, artworks, newspapers, monuments, etc.) tagged with the public domain license. This will involve using Wikimedia Commons categories, license tags, and linked Wikidata items. Corresponding metadata — including upload date, category, creator, license details, and geolocation tags — will be collected using Wikimedia Commons API and Wikidata SPARQL queries.

- **Usage and Reuse Tracking:**
  File reuse data will be gathered from Wikimedia Commons file usage reports and SPARQL queries tracking file transclusions and external linkbacks on Wikimedia projects (Wikipedia, Wikivoyage, Wikisource, etc.). This will help identify which files are most reused and in what contexts.

- **Geographic Mapping:**
  Using geotags embedded in Commons metadata and Wikidata location properties, we will map the geographical distribution of public domain works. This spatial data will allow for visualisations of content concentration by region, with a

particular focus on underrepresented African countries.

- **Tool Development:**
  We will design a prototype public domain tracking tool powered by Wikidata. The tool will automatically calculate the public domain status of works based on metadata such as author death dates, publication years, and country-specific copyright laws. It will append the appropriate public domain tags for different jurisdictions (e.g. Nigeria, US, UK) using predefined legal rules. The tool will leverage Wikidata's structured metadata and existing public domain tags on Commons.

- **Analysis:**
  Data analysis will include:
  - Descriptive statistics on the number of works per category, geography, and reuse frequency.
  - Network analysis to visualize the spread and linkage of public domain works across Wikimedia projects.
  - Thematic content analysis of file categories to identify areas of concentration and underrepresentation.
  - User feedback from community engagement sessions to inform recommendations for better tagging, metadata curation, and public domain management on Wikimedia platforms.

The combination of data extraction, mapping, tool development, and community feedback will enable us to provide actionable, evidence-based insights and a working prototype for Wikimedia Commons and Wikidata contributors.

Expected output

Expected Outputs

1. **Comprehensive Research Report**
   - **Description:** A detailed report documenting the reuse patterns, categories, and geographic distribution of public domain works on Wikimedia Commons, with a particular focus on African contributions.

   - **Audience:** Wikimedia affiliates, GLAM professionals, digital humanities researchers, and open knowledge advocates.
   - **Benefit:** It will provide evidence-based insights to inform decision-making, identify gaps, and support advocacy for digitization and open access in African and global Wikimedia communities.
   - **Target Venues:** Wikimania, OpenGLAM platforms, and Wikimedia Diff blog.

2. **Public Dataset**
   - **Description:** A structured, open-access dataset of public domain works on Wikimedia Commons, categorized by geography, type, and reuse frequency, with accompanying metadata.
   - **Audience:** Wikimedia Commons contributors, Wikidata editors, academic researchers, and GLAM institutions.
   - **Benefit:** Enables further research, improves content

discovery, and supports Wikimedia campaigns and GLAM digitization efforts.
   - **Target Platforms:** Wikimedia Commons, Wikidata, Zenodo, and the Wikimedia Research portal.

3. **Prototype Public Domain Tracker**
   - **Description:** A functional prototype tool powered by Wikidata that predicts when works will enter the public domain based on metadata (creator death dates, release dates, national copyright laws) and recommends country-specific public domain tags.
   - **Audience:** Wikimedia Commons contributors, Wikidata editors, GLAM institutions, and open culture technologists.
   - **Benefit:** Streamlines public domain tagging practices, ensures legal accuracy, and enhances the discoverability of works — particularly valuable for African archives and libraries.
   - **Target Channels:** Wikimedia Commons and Wikidata tool directories, as well as Wikimedia mailing lists.

4. **Policy and Practice Recommendations**
   - **Description:** A document offering recommendations for Wikimedia Commons and Wikidata communities on how to improve public domain uploads, metadata practices, and tagging — especially in African contexts where

copyright terms and documentation vary widely.

- ○ **Audience:** Wikimedia Foundation policy teams, GLAM-Wiki organizers, Wikimedia affiliates, and copyright advocacy networks.
- ○ **Benefit:** Strengthens community workflows, supports legal compliance, and amplifies underrepresented African cultural heritage on Wikimedia projects.
- ○ **Target Venues:** Wikimania, OpenGLAM conference presentations, and Wikimedia Movement Strategy forums.

Risks

The proposed research project anticipates several potential challenges. First, **incomplete or inconsistent metadata** on Wikimedia Commons entries could limit the accuracy of data collection and analysis. This will be mitigated through manual validation and community-driven metadata improvement sprints in collaboration with GLAM-Wiki volunteers. Second, **the complexity and variability of copyright laws** across different countries may affect the reliability of public domain expiry predictions. To address this, the project will consult with legal experts, including Creative Commons Nigeria and the Wikimedia legal team, to validate country-specific copyright terms and ensure accurate tagging logic. Third, there is a risk of **limited community adoption and technical sustainability** for the developed tool. To mitigate this, the project will engage the Wikidata, GLAM-Wiki, and Wikimedia Nigeria communities early in the design process, gather regular feedback, and build awareness through workshops and open documentation.

Community impact plan

This project will directly engage Wikimedia Nigeria User Group, Africa Commons, Creative Commons Nigeria, and GLAM-Wiki communities throughout its lifecycle. We will involve these communities early through consultations, surveys, and collaborative mapping exercises to validate public domain data and improve metadata quality. A dedicated workshop will be organized for Wikimedia editors, Wikidata contributors, GLAM professionals, and public librarians to demonstrate the research findings, train them on the prototype tool, and build capacity in public domain tagging practices, especially for African works.

We also plan to share our findings through presentations at Wikimania, regional Wikimedia events like WikiIndaba, and OpenGLAM webinars, targeting African affiliates, cultural institutions, and policy advocates. The resulting dataset and insights will help inform local digitization priorities, advocacy for better copyright reforms, and targeted Wikimedia campaigns like Public Domain Day in Africa. Additionally, open documentation and how-to guides will empower local Wikimedia volunteers and institutions to adopt the tool and improve public domain metadata practices across Wikimedia Commons and Wikidata.

Evaluation

The success of this project will be evaluated based on both quantitative and qualitative indicators. Key measures will include:

- **Data Coverage**: Number of public domain works identified, categorized, and analyzed from Wikimedia Commons, with particular attention to African contributions.

- **Reuse Mapping**: Number of reuse instances documented across Wikimedia projects and external platforms.
- **Tool Performance**: Functionality, accuracy, and reliability of the prototype public domain tracking tool, tested with sample datasets and reviewed by GLAM-Wiki communities.
- **Community Engagement**: Number of Wikimedia affiliates, GLAM professionals, and Wikidata contributors participating in workshops, surveys, and tool testing activities.
- **Knowledge Dissemination**: Completion and public availability of the final research paper, dataset, and open-source tool, alongside presentations delivered at Wikimania, WikiIndaba, and OpenGLAM events.
- **Adoption Readiness**: Feedback and interest from Wikimedia Commons and Wikidata communities regarding the utility of the tool and datasets, informing potential future improvements.

These metrics will help assess the scholarly, technical, and community value of the research while ensuring alignment with Wikimedia's 2030 Movement Strategy and GLAM priorities.

Budget

Total Budget Request: $40,300

🟩 FKA Research Fund Budget 2025

References

Bollier, David. "The Public Domain: Enclosing the Commons of the Mind." New America Foundation & Public Knowledge, May 2002.

Kelly, E. J. (2019). Reuse of Wikimedia Commons Cultural Heritage Images on the Wider Web. Evidence Based Library and Information Practice, 14(3), 28–51. https://doi.org/10.18438/eblip29575

Oloruntimilehin, Isaac. "Case Study of Open Access Practices: Limitations and Opportunities in Public Libraries in Nigeria." *Medium*, Creative Commons, 2 March 2023, https://medium.com/creative-commons-we-like-to-share/case-study-of-open-access-practices-limitations-and-opportunities-in-public-libraries-in-nigeria-a37a4aab6edb.

Oloruntimilehin, Isaac, and Bart Magnus. "Collaborate, Communicate, and Navigate Ethical Considerations with the Ethical Sharing Card Game." *Medium*, Creative Commons, 26 April 2023,

https://medium.com/creative-commons-we-like-to-share/collaborate-communicate-and-navigate-ethical-considerations-with-the-ethical-sharing-card-game-20623d3f630b.