Biodiversity Knowledge Gaps on Wikipedia: A Cross-Lingual Analysis of Species Coverage and Contribution Patterns

Tiago Lubiana
Biodiversity Heritage Library – Smithsonian Libraries and Archives
Wikimedia Brazil

Abstract

The Wikimedia projects function as a key infrastructure in the open knowledge ecosystem and a crucial role to play in tackling the current biodiversity crisis. To support this mission, this proposal aims to **detect critical biodiversity knowledge gaps analyzing the coverage and contribution patterns for species across multiple Wikipedia language editions**. We will quantify which species are documented (and where), identify patterns of content contribution, and examine how different communities contribute to biodiversity topics. By illuminating underserved areas—such as local species missing from Portuguese Wikipedia or entire taxonomic groups overlooked globally— we will generate actionable strategies to improve and balance content**.** The project is led by a Brazilian post-doctoral researcher and long-time Wikimedian, ensuring both academic rigor and deep community insight and will be advised by an experienced group of Wikimedians and biodiversity scholars at the Biodiversity Heritage Library - Wikimedia working group. By using methods from biodiversity informatics and leveraging Wikidata, the Global Biodiversity Information Facility and the Catalogue of Life, the project will generate generalizable insights on the patterns of Wikipedia contributions while supporting individual volunteers and Wikimedia community members. Identification of critical biodiversity knowledge gaps, thus, is a move towards Wikimedia's 2030 strategic priority to include knowledge and communities that have been left out by structures of power and privilege. With a particular **focus on English, Portuguese and Spanish Wikipedias** and the Latin American communities, the end goal is to provide the means **scaling up the capacity for informed contributions on biodiversity, particularly for neglected species and underrepresented communities** providing the basis for targeted actions for a more equitable "encyclopedia of life", catering for the diversity of human and non-human life.

Introduction

Biodiversity is fundamental to our planet's health and humanity's well-being. From habitat loss in tropical rainforests to coral reefs at risk, the biodiversity crisis increases the need for broad public understanding and appreciation of the beauty, complexity and intrinsic value of the diverse manifestations of life on Earth.

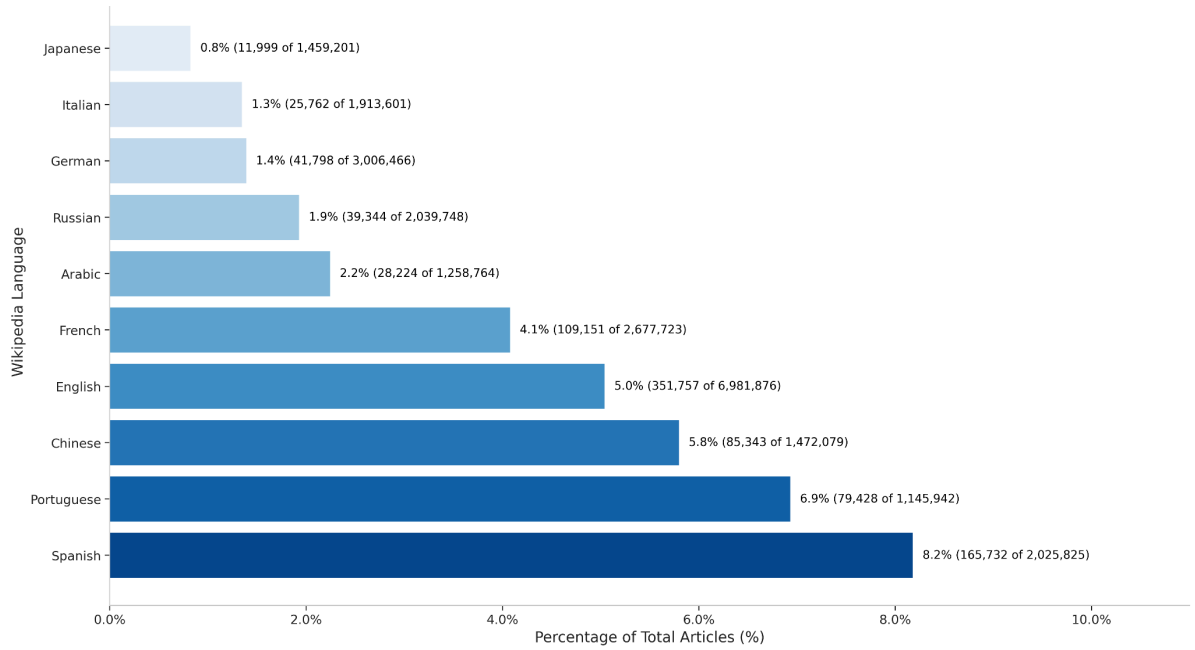


Figure 1: Percentage of species articles on the 10 Wikipedia editions with most users. Values in parenthesis indicate the number of species articles out of the total number of articles. While absolute counts figure quite below English on Portuguese and Spanish Wikipedias, the percentage of articles about species is above, possibly indicating more attention of these communities to biodiversity topics. The source for the figure is available at https://github.com/lubianat/biodiversity_research_fund_2025.

(Cardinale et al., 2012) Wikipedia, as one of the world's most visited sources of information, plays a key role in this ecosystem. (Dearborn, 2023) Ensuring that knowledge about biodiversity is freely available in every major language is crucial for education, conservation awareness, and empowering communities to protect their natural heritage.

Currently, the coverage of species information on Wikipedia varies dramatically between languages. Well-resourced editions like English have hundreds of thousands of articles on individual species, while many other language Wikipedias have far fewer. (**Figure 1**) For example, English Wikipedia features over 350,000 entries about species, while the Portuguese Wikipedia features less than 80,000— in spite of the biodiversity richness of Portugal, Brazil, Angola and other lusophone regions. Moreover, content that does exist might be skewed toward charismatic megafauna and other flagship species, whereas lesser-known or locally significant species remain undocumented in most languages. **The extent and nature of this underrepresentation, however, is unknown, as systematic research on these knowledge gaps is missing.**

These gaps and biases undermine the principle of knowledge equity, which is at the heart of Wikimedia's 2030 strategic direction. (Wikimedia Foundation, n.d.) Knowledge equity calls for focusing our efforts on the knowledge and communities that have been left out by structures of power and privilege. In the context of biodiversity, this translates to elevating knowledge of species and ecosystems from the Global South, indigenous knowledge systems, and non-English sources that historically have

been marginalized. Studying biodiversity knowledge gaps is as much about topic coverage as it is about empowering underrepresented communities with access to local biodiversity information in their own languages .

Our proposed research, *"Biodiversity Knowledge Gaps on Wikipedia,"* directly addresses this challenge. We will conduct a cross-lingual analysis of taxonomic coverage, quality, and patterns community contributions on Wikipedia to provide insights on the current state of biodiversity content across different editions. Reassuringly, previous research has shown a correlation of content for particular Wikipedia language editions and the cultural background of the communities, (Miquel-Ribé & Laniado, 2018) which hints at a likely variation of content about species depending on the geographical location of contributors for each edition. To find out these patterns, our proposal will combine Wikidata with biodiversity informatics resources, such as the Global Biodiversity Information Facility (GBIF) and the Catalogue of Life (CoL), which will provide the missing pieces for a detailed analysis (Bánki et al., 2025; The Global Biodiversity Information Facility, 2025)

Encouragingly, **our work aligns with and will support existing movement initiatives focusing on biodiversity knowledge and underrepresented communities.** The annual WikiForHumanRights campaign, for instance, emphasizes the right to a healthy environment and has mobilized Wikimedians to create and improve articles on climate change, pollution, and biodiversity as a human right. (*WikiForHumanRights* , n.d.) These perspectives are also echoed in Latin America directly by the Climate Justice Working Group (Scann WDU & Wikimedistas de Uruguay, 2023). The recent #1Pic1Bio events, hosted by the Wikimedia Foundation in partnership with the Biodiversity Heritage Library,[1] were organized in early 2025 focusing on the bridging knowledge gaps in biodiversity content in Africa and South America.

To gather insights that cater both to researchers and these Wikimedia communities, we tailored **our research questions around two main axes: (1) coverage and representation and (2) contribution patterns.** Both subjects are rich in complexity and possible research paths as well as in gaps in available information. For the 3 Wikipedia editions under study (Spanish, English and Portuguese), some of the leading questions related to the first axis include

- Are any taxonomic groups systematically underrepresented on particular Wikipedia editions?
- How are featured/good articles for taxa distributed in the tree of life?
- Are endemic species from the Global South less represented on Wikipedia?
- Is there a correlation between species occurrence patterns and content on particular Wikipedia language editions? An example for bird families in Latin America is shown on **Figure 2.**

For the same set of Wikipedias, some of the questions related to the second axis include:

- Do the contributors creating species articles focus on particular biological groups?
- Do they edit only about species, or also about other topics?

[1] https://meta.wikimedia.org/wiki/Event:1Pic1Bio_(Spanish)

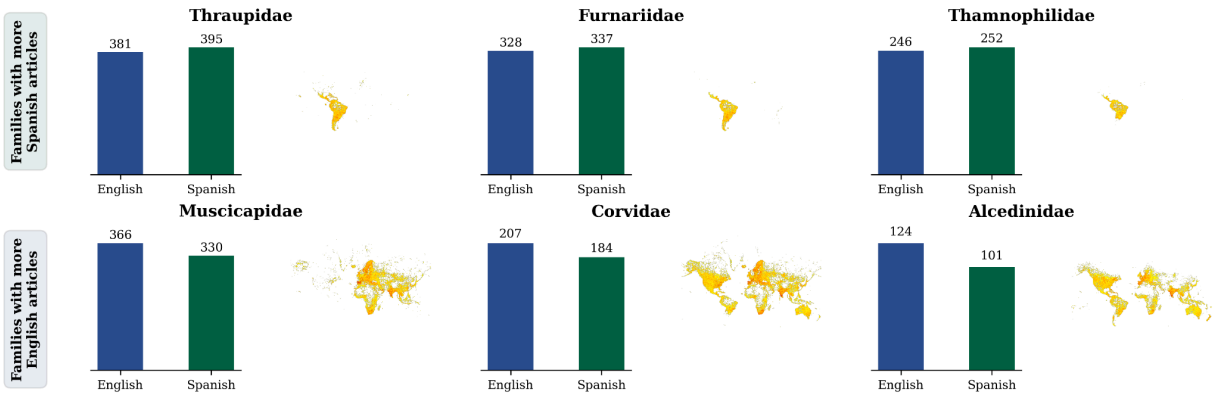


Figure 2: Bird families from Latin America seem to have better coverage on Spanish Wikipedia. Bar plots show the sitelink counts (via QLever) for species of birds on either English or Spanish Wikipedia. Maps display the occurrence patterns for bird families as available via GBIF. The 3 families with biggest differences between editions are shown. The patterns indicate a correlation between coverage on Spanish Wikipedia and species occurrence in Latin America and the Caribbean. The source for the figure is available at https://github.com/lubianat/biodiversity_research_fund_2025.

- Do these editors focus on a single language edition or work across multiple languages?
- Do the Wikipedia communities under study rely on automated systems (Wikidata, bots) for taxonomic information?

By targeting this range of questions, associated with iterations and feedback from the Wikimedia Biodiversity community, our project outputs will be both of academic significance and of immediate usefulness for Wikimedians.

Date: The planned period for this research project is from July 1, 2025 to March 31, 2026 (nine months).

Related work

Addressing Wikipedia's knowledge gaps requires building on insights from both the Wikimedia research community and external scientific studies. The Wikimedia Foundation's Research team has developed a taxonomy of knowledge gaps that spans dimensions of content, representation, and contributors. (Redi et al., 2021) **Of particular relevance to our proposal are *content gaps* (topics not adequately covered on Wikimedia) and *representation gaps* (knowledge from certain regions or languages being underrepresented).** Biodiversity sits at the intersection of these: it's a topic area where content can be missing and where knowledge is often skewed towards certain languages or regions (a representation gap—e.g., South American biodiversity described mostly by English sources). By framing our work in this taxonomy, we ensure a systematic approach to what "gaps" mean, moving beyond anecdotal observations to measurable gaps that we can target. Moreover, besides analyzing presence/absence of articles as measures of content gap, prior research has also introduced language-independent quality models which we can leverage to score articles uniformly across Wikipedias. (Das et al., 2024a)

The academic community has also turned attention to Wikipedia as a lens for understanding public interest in biodiversity

and conservation. Researchers have studied the mining of digital data for conservation insights, and Wikipedia has proven to be a rich source of such data. (Correia et al., 2021). A study by Roll et al. in 2016 analyzed Wikipedia page views for reptile species across many languages and found that Wikipedia could serve as a proxy for which species people find interesting or valuable, revealing biases— venomous snakes, for example, receive more attention than other species. (Roll et al., 2016) Building on that, Mittermeier and colleagues examined seasonal patterns in Wikipedia interest, discovering that page views for certain species follow annual cycles, and that these patterns vary across language editions, reflecting local cultural and geographic context. (Mittermeier et al., 2019) In 2021, the research was extended to analyze public interest in bird species across 251 Wikipedia languages, showing that while an impressive 95% of bird species had at least one Wikipedia article, the level of detail and the amount of page views differed widely. (Mittermeier et al., 2021) **Our project takes inspiration from these works and focuses not just on the demand side (readership patterns) but on the supply side (editorship and content creation).**

The use of Wikidata as a hub for connecting biodiversity informatics and Wikipedia is also supported by a growing corpus. (Mering et al., 2024; Mittermeier et al., 2021; Page, 2022; Paul, 2020; Taraborelli, 2019) Of note, this research intersects with other known content gaps of the Wikimedia ecosystem, namely the Gender Gap, as Sabine von Mering and colleagues used Wikidata for building a case study of plant genera named for women, and subsequently using the insights for bridging the gaps on English Wikipedia. (Mering et al., 2024) These studies highlight the value of Wikidata for biodiversity informatics as a whole and the value of leveraging it as a source of connection between the Wikimedia ecosystem and external taxonomic databases.

Besides traditional research outputs, The Wikimedia movement has also seen GLAM (Galleries, Libraries, Archives, Museums) collaborations aimed at enriching biodiversity content. A flagship example is the partnership with the Biodiversity Heritage Library (BHL), the world's largest open-access digital library of biodiversity literature. It has embraced a "mutualistic" partnership with Wikimedia to make biodiversity knowledge accessible to all.[2] This has included hosting, with support of the Smithsonian Institution, a Wikimedian-in-Residence (the proponent of this research project) and forming a BHL-Wiki Working Group. (Dearborn, 2024) One outcome of the Wikimedian-in-Residentship has been the BHL image reuse tool, which helps volunteers to find where BHL-sourced images could be added on Wikipedia, focusing on bridging geographical disparities.[3]

Methods

To achieve our objectives, we will rely on a combination of data science and biodiversity informatics techniques with the data landscape of Wikimedia available via Wikidata and Quarry. At the core, our methodology involves **quantitative analysis of Wikimedia content and contributor data across multiple languages for biodiversity topics** Here we outline our planned steps in detail.

1. Taxonomic Coverage - Data Collection and Analysis:

We will start by building a comprehensive list of taxonomic entries to analyze. Using Wikidata as a base reference, we can query all items that are

[2] https://meta.wikimedia.org/wiki/BHL
[3] https://bhl-gallery.toolforge.org/

instances of “taxon” (Q16521) with taxon rank “species” (Q7432). This query currently yields on the order of 3.1 million items[4] representing species. A particular **focus will be given to the 2 million species with both a Global Biodiversity Information Facility (GBIF) and a Catalogue of Life identifier on Wikidata**. [5] The GBIF data, available on a public API[6] , which will be used to detect gaps for key groups based on geographical location of species, as well as the conservation status (the IUCN Red List is integrated into the GBIF’s data). The latest version of the Catalogue of Life (April 2025[7]) will be downloaded and parsed locally . For speed, the QLever Wikidata endpoint will be used as a source of links between taxa, GBIF/Catalogue of Life, and different Wikipedia pages. From the set of Wikidata ids, we will extract the set of Wikipedia sitelinks associated with each taxon item, essentially mapping which taxa has an article in which target language edition (Portuguese, Spanish and English).

With the data in hand, our first analytical step is to quantify coverage. For each Wikipedia language edition in our study we will calculate: (a) the on-wiki coverage related to a series of taxonomic groups of interest, (b) the coverage for species endemic to certain regions (e.g. South America or Africa) and (c) the number of “Featured” or “Good” articles for species. These different pieces of information may be crossed to yield combinatory insights (e.g. occurrence of “Featured” articles about birds in different Wikipedias).

The use of “Featured” or “Good” by the Wikipedia communities will be used as a proxy for articles of exceptional quality. This information is available as Badges on Wikidata[8]and will be retrieved via QLever. There are currently around 4.1 thousand articles about species tagged as either “Featured” or “Good”[9] over all Wikipedia editions. The data from the badges will be complemented by quality scores from automated models, namely the 2024 language-agnostic model by Das and colleagues, (Das et al., 2024b), which will be used as a proxy to compare the quality of articles.

The Catalogue of Life (CoL) will serve as the gold standard for taxonomic classification in determining which species belong to which groups. While Wikidata’s species hierarchy may be used as a baseline,CoL will be the ultimate reference source of truth, guaranteeing results are aligned with the academic standard.

To assess whether differences in taxonomic coverage between language editions (e.g., English vs. Portuguese Wikipedias) are statistically significant, we may employ tests such as the chi-squared test of independence. For example, we may construct contingency tables comparing the proportions of articles across different taxonomic groups for each language edition. A significant result ($p < 0.05$) would indicate that the proportions of species coverage differ substantially between the languages, highlighting potential taxonomic biases or gaps specific to each Wikipedia edition. The final statistical methodology will be tailored to fit the needs of each hypothesis testing event.

3. Contributing Landscape - Data Collection and Analysis:

We will analyze contributor patterns for biodiversity content, with an initial focus on article creation. For each Wikipedia language,

[4] https://qlever.cs.uni-freiburg.de/wikidata/1yVwry
[5] https://qlever.cs.uni-freiburg.de/wikidata/L82LEu
[6] https://techdocs.gbif.org/en/openapi/
[7] https://www.checklistbank.org/dataset/309120
[8] https://www.wikidata.org/wiki/Help:Badges
[9] https://qlever.cs.uni-freiburg.de/wikidata/o8ISTg

we will identify creators for species articles either via Quarry[10] or Wikimedia dumps.

We will cross this information with particular taxonomic groups and with cross-language data to detect patterns of activity (e.g., focus on particular taxonomic groups, cross-wiki editing or focus solely on biodiversity or wider Wikipedia contributions)

Another aspect we will evaluate is the use and design of species infoboxes and, in particular, on how they rely on information from Wikidata There are two templates used across language editions, Template:Taxobox (Q52496) and Template:Speciesbox (Q14449650). We will assess the prevalence (via Quarry) and monitor their reliance on Wikidata on the different Wikipedias (e.g. for images, taxon ranges and identifiers). This analysis will base the development of scripts to parse across the wider set of Wikipedia editions **detecting the extent in which coverage of biodiversity on Wikidata influences content on Wikipedias**. Having a better idea of this knowledge flow may support Wiki stakeholders with insights on whether and how contributing to Wikidata may impact content on their home Wikipedias.

Tool Development

Besides data publication and written reports, we will also prototype a dashboard, to be deployed on Toolforge, allowing the Wikimedia community to explore the biodiversity coverage statistics we compiled in a less technical way. The dashboard will be written in Python using the Flask framework, similar to work previously done by the proponent (https://bhl-gallery.toolforge.org). The dashboard will serve a dual use of navigating the data landscape in an autonomous way, as well as offer direct suggestions for contributions.

[10] https://quarry.wmcloud.org/query/85509 (example for Aug 2024, with authors of 330,000 species articles)

Expected outputs

Research Report and Publication

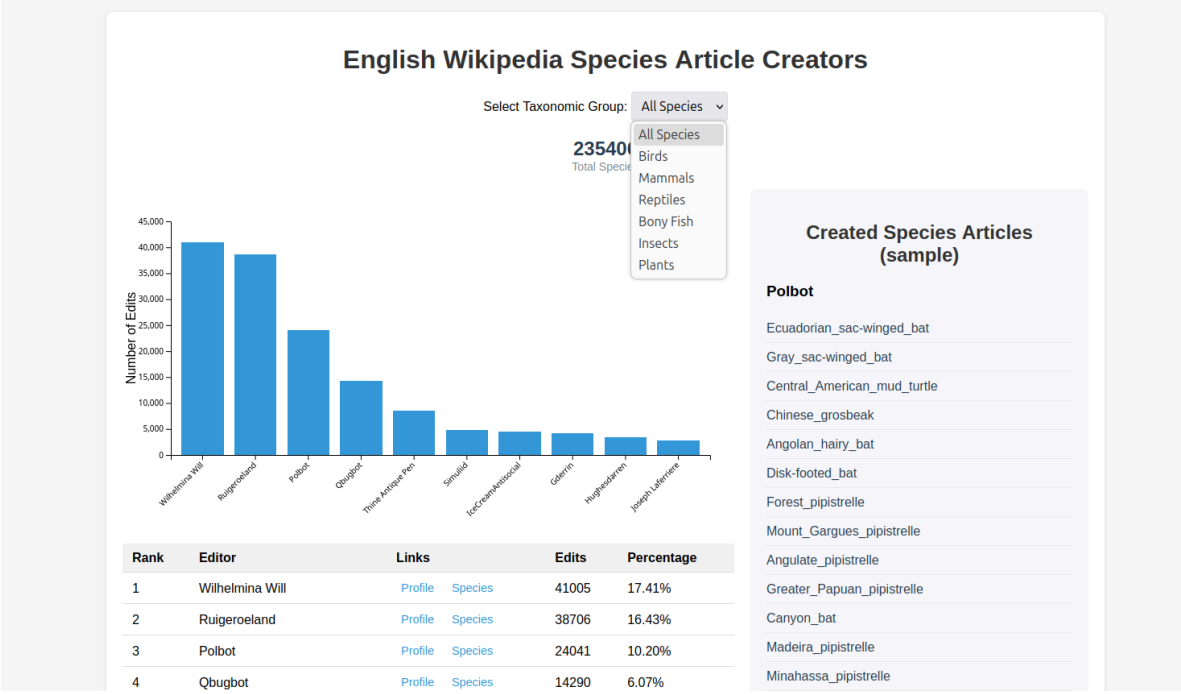
The primary academic product will be a comprehensive research paper detailing our findings. The report will include all data analysis results, figures, and case studies, providing the first in-depth look at Wikimedia's biodiversity coverage and knowledge gaps at scale. This paper will be tailored in the format for **submission to the *Biodiversity Data Journal*, a peer-reviewed, open-access journal prized by the biodiversity informatics community**.

In addition, an executive summary will be prepared for a general Wikimedia audience, to be posted on Diff or a similar community-oriented space. This summary will help volunteers and affiliate members to quickly grasp the main insights of the project.

Open Datasets and Visualizations

We will publish the datasets generated – for example, the table of authors of species articles on different languages – under an open license (CC0 where possible) on Zenodo. An interactive, web-based dashboard will be created to allow stakeholders to explore the data, as exemplified on **Figure 3**.

These dashboards will provide access for the community to data-driven lists pointing to critical gaps of species for the Wikimedia language community. These might be particular endemic species, critically endangered or of particular cultural significance. The exact content of the lists will depend on the research results and from the interaction with the Latin America community, as well as on the availability of structured information on taxonomic databases.



Rank	Editor	Links	Edits	Percentage
1	Wilhelmina Will	Profile Species	41005	17.41%
2	Ruigeroeland	Profile Species	38706	16.43%
3	Polbot	Profile Species	24041	10.20%
4	Qbugbot	Profile Species	14290	6.07%

Figure 3: Interactive tool to explore creators of species articles on English Wikipedia. Based on a Quarry query by User:BilledMammal, this dashboard exemplifies the kinds of interactive datasets this research project aims to make available for the community. The prototype illustrates how one may use the quantitative datasets to gather insights on contributions showing insights, for example, on bot contributions for biodiversity content on English WIkipedia. An interactive version of theprototype is available at http://tiago.bio.br/biodiversity_research_fund_2025/authors.

Community Engagement Outputs

As part of the project, we will also produce periodic updates on project progress on Meta (and shared via different on-wiki and off-wiki platforms) to keep interest and invite feedback on the ongoing project. These will be modeled after the biweekly updates from the Biodiversity Heritage Library Wikimedian-in-Residentship [11] and serve both for ongoing engagement with the interested communities and as a persistent log of the activities under this grant.

Conference and events

The expected outputs include presentations on several Wikimedia events and academic conferences. We will submit a presentation or poster to **Wikimania 2026** to present either virtually or in-person to the global Wikimedia community. Furthermore, the proponent will attend **WikiCon Brasil 2025** (July 2025), and use this opportunity for presenting and gathering feedback from the Brazilian Wikimedia community. Finally, the proponent will submit a paper for the **Wiki Workshop 2026**, ensuring a tighter connection with the Wikimedia Research community.

A core output of this research project is hosting a **Wikimedia and Biodiversity session at Living Data 2025**, in Bogotá, Colombia (October

11 https://meta.wikimedia.org/wiki/Biodiversity_Heritage_Library/Our_outcomes/WiR/Status_updates

2025)[12], a joint conference of four major biodiversity informatics networks (TDWG, GBIF, OBIS and GEO BON). The session, titled "Wikimedia and Biodiversity Data: A Mutualistic Relationship in the Open Knowledge Ecosystem", is **organized by the proponent with Anabela Plos, from GBIF Argentina, and has already been approved.** It will be an outstanding opportunity for fostering the intersection of Wikimedia Research and Biodiversity Informatics, and interact with key collaborators of this project.

In summary, by the end of the project we will deliver an academic paper outlining the results, interactive data-driven visuals and datasets and community-friendly lists and a set of presentations in Wikimedia- and Biodiversity-events. These outputs are designed to be self-contained and widely accessible, ensuring they can immediately serve the dual goals of advancing scholarly understanding and supporting the Wikimedia community.

Risks

Data Completeness and Accuracy

One inherent risk is relying on Wikidata for analysis, which themselves might contain errors or omissions. Though we are confident on the overall quality of Wikidata, if its taxonomy happens to be missing certain species, has duplicate entries, or does not properly hold the links to GBIF or the target Wikipedias, our coverage metrics could be skewed. To mitigate, we will cross-verify the Wikidata taxonomy with the Catalogue of Life, ensuring a validated reference species taxonomy is used. Furthermore, data-driven results will not be taken for granted, but cross-checked and manually verified for consistency at every step.

Notability and Cultural Differences

One risk is assuming that all species "should" have a Wikipedia article. Additionally, English, Portuguese and Spanish might not be representative of the pattern of gaps across the board. To mitigate, we are focusing cautiously on Wikipedia editions for which the proponent has contributed first-hand and is in close contact with multiple key editors. While methods will be designed for generalizability, communication with a larger set of stakeholders will be key if results are to be interpreted in a wider fashion.

Overemphasis on Quantitative Metrics

There is a risk of focusing on what can be measured (number of articles, article creations, etc.) and potentially undervaluing qualitative aspects of knowledge (like how well an article captures cultural significance of a species). Our study might not directly capture those dimensions if we only look at length or references. To mitigate, we explicitly acknowledge these limitations and emphasize the specific research questions at hand, balancing ambition and caution. In our discussions, we will stress that bridging knowledge gaps isn't just about counting articles but also about improving content depth and inclusivity of knowledge.

Scope Creep and Data Overload

With such a broad subject (millions of species, uncountable ways to measure *gaps*), there's a risk of being swamped with data or trying to tackle too much, especially with unexpected events that might delay parts of the project. To mitigate, our analysis will keep focused on the key languages (Portuguese, English and Spanish), continuously evaluating progress and prioritizing tasks to ensure core analyses are as delivered as planned.

[12] https://livingdata2025.com/sessions.html

Community impact plan

A key goal of this research project is to **support the Wikimedia community to actually fill the identified biodiversity knowledge gaps.**

For example, discussions stemming from the BHL-Wiki Working Group are underway to form a **Wikimedians for Biodiversity** User Group.[13] Our project will act as a catalyst for this group, feeding it with = data to support its work.

The research project is also aligned with several Wikimedia campaigns. Besides the previously mentioned WikiForHumanRights and #1pic1bio events, Wikimedia Brasil organizes annual campaigns focused on Brazilian states, always including lists of articles for fauna and flora, for example listing 460 different target species for the campaign this year, focusing on the Brazilian state of Maranhão. (Wikimedia Brasil, 2025). Different Spanish Wikipedia affiliates have gotten together recently for a campaign on Sharks and Rays on Wikipedia, increasing the knowledge on some neglected species. (Wikimedia España, 2024). Importantly, these are communities in which the proponent has been active for a long time, paving the way for solid collaborations.

A key piece of this plan is to maintain the communication channels — which are already open and functional. This will allow community members to give feedback on what they need, reporting errors, asking questions or even bringing up ideas for new analyses. This responsiveness will build trust and make our outputs more directly useful. The network includes not only the Wikimedia community spaces, mailing lists, Slack channels and Telegram groups, ensuring a wide net is cast. For example, sharing preliminary results from this project on the WikiProject Biodiversity Telegram group, for example, has sparked spontaneous suggestions and feedback from dedicated Wikimedians.

In summary, our Community Impact Plan is about turning research into action. By anchoring our work in real campaigns, and supporting ongoing work with information and tools, we aim to translate the analytical findings into measurable content improvements and sustained community engagement.

Evaluation

Success will be evaluated by the completion of the academic outputs and dissemination activities detailed previously. Key performance indicators include:

- Publication of the species coverage datasets with high-quality metadata.
- Publication of biweekly progress reports on Meta, with links and preliminary results.
- Timely submission of the research paper. A successful outcome would be acceptance of a paper at a journal, preceded by the publication on a preprint server with accompanying reviews (e.g. via PREreview[14]).
- Presentations at events and feedback from both the research community and the Wikimedia community at large.

[13] https://outreach.wikimedia.org/wiki/GLAM/Newsletter/March_2025/Contents/Biodiversity_Heritage_Library_report

[14] https://prereview.org/

Budget

Description	Value (USD)
Research Stipend for Project Coordinator – 0.8 FTE (9 months x$2,044.00)	$18,396
Open access publishing cost at *Biodiversity Data Journal*	$904
Travel and registration for Living Data 2025 (Bogota, Colombia)	$1,000
Total	**$20,300**

Biodiversity Knowledge Gaps - Budget

References

Bánki, O., Roskov, Y., Döring, M., Ower, G., Robles, D. R. H., Corredor, C. A. P., Jeppesen, T. S., Örn, A., Pape, T., Hobern, D., & others. (2025). *Catalogue of Life*. https://doi.org/10.48580/dgplc

Cardinale, B. J., Duffy, J. E., Gonzalez, A., Hooper, D. U., Perrings, C., Venail, P., Narwani, A., Mace, G. M., Tilman, D., Wardle, D. A., Kinzig, A. P., Daily, G. C., Loreau, M., Grace, J. B., Larigauderie, A., Srivastava, D. S., & Naeem, S. (2012). Biodiversity loss and its impact on humanity. *Nature*, *486*(7401), 59–67. https://doi.org/10.1038/nature11148

Correia, R. A., Ladle, R., Jarić, I., Malhado, A. C. M., Mittermeier, J. C., Roll, U., Soriano-Redondo, A., Veríssimo, D., Fink, C., Hausmann, A., Guedes-Santos, J., Vardi, R., & Di Minin, E. (2021). Digital data sources and methods for conservation culturomics. *Conservation Biology*, *35*(2), 398–411. https://doi.org/10.1111/cobi.13706

Das, P., Johnson, I., Saez-Trumper, D., & Aragón, P. (2024a). *Language-Agnostic Modeling of Wikipedia Articles for Content Quality Assessment across Languages* (No. arXiv:2404.09764). arXiv. https://doi.org/10.48550/arXiv.2404.09764

Das, P., Johnson, I., Saez-Trumper, D., & Aragón, P. (2024b). *Language-Agnostic Modeling of Wikipedia Articles for Content Quality Assessment across Languages* (No. arXiv:2404.09764). arXiv. https://doi.org/10.48550/arXiv.2404.09764

Dearborn, J. (2023). Unifying Biodiversity Knowledge to Support Life on a Sustainable Planet. *The Biodiversity Heritage Library*. https://doi.org/10.21428/bcf8962c.699434fb

Dearborn, J. (2024, November 20). *Advancing BHL's Data for a Sustainable Future: Meet Tiago, Our New Wikimedian-in-Residence – Biodiversity Heritage Library*.

Mering, S. von, Cubey, R., Endresen, D., Hendriksen, A., Leachman, S., Mietchen, D., & Santos, J. (2024). Advancing Community Curation of Research Expeditions: A Collaborative Journey with Wikidata and Biodiversity Information Standards. *Biodiversity Information Science and Standards*, *8*, e138921. https://doi.org/10.3897/biss.8.138921

Miquel-Ribé, M., & Laniado, D. (2018). Wikipedia Culture Gap: Quantifying Content Imbalances Across 40 Language Editions. *Frontiers in Physics*, *6*. https://doi.org/10.3389/fphy.2018.00054

Mittermeier, J. C., Correia, R., Grenyer, R., Toivonen, T., & Roll, U. (2021). Using Wikipedia to measure public interest in biodiversity and conservation. *Conservation Biology: The Journal of the Society for Conservation Biology*, *35*(2), 412–423. https://doi.org/10.1111/cobi.13702

Mittermeier, J. C., Roll, U., Matthews, T. J., & Grenyer, R. (2019). A season for all things: Phenological imprints in Wikipedia usage and their relevance to conservation. *PLOS Biology*, *17*(3),

e3000146. https://doi.org/10.1371/journal.pbio.3000146

Page, R. D. M. (2022). Wikidata and the bibliography of life. *PeerJ*, *10*, e13712. https://doi.org/10.7717/peerj.13712

Paul, Q. G. & D. (2020, February 21). *Does Biodiversity Informatics 💘 Wikidata?* https://data-blog.gbif.org/post/wikidata/

Redi, M., Gerlach, M., Johnson, I., Morgan, J., & Zia, L. (2021). *A Taxonomy of Knowledge Gaps for Wikimedia Projects (Second Draft)* (No. arXiv:2008.12314). arXiv. https://doi.org/10.48550/arXiv.2008.12314

Roll, U., Mittermeier, J. C., Diaz, G. I., Novosolov, M., Feldman, A., Itescu, Y., Meiri, S., & Grenyer, R. (2016). Using Wikipedia page views to explore the cultural importance of global reptiles. *Biological Conservation*, *204*, 42–50. https://doi.org/10.1016/j.biocon.2016.03.037

Scann WDU, & Wikimedistas de Uruguay. (2023, December 14). Meet the new Wikimedia working group focused on climate change and environmental topics on Latin America! *Diff*. https://diff.wikimedia.org/2023/12/14/meet-the-new-wikmedia-working-group-focused-on-climate-change-and-environmental-topics-on-latin-america/

Taraborelli, D. (2019, February 27). *Open curation of knowledge on biodiversity: An introduction to Wikidata*. https://figshare.com/articles/journal_contribution/Open_curation_of_knowledge_on_biodiversity_An_introduction_to_Wikidata/7780994?file=14486276

The Global Biodiversity Information Facility. (2025). *What is GBIF?* https://www.gbif.org/what-is-gbif

WikiForHumanRights—Meta. (n.d.). Retrieved April 15, 2025, from https://meta.wikimedia.org/wiki/WikiForHumanRights

Wikimedia Brasil. (2025). Wikiconcurso Wiki Loves Maranhão. In *Wikipédia, a enciclopédia livre*. https://pt.wikipedia.org/w/index.php?title=Wikip%C3%A9dia:Wikiconcurso_Wiki_Loves_Maranh%C3%A3o&oldid=69929824

Wikimedia España. (2024, July 18). Más de 100 artículos tiburones y rayas en Wikipedia. *WIKIMEDIA ESPAÑA*. https://wikimedia.es/mas-de-100-articulos-sobre-tiburones-y-rayas-enriquecen-la-wikipedia-en-espanol/

Wikimedia Foundation. (n.d.). *Achieve knowledge equity*. Wikimedia Foundation. Retrieved April 15, 2025, from https://wikimediafoundation.org/our-work/open-the-knowledge/otk-help-us-achieve-knowledge-equity/