

The Knowledge Gap Index and the LGBTQ+ Communities

Miriam Redi and Leila Zia, Research Team



WIKIMEDIA
FOUNDATION

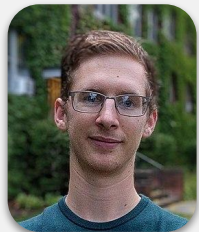
Research team



Pablo Aragón



Martin Gerlach



Isaac Johnson



Fabian Kaelin



Miriam Redi



Diego
Sáez-Trumper



Leila Zia



4 contractors, 1 Intern, 1 Research Fellow, [17 Formal Collaborators](#)

Mission

Develop models and insights utilizing scientific methods, and grow the Wikimedia research communities.

We do this in order to:

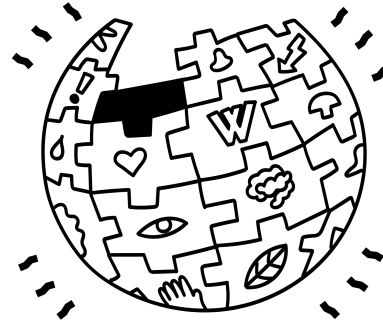
- Support technology and policy needs of the Wikimedia projects,
- advance the understanding of the Wikimedia projects.

Audiences

- Wikimedia Foundation
- **Wikimedia affiliates and user groups**
- Wikimedia volunteer developers
- Wikimedia researchers

Our Programs

**Address
Knowledge
Gaps**



**Improve
Knowledge
Integrity**



Grow the Research Community

Our Programs

**Address
Knowledge
Gaps**



**Improve
Knowledge
Integrity**



Grow the Research Community

The strategic direction: Knowledge Equity

[from Movement Strategy]

Knowledge equity: As a social movement, we will focus our efforts on the **knowledge and communities that have been left out by structures of power and privilege**. We will welcome people from every background to build strong and diverse communities. We will **break down the social, political, and technical barriers** preventing people from accessing and contributing to free knowledge.

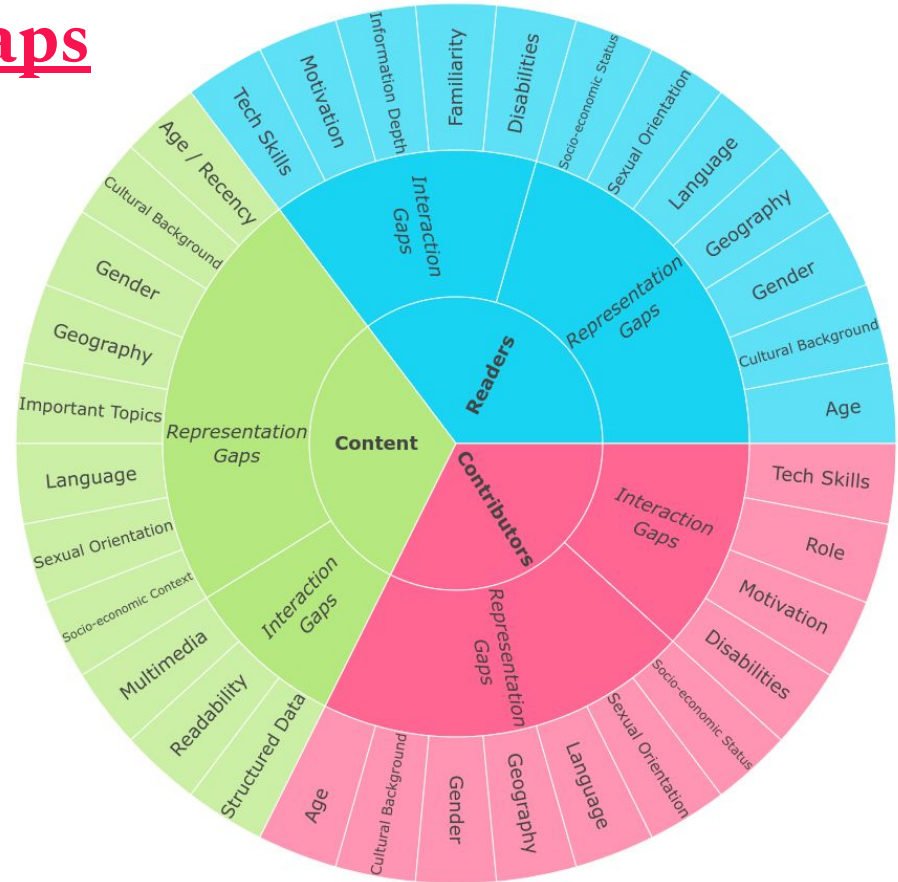
Our Goal:

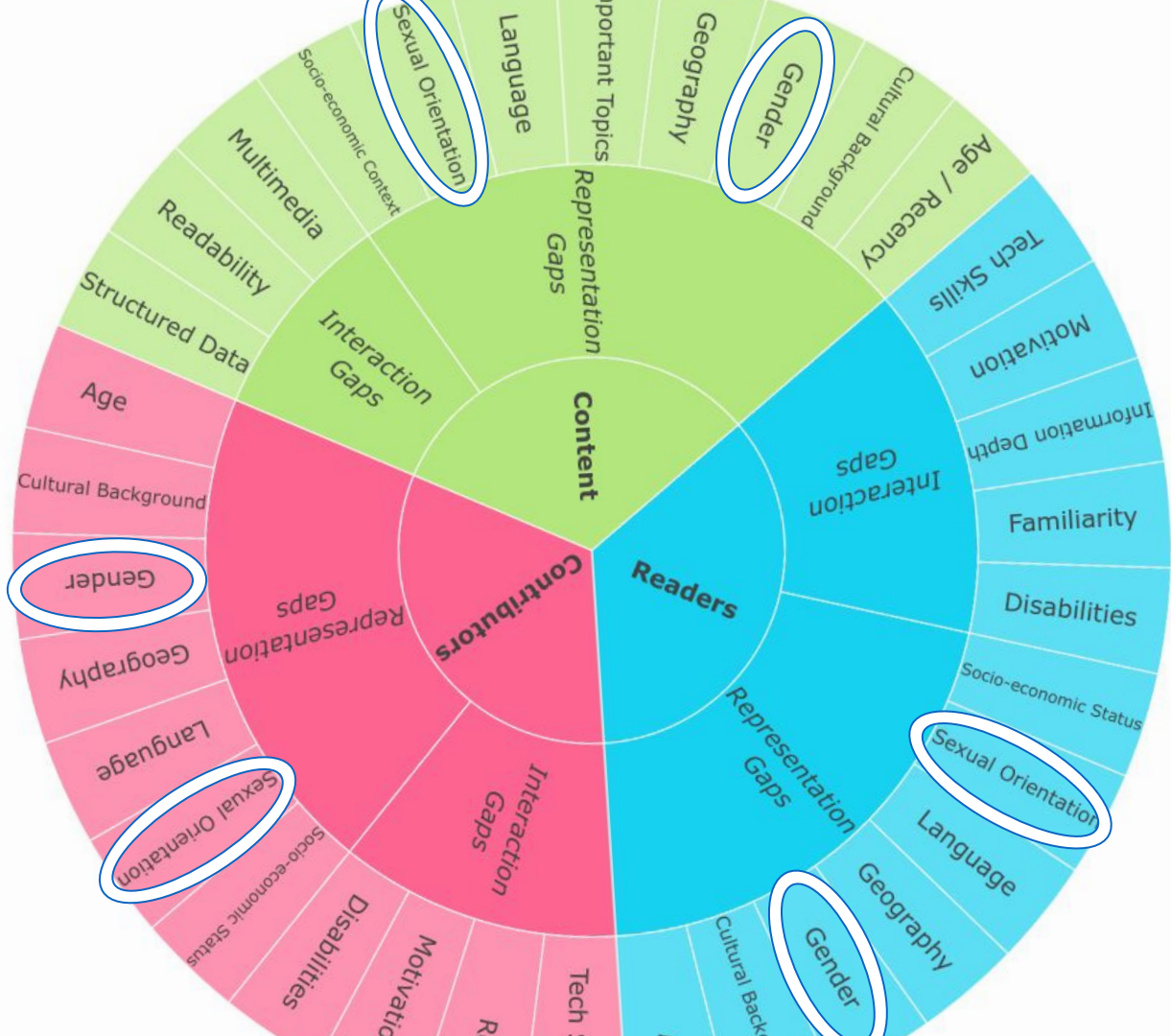
Identify, measure and bridge the individual components **(knowledge gaps)** that prevent us from reaching knowledge equity.

What are knowledge gaps?

Taxonomy of Knowledge gaps

- 200+ references
- Finding and classifying evidence of inequalities in Wikimedia projects.
- Grouping gaps into Content, Readers, and Contributors

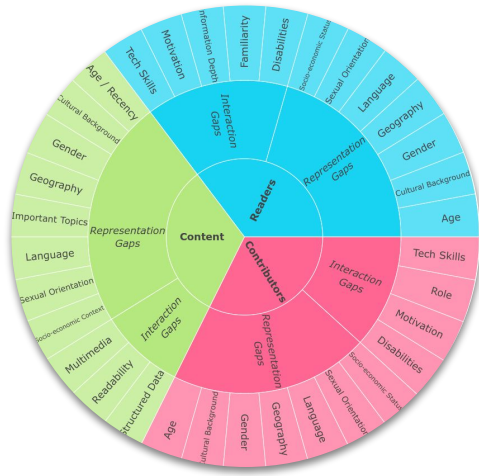




Address Knowledge Gaps

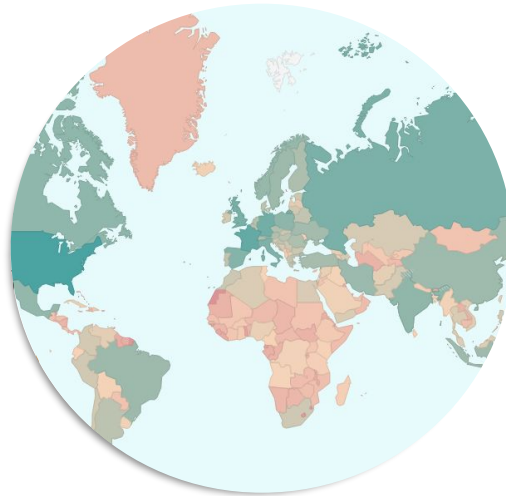
Identify gaps

<https://w.wiki/4DWU>



Measure gaps

<https://w.wiki/4Exh>



Bridge gaps

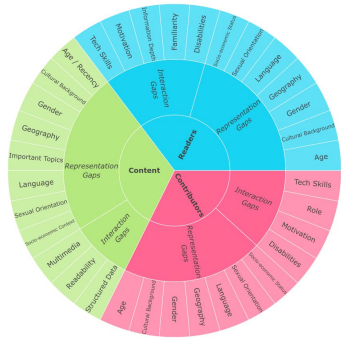
<https://w.wiki/4DWe>



Step 2: Quantify knowledge gaps

Define metrics based on large-scale data analysis

Stakeholder consultations



Data Collection and
(automatic)
categorization



Metrics Generation

Existing tools...

Global Gender Gap

Distribution of content of humans in all Wikimedia Projects

755.5k

Female Content

3.4M

Male Content

2.4k

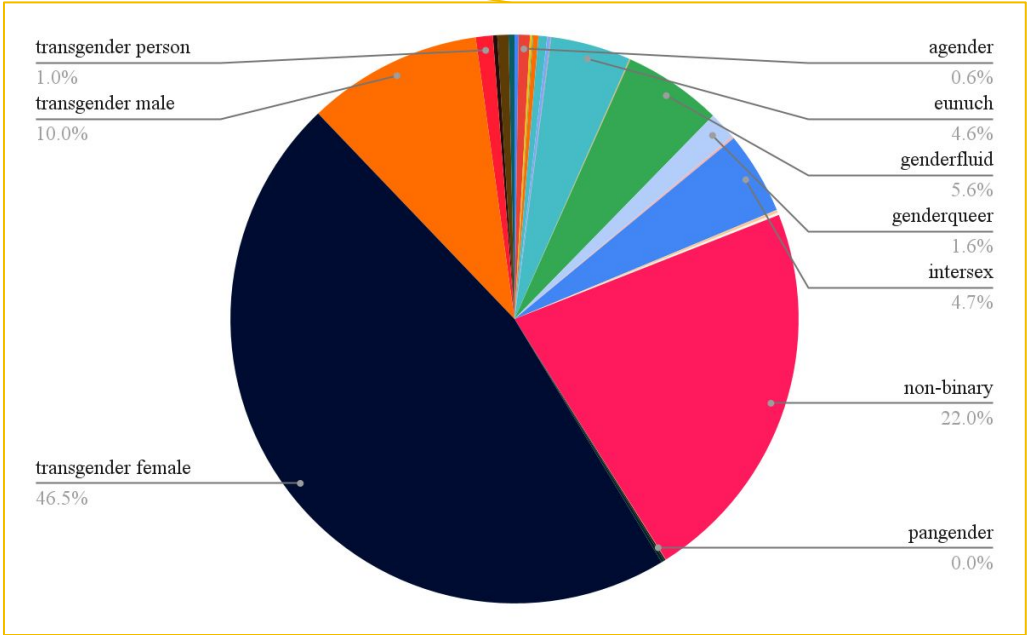
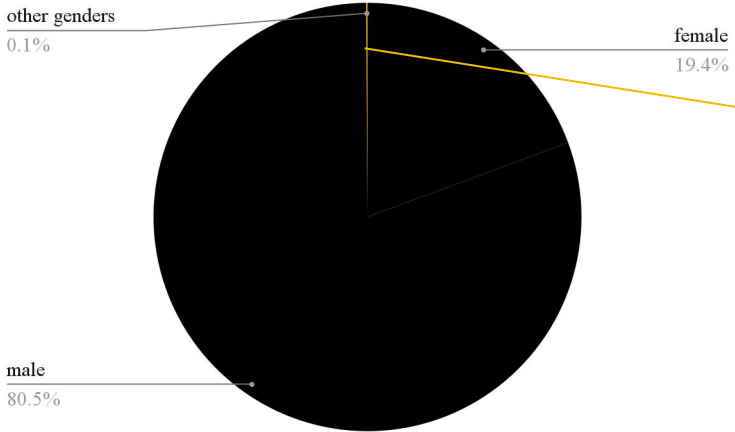
Σ Other Genders Content (sum)



All time, as of May 2022

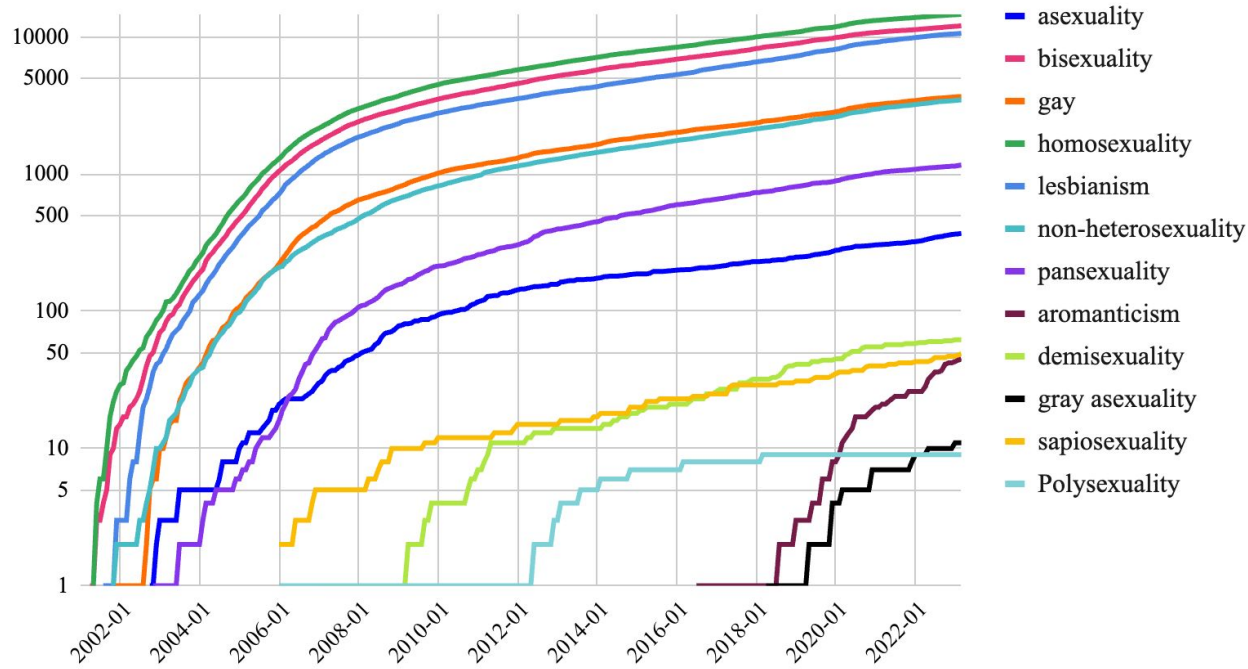
Data from humaniki: <https://whgi.wmflabs.org/>

Our data gets all gender labels from Wikidata



We look at **sexual orientation** (but not many “heterosexual” annotations from Wikidata)

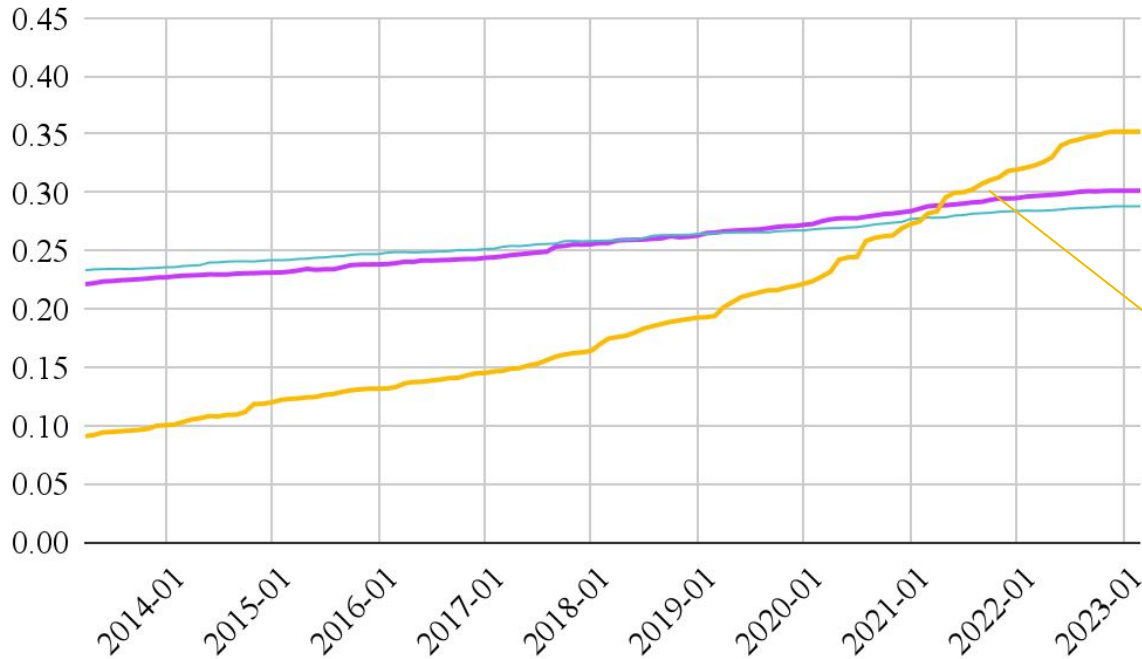
Articles created, non heterosexual sexualities



Or, the evolution of **article quality** over time:

Predicted article quality, 3 genders, over time

— female — male — other genders



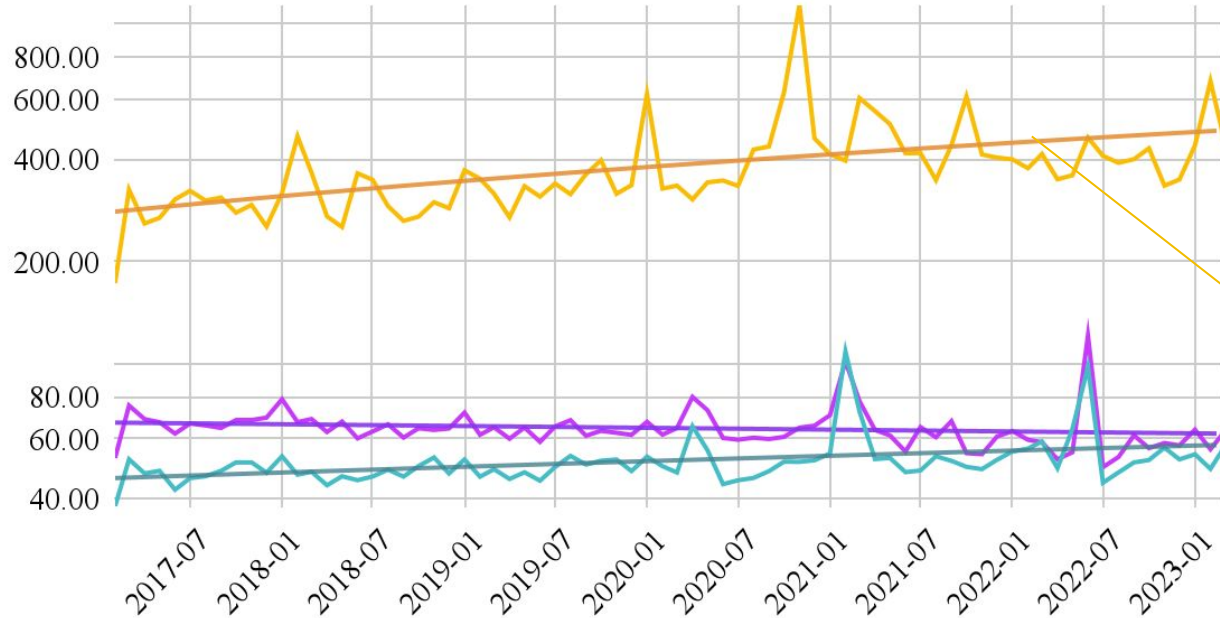
The quality of articles about non-binary folks is on average higher than other genders



Or, average number of pageviews over time:

Average pageviews per page, 3 genders

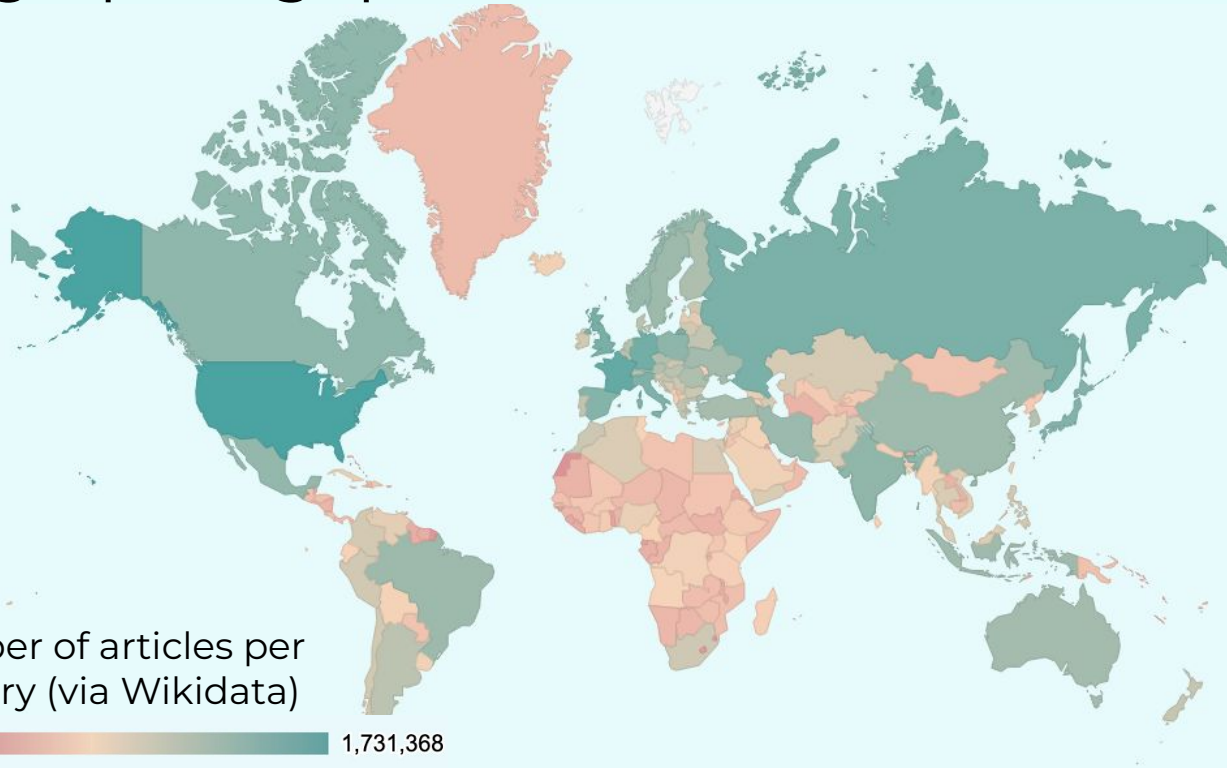
— female — male — other genders



On average, articles about non-binary folks get way more pageviews than other genders!



More data beyond gender! For example, distribution of articles across countries to describe the geographic gap.



Knowledge Gap Metrics **Datasets now Available!** API and tools are coming!

Article **Quantity**

Article **Quality**

Monthly **Pageviews**

Monthly **Revisions**



Gender of the article subject

Geography countries related to the article

Sexual Orientation of the article subject

Time the temporal setting of the article

Multimedia whether the article has image

More info and data: [Research:Knowledge_Gaps_Index/Measurement/Content](#)

Plots in this presentation: [Knowledge Gaps Plots - Queering Wikipedia](#)

Notebook: [https://public-paws.wmcloud.org/User:Miriam_\(WMF\)/queering-wp.ipynb](https://public-paws.wmcloud.org/User:Miriam_(WMF)/queering-wp.ipynb)

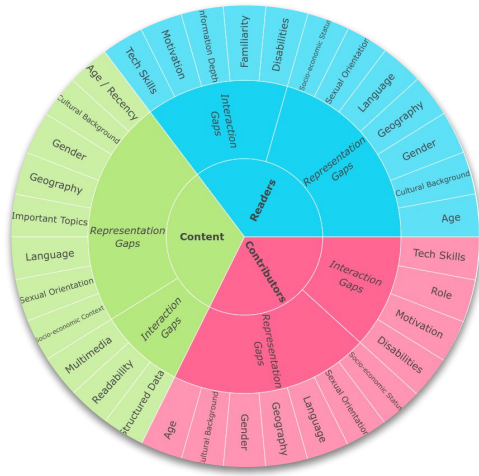
The **knowledge gap index** will be a tool to explore all these measurements.

- Let's start compiling some of the questions that you want to answer using these metrics and measurement.
- What are the functionalities that you would like to see?
- What are other ways in which we can report the data so that is useful to you?
- Are there issues with Wikidata labels that you think we should be aware of?
 - Missing annotation for the “dominant” category
 - Sex and gender are conflated in some cases

Address Knowledge Gaps

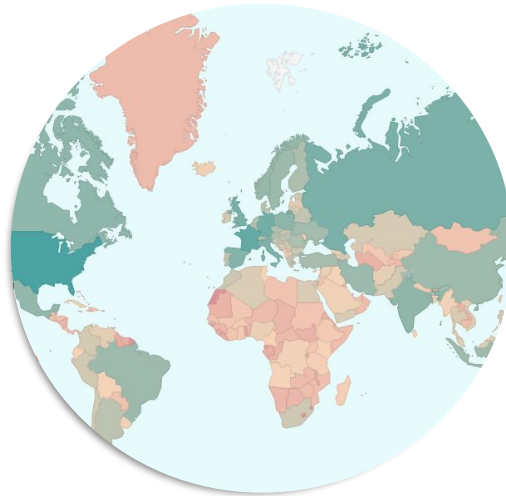
Identify gaps

<https://w.wiki/4DWU>



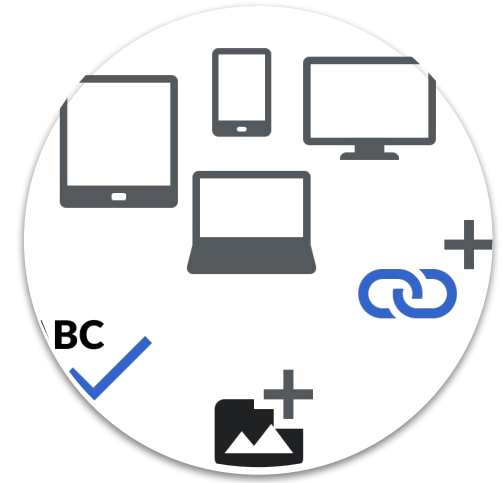
Measure gaps

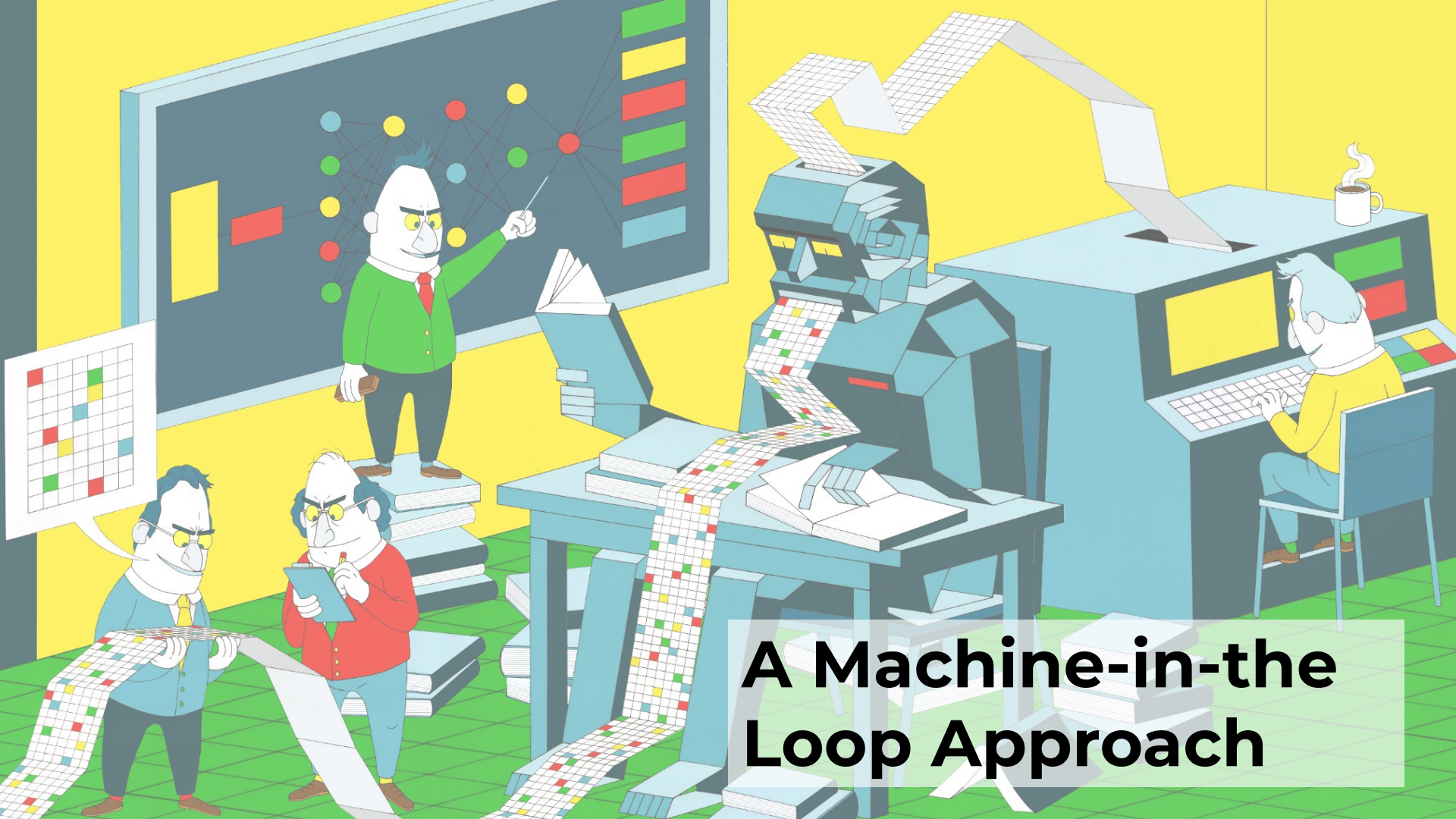
<https://w.wiki/4Exh>



Bridge gaps

<https://w.wiki/4DWe>

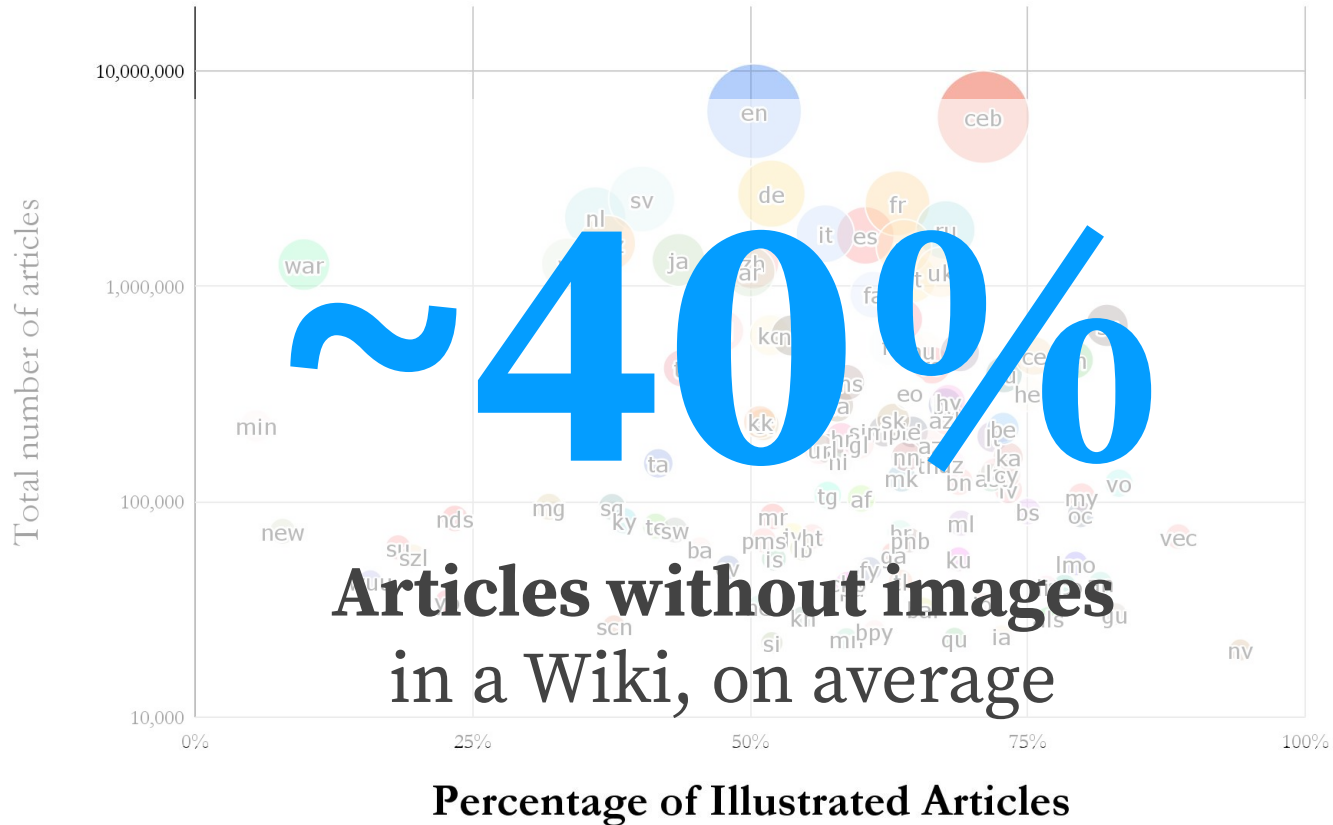




A Machine-in-the Loop Approach

Problem

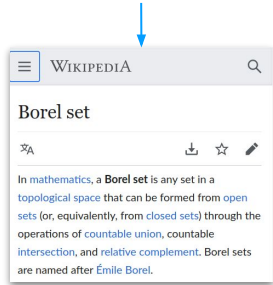
Many low-quality articles are missing images!



Solution

Automated method to match images & articles

Discovering unillustrated articles as articles without images or with *icons* only.

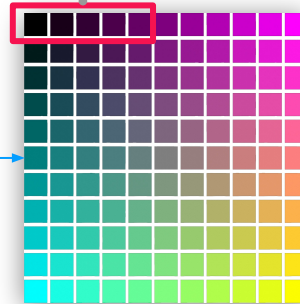


Unillustrated Article

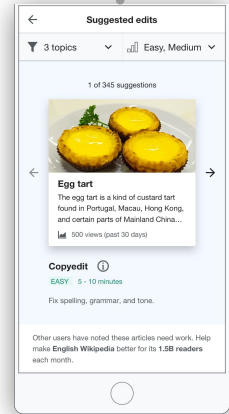


Discovering relevant images from Wikidata item or articles in other languages

Filtering



Ranking images according to Relevance and Quality



Solution

And a tool for editors to easily add images

The image displays four sequential mockups of a Wikipedia editor's interface, illustrating the process of adding an image to an article. Each mockup is titled "B-Onboarding-*i*" and shows a "Church order" article.

- Mockup I:** Shows the "Suggestions" panel. A message box says "Look at the image suggested" and provides information about the image's source (Wikimedia Commons). Below the message is a small image of Sofia Church in Stockholm and a "Skip" button.
- Mockup II:** Shows the "Suggestions" panel with a question: "Is it suitable for this article?". It includes a review prompt: "Review the article for which this image is being suggested. Does it seem like the image is suitable and helpful for readers to understand this topic?". Below this, it states "1 suggested image found" and shows the image with its caption: "Sofia kyrka.jpg Sofia Church in Stockholm. The description will wrap after a few more lines until it reaches the image heig...".
- Mockup III:** Shows the "Suggestions" panel with a question: "Use the image details to help with your decision". It includes a review prompt: "To help evaluate the image suitability, review the image by looking at its file name, description, the reason it was suggested, and zooming in to see more details.". Below this, it states "Suggestion reason: Image was found on the same article in Indonesian Wikipedia." and shows the image with its caption.
- Mockup IV:** Shows the "Add a caption" panel. It includes a review prompt: "Write a short caption here to help explain why the image is relevant to the article." and a message: "Write a caption that gives context to the image". It includes guidelines: "Describe the image so it's clear how it relates to the article. A few guidelines: • Avoid making value judgements like 'good', 'ugly', or 'beautiful'. • Review the description and filename for information that is relevant to include. • Write in the language of the Wikipedia you are in." Below the guidelines is a "Skip" button.

By RHo (WMF) - Mockups from WMF designer, CC BY-SA 4.0,
<https://commons.wikimedia.org/w/index.php?curid=108208100>

You address **knowledge gaps**

- What are the major issues you encounter when addressing knowledge gaps in your spaces?
- What tools do you need other than measurements to address knowledge gaps in your spaces?
- How can automated tools support your work?
- What is the visual knowledge gap in your space?

Next steps ...

We need more and more research to Address Knowledge Gaps!

