Knowledge Gaps Index: Measuring Content Gaps

Marc Miquel

mmiquel-ctr@wikimedia.org

Wikimedia Research

**Martin Gerlach**

mgerlach@wikimedia.org

Wikimedia Research

The strategic direction: Knowledge Equity
[from Movement Strategy 2030]

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.
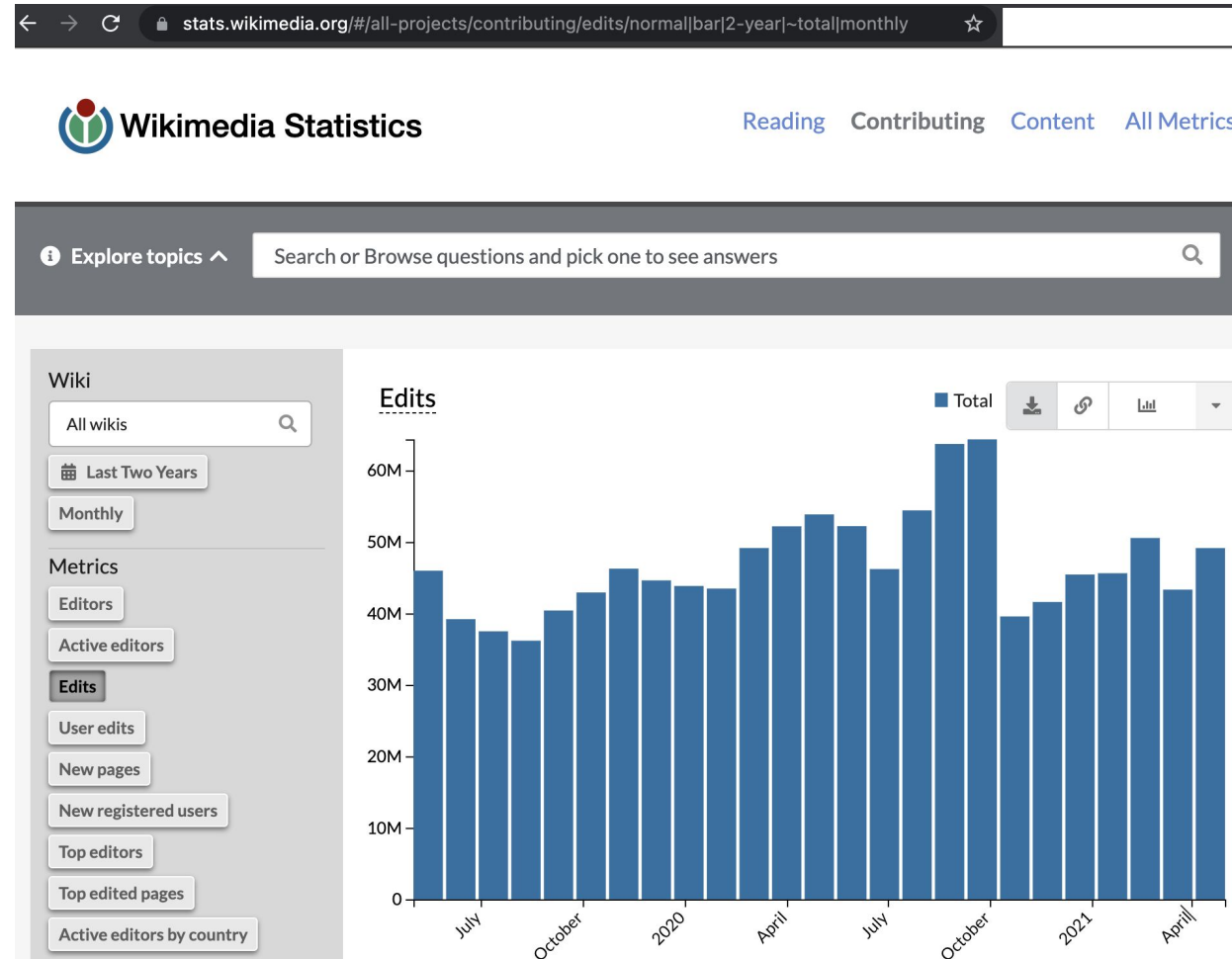
How far are we from reaching knowledge equity?
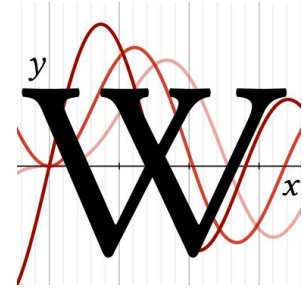
Identify and measure the individual components based on which we can track our progress towards this goal.
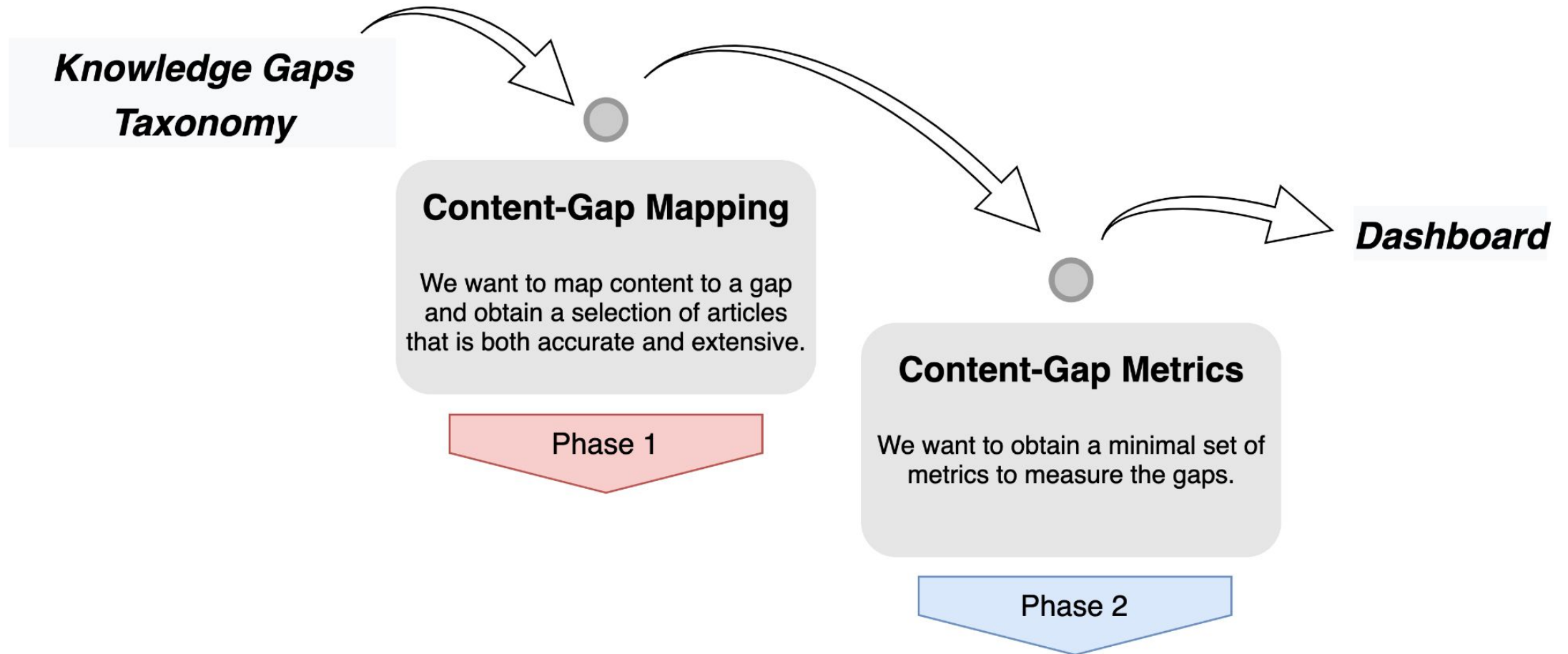
Measuring Content Gaps.

Based on community interest, while creating the Knowledge Gaps Taxonomy, we choose five content gaps to study:

- **gender, sexual orientation, geography, cultural context, and time.**

We want to measure content gaps and eventually have monitoring tools similar to stats.wikimedia.org.

We need to map content to a gap and design metrics that explain them.

Knowledge Gaps Taxonomy

Content-Gap Mapping

We want to map content to a gap and obtain a selection of articles that is both accurate and extensive.

Phase 1

Content-Gap Metrics

We want to obtain a minimal set of metrics to measure the gaps.

Phase 2

Dashboard

Open research: project presentation and discussion

- **Review of communities and user groups documentation** on the five different gaps (gender, sexual orientation, geography, etc.) to understand (1) the type of content that relates to the gap, and (2) the practices, needs and goals that appear in the process of bridging these gaps.

- **Interviews with user group and community leaders** (e.g, Gender and LGBT+) to uncover the most common information needs.

Thank you!

Created by basith ibrahim
from Noun Project

We are creating metrics together to support the whole movement.

Thank you for the interviews

Viquidones, Wikimujeres, Art+Feminism, Wikiesfera, Wikidonne, Wikimedia LGBT+, and other individual Wikimedians.

Chosen metrics

- The final chosen three metrics, based on community interest and scientific maturity are:

  1. **Selection** (Number of articles for each category of the gap). e.g., number of articles for each country for the geography gap.

  2. **Extent-Score** (indicator similar to wikirank to explain the degree of completion/quality of articles based on length, # sections, # images). The extent-score will explain "how good" the articles in each category are.

  3. **Visibility** (Percentage of articles for each category in spaces like the "Main Page" or the group of "Featured articles").

Future metrics

Some very valuable aspects of the content gaps were mentioned in the interviews and have not been addressed.

They would require new original research beyond the scope of this project and will be addressed in the future.

- Deletion process
- Engagement (power dynamics behind the article edit history)
- Different types of references

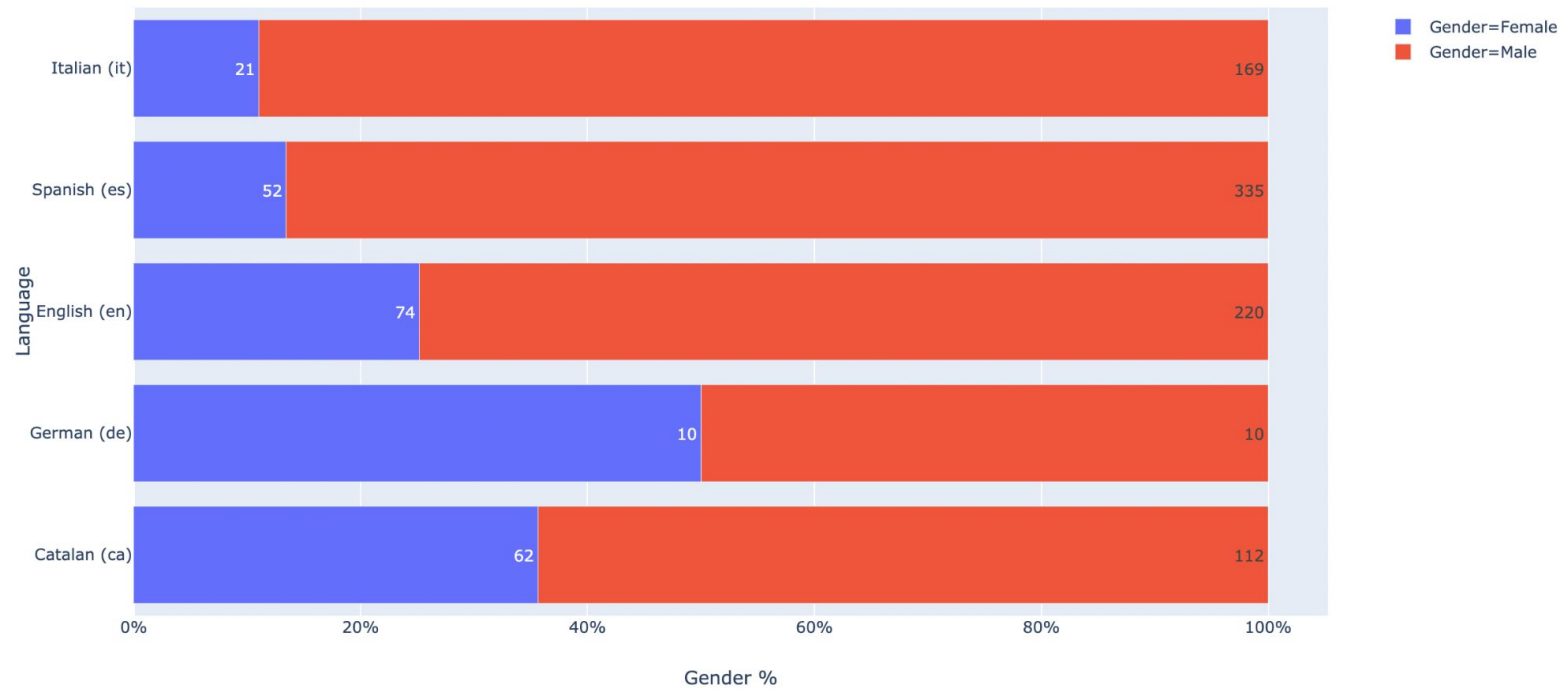# We are implementing the three metrics.

Gender Gap Visibility

Geography Gap Selection - Continents

Language	Continent	Number of articles (%)
ca	Europe	129,193 — 76.20%
	Americas	
	Asia	
	Africa	
	Oceania	24
	Antarctica	0.01%
de	Europe	325,453 — 78.67%
	Americas	
	Asia	
	Africa	
	Oceania	
	Antarctica	
en	Europe	377,964 — 44.51%
	Americas	
	Asia	
	Africa	
	Oceania	33,908 — 3.99%
	Antarctica	
es	Europe	161,168 — 58.17%
	Americas	
	Asia	
	Africa	
	Oceania	87 — 0.03%
	Antarctica	
it	Europe	181,757 — 67.01%
	Americas	
	Asia	
	Africa	
	Oceania	
	Antarctica	
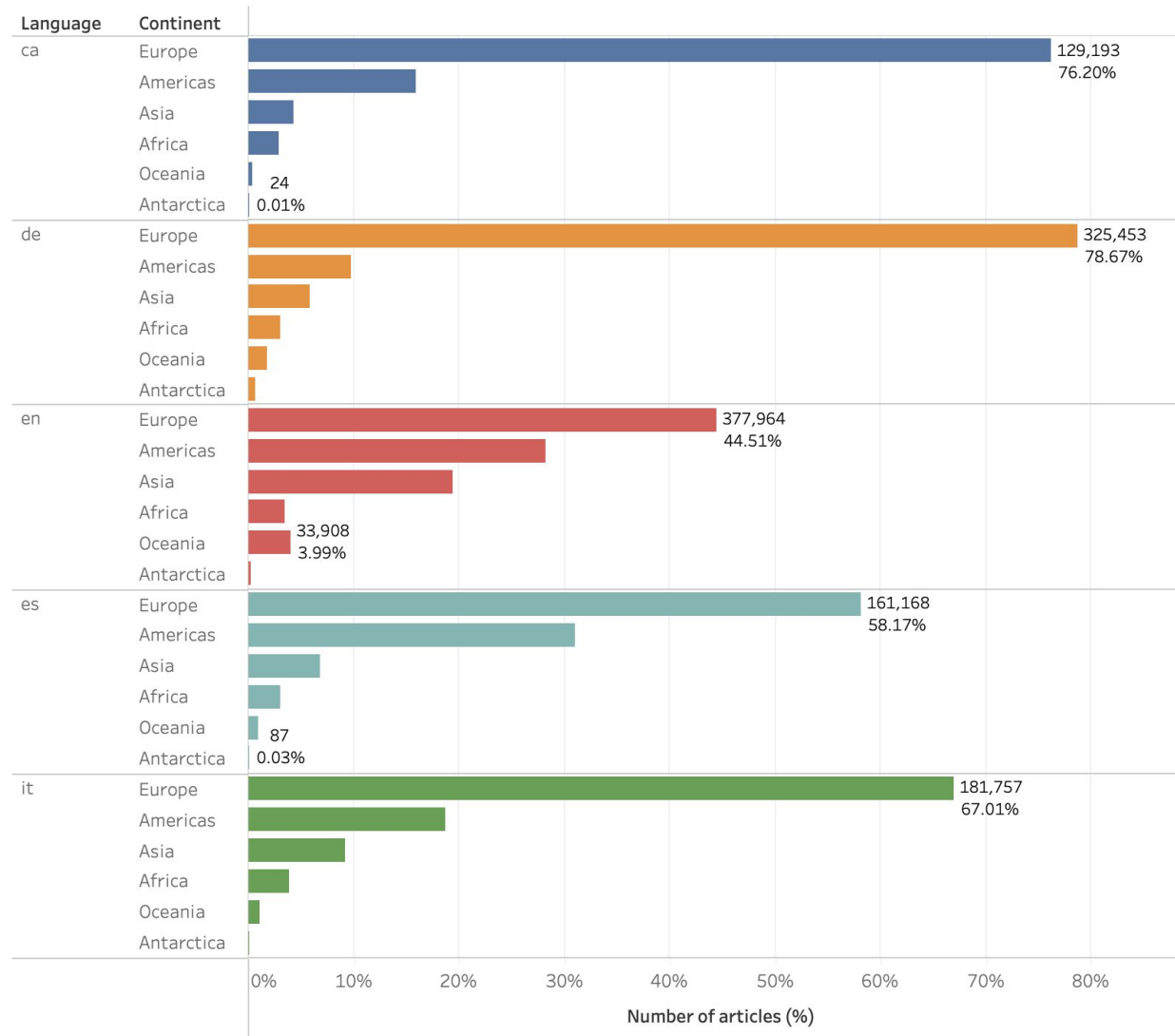
Geography Gap Selection (Continents)

Next steps

- **More documentation** on the research process, methodology and code will be available on Meta in the next weeks.

- Creating a "Stats" dashboard-website for knowledge gaps.
  - If you want to stay involved as participant in the design process, feedback will be very welcome along the design and development process. Please, e-mail me at mmiquel-ctr@wikimedia.org

- You can attend any **Wikimedia Research Office hours** session and ask research related questions that you think we could try to answer.

Thank you very much!

Marc Miquel

mmiquel-ctr@wikimedia.org

Wikimedia Research

Martin Gerlach

mgerlach@wikimedia.org

Wikimedia Research