

Editing and Diversity Funnel:

A Metaphor To Study Wikipedia and Diversity Barriers

Marc Miquel, PhD

{marcmiquel@gmail.com}

Username:marcmiquel

Pompeu Fabra University, Barcelona, Catalonia

Amical Wikimedia (Catalan Wikipedia)

WG Diversity – Strategy 2030



WIKIMANIA
STOCKHOLM
AUGUST 14-18, 2019



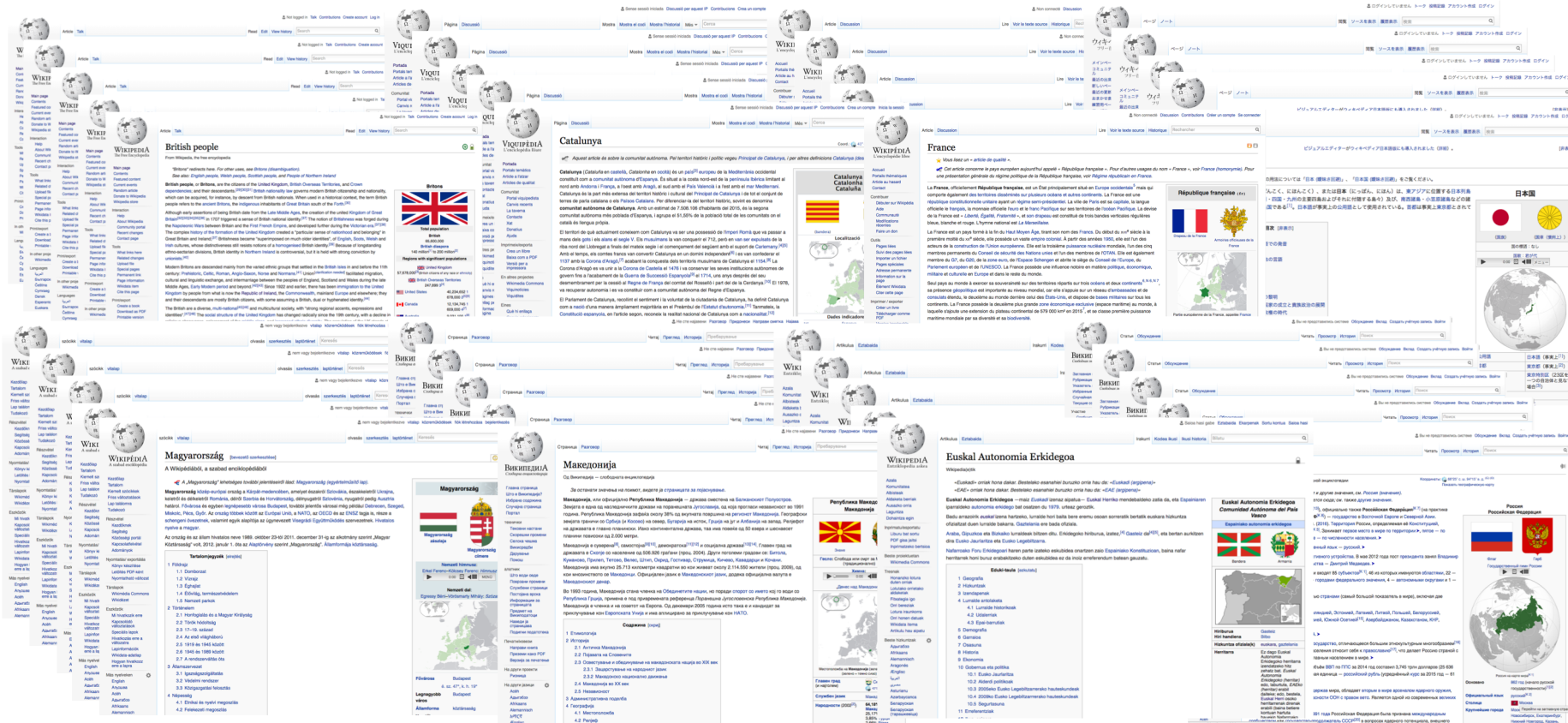
The Problem

Wikipedia project does not reflect enough the world's cultural diversity.



The Solution

Cartography for every Wikipedia



For each Wikipedia language edition, we aim at selecting the **Cultural Context Content (CCC)**, i.e. traditions, language, politics, agriculture, biographies, places, events, etcetera, related to the territories where the language is spoken.

Taking into account the largest Wikipedia language editions, CCC is in average about a quarter of each Wikipedia (Miquel-Ribé and Laniado 2016).



We have a **problem of representation** considering that the CCC extent in non-western languages (African and Asian) is on average much smaller (Miquel-Ribé and Laniado, 2019).

What are the causes to this?

Why is the English CCC 44.23% and the Chinese CCC 14.39%?

It is not just related to population, number of speakers, geographical extension... It has to do with many other variables.

It has to do with the barriers to editing Wikipedia too.

Marketing and Sales Funnels Model

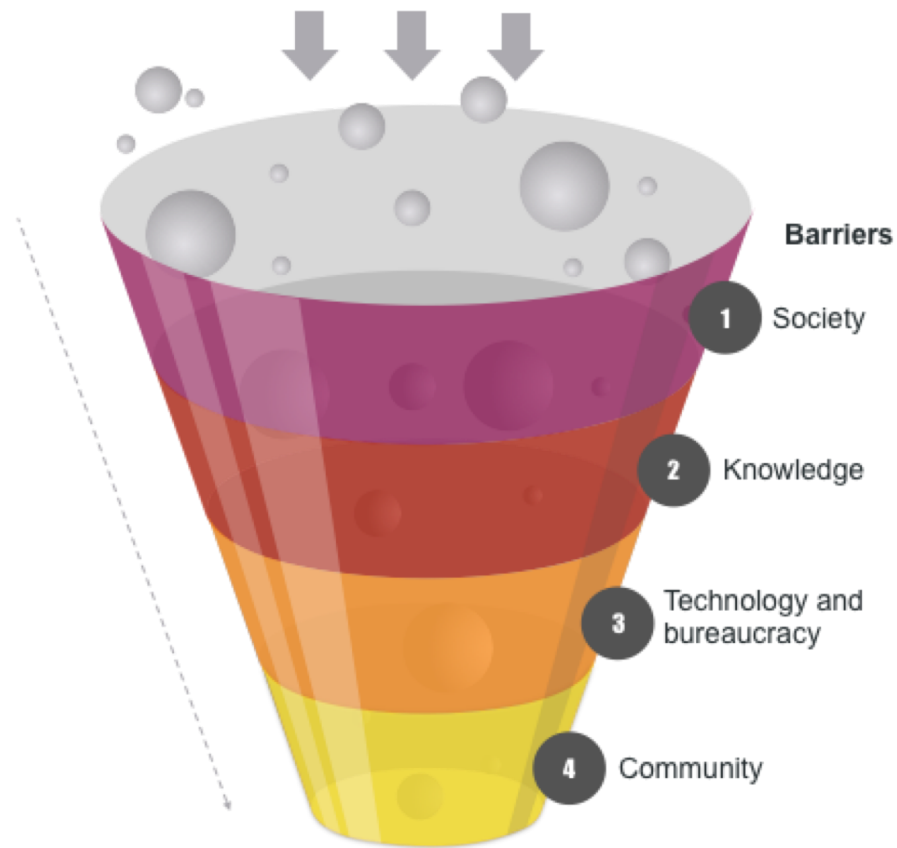
Depending on the interaction, the user can become a customer.



Could it be useful to help us think about Wikipedia and diversity?

Speakers of a language in a context (country or region)

Diversity group: context editors, gender, LGBT+, etc.



Editors representing their **knowledge** in a **Wikipedia language edition**



Barriers

Society

- Lack of access to Internet
- Lack of economical conditions
- Lack of welfare
- Authorities interference on community dynamics
- Legal barriers to publish online

Knowledge

- Lack of literacy
- Lack of sources
- Lack of education to access sources
- Lack of language social status
- Lack of language grammar
- Lack of language localization
- Lack of local knowledge self-recognition

Technology and bureaucracy

- Lack of Wikipedia brand awareness
- Lack of usability in Wikipedia tools
- Lack of policies enabling content
- Restricting content policies

Community

- Lack of mentors availability
- Lack of acknowledgement
- Lack of positive communication (harassment)
- Lack of community initiatives (e.g. GLAM, Wiki Loves Monuments, etcetera.)
- Lack of readers

Diversity group: context editors.
Other groups: gender, LGBT+, etc.

Speakers of a language in a context (country or region)

A: Context Characteristics

- Number of speakers
- Geographical area
- Number of urban areas
- Development indexes

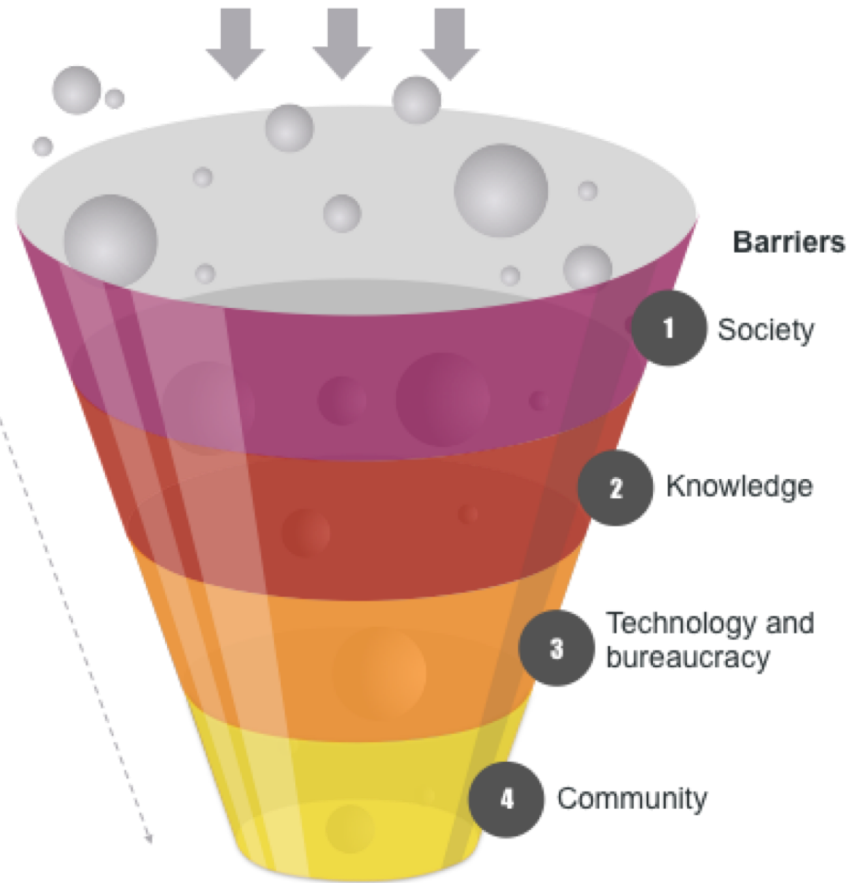
C: Engagement (Editors)

- Number of readers
- Number of editors
- Number of active editors

- Number of editors editing CCC and %
- Number of active editors editing CCC
- Percentage of edits in CCC

D: Cultural Context Content in Wikipedia

- Number of articles and their %
- Number of geolocated articles
- Density of geolocated articles by Km2
- Degree of completion of Top CCC articles
- Number of potentially missing CCC articles existing in a higher status language (only marginalized languages)
- Etcetera.



Editors representing their cultural context in a Wikipedia language edition



B: Society and Wikipedia (Barriers)

Society

- Lack of access to Internet
- Lack of economical conditions
- Lack of welfare
- Authorities interference on community dynamics
- Legal barriers to publish online

Knowledge

- Lack of literacy
- Lack of sources
- Lack of education to access sources
- Lack of language social status
- Lack of language grammar
- Lack of language localization
- Lack of local knowledge self-recognition

Technology and bureaucracy

- Lack of Wikipedia brand awareness
- Lack of usability in Wikipedia tools
- Lack of policies enabling content
- Restricting content policies

Community

- Lack of mentors availability
- Lack of acknowledgement
- Lack of positive communication (harassment)
- Lack of community initiatives (e.g. GLAM, Wiki Loves Monuments, etcetera.)
- Lack of readers

The more we know about the **context** and the **barriers**, the more we will be able to move forward in the understanding and fostering of different knowledge representation.

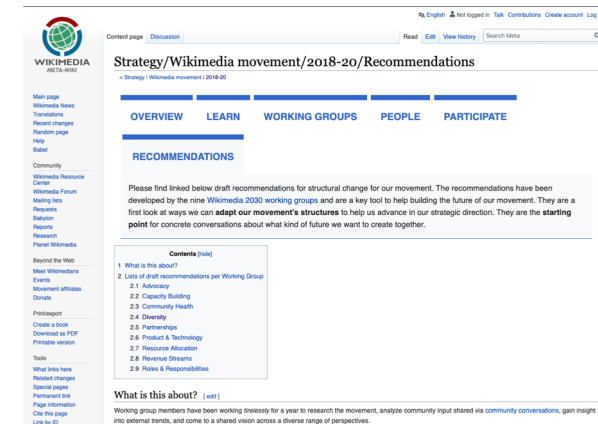


Recommendation for Wikimedia 2030 Strategy: Barriers to Wikipedia

<https://tinyurl.com/wikipediabarriers>

Identifying the Wikimedia Editing and Community Diversity Barriers for Users in Each Country and Introduce Them in Wikidata

To identify and look for indicators related to the different barriers to editing on Wikimedia projects and representing knowledge on the different diversity groups (i.e. women, geographical, marginalized groups, etc.) and store them in one place where they can be consulted (e.g. as properties/values in Wikidata to link them to other stakeholder projects).



<https://tinyurl.com/wikipediabarriers>

https://meta.wikimedia.org/wiki/Strategy/Wikimedia_movement/2018-20/Recommendations

Identifying the Wikimedia Editing and Community Diversity Barriers for Users in Each Country and Introduce Them in Wikidata

To identify and look for barriers

Research Papers
Community Feedback



To codify as indicators

e.g. Digital Divide
(number of Cellphones)



Store them (e.g. as properties/values in Wikidata)

Create an external interface (web?) to check the barriers for a particular country or region.

URLs	Change	Trend	Hint
225	None		Referer-Policy HTTP header is missing
225	None		Strict-Transport-Security HTTP (HSTS) header is missing
225	None		X-Content-Type-Options HTTP header is missing
229	None		X-Frame-Options HTTP header is missing or invalid
229	None		X-XSS-Protection HTTP header is missing or invalid
228	-1		Content-Security-Policy HTTP header is missing or invalid
228	-1		Leaks server information useful for compromising servers
None	None		Has deprecated Public-Key-Pins HTTP header
None	None		Has deprecated Public-Key-Pins-Report-Only HTTP header

URLs	Change	Trend	Hint
228	-1		Has JavaScript served via a CDN without subresource integrity

Groups of Barriers

Society

- Lack of access to Internet
- Lack of economical conditions
- Lack of welfare
- Authorities interference on community dynamics
- Legal barriers to publish online

Technology and bureaucracy

- Lack of Wikipedia brand awareness
- Lack of usability in Wikipedia tools
- Lack of policies enabling content
- Restricting content policies

Knowledge

- Lack of literacy
- Lack of sources
- Lack of education to access sources
- Lack of language social status
- Lack of language grammar
- Lack of language localization
- Lack of local knowledge self-recognition

Community

- Lack of mentors availability
- Lack of acknowledgement
- Lack of positive communication (harassment)
- Lack of community initiatives (e.g. GLAM, Wiki Loves Monuments, etcetera.)
- Lack of readers

Engagement Question:

What editing or diversity barriers are the hardest to codify as indicators?

Why this “Barriers to Wikipedia”?

- To understand better the relationship between context and knowledge representation in each language edition.
- Identifying the barriers in each territory (country or region) is key to finding solutions (partnerships, technology donations, etc.) and prioritize those with a greater probability of success. Replicate what works.
- To make this information accessible to Wikipedians in one place (an eternal site? Wikidata?) can be useful to avoid some circular discussions in events.

Thank you very much!

You can provide feedback on the recommendation:
<https://tinyurl.com/wikipediabarriers>

Marc Miquel

{marcmiquel@gmail.com}

Username:marcmiquel

Pompeu Fabra University, Barcelona, Catalonia

Amical Wikimedia (Catalan Wikipedia)

Wikimedia WG Diversity 2030

Diversity Observatory



References (if you want to know more)

Miquel-Ribé, M., & Laniado, D. **(2019)**. Wikipedia Cultural Diversity Dataset: A Complete Cartography for 300 Language Editions. *Proceedings of the 13th International AAAI Conference on Web and Social Media. ICWSM. ACM.*

Miquel-Ribé, M., & Laniado, D. **(2018)**. Wikipedia Culture Gap: Quantifying Content Imbalances Across 40 Language Editions. *Frontiers in Physics, 5, 12.* (CC BY) Open Access.

Miquel-Ribé, M. **(2017)**. *Identity-based motivation in digital engagement: the influence of community and cultural identity on participation in wikipedia* (Doctoral dissertation, Universitat Pompeu Fabra).

Miquel-Ribé, M., & Laniado, D. **(2016)**. Cultural identities in wikipedias. In *Proceedings of the 7th 2016 International Conference on Social Media & Society* (p. 24). ACM.