**Measuring Arabic
Wikipedia Community Health**

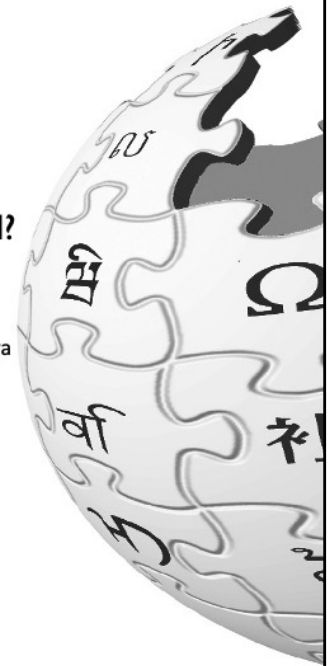Are We "Open" to Community Growth and Renewal?

Marc Miquel i Ribé

Eurecat, Centre Tecnològic de Catalunya
marc.miquel@ce.eurecat.org

Cristian Consonni

Eurecat, Centre Tecnològic de Catalunya
cristian.consonni@acm.org

David Laniado

Eurecat, Centre Tecnològic de Catalunya
david.laniado@eurecat.org

Hi everyone!

Welcome to this presentation. It is called "Measuring Arabic Wikipedia Community Health" with the subtitle question:

Are We "Open" to Community Growth and Renewal?

Hello! My name is Marc Miquel,
I'm a Wikimedia researcher.

"Not this kind of doctor"

My name is Marc Miquel and I'm a Wikimedia researcher. I joined the Movement in 2011.

I'm a PhD. PhDs usually have to say "I'm not this kind of doctor", because they do not deal with health.

In this case, however, I'm studying health, in particular, Wikipedia Community Health.

Community Health Metrics (CHM) is a research project to measure and understand Wikipedia communities' health, and provide tools and recommendations to improve it.

"We **envision** a community who is **aware of its current state of health** and both accompanies newcomers in the process of becoming Wikipedians, sustains and takes care of itself, and ultimately, farewells and recognizes editors in their process of retirement."

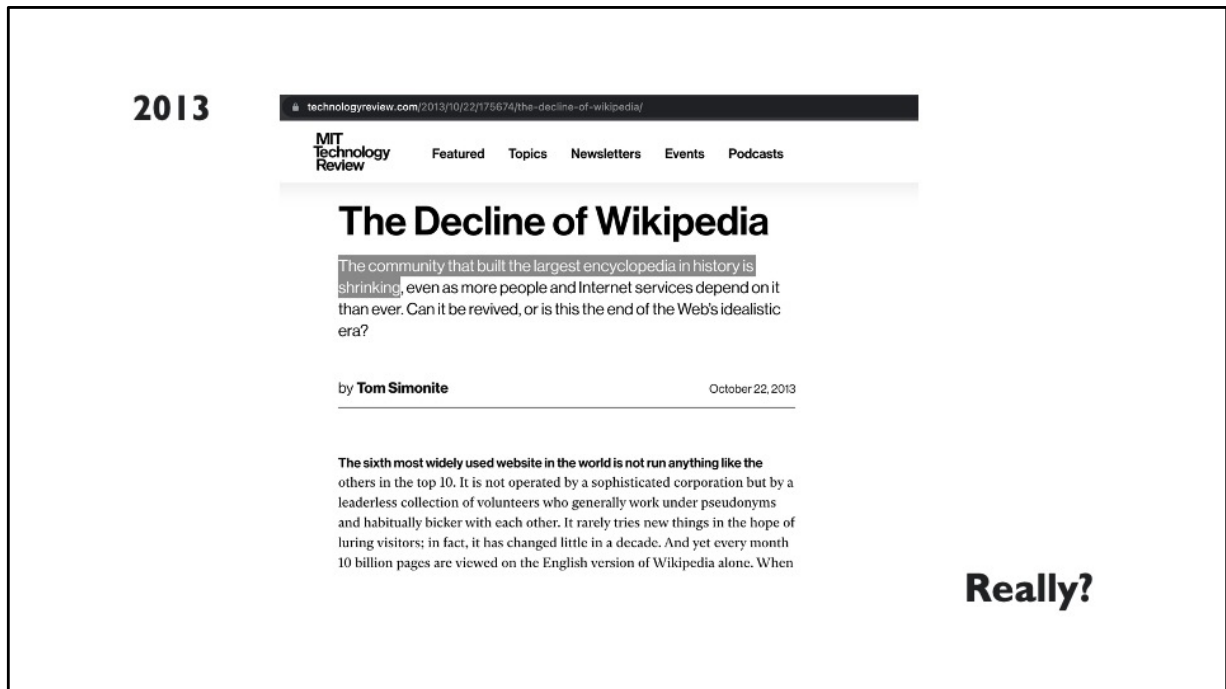https://meta.wikimedia.org/wiki/Community_Health_Metrics

Along with David Laniado and Cristian Consonni, we work on a research project called **Community Health Metrics (CHM).**

We want to measure, understand, raise awareness, and make recommendations to improve Wikimedia communities health.

"We **envision** a community who is **aware of its current state of health** and both accompanies newcomers in the process of becoming Wikipedians, sustains and takes care of itself, and ultimately, farewells and recognizes editors in their process of retirement."

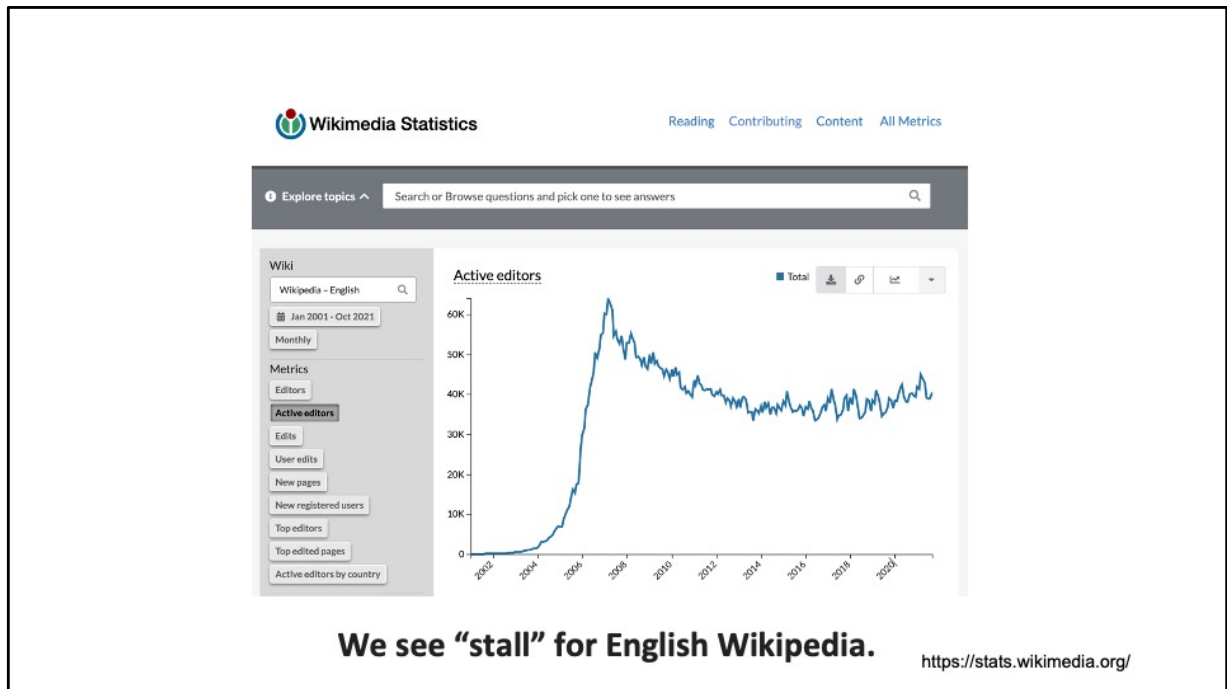Please, take a look at the page. It's under construction though.

In this presentation, we want to talk about **Communities Growth** and **Renewal**.

Wikipedia has reached its second decade, being the largest multilingual and collaborative free knowledge repository in human history. But, scientific studies over the past ten years have shown that Wikipedia has been unable to continue growing its editor communities.

Ever since 2016, community health is a recurring topic of debate among Wikipedians.

But is this decline of Wikipedia communities true? Generalizable?

We see "stall" for English Wikipedia.

https://stats.wikimedia.org/

It is true. We see growth and "stall" for English Wikipedia number of active editors, with a little growth back again in the current year.

You can check this at stats.wikimedia.org.

But also for other languages?

We will *zoom* on the "active editor community" using a <u>quantitative analysis approach</u>.
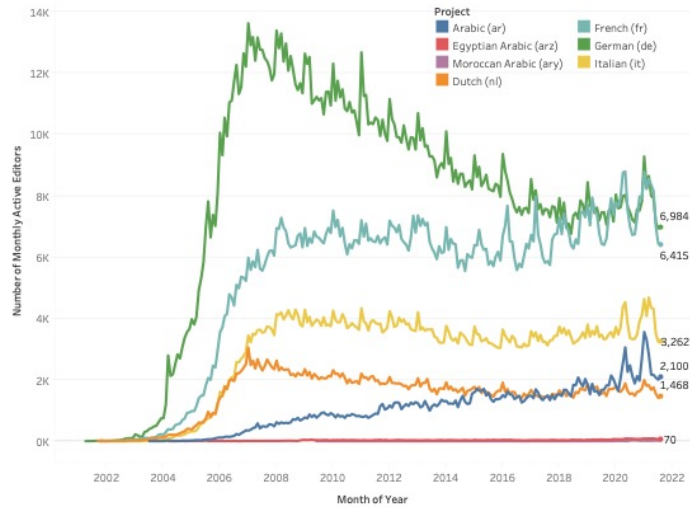
ويكيبيديا
الموسوعة الحرة

In this talk, we'll focus on Arabic Wikipedias (Arabic and the two dialects, Egyptian and Moroccan).

For comparison, we've chosen French, German, Italian and Dutch.

We will *zoom* on the "active editor community" using a <u>quantitative analysis approach</u>.

I apologize in advance for the quantity of graphs I'll show you and I invite you to take another look at the slides later with more calm.

Arabic Wikipedia: 2100 active editors* in January 2021.

* At least 5 edits per month

So how does the curve of active editors look like for other languages? Well, for European languages, we see similar curves of growth and stall to English Wikipedia.

But for Arabic Wikipedia (in dark blue), we see a slow but steady growth, with some high peaks last year.

7

The question we want to answer for **Arabic Wikipedias**

Are We "Open" to Community Growth and Renewal?

We want to answer this question on growth and renewal.

Let's study its current state in different ways.

6 Vital Signs:

1. Retention
2. Continuity
3. Balance

4. Specialists
5. Admins
6. Global

In general

In specific functions

To do so, we will rely on the medical concept "Vital Signs".
Vital Signs indicate the status of the body's vital (life-sustaining) functions.

These measurements are taken to help assess the general physical health of a person, give clues to possible diseases, and show progress toward recovery.

For a community, defining and measuring Vital Signs will allow us to prevent "situations of risk" for its future and even plan for growth.

For example, we want to avoid the classic "bus factor", which asks us what would happen to a project if its responsible people would be run over by a bus.

The indicators we have chosen are: community retention, continuity, and balance, in general for the active community. Taking into account more specific functions, we will observe specialists, flags, and global participation.
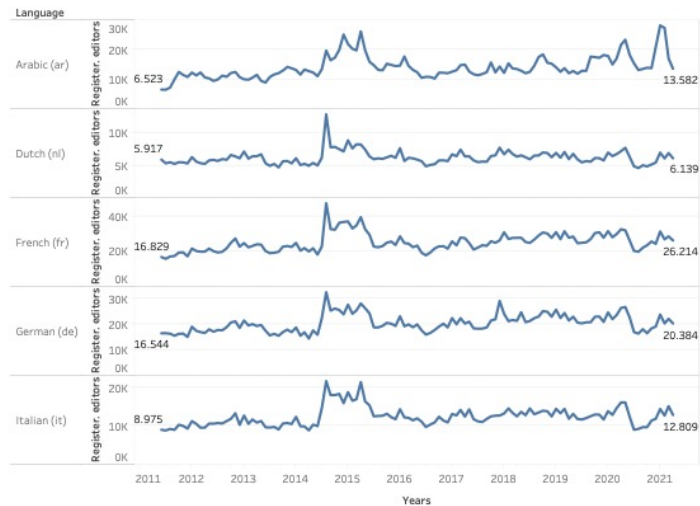
I. Retention

Community retention rate is computed as the % of new editors who survive 60 days after the first edit.

https://meta.wikimedia.org/wiki/Research:Surviving_new_editor

The first vital is retention. **Community retention rate is computed as the % of new editors who survive 60 days after the first edit and edit again.**
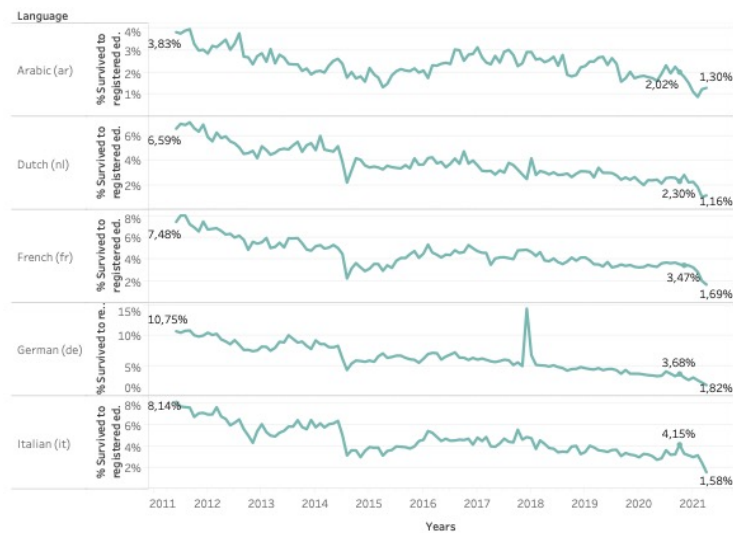
The number of registered editors has been quite stable for the past ten years.

Even growing! This is good! Wikipedia draws people interest. Some languages more than others, but they get people's attention.

Retention rate (%) has been going down in all languages.

What is not so good is…

That the percentage of editors who edited after registering and after 60 days… has been going down for the last ten years.

In Arabic Wikipedia, this is oscillating between 2 and 1.3% of new registered editors. This percentage is higher if we only take into account those who edit.
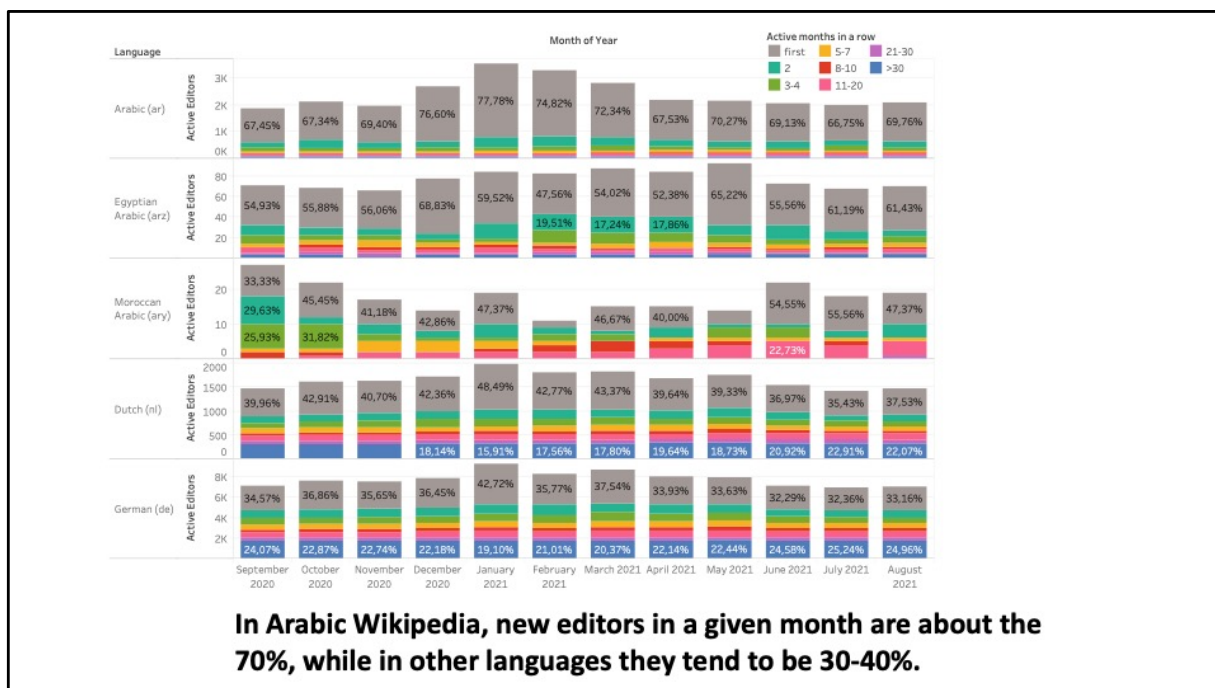
2. Continuity

Community continuity is the persistence of active editors as well as the succession of groups of editors over time.

The second vital sign is continuity. **Community continuity is the persistence of active editors as well as the succession of groups of editors over time.**

You want to ensure that there are fresh editors every month who had not edited on the previous month and that there are others who have edited for many more months.

In Arabic Wikipedia, new editors in a given month are about the 70%, while in other languages they tend to be 30-40%.

In this graph, we see the number of active editors on a monthly basis, andin color the number of months in a row they have been editing for each month. Grey means the first month, green 2 or 3-4.

In Arabic Wikipedia, fresh editors in a given month are about the 70% (in grey), while in others like Dutch and German at the bottom, they tend to be 30-40%.

This means that Arabic Wikipedia is more volatile. It is engaging fresh editors every month, but they do not continue the following months.

Ideally, you want editors who can edit several months in a row (for matters of collaboration or for a Wikiproject)
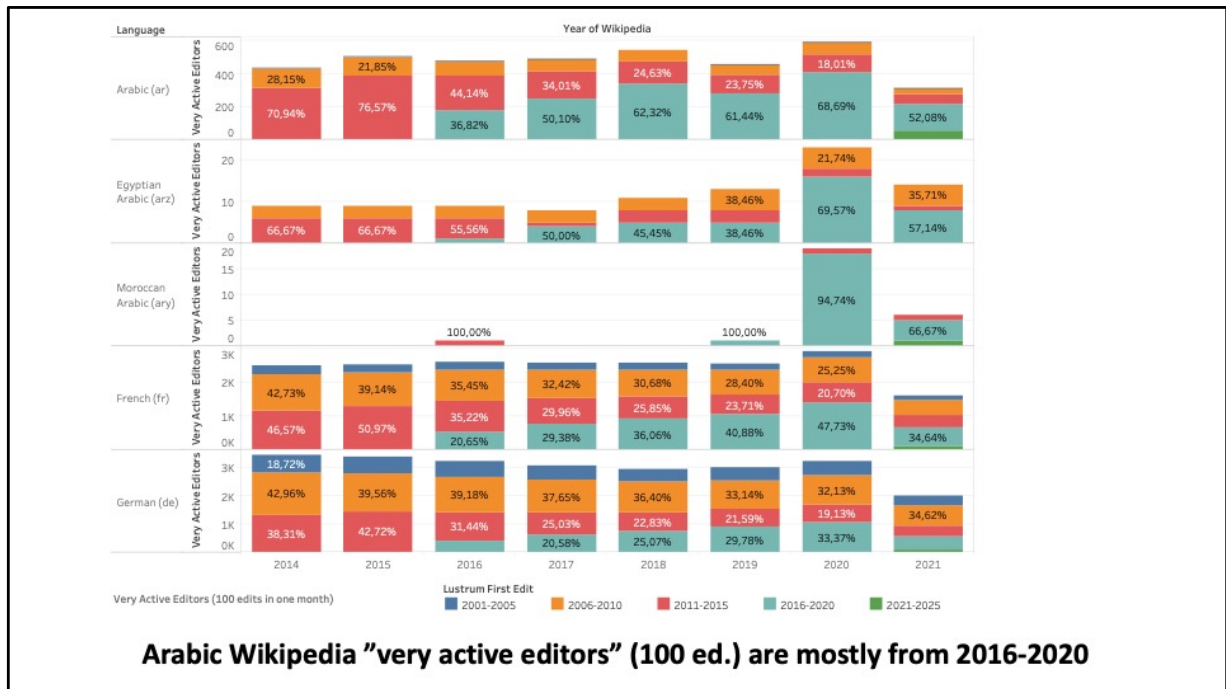
14

3. Balance

Community balance is being able to maintain an equitable proportion of old and new editors.

The third vital sign is balance. **Community balance is being able to maintain an equitable proportion of old and new editors.**

We want to benefit from experience but also be able to stay open to new generations. This is a key sign indicating renewal.

Arabic Wikipedia "very active editors" (100 ed.) are mostly from 2016-2020

In this graph, we see for five languages, the composition of the "very active editors" by lustrum of first edit.

2001-2006, 2006-2010, 2011-2015, 2016-2020, and 2021-2025. We could say that they're "generations".

Very active editors are those who make at least 100 edits per month.

A little more than 80% of the monthly human edits are done by very active editors. So we want to keep them, and we want a balanced community.

We could say that it would be recommendable that each generation had a similar proportion.

We see in European Wikipedias, like French and German, that the generation 2006-2010 survives incredibly well… and the generation of 2016-2020 has been incorporated.

What happens in Arabic Wikipedias is that the 2016-2020 is the main
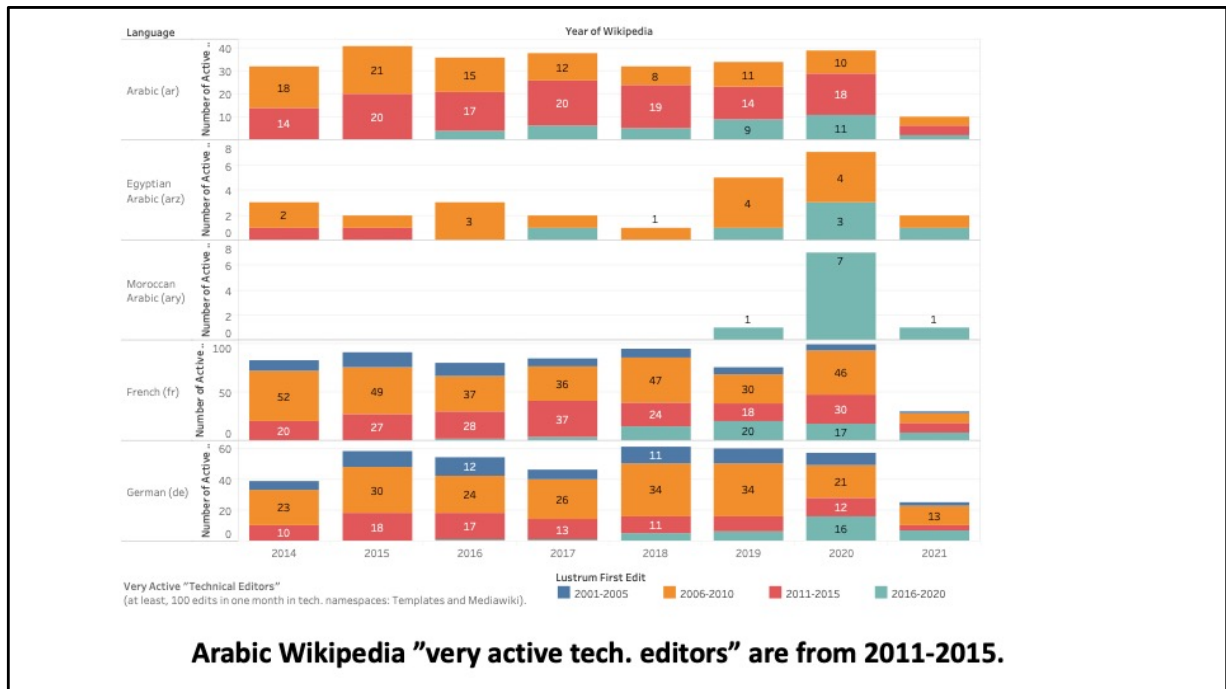
generation. There's barely no 2006-2010, not to mention 2001-2005. This means that the community is younger.

4. Special functions

Community technical and coordination functions undertaken by editors are essential for the project.

The fourth vital sign is special functions. **Community technical and coordination functions undertaken by editors are essential for the project.**
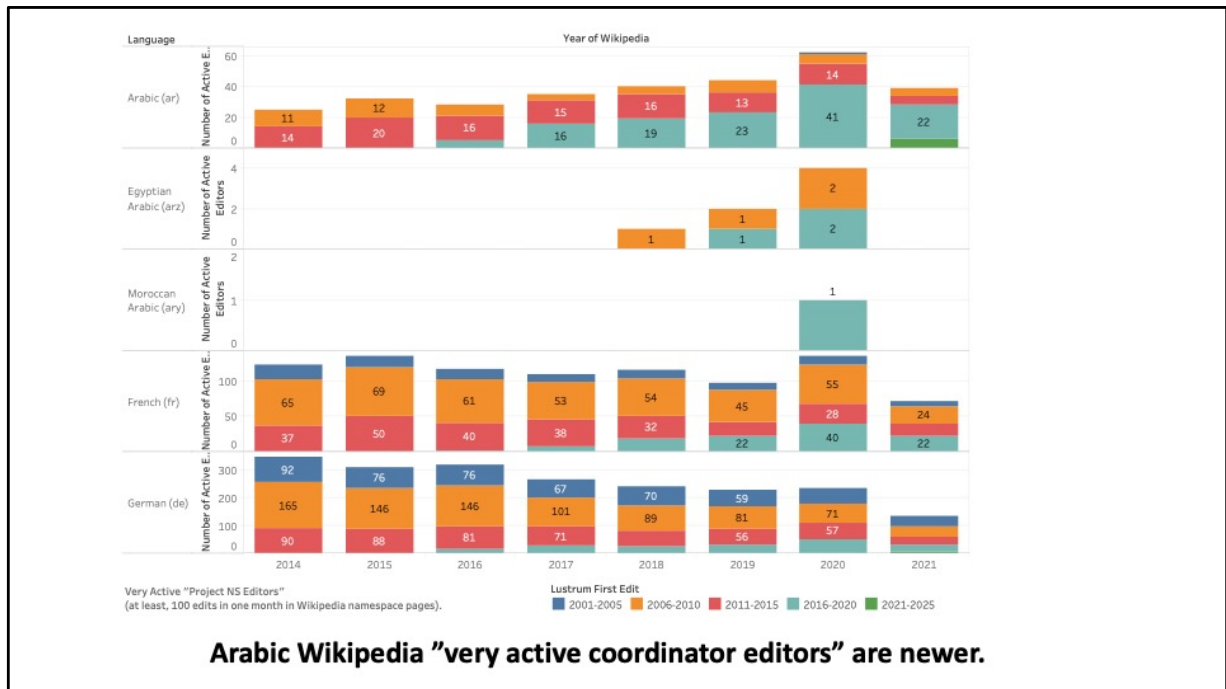
Arabic Wikipedia "very active tech. editors" are from 2011-2015.

Similarly with the previous measurement, here we focus on the very active technical editors. Editors who have done more than 100 edits in one month in technical namespaces (that is templates and mediawiki namespaces).

We check the generation they belong to. Arabic Wikipedia technical editors are much more balanced than the "normal" very active editors. The majority is from 2011-2015.
In general, technical editors are from previous generations in other Wikipedias too.

By the way, pay attention to the bus effect. In Arabic Wikipedia, there are only 40 very active technical editors.

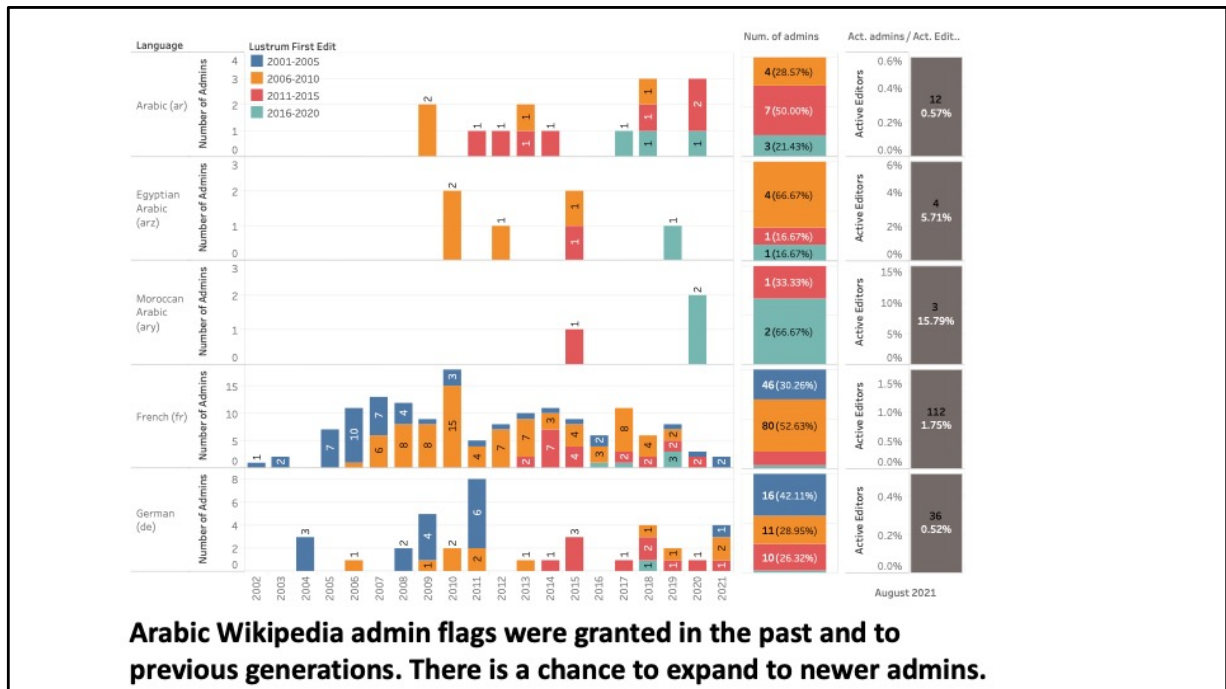**Arabic Wikipedia "very active coordinator editors" are newer.**

In this other case, we do exactly the same analysis but for those very active editors in Wikipedia namespace (this is, Wikiprojects, Village Pump, among others). The number of editors is higher… and at the same time, we see in Arabic Wikipedia that here that the majority is from the 2016-2020 generation.

5. Admins

Admins have rights and responsibilities in performing actions over content and take a key function for the community.

The fifth vital sign is admins. **Admins have rights and responsibilities in performing actions over content and take a key function for the community.**

Arabic Wikipedia admin flags were granted in the past and to previous generations. There is a chance to expand to newer admins.

In this complete graph, we see the admins flags granted by year and the color represents the generation they belong to.

Most flags were granted in 2011 to 2016 and to previous generations. We almost do not see flags from editors 2016-2020.

There is a chance to expand to newer admins.

On the very right we see the number of active admins in the past August. The percentage is the ratio between number of active admins by number of active editors. **This percentage varies according to the language.**

It might be useful to understand the weight an admin carries. We see that the percentage for Arabic is 0.57%, which in comparison with others could be higher, and therefore, expanding the number of admins might be a good idea.
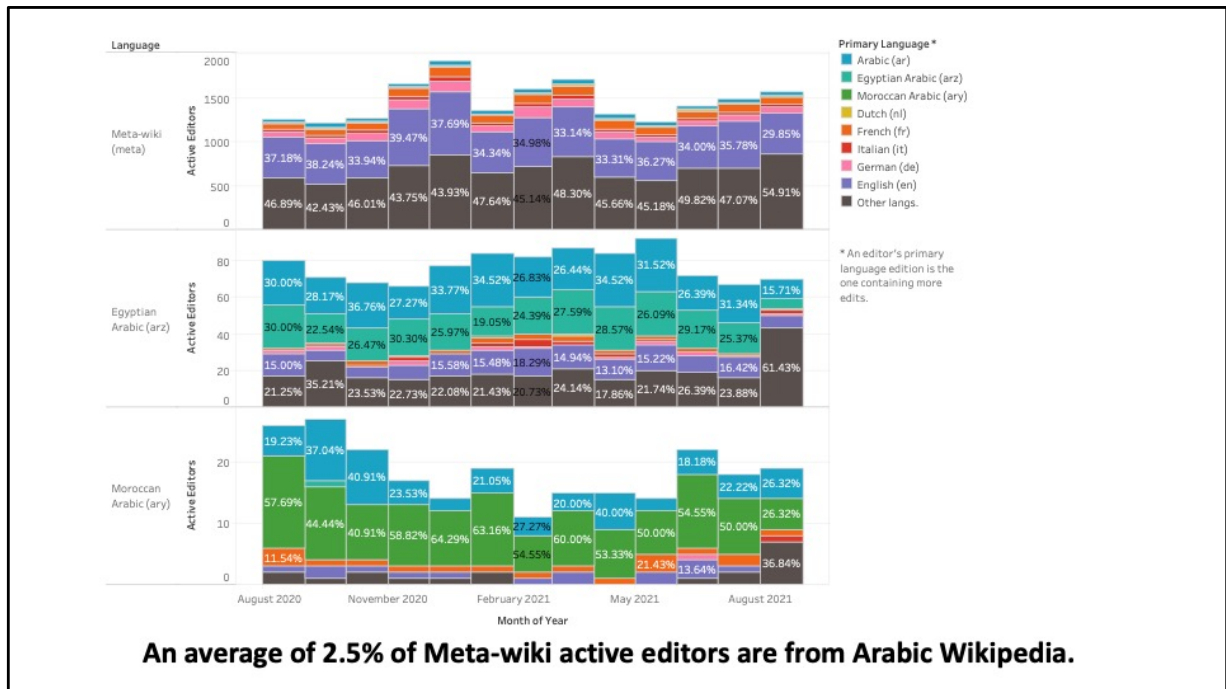
The sixth and last vital sign is global. **Communities participation in the "global community" is key to make their voice heard and learn from others.**

An average of 2.5% of Meta-wiki active editors are from Arabic Wikipedia.

In this graph we see the active editors of Meta-Wiki and the two Arabic dialects by their primary language.

An editor primary language edition is the one containing more edits. So, an editor may regularly edit German Wikipedia and more spontaneously English Wikipedia. Since she has more edits in the German Wikipedia, that's her primary language edition.

So, among the active editors in Meta-wiki, we only see an average of 2.5% whose primary language is Arabic Wikipedia. English is the primary language of more than a third of the active editors.

Curiously, we see that around a 30% of Egyptian Arabic Wikipedia primary language is Arabic Wikipedia. The situation is similar in Moroccan Arabic Wikipedia but with a lower percentage. This means that for a considerable part of the editors of this language, it is their secondary project.

Conclusions

6 Vital Signs:

1. Retention
2. Continuity
3. Balance

4. Specialists
5. Admins
6. Global

In summary, Arabic Wikipedia is a younger community than others. Even though the retention is not so good, we see that the community is composed mainly by editors of the 2016-2020 generation.

There is continuity between editors on a monthly basis, but the proportion of long-term engaged editors (more than an year editing on a monthly basis) is very little.

The "core" positions, technical and admin flags, are usually from previous generations. In comparison with the community in general, we see that admins are from previous generations before 2016.

What can you do as Arabic communities?

There is an obvious margin of improvement to stay **open** in **retention, balance, admin flags, and Meta-wiki participation.**

- Retention is key. Collaborate with affiliates to set **mentorship** strategies and be more ready to embrace **changes in UI** (WMF Growth team).
- Balance is important. To guarantee that a proportion of older generation of very active editors do not drop-off, give them recognition and appreciation.
- Plan for flag renewal. Arabic Wikipedia can increase the number of admins. Open pathways to new editors and diversity of profiles.
- Participate in Meta-Wiki. Encourage editors to share there and participate in the discussions.

There is an obvious margin of improvement to stay **open** in **retention, balance, admin flags, and Meta-wiki participation.**

Retention is key. Collaborate with affiliates to set **mentorship** strategies and be more ready to embrace **changes in UI** (WMF Growth team).

Balance is important. To guarantee that a proportion of older generation of very active editors do not drop-off, give them recognition and appreciation.

Plan for flag renewal. Arabic Wikipedia can increase the number of admins. Open pathways to new editors and diversity of profiles.

Participate in Meta-Wiki. Encourage editors to share there and participate in the discussions.

These analyses will eventually become **dashboards** on a website (work in progress). Your feedback is very valuable!

We also provide data, code and research behind these **Vital Signs**.

Another area of Community Health we are working on is **editor drop off.** We want to understand its causes, examining reverts, discussion patterns and emotions among other factors.

Thank you very much!

Marc Miquel i Ribé

Eurecat, Centre Tecnològic de Catalunya
marc.miquel@ce.eurecat.org

Cristian Consonni

Eurecat, Centre Tecnològic de Catalunya
cristian.consonni@acm.org

David Laniado

Eurecat, Centre Tecnològic de Catalunya
david.laniado@eurecat.org