Measuring African Wikipedias 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 African Wikipedias Growth and Renewal"

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

"Rethink + Reset: Visions of the future"

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

Rethink + Reset: Visions of the future is the title of this conference. I want to talk about the future, I want to talk about 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.

One aspect of Community Health is **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.

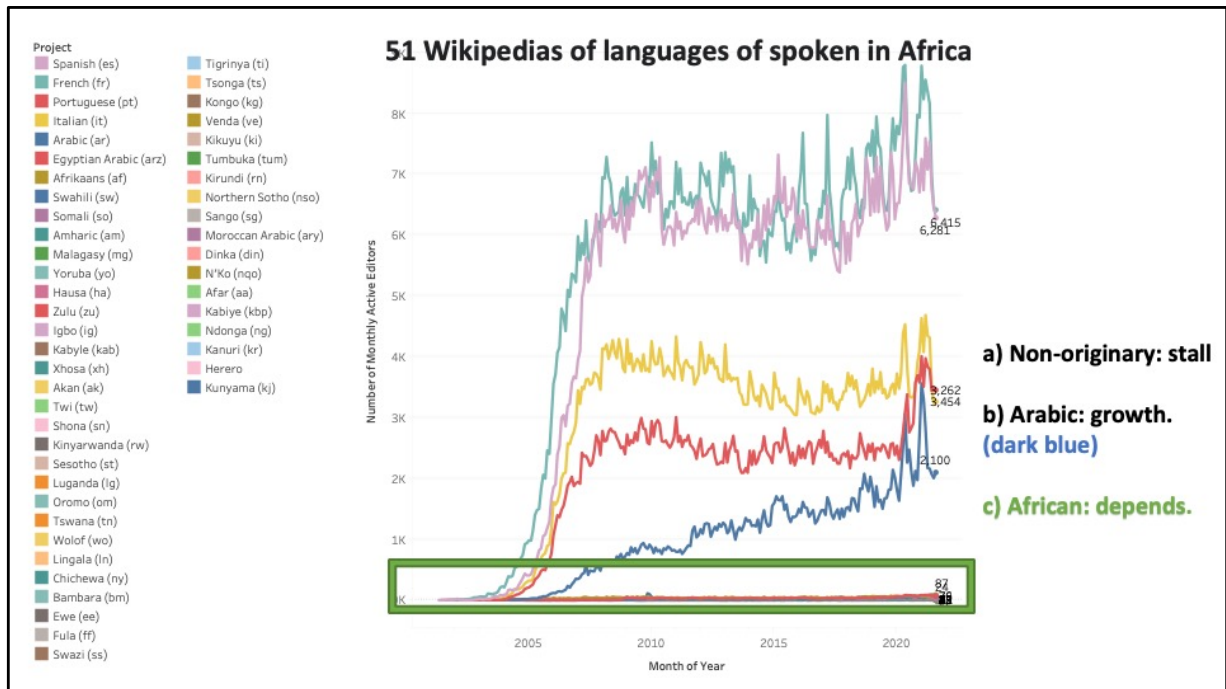Ever since 2016, community health is a recurring topic of debate among Wikipedians.

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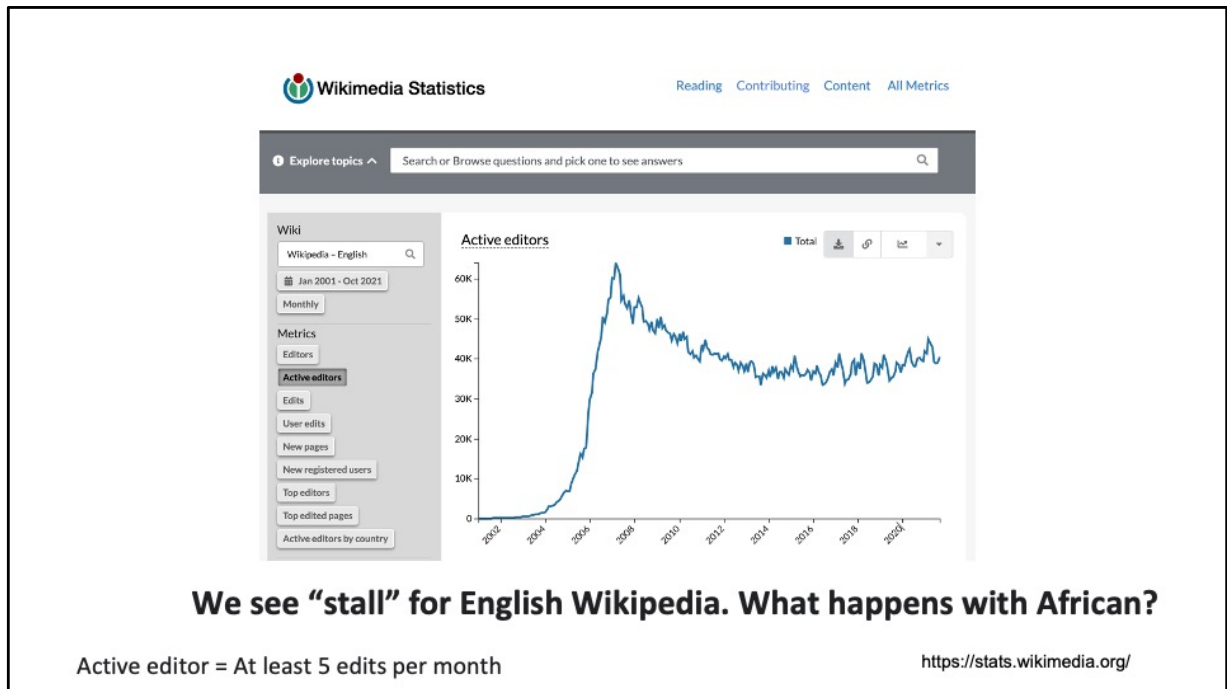
It is true. We see growth and "stall" for English Wikipedia in the number of active editors, with a little growth back again in the current year.

You can check this at stats.wikimedia.org.

But also for other languages? What about the African?

51 Wikipedias of languages of spoken in Africa

a) Non-originary: stall
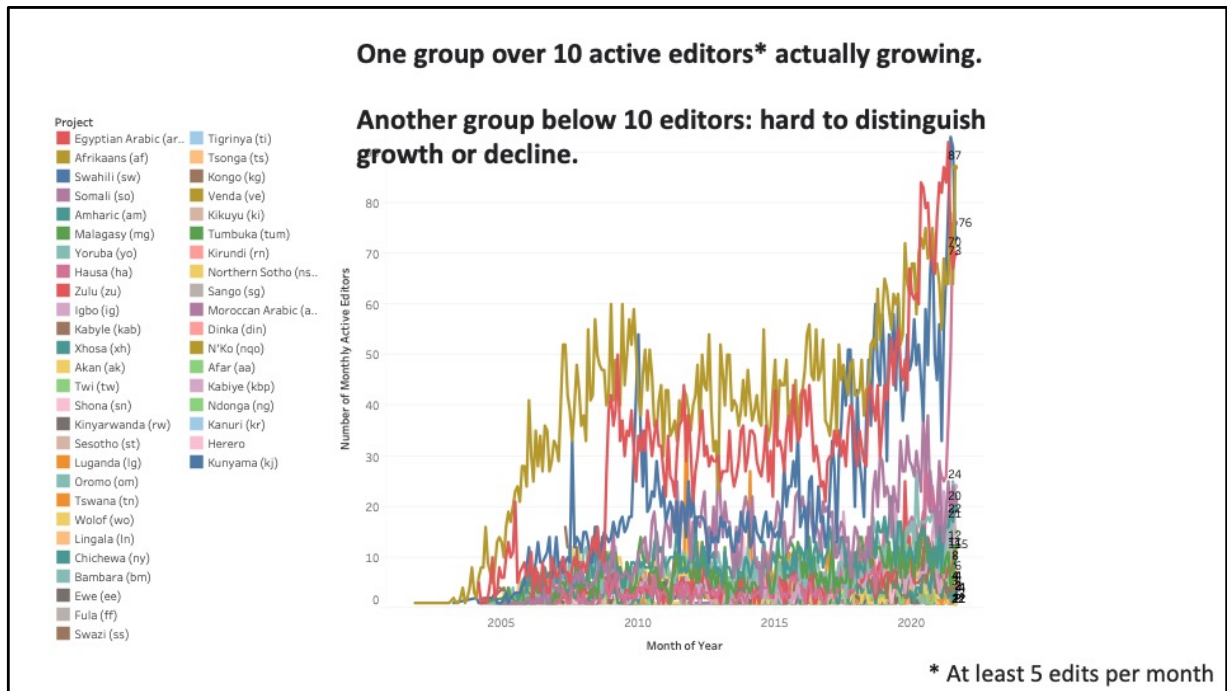
b) Arabic: growth.
(dark blue)

c) African: depends.

Well, there are actually 51 African languages. We have three stories.

One story for the non-originary languages: English, French, Portuguese and Italian. They are in an order of 3 to 7-8 thousand active editors. They stall.

A second story for another big language, Arabic: grows.

A third story for the African originary languages: it depends.

In the graph we see them all. The first are in the order of thousands of active editors. Arabic is at 2 thousand at the moment. African languages, depends.

One group over 10 active editors* actually growing.

Another group below 10 editors: hard to distinguish growth or decline.

* At least 5 edits per month

Among this third group, the originary African languages, we have:

One group over 10 active editors* actually growing.
Another group below 10 editors: hard to distinguish growth or decline.
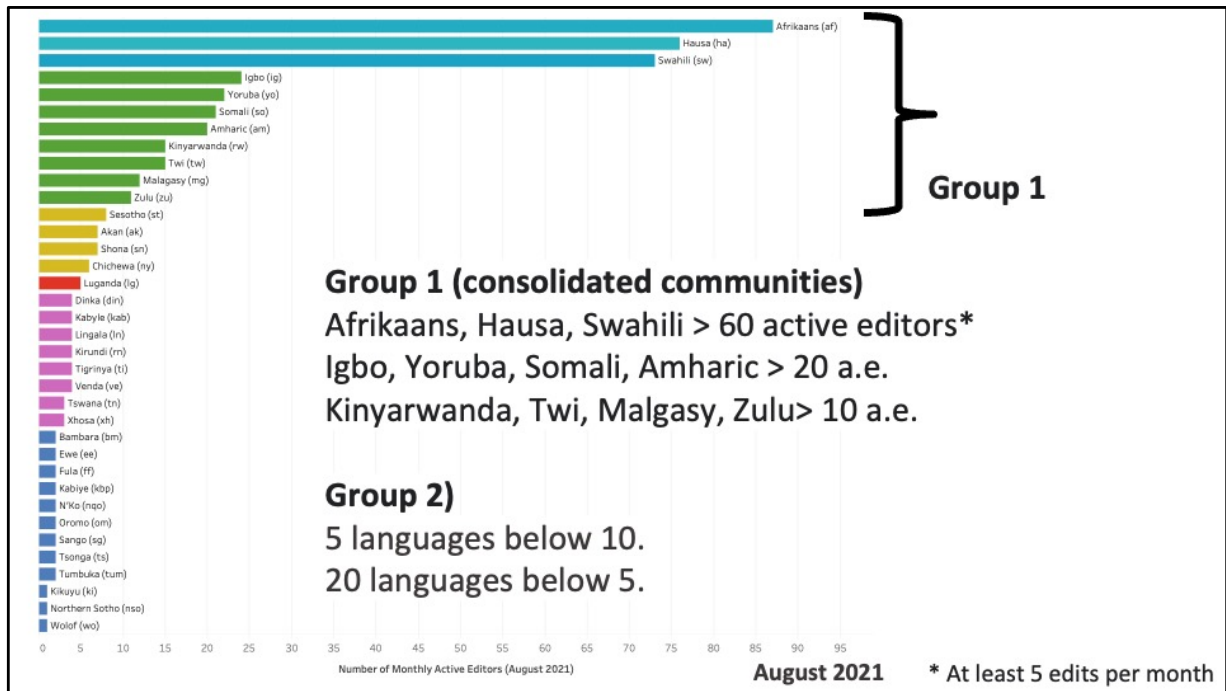
So, we have a group 1 of consolidated communities.
Afrikaans, Hausa, Swahili > 60 active editors
Igbo, Yoruba, Somali, Amharic > 20 a.e.
Kinyarwanda, Twi, Malgasy, Zulu> 10 a.e.

Then, we have group 2 with 5 languages.

9 consolidated
25 unconsolidated.

I do not know if there is a threshold for a minimum number of members for a community, but there are 25 languages

We will focus on African originary languages

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

In this talk, we'll focus on African originary Wikipedias.

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.

The question we want to answer for **African Wikipedias**:

Is there growth or renewal of the active editor community?
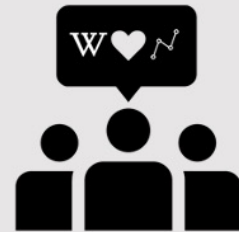
We want to answer this question on growth and renewal.

Let's study the active editors communities 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.

We are creating an "Active Community" Vital Signs framework.

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 six: 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.
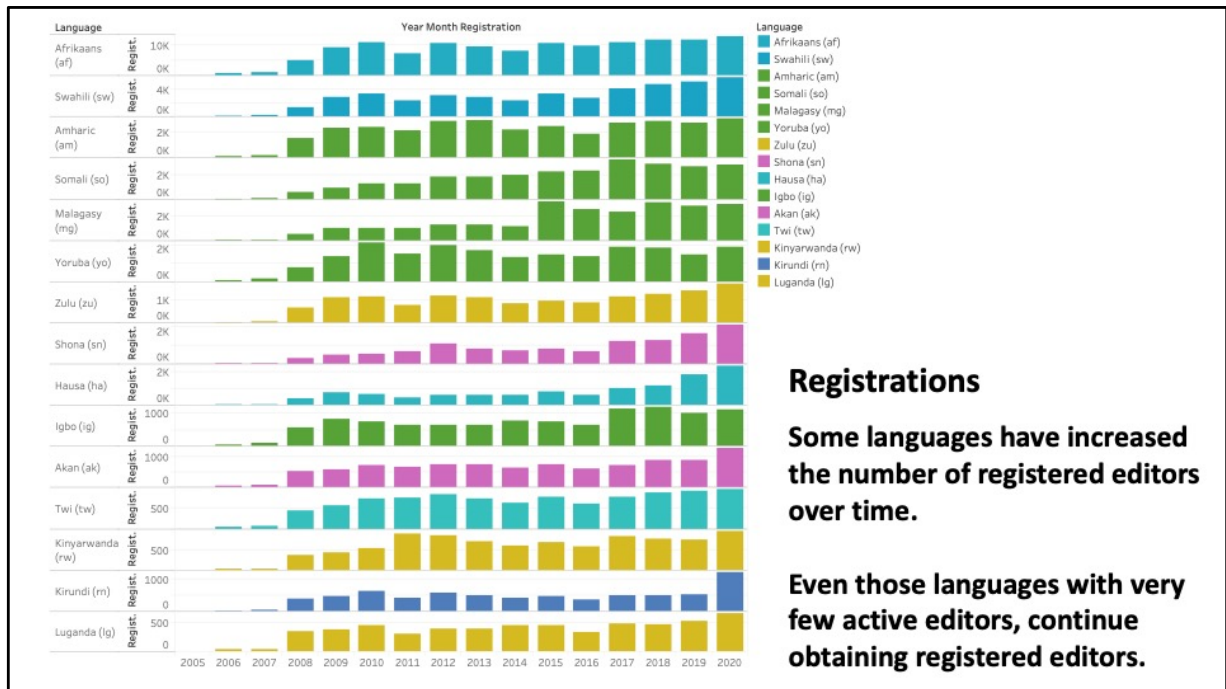
I. Retention

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 stable or growing.

Here in the graph we see some of the African originary languages.

Some languages have increased the number of registered editors over time.

Here on the bottom we see the Luganda Wikipedia for example. Hausa in the middle is also growing a lot.

Also languages with very few active editors keep obtaining new registered editors.

Even growing! This is good! Wikipedia draws people interest. Some languages more than others, but they get people's attention.
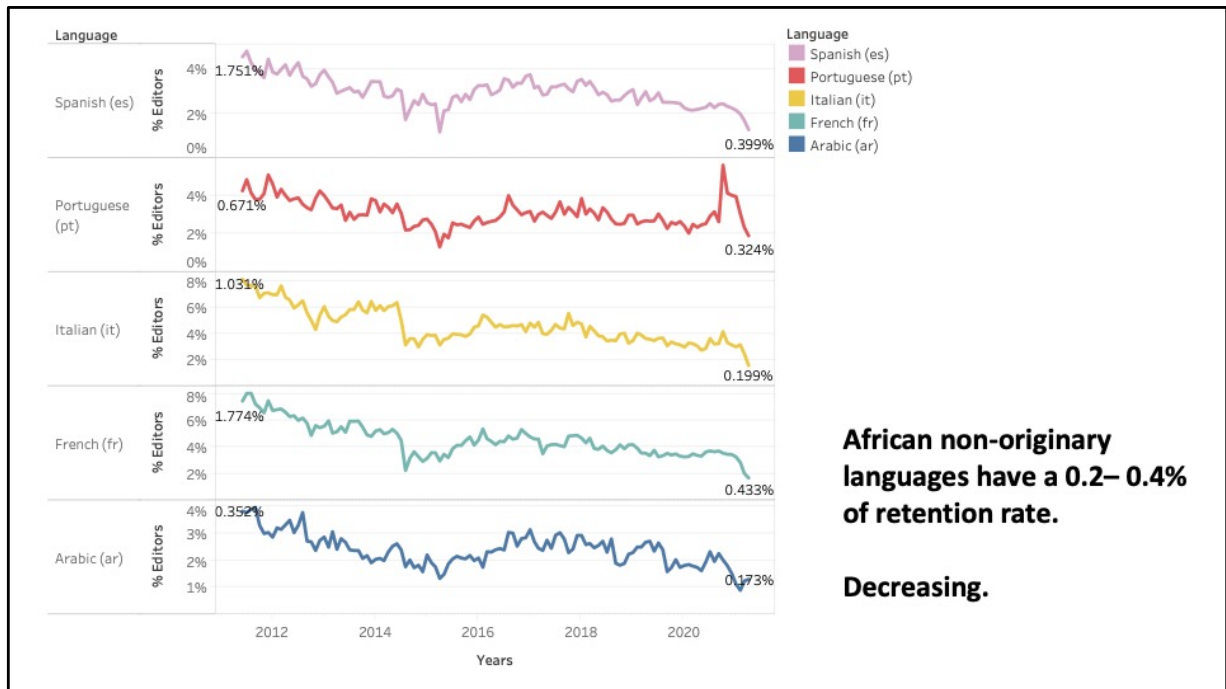
What is not so good is…the retention rate.

For comparison, I show you the non-originary African languages. They have a 0.2 – 0.4% retention rate. These are editors who edit 60 days.

Instead, the is decreasing for all languages, which means that many of the editors we saw in the previous graph just do one single edit and do not return.
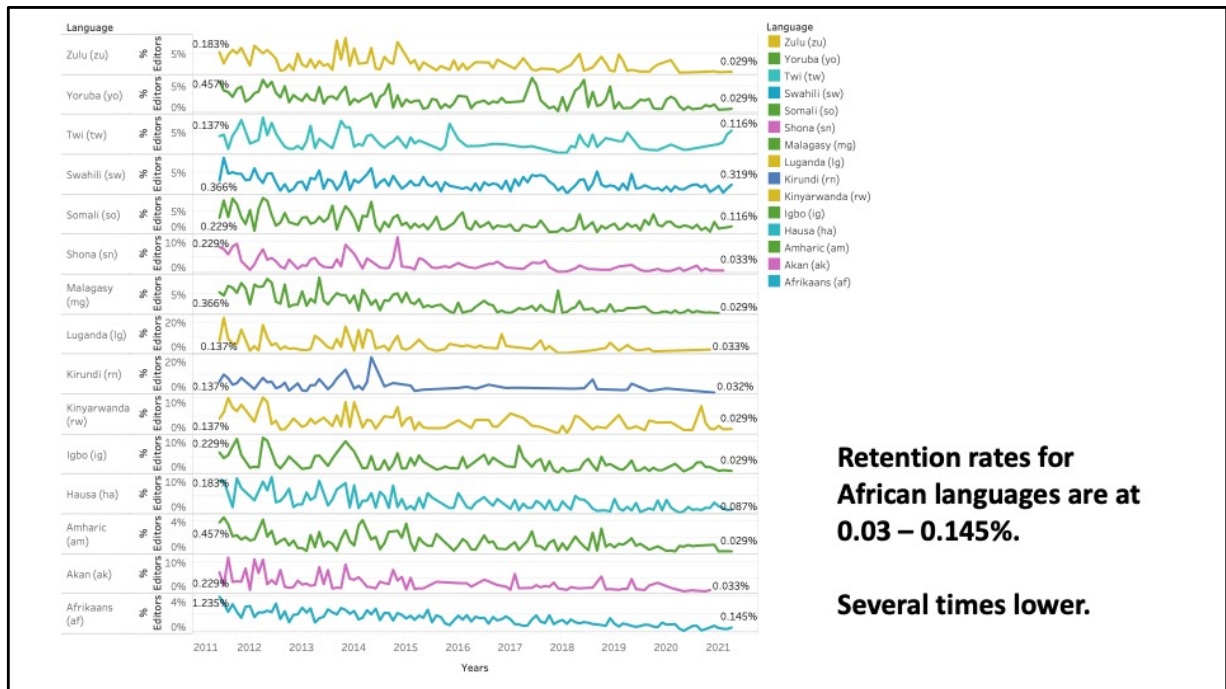
What surprised me is that the retention rate for the African originary languages is much lower. We see it ranges from 0.03 to 0.145%. Between 5 and 10 times lower.

This explains the few active editors.

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.
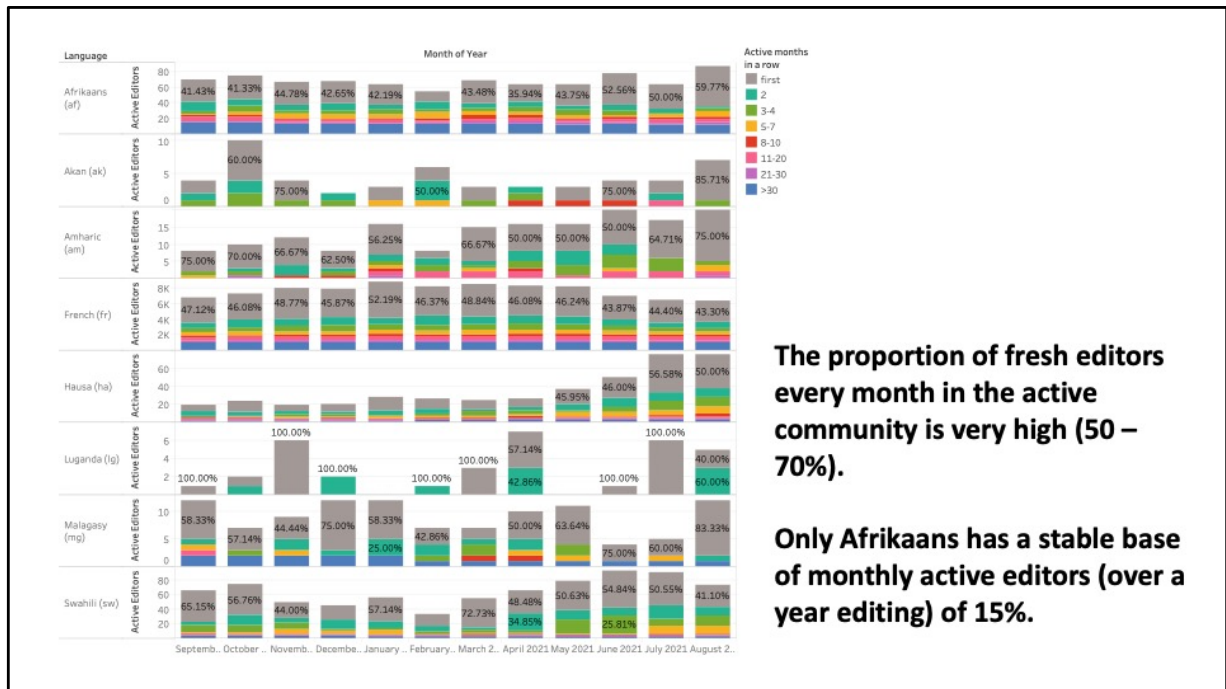
The proportion of fresh editors every month in the active community is very high (50 – 70%).

Only Afrikaans has a stable base of monthly active editors (over a year editing) of 15%.

We see a lot of volatility.

The colours show the number of active months an editor edits in a row. Brown is the first or fresh editor. The proportion of fresh editors every month in the active community is very high (50 – 70%).

Fresh editors
French is 40-45%;

The fresh are in brown. But in blue we have > 30 months (two years and half) editing every month.
This is usually 15% of the active community for consolidated communities (French). This only happens with Afrikaaans. The second African language in number of active editors, Swahili, is very volatile.

Ideally, you want a community with editors who can edit several months in a row up to a year (for matters of collaboration or for a Wikiproject, for keeping

17

memory of certain conflicts or situations, for offering mentorship, etc.). We see that the "stable community" which edits more than 6 months in a row is nearly 50% of the active community of editors; those who have edited more than 30 months in a row (2 years and half) are almost a quarter. This means that these communities contain a core of very committed editors

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.
It is connected to the previous ones.

Swahili and Malagasy keep a balanced community of very active editors.
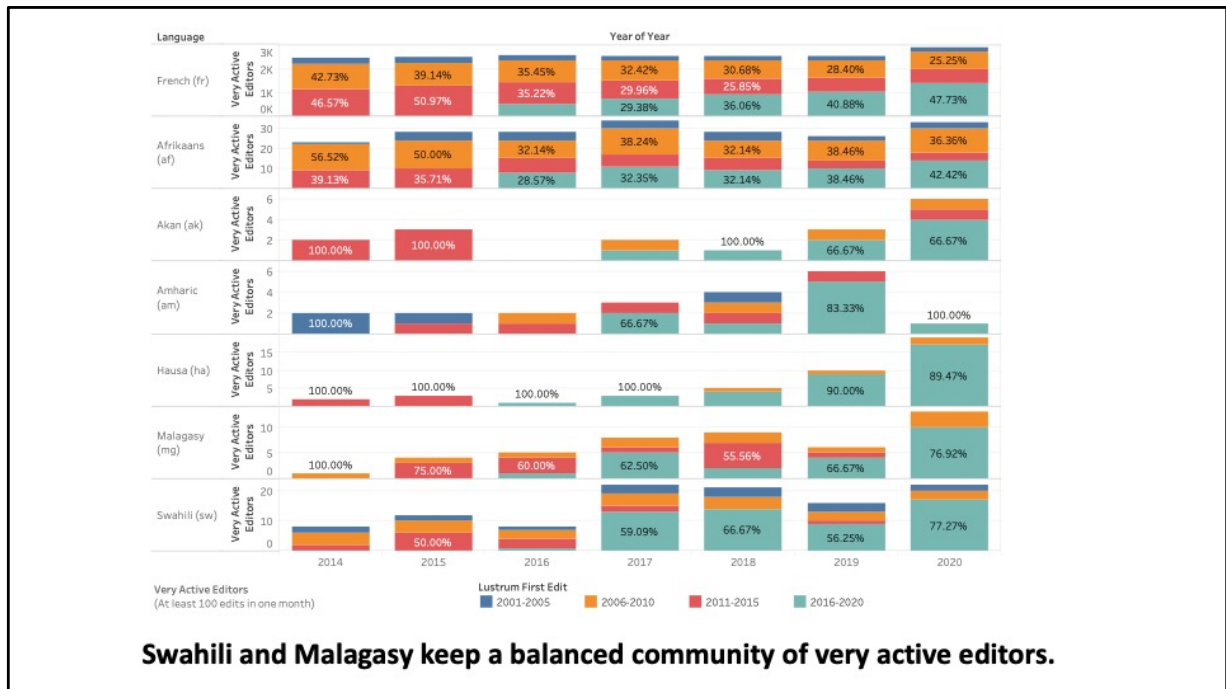
In the following graph, we see for seven African language editions and their "very active editors" over the past 7 years. By definition, very active editors are those who make at least 100 edits per month. So the graph shows the yearly number of editors who have been "very active" at least one month.

Typically, the "very active editors" account for an 80% of the human monthly edits in a Wikipedia. This is why we consider them so important, they are the productive of the community.

The colour shows within each year the number of edits by the lustrum in which they did their first edit (2001-2006, 2006-2010, 2011-2015, 2016-2020, and 2021-2025.) We could say that they are "generations".

I also include the French here, for comparison.

We basically see that the very active editors in African communities are from this past lustrum 2016 to 2020. They are new.

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.**

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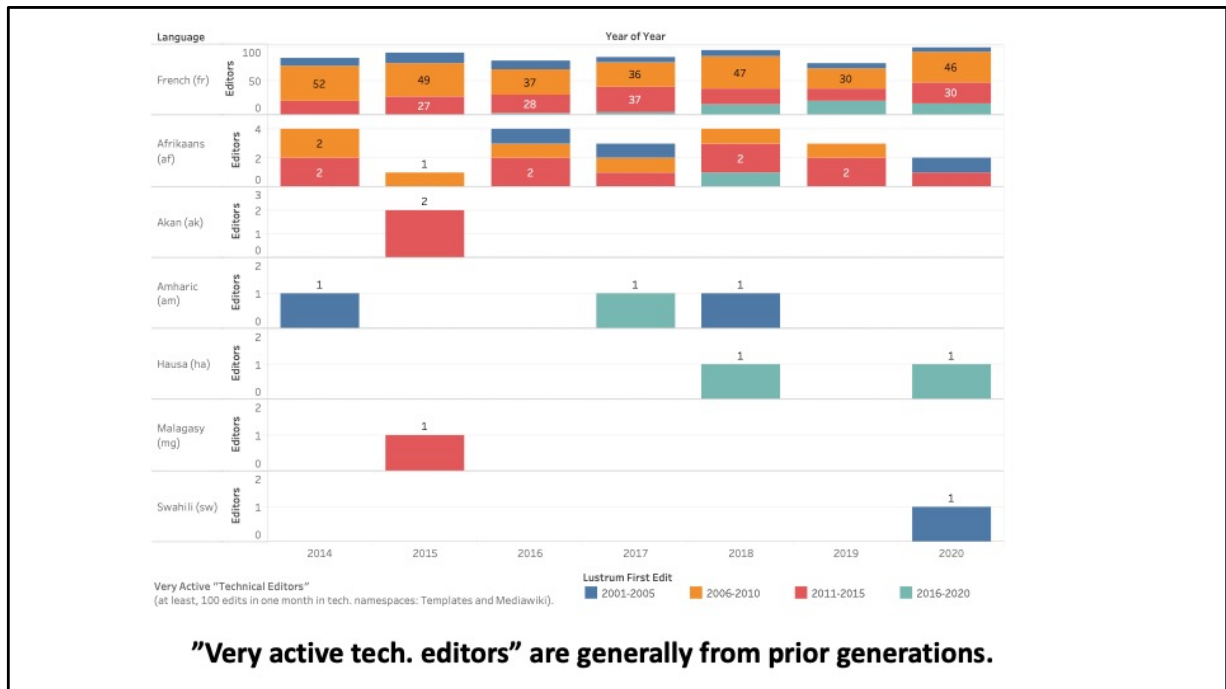
Similarly with the previous measurement, here we focus on the very active technical editors. Editors who have done more than 100 edits in one month in technical namespaces (that is templates and mediawiki namespaces).

In the graph, we see the number of "Very active technical contributors" for the selected African languages. They are very few.

Even languages like Hausa and Swahili only have one "very active technical editor" and it is in 2020.

There are a few "very active coordinator editors" and they are from prior generations.
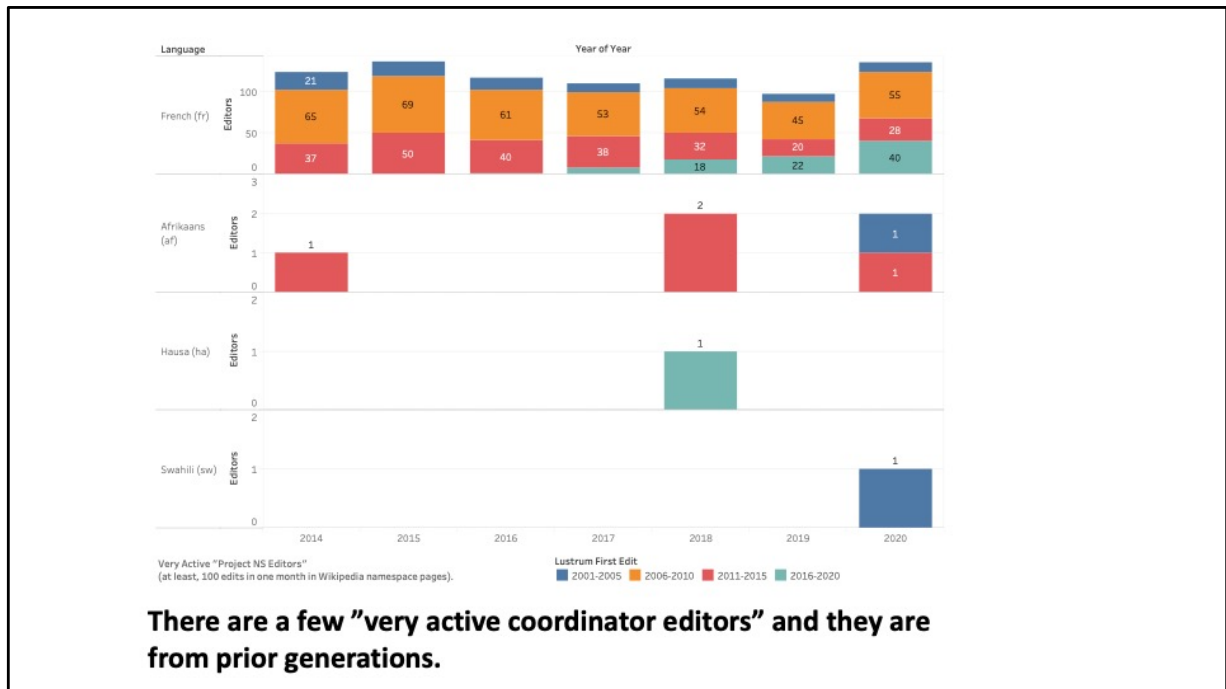
In this other case, we do exactly the same analysis but for very active editors in Wikipedia namespace (this is, Wikiprojects, Village Pump, among others).

In the graph, we see the number of "Very active project coordinators". For French Wikipedia, they are 120 in 2020, many more compared with the technical (100), but many less than the overall group of very active editors or active editors.
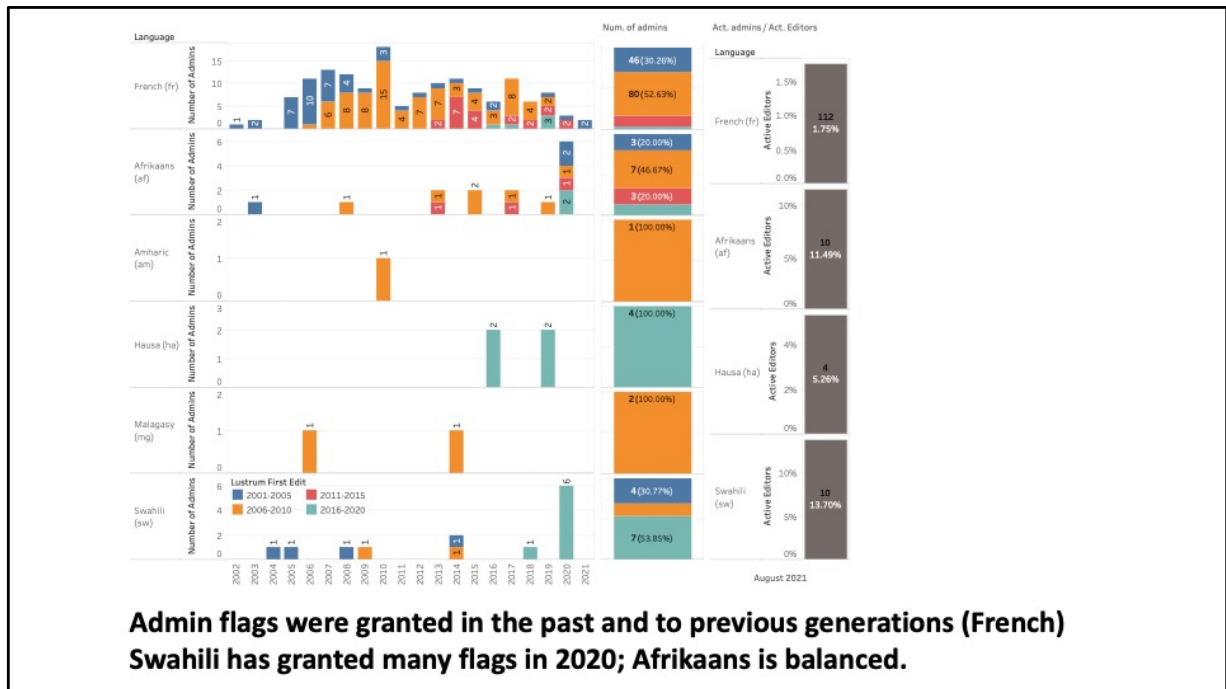
For the African originary languages, the situation is very similar. There are very few or no coordinators.

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.**

Admin flags were granted in the past and to previous generations (French)
Swahili has granted many flags in 2020; Afrikaans is balanced.

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

In many Wikipedias, most flags were granted in 2011 to 2016 and to previous generations. In large language editions like French, the admins tend to be from previous generations.

However, in a younger Wikipedia like the **Swahili, we see it has granted many flags in 2020; Afrikaans is balanced.**

On the very right we see the number of active admins in the past August. The percentage is the proportion of 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.

French admins are 1.75%, while admins of Swahili are 13.70% of the active editors. This means that they have to carry a lower load.

However, we need to take into account that the group of admins are 10 in total. It seems reasonable to have a minimal number of admins for safety reasons, even though this means the proportion of higher.
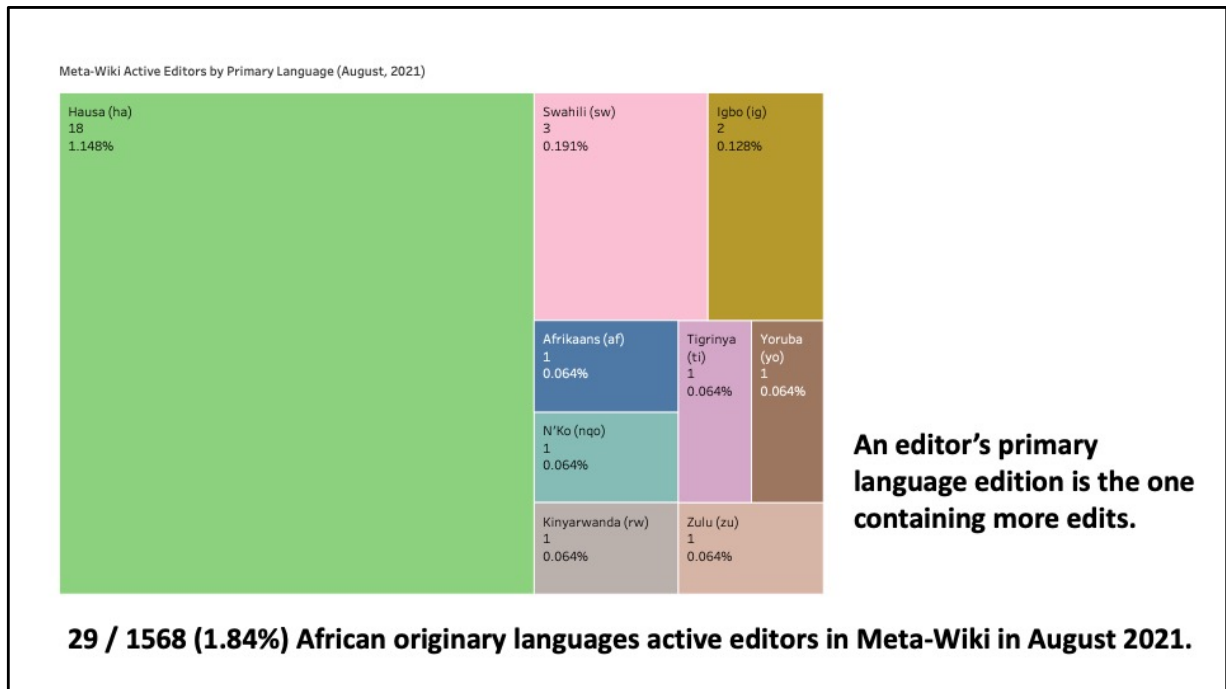
The sixth and last vital sign is global. **Communities participation in the "global community" is key to make their voice heard and learn from others.**

Meta-Wiki Active Editors by Primary Language (August, 2021)

Hausa (ha) 18 1.148%	Swahili (sw) 3 0.191%	Igbo (ig) 2 0.128%
Afrikaans (af) 1 0.064%	Tigrinya (ti) 1 0.064%	Yoruba (yo) 1 0.064%
N'Ko (nqo) 1 0.064%		
Kinyarwanda (rw) 1 0.064%	Zulu (zu) 1 0.064%	

An editor's primary language edition is the one containing more edits.

29 / 1568 (1.84%) African originary languages active editors in Meta-Wiki in August 2021.

In this graph we see the active editors of Meta-Wiki whose primary language is an African originary language.

An editor's 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 1.86% whose primary language is an African originary Wikipedia. 29 Wikipedians.

English is the primary language of more than a third of the active editors in Meta-Wiki.

August 2021 active editors
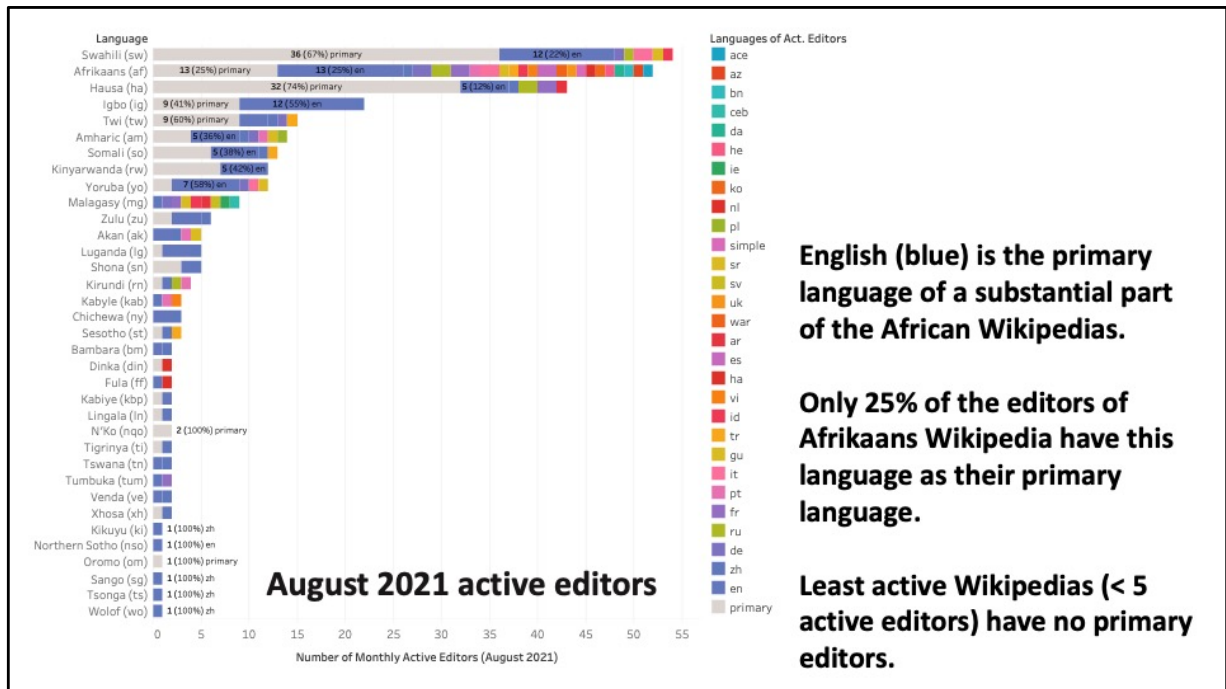
In the following and last graph we can see the number of active editors in African Wikipedia language editions in August 2021, it is very similar to the graph I showed you before, and the percentage of editors who edit primarily that language edition (primary editors).

English (blue) is the primary language of a substantial part of the African Wikipedias.

It is surprising to me that only 25% of the editors of Afrikaans Wikipedia have this language as their primary language.
Only 41% of the active editors in Igbo (9).

Least active Wikipedias (< 5 active editors) have no primary editors.
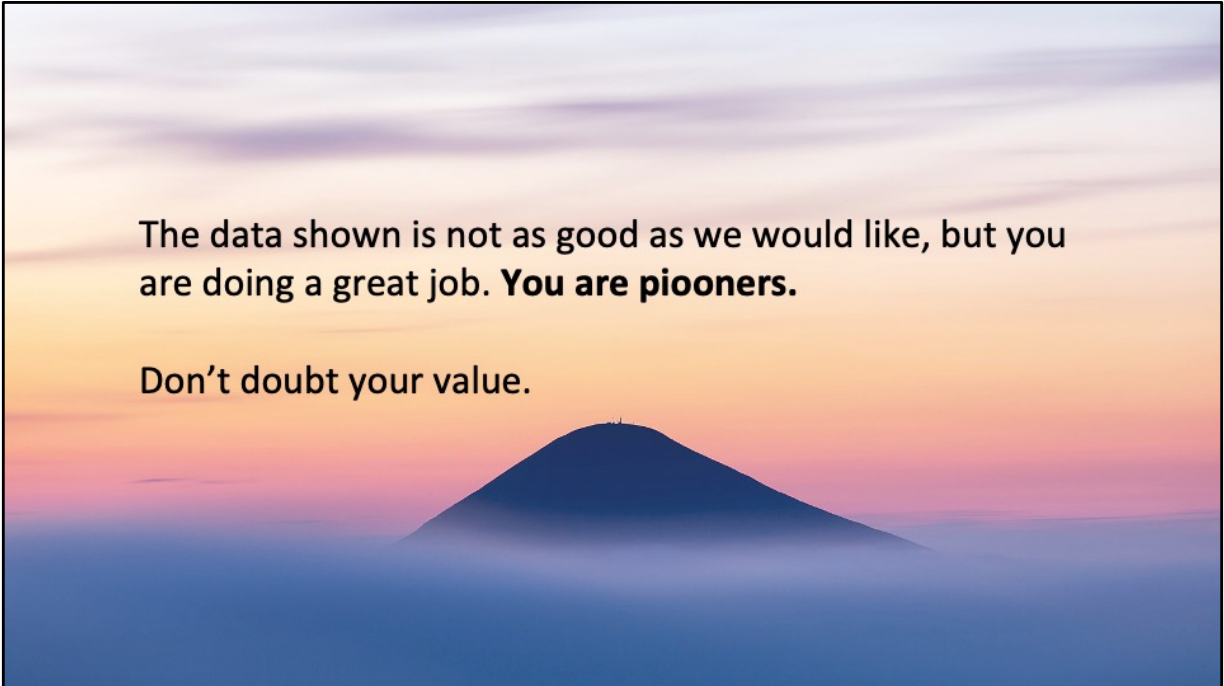
Conclusions

Conclusions.

So, given the results, the main conclusion for African originary languages is that you need to focus on retention, continuity and balance.

What can you do as **African communities**?

Main objective: consolidate communities

- Good thing: there are registrations.
- Bad thing: retention rate is twice lower than large language editions.
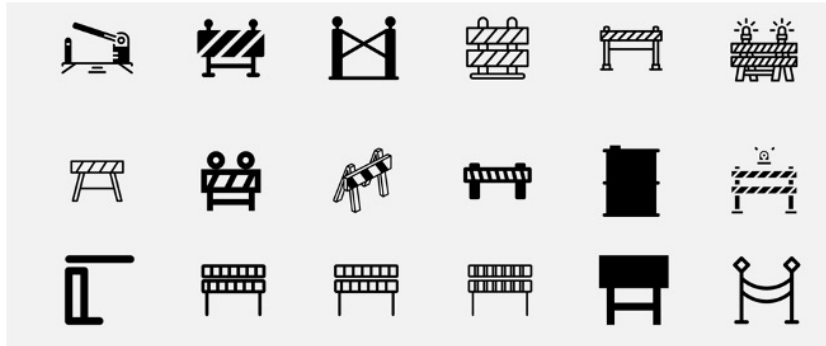
Help active editors identify with roles (technical, admins, etc.). This will strengthen the community.

The data shown is not as good as we would like, but you are doing a great job. **You are piooners.**

Don't doubt your value.

The barriers we need to tear down are economical, technical, connectivity

This is what **knowledge equity** from the strategic direction is about.

Let's focus on the <u>internal ones</u>. You should volunteer for all changes **(piooner):**

- Easier interface (hide the wikitext, visual use of templates, etc.).
- Mobile first.
- …

The potential increase of retention is very high.
Risk of rejection among the current community members is low.

Let's be ready for when external barriers fall.

There are <u>over a hundred</u> Wikipedia language editions with communities below 20 active editors per month.

Joining efforts to ask for a big leap in terms of new interfaces is possibly the best way to go to improve the retention rate and consolidate the communities (ping Growth Team, WMF).

You are not alone.

There are <u>over a hundred</u> Wikipedia language editions with communities below 20 active editors per month.

Joining efforts to ask for a big leap in terms of new interfaces is possibly the best way to go to improve the retention rate and consolidate the communities (ping Growth Team, WMF).

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