Hi everyone!

Welcome to this presentation. It is called "Measuring Central and Eastern Europe Wikipedias Growth and Renewal" with the subtitle question:

https://meta.wikimedia.org/wiki/Wikimedia_CEE_Online_Meeting_2021/Programme/Submissions/Measuring_Central_and_Eastern_Europe_Wikipedias_Growth_and_Renewal

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

Community Health

My name is Marc Miquel and I'm a Wikimedia researcher. I joined the Movement in 2011. Been doing several things on diversity.

Today 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.

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. Is it the same for CEE?

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 does it occur also for other languages? In particular, the CEE languages?

Top 15 CEE languages in
number of active editors*

* At least 5 edits per month
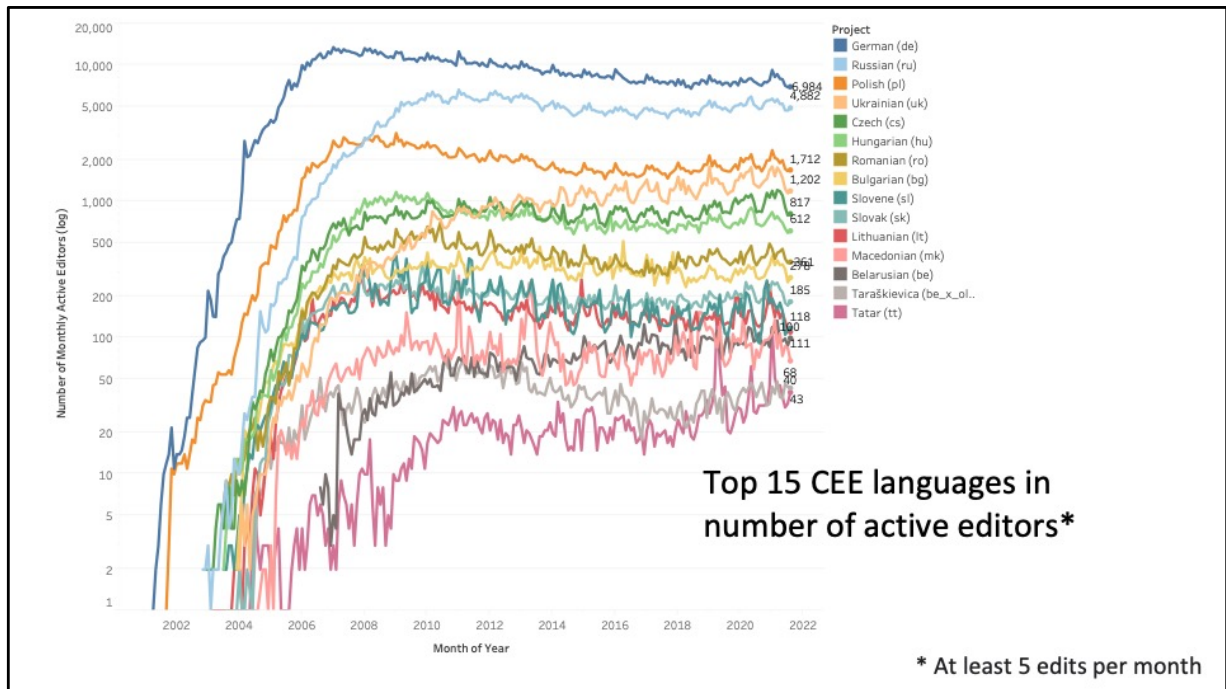
So how does the curve of active editors look like for other languages?

These are the top 15 CEE languages in number of active editors. For most of them, we see similar curves of growth and stall to English Wikipedia.

Few of them like Ukrainian (light orange) or Tatar (pink) show a slight increase in the past five years.

32 other CEE languages editions.
Mostly below 10 active editors* per month.
(August 2021)

Question in the air:
how many "active editors" does it take to have a consolidated community?

* At least 5 edits per month
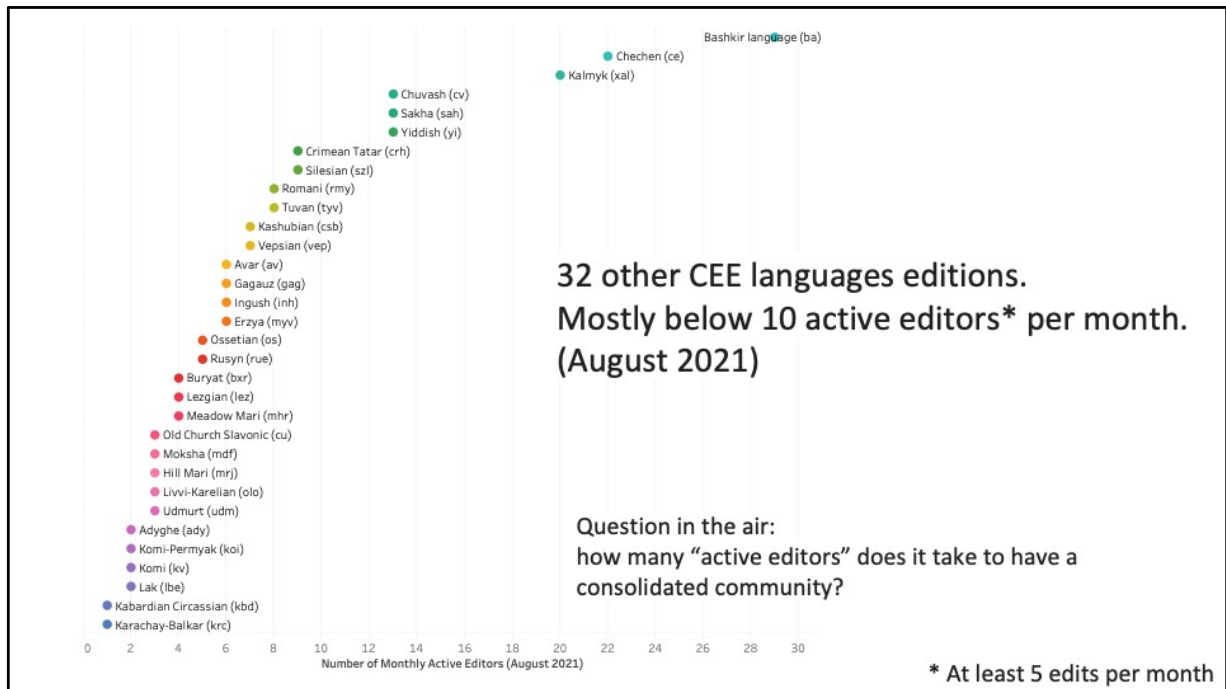
We must acknowledge however that there are 32 other CEE language editions. Mostly below 10 active editors per month. (August 2021)

Question in the air:
how many "active editors" does it take to have a consolidated community?

Because there are more than 100 Wikipedias below 20 active editors per month.

In this talk, we'll focus on CEE 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 **CEE Wikipedias**

Is there growth or renewal of editors?
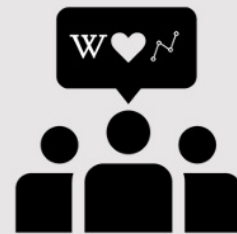
We want to answer this question on growth and renewal.

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.

We are creating a "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: 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.

Wikimedia Poland contacted us and showed us three specific profiles: "Advanced Editor", "Project Maintenance", and "TechWizard". As you can imagine, these are very related to the active editors we will analyze and also more specifically to the specialist profiles.

So, in a way this framework has been inspired by the need to monitor these profiles and help communities like the Polish. Possibly, we'll pay a little more attention to this language edition in particular.
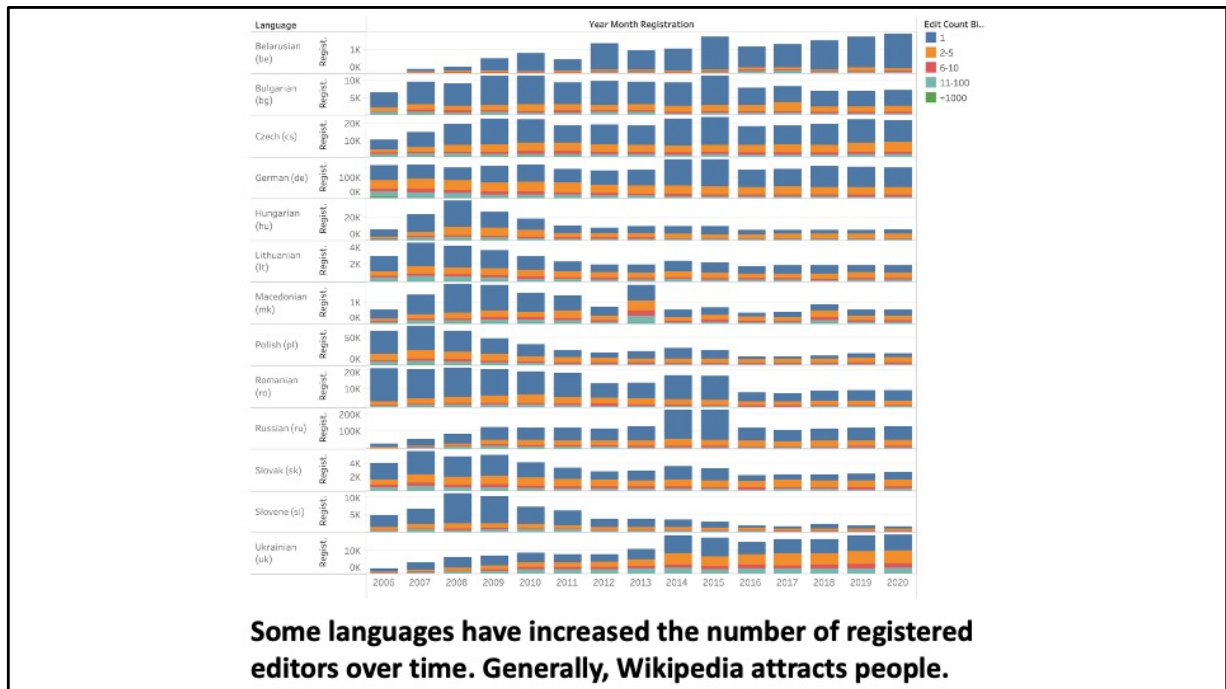
I. Retention

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.**

Some languages have increased the number of registered editors over time. Generally, Wikipedia attracts people.

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.

In this graph we see the number of editors by the year in which they made their first edit and by their total number of edits (binned in 1, 2-5, 6-10, 11-100, 101-500, 501-1000, and 1000+ edits). This graph tells us about the interest to become new editors, which is stable and even growing over the years.

Retention rate (%) has been going down in all languages. In ten years, it divided by 2.
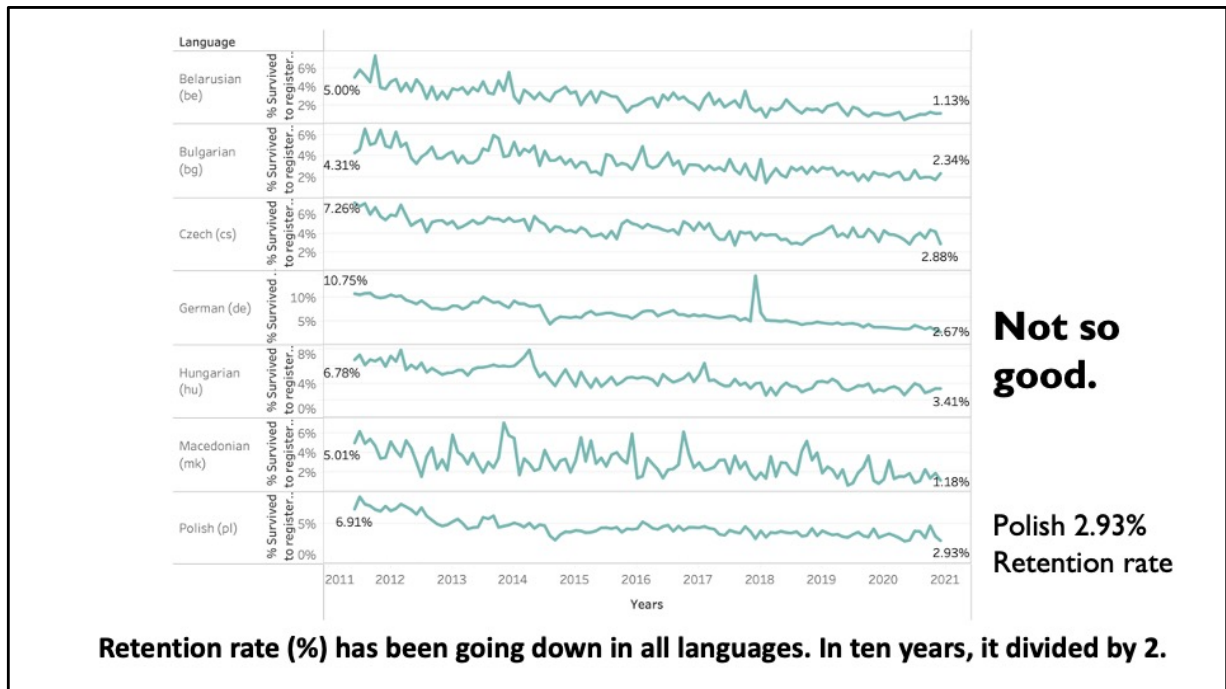
What is not so good is…

Instead, the retention rate 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.

In the following graph, we see the retention rate, which is several times smaller than what it was ten years ago. This means that while new editors get registered and do their first edit, they prefer not editing again. In particular, the Polish Wikipedia does not experiment the biggest decrease (see German or Belarusian), but the percentage is clearly going down.

2. Continuity

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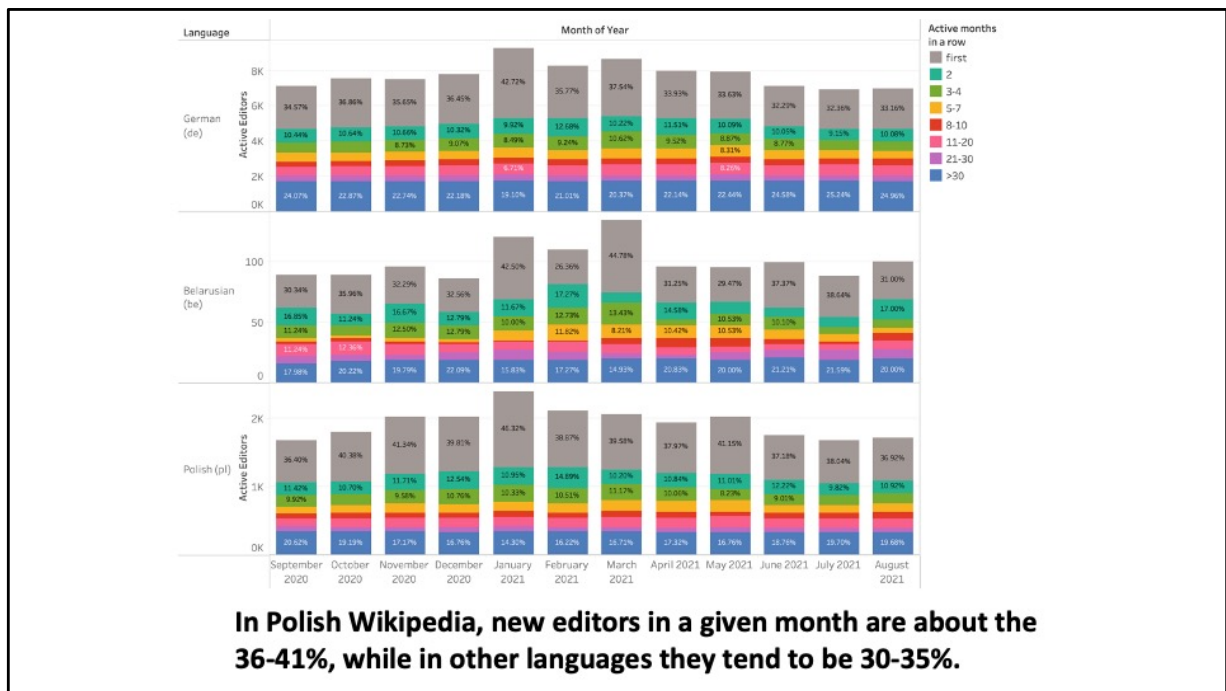
The second vital sign is continuity. **Community continuity is the persistence of active editors as well as the succession of groups of editors over time.**

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 Polish Wikipedia, new editors in a given month are about the 36-41%, while in other languages they tend to be 30-35%.

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

In Polish Wikipedia, fresh editors in a given month are about the 35-40% (in grey), while in others like German, is around 33-35%. This means that Polish Wikipedia is less volatile than German. Polish is engaging fresh editors every month, but the active community is mostly composed by editors who continue one month after another.

Even though they are not present in the graph, younger or less settled communities sometimes have a percentage of 60-70% of fresh editors. This may be seen positively as long as the overall community is growing. However, since for some of them it is not the case, it only indicates their difficulties in consolidating stable communities.

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 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 editor

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.

CEE Wikipedia "very active editors" (100 ed.) have been renewed over time.
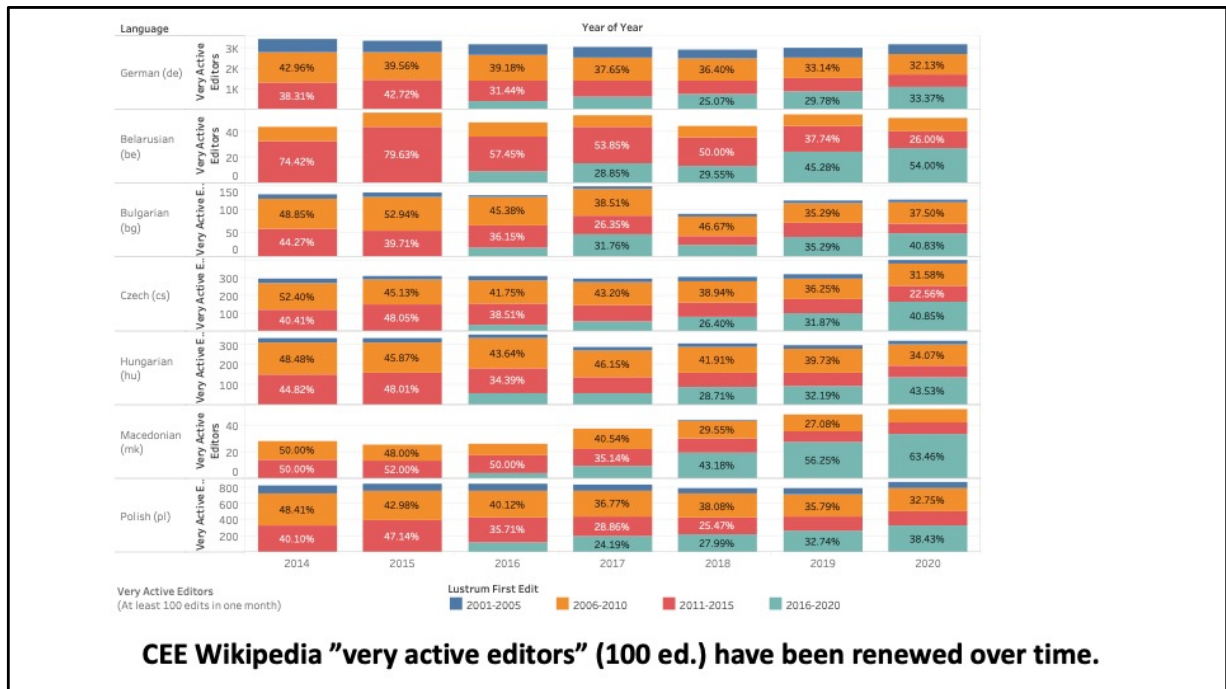
In the following graph, we see for seven CEE language editions, the composition of the "very active editors" every year by lustrum of first edit (2001-2006, 2006-2010, 2011-2015, 2016-2020, and 2021-2025.) We could say that they are "generations".

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

Very active editors account for 80% of the edits made by humans every month. So, we could say that they are the group of very valuable editors. This reaffirms our choice of not setting extra requirement to select the group of experienced editors.

The graph shows this third Vital sign we called "Balance". It gives account of the "stagnation" in terms of growth. Although it is true that 2020 seems a better year than 2019 in number of editors, maybe due to the Global pandemic and the lock-down that took place in many countries. But more importantly, graph also shows that the renewal of editors is occurring over time. We see that

every year the percentage of editors who started editing during 2016-2020 is growing, mostly at expenses of the 2006-2010 and 2011-2015 generations. The percentage of very active editors who started editing during the 2001-2005 has not varied. In Polish Wikipedia, this is 9.13% in 2020, but 9.14% in 2019, 8.45% in 2018, and 10.18% in 2017. We could possibly say that the "founder-factor" ensures more engagement.

Polish Wikipedia does not present percentages that are very different from the other CEE Wikipedia language editions. For example, compared to Macedonian and Belarusian, the proportions between the different generations is more balanced, which we would intuitively say to be better for the community. As it may not be desirable that the "productivity" relies too much on an older generation, but neither that it would depend mostly on the last one. The active communities (or at least, the number of very active editors) do not grow, but at least, they renew over time.

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 "Technical Editors"
(at least, 100 edits in one month in tech. namespaces: Templates and Mediawiki).

Lustrum First Edit: 2001-2005, 2006-2010, 2011-2015, 2016-2020

"Very active tech. editors" are mostly from 2006-2010.
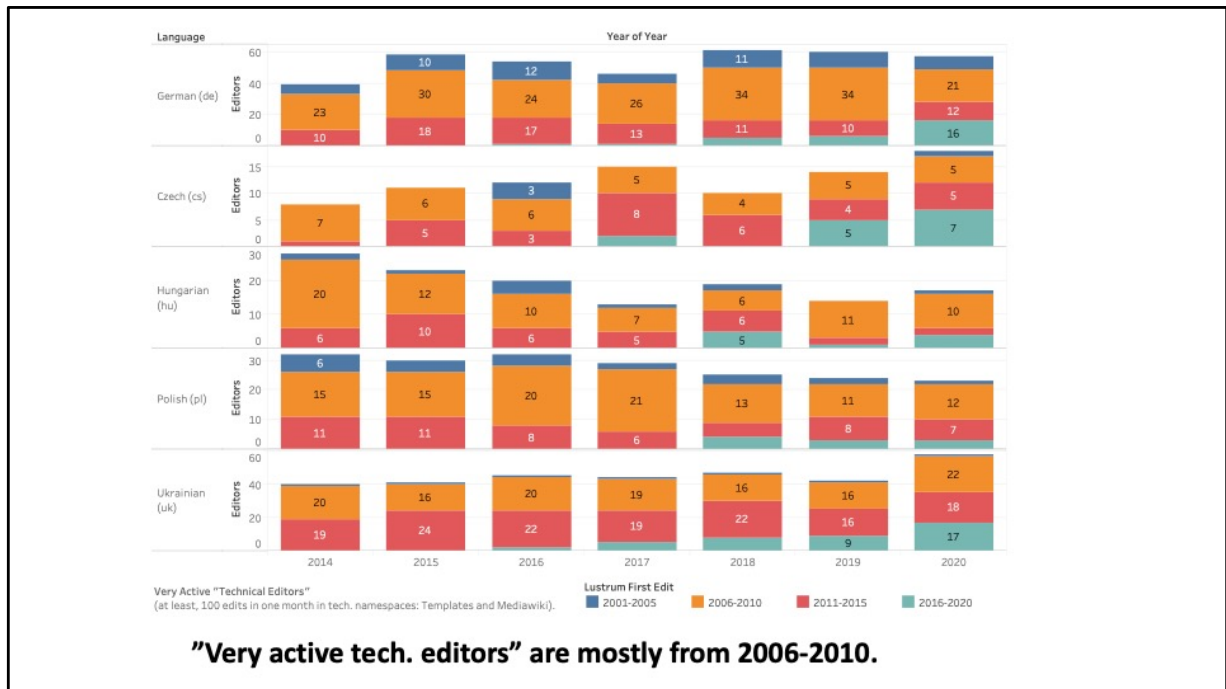
Similarly as with the previous measurement, here we focus on the very active technical editors. Editors who have done more than 100 edits in one month in technical namespaces (that is templates and mediawiki namespaces).

In the graph, we see the number of "Very active technical contributors". For Polish Wikipedia, they are 25 in 2020, very few, especially when compared with the overall group of very active editors (864 in 2020) and the active editors any month this year (1341 in August 2021). Community building on this group of contributors is highly encouraged, given the very few number of editors. In Polish Wikipedia, the group of editors who started editing in 2016-2020 is even decreasing (4 in 2018 and 3 in 2019 and 2020).

Polish Wikipedia technical editors are much less balanced than the overall group of very active editors. The majority is from 2006-2010: 12 editors in 2020, which account for a 52.17% of the very active technical editors that year. The prominence of older generations is also visible in other languages like Hungarian and Belarusian, and other languages not in this graph. In the figure, we see that Czech, German and Ukrainian seems to renew more effectively.

19

Similarly as with the previous analysis, we would encouraged that the groups are balanced, so that they can pass the knowledge/projects from one generation to another.

The "very active coordinator editors" have renewed better.
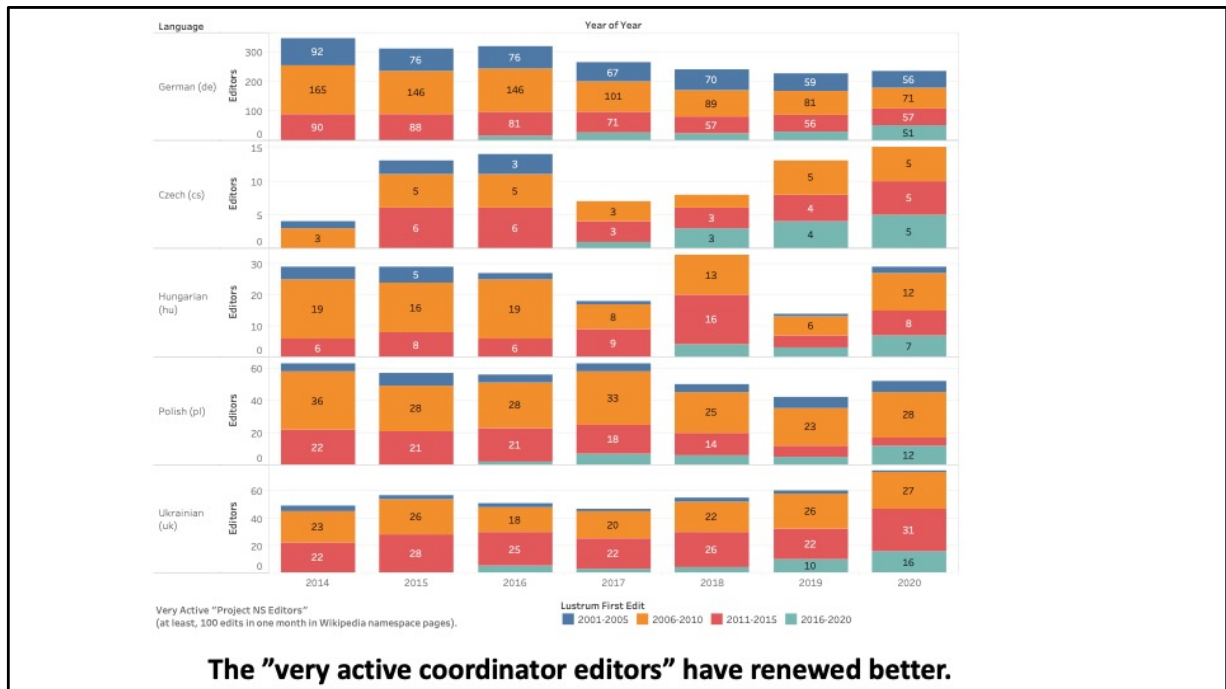
In this other case, we do exactly the same analysis but for those very active editors in the Wikipedia namespace (this is, Wikiprojects, Village Pump, among others). The number of editors is higher.

In the graph, we see the number of "Very active project coordinators". For Polish Wikipedia, they are 52 in 2020, many more compared with the technical (27), but many less than the overall group of very active editors or active editors. Differently than with very active technical editors, we see that there is more renewal (the generation 2016-2020 is actually growing). However, the Polish Wikipedia is not as balanced as Wikipedias such as Ukrainian or Czech in terms of different generations of project coordinators.

It is interesting to see that some languages like German or Polish show a decrease in overall number of project coordinators, while others like Ukrainian and Czech are growing. This may be due to many factors. Nonetheless, since the project coordinators are editors contributing to Wikipedia namespace, we would say that this metric relates to how open is a Wikipedia in terms of inviting newer editors to the very centre of activity.

5. Admins

Admins have special 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.
The group has not been renewed. New admins should be encouraged.
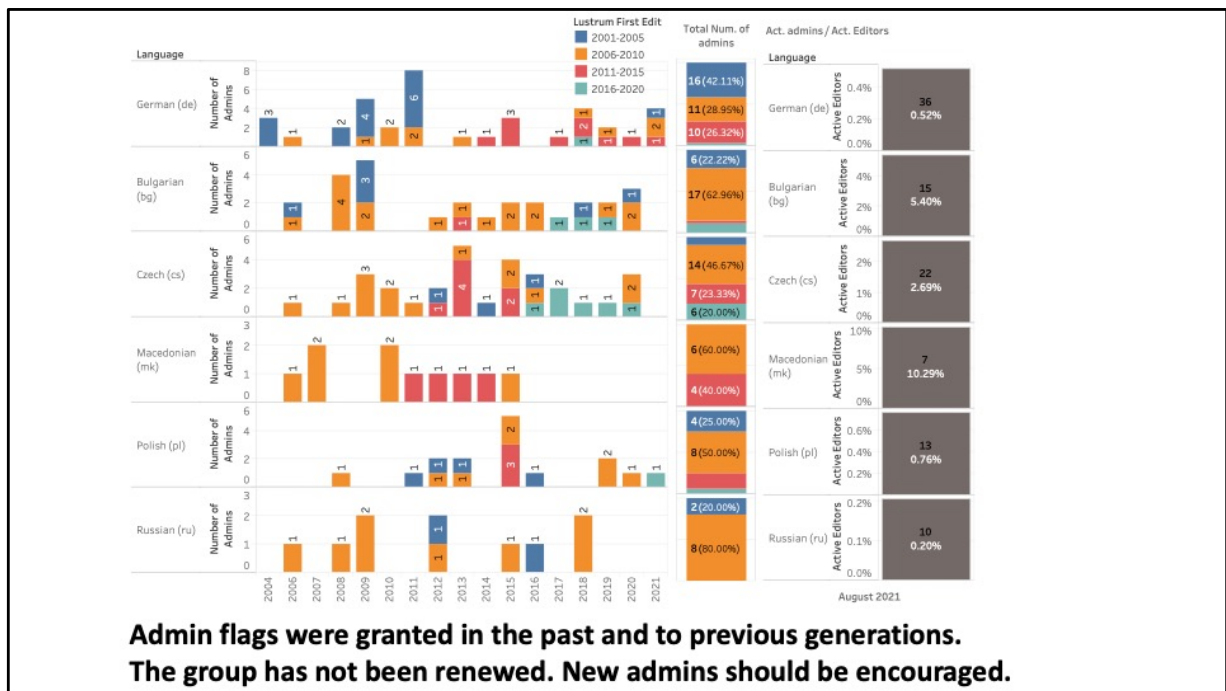
In this complete graph, on the left subgrap, we see the admins flags granted by year and the color represents the generation they belong to. Most flags were granted in 2011 to 2016 and to previous generations. We almost do not see flags from editors 2016-2020 but in a couple of languages (Bulgarian and Czech). In the middle graph, we see the total number of admins: Polish has 16 admins.

No flags have been granted in the Macedonian Wikipedia after 2015, and in the Russian Wikipedia after 2018.

On the very right, we see the number of active admins in the past August 2021. The percentage is the ratio between number of active admins by the number of active editors. This percentage varies according to the language, but it tells the "load" each admin is carrying, given that their task is to patrol the production and act when necessary. The lower the percentage, the higher the load they take.

In languages like German, active admins are a 0.52% of the active community (a minimum of 5 edits per month). In Polish we see there are 13 active admins, which represent a 0.76% of the active editors. Other languages have higher proportions: Belarusian 1%, Bulgarian 5.40% and Macedonian 10.29%, even though the group is not very large. We should say that for smaller languages, the group of admins should at least be big enough not to risk its continuity.

Polish Wikipedia does not present an important risk given the number of active admins. However, given the low percentage in relation to active editors and also the lack of new flags granted (only 4 of the 16 were granted in the past five years), it would be advisable that new flags were granted in order to ensure renewal.
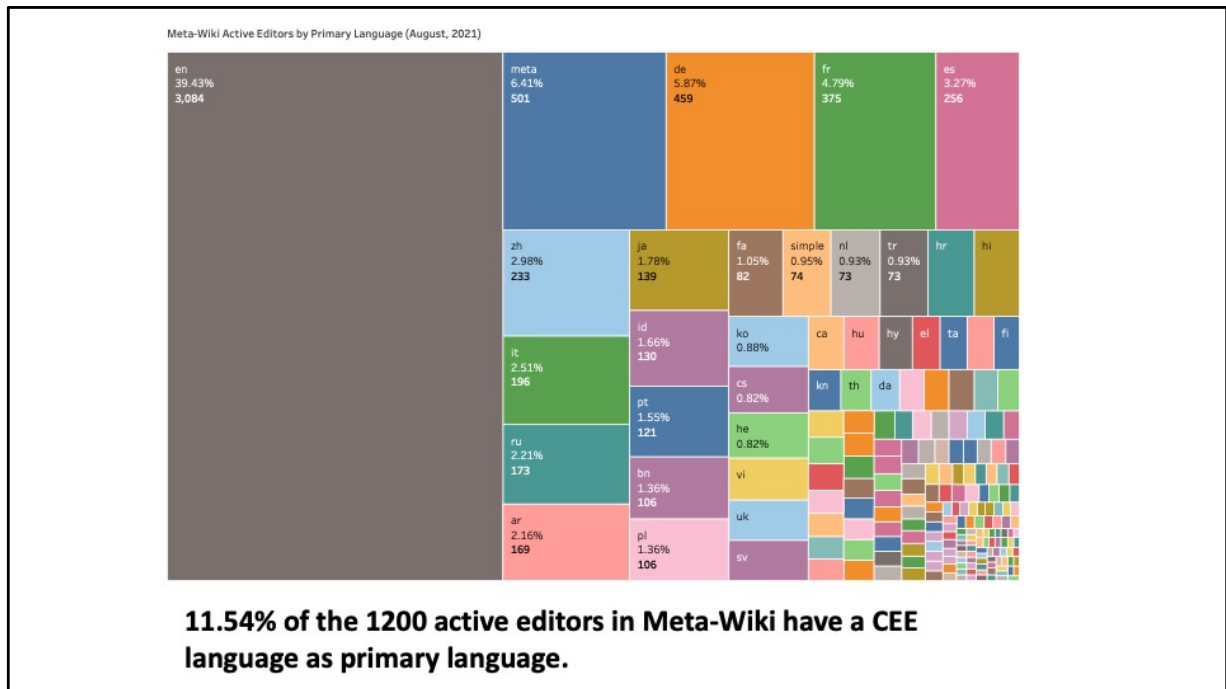
There is a chance to expand to newer admins.

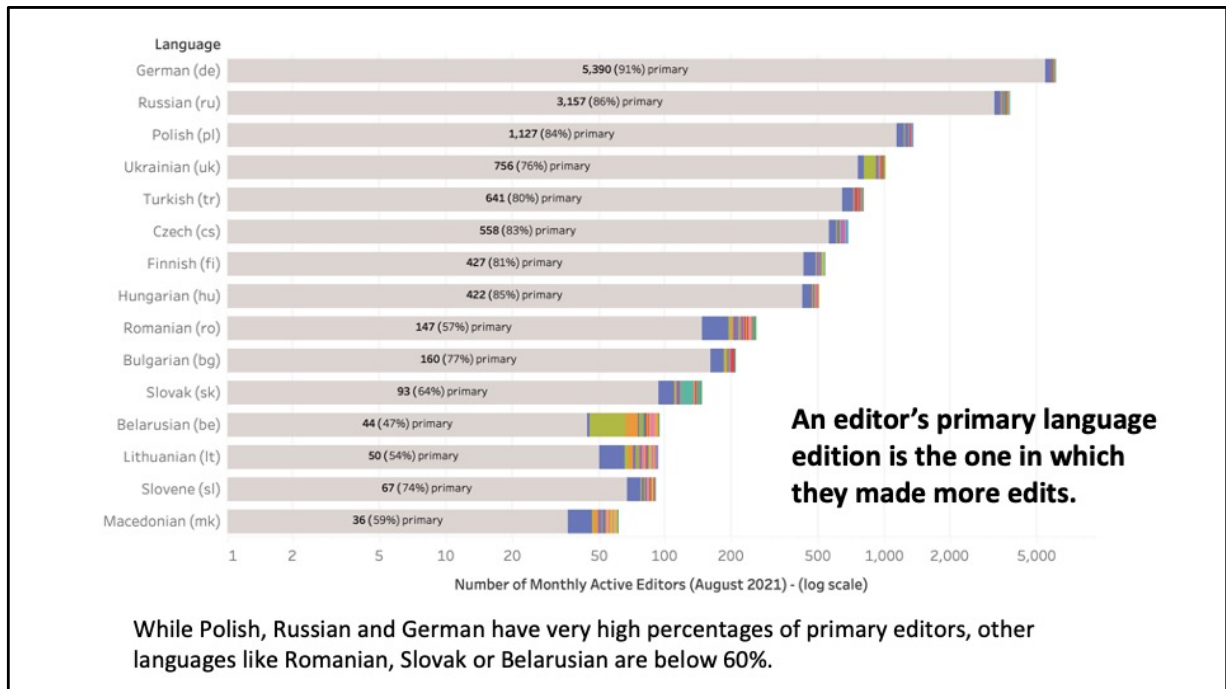The sixth and last vital sign is global. **Communities participation in the "global community" is key to make their voice heard and learn from others.**

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

11.54% of the 1200 active editors in Meta-Wiki have a CEE language as primary language.

In this graph we see the number of active editors in meta-wiki in August 2021 (usually between 1,200 and 1,900 in 2021) by their primary language. An editor's primary language edition is the one in which they made 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 see that only a 6.41% of editors are "local" to that project, which makes sense because of its purpose. Meta-wiki is a coordination space across languages. The Meta-wiki active editors who have Polish Wikipedia as their primary language are 106 (1.36%) of all editors. Polish Wikipedia compares to Portuguese or Bengali. The total number of active editors in Polish Wikipedia is around 1,200, which means that a little less than 10% edit Meta-wiki, which is a reasonable proportion.

Language		
German (de)	5,390 (91%) primary	
Russian (ru)	3,157 (86%) primary	
Polish (pl)	1,127 (84%) primary	
Ukrainian (uk)	756 (76%) primary	
Turkish (tr)	641 (80%) primary	
Czech (cs)	558 (83%) primary	
Finnish (fi)	427 (81%) primary	
Hungarian (hu)	422 (85%) primary	
Romanian (ro)	147 (57%) primary	
Bulgarian (bg)	160 (77%) primary	
Slovak (sk)	93 (64%) primary	
Belarusian (be)	44 (47%) primary	
Lithuanian (lt)	50 (54%) primary	
Slovene (sl)	67 (74%) primary	
Macedonian (mk)	36 (59%) primary	

An editor's primary language edition is the one in which they made more edits.

Number of Monthly Active Editors (August 2021) - (log scale)

While Polish, Russian and German have very high percentages of primary editors, other languages like Romanian, Slovak or Belarusian are below 60%.

In the following and last graph we can see the number of active editors in CEE Wikipedia language editions in August 2021, and the percentage of editors who are from the same language (primary). While Polish, Russian and German have very high percentages of primary editors, other languages like Romanian, Slovak or Belarusian are below 60%. Even though the non-primary editors surely make some useful edits, one could say that in terms of developing Wikiprojects and taking responsibility, the *real* community tends to be the one composed by primary editors.

The chart shows a horizontal bar chart titled "Number of Monthly Active Editors (August 2021)" with languages listed vertically. Selected readable values include:

Language

- German (de): 5,390 (91%) primary
- Russian (ru): 3,157 (86%) primary
- Polish (pl): 1,127 (84%) primary
- Ukrainian (uk): 756 (76%) primary
- Turkish (tr): 641 (80%) primary
- Czech (cs): 558 (83%) primary
- Hungarian (hu): 422 (85%) primary
- Romanian (ro): 147 (57%) primary
- Bulgarian (bg): 160 (77%) primary
- Slovak (sk): 93 (64%) primary
- Belarusian (be): 44 (47%) primary
- Lithuanian (lt): 50 (54%) primary
- Slovene (sl): 67 (74%) primary
- Macedonian (mk): 36 (59%) primary
- Taraškievica (be_x_old): 17 (40%) primary
- Tatar (tt): 16 (48%) primary
- Bashkir language (ba): 17 (63%) primary
- Chechen (ce): 4 (22%) primary
- Yiddish (yi): 6 (50%) primary
- Chuvash (cv): 4 (36%) primary
- Sakha (sah): 2 (20%) en
- Kalmyk (xal)
- Crimean Tatar (crh): 2 (22%) en 3 (33%) ru
- Silesian (szl)
- Tuvan (tyv): 4 (50%) primary
- Vepsian (vep)
- Avar (av): 2 (33%) en
- Erzya (myv): 1 (17%) en 2 (33%) ru
- Romani (rmy): 2 (33%) en
- Kashubian (csb): 2 (40%) pl
- Gagauz (gag): 1 (20%) ow
- Ingush (inh)
- Ossetian (os): 2 (40%) en
- Lezgian (lez): 2 (50%) ru
- Rusyn (rue)
- Buryat (bxr): 1 (33%) ru
- Old Church Slavonic (cu): 1 (33%) en
- Udmurt (udm): 1 (33%) zh
- Adyghe (ady): 1 (50%) tt
- Komi-Permyak (koi): 1 (50%) tt
- Komi (kv): 1 (50%) tt
- Lak (lbe): 1 (50%) zh
- Moksha (mdf): 1 (50%) tt
- Meadow Mari (mhr): 1 (50%) tt
- Hill Mari (mrj): 1 (50%) tt
- Livvi-Karelian (olo): 1 (50%) tt
- Kabardian Circassian (kbd): 1 (100%) fr
- Karachay-Balkar (krc): 1 (100%) m

We see that those CEE languages below 10 active editors, in their composition show an important part of 'foreign editors' (non-primary).

English (dark blue) and Russian (dark green) tend to be the primary languages of these foreign editors.

Here with the other languages. We see that those below 10 active editors, in their composition show an important part of 'foreign editors'.

Especially in small Wikipedias, some people may edit more a large Wikipedia (e.g. Russian or English), because there is more content, more audience etc, but their primary language may typically be the one of their "secondary Wikipedia", e.g. Belorussian.

Conclusions

6 Vital Signs:

1. Retention
2. Continuity
3. Balance

4. Specialists
5. Admins
6. Global

Quick summary, there are more than 45 CEE Wikipedia language editions, but only 15 of them have more than 30 active editors. 30 struggle to have 10 active editors per month, so we could think that they are not consolidated communities, as the definition of active editor is very lax.

For the top 15, we applied the Vital Signs framework to understand the degree of growth and renewal of each community.

Even though languages have different patterns of registered editors (most of them increasing or stable), the retention rate is decreasing, which is quite worrying.

There is continuity between editors on a monthly basis, and the proportion of long-term engaged editors (more than an year editing on a monthly basis) is quite large (a third). The proportion of fresh editors is also a third.

The communities are quite balanced in terms of the different generations of editors which compose them, which means that generally even though there is not growth at least there is renewal.

However, 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. This means that possibly we are not "open" enough. Even though this may be few editors, the message of closedness it delivers to everyone is possibly not helpful.

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. CEE 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.

In case you do not have questions, we invite you to think about what actions you could do in your affiliates to improve on each of the Vital Signs.

1. Retention
2. Continuity
3. Balance
4. Specialists
5. Admins
6. Global

If you want to share your thoughts, we would love to listen to you.

WIKIMEDIA CEE
ONLINE MEETING
2021