Future of Language Incubation

Srishti Sethi and Caroline Myrick
Wikimedia Foundation

WIKIMEDIA LANGUAGE CONFERENCE
CELTIC KNOT '24

This presentation will cover...

- Context: language gap, Incubator

- Incubation Tools, Journey, and Challenges

- Recommendations for Future Incubation

Language Gap

- **334 languages** with at least one hosted project (Sep. 2024)

- **>6,600 living languages** used by millions of people (per Ethnologue, LSA) for which there is no hosted Wikipedia, Wiktionary, Wikisource, Wikibooks, etc.

- Directly relates to fulfilling **our vision** that every single human being can freely share in the sum of all knowledge.
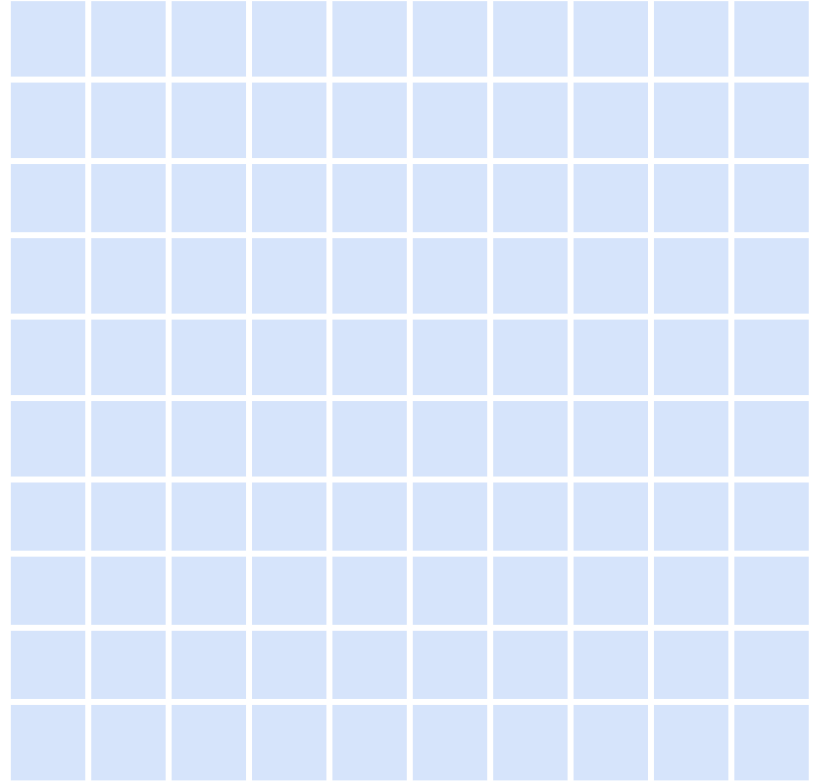
Wikimedia Incubator

Since **2006**, a centralized platform for incubating new language versions of **Wikipedia, Wikibooks, Wikinews, Wikiquote, Wiktionary** & **Wikivoyage**.
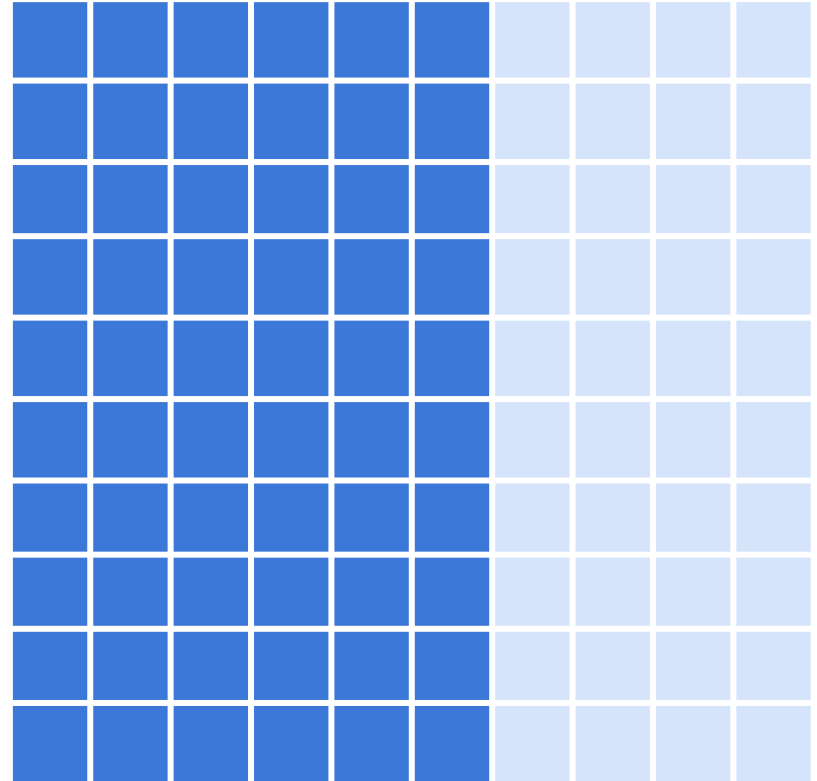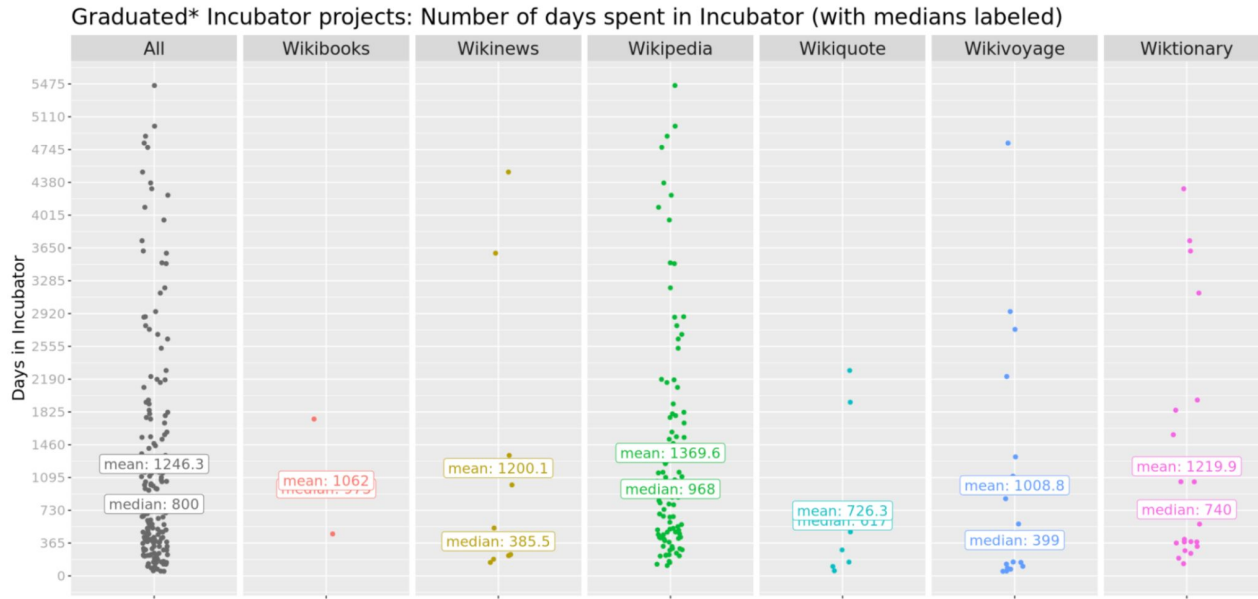
Currently >1000 languages have at least 1 test projects in the Incubator.*

~600 languages have **active** and/or **substantial** test projects in the Incubator.

~400 of these languages have __no__ hosted projects

e.g.,
- *American Sign Language*
- *Ibibio*
- *Marwari*

e.g.,
- *Tiwi*
- *Komering*
- *Western Baluchi*
- *Yucatec Maya*

Time-to-graduation varies widely for graduated projects, from 51 days to 15 years

Graduated* Incubator projects: Number of days spent in Incubator (with medians labeled)

| All | Wikibooks | Wikinews | Wikipedia | Wikiquote | Wikivoyage | Wiktionary |

mean: 1246.3
median: 800

mean: 1062
median: 575

mean: 1200.1
median: 385.5

mean: 1369.6
median: 968

mean: 726.3
median: 617

mean: 1008.8
median: 399

mean: 1219.9
median: 740

Days in Incubator

July 2024

* 'Graduated' refers to projects that have left the Incubator test space, per Language Committee approval, and received their own domain hosted by the Wikimedia Foundation

* Excludes projects that graduated in 2006, the year the Incubator was created.

Time-in-Incubator varies widely for current Incubator projects, from 12 days to 18 years

Current Incubator projects: Number of days in Incubator (with medians labeled)
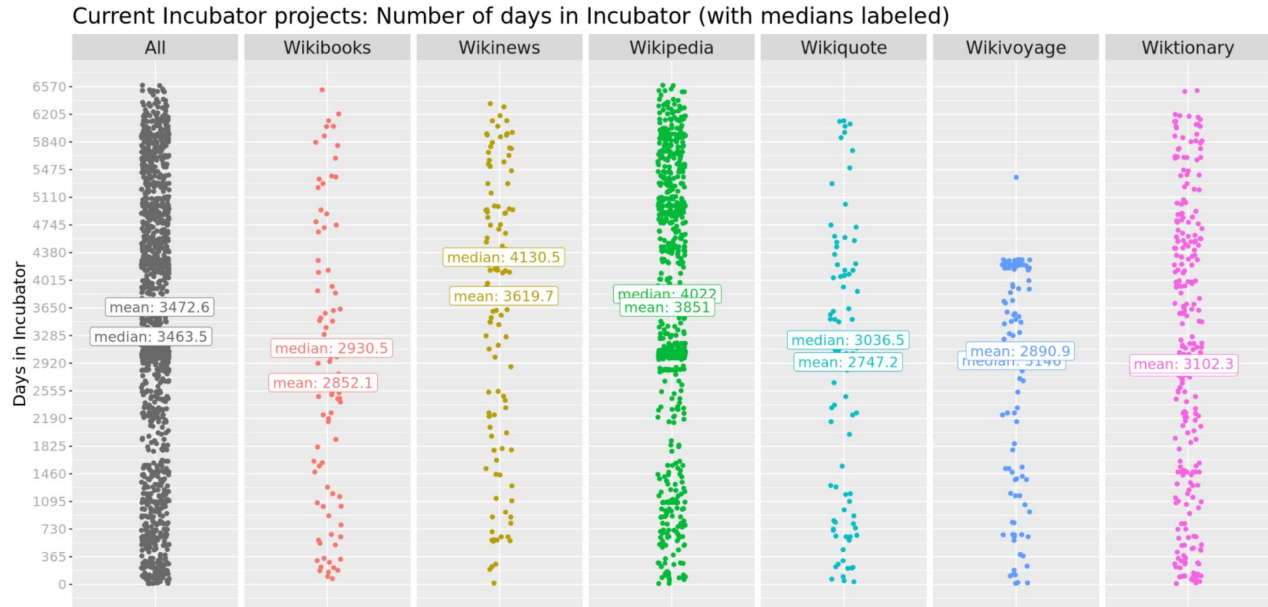
More <mark>test projects[1]</mark> than <mark>hosted projects</mark>

test	hosted	closed
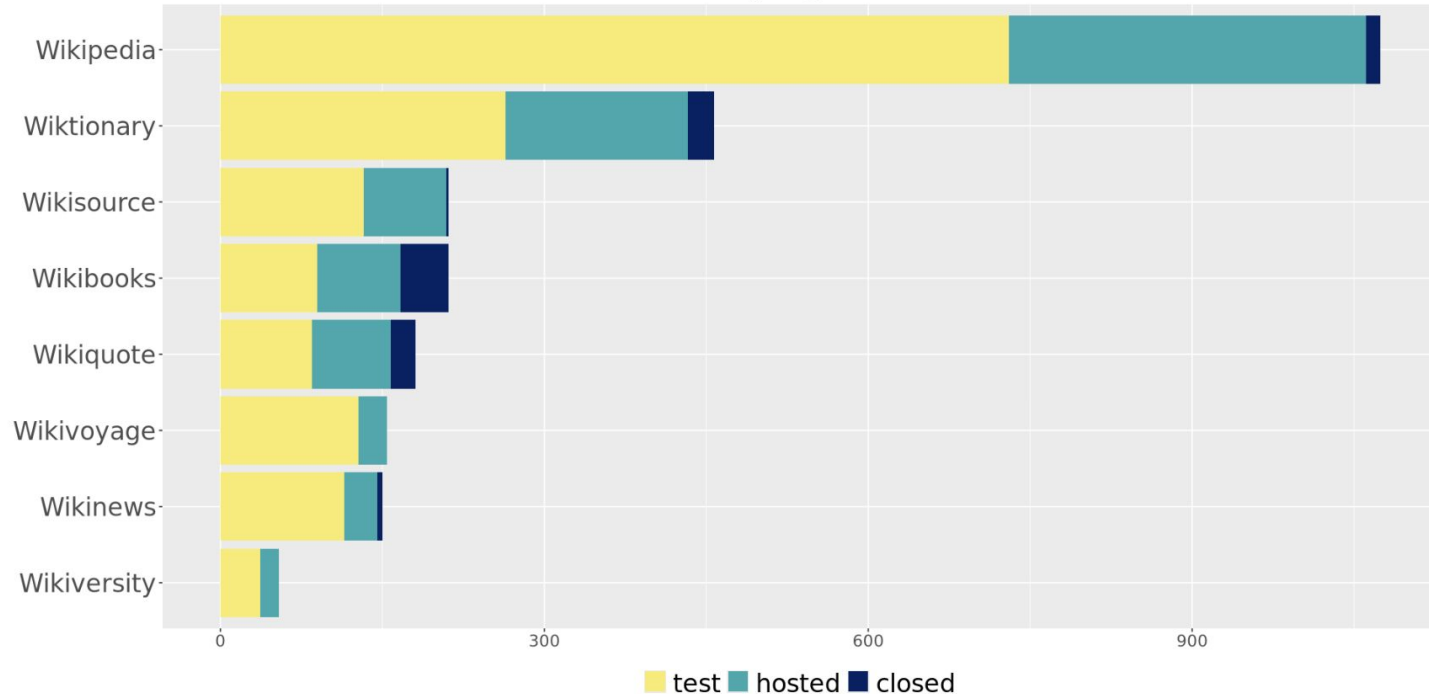
June 2024

[1] Test projects in the Wikimedia Incubator, Wikiversity Beta, and Multilingual Wikisource
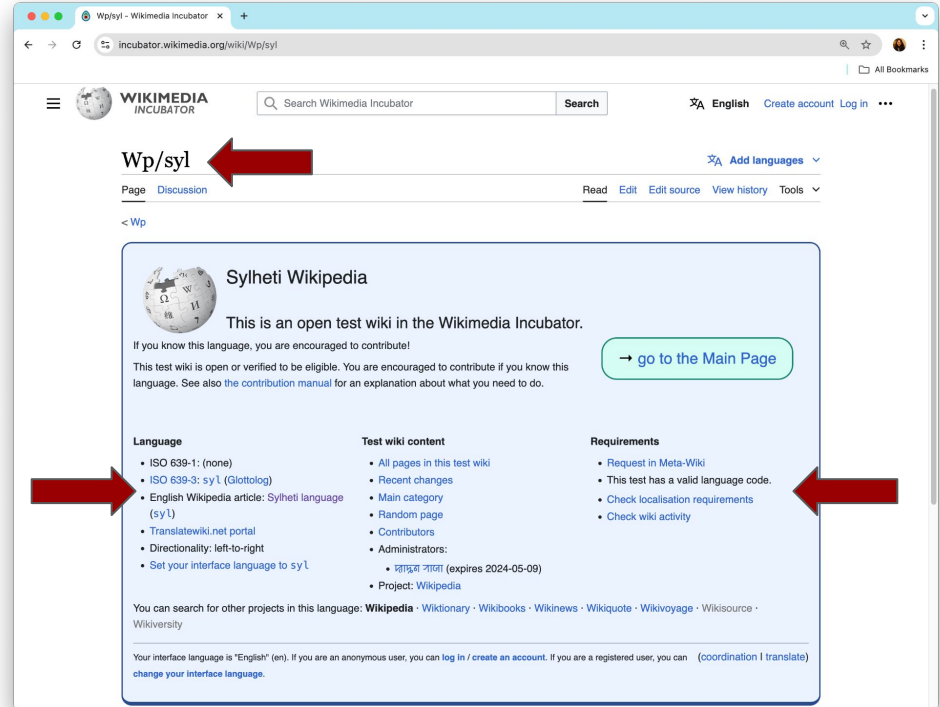
Wikimedia Incubator

Since **2006**, a centralized platform for incubating new language versions of **Wikipedia, Wikibooks, Wikinews, Wiktionary** and **Wikivoyage**.

https://incubator.wikimedia.org/
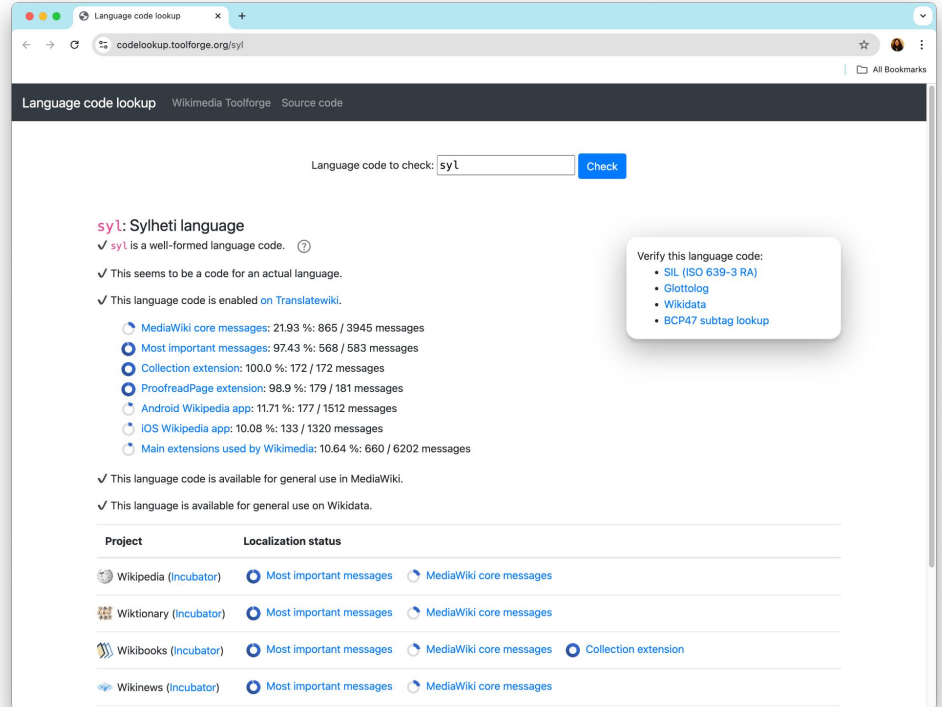
Example Landing Page

Language details, test wiki content, localization requirements, activity and status for **Sylheti Wikipedia**.

https://incubator.wikimedia.org/wiki/Wp/syl

Language Status Tools

Code Lookup verifies language code use on MediaWiki sites, localization status for message groups on Translatewiki.net, and existing Wikimedia sites in that language.

https://codelookup.toolforge.org/

Language Analysis Tools

Catanalysis performs test project analysis such as edits, new pages, editors, bytes added per month, used by Language Committee.

https://meta.toolforge.org/catanalysis

Incubation Journey

Verify principles and requisites → Request creation on Meta-Wiki → Language committee approves language creation → Content development in Incubator → Interface messages translation on Translatewiki.net

Incubator content imported to site ← Developers create the site ← Language committee approves site creation

Incubation Challenges

❖ **Before**: Labor-intensive manual processes.

❖ **During**: Incubator has technical limitations; offers a poor editing experience; lacks modern features (**Content Translation**, **Wikidata**, **Newcomers' Homepage, Machine Translation**); lacks support for keyboards, dictionaries, etc.

❖ **After**: Manual site setup involving multiple parties; may take days or weeks.

Language Committee

❖ Sets the **Language Proposal Policy** and is in charge of its implementation

❖ Formed in 2006 as a subcommittee of the Special Projects Committee

❖ Became stand-alone committee in 2012

❖ Currently **13 members** with diverse geographic and linguistic backgrounds

Future of Language Incubation:
A New Experiment

Visit project page:
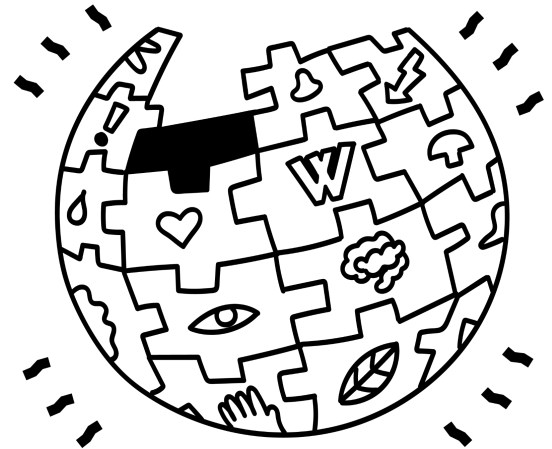
Curious Case of Kadazandusun language Wikipedia

Inactive for over 10 years since launch.

Incubator reintroduced at **Wikimania Singapore**, leading to the 2024 launch through active development.

Wiki Kent Club's meeting. CC BY SA 4.0 User:Jjurieee

Experimental Idea 1

Provide Production Wiki Access to 5 New Languages

Purpose

Evaluate if access to a full-featured wiki (Content Translation, Wikidata support, advanced editing, and search) enhances editor productivity.

Outcome

Assess if this approach is effective for language incubation and if it should be explored further.

Participant Quotes

"We should forget about Incubator completely. And, find another way of starting wiki. Because of the complexities around it, it might take time to improve the technical side of it."

"Editing on Incubator should feel similar to editing on normal wikis, but we are far from achieving this goal."

Language Diversity Hub's Research

Barriers experienced by contributors to small language versions of Wikipedia

A Movement Strategy Implementation Grant Research Report by the Language Diversity Hub

https://meta.wikimedia.org/wiki/File:Barriers_experienced_by_contributors_to_small_language_versions_of_Wikipedia.pdf

Interesting Fact!

Until 2006, wikis were created using an automated system. But, in non-standard ways. To address this issue, the **Incubator** and **Language Committee** were established.

[Intlwiki-l] russian wikipedia

elian intlwiki-l@wikipedia.org
07 Nov 2002 20:36:30 +0100

- Previous message: [Intlwiki-l] Re: [Wikipedia-l] United Wikipedias Forum
- Next message: [Intlwiki-l] russian wikipedia
- **Messages sorted by:** [date] [thread] [subject] [author]

```
Hello,

Brion set up Russian wikipedia at http://ru.wikipedia.org/ with UTF-8
encoding.

Everybody who is speaking Russian please help testing the new software and
getting the Russian Wikipedia running!

greetings,
elian
```

Implementation of the experiment

- **Identify** a set of 5 test wikis in Incubator

- Create the wiki in **production infrastructure** (along with transfer of content from the Incubator) so that the community can quickly move to editing on the production wiki.

- **Monitor content growth** progress of these prioritized wikis on a production wiki vs. Incubator.

- Clearly **mark these wikis as new** to distinguish them from other production wikis during the pilot period.

- **Review status** at the end of a pilot period to ascertain whether any change is needed for the wiki.

Inclusion criteria

- Test wiki is a **Wikipedia**

- Test wiki has at least **2 active editors** in the last 3 months

- Test wiki shows at least **30 edits** in the last 3 months

- Test wiki is either **RTL** or **LTR** language

Selection

❖ Each wiki will be chosen from 4 **data clusters**

❖ Categorized by **3-month activity**:

   ○ Edits or new pages in the content namespace

   ○ Average monthly active editors

   ○ Time spent by language on Incubator
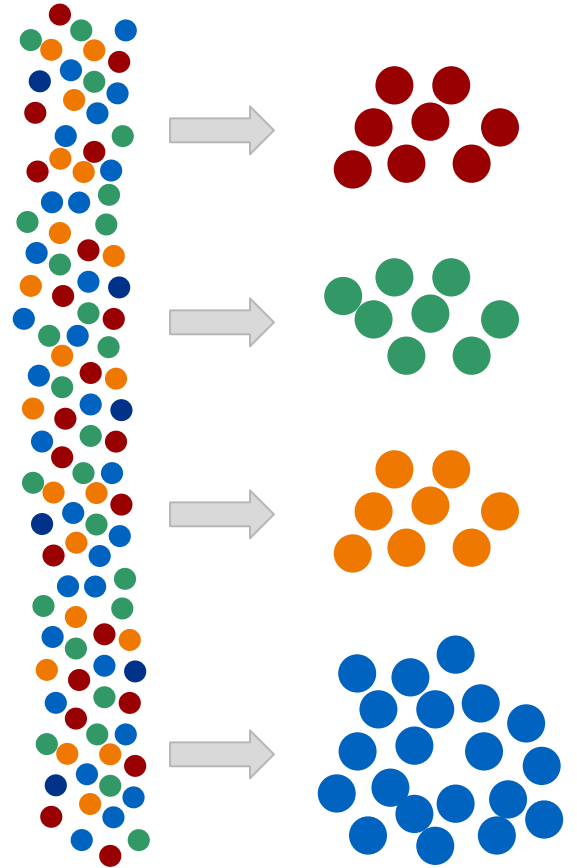
Selection (Continued)

❖ **Human review** after sampling, which ensures that the languages

- are **valid**

- could **potentially graduate**

- represent **a diverse mix of regions**

- have **no major concerns** about having a production wiki
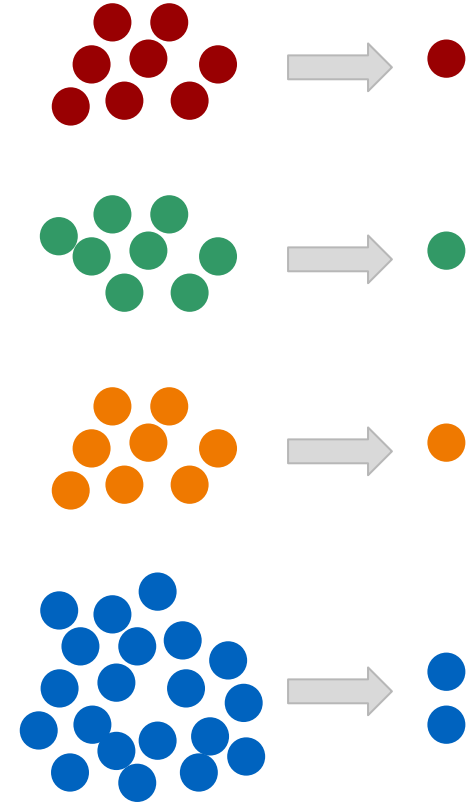
Five Wikipedias [edit source]

🔔 Subscribe

This is an unusual notice. Langcom intends to approve incubator:wp/kri, incubator:wp/nr, incubator:wp/ann, incubator:wp/tdd and incubator:wp/rsk. The last two clearly fulfill the requirements of the Language Proposal Policy. The others are deficient in various degrees. All approvals are still conditional on verification of the content being obtained (the projects posted in this section otherwise already received it) and - certainly for the first three - the community of the test-projects agreeing. This is part of an experiment proposed by the WMF Language and Product Localization team, which you can read about here ↗. At mw:Future of Language Incubation#Hypothesis statement, the idea behind it is summarized. User:SSethi (WMF) expressed the wish that these projects could be approved on 10 September, so if the circumstances work out, this will be a somewhat shorter notification time than the usual 7 days. --MF-W 00:29, 7 September 2024 (UTC) **Reply**

Update: This notification period will run the usual 7 days, so till 14 September. --MF-W 20:15, 8 September 2024 (UTC) **Reply**

It is understandable to feel **disappointed** about seeing projects with **ZERO** ↗ active community being approved for their own Wikipedia subdomains while Sylheti Wikipedia remains in the incubator despite having an active community. This does seem unfair and might discourage future contributors who wants to invest time and effort into building content and community. It is crucial for us to voice our concerns constructively to the Language Committee and ensure that active community, its engagement, and its needs are recognized. Please ensure fair treatment for all deserving language communities. -- বাঙ্গলা নালা (talk) 17:13, 9 September 2024 (UTC) **Reply**

As you can see from the other proposed approvals in this section, the approval of projects does not stop just because this experiment is being conducted. -- MF-W 11:20, 12 September 2024 (UTC) **Reply**

"The others are deficient in various degrees." I will like to know in what way Wp/ann is deficient so we can see what to do to fix it. Before the notification about this experiment, the project was listed on the incubator homepage among the active wikis that will soon get their own site. Before now, this project has met the requirement twice but it was not approved. So some people lost momentum, and some contributors gave up on it. We are however glad that the project has been selected for this experiment. We hope to see it approved soon. Katelem (talk) 22:35, 18 September 2024 (UTC) **Reply**

It is on the low side of the activity spectrum where we usually look for whether to approve a project. For example, in September so far you are the only editor who made more than 2 edits. I hope that the approval of the wiki can help your contributors gain the momentum again. --MF-W 12:21, 19 September 2024 (UTC) **Reply**

- Current (19:45, 23 September 2024 (UTC)) state:
  - Pannonian Rusyn (rsk): phabricator:T374963
  - Southern Ndebele (nr): phabricator:T375087
  - Tai Nüa (tdd): phabricator:T375422
  - Obolo (ann) pending verification of the content
  - Krio (kri) will be left out as there is no active community at all, a replacement is being considered.

Timeline

Oct-Dec 2023	Jan-Mar 2024	Apr-Jun 2024	Jul-Sep 2024	Oct-Dec 2024
Incubator conversations kick-off	**Gather recommend-ations**	**Identify solutions**	**Experiment kick-off**	**Analysis and next steps**
Discussions with stakeholders on key questions.	Documenting and collecting feedback.	Small ideas to be undertaken as experiments.	Selected wikis are invited to participate, followed by a review, content verification, and final approval by the language committee.	Production wikis setup, analysis conducted on the editing activity and next steps defined.

Experimental Idea 2

Automate Backend Site Creation

Purpose

Evaluate whether moving **addwiki.php** to **MediaWiki Core** and improving its code quality will simplify wiki creation for new creators.

Outcome

Assess whether this approach is effective for incubation and if similar **automation** should apply for other **manual wiki creation processes**.

Other Recommendations

1. Streamlining Technical Infrastructure

1.1. Automating New Language Addition and Approval in Incubator

1.2. Providing Access to Modern Wiki Features Beyond Incubator

1.3. Improving Editing Experience Within the Incubator

2. Exploring Social Pathways

1.4. Automate Backend Site Creation

Discussion

- What has the **onboarding journey** been like for your l**anguage community**?

- What **challenges** did you encounter **before joining, during your time in the Incubator, or after graduating**?

- What ideas do you have for addressing these challenges and **improving the current language incubation process**?

Thanks for your attention!

Get in touch with us:

Srishti Sethi
User:SSethi (WMF)
ssethi@wikimedia.org

Caroline Myrick
User:CMyrick-WMF
cmyrick@wikimedia.org

WIKIMEDIA
FOUNDATION

WIKIMEDIA LANGUAGE CONFERENCE
CELTIC KNOT '24