Supporting moderators at the Wikimedia Foundation

Supporting moderators in 2023-2024

- WMF has focused a lot on growing the number of new contributors over recent years

- We recognise that growing the capacity for content moderation is also important

- We can't keep adding more content and more editors without also improving the other side of the process

Improve experience of editors with extended rights: *Complete improvements to four workflows that improve the experience of editors with extended rights (admins, patrollers, functionaries, and moderators of all kinds); extend their creativity; impact at least four different wikis, and meet KRs for each improvement set collaboratively with the volunteers.*

Part of the Wikimedia Foundation's annual plan this year was to to improve workflows of editors with extended rights.

"Moderators"?

Users with extended rights, e.g.

- **Stewards**
- **Functionaries**
- **Administrators**
- **Patrollers**
- **Rollbackers**

Who are we talking about when we say "moderators" throughout this presentation? Specifically for this feature, it will only be available to users with **Rollback rights.**

Android: Patrolling Task

We're excited to discuss the steps we are taking to reduce barriers for patrolling on mobile devices for users with extended rights, specifically those with **sysops and rollback rights.** What we'll cover today is the Patrolling tasks or Edit Patrol tasks on the Android app.

It all started with Watchlist.
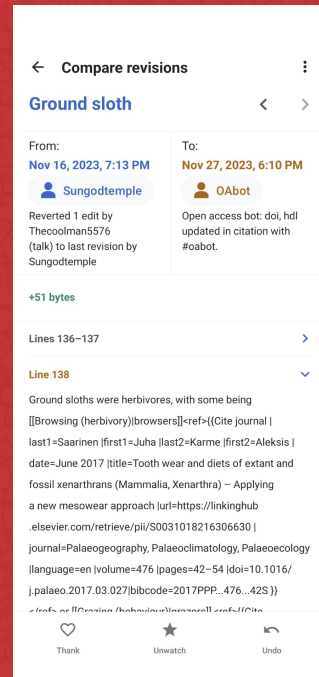
This all started with Watchlist.
- Our motivation for bringing patrolling to the app has been driven by feedback from our communities.
- Advocacy for Watchlist in the app has been heard through the App store, our support email, and community wishlists.
- We're happy to share that Watchlist is available on both the Android and iOS apps.
- As we poured over requests for adding Watchlist, a theme continued to emerge of people sharing what elements were most important. Rollback, Undo and Thank were functions that were highlighted as important actions to take.
- When we dug further, we also noticed a general theme of people wanting to take action on vandalism when they may be in transit, or generally from their mobile device, which expanded our thinking beyond Watchlist.

Opportunity
~~Problem~~ Statement

- Empowering **Mobile** Users
- **Streamlining** Vandalism Action and enable focusing on nuanced edits
- Fostering **Positive and Productive Community** Interaction
- Enhancing **Accessibility** and **Customization** for Non-Technical Users
- Enhancing our **detection models**

To develop our opportunity statement, we understood tools like Huggle, Twinkle and SWViewer (just to name a few) existed already.

There were specific requests to take some of those elements and integrate them into the Wikipedia app.

Based on the feedback from members of the community and intensive comparative analysis we conducted on existing patrolling tools we found the following opportunities.

We had the opportunity to:
- **Empower mobile users:**
    - It can be challenging to have multiple apps for multiple purposes, which is why we received requests to bring a tool in the official Wikipedia app.
    - Only one of the existing tools, SWViewer, worked on Mobile
- **Streamlining Vandalism Action and enable focusing on nuanced edits**
    - Our toolbar also increases access to take swift action on a feed of recent edits even if you are in transit.
    - More importantly we hope users of the tool are able to review more nuanced edits and provide the type of guidance that only humans can
    - Our interface surfaces the quality score of the edit to aid in determining if an edit is potentially damaging and allows filtering.
- **Enhancing Accessibility and Customization for Non-Technical Users**
    - Our research showed that having to create custom scripts for patrolling is a barrier for some folks.

- ○ There is also the challenge that some language wikis did not feel supported with current tools.
- **Enhancing our detection models**
  - ○ Finally, we wanted to ensure that users could provide feedback to the model we are using to detect edits that are at risk of reverts so that it can continue to be improved over time.
  - ○ The feature is currently using ORES scores, but will be migrated to RevertRisk in the coming months.

Hypothesis

Implementing Patroller Tasks in the Wikipedia apps will increase moderator velocity across several language wikis when particularly for mobile device users. Simultaneously, this initiative will reduce the barrier to patrol for less technical experienced editors.

Secondarily, the feedback provided through the Patroller Suggested Edits tasks will refine our edit quality models.

- We believe by building this feature we can increase moderator velocity across several language wikis, specifically when users are primarily using mobile devices.
- We recognize that not all patrolling is currently suitable for mobile so we are focused on actions that are.
- We also believe we will reduce barriers for patrolling for less technical editors with extended rights.
- Secondarily, that feedback provided to the model when the score isn't quite right will serve as an input to help improve its accuracy over time.

Evaluating Success

- **Satisfaction**
- **Engagement**
- **Impact as compared to other tools**

In the future
- **Education**

With all of these ambitious goals, how will we evaluate success? We plan to monitor
- Satisfaction of people using the tool as well as those that receive engagements via the tool.
  - There will be an indicator in the edit summary when an action is taken via the app patrolling tool.
- Engagement
  - How often are people using it and then abandoning it as well as what features folks are using.
- Impact:
  - We will view MediaWiki history to get an idea of what impact the tool is having quantitatively on content as compared to other patrolling tools.
- Finally education, in a future version, we hope to build a training gym for expert editors who do not have extended rights yet, so that they can gain experience patrolling recent changes learn how to request rights. That training task is something we absolutely must do in partnership with current patrollers with extended rights once we ensure the patrolling task for users with extended rights is nailed down

We have released V1 of Patroller tasks to Indonesian, French, Spanish Wikipedia and are continuing to iterate and improve the feature. We hope to make the feature available for sysops and rollbackers everywhere in April.
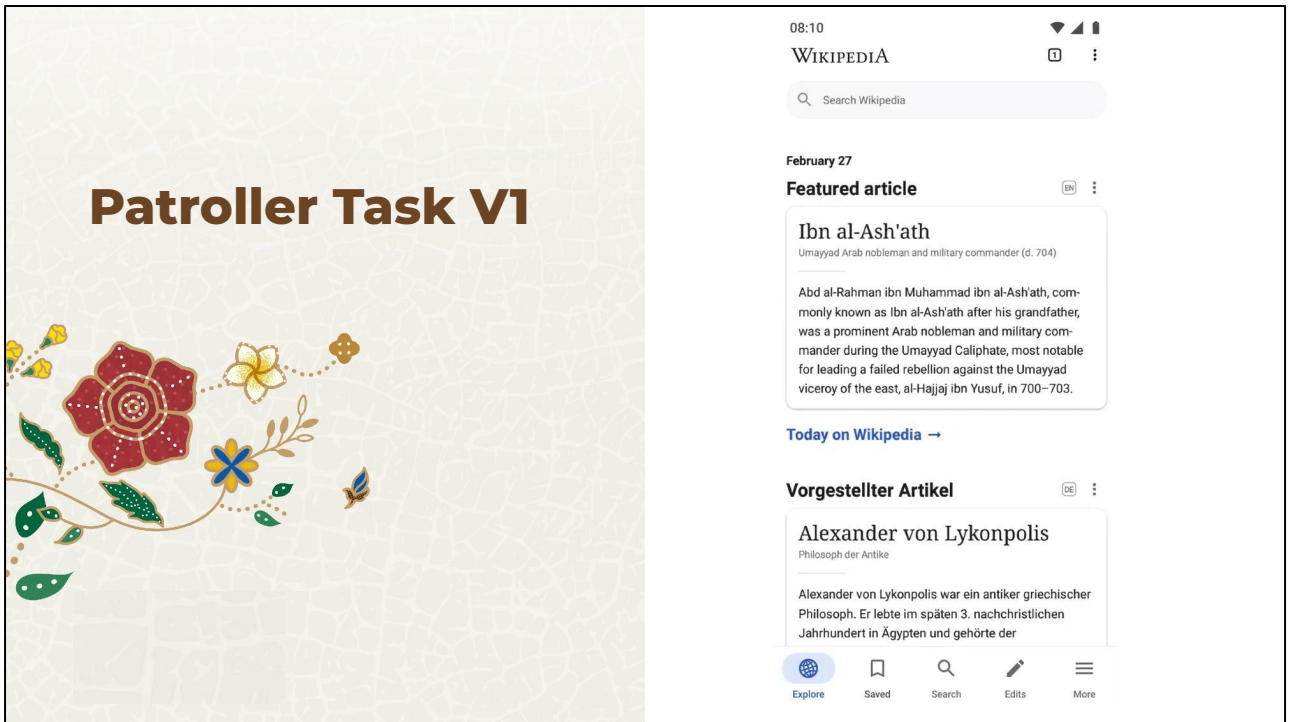
Patroller Task V1

Here is a demo of the first version of the tool
- Users with Sysops and Rollback rights would access the patroller task through the Suggested Edits entry point.
- You will be able to filter edits, and some of these options should look familiar to you based on recent edits filters on Web.
  - Here You can also easily switch between languages should have extended rights on more than one language wiki.
- Reporting issues and accessing user information is readily available from the list view
  - And if you click into an edit, you can see user info as well in the diff view.
- You have the ability to create and store warning messages locally on your device in V1.
  - In future versions we will explore adding preloaded templates based on what is commonly used and agreed upon by communities.
  - Warning messages are posted to the user's talk page.
- We've nested Rollback in undo based on feedback to not make it too easy to rollback.
  - The two clicks add a visual barrier.
  - If you want to continue to monitor an article you can add it to your watchlist.
- You can easily add that user to your watchlist from here as well.
- You can swipe or click the buttons at the top to move to the next edit in your list.
- Thank is also readily available which is helpful for users that may have gotten to reverting something before you did.
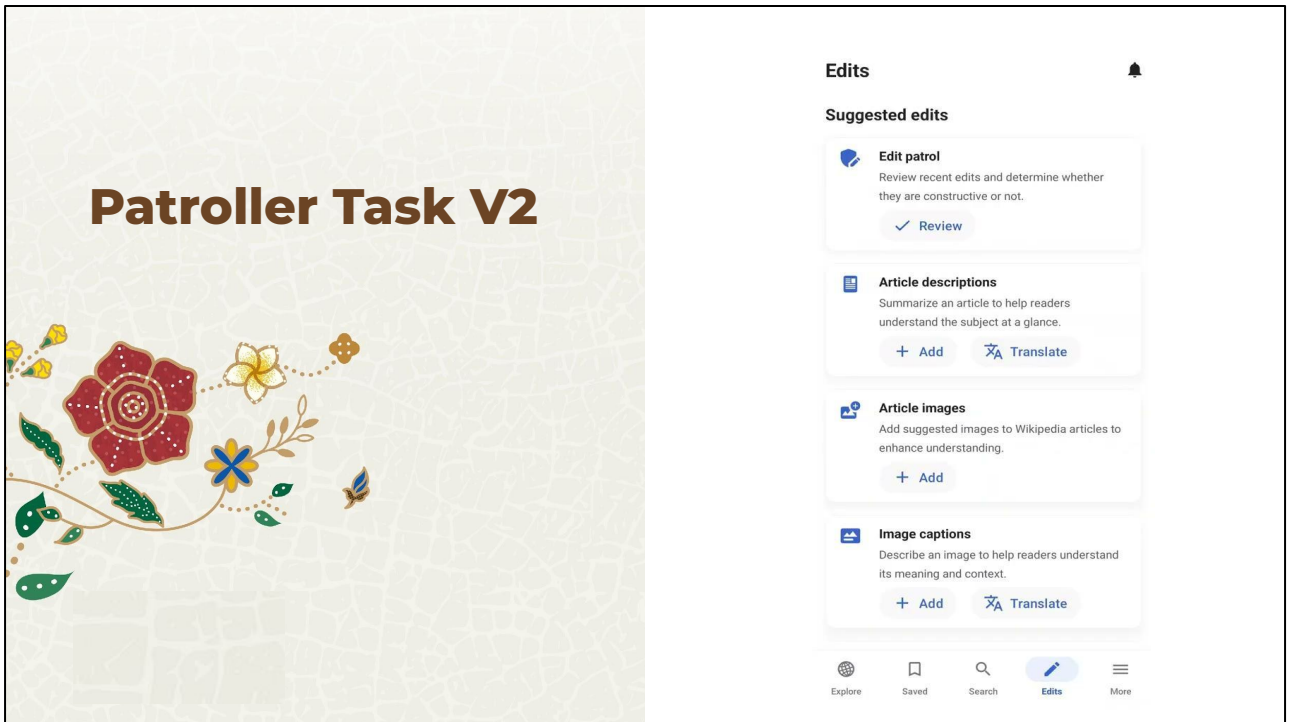
Here is a demo of the most recent version, V2:
https://upload.wikimedia.org/wikipedia/commons/5/52/Edit_Patrol_V2_Demo.webm

After hearing feedback from our users on V1, we built two improvements: Support for templates and template search, and Example messages. We also changed the wording for the talk page message flow, removing "Warn" language/iconography.

- This will be accessible from the Edits tab, to Users with Rollback rights on that Wikipedia (their primary language on the app), or global rights
- We see the recent changes feed for this language wiki (in my case, test.wiki)
- You will be able to filter edits, and some of these options should look familiar to you based on recent edits filters on Web
- You can also search in the recent changes
- And if you click into an edit, you can see user info as well in the diff view.
- You can swipe or click the buttons at the top to move to the next edit in your list.
- In Diff view,
    - You can click on the user to view more information about them
    - You can thank
    - You can Watch the user for a period of time
    - You can send a talk page message
    - And you can undo / rollback. Rollback is nested under undo to prevent accidental rollbacks and requires confirmation
        - Confirmation for undo and rollback with edit summary
- Going into the flow for leaving a talk page message, you have a few options
    - **Write a new message**
        - You have the option to save it for reuse
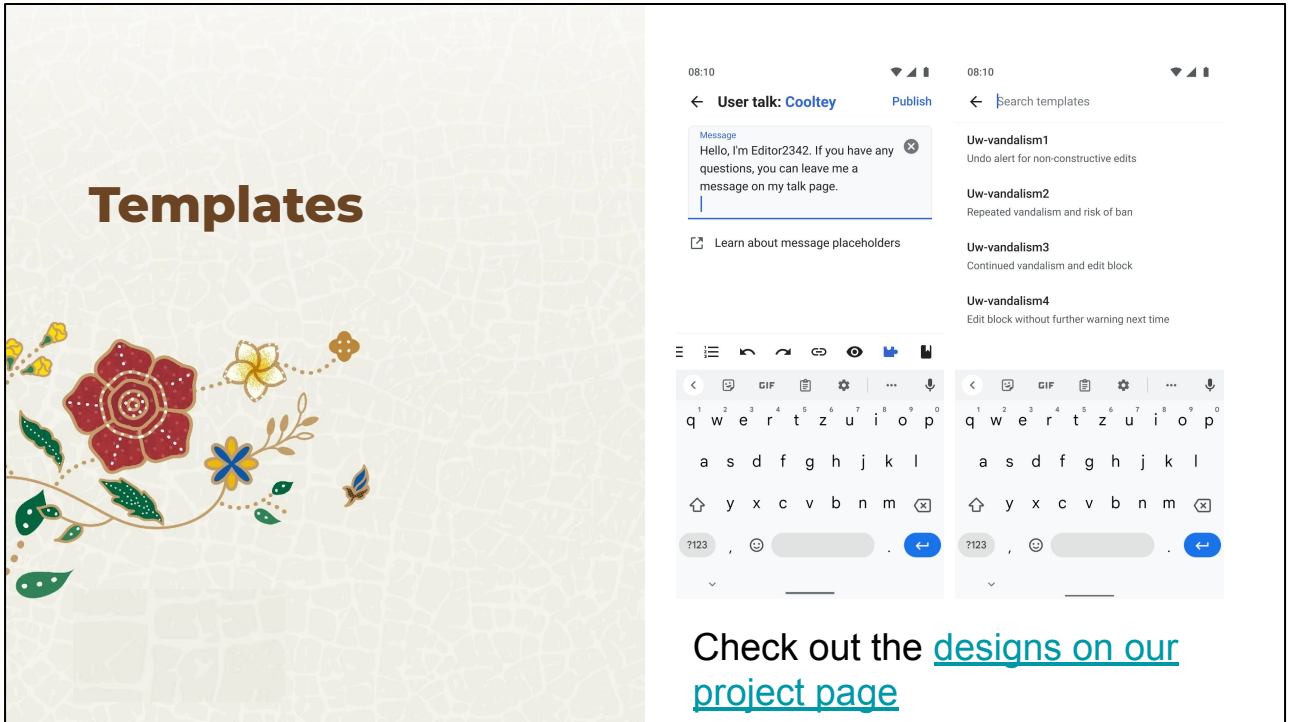   - *Thank you for the great work* example

- ○ **Use a template**
    - ■ You have the option to save it for reuse
    - ■ *Vandalism Second Warning* example
- ○ **Use one of your saved messages, "Your Messages"**
    - ■ *Welcome to Wikipedia* example
- ○ **Start from one of the Example messages:** inspired by a review of 5 languages' user warning templates. These are intended to give a starting place, especially for languages that don't yet have existing templates for patrolling purposes.
    - ■ These are shown by default, and the hope is they can eventually be configurable by each community.
    - ■ If I choose an example message, I can send it as-is, or make edits and save it to "Your Messages"
    - ■ In the example messages, we also have placeholders which refer to the user name of the sender, receiver
    - ■ *Vandalism warning* example
- ● You can always get to your Saved Messages library through the overflow menu
- ● Reporting issues and accessing user information is readily available from the list view

Saved Messages

Check out the [designs on our project page](#)
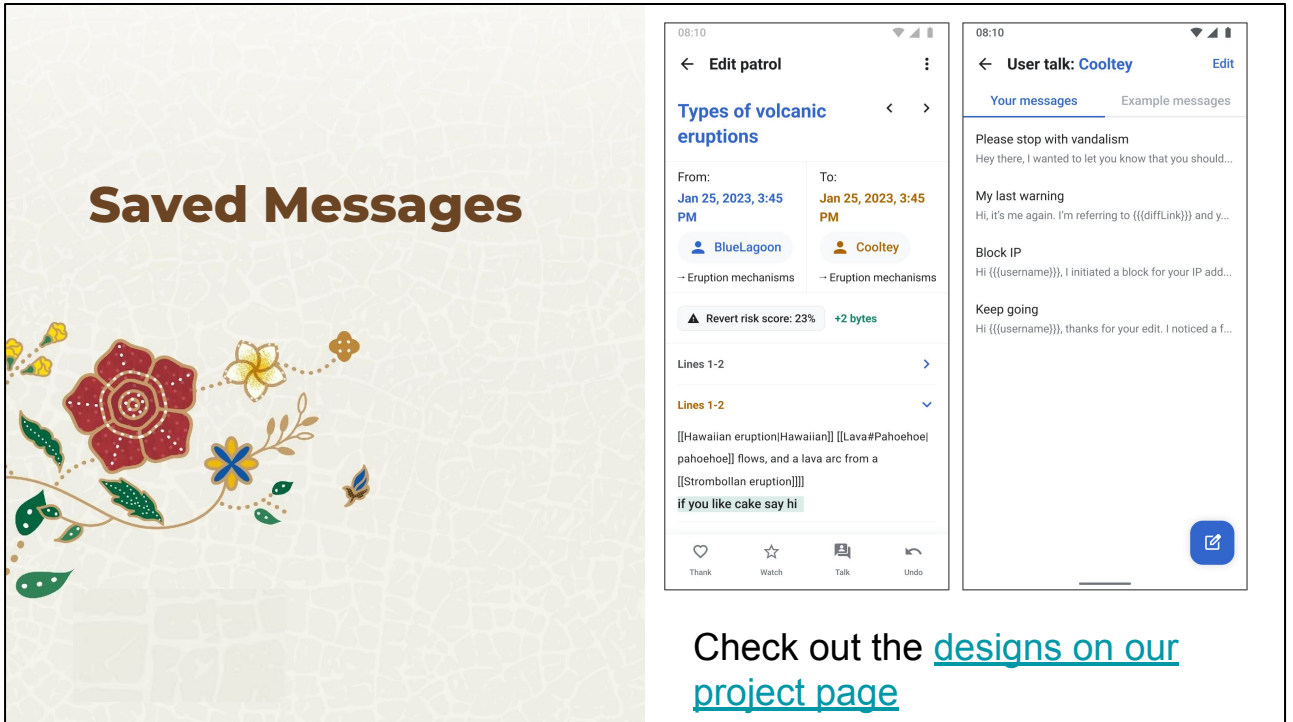
Screenshots of the Saved Messages interface for talk page communication
Example Messages can be viewed
https://www.mediawiki.org/wiki/Wikimedia_Apps/Team/Android/Anti_Vandalism/Saved_Messages

Templates

08:10 ▼◢▮	08:10 ▼◢▮
← **User talk: Cooltey** Publish	← Search templates
Message Hello, I'm Editor2342. If you have any questions, you can leave me a message on my talk page. ⊗	**Uw-vandalism1** Undo alert for non-constructive edits
	Uw-vandalism2 Repeated vandalism and risk of ban
⬈ Learn about message placeholders	**Uw-vandalism3** Continued vandalism and edit block
	Uw-vandalism4 Edit block without further warning next time

Check out the [designs on our project page](#)

Screenshots of using Template Search to create a talk page message

Discussion

- Let us know your thoughts and feedback on the Project Page
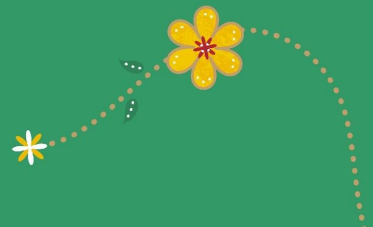
Feedback is a gift and we deeply value connection, so thank you in advance for anything you may share in response to this demonstration.

We'd like to enable this feature for your Wikipedia in the Android app with your support. Please write your reviews and feedback on the project's discussion page: https://www.mediawiki.org/w/index.php?title=Talk:Wikimedia_Apps/Team/Android/Anti_Vandalism
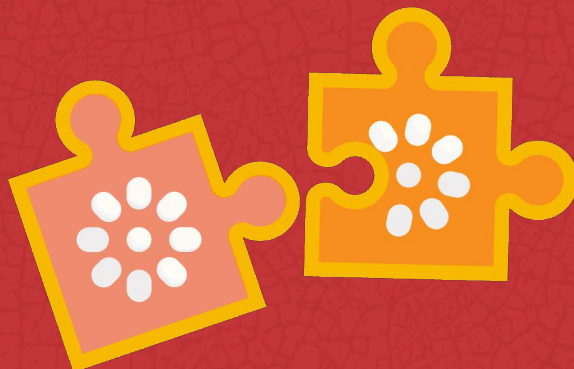
CONNECT WITH US.

MediaWiki:
[[Wikimedia_Apps/Team/Android/Anti_Vandalism]

MetaWiki:
[[Wikimedia_Apps/Newsletter]]

If you'd like to stay up to date with the project or apps work in general, feel free to visit our MediaWiki and Meta pages.