Aligning Technical Solutions & Policy to Combat Harassment on Wikimedia Projects

October 2018

WikiConference North America
~Community Health Initiative~
~Trust and Safety Team~
~Anti-Harassment Tools Team~
Sydney Poore, Wikimedia Foundation Trust & Safety Specialist
Primary purpose of the session is to discuss possible strategies for reviewing or revising community policies and workflows.
Overview of Tools by Wikimedia Foundation
Anti-Harassment Tools Team
User mute features
Notifications mute

- Stops all notifications from particular users
- The Notifications Mute feature can be found at the bottom of the "Notifications" tab of user preferences.
Special:EmailUser Mute

- Stops emails from particular users
- The EmailUser Mute feature can be found in the “User profile” tab of user preferences.
Allow users to restrict which user groups can send them direct emails

- A user can restrict new accounts from contacting them through “Email this user”.
- Available under the 'Email options' section of the 'User profile' tab of Special:Preferences.
- A tick box that when checked disallow users who are not autoconfirmed from seeing your “Email this user” link.
Anonymous User Cookie Blocking

- Addresses the issue of logged out blocked users easily avoiding site-wide blocks.

By Erin Silversmith
New Tools Under Development
Interaction Timeline

https://tools.wmflabs.org/interaction-timeline

- Investigate user misconduct
- Ahows a chronologic history for two users on pages where they have both made edits.
Look at the Interaction Timeline and then tell Claudia Lo or Trevor Bolliger how to improve it so that it works better in actual noticeboard cases.
Partial blocks

- Per-user page, namespace, and upload blocking
- For times when a full sitewide block is not needed.
- Sitewide blocks stay the same

By Alex1ruff [CC BY-SA 4.0 (https://creativecommons.org/licenses/by-sa/4.0)], from Wikimedia Commons
Partial blocks

**Call to Action**

- See Trevor Bolliger for clickable mockup of designs to give feedback.
Partial blocks

Call to Action

- In this session today discuss the current on wiki policies, processes, and workflows that might need to change if your wiki does partial blocks.
- Consider when and how communities will make these needed changes.
- What will the next steps be for consulting with communities?
Research about the effectiveness of blocks

See page on Meta
<table>
<thead>
<tr>
<th>Wiki</th>
<th>Full</th>
<th>Semi</th>
<th>Extended</th>
</tr>
</thead>
<tbody>
<tr>
<td>dewiki</td>
<td>129</td>
<td>2,257</td>
<td>0</td>
</tr>
<tr>
<td>enwiki</td>
<td>2,012</td>
<td>11,305</td>
<td>1,194</td>
</tr>
<tr>
<td>eswiki</td>
<td>133</td>
<td>910</td>
<td>0</td>
</tr>
<tr>
<td>frwiki</td>
<td>70</td>
<td>626</td>
<td>0</td>
</tr>
<tr>
<td>itwiki</td>
<td>27</td>
<td>233</td>
<td>0</td>
</tr>
<tr>
<td>Wiki</td>
<td>Full protection (only admins can edit)</td>
<td>Semi protection (autoconfirmed can edit)</td>
<td>Extended protection (enwiki only, 500+ edits can edit)</td>
</tr>
<tr>
<td>------</td>
<td>--------------------------------------</td>
<td>------------------------------------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>dewiki</td>
<td>12</td>
<td>101</td>
<td>0</td>
</tr>
<tr>
<td>enwiki</td>
<td>22</td>
<td>417</td>
<td>18</td>
</tr>
<tr>
<td>eswiki</td>
<td>1</td>
<td>70</td>
<td>0</td>
</tr>
<tr>
<td>frwiki</td>
<td>1</td>
<td>49</td>
<td>0</td>
</tr>
<tr>
<td>itwiki</td>
<td>1</td>
<td>31</td>
<td>0</td>
</tr>
</tbody>
</table>
A user reporting system for user misconduct

- Help users quickly & accurately submit reports of harassment to the correct place.

By Pete O'Shea [CC BY 2.0 (https://creativecommons.org/licenses/by/2.0)], via Wikimedia Commons
A user reporting system for user misconduct

- For people handling harassment, the software can help make faster, more accurate decisions that stand up to scrutiny.

By Adrian Jack Bunsby (Flickr: Dushi and Briti disagree) [CC BY 2.0 (https://creativecommons.org/licenses/by/2.0)], via Wikimedia Commons
A user reporting system for user misconduct

Call to Action

● Join on and off wiki discussions about a new reporting system.
● Review and discuss the initial research about Administrators noticeboard
● Review and discuss the Harvard Law School Mediation Clinic Study.
● Discuss in session today.
The English-language Wikipedia community (and most other Wikimedia projects) have conduct policies for their users to follow including policies on civility, harassment, personal attacks, and dispute resolution with established processes and workflow to address these.
Divide into groups by area of interest

Group 1: Partial blocks

Group 2: User reporting system
Questions to answer

Partial blocks

- How will your wiki community decide when to use partial blocks?
- Identify policies, processes & workflows that need to change.
- What are next steps for making these changes?
Questions to answer

User reporting system

- How will your wiki community consult with AHT team about a new user reporting system.
- Which people and groups are essential to include in a consultation?
Questions to answer

User reporting system

● What is best way to include user who were harassed and left?
● Who will lead the discussion about changes to policy, processes and workflows for various types of users who are involved with handling cases?
Thank you!