Evolution of blocking policies and events in the Wikipedia English mainspace

We are a team of 6 students from the Interdisciplinary Approaches in Research and Education master in Learning Planet Institute (France). Our small team was formed with students from the different tracks (Learning sciences, Digital sciences and Life sciences) to carry an open science project over the course of 3 months.

We were interested in exploring the possible impacts of policy development on Wikipedia's regulatory processes. Focusing on the blocking mechanism, we conducted a short study using the block log data from Wikipedia English: Wikipedia:Block_log.

Due to time constraints, we could not pursue all of our original goals, but still found some results that you might find interesting. So we would like to share this preliminary work. We feel like much more could be done, especially with from your point of view, adding much qualitative and technical knowledge, unfortunately we are all now entering internships and other projects.

We are sharing the data and our results with you so that you can dig deeper, if it's of any interest for the Wikipedia community. Here is an outline of the work we did.

Key contributions:

- We provide collected data from Wikipedia English mainspace “block log”, blocking-related policies, and scripts to reproduce our data analysis
- We found that the % of users being blocked grew overtime in spite of new policies and guidelines being developed regarding blocking and disruptive behavior
- We found that the reasons for blocking registered by the administrators gained internal consistency along with the development of a dropdown menu

Introduction

Disruptive behavior from Wikipedia users occurs on a regular basis and can take different forms, such as vandalism of a page, personal attacks etc.

The Wikipedia community designed enforcement policies to protect the encyclopedias’ content (Forte, Larco & Bruckman, 2009).

The main regulatory action is blocking of users, which technically prevents users from editing Wikipedia pages. Blocking is enforced by admins upon reports from users (Jemielniak, 2014).

Research question

How did the Wikipedians' protection from disruptive behavior using blocking evolve on the English Wikipedia along with the development of regulation policies?

Hypotheses

We expected to find:

- H1: from better understanding of policies by the users:
    - less blocked editors/total editors
- H2: from better codification of regulation
    - less variance in reason of block

Understanding the development of blocking-related policies

To frame the scope of our research we first wanted to investigate the development of the policies related to blocking.

From the Blocking policy page, we indexed all outgoing links. For each of them we retrieved the category of the page (policy, guideline, essay...) and the date of creation. When the target was a page's subsection, we looked at the addition of the section's title using the Wikiblame tool and registered it as date of creation as well.

This work can be reviewed on this document (first tab). Looking at the type of page, we chose to retain only Policies and Guidelines pages, as they are the ones holding regulation-oriented content.

Looking at the frequencies of those Policies and Guidelines over time, we noticed that a plateau was reached around 2010-2011. We decided to look at the block log data up until 2017 to be able to observe changes in blocking events long after the plateau was reached and account for any "latency" in policies implementation.

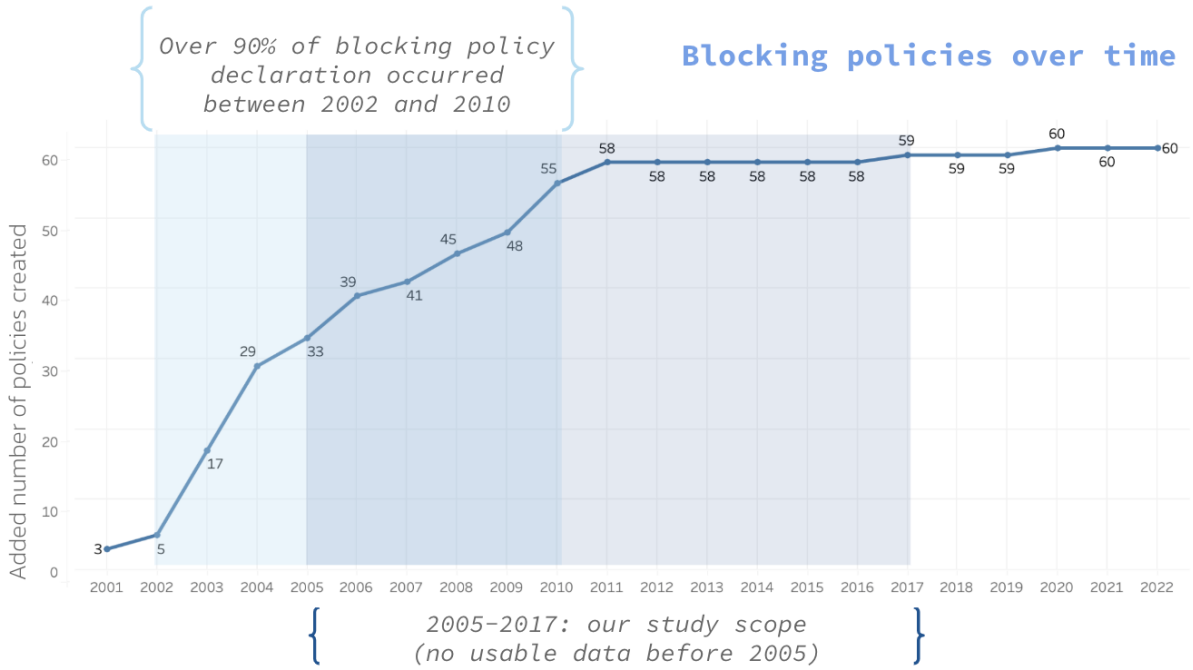

Figure 1: cumulative frequency of blocking-related policy pages creation. As it will be explained in the data section, we were not able to use the data before Dec 2004, so our scope was reduced to 2005-2017

Additional remarks:

- No qualitative analysis was conducted to investigate the content of the policies (content analysis over time)
- Some policy pages are referred on multiple occasions through their subsections, we didn't merge those, assuming that if the topic was considered important enough to be referenced independently, then it could be considered as a separate element of policies

Gathering block log data

To gather the required data we utilized the Wikipedia API

`https://en.wikipedia.org/w/api.php`

With the following request parameters:

```
action:     "query",
format:     "json",
list:       "logevents",
```

```
    lelimit:    "500",
    Letype:          "block",
    Lestart:    startDate,
    Leend:      endDate,
    Ledir:      "newer"
    Lecontinue:      continueRecords
```

Where the variables **`startDate`** & **`endDate`** were used to handle requests for yearly data e.g 2017-01-01T00:00:00Z till 2018-01-01T00:00:00Z respectively shall fetch the newest 500 records of data data for the year 2017.

The API limits the results to a max of 500 records per call, the response JSON contains the following keys:

```
{
     batchcomplete : If batch of pages is complete
     continue : If there are more records to continue pagination
     query : {
               logevents : List of log event records
}
```

Hence, the continue variable is used to fetch the next 500 records by using this value in **`continueRecords`** in the next call. The process continues until the continue variable is NULL which indicates the end of records. For each call we progressively appended the raw data in csv format.

For the scope of this project we selected the following features returned from the API:

- title : User name (or IP Address if not available) of the user that was blocked
- action : Whether the action was **block** or **unblock**
- user : The user that blocked the title user
- timestamp : Date and Time stamp of the log event
- comment : Comment provided by the user while blocking
- expiry : Date and Time stamp till which the title user is blocked
- flags : Flags selected during the block e.g nocreate, noemail, noautoblock etc.

The script to fetch any years complete log data can be found here:
Fetch_Block_Log_data.ipynb.

Computing quantitative indicators & analysis of results

The script used to filter the data and compute the indicators can be found here: Analyze_Blocklog data.ipynb.

Testing H1: Blocked users vs total editors

From Wikimedia Stats we retrieved the monthly number of editors on English Wikipedia, which we identified as active users. From the block log we retrieved the monthly number of blocked users, assuming that only active users would be blocked. That allowed us to calculate a monthly percentage of blocked users on English Wikipedia (see this document - "table_1_monthly").

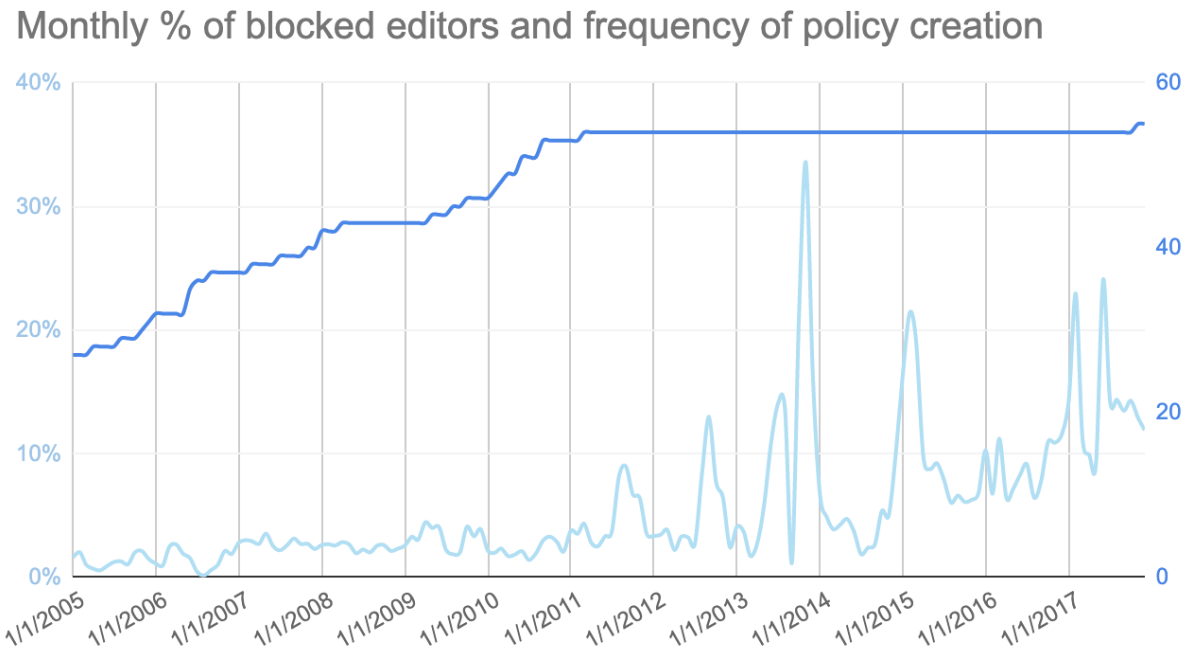

Figure 2: evolution of monthly percentage of blocked users over number of editors compared to the cumulative frequency of policy creation

The results show a rising percentage of users being blocked after the development of blocking policies. Therefore, H1 was not supported by our findings (we expected to find less blocked editors/total editors, from better understanding of policies by the users).

This could be due to the fact that policies define a better blocking criteria for penalizing or eliminating users who do not follow the wikipedia guidelines. Another

possible explanation could be that the enforcement of protection policies by the admins was made more efficient (technical and human means). Or that the community of editors extended over time with new members who were less committed to learn about policies or follow guidelines.

Other point of interest: the data showed surprising peaks, especially in Nov 2013. We thought maybe the Wikipedia community might have clues to explain this data (testing of a new bot maybe?...).

Testing H2: Rationales for blocking

Block events features such as reason for blocking or duration of block are documented in the block log. Our second hypothesis was that the development of policies helped build consistency in terms of blocking features. Over time, we expected to observe :

- less variance in reasons of block
- less variance in block duration
- less discussions related to block events

Qualitative work to analyze block duration depending on reason for blocking or length of discussions would have required more time than available, so to test this hypothesis we decided to focus on the first criteria: less variance in reason of block.

The reason for blocking was registered as "Comment" in the fetched data. Initially, the administrator would type the reason manually. A dropdown menu was added later, allowing the admin to select a reason from a list when blocking a user. The content of this menu evolved overtime and can be found here [MediaWiki:Ipbreason-dropdown]

From the Ipbreason-dropdown page, we retrieved the list of the 41 currently existing reasons for blocking and named them “tags” in our work.

```
tags =
'Vandalism**Vandalism-only**Copyright**Attack**Biographies**Citing**Patent
**Spam**promotion**NOTADVERTISING**advertising**Spam**Promotion**Edit
warring**Three-revert**Disruptive editing**Harassment**No legal
threats**legal threats**Arbitration**Evasion**Sockpuppetry**multiple
accounts**Edit filter**Long-term abuse**NOTHERE**{{anonblock}}**{{school
block}}**{{rangeblock}}**{{blocked
proxy}}**{{uw-ublock}}**{{uw-uhblock}}**{{uw-causeblock}}**{{uw-ublock-wel
lknown}}**{{uw-ublock-double}}**{{uw-uhblock-double}}**{{uw-softerblock}}*
*{{uw-spamublock}}**{{Uw-spamblacklistblock}}**{{uw-vaublock}}**{{CheckUse
r
```

```
block}}**{{checkuserblock-wide}}**{{checkuserblock-account}}**{{Tor}}**{{w
ebhostblock}}**{{colocationwebhost}}**{{OversightBlock}}'
```

We retrieved the [date of creation for all of them] using the Wikiblame tool, which showed that the first were added in 2007 - probably the first year of implementation of the dropdown menu.

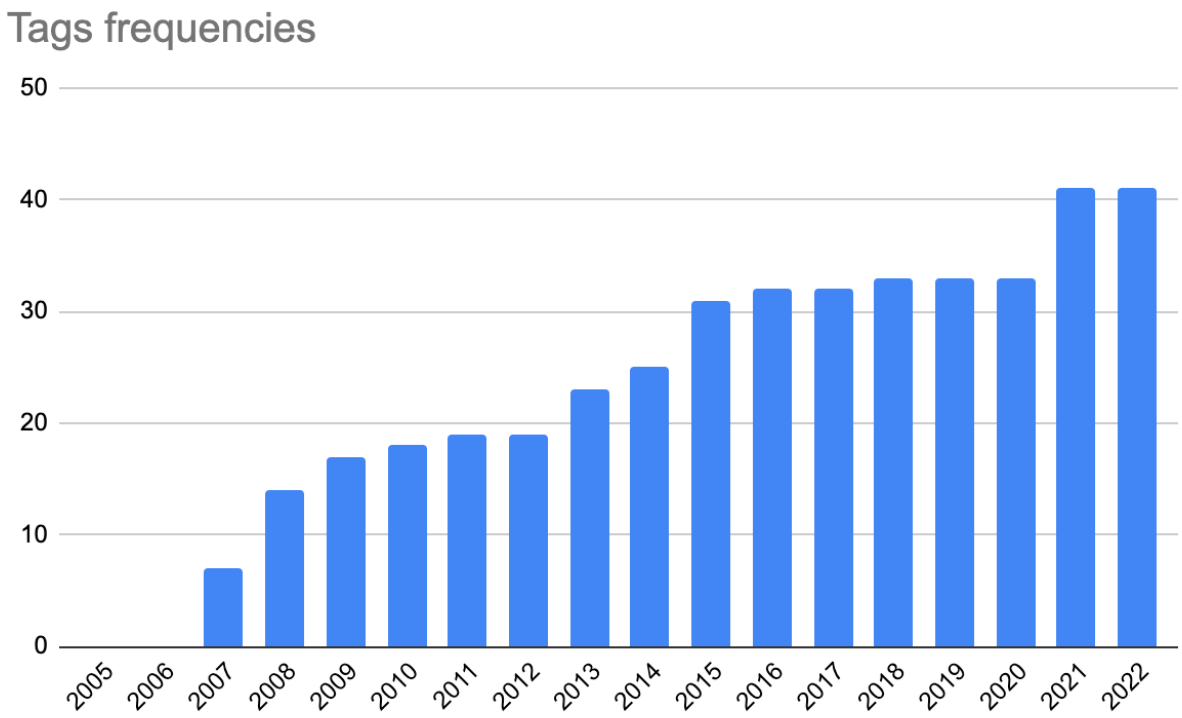

Figure 3: cumulative frequency of tags in the dropdown menu [Ipbreason-dropdown]

We matched the presence of tags within the "Comment" column for all blocks events. In the absence of a match, a "None" value was applied. A "None" value would either mean that no tag was found or that no reason was registered by the administrator as a reason for blocking.

As shown in Figure 4, the ratio for "None" values dropped as more reasons for blocking were being implemented into the dropdown menu. This observation tends to support H2 (less variance in reasons of block) but more qualitative work remains to be done on the content of "None" tagged comments to verify this hypothesis.

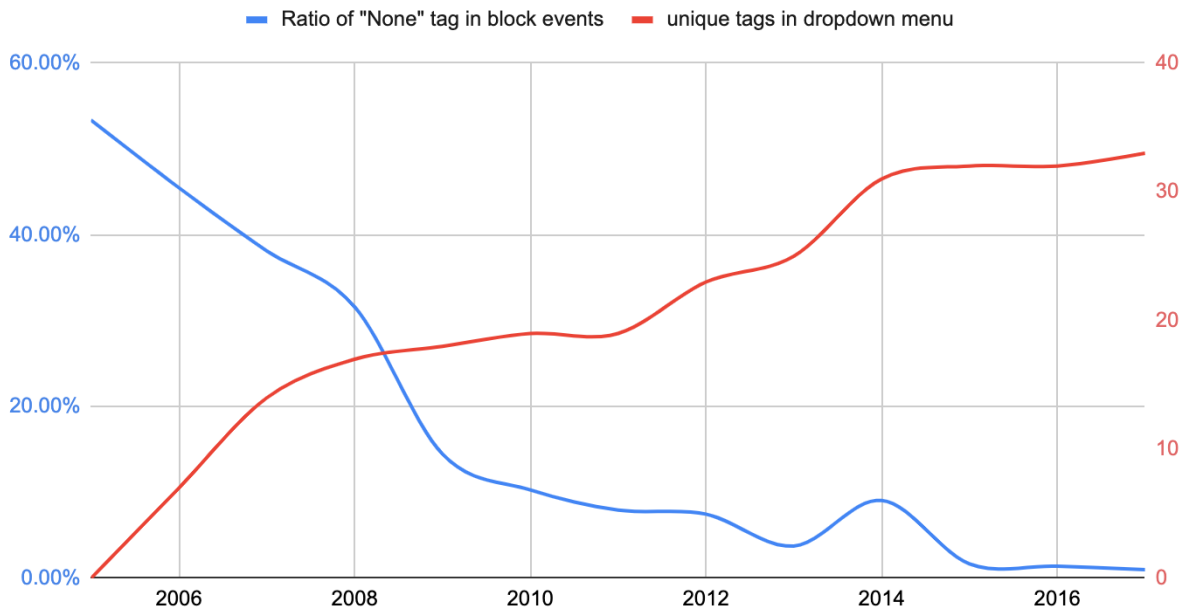

Figure 4: ratio of "None" tags for all block events compared to the number of available blocking reasons in the dropdown menu, from year 2005 to 2017 (scope of our work)

Conclusion and further work

During this project, we gathered data about blocking-related policies development by collecting the creation dates for all pages linked from the Blocking policy page. Further research could focus on tracing the history of those pages content development to observe their development after having been created.

We produced scripts to fetch the Wikipedia English mainspace "block log" data and to conduct analysis on it. We provide the data and the results of our analysis.

Analysing the data from this preliminary work, we found that, between 2005 and 2017, a growing portion of users was blocked. This happened in spite of the development of new policies and guidelines regarding blocking and disruptive behavior.

We lacked the time and knowledge to investigate the spikes observed in the ratio of blocked users over total number of editors (see figure 2).

We also found that the reasons for blocking provided by the administrators gained internal consistency over time, along with the implementation of a dropdown menu showing a predetermined list of rationales. Further qualitative investigation should

look into the content of comments for block events tagged as “None” after this preliminary work.

Then an exploratory analysis could consist in analyzing the evolution of correlation between reason for blocking and duration of block, to find out whether or not these durations gained consistency over time as well. There are probably other source of information which could be taken into account to inform such an analysis, eg. the discussion pages around block events or policy development.

We feel like more could be done, especially adding qualitative and technical knowledge from the Wikipedia community. We hope that sharing our working material will allow Wikipedians to further investigate the matter.

Team members	ORCID ID	CRI AIRE Master Track
Saffian Asghar	0000-0002-5854-1729	Disc
Mariia BAI	0000-0003-1338-2004	DiSc
Maria Bronskaia		LeSc
Anne-Flore Cabouat	0000-0002-3327-2729	LeSc
Thelma BIARD	0000-0002-1393-8239	LeSc
Mostafa Elraies	0000-0001-6666-7404	LiSc