Wikimedia Foundation v. NSA
No. 15-cv-0062-TSE (D. Md.)

Plaintiff's Exhibit 12

Handling our user data - an appeal and a response

(Today we are posting an English translation of a blog post from German Wikipedians outlining concerns about the handling of Wikipedia user data, or metadata. Above that post you will find the Foundation's response to those concerns.)

Response to user appeal

In June this year, the Wikimedia Foundation (WMF) started to solicit community input on our privacy policy, and since September we have been inviting participation in a discussion of the draft for a new privacy policy. The purpose of this discussion has been to review and improve our privacy policy, and ensure that all members of the Wikimedia community have an opportunity to be heard and contribute.

This discussion has already helped us to understand the diverse range of views in our large, international community (each month, more than 75,000 users contribute to Wikimedia projects in more than 200 languages). As part of this discussion, about 120 German Wikipedia contributors who advocate for more stringent privacy rules have made a statement and published it on the German chapter's blog (English translation below). We welcome the contribution of these editors, and hope that the resulting discussion will strengthen the policy. However, while we hear and respect these concerns, the WMF was not invited to explain its position during the drafting of the statement, and so we'd like to do so here.

Existing practices

As the authors of the statement mention, the past year has seen increased global concern about privacy and the activities of intelligence agencies in both the US and Europe. The Wikimedia Foundation is extremely sensitive to those concerns, and we have taken several steps to address them, including joining activism here in the US, encrypting more traffic to and from the Wikimedia sites, and assuring readers that we have not been contacted under the surveillance programs at issue.

The Wikimedia Foundation also protects its readers by collecting very little information, particularly relative to most major websites. Editors who create an account do not have to connect their account to a real-world identity unless they choose to do so. It is possible to read and use the Wikimedia sites without providing your real name, home address, email address, gender, credit card or financial information. In all but a few cases (related to abuse prevention), we delete IP addresses of logged-in editors after 90 days. All in all, there is small incentive for governments to contact WMF and request information about Wikimedia users.

Requested changes

As part of the normal operation of a wiki, the Wikimedia sites have always published certain information about edits, particularly when the edit was made, and what page was edited. This information can be collected to make educated guesses about an account, such as what time zone an account is in (based on when edits occur).

One part of the statement asks that we limit public access to this editing information. This information is used in a variety of places, many of which are important to the health and

functioning of our projects:

- Protecting against vandalism, incorrect and inappropriate content: There are several bots that patrol Wikipedia's articles that protect the site. Without public access to metadata, the effectiveness of these bots will be much reduced, and it is impossible for humans to perform these tasks at scale.
- Community workflows: Processes that contribute to the quality and governance of the project will also be affected: blocking users, assessing adminship nominations, determining eligible participants in article deletion discussions.
- Automated tools: Certain community-created tools that help perform high-volume editing (such as Huggle, for vandalism fighting on several wikis) will be broken without public access to this metadata.
- Research: Researchers around the world use this public metadata for analysis that is essential to the site and the movement's understanding of itself.
- Forking: Allowing others to fork is an important principle of the movement, and that requires some exposure of metadata about how articles were built, and by whom.

The Foundation has been open and transparent about these data publication practices for years, so we do not currently plan to make the requested changes. Nevertheless, we welcome the appeal as part of the wider community discussion regarding the Foundation's privacy policy.

The statement also asks that we implement a new policy on the Wikimedia Labs experimental development servers. The predecessor to Labs, called Toolserver, had a policy that prohibited volunteer-developed software if the software aggregated certain types of account information without consent. The terms of use for Labs allows such software to be taken down at the WMF's discretion, but does not prohibit it explicitly. We have suggested a clarification to the Labs terms of use in the Privacy Policy discussion, and will continue to discuss that there.

We invite anyone interested in Wikimedia Foundation's privacy policy to get involved in the ongoing consultation with the Wikimedia community. You can read more about that process in the blog post that announced it. The consultation process will continue through January 15, 2013.

Luis Villa

Deputy General Counsel, Wikimedia Foundation

(Translated blog post from Wikimedia Deutschalnd follows - from https://meta.wikimedia.org/wiki/Wikimedia_Blog/Drafts/Handling_our_user_data_-_an_appeal)

Handling our user data - an appeal

Preface (Wikimedia Deutschland)

For several months, there have been regular discussions on data protection and the way Wikimedia deals

with it, in the German-speaking community – one of the largest non-English-speaking communities in the Wikimedia movement. Of course, this particularly concerns people actively involved in Wikipedia, but also those active on other Wikimedia projects.

The German-speaking community has always been interested in data protection. However, this particular discussion was triggered when the Deep User Inspector *tool on Tool Labs nullified a long-respected agreement in the Toolserver, that aggregated personalized data would only be available after an opt-in by the user.*

As the Wikimedia Foundation is currently reviewing its privacy policy and has requested feedback and discussion her by 15 January, Wikimedia Deutschland has asked the community to draft a statement. The text presented below was largely written by User:NordNordWest and signed by almost 120 people involved in German Wikimedia projects. It highlights the many concerns and worries of the German-speaking community, so we believe it can enhance the discussion on these issues. We would like to thank everyone involved.

This text was published in German simultaneously in the Wikimedia Deutschland-blog and in the Kurier, an analogue to the English "Signpost". This translation has been additionally placed on the talkpage of the WMF-privacy-policy-draft at Meta.

(preface Denis Barthel (WMDE) (talk), 20.12.)

Starting position

The revelations by Edward Snowden and the migration of programs from the Toolserver to Tool Labs prompted discussions among the community on the subject of user data and how to deal with it. On the one hand, a diverse range of security features are available to registered users:

- Users can register under a pseudonym.
- The IP address of registered users is not shown. Only users with CheckUser permission can see IP addresses.
- Users have a right to anonymity. This includes all types of personal data: names, age, background, gender, family status, occupation, level of education, religion, political views, sexual orientation, etc.
- As a direct reaction to Snowden's revelations, the HTTPS protocol has been used as standard since summer 2013 (see m:HTTPS), so that, among other things, it should no longer be visible from outside which pages are called up by which users and what information is sent by a user.

On the other hand, however, all of a user's contributions are recorded with exact timestamps. Access to this data is available to everyone and allows the creation of user profiles. While the tools were running on the Toolserver, user profiles could only be created from aggregated data with the consent of the user concerned (opt-in procedure). This was because the Toolserver was operated by Wikimedia Deutschland and therefore subject to German data protection law, one of the strictest in the world. However, evaluation tools that were independent of the Foundation and any of its chapters already existed.

One example is Wikichecker, which, however, only concerns English-language Wikipedia. The migration of programs to ToolLabs, which means that they no longer have to function in accordance with German data protection law, prompted a survey of whether a voluntary opt-in system should still be mandatory for

X!'s Edit Counter or whether opt-in should be abandoned altogether. The survey resulted in a majority of 259 votes for keeping opt-in, with 26 users voting for replacing it with an opt-out solution and 195 in favor of removing it completely. As a direct reaction to these results, a new tool – Deep User Inspector – was programmed to provide aggregated user data across projects without giving users a chance to object. Alongside basic numbers of contributions, the tool also provides statistics on, for example, the times on weekdays when a user was active, lists of voting behavior, or a map showing the location of subjects on which the user has edited articles. This aggregation of data allows simple inferences to be made about each individual user. A cluster of edits on articles relating to a certain region, for example, makes it possible to deduce where the user most probably lives.

Problems

Every user knows that user data is recorded every time something is edited. However, there is a significant difference between a single data set and the aggregated presentation of this data. Aggregated data means that the user's right to anonymity can be reduced, or, in the worst case, lost altogether. Here are some examples:

- A list of the times that a user edits often allows a deduction to be made as to the time zone where he or she lives.
- From the coordinates of articles that a user has edited, it is generally possible to determine the user's location even more precisely. It would be rare for people to solely edit area X, when in fact they came from area Y.
- The most precise deductions can be made by analyzing the coordinates of a photo location, as it stands to reason that the user must have been physically present to take the photo.
- Places of origin and photo locations can reveal information on the user's means of transport (e.g. whether someone owns a car), as well as on his or her routes and times of travel. This makes it possible to create movement profiles on users who upload a large number of photos.
- Time analyses of certain days of the year allow inferences to be drawn about a user's family status. It is probable, for example, that those who tend not to edit during the school holidays are students, parents or teachers.
- Assumptions on religious orientation can also be made if a user tends not to edit on particular religious holidays.
- Foreign photo locations either reveal information about a user's holiday destination, and therefore perhaps disclose something about his or her financial situation, or suggest that the user is a photographer.
- If users work in a country or a company where editing is prohibited during working hours, they are particularly vulnerable if the recorded time reveals that they have been editing during these hours. In the worst-case scenario, somebody who wishes to harm the user and knows extra information about his or her life (which is not unusual if someone has been an editor for several years) could pass this information on to the user's employer. Disputes within Wikipedia would thus be carried over into real life.

Suggestions

Wikipedia is the fifth most visited website in the world. The way it treats its users therefore serves as an important example to others. It would be illogical and ridiculous to increase user protection on the one

hand but, on the other hand, to allow users' right to anonymity to be eroded. The most important asset that Wikipedia, Commons and other projects have is their users. They create the content that has ensured these projects' success. But users are not content, and we should make sure that we protect them. The Wikimedia Foundation should commit to making the protection of its registered users a higher priority and should take the necessary steps to achieve this. Similarly to the regulations for the Toolserver, it should first require an opt-in for all the tools on its own servers that compile detailed aggregations of user data. Users could do this via their personal settings, for example. Since Wikipedia was founded in 2001, the project has grown without any urgent need for these kinds of tools, and at present there seems to be no reason why this should change in the future. By creating free content, the community enables Wikimedia to collect the donations needed to run WikiLabs. That this should lead to users loosing their right of anonymity, although the majority opposes this, is absurd. To ensure that user data are not evaluated on non-Wikimedia servers, the Foundation is asked to take the following steps:

- Wikipedia dumps should no longer contain any detailed user information. The license only requires the name of the author and not the time or the day when they edited.
- There should only be limited access to user data on the API.
- It might be worth considering whether or not it is necessary or consistent with project targets to store and display the IP addresses of registered users (if they are stored), as well as precise timestamps that are accurate to the minute of all their actions. The time limit here could be how long it reasonably takes CheckUsers to make a query. After all, data that are not available cannot be misused for other purposes.

submitted by Silke WMDE (talk) 16:21, 20 December 2013 (UTC)