Participation in Wikipedia's Article Deletion Processes

R. Stuart Geiger
School of Information
University of California, Berkeley
102 South Hall, Berkeley CA 94703
sgeiger@berkeley.edu

Heather Ford
School of Information
University of California, Berkeley
102 South Hall, Berkeley CA 94703
hford@ischool.berkeley.edu

ABSTRACT

We present results on a study of two levels of Wikipedia's article deletion process: speedy deletions (or CSDs) and articles for deletions (or AfDs). Our findings indicate that the deletion process is heavily frequented by a relatively small number of longstanding users. In analyzing the rationales given for such deletions, it is apparent that the vast majority of such deleted articles are not spam, vandalism, or 'patent nonsense,' but rather articles which could be considered encyclopedic, but do not fit the project's standards.

Keywords

Wikipedia, governance, community, administration, bureaucracy

1. INTRODUCTION

On the English-language version of Wikipedia, the self-proclaimed "free encyclopedia anyone can edit," hundreds of articles are created each day, only to be 'deleted' by one of the project's 1,750 or so administrators. Any administrator has the technical ability to delete any article, but any other administrator can also undelete any deleted article; as such, there have evolved complex norms, standards, venues, and modes of discourse for determining what ought to be kept and what ought to be removed. At present, the deletion process is comprised of several distinct yet interrelated bureaucratic procedures for determining whether an article fits the project's encyclopedic standards. From AfDs and CSDs to PRODs and DRVs, each of the different procedures have their own discursive and normative standards in which Wikipedia's administrative apparatus assembles in response to allegedly unencyclopedic articles. Previous research has focused on Wikipedia's deletion processes, but are typically concerned with decision making and decision [9,13]. In contrast, this study is primarily concerned with the accessibility of such processes, particularly for newcomers and content creators.

Wikipedians have a complex schema defining 24 reasons why an article may be unilaterally deleted by an administrator, with no need for a discussion or debate. These 'criteria for speedy deletions' (or 'CSDs') include ostensibly non-controversial justifications like "G1: Patent nonsense, meaningless, or incomprehensible" or "A10. Recently created article that duplicates an existing topic." For articles that do not fit any of these categories – or for users who are not administrators – the article must be nominated for deletion. This process, called AfD, involves creating a new sub-page in a discussion space called "Articles for Deletion" for administrators and editors to discuss why they think the article topic is or is not encyclopedic. After typically seven days of open discussion, a typically non-involved administrator will 'close' the debate, announcing the outcome, and typically carrying out the deletion if

such is the result. These discussions can range from as few as four or five participants in cases of banal articles to as many as a hundred or more participants when the topic is contested or popular. Yet in both the small and large-scale debates, Wikipedians invoke and debate fundamental conceptions about social, cultural, political, and epistemological issues about how Wikipedia as both a global community and a universal encyclopedia is and ought to be.

2. METHODOLOGY

As in Geiger and Ribes's "trace ethnography" [3,4] of Wikipedian vandal fighters, we learned through long-term ethnographic fieldwork that veteran editors and administrators frequently leave specialized traces in revision history and log metadata, specifically the edit summary field. This is particularly the case in deletion debates: when administrators unilaterally delete articles based on a speedy deletion criteria, for example, they routinely leave a specific marker that includes the text 'CSD' and the specific criteria, such as 'G1' for "patent nonsense" or 'A7' for "No indication of importance (individuals, animals, organizations, web content)." The regularity of both traces themselves and the modes of tracing which proliferate in environments like Wikipedia make it possible to produce aggregate level analytics of phenomena like deletion processes. In fact, such an approach is a mainstay of research into online communities where such trace data is widely or publicly available, as it is in Wikipedia. For example, research on barnstars [7] and especially policy citations [1,2,8,12] has both qualitatively demonstrated and quantitatively relied on specific kinds of traces Wikipedians leave when they want to make a socially-relevant action. By examining contribution histories, for example, researchers have been able to produce sophisticated analyses of group and leadership dynamics [5,6], and the lifecycles of new and 'lurking' users [10,11].

Our analysis was performed by obtaining the publically-accessible records of every AfD discussion and every article deleted from the English-language version of Wikipedia between 1 June 2007 and 1 July 2011. This was made rather easy because of the MediaWiki API, which provides structured access to various records and logs of user activity. The revision histories obtained contained a series of human and platform-authored traces in the edit summaries which allowed us to automatically code for the criteria for deletion used, and revision history metadata was used to determine participation in these administrative spaces, as well as the creator of the nominated article. However, it is possible that some of this trace data is inaccurate: e.g. we found that in 1.59% of all deletions, administrators noted that the deletion was a CSD but did not properly specify a coded rationale.

3. RESULTS

We found that over half (59.66%) of all deleted articles from June 2007 to July 2011 were unilaterally deleted by administrators via the CSD process. While it is logical to assume that the majority of such deletions would be deleting spam or vandalism, our analysis of the criteria for speedy deletion used showed differently. "A7: No indication of importance" is overwhelmingly the most used

Chart 1: Number of participants in AfD debates

Rationales (some rationales truncated due to space limits)	Deleted articles	% of all CSDs	% of all deletions
A7: No indication of importance	497397	37.13%	22.16%
G11: Unambiguous advertising	101723	7.59%	4.53%
G1: Patent nonsense	88084	6.58%	3.92%
A1: No context	79139	5.91%	3.53%
G3: Pure vandalism	68454	5.11%	3.05%
G10: Attack pages	62449	4.66%	2.78%
A3: No content.	59889	4.47%	2.67%
G12: Copyright infringement.	58254	4.35%	2.59%
R1: Redirects to non-existent pages	57091	4.26%	2.54%
G6: Technical deletions.	54750	4.09%	2.44%
G7: Author requests deletion.	53595	4.00%	2.39%
G8: Pages dependent on a deleted page.	45331	3.38%	2.02%
G2: Test pages.	24388	1.82%	1.09%
R3: Implausible typos.	23343	1.74%	1.04%
R2: Redirects from the main namespace to any other namespace	9152	0.68%	0.41%
G5: Creations by banned users.	5782	0.43%	0.26%
Other named criteria	14,943	1.11%	0.66%
CSD, but criteria not specified	35685	2.66%	1.59%
All CSDs	1339449	100.00%	59.66%
All deleted articles	2244952	--	100%
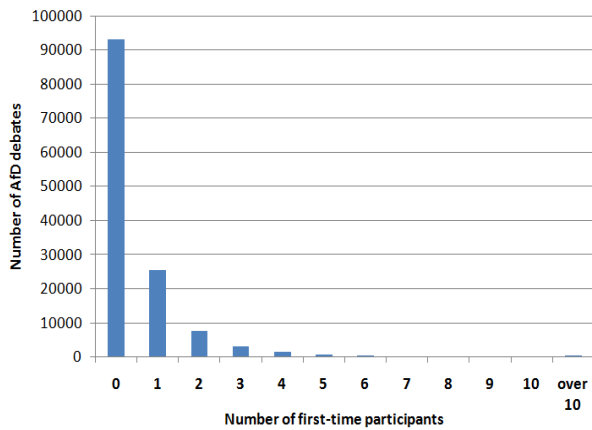
Figure 1: AfD debates, grouped by number of users who have never previously participated in an AfD

category, making up 37.13% of all CSD criteria, while spam, vandalism, and patent nonsense made up 7.59%, 5.11% and 6.58% respectively. In our analysis of AfD debates, we found that as with the study by Lam et al [9], such processes have few participants, and those participants are overwhelmingly regulars to the process. We found that 83.62% of all AfDs have between 4 and 12 participants, and on average, 95.8% of the participants in an AfD have participated in at least one previous AfD (Figure 1). In addition, we found that 74% of all AfDs are comprised entirely of users who have previously participated in an AfD, and 18% of all AfDs only have one newcomer. Finally, we found that only 17.59% of the creators of nominated articles ever participated in that article's deletion discussion.

4. ANALYSIS AND DISCUSSION

There is a strong argument to be made that new users have significant difficulty in navigating these organizational processes – a claim Wikipedians and critics alike have made over the years. According to many popular and anecdotal accounts, the deletion process is plagued by highly-nuanced standards and norms, substantial use of jargon and categorization, compartmentalization of related processes, and a significant imbalance between the number of procedurally-oriented administrators and procedurally-unaware newcomers. While further study is certainly necessary on this issue, our preliminary findings indicate that the deletion process is heavily frequented by longstanding users, even though one would assume that an article being nominated for deletion would bring in editors who worked on that article to defend it. In analyzing the rationales given for speedy deletions, it is apparent that the vast majority of articles deleted in such a manner are not spam, vandalism, or 'patent nonsense,' but rather articles which could be considered encyclopedic, but do not fit the project's standards. The A7 CSD criteria in particular shows how many articles are deleted not because the topic was explicitly judged to be unimportant, but rather that such importance was simply not indicated. Yet the participation rates in AfDs suggest that new users who likely do not know that they must specify such importance (or the discursive techniques and technical practices required to, say, add citations) are certainly not entering AfD debates to defend the legitimacy of their newly-created articles.

5. REFERENCES

[1] Beschastnikh, I., Kriplean, T., and McDonald, D. Wikipedian Self-Governance in Action: Motivating the Policy Lens. *Proc ICWSM 2008*, (2008).

[2] Bruckman, A. and Forte, A. Scaling Consensus: Increasing Decentralization in Wikipedia Governance. *Proc HICSS 2008*, IEEE (2008), 157.

[3] Geiger, R.S. and Ribes, D. The Work of Sustaining Order in Wikipedia: The Banning of A Vandal. *Proc CSCW 2010*, ACM (2010).

[4] Geiger, R.S. and Ribes, D. Trace Ethnography: Following Coordination Through Documentary Practices. *Proc HICSS*, IEEE (2011).

[5] Kittur, A., Suh, B., Pendleton, B.A., and Chi, E.H. He says, she says: conflict and coordination in Wikipedia. *Proc CHI 2008*, ACM (2008), 453-462.

[6] Kraut, R. and Burke, M. Taking up the mop: identifying future wikipedia administrators. *Proc CHI 2008*, ACM (2008), 3441-3446.

[7] Kriplean, T., Beschastnikh, I., and McDonald, D. Articulations of wikiwork: uncovering valued work in wikipedia through barnstars. *Proc CSCW 2008*, ACM (2008), 47-56.

[8] Kriplean, T., McDonald, D.W., Beschastnikh, I., and Golder, S.A. Community, consensus, coercion, control: cs*w or how policy mediates mass participation. *Proc GROUP 2007*, ACM (2007), 167-176.

[9] Lam, S.K., Karim, J., and Riedl, J. The effects of group composition on decision quality in a social production community. *Proc GROUP 2010*, ACM Press (2010), 55.

[10] Panciera, K., Halfaker, A., and Terveen, L. Wikipedians are born, not made: a study of power editors on Wikipedia. *Proc GROUP 2009*, ACM (2009), 51-60.

[11] Panciera, K., Priedhorsky, R., Erickson, T., and Terveen, L. Lurking? cyclopaths? a quantitative lifecycle analysis of user behavior in a geowiki. *Proc CHI 2010*, ACM Press (2010).

[12] Pike, J., Joyce, E., and Butler, B. Don't look now, but we've created a bureaucracy: the nature and roles of policies and rules in wikipedia. *Proc CHI 2008*, ACM (2008), 1101-1110.

[13] Taraborelli, D. and Ciampaglia, G.L. Beyond notability. Collective deliberation on content inclusion in Wikipedia. *Proc SASOW 2010*, IEEE (2010), 122-125.