

# Navigating between Chaos and Bureaucracy: Backgrounding Trust in Open-Content Communities

Paul B. de Laat

Faculty of Philosophy, University of Groningen, Groningen, The Netherlands

p.b.de.laat@rug.nl

**Abstract.** Many virtual communities that rely on user-generated content (such as social news sites, citizen journals, and encyclopedias in particular) offer unrestricted and immediate ‘write access’ to every contributor. It is argued that these communities do not just assume that the trust granted by that policy is well-placed; they have developed extensive mechanisms that underpin the trust involved (‘backgrounding’). These target contributors (stipulating legal terms of use and developing etiquette, both underscored by sanctions) as well as the contents contributed by them (patrolling for illegal and/or vandalistic content, variously performed by humans and bots; voting schemes). Backgrounding trust is argued to be important since it facilitates the avoidance of bureaucratic measures that may easily cause unrest among community members and chase them away.

**Keywords:** access, bots, bureaucracy, etiquette, open content, trust, vandalism, Wikipedia

## 1 Introduction

Online communities that thrive on user-generated content come in various formats. Contents may vary considerably—from text, photographs, videos, designs and logos to source code. Furthermore, cooperation may range from ‘loose’ interaction: uploaded contents are presented as-is—to ‘tight’ interaction: an evolving product is being worked on collectively. This distinction in patterns of cooperation is referred to by Dutton (2008) as ‘contributing 2.0’ vs. ‘co-creation 3.0’. Typical examples of the former are Flickr and YouTube, of the latter Wikipedia and open source software.

These communities face the dilemma of determining which contributors are to be accepted as members and how contributions are to be processed and published. Some communities take a cautious approach: only some categories of people are allowed to contribute, and their contributions are critically examined, by filtering before reception or moderating afterwards. A typical example is the Encyclopedia of Earth which only accepts input from acknowledged experts. Moreover, their appointed ‘topic editors’ decide who is to write the entries and who is to participate in reviewing them. In the end they have to approve of entries appearing in a public version. Other communities, though, prefer to hand out a generous invitation to their ‘crowds’ in order to maximize possible returns. This consists of two parts: (1) Anyone is invited to contribute content without any restrictions on entry; accordingly, access is fully open to anyone who cares to contribute; (2) Contents contributed are subsequently accepted with no questions asked and appear right away in the appropriate place. Publication proceeds without review and without delay. In terms of Goldman (2008): no filtering is applied at the reception stage.

Which communities typically practice this two-fold institutional gesture? Let me mention some of them in so far as they predominantly revolve around soliciting and reworking of *text*. I select these since it seems especially with text that the whole spectrum from contribution (2.0) to co-creation (3.0) unfolds; activities in communities which focus on other kinds of content most often remain at the level of contributing.<sup>1</sup> The first category is ‘*social news*’ sites that focus on creating a collective discussion about topics in the news that are deemed to be relevant. The formula is basically the same for all: users are invited to submit news stories and/or news links that will be put up for public discussion (comments). In this category we find Digg (<http://digg.com>; started in 2004) and Reddit (<http://www.reddit.com>; 2005) which focus on news of all kinds, and Slashdot (<http://slashdot.org>; 1997) and Hacker News (<http://news.ycombinator.com>; 2007) which focus on technology-related issues.<sup>2</sup>

The second category is user-generated *newspapers* that have been around since 2004. NowPublic (<http://www.nowpublic.com>; 2005), Digital Journal (<http://www.digitaljournal.com>; 2006) and GroundReport (<http://www.groundreport.com>; 2006) invite everybody to become a citizen journalist and contribute their own articles, blog entries and/or images to the site, as well as leave comments on

---

<sup>1</sup> These communities will be enumerated with their URL and year of foundation in brackets.

<sup>2</sup> All websites referred to in text, footnotes, and references were last accessed on 21 September 2012.

those of others. These contributions essentially remain unaltered. Wikinews (<http://en.wikinews.org>; also in language versions other than English; 2004) goes one step further: in the so-called ‘news room’ articles which have been submitted are polished further by fellow contributors (by means of a wiki). As soon as criticisms have been met, the article can officially appear on the ‘front page’.

The third and final category consists of user-generated *encyclopedias*. Many such communities exist (cf. de Laat 2012a), but only a few have adopted policies of open access and immediate publication. British h2g2 (<http://h2g2.com>; 2001) invites everybody to compose entries; these are put up on the site for public commentary. Wikipedia (<http://en.wikipedia.org>; also in many languages other than English; 2001) and Citizendium (<http://en.citizendium.org>; 2007) lean more towards co-creation by publishing new entries in an open-access wiki that allows other participants to instantaneously insert their own textual changes.<sup>3</sup>

## 2 Trust

This gesture of unrestricted and immediate access to the community platform (to be denoted ‘write access’)<sup>4</sup> can be interpreted as a form of ‘*institutionalized*’ trust towards prospective participants.<sup>5</sup> The italics are employed in order to stress two particular points. On the one hand, the gesture is an institutional one: we are dealing here with the ways in which an institution approaches the members it depends on, not with interpersonal trust. On the other hand, the gesture embodies the presumption that prospective participants are willing to contribute content with good intentions and to the best of their capabilities. Their trustworthiness in terms of moral intentions and capabilities is taken for granted. Notice that different capabilities are involved across the various communities. Social news sites aim to stimulate lively debates about current issues. Therefore, the required capabilities concern being able to take a normative stand and provide supporting arguments. Mastering the rules for discussion (rhetoric) is vital. Encyclopedic projects, on the other hand, aim to produce state-of-the-art knowledge about issues that are deemed relevant. Hence, participants should have adequate capabilities to contribute knowledge; their epistemic qualities are sought after. Citizen journals occupy a position in between. They are looking for both kinds of capabilities, since journal articles usually require both reporting of the facts and commenting on them. Good news is a judicious blend of fact and opinion.

That trust is at issue here can easily be seen from the fact that all the communities concerned are exposing their respective repositories of content and *entrusting* them as it were to the whims of the masses. They have decided to rely fully on their volunteers, thereby making themselves vulnerable and taking risks. Discussion sites, published news reports and encyclopedic entries can easily be polluted and spoiled by all kinds of disruptive actions. As Wikipedia defines the matter, ‘cranks’ may insert nonsense, ‘flamers’ and ‘trolls’ may enjoy fomenting trouble, ‘amateurs’ may ruin factual reporting, ‘partisans’ may smuggle in their personal opinion where this is inappropriate, and ‘advertisers’ may just try to promote their products anywhere (<http://en.wikipedia.org/wiki/Wikipedia:RCO>). Repositories polluted in this way undermine the viability of any community, and necessitate laborious cleanups.

Given this gesture of fully trusting potential participants and giving them write access accordingly, what mechanisms of trusting others may be relied on in the process? What processes possibly lie behind it? In the sequel I discuss three well-known mechanisms to handle the trust problem: the assumption, inference, and substitution of trust. Subsequently, I argue for a fourth mechanism that seems to have been neglected in the literature thus far: *backgrounding* trust. In this approach the gesture of full trust is underpinned by developing support mechanisms in the background that render the trust-as-default rule rational in a reductionist way.

First and foremost, the trust involved may be the simple *assumption* that the crowds are trustworthy. Trustworthiness is assumed without any particular evidence to support that assumption. This assumption is not made without reason; its rationale, as observed by Gambetta decades ago, is that precisely by acting *as if* trust is

---

<sup>3</sup> These communities will serve as cases to be analysed further on in this article. Note that while they practiced unrestricted and immediate access from the outset, some of them recently have been pondering—or actually resorted to—more restrictive editorial policies: filtering before reception (to be commented on below).

<sup>4</sup> This term is in use among developers working together on open source software. As a rule, anyone may access the site and inspect the contents (‘read access’). When participants have proven their skills, they may acquire the additional right to directly contribute code to a project’s source code tree: they have obtained ‘write access’.

<sup>5</sup> Note that such usage of the term ‘trust’ as exercised by institutional actors harks back to Alan Fox, who proposed to interpret organizational regimes (‘work role patterns’) in terms of the amount of trust granted by organizations to their members (Fox 1974). In this instance the institution is neither conceived of as a *producer* of trust (‘institutional-based trust’; Zucker 1986), nor as an *object* of trust (‘system trust’; Luhmann 1979). Instead, the analysis casts the institution in the role of a trusting party, a ‘trustor’.

present, one may actually produce it in the process (Gambetta 1988). In Luhmannian terms: the gesture of trust creates a normative pressure to respond likewise. The act of trust can thus be seen as an investment that it is hoped will pay off (Luhmann 1979). Can any good reasons be advanced for the assumption? What mechanism may be argued to underlie the said normative pressure?

In line with Luhmann, Pettit (1995) argued that esteem is the driving force. Since people are sensitive to the esteem of others, they will answer an act of trust with trust as it enables them to reap the esteem that is being offered to them. As elaborated before (de Laat 2010: 332), this interpretation of the normative force of trust does not seem wholly convincing in the case of open-content communities. While esteem surely is a driving force, it would seem to be an underlying one, not a paramount one. A more forceful interpretation obtains if we move away from this calculating conception of as-if trust to another conception that is based on a vision of and hope in the capabilities of others. As argued by McGeer (2008), showing trust may be rooted in the hope of challenging others to apply their capabilities in return. These others are not manipulated but empowered to show their capacities and further develop them. The trusting party puts his/her bets on a utopian future.<sup>6</sup> Such reasoning can in a straightforward fashion be applied to our open-content communities since the capabilities that are the cornerstone of this McGeerian vision have quite specific connotations here. By granting unrestricted and immediate access, crowd members are challenged to show their capacities of commenting, reporting news, or contributing reliable knowledge. They are invited to fulfil the promise of a community of exciting, newsworthy, or encyclopedic content.

A second way to handle the tensions that a trusting gesture generates is to *infer* trustworthiness. One looks for indicators that inspire confidence in the other(s) as a trusted partner: perceived individual characteristics like family background, sex, or ethnicity, belonging to a shared culture, connection(s) to respected institutions, or reputation based on performance in the past (this argument can be traced back to Zucker (1986). Moreover, the calculative balance of costs and benefits may seem to preclude a non-cooperative outcome. As argued before (in de Laat 2010: 330-31), I do not believe that an open-content community operating in cyberspace has many reliable indicators to cling to. Virtual identities are always precarious; the anonymity of contributors only aggravates this problem. Even the common requirement to register and choose a user name (or even disclose one's real name) hardly alleviates the problem (cf. de Laat 2012a). Moreover, contributors often just enter and leave, precluding any stable identity let alone the formation of a reputation. To sum up: signalling trustworthiness cannot be implemented in a reliable way. While the inference of trust has rightly been regarded a central component of processes of trust formation in *real life*, I do not think it has much value in the virtual surroundings of open-content communities.<sup>7</sup>

A third way to handle the problem of trust may be referred to as the *substitution* of trust. Wherever people interact continuously and some kind of community emerges, rules, regulations, and procedures tend to be introduced. Often these enact restrictions on behavioural possibilities. As a result, reliance on participants' wisdom and judgment in contributing is reduced; their actions become less discretionary. As a corollary, the need to grant them trust is lessened; the problem of trust is partly eliminated. The introduction of a bureaucratic structure of the kind effectively substitutes for the need to estimate—or assume—participants as being trustworthy. Below, evidence is presented on some of our open-content communities recently instituting restrictive rules and regulations: filtering incoming content prior to publication. Write access thus becomes circumscribed and regulated.

However, a fourth mechanism to deal with the tensions of an all-out policy of trust is to be distinguished. It embodies efforts, in the absence of reliable inference, to create a middle road between relying on the normative power of trust on the one hand, and (partly) eliminating the problem by substitution on the other hand. In this approach the default rule of all-out trust is kept intact by underpinning it in the background with corrective mechanisms that contain the possible damage inflicted by malevolent and/or incapable contributors. To my knowledge, this approach, to be referred to as *backgrounding* trust, has been neglected in the literature up to the present. As we will see, the supportive mechanisms themselves are not unknown, but their corrective function for keeping the default rule of trust intact has largely gone unnoticed.

### 3 Backgrounding Trust

---

<sup>6</sup> McGeer uses the term 'substantial' trust, as opposed to the shallow trust Pettit is supposed to refer to. I prefer to avoid the former term since, in my view, not another type of trust is being defined, but just a different mechanism for generating trust *ex post* that actors may supposedly rely on *ex ante*.

<sup>7</sup> To be fair, though, it should be remarked that in many virtual *trading* communities reputation systems have been built that do provide more solid grounds for inferring trust.

I propose that several types of backgrounding can be distinguished (to be elaborated below in further detail). First, a cultural offensive can be launched to curb potential vandals: legal terms of use and an etiquette of sorts that defines proper behaviour are developed and propagated. Secondly, these standards of behaviour can be underscored by defining sanctions and disciplinary measures. Participants who deviate too much from the ground rules for constructive cooperation may be punished and ultimately expelled from the community. Thirdly, structural schemes can be introduced that aim to guarantee the quality of the community's contents. These range from relatively simple vandalism patrol schemes up to voting and quality enhancement programs. The bottom line for all three activities is that they may—at least partly—contribute to sustaining the rationality of the decision to maintain an editorial policy of all-out trust. They serve to keep the default rule of full trust in place.

### 3.1 Legal Terms and Etiquette

As a consequence of their full-trust write access policy, our open-content communities are quite vulnerable to disruptive behaviour, from posting illegal content to vandalist actions. As a way of defence they are first of all trying to lay down legal guidelines. Plagiarism, libel, defamation, illegal content and the like are strictly forbidden. This is considered the baseline for proper behaviour since deviations from them would land the site in legal trouble.

Interestingly, though, our communities under study also promote 'good manners' *beyond* these legal terms of use. An etiquette is formulated for regulating mutual interactions on their sites. Leaving Wikinews and Wikipedia aside for the moment (see below), all of them stress the same kind of exhortations in their 'community guidelines', 'house rules', 'netiquette', or 'rediquette'—albeit to varying degrees.<sup>8</sup> On the positive side, members are urged to always remain respectful, polite, and civil; to stay calm; to be patient, tolerant, and forgiving; to behave responsibly; and/or to stay on topic at all times. On the negative side, the list of interdictions is much longer. One is urged to refrain from calling names, offensive language, harassment, and hate speech. Flaming and trolling are sharply condemned. Commercial spam and advertisements are declared out of bounds. Flooding a site with materials that are offensive, objectionable, misleading, or simply false only amount to an objectionable waste of the site's resources (nicknamed 'crapflooding').

Finally, let us consider Wikinews and Wikipedia. Both under the umbrella of the Wikimedia Foundation, they have adopted virtually the same etiquette (called: Wiki etiquette). It is in fact the most extended set of rules for polite behaviour in open-content communities to be found anywhere on the Net. Assuming good faith on the part of others—and showing it yourself—is the starting point. Help others in correcting their mistakes and always work towards agreement. Remain civil and polite at all times: discuss and argue, instead of insulting, harassing or personally attacking people. Be open and warm. Give praise, and forgive and forget where necessary. Overall, several pages are devoted to the subject (<http://en.wikinews.org/wiki/Wikinews:Etiquette>; [http://en.wikipedia.org/wiki/Wikipedia:Wiki\\_etiquette](http://en.wikipedia.org/wiki/Wikipedia:Wiki_etiquette)).

### 3.2 Enforcement

Both legal rules and etiquette cannot operate without some mechanism of enforcement. With all the communities above, without exception, sanctioning of deviant users has become the normal state of affairs. Users who (repeatedly) flout the rules of etiquette—let alone the legal rules—can be banned from the community for some period of time, or even forever. As a rule the professional editors employed by the site ('editorial team') simply assume these judicial powers themselves. With others, site volunteers are entrusted with the task. At h2g2, these are appointed for the job (as 'moderators') by the staff of the company which owns the site (formerly the BBC). The pair of Wikipedia and Wikinews appoints candidates with a procedure that relies on public consultation of the community ('administrators'). Citizendium does likewise ('constables').

---

<sup>8</sup> The observations that follow are based on a range of sources: [digg.com.tos](http://digg.com/tos), <http://www.reddit.com/help/reddiquette>, <http://slashdot.org/faq>, <http://ycombinator.com/newsguidelines.html>, [http://www.nowpublic.com/newsroom/tips/fine\\_print/flaming\\_policy](http://www.nowpublic.com/newsroom/tips/fine_print/flaming_policy), [http://www.nowpublic.com/newsroom/community/code\\_of\\_conduct](http://www.nowpublic.com/newsroom/community/code_of_conduct), <http://digitaljournal.com/article/179808>, <http://www.groundreport.com/content.php?section=editorial>, <http://en.citizendium.org/wiki/CZ:About>, <http://h2g2.com/dna/h2g2/A901838>, and <http://h2g2.com/dna/h2g2/A87523211>.

The mechanisms of rules and sanctions taken together send the message: respect legal terms of use and be civil and polite—otherwise risk expulsion. Notice how these may impact on the employed policy of unrestricted and immediate access. That policy assumes the trustworthiness of the participants from the outset. Inculcating respect for legal issues and rules of etiquette then may serve to *create* trustworthiness where it is found to be lacking—*afterwards*. Whenever the assumption of trustworthiness appears unwarranted, that defect can (at least partly) be repaired afterwards. As a result, the full write-access policy is underpinned and can possibly remain in force after all. ‘Backgrounding’, as I shall call this phenomenon, keeps confidence in full-trust as the default intact.

I would argue, however, that these mechanisms can do just so much. They can only possibly ‘educate’ participants who are staying longer. Newcomers, who are the most likely source of mischief, can hardly be supposed to have read let alone internalized the rules involved upon entry. As a result, the campaign for legal and civil conscience has no effect on them, and the full-trust policy remains vulnerable to their abuse. Therefore we now turn to structural means that may support the full-trust policy. No longer the dispositions of people but the contents they actually contribute come into focus. I shall argue that these tools are ultimately able to do a more powerful job of sustaining that policy.

### 3.3 Quality Management

The term ‘quality management’ is used with quite a broad meaning: it refers to both *rating* and (for dynamic entries) *raising* the quality of contributed content, throughout the whole quality range, from low to high. At the lower end, the mess of clearly inappropriate content that flouts basic legal terms of use or etiquette has to be cleaned up. Beyond these tasks of ‘basic cleaning’ (as I shall label them) the quality of the content—as far as it has passed the former test of scrutiny—can be monitored continuously and (in case of dynamic content) raised ever further. Such quality schemes may already be the normal *modus operandi* (cf. the wiki format); they may also be developed as additional mechanisms by the communities involved since they consider their basic mode to be an insufficient guarantee of quality.

**Social News Sites and Citizen Journals.** Social news sites and citizen journals (apart from Wikinews) are usefully treated together since all operate in the ‘contributing 2.0’ mode. These solicit stories (whether existing—for social news sites, or newly composed—for citizen journals) and comments on them. The tasks of basic cleaning are performed (afterwards) by the editorial teams involved: they scout their sites continuously for illegal and inappropriate content. Usually, site visitors are also solicited to report ‘violations’. Any content of the kind—whether illegal content, flooding, spamming, advertising, hate speech or abusive language—is immediately dealt with and deleted; those who posted them are reprimanded or, after repeated violations, banned from the site.<sup>9</sup> Such basic cleaning can however just achieve so much: the quality of the contents *above* the baseline of appropriate content remains an issue.

In order to tackle this thornier problem these sites have pioneered a novel approach: stories and comments can be *voted on*, usually as either a plus or a minus. As a rule, all users are entitled to vote. Note, though, that some communities require registration, and in Slashdot the right to vote obtains for a limited amount of time only. Let me elaborate these schemes. Digg has pioneered ‘digging’: if a user ‘likes’ the content, it is digged (+1), if (s)he ‘dislikes’ it, it is buried (-1). GroundReport has adopted the very same scheme. Reddit, Hacker News, and Slashdot use the more neutral wording of voting for the process: a plus if entries are found to be ‘helpful’, ‘interesting’, or ‘constructive’, a minus if they are not. Finally, NowPublic and Digital Journal only allow plus votes, for articles deemed ‘newsworthy’.

The sum total of votes then determines the *prominence* of articles on the site. By default, stories (on the front page) and comments on them (below each story) are displayed in chronological order of submission, with the most recent ones on top. Entries thus have a natural rate of decay. Voting data, fed into one algorithm or another, then force the liked items to remain longer on top of the page (countering natural decay), while at the same time forcing the disliked items—at least as far as ‘dislikes’ are part of the scheme—to plunge down the page quicker (accelerating natural decay).<sup>10</sup> Slashdot uses a slight variation: with vote totals for items being limited to the range -1 to +5, readers can choose their own personal threshold level to determine whether items become *visible* to them or not when they enter the site. Thus articles of bad repute are no longer punished by being pushed down the page, but by being ‘deleted’ for all practical purposes.

---

<sup>9</sup> In Reddit, those who start a ‘subreddit’ usually are awarded the same powers for their particular subreddit.

<sup>10</sup> Some basics of these algorithms are elaborated in <http://www.seomoz.org/blog/reddit-stumbleupon-delicious-and-hacker-news-algorithms-exposed>.

**Encyclopedias and Wikinews.** The remaining communities in my sample operate in proper ‘co-creation 3.0’ mode (Wikinews and encyclopedias). They also resort to basic cleaning concerning illegal or inappropriate content; in addition they have introduced elaborate quality schemes that go beyond simple voting. Let me start with h2g2 that does not use the wiki format, but just old-fashioned commenting. Tasks of basic cleaning are executed by the aforementioned volunteer ‘moderators’ (as appointed by the owner). As they phrase it, someone has to ‘clean the flotsam’. In addition, they decide on banning users who are found to be in violation. Higher up the quality scale, authors may strive for their article to appear in the ‘edited guide’. To that end, it has to be put up for public review, be recommended by a ‘scout’, and edited by ‘subeditors’. Notice that these two roles (volunteer roles one has to apply for) are intended to support authors, as opposed to control them. They are urged to operate as ‘first among equals’.

Citizenium, Wikipedia, and Wikinews have the wiki mode of production in common. This wiki is *the* place to carry out basic cleaning of illegal and inappropriate contents. Users are always on the alert regarding the content, allowed to immediately correct new edits in the wiki, and invited to ‘report’ any transgressor to the authorities concerned (constables and administrators respectively). The three communities have quite similar procedures as well for identifying and promoting high quality content (apart from normal ‘wikiing’). In Citizenium an entry may gain the status of ‘approved’. To that end, an appointed moderator (denoted ‘editor’) has to give his/her approval. This role incumbent is also to exercise ‘gentle oversight’ concerning matters of evolving content. So here again, as in h2g2, a non-authoritarian role, a ‘primus inter pares’. Wikinews and Wikipedia, on their part, elaborated wholly public procedures for entries to gain the status of ‘good’ article, or even ‘featured’ article: an article that meets ‘professional standards of writing, presentation and sourcing’ ([http://en.wikipedia.org/wiki/Wikipedia:Featured\\_article\\_criteria](http://en.wikipedia.org/wiki/Wikipedia:Featured_article_criteria)). As a preliminary step towards acquiring such a status an entry may be put up for public ‘peer review’ first.

Wikipedia in particular, though, over time has come to develop *additional* efforts of quality management that supplement the basic wiki mode of production. The most extended quality-watch program anywhere in our communities is to be found here. It revolves around a kind of permanent mobilisation of Wikipedians who are invited to focus their energies on quality enhancement. In their fight against ‘vandalism’ basic cleaning is high on the agenda. Users can maintain personal ‘watch lists’: listed entries are kept under surveillance for new edits coming in. ‘New Pages Patrol’ is a system for users to scan newly created entries for potential problems right after they are submitted. Furthermore hundreds of software bots have been developed for the purpose. After severe testing and public discussion within the Wikipedian community, these may be ‘let loose’ on a 24 hours basis. A famous example is Cluebot, which is instructed to intervene whenever suspicious words are inserted (‘black lists’) or whole pages deleted (<http://www.acm.uiuc.edu/~carter11/ClueBot.pdf>). The ‘new generation’ CluebotNG operates along quite different lines: as a neural network. The bot has to be fed with both constructive and vandalist edits. By interpreting those data it is hoped that it will learn in the long run to correctly diagnose instances of vandalism ([http://en.wikipedia.org/wiki/User:ClueBot\\_NG](http://en.wikipedia.org/wiki/User:ClueBot_NG)).

Close watch also extends beyond the issue of vandalism. Wikipedian pages and articles are under constant surveillance whether they should be kept, deleted, merged, redirected, or ‘transwikied’ (meaning: transferred to another Wikimedia project). More importantly, in order to raise the quality of entries further, ‘WikiProjects’ (with subordinate ‘taskforces’) are formed in which people focus on specific themes (such as classical music or Australia). Each project takes the relevant entries under its wing and promotes improvement. In particular they are entrusted with the task of grading the articles in their purview by quality (7 degrees, the highest being featured and good, cf. above) and importance (4 degrees) ([http://en.wikipedia.org/wiki/Wikipedia:WikiProject\\_Council/Guide](http://en.wikipedia.org/wiki/Wikipedia:WikiProject_Council/Guide)).

Last but not least, tools are made available to users which allow judging the credibility of entries from their revision histories. The WikiDashboard displays the edit trends of an article, and the editing activities of the most active contributors to it (Suh et al. 2008). Furthermore, the WikiTrust extension colours words in an entry according to their ‘age’, as proxy for their credibility (Adler et al. 2008). The colour chosen is orange: the ‘younger’ the text, the darker the orange it is rendered in. Contributors to Wikipedia may use these indicators for focussed quality enhancement of entries.

**Intensity of Quality Control.** Before embarking on a discussion of the relationship between measures of quality control and trust, let me first put them in a comparative perspective across the whole range of the open-content communities under study. Legal rules and etiquette (3.1 and 3.2) seem to be emphasized throughout, in about equal measure. This stands to reason, since these revolve around behavioural norms of trust and respect which are universally applicable to all communities of open textual content. Not so however for quality management

efforts: these are clearly intensifying if we move towards the encyclopedic end of the range. For one thing, patrolling for improper content is increasing. For another, voting schemes make way for a variety of teams that focus on quality within the wiki mode. Why this more intense mobilisation?

I want to argue that this is mainly due to the different types of content involved. Social news sites aim to foster discussions; an exciting exchange of *opinions* is what they are after. These discussions, moreover, have a kind of *topicality*—in the long run their importance simply fades away. To that end, a ‘contributing 2.0’ mode is sufficient. In order to guarantee quality in this mode, scouting for inappropriate content combined with voting schemes is good enough: good discussions will remain in view (longer), while bad discussions will disappear out of sight (quicker). The natural tendency for time to produce ‘decay’ is intensified. To citizen journals, furthermore, similar arguments apply.

Encyclopedias, however, aim to render the ‘*facts*’ about particular matters. Such entries cannot be produced in one go, but have to evolve over time. Moreover, such entries are to remain *permanently* visible, ready to be consulted. For the purpose, ‘co-creation 3.0’ is the preferred mode: Wikipedia, Wikinews, and Citizendium have chosen the interactive wiki format as their mode of production (which does not necessarily have to be so: h2g2 prefers a ‘contributing’ approach). Obviously, such a dynamic mode is susceptible to disruptions. Watching over quality therefore becomes a more urgent and permanent task. For that purpose, the wiki is turned into a space of intense patrolling and quality enhancement efforts.

**Backgrounding Trust by Quality Control.** After this assessment of quality management efforts across our sample of open-content communities, their connection with the default rule of full trust concerning write access finally remains to be specified. To what extent may this institutionalized trust be said to be ‘backgrounded’ by quality control? As far as this control is concerned with basic cleaning tasks, there *is* a connection. Scouting for inappropriate or outright vandal contributions—whether inside a wiki or not, whether by special volunteer patrol teams or the editorial team only, whether by humans or bots—combined with appropriate corrective action and disciplining of transgressors, is a contribution to keep the policy of full write access viable. Since disruptive contributions can always be sifted out afterwards, the gates may remain open to all. ‘Backgrounding’ of the kind may effectively allow unrestricted and immediate write access to remain the default. The same may apply to voting schemes in order to push high quality articles to a prominent and/or visible position (social news and citizen journals). To the extent that the communities involved consider it a basic aspect of quality that contributions display a minimum amount of decency and relevance, such schemes do contribute to keep their practice of full write access intact.<sup>11</sup>

However, remaining efforts under the rubric of quality control—which aim to promote really high quality—are *not* likewise connected to trust: the efforts to promote articles to the ‘edited guide’ (h2g2), to develop ‘approved’ articles (Citizendium), or to produce ‘good’ or ‘featured’ articles (Wikinews, Wikipedia). These on-going initiatives cannot be considered to support the institutional trust exhibited. Instead, rather the reverse applies: they profit from and thrive on this policy of full-write access for everybody, since it solicits a maximum inflow of contributions.

## 4 Discussion

As regards quality management (3.3) critics may object that the relevant rules, regulations, and procedures cannot neatly be sorted into those that either substitute or background trust (or, in reverse fashion, profit from it); they are just variations on the same theme of concern for quality that only differ in the temporality of their application. I would argue, however, that the distinction is sound and important. My argument proceeds along the following lines.

On the one hand, schemes for quality control can aim directly at the discretion of participants and reduce it (e.g., filtering). This reduction of discretion by definition leaves less-than-full-trust to participants. As a corollary, hierarchical distinctions among participants need to be defined (such as determining who is entitled to carry out

---

<sup>11</sup> Backgrounding trust in this fashion has an analogue in the epistemology of testimony. A default rule of accepting speakers’ utterances as true (under normal conditions) may be adhered to for non-reductionist, a priori reasons (cf. the acceptance principle). Reductionist reasoning though may also support the default rule: background evidence from our testimonial practice (like truthfulness as the norm, or reputations and sanctions) is considered to provide sufficient reasons for acceptance (for all this cf. Adler 2006). Note though that in the classic epistemological case, backgrounding has to do with the perception of mechanisms that operate within the community of speakers who *send* the messages. In our case, it has to do with the active creation of filtering and grading mechanisms within the community of readers who *receive* the messages.

filtering, and who is to be subjected to it).<sup>12</sup> If so, some amount of bureaucracy proper has been introduced into the community. Note finally, that the substitution of trust as effectuated is precisely the intention of such schemes. On the other hand, measures of quality control can also buttress policies of write access for all (e.g., scouting and patrolling for vandalism, whether by humans or bots; voting schemes). Institutionalized full trust remains a viable option because of the ‘damage repair options’ that are unfolding. As long as these schemes take care to mobilize the whole community, they can avoid introducing hierarchical distinctions. Furthermore, the supporting effect on institutionalized trust towards participants is more properly a side effect; the main focus of such campaigns is quality overall. Obviously, besides these two categories, quality management initiatives can be discerned that do *not* likewise touch upon our issue of institutional trust. The above mentioned quality rating schemes are cases in point: they more properly thrive on the full-trust-policy.

The contrast can best be captured in terms of the trust assumptions embodied in the various write access policies involved. In the case of patrolling new inputs and new contributors and of voting schemes (as well as quality watch schemes more generally), the assumption of full trust of potential participants is left intact and untouched. The default remains: ‘we trust your inputs, unless proved otherwise.’ In the case of filtering which reduces the trust offered, this default is exchanged for quite another one: ‘we can no longer afford to trust your inputs, and accordingly first have to check them carefully.’

In line with the above I want to underline that backgrounding trust in open-content communities is very important for their functioning. The mechanism allows the full-trust write access policy to remain in force. By the same token, other available mechanisms to manage the trust problem do *not* have to be resorted to. In particular, the substitution of trust by installing bureaucratic measures can be avoided. Before elaborating this point let me first provide some examples of steps towards bureaucracy as considered or actually taken by our communities. The Slashdot editorial team routinely scans incoming stories and only accepts the ‘most interesting, timely, and relevant’ ones for posting to the homepage (<http://slashdot.org/faq>). Furthermore, since 2009, Now Public and GroundReport filter incoming news before publication (<http://www.pbs.org/idealab/2009/06/citizen-journalism-networks-stepping-up-editorial-standards158.html>). With the former, first articles from aspiring journalists are thoroughly checked by the editorial team; subsequent ones may go live immediately and are only checked afterwards (<http://www.nowpublic.com/newsroom/community/faq>). With the latter, the site’s editors have to give their approval to all proposed articles prior to publication. Only reporters with a ‘strong track record’ have full write access (<http://www.groundreport.com/info.php?action=faq&questionID=1>). In the Wikimedia circuit, finally, proposals for checking incoming edits for vandalism before publication have been circulating for several years; only after approval are edits to become publicly visible. Such review is to be carried out by experienced users. In this fashion, evidently, trust in newcomers gets restricted. The proposal is actually in force in a number of their projects from 2008 onwards: Wikipedia and Wiktionary (German versions), as well as Wikinews and Wikibooks (English versions).<sup>13 14</sup>

Why then would it be important to avoid bureaucracy? The answer is that such measures may meet a chilly reception and cause unrest and trouble among community members. A conspicuous example of such unrest is the heavy contestation of the system of reviewing edits prior to publication (called Flagged Revisions) in English Wikipedia: the proposal encountered fierce resistance and finally had to be abandoned (cf. de Laat 2012b). Community members may simply detest bureaucratic rules and threaten to withdraw their commitment accordingly. That is why backgrounding trust is such an important mechanism.<sup>15</sup> Note also in this context the conspicuous role of software bots in Wikipedia. These have been and still are very active in detecting vandalism—often ahead of flesh and blood patrollers. The home page of Cluebot is full of ‘barn stars’ from co-Wikipedians, awarded since the bot had detected vandal edits before them, in just a few seconds. Reportedly it

---

<sup>12</sup> Cf. by way of analogy the common distinction between developers and observers in open source software projects.

<sup>13</sup> The proposal is also in force in several smaller language versions other than English, German, or French (cf. [http://meta.wikimedia.org/wiki/Flagged\\_Revisions](http://meta.wikimedia.org/wiki/Flagged_Revisions)).

<sup>14</sup> In our sample it is editorial teams (social news sites, citizen journals), moderators (h2g2), constables (Citizendium) and administrators (Wikipedia, Wikinews) who hold the powers to clean up messy content and/or to discipline members. Obviously, these power holders also represent bureaucracy—the difference from the filtering measures mentioned being, that no community members seem to be opposed to such a baseline of bureaucracy.

<sup>15</sup> Note in this respect how some of our communities try to bolster the quality process by introducing specific supportive roles that are intended as ‘prime among equals’ (cf. ‘editors’ in Citizendium, and ‘subeditors’ in h2g2). Their intention is clearly to *avoid* introducing hierarchical relations in this fashion. But trying to operate as such a ‘primus’ is walking a tight rope: in his/her performance, the role occupant may easily come to be perceived as an ordinary boss.



identifies, overall, about one vandalist edit per minute (over a thousand per day). Due to Cluebot and its like, Flagged Revisions were not inevitable and the plans could be shelved.

Recently both Simon (2010) and Tollefsen (2009) asked themselves the question: can users rely on Wikipedia? In their affirmative answers they pointed to editorial mechanisms in place that may ensure *high quality*: the wiki format with associated talk pages (Simon 2010: 348), and the procedure for acquiring ‘good’ or ‘featured’ status (Tollefsen 2009: 22). My question has been a slightly different one: can Wikipedia trust their users and grant them unrestricted and immediate write access? No wonder my answer—though equally affirmative—turns out to be slightly different. Contributors can fully be trusted since swift procedures to filter *low quality* submissions afterwards are in place. In complementary fashion, a continuous campaign among participants promotes respect for etiquette and basic rules of law.

## References

- Adler, B. T., Chatterjee, K., de Alfaro, L., Faella, M., Pye, I., Raman, V. Assigning trust to Wikipedia content. Proceedings of the 4th International Symposium on Wikis (WikiSym '08), 8–10 September 2008, Porto, Portugal. Obtained from <http://dx.doi.org/10.1145/1822258.1822293> (2008).
- Adler, J. Epistemological problems of testimony. *The Stanford Encyclopedia of Philosophy*. Obtained from <http://plato.stanford.edu/entries/testimony-episprob/> (2006).
- Dutton, W.H. The wisdom of collaborative network organizations: Capturing the value of networked individuals. *Prometheus*, 26(3): 211-230 (2008).
- Fox, A. *Beyond Contract: Work, Power and Trust Relations*. Faber and Faber, London (1974).
- Gambetta, D. Can we trust trust? In Gambetta, D. (ed.), *Trust: Making and breaking cooperative relations*, Blackwell: Oxford, pp. 213-237 (1988).
- Goldman, A.I. The social epistemology of blogging. In van den Hoven, J., Weckert, J. (eds.), *Information Technology and Moral Philosophy*, Cambridge University Press: Cambridge etc., pp. 111-22 (2008).
- de Laat, P.B. How can contributors to open-source communities be trusted? On the assumption, inference, and substitution of trust. *Ethics and Information Technology*, 12(4): 327-341 (2010).
- de Laat, P.B. Open source production of encyclopedias: Editorial policies at the intersection of organizational and epistemological trust. *Social Epistemology*, 26(1): 71-103 (2012a).
- de Laat, P.B. Coercion or empowerment? Moderation of content in Wikipedia as ‘essentially contested’ bureaucratic rules. *Ethics and Information Technology*, 14(2): 123-135 (2012b).
- Luhmann, N. *Vertrauen: Ein Mechanismus der Reduktion sozialer Komplexität*. 4th edition, Lucius & Lucius, Stuttgart (2000; originally 1968). English translation published in Luhmann, N. *Trust and Power*. John Wiley, Chichester (1979).
- McGeer, V. Trust, hope and empowerment. *Australasian Journal of Philosophy*, 86(2): 237-254 (2008).
- Pettit, Ph. The cunning of trust. *Philosophy and Public Affairs*, 24(3): 202-225 (1995).
- Simon, J. The entanglement of trust and knowledge on the Web. *Ethics and Information Technology*, 12(4): 343-355 (2010).
- Suh, B., Chi, E.H., Kittur, A., Pendleton, B.A. Lifting the veil: Improving accountability and social transparency in Wikipedia with WikiDashboard. Proceedings of the Twenty-sixth Annual SIGCHI Conference on Human Factors in Computing Systems, 5–10 April 2008, Florence, Italy. Obtained from <http://dx.doi.org/10.1145/1357054.1357214> (2008).
- Tollefsen, D.P. Wikipedia and the epistemology of testimony. *Episteme*, 6(1): 8-24 (2009).
- Zucker, L.G. Production of trust: Institutional sources of economic structure, 1840-1920. *Research in Organizational Behaviour*, 8: 53-111 (1986).