

## Learning to Trust the Crowd: Some Lessons from Wikipedia

F. Xavier Olleros  
Université du Québec à Montréal  
olleros.xavier@uqam.ca

### Abstract

*Inspired by the open source software (OSS) movement, Wikipedia has gone further than any OSS project in decentralizing its quality control task. This is seen by many as a fatal flaw. In this short paper, I will try to show that it is rather a shrewd and fertile design choice. First, I will describe the precise way in which Wikipedia is more decentralized than OSS projects. Secondly, I will explain why Wikipedia's quality control can be and must be decentralized. Thirdly, I will show why it is wise for Wikipedia to welcome anonymous amateurs. Finally, I will argue that concerns about Wikipedia's quality and sustainable success have to be tempered by the fact that, as disruptive innovations tend to do, Wikipedia is in the process of redefining the pertinent dimensions of quality and value for general encyclopedias.*

### 1. Introduction

Despite its success, Wikipedia continues to have some persistent critics, often among OSS advocates [1]. According to these critics, Wikipedia has two glaring weaknesses. First, its quality control is too sloppy and decentralized: without quality control by a select group of experts -they say- Wikipedia will never attain the credibility and legitimacy proper to a respectable encyclopedia. Secondly, it has a stubborn policy of accepting contributions from anonymous people. Such a policy -we are told- cannot but encourage 'vandalism' of articles and discourage potential expert contributors from investing time in improving the encyclopedia. Experts will naturally resent having to defend their ideas against anonymous, often obdurate, amateurs.

In this short paper, I intend to prove that these criticisms are not well founded. After showing that Wikipedia's quality control is indeed sloppier and more decentralized than that of your average OSS project, I will explain why this must be so. Next, I will show why it is wise for Wikipedia to welcome anonymous contributors. I will then argue that concerns about Wikipedia's quality and legitimacy have to be nuanced by

the fact that Wikipedia is in the process of redefining what a useful general encyclopedia should be like.

### 2. How is Wikipedia more decentralized than open source software projects?

The OSS process is driven by the interplay between two sharply differentiated constituencies: a small core of committed developers and a much larger periphery of occasional contributors [2, 3]. Software code evolves as a result of a sequence of repeated iterations between these two groups. Typically, the core strives for proper integration and insures overall coherence and functionality before launching new software releases, while the periphery contributes new code and tests new releases.

Both code production and quality control are quite decentralized in OSS projects, in as much as anyone can propose new code additions or corrections to code already incorporated in the kernel. The small cadre of core overseers, however, has irrevocable veto power as to what gets or does not get incorporated into the common platform. Thus, the decentralization of quality control in OSS projects is only relative: proposals for new code and corrections may come from all sides, but nothing becomes operational without the stamp of approval of the core developers.

Needless to say, in this crucial selection task, the core developers get a big help from the computer. In the final analysis, the 'silicon test' has to be passed: the code has to 'run' and produce adequate results. If the computer spews garbage, the verdict cannot be tampered with. This gives the selection task a heavy dose of objectivity, urgency and finality.

In addition to its basic commitment to communitarian collaboration and transparency, Wikipedia has imitated, to some extent, the core-periphery structure of OSS projects. As in OSS projects, Wikipedia's core is composed of a relatively small group of committed agents (the 'administrators') who have come to know and respect each other, while the periphery is populated by a very large number of

occasional contributors who may never interact directly<sup>1</sup>.

In my opinion, two frequent criticisms aimed at Wikipedia lack merit. Much has been made by some authors of Wikipedia's egalitarian treatment of all peripheral players [4, 5, 6]. These authors point out that in Wikipedia's periphery, authorial credentials count for nothing, since the views of amateurs are given the same weight as those of experts. But this is also true of OSS projects. A new piece of software code is not accepted because of its author's credentials, but because of its quality and pertinence. The proof is in the 'pudding' and credentials are quite irrelevant.

Likewise, Wikipedia has often been criticized for opening the door of its periphery not just to amateurs, but even to anonymous amateurs [1, 5]. But this, in itself, is neither new nor problematic. OSS projects do it all the time -fictional nicknames are almost de rigueur in that subculture- and no one is the worse for it.

Wikipedia parts ways with the OSS approach not by its egalitarian treatment of all peripheral players or by the anonymity of many such players, but by the fact that all peripheral players -experts or not- are given direct access to the common platform. In OSS projects, peripheral contributors can only make proposals for code additions and changes in the common platform. In Wikipedia, on the other hand, peripheral contributors do not need to bother with proposing changes. They can make the changes themselves, and these are immediately visible to all. Unlike in OSS projects, in Wikipedia no screening process by core players protects the common platform from vandals, spammers or clueless contributors.

Thus, what really sets Wikipedia apart from your typical OSS project is the soft power wielded by its core administrators. First, because quality control can only be done after the fact: nothing can be corrected or dismissed if it has not been publicly posted. And secondly, because the administrators' quality-control tools lack teeth. True, they can freeze problematic pages and block obstinate vandals (or at least those unwilling or unable to switch IP addresses) from accessing Wikipedia for indefinite periods of time. But they never really get to say the last word about a debated point, a persistent bias or a recurrent graffiti. There is a finality to quality-control decisions by core developers in OSS projects which is totally absent in Wikipedia. Regardless of its degree of completeness and polish, every Wikipedia article is in permanent open-ended mode.

### **3. Wikipedia's quality control can be decentralized**

As we have just seen, quality control in Wikipedia is much more decentralized than in OSS projects. In OSS projects, all contributions are sifted through the combined funnel of the core developers' expert judgement and the

---

<sup>1</sup> As in OSS projects, outside a small group of founders, Wikipedia's core players have emerged from the periphery, by dint of a dedication and competence superior to those of the average contributor.

computer test's verdict. There is no such central and final funnel in Wikipedia.

As it turns out, there is a very good reason for this, one that Wikipedia bashers tend to forget: encyclopedias are nothing like software programs. Whereas the latter are vertical assemblages, the former are horizontal assemblages. This difference has important consequences.

Being a vertical assemblage, a software program (much like a novel or a poem) requires a high degree of internal coherence; without it, it will not run. Software is therefore highly sensitive to incoherence. It is also highly sensitive to error. Even relatively minor errors can be fatal and block the program from running properly. The implication of this low tolerance for error and incoherence is that, even in highly modular programs (such as Linux, for example), overall quality control is imperative. Indeed, it is somewhat embedded: the computer test is inexorable.

Therefore, the quality control of an OSS platform cannot be decentralized, even if contributions come from many sources. You need a small and tight team of assemblers and quality controllers at the top of the chain to insure overall performance.

Being a horizontal assemblage, on the other hand, an encyclopedia has a natural modularity that makes internal coherence less imperative and keeps individual errors from becoming systemic and fatal. Being loosely coupled, encyclopedias have natural 'firewalls'. Bad quality is always local and stays local. A bad apple cannot spoil the cart. As long as only a small percentage of entries are of poor quality, they cannot possibly bring the system down.

Therefore, an encyclopedia's quality control is not as critical an issue as that of a software program, and you do not need a small and tight team of quality controllers at the top. In an online encyclopedia, quality control can be as decentralized as production.

### **4. Wikipedia's quality control must be decentralized**

In his speeches and interviews, Jimmy Wales has said repeatedly that Wikipedia is trying to put together the sum total of all human knowledge and make it freely available to all people<sup>2</sup>.

Try to picture the sum total of all human knowledge in terms of encyclopedia entries, ranked in order of search frequency. It is going to look something like figure 1. At the far left (the head of the curve), you will have general-knowledge, frequently sought entries, from DNA and the Amazon River to cholesterol and traffic signs - the kinds of topics that we all need to know something about. At the far right (the tail of the curve), you will find very specialized, and therefore

---

<sup>2</sup> For example, in a speech in Melbourne, on April 27<sup>th</sup>, 2007. (<http://www.nickhodge.com/blog/archives/1907>)

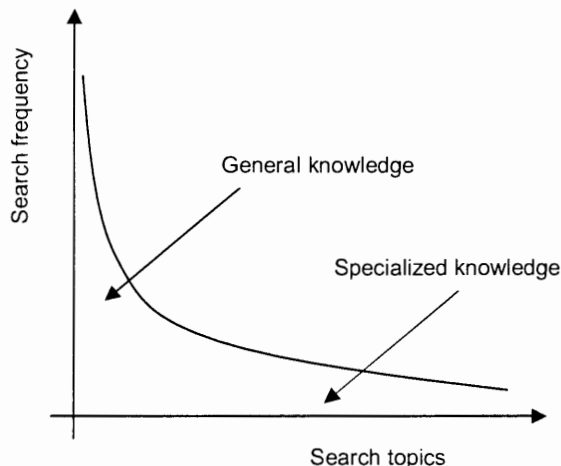


Figure 1

rarely sought entries. The head of the curve is the natural space for general encyclopedias, whereas the tail of the curve is the natural space for specialized encyclopedias.

A specialized encyclopedia, even if online and free of charge, can afford to centralize its quality control function. The Stanford Encyclopedia of Philosophy, for example, is online and free of charge, but it follows a very conventional formula: every article is written by a reputed author and then reviewed by his or her peers and proof-read by professional editors before posting. This is so partly because in a specialized encyclopedia the total number of articles is quite manageable and partly because, given the relatively narrow range of topics treated, the editorial team and their evaluators are well qualified to perform the quality control tasks.

By contrast, given the ever increasing range of topics treated in a general encyclopedia with Wikipedia's ambitious goals, top administrators are not particularly well placed to spot errors or even graffiti. They are certainly not more adept at spotting subtle biases - those can only be detected by experts in the topic at hand. This means that the team of dedicated editors to oversee a wide ranging encyclopedia project, such as Wikipedia, would have to be extremely large and diverse.

Clearly, a free-of-charge general encyclopedia has no other option but to leverage the power of the crowd maximally. This was proven by the fiasco of Nupedia, the predecessor of Wikipedia.

In March 2000, Jimmy Wales and Larry Sanger launched Nupedia, a free, collaborative encyclopedia very much like Wikipedia, but with a demanding editorial policy: every article had to be authored by a recognized expert in the pertinent field and had to survive a rigorous peer-review process. A year and several thousand dollars later, Nupedia contained only 24 finished articles. In the meantime, Wales and Sanger had come across Ward Cunningham's Wiki Wiki, a software that allowed for a

radically open approach to collaborative writing. In January 2001, they decided to launch Wikipedia as a project complementary to Nupedia -the former acting as a feeder of promising articles for the latter. In March 2003, however, in view of Wikipedia's explosive growth and Nupedia's persistent crawl, the latter project was discontinued.

Nupedia's main drawback was that by burdening the editorial team with the task of screening aspiring contributors and monitoring the review and evaluation of their work by their peers, Nupedia's designers radically undermined their project's scalability and rendered it totally impractical.

Nupedia's designers clearly underestimated the wealth of resources they could tap by opening the door to mere amateurs. But they also underestimated the transaction costs that their formula would impose on the platform's core constituency. The second of these two mistakes was the more serious one<sup>3</sup>.

In the Internet economy, in order to succeed, a new platform needs to be able to scale up and evolve fast. But in order to scale up and evolve rapidly, a platform needs to be loosely coupled and radically decentralized [7]. The Nupedia flop proved this beyond any doubt. It convinced Jimmy Wales that an online encyclopedia could not be at the same time general, free-of-charge and centralized.

## 5. The wisdom of welcoming the anonymous amateur

Why Wikipedia's insistence -so far, at least- on opening its platform even to anonymous contributors? It would seem that requiring some prior identification should deter the ill-willed rascal far more than the sensible contributor. It would also seem that Wikipedia, unlike OSS projects, could use this bit of protection<sup>4</sup>. Wikipedia's radical openness does make its pages a strong attractor to people with a destructive bend.

And yet, it is widely acknowledged that respect for contributors' anonymity has played a critical role in the precarious early stages of Wikipedia's growth [9]. Few who could contribute in that early hour resisted the pull to do so, in large part because the downside for them was close to nil. The door was wide open to them, their

<sup>3</sup> Larry Sanger, who was always uneasy about the anarchic elements in Wikipedia and left that project in January 2003, has recently argued that Nupedia's experts-only policy was not a mistake. According to him, if Nupedia had evolved into an expert-only encyclopedia without centralized editorial supervision, it might have survived and become even more useful than Wikipedia is today [8]. Sanger has in fact decided to act on this conviction and recently launched Citizendium, a new online collaborative encyclopedia moderated by recognized experts and closed to anonymous contributions. Only time will tell whether this hybrid project, part Wikipedia and part Nupedia, will be a success.

<sup>4</sup> In many OSS projects anonymity is rampant, but it is less of a problem because the platform is not open to the entire world (given the nature of the task, the pool of contributors is effectively restricted to software programmers) and because peripheral players never get direct access to the common platform.

freedom of action was maximal and their reputation was never on the line. Unconcerned about their reputation and unafraid of making errors in plain view of everybody, thousands of amateurs working away in anonymity helped to propel Wikipedia well beyond critical mass and into self-sustained growth. It is improbable that Wikipedia would have taken off so rapidly -indeed, that it would have taken off at all- had it enforced a policy of compulsory and truthful registration earlier on.

Whether or not embracing the anonymous crowd unreservedly still is optimal and will continue to be so for long is unclear. Now that its growth is self-propelled, Wikipedia could well change policy and start demanding prior identification from potential contributors. But before concluding that radical openness is now somewhat dysfunctional and that it is time to start giving more protection to the common platform, we should bear in mind four additional points:

First, Wikipedia's designers have a strong confidence in the capacity of the Wikipedia community to cope with the possible irritants of anonymous contributions. They consider the wisdom of *The Wiki Principle* ("Facilitate the correction of errors, rather than trying to block them") as applicable to graffiti and biases as to typos. Given the undeniable success of their open-door policy, the burden of proof is on the side of the critics.

Secondly, the efficacy of a universal policy of compulsory and truthful registration is less than certain. Ensuring the truthfulness of all registrations might be trivial for a specialized encyclopedia, but it would be next to impossible for an encyclopedia of Wikipedia's scale and scope.

Thirdly, there are reasons to argue that, although they can still win some battles, vandals have already lost the war. This is so because vandals are greatly outnumbered by the good guys; because they cannot inflict systemic, irreparable damage; because their victories (always local and temporary) usually go unnoticed; because the worse the graffiti, the easier it is to spot it and correct it; and because with time, the proportion of vandalized pages cannot but continue to decrease.

Lastly, as it turns out, vandals are easier to spot and their work easier to erase if they are allowed to enter anonymously. As explained recently by one of Wikipedia's top administrators, the policing task by both humans and software robots is considerably facilitated by the fact that most vandalism in Wikipedia is done anonymously. A policy change blocking anonymous contributions would only result in 100% of graffiti coming from registered users, thus making its removal slower and more laborious [10].

## 6. Redefining what a useful encyclopedia should be

In an interview for the New York Times, Jorge Cauz, Britannica's president, has dismissed Wikipedia with

these words: "Wikipedia is to Britannica as 'American Idol' is to the Juilliard School" [11].

Britannica managers seem unable to let go of their elitist worldview. This worldview led them to think themselves unassailable to challenges 'from below'. If you are Rolls-Royce, you will think that only a more luxurious car can threaten your market niche. Likewise, if you are Britannica, only an Encyclopedia written by more reputed authors may seem threatening.

Clayton Christensen has shown how dangerous this kind of mindset can be when you face a disruptive technology [12]. Typically, disruptive technologies spread and conquer not by targeting current users with a new technology that outperforms the incumbent technology along the established dimensions of value, but rather by targeting mostly non users with a new technology that underperforms along the established dimensions but outperforms along new dimensions of value. In the process, it redefines the pertinent dimensions of value. That is exactly what Encarta did in the 1990s and what Wikipedia is doing today, much to the dismay of the incumbent encyclopedias, notably Britannica.

Britannica stood, and still stands, for editorial rigor and impeccable authorial credentials. Unfortunately for them, all along, they have been overshooting the needs of most users along that dimension. Indeed, their 32-volume set had more of a symbolic and decorative value than anything else<sup>5</sup>.

In the mid 1990s, Encarta used the CD-ROM to undermine Britannica's profitable franchise. It did so by introducing an encyclopedia with barely adequate text but much more affordable, updateable online, rich in multimedia and rich also in internal hyperlinks that made navigation a breeze.

That was round one. Britannica had not yet recovered from the Encarta shock, when Wikipedia came along to continue redefining what a useful general encyclopedia is supposed to be. It has done so by improving on Britannica, and even on Encarta, along several dimensions of value (breadth of coverage, affordability, currency and internal and outgoing links), as well as by adding two new dimensions: multiple languages and a radical openness to non-contractual contributions. In addition, and quite unexpectedly, a *Google Effect* has added another critical dimension of value to Wikipedia: an ever greater visibility, thanks to the increasingly numerous hyperlinks -particularly from blogs- making reference to its articles. Users clearly feel that these advantages more than compensate for the current Wikipedia shortcomings - often low on accuracy and completeness, as well as on embedded multimedia content.

In the final analysis, Britannica has been the victim of a reductionist view of the encyclopedia, according to which its value is a simple positive function of authorial credentials and editorial rigor.

---

<sup>5</sup> According to Britannica's own market research, the typical encyclopedia was opened less than once a year [13].

This view would have been right for a specialized encyclopedia aimed at a narrow niche of experts, but it is entirely wrong for a general encyclopedia aimed at a worldwide constituency. In the world market of general online encyclopedias, several other dimensions are far more critical than rigor and credentials. Wikipedia users attest to that every day.

## 7. Conclusion

Millions of people across the world who never bought an encyclopedia and rarely consulted one are today, almost on a daily basis, consulting an encyclopedia that was launched seven years ago. This is the great achievement of Wikipedia.

In this paper, I have tried to show the wisdom of Wikipedia's radical decentralization and its openness to anonymous contributions. I have also argued that concerns about Wikipedia's quality and legitimacy have to be tempered by the fact that, as disruptive innovations tend to do, Wikipedia has redefined the pertinent dimensions of quality and value for general encyclopedias. Editorial rigor and authorial credentials, the old criteria of quality and legitimacy for general encyclopedias, are no longer pre-eminent.

## REFERENCES

- [1] R. Stross, "Anonymous source is not the same as Open Source", *The New York Times*. March 12, 2006. <http://www.nytimes.com/2006/03/12/business/yourmoney/12digi.html?ex=1148443200&en=d45d9b851758e076&ei=5070>.
- [2] C. Gacek and B. Arief, "The many meanings of Open Source". *IEEE Software*, January/February, 2004, pp. 34-40.
- [3] K. Crowston, K. Wei, Q. Li and J. Howison, "Core and periphery in free and open source software team communications". *Proceedings of the 39th Annual Hawaii International Conference on System Sciences*, 2006, Track 6, pp. 118a. <http://csdl2.computer.org/persagen/DLAbsToc.jsp?resourcePath=/dl/proceedings/&toc=comp/proceedings/hicss/2006/2507/06/25076toc.xml&DOI=10.1109/HICSS.2006.101>
- [4] L. Sanger, "Why Wikipedia must jettison its anti-elitism", 2004. <http://www.kuro5hin.org/story/2004/12/30/142458/25>.
- [5] S. Vaknin, "The six sins of the Wikipedia", 2006 [http://www.yoursdaily.com/culture\\_media/media/the\\_future\\_of\\_wikipedia](http://www.yoursdaily.com/culture_media/media/the_future_of_wikipedia).
- [6] A. Keen, *The Cult of the Amateur*, Currency, New York, 2007.
- [7] X. Olleros, "The lean core in digital platforms", *Technovation*, 2008 (in press).
- [8] L. Sanger, L., "The early history of Nupedia and Wikipedia: A memoir", in: C. DiBona, D. Cooper and M. Stone (eds.) *Open Sources 2.0: The Continuing Evolution*. O'Reilly Press, Sebastopol, CA, 2006, pp. 307-338.
- [9] J. Udell, "Wikipedia, competition and the future", *Infoworld*. January 4, 2006. [http://www.infoworld.com/article/06/01/04/02OPstrategic\\_1.html](http://www.infoworld.com/article/06/01/04/02OPstrategic_1.html)
- [10] A. Beesley, "What percentage of Wikipedia vandalism is done by registered users?", March 29, 2007. <http://wikiangela.com/blog/wikipedia-vandalism-study/>.
- [11] S. Schiff, "Know it All". *The New York Times*. July 31, 2006.
- [12] C. Christensen, *The Innovator's Dilemma*, Harvard Business School Press, Cambridge, MA, 1997.
- [13] P. Evans and T. S. Wurster, *Blown to Bits: How the New Economics of Information Transform Strategy*. Harvard Business School Press. Boston, MA. 2000, p. 4.