

Wikipedia and science publishing. Has the time come to end the *liaisons dangereuses*?

Gaëll Mainguy^{a,b,c}

^aWorld Academy of Young Scientists, Budapest, Hungary

^bInstitut Veolia Environnement, Paris, France

Abstract. Structuring information into knowledge is an important challenge for the 21st century. The emergence of internet and the diffusion of collaborative practices provide new tools with which to build and share knowledge. Scientists are seeking efficient ways to get recognition and to diffuse their work while Wikipedia is seeking well grounded contributors to shape in-depth articles. Science publishing and Wikipedia are thus profoundly modifying access to knowledge and may provide suitable conditions for a reorganization of the academic landscape.

Keywords. Science publishing, Wikipedia, open access, knowledge management

"All genuine scientific procedures of thought and argument are essentially the same as those of everyday life"(Ziman 1968).

In October 2006, Wikipedia, the multilingual, Web-based free content encyclopaedia¹ featured more than 5 million articles in 229 languages (1.4 million in English) and had entered the Alexa's top20² with more than 2000 accesses a second. Wikipedia was started as a strictly controlled, free encyclopaedia edited by PhD students, but the project failed and resulted in only a few hundred articles. Since then popularity has not really taken off amongst scientists. Last year, *Nature* surveyed more than 1,000 *Nature* authors and found that although more than 70% had heard of Wikipedia, only 17% of those consulted it on a weekly basis and less than 10% help to update it³.

¹ <http://en.wikipedia.org/wiki/Wikipedia>

² Alexa (alexa.com) is a website that computes traffic rankings by analyzing the Web usage of millions of users and thereby provides a metric of popularity

³ <http://www.nature.com/news/2005/051212/full/438900a.html>

^c Corresponding Author: Gaëll Mainguy, Institut Veolia Environnement, 15 rue des Sablons 75016, Paris, France
E-mail: president(at)waysnet.org

Despite the lukewarm interest from scientists, most of Wikipedia is accurate. In 2005, *Nature* found that the number of errors in a typical Wikipedia science article was not substantially different from Encyclopaedia Britannica's⁴. Ironically, scientists may well be one of the last communities to embrace Wikipedia's potential. Wikipedia is far ahead of all other reference websites including Library of Congress, NIH and Encyclopaedia Britannica. One figure gives the magnitude of the phenomenon: Since January 2006 Britannica's average daily page views score has been less than 1% of Wikipedia's⁵. It is also ahead of all university sites and all English language news and media sites⁶. Accordingly, it is increasingly used as a professional reference for journalist⁷ and it has been cited in several court cases since 2003⁸.

So Wikipedia is highly and increasingly popular and the dream of its founder, Jimmy Wales "*to create and distribute a multilingual free encyclopaedia of the highest possible quality to every single person on the planet in their own language*" is in a sense becoming true. However the problem of how to define and attain and control "*the highest possible quality*" remains challenging.

Quality control for Wikipedia : from collective work to Peer review

Although Wikipedia is used extensively as a reference, in its current state, it lacks the respect given to traditional encyclopaedias. As content can be added or changed at any time by anyone, Wikipedia is vulnerable to ignorance and malice and faces serious problems to be a serious reference tool. Since its inception, Wikipedia has been regularly under attack by academics as being tawdry and full of inaccuracies. A fairly detailed list of criticisms is available on the Wikipedia website itself⁹. Without a formal peer review process for fact-checking, and due to the lack of requiring qualifications to edit any article, the contributors themselves may not be well-versed in the topics they write about (not to mention the deliberate liars and vandals, see for instance the Seigenthaler case¹⁰). The site's increasing influence is regularly driving educators to crusade and this year, Jimmy Wales had to publicly warn college students that they shouldn't use it for class projects or serious research¹¹.

⁴ <http://www.nature.com/news/2005/051212/full/438890a.html> (Wiki's Wild World)

⁵ http://meta.wikimedia.org/wiki/Wikipedia.org_is_more_popular_than...

⁶ *ibidem*

⁷ A.Lih *Wikipedia and the rise of Participatory Journalism* <http://jmsc.hku.hk/faculty/alih/>

⁸ http://en.wikipedia.org/wiki/Wikipedia:Wikipedia_as_a_court_source

⁹ http://en.wikipedia.org/wiki/Criticism_of_Wikipedia

¹⁰ Wikipedia had hosted for nearly four months an article about John Seigenthaler falsely claiming that "for a brief time, he was thought to have been directly involved in the Kennedy assassinations of both John, and his brother, Bobby." See: <http://writ.news.findlaw.com/ramasastry/20051212.html>

¹¹ <http://chronicle.com/wiredcampus/article/1328/>

Quality control is thus the first issue and Wikipedia is constantly adapting its editorial strategy to improve it. Here are three examples:

- Users are encouraged to identify themselves before contributing. Those who prefer to remain anonymous are nonetheless identified and monitored by an IP tracking procedure. Users who do not comply with the editorial policy can be banned and blocked even if they never properly registered¹². In addition sensitive articles can be partially locked.¹³
- Since 2004 Wikipedia is running a reviewing procedure to certify the quality of exceptional articles. These so-called “featured articles” are reviewed for accuracy, neutrality, completeness, and style¹⁴ and are labelled with a small bronze star on the top right corner. (Bad quality articles are also increasingly labelled with warnings such as “*This article or section does not cite its references or sources.*”) So far this measure had a marginal impact: in September 2006, featured articles represented only 0.08% of all the articles in both English and French. Given the exponential growth of Wikipedia’s content¹⁵, it is likely to remain so, as the proportion of “featured article” is constantly decreasing with time¹⁶.
- Wikipedia has also started a project focused on checking facts and references using multiple sources “*to be the most cross-referenced body of knowledge.*”¹⁷

These procedures (and a lot of others), designed to increase the standards of Wikipedia’s quality, form a complex set of rules. The knowledge of these rules gives a certain authority and establishes *de facto* a hierarchy among Wikipedians. A proposal “*to give the Arbitration Committee the ability to consult Wikipedia users who are knowledgeable in subject-areas that apply to cases before them*”¹⁸ suggests that this trend is gaining ground.

The other major issue is the ephemeral existence of Wikipedia’s content. Even when an article is complete and accurate, the dynamism of Wikipedia would prevent its use for scholar publications. In principle, the history of a particular page could be searched for (on the history page of each entry) but in reality retrieving information from an old version is cumbersome and cannot be envisioned for decades. To tackle this issue, Wikipedia is investigating how to create stable versions and establish reliable sources¹⁹.

¹² http://en.wikipedia.org/wiki/Wikipedia:Banning_policy

¹³ Seigenthaler entry for instance features: “*Because of recent [vandalism](#) or other disruption, editing of this article by anonymous or newly registered users is currently [disabled](#). Such users may [discuss changes](#), [request unprotection](#), or [create an account](#).*” http://en.wikipedia.org/wiki/John_Seigenthaler_Sr

¹⁴ http://en.wikipedia.org/wiki/Wikipedia:Featured_article_candidates

¹⁵ Since 2003 see http://en.wikipedia.org/wiki/Wikipedia:Modelling_Wikipedia's_growth

¹⁶ http://en.wikipedia.org/wiki/Wikipedia:Featured_article_statistics

¹⁷ http://en.wikipedia.org/wiki/Wikipedia:WikiProject_Fact_and_Reference_Check

¹⁸ http://en.wikipedia.org/wiki/Wikipedia:Requests_for_arbitration/RFC#Alternate_solution_.239_by_mav._Content_subcommittee

¹⁹ http://en.wikipedia.org/wiki/Wikipedia:Why_stable_versions

In addition, the arrival of WebCite (webcitation.org) may circumvent this citation problem. WebCite²⁰ is a free tool that archives the content of a webpage to allow readers in the future to retrieve quoted web pages as they were at the time of the citation. Any page at any time can now be identified, archived, quoted and retrieved as a separate document. As a proof of principle, this article has been almost entirely documented using online sources and quoted pages have been cached and archived using WebCite (see reference section).

In sum, contributions are less and less anonymous, certification and refereeing procedures are increasingly complex and indefinite access to stable sources is now possible. In several ways, Wikipedia is getting closer to science production mechanisms. Concomitantly, during the last decade, science publishing has also been confronted to problems of quality control and electronic publication allows less authoritative and more collective feedbacks.

Quality control in science from Peer reviewing to collective work

Information, knowledge and science can be viewed as three polysemous worlds used to designate a collection of facts and data. The difference lies not in the number of items but on their interactions and their degree of trustworthiness. Information is purely factual and does not need to be true or false. Science, on the other side, is an organized corpus of trustworthy and verifiable facts. The word science originates from *scire*, the Latin word for “to know” and designates an organized corpus of knowledge as much as a method for building knowledge based on rational criticism, an interwoven association captured in Herbert Spencer’s aphorism “*Science is organized knowledge*”. A key aspect of the “method of science” resides in interpersonal communication. The intellectual form of scientific knowledge is determined by the absolute need for the scientist to communicate his findings and to make them acceptable to other people.²¹ One can “know” through faith, authority, intuition, or science but the scientific method allows other persons to ascertain the truth content of a given statement. Scientific knowledge is the product of a collective human enterprise to which scientists make individual contributions which are purified and extended by mutual criticism and intellectual cooperation²². The internal social relations of the scientific community are therefore all-important. *The goal of science is a consensus of rational opinion over the widest possible field.*²³

²⁰ Gunther Eysenbach and Mathieu Trudel (2005). “[Going, going, still there: using the WebCite service to permanently archive cited web pages](http://www.jmir.org/2005/5/e60/)”. *Journal of Medical Internet Research* 7 (5).
<http://www.jmir.org/2005/5/e60/>

²¹ Ziman, J. (1968), *Public Knowledge*, Cambridge, U.K.: Cambridge University Press.

²² Malhotra, Yogesh. (1994). On Science, Scientific Method And Evolution Of Scientific Thought: A Philosophy Of Science Perspective Of Quasi-Experimentation
<http://www.brint.com/papers/science.htm>

²³ Ziman, J. (1968), *Public Knowledge*, Cambridge, U.K.: Cambridge University Press.

A recent estimation reported about 24,000 peer-reviewed journals, publishing about 2.5 million articles a year, across all languages and all scholarly and scientific research disciplines²⁴. And yet, the “widest possible fields” are each steadily getting smaller as disciplines are heading towards balkanisation. This fragmentation is favourable to the reinvention of wheels, to work duplication work and to maintaining inefficient techniques, not to mention wrong usage. PubMed²⁵, as a universal platform to integrate most of the biomedical publications, proved useful in resisting and presumably reversing this trend for Life Sciences.

Mutual criticism and quality of scientific knowledge rely heavily on peer reviewing and selection of authoritative persons (for a presentation of peer review see²⁶). However, neither peer-review, nor authority, is necessary for scientific production and diffusion. First, historically the practice of editorial peer reviewing is recent and did not become general until sometime after World War II.²⁷ Second, nowadays absence of peer-reviewing is not mandatory for certain scientific communities and can even concern very influential papers. The most striking example is provided by Grigori Perelman who was recently awarded the Fields Medal²⁸ for a work uploaded on arXiv²⁹ but never published in a peer reviewed journal (see the Wikipedia entry³⁰ for a detailed presentation of Grigori Perelman’s case).

The virtue of peer review is that it is rare for an individual author or research team to spot every mistake or flaw in a complicated piece of work. Reviewing should not be more than an argumentative process where reviewers are engaging in an imaginary debate with distant authors³¹. So the same results can, at least theoretically, be achieved better, faster, more dynamically and on a permanent ground through ongoing public debate involving a large number of contributors. Indeed, several journals have recently adopted publishing policies along this line. The new journal *Philica*³² for instance, publishes articles upon reception and the review process takes place afterwards. Readers use reviews and other users’ evaluation to spot popular or unpopular work. *Naboj*³³ website provides a dynamical peer review service for users to

²⁴ Harnad, S. (2005) On Maximizing Journal Article Access, Usage and Impact. *Haworth Press (occasional column)*. <http://eprints.ecs.soton.ac.uk/10793/>

²⁵ <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?DB=pubmed>

²⁶ http://en.wikipedia.org/wiki/Peer_review

²⁷ Burnham JC. The evolution of editorial peer review. *JAMA*. 1990;263:1323-1329.

²⁸ <http://www.mathunion.org/medals/2006/>

²⁹ arXiv (<http://arxiv.org/>) is a website maintained by Cornell University that contains ca. 400’000 e-prints articles in Physics, Mathematics, Computer Science and Quantitative Biology. A majority of the e-prints are submitted to journals for publication, but some work remains purely as e-prints and is never published in a peer-reviewed journal.

³⁰ http://en.wikipedia.org/wiki/Grigori_Perelman

³¹ Open Peer Review & Argumentation: Loosening the Paper Chains on Journals
<http://www.ariadne.ac.uk/issue5/jime/>

³² <http://philica.com/>

³³ <http://www.naboj.com/>

write peer reviews of preprints from arXiv. The review system is based on Amazon ranking and users have the opportunity to evaluate articles and reviews. These experiments are still in infancy and it is hard to predict where science publishing is heading. It is nonetheless striking to note how the present situation is fulfilling the prediction Horace Freeland Judson made 12 years ago: *peer review and refereeing are not laws of nature, nor of epistemology, but social constructs of recent date (...) and (...) are likely to evolve towards a form of publication that will be a continuing open dialogue and collaboration among contributing scientists, editors, expert commentators and readers*³⁴.

Can Wikipedia and Science publishing meet half way?

We have seen that Wikipedia is not really popular among scientists. The reciprocal is also true, as illustrated by the fact that “*a call for higher academic standards*” is denoted as *Academic standards disease*³⁵ in Wikipedian jargon (this entry provides an overview of the disagreement). Another problem is that, at the time being, “*No original research*”³⁶ is one of the three content-governing policies of Wikipedia. The original motivation for this policy was to combat people with personal theories, who would attempt to use Wikipedia for their personal opinion and agenda.

Nonetheless, mechanisms for aggregation and integration of knowledge in a scholar paper and on a Wikipedia entry are converging and suggest that they could adopt similar procedures in the near future. It seems that there is no irreducible difference between a reviewing procedure of a scientific article and a discussion on a talk page of a controversial entry. A greater involvement by scientists has been advocated by Jimmy Wales and stable versions are explicitly designed to foster an environment of academic quality for serious researchers. Wikibooks are already likely to revolutionize the way textbooks are written, updated and distributed, especially in the developing world³⁷. This initiative is likely to be a resounding combination of collaborative work and authoritative figures.

Wikipedia bears tremendous opportunities to store, concentrate and build “scientific knowledge”. It may be time for scientific communities to not only recognize the potential of Wikipedia but also to think in terms of a common future. A dynamical peer review system based on comments such as the one provided by *Naboj* would not be so different from the Wikipedia policies used to validate information. More and more books are scanned and accessible online and self-archiving is gaining ground, the future of source identification and citation resides online. E-prints and e-books are building blocks of knowledge (and thus as the only pieces not accessible for modifications). Comments and citation provide alternate quality indicators to evaluate

³⁴ Horace Freeland Judson. Structural Transformations of the Sciences and the End of Peer Review JAMA. 1994;272:92-94 available at http://www.ama-assn.org/public/peer/7_13_94/pv3112x.htm

³⁵ http://meta.wikimedia.org/wiki/Academic_standards_disease

³⁶ http://en.wikipedia.org/wiki/Wikipedia:No_original_research

³⁷ <http://www.scidev.net/content/news/eng/free-wiki-textbooks-planned-for-developing-nations.cfm>

a work and prevent it from quickly sinking into oblivion. This could allow newcomers to spot important publications, integrate knowledge and structure new scientific fields. Anybody would be free to start his own monument based on those building blocks, and collaborative work will do the rest. Considering what the mathematician's community has been able to gather and organize in a few years³⁸ what could be hosted and integrated by Wikipedia seems limitless. Based on the original -thus stable- work available, Wikipedia can aggregate, display and synthesize all comments and opinions at any scale reflecting the nested organization of human knowledge. This constantly changing architecture would be able to accommodate science dynamism.

Open source systems give educators and students the opportunity to create courses according to their needs³⁹ and dynamic information is a fantastic tool to challenge the reliability of an argument and develop critical thinking. Scientists and teachers must contribute massively, and must teach students how to use Wikipedia efficiently. In his novel, '*les liaisons dangeureuses*'⁴⁰, Choderlos de Laclos depicted the tragic consequences of blind rivalry. The time may have come for Wikipedians and scientists to join their efforts and design a universal integrator of human knowledge providing *the highest possible quality to every single person on the planet in their own language*.

³⁸ http://en.wikipedia.org/wiki/List_of_mathematics_lists

³⁹ Moodle (<http://moodle.org/>) is based on a sound pedagogical principle: people construct new knowledge as they interact with each other.

⁴⁰ The complete edition of 1782 is available at http://fr.wikisource.org/wiki/Les_Liaisons_dangereuses

References :

Numbers refer to footnotes where webpages are cited.

URLs from Webcitation provide a permanent access to cached copy of cited sources.

- [1] <http://en.Wikipedia.org/wiki/Wikipedia>
Please see <http://www.webcitation.org/5JYNWhHZA>
- [2] <http://www.alex.com/>
Please see <http://www.webcitation.org/5JYNWhHZA>
- [3] <http://www.nature.com/news/2005/051212/full/438900a.html>
Please see <http://www.webcitation.org/5JYNWhHZm>
- [4] <http://www.nature.com/news/2005/051212/full/43890a.html>
Please see <http://www.webcitation.org/5JYNWhHZw>
- [5] http://meta.wikimedia.org/wiki/Wikipedia.org_is_more_popular_than
Please see <http://www.webcitation.org/5JYNWhHa6>
- [7] <http://jmsc.hku.hk/faculty/alih/>
Please see <http://www.webcitation.org/5JYNWhHaG>
- [8] http://en.Wikipedia.org/wiki/Wikipedia:Wikipedia_as_a_court_source
Please see <http://www.webcitation.org/5JYNWhHaQ>
- [9] http://en.Wikipedia.org/wiki/Criticism_of_Wikipedia
Please see <http://www.webcitation.org/5JYXfzj8Z>
- [10] <http://writ.news.findlaw.com/ramasastry/20051212.html>
Please see <http://www.webcitation.org/5JYXfzj91>
- [11] <http://chronicle.com/wiredcampus/article/1328/>
Please see <http://www.webcitation.org/5JYXfzj9C>
- [12] http://en.Wikipedia.org/wiki/Wikipedia:Banning_policy
Please see <http://www.webcitation.org/5JYXfzj9M>
- [13] http://en.Wikipedia.org/wiki/John_Seigenthaler_Sr
Please see <http://www.webcitation.org/5JYXfzj9Y>
- [14] http://en.Wikipedia.org/wiki/Wikipedia:Featured_article_candidates
Please see <http://www.webcitation.org/5JYXfzj9j>
- [15] http://en.Wikipedia.org/wiki/Wikipedia:Modelling_Wikipedia's_growth
Please see <http://www.webcitation.org/5JYYwabwF>
- [16] http://en.Wikipedia.org/wiki/Wikipedia:Featured_article_statistics
Please see <http://www.webcitation.org/5JYXfzj9s>
- [17] http://en.Wikipedia.org/wiki/Wikipedia:WikiProject_Fact_and_Reference_Check
Please see <http://www.webcitation.org/5JYXfzjA2>
- [18] http://en.Wikipedia.org/wiki/Wikipedia:Requests_for_arbitration/RFC#Alternate_solution_.239_by_mav._C
[ontent_subcommittee](http://en.Wikipedia.org/wiki/Wikipedia:Requests_for_arbitration/RFC#Alternate_solution_.239_by_mav._C)
Please see <http://www.webcitation.org/5JYXfzjAB>
- [19] http://en.Wikipedia.org/wiki/Wikipedia:Why_stable_versions
Please see <http://www.webcitation.org/5JYNWhHaZ>
- [22] <http://www.brint.com/papers/science.htm>
Please see <http://www.webcitation.org/5JbYdIIOY>
- [24] <http://eprints.ecs.soton.ac.uk/10793/>
Please see <http://www.webcitation.org/5JYNWhHai>
- [25] <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?DB=pubmed>
Please see <http://www.webcitation.org/5JYNWhHat>
- [26] http://en.Wikipedia.org/wiki/Peer_review
Please see <http://www.webcitation.org/5JYNWhHb3>
- [28] <http://www.mathunion.org/medals/2006/>
Please see <http://www.webcitation.org/5JYNWhHbC>
- [30] http://en.Wikipedia.org/wiki/Grigori_Perelman
Please see <http://www.webcitation.org/5JYNWhHbV>
- [31] <http://www.ariadne.ac.uk/issue5/jime/>
Please see <http://www.webcitation.org/5JYVJwheQ>

- [32] <http://philica.com/>
Please see <http://www.webcitation.org/5JYNWhHbn>
- [33] <http://www.naboj.com/>
Please see <http://www.webcitation.org/5JYNWhHbx>
- [34] http://www.ama-assn.org/public/peer/7_13_94/pv3112x.htm
Please see <http://www.webcitation.org/5JYNWhHc7>
- [35] http://meta.wikimedia.org/wiki/Academic_standards_disease
Please see <http://www.webcitation.org/5JYNWhHcH>
- [36] http://en.Wikipedia.org/wiki/Wikipedia:No_original_research
Please see <http://www.webcitation.org/5JYNWhHcQ>
- [37] <http://www.scidev.net/content/news/eng/free-wiki-textbooks-planned-for-developing-nations.cfm>
Please see <http://www.webcitation.org/5JYNWhHcZ>
- [38] http://en.wikipedia.org/wiki/List_of_mathematics_lists
Please see <http://www.webcitation.org/5JaEyKoSN>
- [40] http://fr.wikisource.org/wiki/Les_Liaisons_dangereuses
Please see <http://www.webcitation.org/5JYNWhHcj>