

Wikipedia, Academia, and Science

Eduard Aibar

Universitat Oberta de Catalunya
eaibar@uoc.edu

Abstract

This paper presents some results of an ongoing research project on the interactions between Wikipedia and academia. The first part of the project analyses the perception and practices of academic faculty about Wikipedia. It also includes a study on 52 teaching experiences at universities around the world where lecturers have used Wikipedia for designing assignments involving students to edit the free encyclopedia. The second part of the project, still ongoing, is aimed at analyzing the scientific content of Wikipedia, in order to see how the non-expert nature of most of its editors shapes the way science and technology issues are presented in its articles.

Introduction

Though so-called commons-based peer production (Benkler 2006) projects share many features with traditional scientific practices—peer review, open publication of results, intensive collaboration and commons-oriented production—there are also important differences regarding authorship, publishing, collaboration styles, and laypeople involvement. The present movement towards Open Science and Open Research pleads for importing some peer production mechanisms into the realm of science and academia (Nielsen 2011), but it is not clear whether both cultures and ways of knowledge production are fully compatible.

Our research project tries to explore the interactions between these two realms—science and Wikipedia—and analyze possible conflicts, tensions or mutual enrichment processes. To put it bluntly we are interested to see both (1) what scientists do and think about Wikipedia and (2) what Wikipedia does to science. The first part of the project, already completed, studies the perception and practices of academic faculty—taken as a good proxy for ‘scientists’—about Wikipedia. The second part, still ongoing, is aimed at analyzing the scientific content of Wikipedia, i.e. the subset of articles dealing with science and technology

issues, in order to see how the non-expert nature of most of its editors and the open collaborative style of Wikipedia shapes the way science and technology issues are depicted in its articles.

Wikipedia and Academic Faculty

Recent empirical studies show that Wikipedia is heavily and frequently used by a large majority of university students to carry out different assignments and tasks (Wannemacher and Schulenburg 2010). Though there are very few studies with substantive empirical data on the attitude of university faculty members about Wikipedia—see, as a recent exception, Soules (2015)—the common view is that they do not seem to be so positive: academics and scientists are thought to perceive Wikipedia with skepticism or cynicism and very few become actual editors, improving or creating articles.

Our study is based on a large online survey (913 valid responses) to all faculty members (3,639 people) in two large public universities in Catalonia, Spain (Universitat Oberta de Catalunya and Universitat Pompeu Fabra) in order to analyze their perceptions, attitudes and practices on Wikipedia. Our results show ($\pm 2.81\%$ for overall data in the case of maximum uncertainty; $p=q=0.5$; confidence level 95%) that the overall quality of Wikipedia articles is highly valued and most faculty members are also regular users, mainly for information seeking.

The results of our survey do not support, contrary to widespread opinion, an overwhelming negative or skeptical attitude among university faculty towards Wikipedia. On the one hand, not only do most of them see it as a useful teaching recourse (46.8%, while only 18.8% don’t) but few feel uncomfortable about students using it as a source of information (23.1%). On the other hand, the overall quality of Wikipedia articles is rather positively valued. From the three questions asking about quality, reliability and updating got a striking majority of positive answers—only articles’ comprehensiveness received a slightly negative evaluation. The common assumption that most faculty members perceive Wikipedia as an inaccurate and unreliable source of information is not supported by our survey.

When considering their level of use of Wikipedia both for professional and personal matters but without focusing specifically on teaching activities, faculty members show a similar behavior to that reported for students. Most of them are regular users of Wikipedia: 62.6% for personal matters and 55.3% for academic matters. In fact, our control data even show that a relatively high percentage of them (13.5%) are even registered users of Wikipedia—far beyond the average rate of registered users for the general population of Catalonia (0.4%). All in all, our survey depicts faculty members as frequent users of Wikipedia as long as passive use—information seeking mainly—is concerned.

Another remarkable finding is that private instances of use—whether professional or personal—are not matched by public uses (i.e., those that require some sort of publicly stated commitment). While frequently using it in the private sphere, most faculty members think Wikipedia is not well regarded by their colleagues as a respectable source of information (53.2%). Although they find it useful and rich, they do not tend to recommend its use to students (27.2%) and even less to their colleagues (23.1%). They are frequent users but prefer not to talk about it. Wikipedia seems to be for academics what porn is for polite dinner conversations.

In general, this deep dependence on colleagues' opinion and behavior, together with the preference for private uses of Wikipedia—those not involving public acknowledgement—create a negative feedback loop. Since colleagues do not talk much about it, most faculty members tend to think their colleagues do not use it because they find it inappropriate or unreliable and that prevents them to publicly expose their own positive opinion and intensive use. Though most respondents (N=856) do not take a clear side (40.0% choosing the midpoint), those that think colleagues do not use it much (34.0%) significantly outnumber those who think they do (26.0%).

In fact, colleagues seem to act as strong role models for most faculty members on this issue, whereas the institutional context—their own university policies and culture—seems less important. Even quality assessment of Wikipedia content seems to be dependent on their perception of colleagues' position about it (Aibar et al. forthcoming). Consistent with that, belonging to a specific area of expertise seems to be more decisive than formal institutional affiliation. Like some of the past literature (Eijkman 2010), our study finds that faculty from hard sciences show a more positive assessment and use of Wikipedia than their social sciences and humanities' colleagues. The general clash between scientific culture and peer production we had hypothesized is eventually modulated by the particular subcultures of more specific scientific disciplines. As former STS (Science & Technology Studies) works on science culture have demonstrated (Knorr-Cetina 1992), there are enormous disparities—in methods, practices and epistemological styles—between different sciences.

Science and Technology in Wikipedia

The aim of this second part of our project is to develop an analysis of the scientific content of Wikipedia, that is, the subset of articles dealing with topics of science and technology (our initial estimate is that about 10% of the 1.15 million articles of the Spanish version could belong to it), from an STS point of view.

The study is based on the fact that Wikipedia has recently become—among other things—the main platform for the public communication of science. Recent studies on communication and public perception of science agree that the Internet has become, for most people, the main source of scientific information (Brossard and Scheufele 2013; FECYT 2012). In recent years the Internet has surpassed traditional media in this regard: newspapers, radio and television. According to a study by the National Science Foundation (USA) more than 60% of citizens seeking scientific information on specific topics, turn first to the Internet while only 12% are still using the online versions of traditional media—newspapers or magazines (National Science Board 2012).

Data from the Spanish *Survey on the Social Perception of Science* (FECYT 2012) show that the Internet is also the main source of scientific information for the Spanish public—for 40.9% of respondents. For the first time in Spain, the Internet is above TV (31%) and well above the rest. When asked about the type of Internet resources used for scientific information, 21.7% say they use Wikipedia as their main source. Only blogs and social media are above, but since both include a large variety of instances, Wikipedia can actually be considered the most consulted singular source and, therefore, the most important channel for the public communication of science nowadays.

Our analysis departs from the non-trivial task of selecting the corpus of science and technology articles in the Spanish version of Wikipedia. Since Wikipedia categories cannot be used for that purpose (they are neither exclusive nor exhaustive) we have built a preliminary corpus using an algorithm that produces clusters of articles through the analysis of the links between them. We have then filtered those clusters using the 6-digit code of the UNESCO nomenclature for fields of science and technology, by selecting those clusters with an existing article matching any UNESCO field. We have finally carried out a manual selection to determine the final corpus. Since the UNESCO classification is too broad and generous, we have left out those clusters mainly devoted to places (cities, countries, etc.), arts (music, literature, etc.) or religion. The rest of social sciences and humanities have been included.

The final corpus gathers 340 clusters covering a total of 60,108 articles, which means 5% of all articles in the Spanish Wikipedia. Since this is half of our initial expectation we are right now trying to devise methods to include other clusters (we are fairly sure that almost all articles in our corpus are of a clear scientific or technological nature, but

we also know that some scientific clusters have been left out).

Our basic research question is whether the non-expert or lay character of the average Wikipedia editor is actually shaping the way science is depicted in the free encyclopedia. Several studies in the field of STS have shown how the involvement of non-standard actors in science and technology issues—whereas in formal participation procedures or in informal settings—may change the way they work and also their outcomes (scientific knowledge or technical artefacts). But STS scholarship on the participation of citizens or laypeople in science and technology matters has not paid much attention to Wikipedia, a platform deliberately designed to foster laypeople involvement in building (encyclopedic) knowledge.

This general basic question has been operationalized into other, more specific, sub questions. Are Wikipedia articles on science and technology more sensible to social issues around science and technology? Do they reflect standard scientific consensus or give some room to minority positions in science? Are the more controversial issues in Wikipedia (in science and technology matters) also those more controversial in society at large? How is the expert/lay divide managed in Wikipedia?

In order to answer those questions we are developing our research into four dimensions. First of all, we are going to explore the quality of articles through the analysis of their bibliographic references. Other authors (Nielsen 2007 for instance) have already carried out studies on Wikipedia references and very recently the Wikimedia Foundation itself has published a data set containing all references to scientific publications extracted from the English Wikipedia (Halfaker and Taraborelli 2015).

Secondly, we are analyzing the relative presence of Wikipedia articles across the different fields of science and technology and comparing it to the distribution of papers in scientific journals in standard academic data bases. Along this line, we are also going to produce a cognitive map of science and technology in Wikipedia, by exploring the links between articles in different fields, in order to compare it to existing cognitive maps of science.

Third, we would like to study the editors' profiles and their editing records, in order to see if science and technology Wikipedia editors are somewhat different from the rest. Finally, we want to pay particular attention to the talk pages and develop a qualitative analysis of controversies between editors in the most disputed articles. We are mostly interested in identifying what these controversies are about, what kind of resources are deployed, how closure is achieved, and what kind of general assumptions about the nature of science and technology are made.

Concluding remarks

A final remark can be made on the benefits of actively using Wikipedia in higher education. Though previous studies have successfully highlighted the positive educational improvement of different students' abilities, such as motivation, collaboration, critical reviewing, writing and referencing skills, much less has been said on the urgency for scientists and scholars to pay attention to Wikipedia as a new and powerful channel for the public communication of science. The widespread social use of Wikipedia as a source for scientific information—including sensitive information on medical and health issues—for the general public, should also encourage a more active and systematic engagement of professional scientists and scholars in improving this free encyclopaedia's content in science-related matters.

Nevertheless the fact is that scientists, scientific institutions—from universities to laboratories or research centres—and public bodies funding research are mostly unaware of this phenomenon. When dealing with public communication of their work they still think of traditional media or, at most, they turn to private social media in the Internet.

References

- Aibar, E.; Lladós, J.; Minguillon, J.; Meseguer, A.; and Lerga, M. (forthcoming). Wikipedia at University: what Faculty Think and Do about it. *The Electronic Library* 33(4). (Preprint available at <http://hdl.handle.net/10609/39442>).
- Benkler, Y. 2006. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. New Haven, CT: Yale University Press.
- Brossard, D.; and Scheufele, D.A. 2013. Science, New Media, and the Public. *Science* 339(6115): 40–41.
- Eijkman, H. 2010. Academics and Wikipedia: Reframing Web 2.0+ as a Disruptor of Traditional Academic Power-Knowledge Arrangements. *Campus-Wide Information Systems* 27(3): 173–185.
- FECYT. 2012. *Encuesta de Percepción Social de la Ciencia*. Report available at: http://icono.fecyt.es/informesypublicaciones/Documents/Percepción Social_2012.pdf. Retrieved Apr 11, 2015.
- Halfaker, A.; and Taraborelli, D. 2015. Scholarly article citations in Wikipedia. *Figshare*. Available at <http://dx.doi.org/10.6084/m9.figshare.1299540>. Retrieved Apr 11, 2015.
- Knorr-Cetina, K. 1992. The Couch, the Cathedral, and the Laboratory: On the Relationship between Experiment and Laboratory in Science. In Pickering, A. (ed.). *Science as Practice and Culture*. Chicago, IL: The University of Chicago Press; 113–138.
- National Science Board. 2012. *Science and Engineering Indicators 2012*. Washington, DC: National Science Foundation.
- Nielsen, F. A. 2007. Scientific Citations in Wikipedia. *First Monday* (12)8. Available at

http://firstmonday.org/issues/issue12_8/nielsen/index.html. Retrieved Apr 11, 2015.

Nielsen, M. 2011. *Reinventing Discovery: the New Era of Networked Science*. Princeton, NJ: Princeton University Press.

Soules, A. 2015. Faculty Perception of Wikipedia in the California State University System. *New Library World* 116(3/4): 213-226.

Wannemacher, K.; and Schulenburg, F. 2010. Wikipedia in Academic Studies: Corrupting or Improving the Quality of Teaching and Learning? In Ebner, M.; and Schiefner, M. (eds.). *Looking Toward the Future of Technology-Enhanced Education: Ubiquitous Learning and the Digital Native*. Hershey, PA: Information Science Reference; 295–311.