Information Product Creation Through Open Source Encyclopedias

Chitu Okoli
John Molson School of Business
Concordia University, Montréal, Canada
Chitu.Okoli@concordia.ca

Abstract

The same open source philosophy that has been traditionally applied to software development can be applied to the collaborative creation of non-software information products, such as encyclopedias, books, and dictionaries. Most notably, the eight-year-old Wikipedia is a comprehensive general encyclopedia, comprising over 12 million articles in over 200 languages. It becomes increasingly important to rigorously investigate the workings of the open source process to understand its benefits and motivations. This paper presents a research program funded by the Social Sciences and Humanities Research Council of Canada with the following objectives: (1) Survey open source encyclopedia participants to understand their motivations for participating and their demographic characteristics, and compare them with participants in traditional open source software projects; (2) investigate the process of open source encyclopedia development in a live community to understand how their motivations interact in the open source framework to create quality information products.

1. Introduction

The emergence of open source software has radically changed the technological landscape of the computing industry, affecting the strategic dynamics involved in various commercial enterprises, including proprietary software developers, hardware manufacturers, and makers of network products alike. OSS can be defined as "software that is made freely available to all" [1], and that is developed and maintained by numerous contributors who are scattered around the world, but who interact through a virtual community on the Internet. The majority of these public-good producers contribute freely to the provision of the project in return for only intangible rewards.

It has normally been believed that high management control is necessary to ensure the quality of a software or information product project. However, through flagship products such as the Linux operating system, the Apache Web server, and the Firefox Web browser, the open source model has demonstrated convincingly that the apparently anarchistic process of open source development can yield products of high quality [2].

2. Open source encyclopedias

Although the open source approach has traditionally been applied only to software products, recent years have demonstrated its applicability to the creation of other information products, most notably the open source encyclopedia (OSE). Such an encyclopedia allows virtually anyone to contribute to it, and permits free (in the open source sense) modification and distribution of its content. Wikipedia (www.wikipedia.org) is by far the most extensive and best-known OSE. In just eight years, this comprehensive general encyclopedia has compiled over nine million articles in over 200 languages (with over 2.7 million in English). Other OSEs, though not nearly as well developed, include the ISPedia (ispedia.terry.uga.edu) of the Association for Information Systems, Enciclopedia Libre Universal en Español (enciclopedia.us.es), Citizendium (www.citizendium.org), and Wikinfo (www.wikinfo.org). OSEs are based on the "wiki", a Web 2.0 technology that allows viewers to add content to Web pages using minimal technical skills. OSEs are released with some sort of license that permits the free redistribution and modification of their content. Wikipedia uses the Free Software Foundation's Free Documentation License (FDL)—the textual complement to the General Public License for traditional open source software. This license permits anyone to freely copy, modify, and distribute the encyclopedia's content under two conditions: the original source must be cited, and everyone else must be accorded the same right of free modification and distribution of all derivative works via the FDL. Other OSEs use a variety of Creative Commons licenses

(www.creativecommons.org), usually compatible with the FDL.

Open source encyclopedias are very similar in philosophy to traditional open source software, in that they aim to create high-quality digital information products through the participation of large numbers of contributors, mostly volunteers, though some are paid. They are structured to eliminate or minimize individual agendas and strive towards public or industry welfare in setting policies for the development of their products. Their organizational structures are generally rather loose, yet there is some central administration that permits the project to survive and flourish [3,4]. Anecdotal evidence abounds as to the quality of open source products, mainly by counting the number of users of such software, and by noting large corporate users for whom quality is a critical issue. For example, the British Broadcasting Corporation's news websites routinely link to Wikipedia, and *Nature* journal has found its accuracy in science articles comparable to that of the venerable Britannica [5]. However, scholastically rigorous evidence is necessary to objectively evaluate the quality of such information products.

When the value of the open source approach to creating information products is appreciated, both for software and other products such as encyclopedias, it becomes important to better understand what motivates people to participate in such altruistic ventures, and to understand the process by which this development methodology leverages people's energy and resources to create valuable products. Studies have found that individual volunteers are motivated for various reasons: personal benefit, belief in the open source philosophy, a desire to contribute socially, and ego gratification [6,7]. In open source projects, it is important to consider the performance of the individual developers who participate in a project, since most open source participants are volunteers with individual motivations for participation. Aligning performance measures with motivations for participation would be helpful for organizers and evaluators of open source projects to increase the intensity and the quality of participation.

However, the Wikipedia approach is not without its criticisms. Okoli [8] summarizes some of the key criticisms that have been levied against this radical concept and against its implementation in Wikipedia. There are doubts that the collaborative editing will result in a quality encyclopedia [9,10], and even occasional errors serious to the point of libel [11].

By looking at the final output of the open source process, it seems that somehow, the model "works". Although there have been studies that have analyzed the philosophy of open source projects and analyzed their critical success factors [1,2], there is a shortage of studies that investigate the intrinsic dynamics of the process; that is, at the individual contributor level, how the open source philosophy, policy structure, and methodology merge the motivations of volunteer contributors with the community's goals of producing high quality information products. The understanding of this process is the ultimate objective of this research program. Specifically, this research program joins the nascent body of scholarly research on the Wikipedia phenomenon [3,4], seeking to objectively and rigorously establish the quality of open-source information products compared to their traditional counterparts, understand contributors' motivations, and analyze the process of the product creation.

3. Reasons and Motivations for Open Source Development

This paper presents a research program funded by the Social Sciences and Humanities Research Council of Canada. This research program draws from multiple research domains, mainly related to open source software and wikis. Building on the literature on reasons and motivations for open source development, and on other aspects of open source development, I will study the process that contribute to effective open source development, in the context of non-software applications.

Research on open source thus far has mainly focused on several dimensions, which include the economical and sociological foundation of open source movement, motivational issues concerning open source participants, and the managerial and legal aspects of open source practice. Von Hippel and von Krogh [1] consider OSS development the "private-collective" innovative model and illustrate how this integrated framework deviates from each of the private investment and the collective action models. They believe that the "private-collective" model offers "the best of both worlds" by resolving the problems each inherently possesses, such as free revealing of innovations developed by private funds and potential free-riding phenomenon. From the perspectives of labour economics and industrial organization, Lerner and Tirole [12] examine the economics of open source initiatives. They argue that signalling incentives (i.e., career concern incentive [future job opportunity] and ego gratification incentive [peer recognition]) are the main drivers of the volunteers' participation. These authors also discuss the importance of leadership due to the unique governance structure of open source organization and competitive dynamics betweens open and commercial software development.

In open source projects, it is important to consider the performance of the individual developers who participate in a project, since most open source participants are volunteers with individual motivations for participation. Aligning performance measures with motivations for participation would be helpful for organizers and evaluators of OSS projects to increase the intensity and the quality of participation. According to the resource-based model of social structure sustainability [13], online social structures should provide members with positive benefits crucial for maintaining traditional organizations, such as opportunities to influence people [14], opportunities for affiliation or championship [15], and the ability to disseminate ideas rapidly [16]. While not absolutely necessary for participation, it is very important for open source participants to be recognized for their contributions [17]; thus they gain status and respect in the community. To better encourage participation in open source projects, it is important to understand the dynamics of recognition as an incentive, and how the attainment of this motivation serves as an indication of a participant's performance, in the sense that they have achieved a desired individual goal.

4. Research methodology

There are two major projects in this research program, which successively evaluate the quality of information products created via the open-source model, and investigate the details of how the process works. Because the Wikipedia database is fully downloadable in accordance with the Free Documentation License, most of this research can be freely conducted even without such cooperation. I already maintain a clone of the 64 GB Wikipedia website, and all the projects described below (except for the demographic survey) could be conducted without the assistance of the Wikimedia Foundation. This research program will comprise the following specific projects:

4.1. Motivations of Open Source Encyclopedia Participants

Participation in open source encyclopedia creation readily shows that non-software open content projects such as Wikipedia have a very different demographic background from the typically white male aged 25 to 45 demographic of open source software development. This different sociological makeup could result in significantly different patterns of behaviour between software- and non-software open source communities.

Thus, an important base study in this research program will involve surveying Wikipedia's community members to understand their personal, organizational, and societal motivations for participating in such a non-software open source project (many participants contribute several hours each week). The study would also obtain their demographic characteristics, which would permit comparison with members of traditional open source software communities, such as SourceForge. There would be web surveys based on questions from the open source motivation literature and based on focus groups of open source participants. A number of studies have investigated the motivations of open source community members, and have developed instruments to measure their demographic characteristics [6,13]. Using these as a foundation, appropriate instruments could be tailored to a population broader than software developers.

4.2. How the Open Source Process Converts Motivations to Quality Performance

Based on the prior study, the major phase of this program will investigate how participants' Wikipedia collaboration based on the identified motivations operates through the open source process to create information products of quality. Focusing on the *process* by which the methodology works, this phase will involve in-depth analysis.

Using social network analysis tools such as UCINET, various measures will be calculated that specify community members' relationships with each other. Algorithms will be developed for quantitatively measuring the quality of articles, and the quality of the contributors. These dependent (predicted) variables will be obtained by querying the cloned Wikipedia database. Using these group and individual quality measures, methods such as multiple linear regression, logistic regression, and partial least square analysis could be employed to determine how contributors' positions in the Wikipedia social structure in relation to other participants affects their individual and group performance in this open source community.

Wikipedia members interact via text-based discussion pages. Thus, there is an enormous corpus of recorded discussion in the database that documents the process they have followed to arrive at consensus and to revise articles collaboratively to move towards products of higher quality. It is necessary to use qualitative research methods to study their collaboration process and understand the workings of the open source method in this case, building on the understanding of contributors' motivations and their

demographic backgrounds—particularly their knowledge domains—and observing how they interact with other contributors in the Wikipedia open source policy structure to create articles. Techniques such as argumentation mapping—particularly pertinent in light of Wikipedia's Neutral Point of View policy—and expertise cues could be valuable, as well as other approaches as appropriate.

5. Conclusion

This research program can make unique valuable contributions in the advancement of knowledge in that the investigation of the open source phenomenon outside the realm of software takes the study outside of its traditional context in the world of software programmers. By studying non-software open source phenomena that involve Internet users from almost all demographic backgrounds, this program will be able to tap into the fundamentally human aspects of the open source phenomenon.

Furthermore, this research program will yield immediate and clear benefit to those outside the academic community because of the intensely participatory nature of the open source phenomenon. It will help society in designing, developing, and maintaining open source projects that effectively create information products such as Wikipedia for the public good. The results will also have ramifications for non-public information products that might use open source development methodologies, such as corporate knowledge management repositories.

6. References

[1] E. von Hippel et G. von Krogh, "Open source software and the "private-collective" innovation model: Issues for organization science," *Organization Science*, vol. 14, 2003, pp. 209-223.

[2] E.S. Raymond, "Linux and Open - Source Success," *IEEE Software*, vol. 16, 1999, pp. 85 - 89.

[3] A. Ciffolilli, "Phantom authority, self–selective recruitment and retention of members in virtual communities: The case of Wikipedia," *First Monday*, vol. 8, 2003.

[4] A. Lih, "Wikipedia as participatory journalism: Reliable sources? Metrics for evaluating collaborative media as a news resource," *5th International Symposium on Online Journalism*, Austin, Texas: University of Texas at Austin, 2004.

[5] J. Giles, "Internet encyclopaedias go head to head," *Nature*, vol. 438, Déc. 2005, pp. 900-901.

[6] G. Hertel, S. Niedner, et S. Herrmann, "Motivation of Software Developers in Open Source Projects: An Internet-Based Survey of Contributors to the Linux Kernel," *Research Policy*, vol. 32, 2003, pp. 1149-1291.

[7] K.R. Lakhani, B. Wolf, J. Bates, et C. DiBona, *The Boston Consulting Group Hacker Survey*, Boston Consulting Group and Open Source Developers Network, 2002.

[8] C. Okoli, "A Brief Review of Studies of Wikipedia in Peer-Reviewed Journals," *Proceedings of the Third International Conference on Digital Society, 2009.*, Cancun, Mexico: 2009, pp. 155-160.

[9] E. Svoboda, "One-Click Content, No Guarantees," *IEEE Spectrum*, vol. 43, Mai. 2006, p. 64.

[10] N.L. Waters, "Why You Can't Cite Wikipedia in My Class," *Association for Computing Machinery. Communications of the ACM*, vol. 50, Sep. 2007, p. 15.

[11] T. Fiedler, "The Web's Pathway to Accuracy," *Nieman Reports*, vol. 62, Summer. 2008, p. 40.

[12] J. Lerner et J. Tirole, "Some simple economics of open source," *Journal of Industrial Economics*, vol. 50, 2002, pp. 197-234.

[13] B. Butler, "Membership size, communication activity, and sustainability: A resource-based model of online social structures.," *Information Systems Research*, vol. 12, 2001, pp. 346-362.

[14] D.G. Winter, *The Power Motive.*, New York: Free Press, 1973.

[15] D.C. McClelland, "How motives, skills, and values determine what people do," *Amer. Psych.*, vol. 40, 1985, pp. 812-825.

[16] D.S. Kaufer et K.M. Carley, *Communication at a Distance: The Influence of Print on Sociocultural Organization and Change*, Hillsdale, USA: Lawrence Erlbaum Associates, 1993.

[17] R.M. Stallman, *The Free Software Definition*, 1996.

7. Acknowledgements

This study is funded by Social Sciences and Humanities Research Council of Canada grant #410-2006-2276.