Free knowledge sharing in an open internet community: the case of Wikipedia

Applying the concept of Open Source Software innovation to a community based encyclopaedia

Paper presented at the SMS annual Conference 2005
Orlando, Florida

Author:
Stephan Herting
University of St. Gallen
Institute of Management
Dufourstr. 40a
stephan.herting@unisg.ch

Abstract

In this paper, the concept of a private-collective innovation model that has been applied to Open Source Software projects will be used to analyze the case of the online encyclopaedia Wikipedia. The work suggests that the implications drawn in a software development context hold as well in different community based collaboration contexts. Endorsing the generalizability of the concept, this work instigates further research to be conducted using the model in a company setting. Additionally, the paper gives rise to the question how the incentives used to explain knowledge sharing in communities should be under research.

Introduction

Over the last few years, the software development in Open Source Software (OSS) projects has changed the way, management scholars discuss and research knowledge. Open Source Software development shows, how communities that are spread over the whole planet create knowledge intensive products – for free[1]. This gives rise to the question why people invest their time and skill into the development of a product that is free.

The motives for contributing have been recently under research (G. von Krogh, Spaeth, & Lakhani, 2003) and show that it is neither pure altruism nor a distinct cost-benefit relation that drives people to contribute to an OSS project but rather a mixture of several aspects like reputation, personal learning and technology leadership. In this sense, Open Source Software can be seen as based on a private-collective model (von Hippel & von Krogh, 2003) to describe the economics of such community based collaboration methods. Since the focus of research has mostly seen OSS projects as research topic, the transferability and generalizability of the findings to knowledge intensive innovation projects outside the field of software development can be cumbersome.

In this paper I will approach the phenomenon with an analytical tool that is related to the methods used for the research of OSS development but will propose that it be applied to the internet based encyclopaedia Wikipedia[2]. As the people contributing to Wikipedia have a personal background that not necessarily stems onto software programming, I will propose that the benefits arising from being part of a community are key for people to contribute to a free knowledge project.

The paper is structured in the following way: after presenting the recent theory this paper is based on, some hypotheses will be developed. After a case description, a design for a case study that seems suitable to test the derived hypotheses will be proposed. In the conclusion I will focus on what has been achieved, on limitations of this case study and on where more research needs to be done.

Theoretical setting

The literature about the resource based view of the firm has for a longer time been concerned with the question of how to derive competitive advantages from firm specific resources (J. B. Barney, 2001; Jay B. Barney, 2002; Wernerfelt, 1984, 1995) and therein especially knowledge (Grant, 1996). From this theories, the importance of the research on knowledge has arisen. Knowledge, however, is a concept that is deeply related to social structures and processes (Merton, 1973). The literature distinguishes content with the distinction between tacit and explicit knowledge (Polanyi, 1966) as well as process with the distinction between knowledge location, sharing and creation (Georg von Krogh, 2000). The distinction between tacit and explicit knowledge helps to grasp the difference between the part of knowledge that is easily transferable and scalable (explicit knowledge) and the part that needs learning and socialization to transfer (tacit knowledge). As this paper proposal deals with the creation of an online encyclopaedia, at first glance the main process may be seen in sharing explicit knowledge. Nevertheless, a lot of the skills required to write and contribute good online articles has as well a tacit dimension as the coding of a piece of software has. In the development of Open Source Software, people often use skills that they acquired before and use the OSS project to refine them (G. von Krogh et al., 2003) and by doing this create new knowledge.

---

[1] The term Open Source Software in this context is used to describe the "business model" of open innovation and thus also comprises lesser "open" Software that is distributed under licenses like the LGPL or the BSD license where still an open community contributes to a software project, but the openness of the source code may be limited. For a detailed description of the different licenses see the documentation available at the respective publishers web-sites (Free Software Foundation, 2002; The FreeBSD project, 2005).

[2] Please visit http://www.wikipedia.org to access the encyclopaedia online (Wikipedia, 2005b).

A lot of research focuses recently on understanding the individual and social requisites that enable knowledge sharing (Kogut & Zander, 1992; G. von Krogh, 2002; Widen-Wulff & Ginman, 2004). They focus on trying to frame the interactions and the personal incentives of knowledge sharing and all of them stipulate that there are some sort of catalysts or enablers for such a process. One approach chosen is to view the product properties as being constitutive to the personal incentives and examines knowledge sharing from an economic rather than a sociological approach. As Knowledge itself has as well a private as a public good character (Inkpen & Tsang, 2005), there are special issues connected with incentive systems for knowledge sharing. Especially the predictions that are related to the public goods character of knowledge puzzles scholars.

To gain a better understanding of the processes that evolve from the public goods character, recent research has focussed on the understanding of a special sort of knowledge sharing and creation projects – Open Source Software. But the Open Source Software research faces a dilemma. Because of the attributes of the product, open innovation projects should be classified as public goods and thus face the effects that are widely discussed like free-riding, insufficient private supply, etc. (Olson, 2002). But, these effects do not hinder the development of OSS products; in fact they exist and flourish even though according to basic economic theory they "should not". Therefore, different explanatory constructs have to be used. A solution to this problem has been provided by stating that such innovation is in fact a mixture of private innovation and collective action and thus a private-collective model of innovation (von Hippel & von Krogh, 2003).

In a recent case study on an Open Source project, von Krogh et. al. (2003) distinguished project members into different groups – related to the depth of their involvement – and described several selective incentives that are only available for active contribution in such a project like control over technology, reputation and learning opportunities (see also G. von Krogh, 2002). This research shows that there are distinct incentive systems available that enable effective knowledge sharing in communities.

The Open Source Software setting is a good example to study especially the commons-related aspect of knowledge as a public good. But it also has the disadvantage that it is connected with the software development setting. A lot of managerial attention is drawn on finding constructs describing the preconditions for knowledge sharing in a company environment (Grant, 1997). But, basically, the research on Open Source Software stays in the realm outside a company and with a very distinct type of people. Available literature has not been concerned with the question of dependency on a very specific research object. But maybe not the research object OSS is specific, rather the community producing it? Perhaps the affinity to develop "free"-software could as well stem from the properties of the research object "software" and not the research object "contributor"? Thus, it is questionable if these incentive systems also work for a more extensive set of people, not only software programmers. The theory under review does not give evidence that this problem can be neglected in further research. Thus there is a need to make the constructs derived in theory reliant for non-OSS community projects.

Seeing the problem this way, there is some need to extend the theory and enter research objects and cases that are related to Open Source Software in the way that a public good is freely produced by a community that is not a piece of Software. This gives rise to examining a different case using an OSS framework.

Empirical setting and case description

Recently, a world-wide debate has arisen about the public good character of intellectual property. Scholars promote that intellectual property is in fact a public good and the creation and dissemination of intellectual property should be fostered by restricting creators copyrights (Lessig, 1999, 2001, 2004). Sharing such a view, a lot of internet based projects have evolved that try to make intellectual property freely or easily accessible (e.g. Internet archive for the preservation of non-copyrighted material). Especially topics of popular interest are very successful in creating a caring online community that contributes to the extension of the project (e.g. Link Everything Online, 2005)[3]. One of the most renowned projects is Wikipedia that will be the research object of this paper.

Why chose Wikipedia? At Wikipedia, through amalgamation of different existing knowledge, the communal knowledge base is enhanced and thus – mainly via the process of combination of existing knowledge (Nonaka,

[3] Link Everything Online (LEO) is an online dictionary that translates words and terms from german-english, english-german, german-french and french-german. The project is community - based, although a sort of supervisory board checks entries before they are added. The english / german part hosts over 400.000 expressions from colloquial language up to technical terms. Included is a forum where community members actively help members to find the right translation of phrases.

1995) – knowledge is created. In fact, contributing to Wikipedia can be described using the concept of a gift economy as proposed for OSS development by Zeitlyn (2003). Additionally, Wikipedia as well as OSS projects are massive internet-based collaboration projects that affect the way, people use the internet today. So, the case of Wikipedia seems to be comparable to OSS projects with the difference that people contributing to Wikipedia don't need special skills like knowing computer languages to be able to contribute.

Wikipedia is an online encyclopaedia that is based on a software tool called "wiki". A "wiki" is an online collaboration tool where everyone has access to text being displayed on the internet and can change it without the need for approval or special rights to do so (Kleinz, 2004). Wikipedia is a project to create an encyclopaedia that is freely accessible and expandable, based on such a "wiki" software. The project was started in January of 2001 and now includes over 440.000 articles in January 2005. The project has also been spread into different languages where translations as well as new versions of the articles from the English main page can be found. All content made available on Wikipedia is set under the "GNU Free Documentation License" (GFDL) and is thus in the public domain. People contributing to Wikipedia cannot enforce their copyright on the text that they contributed (Free Software Foundation, 2002). The quality of Wikipedia is widely regarded as being satisfactory (Kleinz, 2004) with - in comparison to other encyclopaedias - a sufficient scope.

One big advantage of Wikipedia is seen in the method of "peer review" that can be used to improve article quality. The idea is that if anybody who encounters wrong or insufficient information contained in an article will correct it (Wikipedia, 2005a). Thus, with time the quality of the articles should reach a similar or higher quality than available from commercial encyclopaedias. Another method to foster quality is a "featured article" mechanism that presents articles of high quality directly in a list and on the main page and by this creates of course reputation through visibility for the editor(s).

Overall, this shows that Wikipedia is an attempt to create a public domain encyclopaedia that is driven by individuals that collaborate in an online community to produce a freely available good. So Wikipedia is as well comparable to OSS projects in the way it is fostered, governed and regulated. Additionally, it is based on the product wiki that evolved directly out of the OSS movement as it itself is an OSS project (Wiki.org, 2005).

Hypotheses

Based on the theory review and the specifics of Wikipedia, several hypotheses can be shaped that can be tested using the case.

The transfer of the OSS theory to a project that is – even though being internet based – not a software programming project allows for a generalization of the community contribution mechanisms that until now have only been identified using OSS projects as research objects. So, Wikipedia can serve to give more insight to the specifics of communities and their inherent incentive systems.

Proposition 1: The concept of a private-collective innovation model is also applicable to a non-software-development environment.

The case of Wikipedia can serve to investigate if the same assumptions about the nature of knowledge as it is produced and shared in this community of Wikipedia contributors can be made as for the private-collective innovation model.

Proposition 2: At Wikipedia there is a distinct set of three different types of contributors: heavy contributors, sometimes contributors and lurkers.

Again, the quantitative part of the study will be used show that at Wikipedia, basically the same processes and categories can be applied as are possible in the research on OSS projects. Thus, by showing this, Proposition 2 can foster the statement made in Proposition 1 and in the motivation for this paper that the problem faced with free knowledge sharing is not a specific attribute of software projects. On the other hand, Proposition 2 shows that the theoretical distinction is suitable for describing collective knowledge sharing projects and should be used to generalize theory about this topic.

The qualitative part of the proposed case study will be used to test a hypothesis that then discovers incentives for contributing at Wikipedia. OSS literature states, that especially being part of a community and the incentives this community offers serve as incentives for contributing.

Proposition 3: The main incentives for heavy contributors to share knowledge at Wikipedia are the personal learning possibility, reputation and control over the development of Wikipedia and are connected with being part of an active community that facilitates the project.

The aim of the proposed hypotheses is to shed new light on the collective mechanisms of community knowledge sharing. An empirical study will be used to support the hypotheses and serves as "experiment" for the presented theory.

Methodology

In this paper proposal I will use an approach of applying existing theory to a new – even though related – phenomenon. Thus, the paper will try to validate the theory as well as by doing this expand it out of the realm of OSS development into a wider area of community related knowledge sharing. Recognizing a problem, testing available theory and drawing own conclusions out of it is a widely accepted strategy. In this view, the paper is rather based on the approach using a scientific method of theory testing (Descartes, 1899; Kuhn, 1996; Popper, 1966). The novelty of the hypotheses lies in the change of the research object and conducting the "experiment" using this different research object. Thus, the case will serve as validation instrument for the above stated propositions (Yin, 1994).

Note, however, that proposition 3 stipulates effects that are well found in theory but focuses on a special concept. To justify such a procedure, the adequacy of the research object has to be proven. As contribution and use of Wikipedia is totally free, totally unregulated, independent from software specific expertise and versatile, it is the extreme case of a public domain knowledge sharing project and thus serves well as research object. So, examining the extreme case provides to validate theory even from a single case study (Eisenhardt, 1989).

The case study will have a quantitative as well as a qualitative aspect. The quantitative analysis will be used to prove the suitability of the theoretical founding whereas the qualitative part will serve to validate proposition 3. The proposed methodology is twofold. Firstly, an empirical study about the validity of the case should be conducted to test if the hypothesis set available from theory matches the case study. Next, a qualitative study covering heavy contributors serves to test the propositions. Additional data from the community forum will be used for triangulation of the data to check its validity.

The "way-back-machine"[4] of the web-archive can serve as a starting point to track growth of Wikipedia and present prior versions of the articles in the research sample if necessary. A generic sample of topics should be created that covers various fields of science, history and general knowledge as has shown suitable in practice in Kurzidim (2004). Given the large size of Wikipedia, a set of about 500 articles should be sufficient to gain statistical evidence about the behaviour of contributors (Stier, 1999). As the articles will be chosen by random out of an existing encyclopaedia[5], the gathered data should not be biased by special tastes of Wikipedia contributors. The random selection will also guarantee that the sample does not select articles from a special domain of knowledge like e.g. natural sciences. This set of topics will be researched in the archives of Wikipedia for existence, length, revisions and content. However, examining only a subset of the totality of the available data faces evident inherent dangers. Especially randomly picking a large amount of unfinished articles or articles of minor quality could reveal false contributing patterns. Thus, the proposed subset of articles will be dynamically enlarged if sufficient statistical evidence cannot be found.

Having identified the articles that serve as research object, a deeper research should be conducted that stems especially on the editing of the articles under research. The Wikipedia wiki-system itself keeps track of who made changes and when and additionally of how many accesses to a specific article have been done. The same methodology has been used to build hypotheses in the research of von Krogh et. al. (2003), so the study will be able to measure similar constructs. This should reveal a pattern of heavy contributors, some-times contributors and lurkers as it has been identified earlier (G. von Krogh et al., 2003). This part of the research will serve to test the propositions 1 and 2 and should show that the constructs used in the corresponding theory set are applicable to the case of Wikipedia.

[4] The web archive archives complete web-pages on a regular basis and makes them available today in the same state as they were years ago. So it is a perfect tool to track the development and change of a specific website in a longitudinal manner (Internet Archive, 2005).

[5] I will use Encyclopaedia Britannica to deliver article topics. Encyclopaedia Britannica is a commercial encyclopaedia that is renowned for its richness and is one of the standard references for general knowledge.

Having shown this, the category of heavy contributors will be under further research. Following the quantitative research, heavy contributors as well as project leaders will be identified and interviewed to test their incentives against those identified by von Krogh et. al. (2003).The identification of heavy contributors can be complicated by the fact that Wikipedia allows anonymous entries. Even though those entries are tracked using the IP-Address of the contributor (and thus are not really anonymous), the identification of such contributors can be hard. Nevertheless, as a lot of users choose to use a username with an E-Mail address, the identification of heavy contributors should still be possible. As it is not the intent of this study to interview every heavy contributor but only to identify a few for a qualitative study to give foundation to proposition 3, this fact should not interfere with the research. But, one should take into account that it is a special feature of the identified persons to not contribute anonymously. This may present proof that reputation and responsibility for an article are motives for such persons and will be included in the interview process. For purposes of triangulation of the data, the Online Forum of Wikipedia will be under review for contributions of the identified heavy contributors. Also, the maintainers of the web-site will be interviewed to test the reliability of the acquired data.

Using semi-structured interviews, the key incentives will be identified and coded especially regarding the incentives noted in proposition 3 (Stier, 1999). This potentially reveals learning objectives are the primus-inter-pares of the stipulated incentives for contributing, and thus for knowledge sharing. Of course, the interview data will also contain control variables, especially gender, age, profession and education. Eventually, the analysis could show that again a distinct type of people chose to contribute to Wikipedia (e.g. male, 20 - 30 years old, affiliation to an academic institution and above average education). If such is the case, especially proposition 1 could loose founding as the special case could reveal that generalization to a general theory about community contribution is not possible using the case of Wikipedia. This would result in further research that should be conducted to verify proposition 1 using a more general case.

Altogether, this methodology should be reasonable to test propositions 1 - 3 and reveal the basic hypothesis that free knowledge sharing is a community related concept rather than being specific for OSS projects.

Conclusions

This paper proposes a case study to be conducted to test and expand the incentive system of a private-collective model of innovation. As there is potential for criticism about the phenomenon of open innovation being bound to OSS specific projects, this study will show that an incentive model that has been derived to describe OSS phenomena can lead to an explanation of viable all-purpose community projects. So the emphasis on further research might be drawn to better understand the individual incentives connected with sharing knowledge to a community.

However, this approach has limitations. As stated above, the identified heavy contributors have the characteristic that they want to be recognizable by using a username while contributing. Again, perhaps this is a specific that only the contributors share and may shed light on the fact that the reputation mechanism may be very complex. The approach can well serve to prove that the ideas generated in the context of software programming are in fact widely applicable in any set of intellectual property and knowledge sharing; the findings connected with proposition 3 are to be further examined as this case study examines again a special community. For a generalizability of the concept, there should be research conducted to test the identified constructs in company setting to validate whether they are community inherent or not. If the propositions can be verified, there is still need for more research on similar projects to craft a useful framework for understanding community contribution processes.

The existing literature as well as this papers implications are useful for the theoretical discussion but still do not meet the demands of practitioners. Further research should dwell on the insights presented here and derive a framework that is applicable for managing communities in their knowledge sharing activities. Understanding the duality and implications of such a private-collective model of community participation could help to give practitioners methods at hand to foster their knowledge-related activities. Accepting that contributing to communities is itself a benefit for people will help shaping the discussion about knowledge-management tools and may make the research that has been undertaken in the field of Open Source Software a useful supplemental theoretical framework for firms.

References

Barney, J. B. (2001). Is the resource-based "view" a useful perspective for strategic management research? Yes. *Academy Of Management Review, 26*(1), 41-56.

Barney, J. B. (2002). *Gaining and sustaining competitive advantage* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Descartes, R. (1899). *Discours de la méthode pour bien conduire sa raison et chercher la vérité dans les sciences*. Paris: Mellottée.

Eisenhardt, K. M. (1989). Building theories from case-study research. *Academy Of Management Review, 14*(4), 532-550.

Free Software Foundation. (2002). Gnu free documentation license. Retrieved 11.01.2005, 2005, from http://www.gnu.org/licenses/fdl.html

Grant, R. M. (1996). Toward a knowledge-based theory of the firm. *Strategic Management Journal, 17*, 109-122.

Grant, R. M. (1997). The knowledge based view of the firm: Implications for management practice. *Long Range Planning, 30*(3), 450-454.

Inkpen, A. C., & Tsang, E. W. (2005). Social capital, networks, and knowledge transfer. *Academy Of Management Review, 30*(1), 146 - 165.

Internet Archive. (2005). About the wayback machine. Retrieved 11.01.2005, from http://www.archive.org/about/press_release.php

Kleinz, T. (2004). Qualitätsoffensive. Freie online-enzyklopädie wikipedia stellt die weichen für die zukunft. *c't Magazin für Computer Technik*(14), 38+.

Kogut, B., & Zander, U. (1992). Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science, 3*(3), 383-397.

Kuhn, T. S. (1996). *The structure of scientific revolutions* (3rd ed. Vol. 2,2). Chicago: Chicago University Press.

Kurzidim, M. (2004). Wissenswettsreit. Die kostenlose wikipedia tritt gegen die marktführer encarta und brockhaus an. *c't Magazin für Computer Technik*(21), 132+.

Lessig, L. (1999). *Code and other laws of cyberspace*. New York, NY: Basic Books.

Lessig, L. (2001). *The future of ideas the fate of the commons in a connected world*. New York: Random House.

Lessig, L. (2004). *Free culture how big media uses technology and the law to lock down culture and control creativity*. New York: Penguin Press.

Link Everything Online. (2005). The online dictionary. Retrieved 25.01.2005, from http://www.leo.org

Merton, R. K. (1973). *Sociology of science theoretical and empirical investigations*. Chicago: University of Chicago Press.

Nonaka, I. (1995). *The knowledge-creating company how japanese companies create the dynamics of innovation*. New York, NY: Oxford University Press.

Olson, M. (2002). *The logic of collective action public goods and the theory of groups* (20th printing ed. Vol. 124). Cambridge (Mass.): Harvard University Press.

Polanyi, M. (1966). *The tacit dimension*. Gloucester, Mass.: Smith.

Popper, K. R. (1966). *Logik der forschung* (2., erw. Aufl. ed. Vol. 4). Tübingen: Mohr.

Stier, W. (1999). *Empirische forschungsmethoden* (2., verb. Aufl. ed.). Berlin <etc.>: Springer.

The FreeBSD project. (2005). Freebsd cobyright and legal information. Retrieved 25.01.2005, 2005, from http://www.freebsd.org/copyright/index.html

von Hippel, E., & von Krogh, G. (2003). Open source software and the "private-collective" innovation model: Issues for organization science. *Organization Science, 14*(2), 209-223.

von Krogh, G. (2000). *Enabling knowledge creation how to unlock the mystery of tacit knowledge and release the power of innovation*. New York: Oxford University Press.

von Krogh, G. (2002). The communal resource and information systems. *Journal Of Strategic Information Systems, 11*(2), 85-107.

von Krogh, G., Spaeth, S., & Lakhani, K. R. (2003). Community, joining, and specialization in open source software innovation: A case study. *Research Policy, 32*(7), 1217-1241.

Wernerfelt, B. (1984). A resource-based view of the firm. *Strategic Management Journal, 5*(2), 171-180.

Wernerfelt, B. (1995). The resource-based view of the firm - 10 years after. *Strategic Management Journal, 16*(3), 171-174.

Widen-Wulff, G., & Ginman, M. (2004). Explaining knowledge sharing in organizations through the dimensions of social capital. *Journal Of Information Science, 30*(5), 448-458.

Wiki.org. (2005). Wiki: What is wiki. Retrieved 31.01.2005, from http://www.wiki.org

Wikipedia. (2005a). Peer review. Retrieved 11.01.2005, from http://en.wikipedia.org/wiki/Wikipedia:Peer_review

Wikipedia. (2005b). Wikipedia homepage. Retrieved 31.01.2005, from http://www.wikipedia.org

Yin, R. K. (1994). *Case study research design and methods* (2nd. ed.). Thousand Oaks, Calif.: Sage.

Zeitlyn, D. (2003). Gift economies in the development of open source software: Anthropological reflections. *Research Policy, 32*(7), 1287-1291.