2006

Knowledge Integration and Effectiveness of Open Source Software Development Projects

Annapoornima M. Subramanian
NUS Business School, annapoornima@nus.edu.sg

Pek-Hooi Soh
NUS Business School, bizsohph@nus.edu.sg

Knowledge Integration and Effectiveness of Open Source Software Development Projects

Annapoornima M. Subramanian

NUS Business School

annapoornima@nus.edu.sg

Pek-Hooi Soh

NUS Business School

bizsohph@nus.edu.sg

Abstract

The emerging work on Open Source Software (OSS) field focuses on the various factors that lead to effectiveness in OSS development teams. This paper utilizes the theory on knowledge integration to show how the capability of OSS projects in integrating knowledge influences their effectiveness. The research model is based on the idea that competency of OSS projects along three characteristics of knowledge integration (1) Efficiency of Integration (2) Flexibility of Integration (2) Scope of Integration, enhances the effectiveness of the project. The effectiveness of the project is measured based on the output that the project generates during software development. The impact of the three knowledge integration capabilities on effectiveness of OSS projects are tested using objective data on 152 open source software development projects. Results indicate that efficiency and flexibility of knowledge integration of OSS projects, positively influences the effectiveness of these projects. Theoretical and practical implications are discussed.

Keywords: Knowledge integration, Open source software, Flexibility, Efficiency, Scope

1. Introduction

Open Source Software (OSS) development projects are internet-based communities of software developers who voluntarily collaborate in developing software that are made freely available to everyone. The open source software concept can be dated back to 1960s wherein

academic scientists found it a normal part of their research culture to freely give away software and to build upon other's software individually and collaboratively. This communal behavior has grown tremendously to such an extent that in recent years, tens of thousands of projects are getting registered on Sourceforge, a website that supports open source development (http://www.sourceforge.net). This growth has triggered researchers from various fields ranging from economists who are interested in knowing why software developers and profit organizations give away knowledge free of cost, to sociologists who are interested in understanding how people from diverse background collaborate in virtual community for the public good.

A number of studies on open source software attempt in understanding the various motivations of the software developers to participate in the community (Lerner and Tirole 2002; Lakhani and von Hippel 2003). Intrigued by the informal structure of open source project community, a few researchers have explored how people join the open source community and how the OSS projects possess self-organizing power (Joode and Kemp 2001). The emerging topic in OSS field is to study the various factors that contribute to the effectiveness of OSS projects. Factors that are identified by existing literature are size and level of activity (Crowston et al. 2003), community ideology and trust (Stewart and Gosain 2006), peer-review and quality consciousness (Senyard and Michlmayr 2004). While these studies identify some of the important drivers of OSS project effectiveness, very few studies have attempted to use the knowledge based theory to enrich this branch of research (Krogh et al. 2005; Hemetsberger and Reinhardt 2004). OSS projects are virtual communities that integrate knowledge possessed by developers with various expertises. The scope of different kinds of knowledge that are brought in for development can increase the effectiveness of the project. But at the same time, since the actors who collaborate have diverse goals and perspectives, it might be a challenging task to integrate their knowledge. Motivated by these two contradictory premises, our study analyzes the impact of knowledge integration capabilities of OSS projects on their effectiveness. Following the knowledge integration theory proposed by Grant (1996), our study examines the impact of three important knowledge integration capabilities (1) Efficiency of Integration (2) Flexibility of Integration and (3) Scope of Integration, on the effectiveness of 152 OSS projects in generating output during the software development process.

2. Theory development and Hypotheses

Knowledge integration is the fusion and combination of knowledge from sources that are multiple, distributed and heterogeneous (de Boer et al. 1999). In many situations, knowledge transfer or integration through market contracts can be very difficult because of the high transaction-specific cost involved. Hence if diverse specialized knowledge is a critical input to all production process then the primary role of a firm is to integrate the diverse knowledge possessed by individuals (Grant 1996). This explains that hierarchy is better than market for knowledge integration and that the organizations' effectiveness and competitive advantage depends on their capability in integrating knowledge. Apart from market and hierarchy, a new setup for knowledge creation has evolved in recent times, known as open innovation (Chesbrough 2001). In open innovation, communities of people get involved in the knowledge creation process for the common good. Open source software development is a premier example of an open innovation wherein, communities of software developers virtually collaborate in developing software that are made freely available to everyone. The relative advantage of these communities over firms in knowledge creation process is similar to that of the advantage possessed by hierarchy over market. Software developers, who are specialists in varied background, participate in developing an innovation for the public good. The main advantage of OSS software projects over proprietary software development is the varied set of knowledge that these developers from heterogeneous background bring forth for development. The added benefit of OSS projects is that such heterogeneous knowledge base is available without any resource constraints, because many of these software developers participate in the community development without anticipating any economic benefits. While the knowledge heterogeneity is an advantage, the diversity of background of these developers can be a threat in attaining the benefits that hierarchy enjoys over market. Hence, we posit that the effectiveness of OSS projects depends on the capability of these projects in integrating the knowledge possessed by the specialists. Given that meeting budgets and requirements may not be relevant in OSS context, more appropriate indicators of effectiveness may be those related to the ongoing productivity of the team (Stewart and Gosain 2006). Hence we consider the effectiveness of the open source software projects as the extent to which it produces observable outputs such as addition of new features to the software. The following sections develop the research model shown in Figure 1 that explains how the three

characteristics of knowledge integration help the OSS software projects in achieving effectiveness.

2.1 Efficiency of Knowledge Integration

Efficiency of integration is the extent to which the projects access and utilize the specialized knowledge held by individual members (Grant 1996). One of the important factors in determining the efficiency of integration is the communication (Grant 1996). There are two ways through which communication can help OSS projects in utilizing and harnessing the knowledge from its specialists for efficient integration. Firstly, communication helps the project members in gaining metaknowledge, which is the "knowledge of knowledge" (Majchrzak et al. 2004). Possessing metaknowledge is important for knowledge integration and the reason is, unless the project members get to know the expertise of other project members, the possibility of accessing this knowledge for integration is almost impossible. Secondly, communication helps in developing a common language among the specialists which is important for integrating their knowledge. The general tenet is that, wider the scope of knowledge being integrated, greater the diversity of individuals and hence lower is the level of common knowledge. Effective communication between the participants can mitigate the drawback of diversity by providing a common knowledge base. Hence communication is vital for a virtual community such as OSS that comprises of software developers from diverse background, in efficiently integrating their knowledge for overall productivity. This leads to our first hypothesis

Hypothesis 1: The efficiency of OSS projects in integrating knowledge, as measured by the communication structure, is positively related to the effectiveness of the project, measured in terms of the output generated.
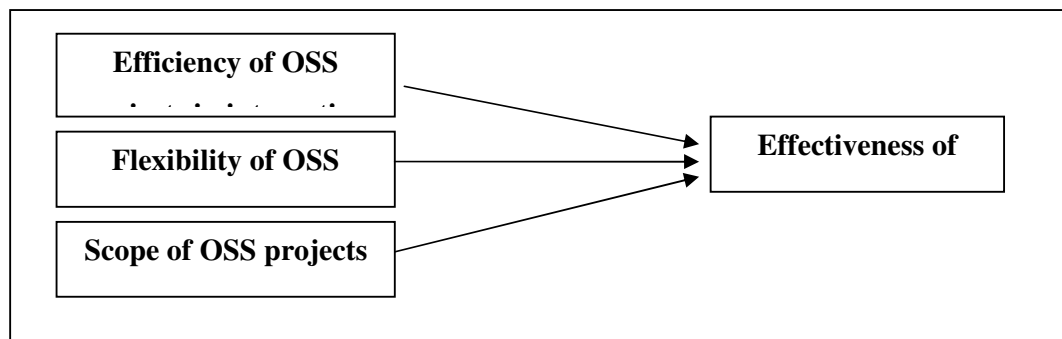

Figure 1. Knowledge integration and effectiveness of OSS projects

917

2.2 Flexibility of Knowledge Integration

Flexibility of integration is the extent to which a capability can access additional knowledge and re-configure existing knowledge (Grant 1996). While efficiently integrating the knowledge that resides within the project is important, it is equally important that the project possesses absorptive capacity to integrate knowledge from outside (Cohen and Levinthal 1990). Superior performance requires the continuous renewal of competitive advantage by incorporating new capabilities from outside (Chesbrough 2001). Since software is notably explicit, flexibility of integration in software context can be referred to as the extent to which the projects can integrate or reuse the code that is available in other projects. The practice of integrating other's work is of particular relevance in software industry because most significant advances in knowledge reuse are made in this field (Markus 2001). With the copyleft license existing in OSS projects, one would expect people to reuse other's software and build on it to speed up the innovation process (von Krogh et al. 2005). Hence OSS projects that have the capability of exploiting the open nature of innovation by integrating code from other OSS projects are expected to be more effective in generating output. But there are a few factors that prevent such code reuse being widely adopted even in an open set up such as OSS. While aspects such as modularity and technological compatibility are widely dealt in literature (von Krogh 1995), compatibility of license, often considered to be a primary issue that inhibits the integration of code from other projects (http://www.advogato.org/article/657.html) has gained very little attention. OSS projects that are license compatible with a number of other OSS projects have increased opportunity in integrating and reusing knowledge from other projects. This flexibility of increased integration and reuse of others code can enhance the effectiveness of the project by speeding up the output generation. Hence

Hypothesis 2: The flexibility of OSS projects in integrating knowledge, as measured by the number of other license compatible OSS projects, is positively related to the effectiveness of the project, measured in terms of the output generated.

2.3 Scope of Knowledge Integration

Scope refers to the breadth of specialized knowledge that the project draws upon (Grant 1996). Breadth in our study pertaining to OSS context refers to the different software

languages and operating systems from which the knowledge is drawn upon in developing the software. OSS projects that increase the scope of knowledge integration, increases its effectiveness in output generation through three sources. Firstly, greater the scope of knowledge, higher is the number of developers in the open community who can participate in the development. This will enhance the input effectiveness of the project by increasing the team size there by resulting in enhanced output generated (Stewart and Gosain 1996). Secondly, different types of specialized knowledge are complements rather than substitutes. Hence an additional knowledge scope increases the capability of the project in accomplishing the tasks. Thirdly, greater the breadth of knowledge integrated, higher is the number of other projects in OSS community from which knowledge can be reused, which increases the effectiveness of the project. Hence

Hypothesis 3: The scope of OSS projects in integrating knowledge, as measured by the number of software languages and operating systems used, is positively related to the effectiveness of the project, measured in terms of the output generated.

3. Research Method

3.1 Sample and Data Collection

Objective data was collected from the Sourceforge website, one of the major sites where most of the open source projects are hosted (von Hippel and von Krogh 2003). The projects in Sourceforge website are divided into various categories and sub-categories. We limited our sample to the algorithm and modeling categories of software development projects for which integration is considered to be of greater importance. Since our model is concerned with knowledge integration, we ensured that the sample projects had at least 2 members.

Team effectiveness was measured based on the output generated by these projects. We calculated the output generated as the percentage of task completed: ((total feature requests-open feature requests)/ total requests)*100, or zero for projects with no task requests (Stewart and Gosain, 1996). *Efficiency of knowledge integration* refers to the communication effectiveness of the OSS projects. Two items were used to measure this construct (1) number of forum messages (2) number of messages in mailing lists. *Flexibility of knowledge integration* was measured using the number of other projects that are license compatible with the focal project of interest. A restricted measure of license compatibility is used in our paper.

It is assumed that a project with a particular license such as BSD is only compatible with projects having the same license (http://www.advogato.org/article/657.html). *Scope of knowledge integration* was measured using the number of (1) software languages and (2) operating systems that the OSS project is based upon. Since the breadth of software languages and operating systems used in the project determines the extent to which the project attracts experts from broad areas, we used this proxy measure to capture the scope of knowledge integration. Apart from the above four main variables, control variables such as *age of the project* (number of years since project registration), *development stage of the project* (as reported in the Sourceforge website) and *number of developers* (as reported in the Sourceforge website) were used in testing the research model.

3.2 Measurement Model

Descriptive statistics are provided in Table 1. To validate our measurement model, two types of validity were assessed: convergent validity and discriminant validity. We assessed the convergent validity by examining composite reliability and average variance extracted from the measure. The reliability values and average variance extracted for our measure are provided in Table1. The composite reliability values are above the recommended 0.7 value (Chin 1998) and the average variance extracted by our measures are above the prescribed value 0.5 (Fornell and Larcker 1981). The result in Table 1 also confirms the discriminant validity: the square root of the average variance extracted for each construct is greater than the correlations involving other constructs (Fornell and Larcker 1981).

Constructs	Mean	S.D.	Reliability	Avg.Var. Extracted	1	2	3	4	5	6	7
Age of the project	2.71	2.09			0.417	0.158	0.202	0.279	1		
Number of developers	30.28	41.54	1	1	0.310	0.485	0.035	0.141	0.227	1	
Development stage	3.64	1.40									
Effectiveness	3.64	4.45	1	1							
Efficiency of KN integration	9.22	46.71	0.839	0.723	0.035 / 0.285	-0.850 / 0.059	0.116	-0.058	0.196	0.061	1
Flexibility of KN integration	1.82	0.53	1	1	0.202	0.113	1				
Scope of KN integration	26.35	25.00	0.781	0.656	0.079	0.116	0.061	0.810			

Table 1. Descriptive Statistics, Reliability Analysis and Inter-Construct Correlations

Paths	Coefficients
Efficiency of KI->Effectiveness	0.177^{***}
Flexibility of KI->Effectiveness	0.156^{**}
Scope of KI->Effectiveness	-0.074
Age of Project->Effectiveness	0.354^{***}
No. of Developers->Effectiveness	0.152^{**}
Development Stage->Effectiveness	-0.051

Table 2. Results (**p<0.05, ***p<0.01)

3.3 Structural Model

The hypotheses were tested using PLS Version 3.00. PLS is a latent structural equation modeling technique that utilizes a component-based approach to estimation (Chin 1998). Assessment of structural model involves estimating the path coefficients and the R^2 value. Paths are interpreted as standardized beta weights in regression analysis and R^2 measures the predictive power of the model for the dependent variable. The path coefficients are provided in Table 2. The structural model could explain 27% of the variance in effectiveness of the project (R^2=0.27).

As hypothesized, the efficiency of knowledge integration significantly predicted the effectiveness of the project. Thus hypothesis 1 is supported. Hypothesis 2 that predicted the positive influence of flexibility of knowledge integration on OSS project's output generation is also supported. The scope of knowledge integration did not have any significant impact on effectiveness of the project, hence rejecting hypothesis 3. Of the three control variables, age of the project and number of developers in the project had significant positive effect on the output generated by the project. Development stage control variable is insignificant in explaining the effectiveness of the OSS project.

4. Discussion and Conclusion

Results support the major thesis of the paper that the knowledge integration capabilities of the OSS projects are important for their effectiveness. Efficiency of knowledge integration as determined by effective communication turned out to be a stronger predictor of the

effectiveness of the project. This reiterates the existing finding that good communication within the OSS projects is essential for generating better output (Stewart and Gosain 2006). The results of our study also support the claim that flexibility of knowledge integration, as measured by the extent to which OSS projects are license compatible with other OSS projects has a significant positive effect on the effectiveness of the project. In an open knowledge sharing environment such as OSS, effectiveness of the project greatly depends on the extent to which the project can integrate and build upon knowledge that is available in public domain. This phenomenon, also known as open innovation (Chesbrough 2001) is essential for technologies such as software, where the knowledge required for building the innovation is widely available. Knowledge integration theory posits that the scope of knowledge integration has positive influence on the effectiveness of the project. But our result does not show any significant influence exerted by the scope of knowledge integration on the output generated by OSS projects. One of the reasons that can be attributed to this discrepancy is the limitation pertaining to the measure. Our study draws on the number of different software languages and operating systems used in the development of OSS project to represent the scope of knowledge integration. The reasoning behind using such a measure is that, greater the breadth of software languages and operating systems used for development, wider would be the expertise of people who come forward in developing the project. Though the assumption may be valid, it should also be recognized that the breadth of software languages and operating system can also result in greater challenge in integrating the project, as a result of which effectiveness of the project can deteriorate. Future studies can incorporate a more direct measure for scope of knowledge integration such as the breadth of knowledge possessed by the developers who participate in the development of the OSS project.

While a number of studies have explained the importance of knowledge integration, there has been no empirical examination of this idea because of the difficulty in observing the phenomenon. Using the open source setting wherein the actions, outputs and communication structure are observable from websites, forum messages and email archives, we contribute to this branch of literature by empirically evaluating the importance of knowledge integration. Our study has also contributed to this branch of literature by emphasizing the significance of license structure. License in OSS context is an important intellectual property protection that determines the extent of compatibility and attractiveness of the project. But issues related to OSS licenses have not gained due attention from researchers and practitioners in the past.

Our study has several practical implications for OSS project managers. OSS project managers need to maintain an open knowledge sharing structure and culture that values ideas for their merit and not on the basis of the source from where the ideas come from. Such open culture can increase the number of developers who collaborate in developing the project, which is found to have significant positive influence on the effectiveness of the project. Open culture can also increase the scope of diverse knowledge that can be integrated for effective output generation. Apart from maintaining a rich source of knowledge within the project, project managers should also foster good communication among the developers that can speed up the knowledge integration process. The communication structure should be capable of fostering healthy factors that induce trust and coordination among the members, and at the same time economize on unhealthy conversations that might impair the capability of the members to generate output. While a good store of internal knowledge augments the effectiveness of the project, it is equally important that the project integrates and reuses knowledge that is residing in other projects. Due to the not-invented syndrome, software developers may not have the inherent motivation to rebuild on other's code. It is the responsibility of the project managers to drive the developers towards integrating knowledge from other projects so as to speed up the process of development and hence increase the effectiveness of the project. Last but not least, project managers should be wary in deciding the license structure. Though most of the prior literature considers OSS environment to embrace open knowledge sharing culture, exploration of some of the forum messages reveal that software developers consider intellectual rights and licensing to be a major factor. Hence in deciding on the license, project managers need to consider the concerns of the developers who are involved in the project and at the same time maintain a license that is compatible with other OSS projects. This can enhance the attractiveness of the project as well as its flexibility in reusing code from other projects.

References

Chesbrough, H. "The Era of Open Innovation," *MIT Sloan Management Review*, 2003, pp 35 -41.

Chin, W. W. "The partial least squares approach to structural equation modeling," *Modern Methods for Business Research*, G. A. Marcoulides, Mahwah, NJ, Lawrence Erlbaum, 1998.

Cohen, W., and Levinthal, D. "Absorptive capacity: A new perspective on learning and Innovation," *Administrative Science Quarterly* (35), 1990, pp 128-152.

Crowston, K., Annabi, H., and Howison, J. "Defining Open Source Software Project Success," *International Conference on Information Systems*, Seattle, WA, 2003.

de Boer, M., Van den Bosch, F. and Volberda, H. "Managing Organizational Knowledge Integration in the Emerging Multimedia Complex," *Journal of Management Studies* (36:3), 1999, pp 379–98.

Fornell, C. and Larcker, D. F. "Structural equation models with unobservable variables and measurement errors," *Journal of Marketing Research* (18:2), 1981, pp 39-50.

Grant, R. M. "Prospering in Dynamically-Competitive Environments: Organizational Capability as Knowledge Integration," *Organization Science* (7:4), 1996, pp 375-387.

Hemetsberger, A., and Reinhardt, C. "Sharing and Creating Knowledge in Open-Source Communities The case of KDE," 2004, http://opensource.mit.edu/papers/hemreinh.pdf

Joode, R., and Kemp, J. "The Strategy Finding Task Within Collaborative Networks, Based on an Exemplary Case of the Linux Community," 3rd IFIP Working Conference on Infrastructures for Virtual Enterprises, 2001.

Lerner, J., and Tirole, J. "Some Simple Economics of Open Source," *Journal of Industrial Economics* (50:2), 2002, pp 197-234.

Majchrzak, A., Cooper, L. P., and Neece, O. E. "Knowledge Reuse for Innovation," *Management Science* (50:2), 2004, pp 174-188.

Markus, M.L. "Towards a theory of knowledge reuse: Types of knowledge reuse situations and factors in reuse success," *Journal of Management Information Systems* (18:1), 2001, pp. 57-93.

Senyard, A., and Michlmayr, M. "How to Have a Successful Free Software Project," 2004, http://opensource.mit.edu/papers/senyardmichlmay.pdf

Stewart, K.J., and Gosain, S. "The Impact of Ideology on Effectiveness in Open Source Software Development Teams," *MIS Quarterly,* forthcoming.

von Hippel, E., and von Krogh, G. "Open Source Software and the Private-Collective Innovation Model: Issues for Organization Science," *Organization Science* (14:2), 2003, pp 209-223.

von Krogh, G., Spaeth, S., and Haefliger, S. "Knowledge Reuse in Open Source Software: An Exploratory Study of 15 Open Source Projects," *Proceedings of the 38th Hawaii International Conference on System Sciences*, 2005.