Intellectual Property Policy and Attractiveness: A Longitudinal Study of Free and Open Source Software Projects

Carlos D. Santos Jr.
FLOSS Competence Center (CCSL)
Dep. of Computer Science (DCC)
Institute of Math and Statistics (IME)
University of Sao Paulo (USP)
Sao Paulo, SP, Brazil
denner@ime.usp.br

Marcos Cavalca
CCSL-DCC-IME-USP
mbonci@linux.ime.usp.br

Fabio Kon
CCSL-DCC-IME-USP
kon@ime.usp.br

Julio Singer
Department of Statistics
Institute of Math and Statistics (IME)
University of Sao Paulo (USP)
Sao Paulo, SP, Brazil
jmsinger@ime.usp.br

Victor Ritter
Department of Statistics
IME-USP
victor.ritter@usp.br

Damaris Regina
Department of Statistics
IME-USP
damaris.regina@usp.br

Tamy Tsujimoto
Department of Statistics
IME-USP
tamy.tsujimoto@usp.br

CSCW 2011, March 19–23, 2011, Hangzhou, China.
ACM 978-1-4503-0556-3/11/03.

Abstract
This paper reports early findings of a longitudinal study designed to evaluate the impact of changes in the intellectual property policy of 756 free and open source projects on their attractiveness over 44 months.

Keywords
Open source software, attractiveness, licenses, intellectual property, GPL, free software, governance.

ACM Classification Keywords
K.6.1. Management of Computing and Information Systems: Project and People Management.

General Terms
Open source software, attractiveness, licenses, intellectual property, GPL, free software, governance.

Introduction
Searching for peers to collaborate with in projects is a common mission in many fields, especially in science and software engineering [5, 11]. Recently, this has been stimulated by the decreasing costs of searching for collaborators and communicating with them via the Internet. Also, the asynchronous dynamic of working over the Web has led many investigators to engage in geographically distributed projects. This phenomenon is

particularly evident in the production of free and open source software (free software, for short). There are hundreds of thousands of free software projects online, each representing a computer supported cooperative work opportunity.

Free software projects (FSP) reflect the intention to share costs of continuous software improvement, user base expansion and visibility growth [8]. Arguably, Linus Torvalds was the first to publicize this intention, bringing forth Linux, a project that has achieved unprecedented size and impacted the IT industry deeply. Unsurprisingly, nowadays, many organizations create FSP as a deliberate strategy, known as opensourcing, an alternative to outsourcing [1]. When successful, FSP involve active communities maintaining an ever evolving public software through a resourceful communication channel between users, developers and sponsors. Nevertheless, in these terms, success has been achieved only by a small fraction of the total number of FSP.

In this scenario, knowledge on how to effectively create and manage FSP to suit better the demands of stakeholders is useful and timely. The objectives include increasing adoption and intention to contribute. A major concern of researchers in this area has been the terms under which the application source code is allowed to be modified, used and distributed. According to previous studies, the intellectual property policy (IPP) delineated by the chosen license has the power to drive people and organizations away from adopting and contributing to FSP, and operates as a governance mechanism, impacting production dynamics [2, 3, 7, 9, 10]. In summary, the predominant thesis in the literature is that IPP influences FSP's attractiveness and, thereby, production dynamics [8]. However, this thesis is yet to pass one of the most rigorous tests of causal statements: a longitudinal study with a large sample observed *in natura* over a wide time frame. This paper closes this gap by asking "Does an Intellectual Property Policy intervention influence FSP Attractiveness?".

Definitions and Model

Free and Open Source Software Projects

In general, projects are endeavors toward goals, such as to write a paper or develop a software. When a software project has its source code freely and publicly available online for use and modification, it may be classified as a free and open source software project [7]. FSP are the object of interest to this study. Several of them have become widely known, such as the GNU/Linux operating system, the R statistical package and the Apache HTTP server. The communities maintaining these systems are large, active and professional, producing first class applications in their domains. However, most FSP have not become successful, never attracting external users and contributors to build a productive network of peers.

The Role of Attractiveness

One way to understand why some FSP are successful and others are not is through the study of their attractiveness [8]. Attractiveness is a common cause of how many visitors a project website receives, how many users it has, and how many contributors it possesses. Higher attractiveness leads to more activeness and efficiency, improving software quality and facilitating innovation via the "more eyeballs effect" [3, 6, 8]. Given that, it is clear how important it is to understand what influences FSP attractiveness.

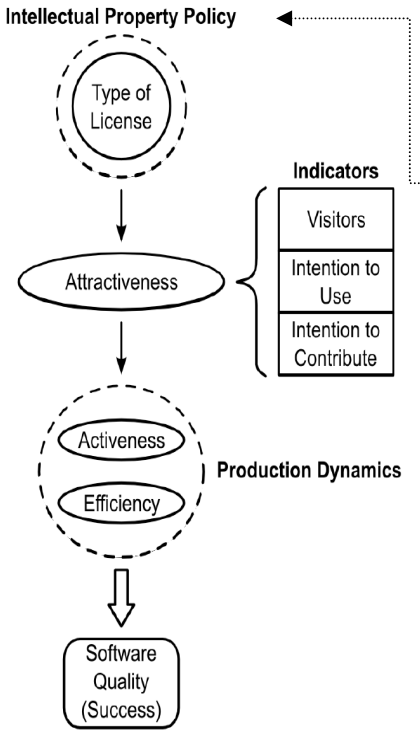

Figure 1. Theoretical Model.

Intellectual Property Policy

The choice of license impacts FSP success because it defines the scope of doing business from distributing the software and may prevent the free-riding/hijacking problem, thereby, impacting the citation mechanism and the perception of control over the technology [7, 10, 12]. With that in mind, people and organizations take the license into consideration on deciding whether to adopt and use a free software and, later, whether to contribute to the project. **Figure 1** depicts this thesis.

This study focus is on four IPPs that may be applied to the source code. The first relates to whether the source code is "restrictive", requiring derivative works to be released under the same license [3]; the second, to whether it is "highly restrictive", forbidding the source code to be mingled with software of different licenses [3]; the third, to whether the code may be relicensed, meaning that "any distributor has the right to grant a license to the software […] directly to third parties" [7, p. 88]; and, the fourth, to whether a project is licensed under the Academic Free License, since it was written to correct problems of important licenses such as MIT and BSD [7] and is understudied (see **Table 1**: Legend for all types of IPP and examples).

Methods: Sampling and Analyzing

The sampling procedure adopted was specifically designed to rigidly test the impact of IPP on FSP attractiveness. To do so, only projects from the largest, most used in research, online repository [2] that had changed their type of license over the years covered by FLOSSmole.org were selected. From this initial sample, projects that had their listing categories or audiences changed were selected out to control for possible confounding effects. In summary, the sample contains FSP that managers intervened only on IPP. The final sample is of 756 FSP over a period of 44 months, from October/2005 to June/2009, missing one.

For each project, licenses were collected and classified based on the IPPs of focus; and monthly data on Web hits (visitors), downloads (intention to use) and members (intention to contribute) were gathered. To calculate "attractiveness", the correlation matrix of a previous study [8] was used in a principal component analysis [4]. The principal component is defined as (0.63*logwebhits + 0.64*logdownloads + 0.43*logmembers) and explains 65% of the sample variance. Data was analyzed using R (r-project.org).

To study the IPP intervention impact on attractiveness, FSP were classified according to the type of intervention they were subjected to, and a ratio of mean attractiveness after/before the intervention was computed. **Table 1** summarizes all ratios observed. For example, **Table 1**: Cell (1, 2) indicates that projects changing from IPP A to B experienced lower levels of attractiveness after the intervention, on average.

Results and Findings

Out of 35 types of intervention observed in the sample, 13 were positive to attractiveness, 21 were negative, and only 1 neutral. In total, 1012 IPP interventions were found (an average of more than one per project).

Taking the initial state (rows of **Table 1**) into account, the most common IPP in interventions is F (417), and it has the largest positive impact on attractiveness (16% on average). The least common origin is C (14), and it impacts attractiveness negatively (4%). The largest negative impact occurs for E (15%), found 49 times.

LEGEND:

A None (e.g., "other" or "adaptive").

B Non-Restrictive and Relicensable (e.g., Public Domain or MIT).

C Academic Free License-AFL (Non-Restrictive and Relicensable).

D Restrictive and Non-Relicensable (e.g., GNU Lesser General Public License-LGPL).

E Restrictive and Relicensable (e.g., Mozilla Public License-MPL).

F Restrictive, Highly Restrictive and Non-Relicensable (e.g., GNU General Public License-GPL).

G Restrictive, Highly Restrictive and Relicensable (e.g., dual licensed: GPL and Apache).

From \ To	A	B	C	D	E	F	G	Mean	Rank
A	--	0.94	1.07	1.06	1.14	1.09	0.87	1.07	2
B	0.96	--	0.97	1.02	1.03	0.98	1.01	0.99	4
C	0.92	0.93	--	--	--	1.05	--	0.96	5
D	0.98	1.05	--	--	0.96	1.03	0.92	0.99	3
E	0.70	0.86	--	0.91	--	0.89	0.89	0.85	7
F	0.89	1.00	2.00	0.98	1.06	--	1.01	1.16	1
G	--	0.85	--	0.98	0.88	0.89	--	0.90	6
Mean	0.89	0.94	1.35	0.99	1.02	0.99	0.94	*0.99*	
Rank	7	6	1	4	2	3	5		

Table 1. IPP Intervention on Attractiveness (Post/Pre Ratio).

Taking the final state (columns of Table 1) into account, the most common intervention is F (298), and it affects attractiveness negatively (1%); the least common target is C (12), and it has the largest positive impact (35%). The largest negative impact relative to the final state relates to A (11%), which was targeted 55 times. By type of intervention (cell), the most common one is from F to G (155), which impacted attractiveness positively (1%); the largest positive impact is from F to C (100%), but it happened only 3 times; and the largest negative impact is from E to A (30%), but it was observed only 3 times too.

Moreover, every intervention that targeted A or originated from E or G impacted attractiveness negatively. Also, although changing from C to B does not change the project IPP, it does impact attractiveness, suggesting that stakeholders prefer AFL to MIT, for instance. Finally, C is the least targeted intervention (12) but has provided the higher benefit (35%), highlighting that AFL is underexplored.

Conclusions

IPP interventions are not always good, and, generally, impact attractiveness. Thus, the importance to carefully select and change the type of license for FSP to (continuously) succeed is reiterated. Future research must persist in pursuing the topic, analyzing this data with inferential statistical techniques to understand the causal relationships deeper and even more rigorously.

Acknowledgements

FAPESP (2009/02046-2) and the QualiPSo Project (EC).

References

[1] Agerfalk, P. and Fitzgerald, B. Outsourcing to an unknown workforce, Exploring Opensourcing as a global sourcing strategy. *MIS Quarterly*. 32, 2 (2008)

[2] Colazo, J. and Fang, Y. Impact of license choice on Open Source Software development activity. *J. of the Am. Society for Inf. Science and Tech*. 60, 5 (2009).

[3] Lerner, J. and Tirole, J. The Scope of Open Source Licensing. *J. of Law, Econ, and Org*. 21, 1 (2005).

[4] Mardia, K. et al. 1980. Multivariate Analysis (Probability and Mathematical Stats). Academic Press.

[5] McCafferty, D. Should code be released? *Communications of the ACM*. 53, 10 (2010).

[6] Raymond, E. The cathedral and the bazaar. *Knowledge, Technology & Policy*. 12, 3 (1999), 23-49.

[7] Rosen, L. 2004. Open Source Licensing: Software Freedom and Intellectual Property Law. Prentice Hall.

[8] Santos Jr., C. et al. Attractiveness of Free and Open Source Software Projects. ECIS, 2010.

[9] Sen, R. et al. Determinants of the Choice of Open Source Software License. *JMIS*. 25, (2008).

[10] Stewart, K.J. et al. Impacts of License Choice and Organizational Sponsorship on User Interest and Development Activity in Open Source Software Projects. *Information Systems Research*. 17, 2 (2006).

[11] Stone, R. Earth-Observation Summit Endorses Global Data Sharing. *Science*. 330, 6006 (2010).

[12] Watson, R.T. et al. The business of open source. *Communications of the ACM*. 51, 4 (2008), 41-46.