

RESEARCH

Open Access



Changes in free and open source software licenses: managerial interventions and variations on project attractiveness

Carlos Denner dos Santos Jr

Abstract

The license adopted by an open source software is associated with its success in terms of attractiveness and maintenance of an active ecosystem of users, bug reporters, developers, and sponsors because what can and cannot be done with the software and its derivatives in terms of improvement and market distribution depends on legal terms there specified. By knowing this licensing effect through scientific publications and their experience, project managers became able to act strategically, loosening up the restrictions associated with their source code due to sponsor interests, for example; or the contrary, tightening restrictions up to guarantee source code openness, adhering to the “forever free” strategy. But, have project managers behaved strategically like that, changing their projects license? Up to this paper, we did not know if and what types of changes in these legal allowances project managers have made and, more importantly, whether such managerial interventions are associated with variations in intervened project attractiveness (i.e., related to their numbers of web hits, downloads and members). This paper accomplishes these two goals and demonstrates that: 1) managers of free and open source software projects do change the distribution rights of their source code through a change in the (group of) license(s) adopted; and 2) variations in attractiveness are associated with the strategic choice of a licensing schema. To reach these conclusions, a unique dataset of open source projects that have changed license was assembled in a comparative form, analyzing *intervened projects* over its monthly periods of different licenses. Based on a sample of more than 3500 active projects over 44 months obtained from the FLOSSmole repository of Sourceforge.net data, 756 projects that had changed their source code distribution allowances and restrictions were identified and analyzed. A dataset on these projects’ type of changes was assembled to enable a descriptive and exploratory analysis of the types of license interventions observed over a period of almost four years anchored on projects’ attractiveness. More than 35 types of interventions were detected. The results indicate that variations in attractiveness after a license intervention are not symmetric; that is, if a change from license schema A to B is beneficial to attractiveness, a change from B to A is not necessarily prejudicial. This and other interesting findings are discussed in detail. In general, the results here reported support the current literature knowledge that the restrictions imposed by the license on the source code distribution are associated with market success vis-a-vis project attractiveness, but they also suggest that the state-of-the-science is superficial in terms of what is known about why these differences in attractiveness can be observed. The complexity of the results indicates to free software managers that no licensing schema should be seen as the right one, and its choice should be carefully made, considering project strategic goals as perceived relevant to stakeholders of the application and its production. These conclusions create awareness of several limitations of our current knowledge, which are discussed along with guidelines to understand them deeper in future research endeavors.

Keywords: Open source software, Attractiveness, Software license, Intellectual property, GPL, Free software, Governance, Project and people management, Information technology, Software project, Open source

Correspondence: carlosdenner@unb.br
Department of Management (PPGA/ADM), University of Brasilia (UnB), Brasília,
Brazil



© The Author(s). 2017 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

1 Introduction: collective production and legal issues

Society and its creations have become increasingly complex as our body of knowledge grew, and information retrieval technologies evolved. Innovating and competing on a global scale is no activity for an individual alone. Searching for partners and peers to collaborate with and in projects is a crucial task in most fields, notably in science, software engineering and public policy management [1–3]. Experts have noticed this and expressed such notion by saying that modern inventors are organizations, not individuals and that production processes are best dealt with in an open and public fashion, as opposed to the proprietary and private economic model for firm production [3–5]. This change, of course, raises concerns on how the rights of such collective goods (properties) should be regulated and managed as to prevent disincentives for entrepreneurship, cooperation and thus maintain the labor market active and sustainable [6–8].

The digitalization of the world has stimulated this trend of working in collectivities by decreasing the costs of searching for collaborators and using communication technologies to coordinate production activities. The asynchronicity of production activities over the web has led many investigators and developers to engage in geographically distributed projects, such as for software development [9, 10]. For at least the last 20 years, this phenomenon of “collective production” has been particularly prominent in the development of free and open source software (free software, for short), reshaping the information technology (IT) industry as it became a strategic player. Nowadays, there are hundreds of thousands of free software projects online, each representing a computer supported cooperative work opportunity for generating an active and growing ecosystem of users and contributors capable of joint development at an unprecedented scale [11, 12].

Free software projects (FSP) reflect the intention of a founder, the original owner of the property rights, to share costs of continuous software improvement, user base expansion, and visibility growth [13–15]. The ability to attract peers to co-create with the founder is understood as the attractiveness of the project [12]. Richard Stallman and Linus Torvalds are among the first and most famous ones to publicize¹ this type of intention, bringing forth the GNU operating system and Linux, a project incredibly successful that alone impacted the IT industry deeply. Unsurprisingly, inspired by the Linux case, many organizations have created FSP as a deliberate organizational strategy, known as open sourcing, an alternative to the classic outsourcing possibility [11]. When successful, FSP involve active communities structured as networks for the evolution of 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, making the investment of releasing intellectual property to the public and assembling a proper IT infrastructure risky and worth of managerial consideration, as a failed attempt wastes organization’s limited resources [12–16].

In this scenario of uncertainty and competition on whether the attention of users and developers will be obtained, knowledge on how to effectively create and manage FSP to suit better the demands and interests of stakeholders, be a sponsor or a co-developer, is useful and timely. Founders and managers should take into account the stakeholders demands and interests as they expect that to translate on increasing software adoption and intention to contribute (i.e., people reporting and developers fixing bugs). One of the central issues in the literature of open source project affecting intention to adopt and contribute, its attractiveness, is the license terms, the legal specifications under which the software has been released to regulate further improvement and distribution [6, 7, 16–18].

The influence of the license choice has been discussed on many grounds, from a legal [6], strategic [3, 8] and sociological [7] standpoints. The main effects can be summarized as related to people’s motivation in getting involved as some in the community (stakeholders) believe that private property should not be a derivative of a public one; a legal restriction that has been found to scare corporations’ investments away from software obliged to be always free and open (e.g., licensed GPL 2.0). This duality of effects creates a tension where the interests of all cannot be met at once, forcing FSP managers to choose a strategic path and “pick a side” in terms of licensing, distribution rights.

A major concern has been the terms under which the application source code is allowed to be modified and re-distributed. Free software can be modified, and the result of that modification distributed in a sold hardware, for example, and the source code of the embedded software kept proprietary, or not, depending on the license chosen. According to previous studies, the intellectual property policy delineated by the chosen license schema has the power to drive people and organizations away from adopting and contributing to FSP, and operates as a governance mechanism, thereby impacting the attractiveness of the project and consequently its production activities [6–8, 12, 17–19].

In a nutshell, the license is believed to influence FSP’s attractiveness, production activities and, thereby, success. As this strategic effect becomes known to FSP founders and managers, assuming their rationality towards the attempt to be successful, an expectation that they should act in practice and change their project licenses to affect

attractiveness is created. This paper represents a methodological advance in comparison to previous studies, as it verifies this theoretically-derived expectation of a relationship between license and attractiveness by performing a longitudinal study with a large sample observed *in natura* over a wide time frame. This methodological approach was specifically developed towards the answers of the following research questions: 1) *Do intellectual property interventions, license changes, occur in practice?* 2) *Are the different licensing schemas chosen by project managers associated with FSP attractiveness?* These questions are answered with a sampling strategy designed to identify the projects that have changed licenses, followed by a statistical analysis of various types of license interventions that FSP managers have decided to make, changing thereby the legal restrictions of their software (and thereby their project attractiveness). Nevertheless, besides this methodological improvement to the literature found here, this paper also contributes in the sense that most previous empirical studies have considered that an open source project has only one type of license, even though many of these projects have more than one. This paper incorporates that in its methodological procedures and improves the classic way of classifying licenses based on Lerner and Tirole²'s work in a more realistic, empirically-based schema. Furthermore, the unique dataset assembled to produce this paper is released open, free of charge along with its publication, which is another form of contribution to future research endeavors Additional file 1.

The scientific basis grounding the theoretical expectations just spelled out are next stated in more details, a foundation followed by a methods section describing the specific steps followed to obtain the sample and results discussed before the conclusions.

1.1 Theoretical foundations: definitions and related work

1.1.1 Free and open source software projects

In general, projects are endeavors toward goals, such as writing a paper or developing software. When a software project has its source code freely and publicly available online for use and modification with a license specifying that attached to it, it may be classified as a free and open source software project [7, 8, 11, 12]. Free software projects (FSP) are the object of interest to this study for their position as key players in the IT industry. Several of them have become widely known, such as the GNU/Linux operating system, the R statistical package, and the Apache web server. The communities maintaining these systems are large, active and professional, producing first class applications in their domains and receiving sponsorship from companies such as IBM and Google. However, beyond these high-class applications, most FSP has not become successful, never attracting external users and contributors to generate a network of peers producing useful, up-to-date public software freely available [12–14].

1.2 The role of attractiveness

One way to understand why some FSP are successful and others are not is through the study of their attractiveness [12], or their “magnetism and stickiness” as some have more informally stated.³ Attractiveness is a common cause of how many visitors a project website receives, how many users it has, or its number of downloads, and how many contributors it possesses. FSP attractiveness is a concept considered responsible for the (lack of) flow of market resources, basically time and money, to the project. Higher attractiveness leads to more intention to adopt (download) and contribute (become a member), motivating and justifying production activities and investments towards the software to improve quality and generate innovation via the “more eyeballs effect” [12, 19, 20]. FSP attractiveness has a vital role in this perspective, and it is evident how important it is to understand what influences or is associated with attractiveness variations.

1.3 The choice of license and FSP success

The choice of license impacts FSP success because it defines the scope of doing business with the distribution of the software and its derivatives, perhaps preventing the source code hijacking, or impacting the reuse or “citation” incentive, but for sure influencing stakeholders’ perception of control and utility over the technology. People and organizations take the license terms into consideration on deciding whether to adopt and use free software and, later, if it is worthy contributing to or reusing the source code [7, 8, 16, 21]. Figure 1 depicts this thesis causal chain, from intellectual property choice to attractiveness and then software quality/project success.

In summary, based on the literature review in which this study is grounded [8, 12], Fig. 1 can be read from left to right as FSP managers select a license that defines the restrictions applied to the source code redistribution, which affects the flow of market resources to the project (visits to the website: visitors, downloads: intention to use, and membership: intention to contribute). As a consequence of an increase in the project attractiveness, with more people thus interested in the software quality, more bugs will be reported and fixed, and new features will be requested and developed, influencing directly in the project long-term success. Accordingly, this causal chain is expected to be “disturbed” by a managerial intervention/change in the project license, as the interests of relevant stakeholders (sponsors, volunteers, etc.) might not be met anymore.

To explore empirically this hypothesis, based on what has been done in previous research [8, 12, 21, 22], this study focuses on four types of legal restrictions that may be applied to the free and open source code. The first relates to whether the source code is “restrictive”, requiring derivative works to be released under the same

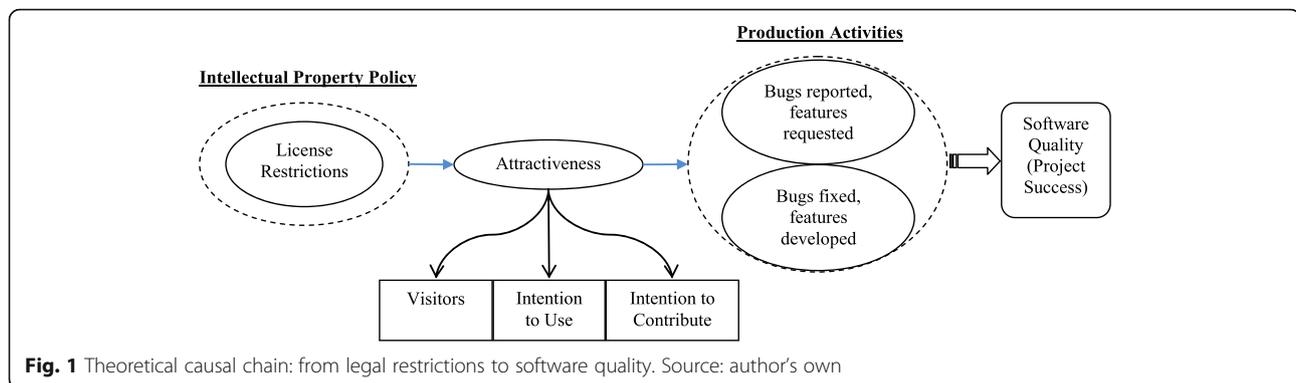


Fig. 1 Theoretical causal chain: from legal restrictions to software quality. Source: author's own

license in case of redistribution [19]; the second, to whether it is “highly restrictive”, which besides being restrictive, forbids the source code to be even mingled for compilation with software of a different license [19]; 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. Methodologically speaking, projects licenses were classified in this basis, including the cases where a project would have more than a license. Therefore, in this schema, a project might not have a restriction for one group of stakeholders, students for example, but do have that restriction for corporations. This methodological choice reflects the reality of open source projects more accurately but has the downside of being more complex, as the results will demonstrate for themselves.

The basic sampling strategy idea that guided this research was to look for projects that have undergone a change in these legal terms during their life-cycle and verify possible associations/variations on the main indicators of attractiveness of such projects. This approach aims to uncover whether FSP managers change legal restrictions over their projects life-cycle (research question #1-RQ1) and evaluate whether the success of FSP is associated with the legal terms change through a before-and-after statistical analysis of a managerial intellectual property intervention (IPI) on project attractiveness (research question #2-RQ2). These intents together have not been addressed in previous research with such methodological approach.

2 Methods: data, sampling and statistical analyses

To obtain data capable of answering the questions of whether FSP managers have performed changes in their schema of licensing over the years (RQ1), and whether these changes are associated with the attractiveness of the project (RQ2), a search on the internet for secondary

data on free software projects was made. A few options popped up, such as the University of Notre Dame based, but the more seemingly straightforward one was chosen, FLOSSmole [23].⁴ Data obtained and released by FLOSSmole on all projects from the largest⁵ free software repository available online [6] at the time of this project data collection efforts was organized in a database for inspection, covering 44 months of activities. This database was filtered down to contain only those projects that have changed their listed licenses over the years covered in the obtained dataset. If this filtered dataset was equal to zero projects, the first research question of this paper would be “no, FSP managers have not changed their license schema, despite the known effect of that on attractiveness found in previous research”. But the empirical answer is yes, FSP managers have made these interventions (aka, IPI) hundreds of times in this research sample.

After obtaining this working sample, a data organization process was performed, classifying the various licenses of projects (many have more than one license at a given point) into the categories described right after Fig. 1 shown above. All information on the project audience (end-user or developer, for example), date of creation, etc. was also kept for sample description, and data on numbers of web hits, downloads, and members were gathered monthly to allow for comparisons on these indicators of attractiveness anchored on the type of licensing schema intervention. The choice of these specific indicators is aligned with previous research [12], where attractiveness was first directly addressed in the specialized literature. A few more details on this data preparation procedure are described below.

The sampling and filtering procedures adopted were specifically designed to detect the changes in license terms adopted by FSP managers and explore if these IPI are associated with FSP attractiveness variations. As the ideal methodological situation of random selection of projects to undergo a license change is not possible due to the impossibility of doing that with other people's project (this is not an experiment), alternatively, to

control for confounding effects, projects that had their listing categories or audiences changed during the period covered by this study were selected out too. Also, any project with missing data on the number of members was also removed from the sample, as this indicates an “orphan” project. The working sample is of 756 FSP with monthly data covering a period of 44 months, from October/2005 to June/2009 (1 month was missing in FLOSSmole, July 2008).

For each project, monthly data on its license were collected for further classification based on the legal restrictions covered in this paper, as explained before. This classification set forth here is based on previous research, which has always treated licenses by their restrictions of 1) compatibility for mingling with a different software during compilation (when not, referred to as “highly restrictive”), 2) whether an improvement of a software must be released as free software as well (when yes, referred to as “restrictive”), and 3) whether a software might be relicensed by third party to a different license originally chosen (referred to as “relicensable”). However, the empirical fact that projects have more than one license challenges a classification that considers a project simply based on one of its licenses. Free software projects choose schemas of licensing, for example, with a “highly restrictive” stamp for non-payers, and a “relicensable” option for who pays for the software. The classification adopted here takes that into account to obtain a more accurate however complex picture of projects licensing schema. All listed projects’ licenses were considered, and so a dual-licensed project might indeed be “Restrictive, Highly Restrictive and Relicensable”, something that at first sight can appear contradictory. This classification was performed per month, and changes in the schema, managerial interventions detected were flagged for further analysis.

3 Results and Findings: descriptive statistics towards RQ1

Table 1 summarizes all interventions detected along with labels given to them (see column “description”), and the number of occurrences of each type of change in legal terms is displayed in the table cells. This table represents the detailed answer to RQ1. One can see, for example, that GPL was involved in the managerial interventions 715 times (being the end-state 298 times, the sum of column F, and a beginning state of the change 417 times, the sum of row F). In the description column, one can see that the GPL is restrictive and highly restrictive, that is, derivative work redistributed must be GPL as well, and source code mingled with it during compilation must be GPL as well (a “viral” license). Further, GPL software cannot be relicensed under a different license. GPL is thus restrictive, highly restrictive and non-relicensable. GPL motivates the most managerial interventions, probably due to its popularity and mixed feelings of the community with its adoption (loved by those who believe in “free software forever” and not so much by those primarily guided by competitive motivations). This GPL leadership is followed by the dual-licensing strategy, where FSP managers decide to release code under different licenses depending on the interest and profile of the user (e.g., whether an individual or a for-profit organization). These interventions ranking and the number of their occurrences can be found on Table 1’s *column* for data related to the new license type chosen to be adopted, and on its *rows* for the data about the license type abandoned by the project (the “from” and “to” indicated in the first cell of the second row).

Additionally, monthly data on Web hits (visitors), downloads (intention to install the software use) and a number of members (intention to contribute reporting bugs or features), besides the type of project and development stage, were gathered. Table 2 contains the

Table 1 Number of occurrences and types of intellectual property interventions

From\To	Description	Count of license type interventions in sample							Sum	Ranking
		A	B	C	D	E	F	G		
A	None (or “other”).	0	22	2	13	3	47	1	88	5
B	Non-Restrictive and Relicensable (e.g., Public Domain or MIT).	8	0	7	20	16	31	45	127	4
C	Academic Free License-AFL (Non-Restrictive and Relicensable).	2	5	0	0	0	7	0	14	7
D	Restrictive and Non-Relicensable (e.g., GNU Lesser General Public License-LGPL).	6	34	0	0	21	67	6	134	3
E	Restrictive and Relicensable (e.g., Mozilla Public License-MPL).	3	19	0	12	0	7	8	49	6
F	Restrictive, Highly Restrictive and Non-Relicensable (e.g., GNU General Public License-GPL).	36	81	3	137	5	0	155	417	1
G	Restrictive, Highly Restrictive and Relicensable (e.g., dual licensed: GPL and Apache).	0	32	0	6	6	139	0	183	2
Sum		55	193	12	188	51	298	215	1012	
Rank		5	3	7	4	6	1	2		

Source: author’s own

Table 2 Illustrative descriptive statistics for the first month

Variable	Minimum	Maximum	Mean	Std. deviation
<i>Registered</i>	11/04/1999	3/13/2009	1/08/2003	-
<i>n_licenses.200510</i>	0	4	1.10	0.39
<i>attractiveness.200510</i>	0	16.12	5.4694	3.33
<i>downloads.200510</i>	0	34,514	378.07	1941.56
<i>webhits.200510</i>	0	836,740	9267.23	48,727.7
<i>members.200510</i>	1	55	3.38	5.22

Source: Author's own

descriptive statistics for the numerical variables, and Table 3 the frequency of projects that have a particular type of license versus their development status in the first month of the dataset, October of 2005. To calculate “attractiveness,” a latent construct, the correlation matrix of a previous study [12] was used in a principal component analysis [24], where a linear combination of three indicators of attractiveness was identified to maximize the explained variance. The first principal component extracted is operationally defined as $(0.63 \cdot \log \text{webhits} + 0.64 \cdot \log \text{downloads} + 0.43 \cdot \log \text{members})$ and explains 65% of the sample variance. This first component extracted was used to calculate a new variable named attractiveness, a result of the multiplied sum of projects log-transformed web hits, downloads and number of members at any given month. This measure of attractiveness expresses the ability of a project to attract these market resources from the environment where it competes with other projects. Attractiveness is thus a common cause of website visits, downloads and membership numbers. Data was organized and statistically analyzed with R.

From Table 2, one can see that in the sample: 1) projects were founded as early as 1999; 2) on average a project had approximately 378 downloads in October 2005; and that at least one project has four different licenses listed at this point. Table 3 depicts a different picture, showing that: 1) 48% of projects, 363, are licensed GPL (restrictive + highly_restrictive + non-relicensable) and out of these 95 are in beta stage; 2) 11% of the 756 projects have no license specified; and 3) only 7 projects have no license and no development status on their file at October/2005. This distribution of projects in the sample demonstrates a wide variability over the various stages of software lifecycle, reducing once more the limitations of non-experimental nature of this study and its potential sampling biases.

4 Results and Findings: preparing to answer RQ2

To explore the IPI associations with attractiveness variations and obtain some if any statistical evidence of variation, FSP were classified according to the type of intervention they were subject to every month, and the working sample was again organized and analyzed in the following fashion.

To allow for statistical comparisons with reasonable sample sizes, the dataset was reorganized to display the seven licensing schemas, from A to G on columns, and attractiveness on the rows. In this new dataset, each cell represents the attractiveness of a project in a specific month, broken by licensing schema with the various columns. This analytical strategy of treating the licensing

Table 3 Frequency cross-table for type of license and development status

Type of license/development status		Alpha	Beta	Mature	None	Planning	Prealpha	Stable	Total
A	#	12	18	2	7	7	7	33	86
	%	1.6%	2.4%	0.3%	0.9%	0.9%	0.9%	4.4%	11.4%
B	#	25	36	4	2	15	11	33	126
	%	3.3%	4.8%	0.5%	0.3%	2.0%	1.5%	4.4%	16.7%
C	#	1	1	0	2	1	0	3	8
	%	0.1%	0.1%	0%	0.3%	0.1%	0%	0.4%	1.1%
D	#	22	33	2	5	15	13	38	128
	%	2.9%	4.4%	0.3%	0.7%	2.0%	1.7%	5.0%	16.9%
E	#	4	10	0	1	1	3	6	25
	%	0.5%	1.3%	0%	0.1%	0.1%	0.4%	0.8%	3.3%
F	#	84	95	9	8	36	38	93	363
	%	11.1%	12.6%	1.2%	1.1%	4.8%	5.0%	12.3%	48.0%
G	#	4	7	0	0	2	3	4	20
	%	0.5%	0.9%	0%	0%	0.3%	0.4%	0.5%	2.6%
TOTAL	#	152	200	17	25	77	75	210	756
	%	20.1%	26.5%	2.2%	3.3%	10.2%	9.9%	27.8%	100%

Source: Author's own

schema and not the specific change of the schema increased the sample size immensely and permitted statistical mean comparisons of attractiveness, as RQ2 required. The classic t-test, robust to violations of assumptions with such large samples,⁶ was performed using the software SPSS.

The descriptive statistics, variable by variable, for this new dataset is shown below in Table 4, and it is possible to see that the smallest sample size is 265, which means that in 33,264 month-projects available (756 projects times 44 months), 265 month-projects could be flagged with a C type of schema.

5 Results and Findings: revisiting RQ1 towards RQ2

A project license or schema of licensing imposes restrictions and allowances to the application adopter and source code contributor, creator of a derivative work. For example, a company that customizes a GPL application and distributes it in the market is obliged to make the source code of the redistributed, improved public software. The license choice is a strategic decision with social and economic impacts on the project, as it can block the interests of people related to the software, that is, users, developers and other relevant stakeholders. A major decision like that is not expected to occur very often, as managers avoid *status quo* changes that harm expectations and turn people’s attention away from the actual work (e.g., into politics and disputes). This tendency to not change strategic matters is known in the organizational literature as structural inertia [25].

In conformance to this organizational inertia, out of thousands of free software projects obtained from FLOSSmole and Sourceforge.net and analyzed in this research, only 756 have decided to change their license type over the period of 44 months covered in this research, from October/2005 to June/2009, missing July/2008. Nevertheless, as it has already been shown in Table 1, these 756 projects that changed licenses have done so 1012 times, a considerable number that validates the theoretical expectation of managerial action through changes in software legal restrictions towards

Table 4 Descriptive statistics for mean comparisons by licensing schema

	Sample size	Minimum	Maximum	Mean	Std. Dev
A_attractiveness	2134	0.44	14.60	6.9037	2.81252
B_attractiveness	5322	0.00	16.12	6.4007	2.75651
C_attractiveness	265	0.00	11.12	6.4196	2.09862
D_attractiveness	5522	0.30	14.46	6.7547	2.31416
E_attractiveness	1073	0.30	15.97	7.4004	2.74449
F_attractiveness	9849	0.00	16.83	6.6443	2.88175
G_attractiveness	1865	0.44	18.01	7.6265	3.08157

Source: Author’s own

meeting stakeholders’ demands and expectations for project success. Previous research has stated that the license affects the probability of project success and, accordingly, FSP managers have indeed attempted changes in legal restrictions.

In terms of specific results, leaving projects exposed and legally unattended, the managerial decision of not having a license specified was detected both ways, as projects left the “none” choice 88 times and, surprisingly, changed their current state of having a license to one where they have no license 55 times (see Table 1, type of license A). In fact, it has been found that projects have had no license specified in every month covered by this research. FSP with no license, the “none” A-category created, have less average attractiveness than restrictive/relicensable and dual-licensed projects often, but have more attractiveness than GPL (F-schema). Let us now move one step further to analyze the data numerically.

To initially explore the statistical associations of attractiveness and license, the ratios of mean attractiveness after/before interventions were computed, considering all projects of a given change in licensing schema (summarized in Table 5). For calculating the ratios, it was summed up for all projects of a specific license, after the attractiveness component was calculated for standardization. It is the sum of the attractiveness of all projects in a state of license change for each type of change. Projects were aggregated and afterwards one ratio was calculated by dividing their mean attractiveness after the change by their mean attractiveness before the change.

To interpret the results in Table 5, for example, one can see that the ratio of 0.94 in the first row indicates that projects changing from type of license A to B experienced lower levels of attractiveness after the intervention, that is, moving away from a status of having no license (A) and going to a status of “public domain” license (B) is on average detrimental to attractiveness (specifically, a reduction

Table 5 Types of license interventions and variations in attractiveness (post/pre ratios)

From*\To	Intervention variations on attractiveness (ratios)						
	A	B	C	D	E	F	G
A ^b	–	0.94	1.07	1.06	1.14	1.09	0.87
B ^{acdf}	0.96	–	0.97	1.02	1.03	0.98	1.01
C ^{bf}	0.92	0.93	–	–	–	1.05	–
D ^{beg}	0.98	1.05	–	–	0.96	1.03	0.92
E ^{dg}	0.70	0.86	–	0.91	–	0.89	0.89
F ^{bc}	0.89	1.00	2.00	0.98	1.06	–	1.01
G ^{de}	–	0.85	–	0.98	0.88	0.89	–

*Superscript letters indicate an asymmetric effect of interventions, that is, going from one license to another has a similar effect (e.g., it is both positive to leave C and go F, and to leave F and go to C)

Source: Author’s own

of 6%). However, that strategic move has been detected only 22 times in the sample (see Table 1), imposing a limitation to any robust statistical analysis of such variation in attractiveness. This limitation is overcome later in the analysis, with the t-tests as described in the methods section.

Moving ahead with this exploratory results interpretation, as for the associations of the odd managerial action of moving away from having a license specified to not having one (type of change with “A” as target) with attractiveness variations, the average attractiveness ratio of projects that have undergone this type of change have been found to be always detrimental to attractiveness (column A of Table 5), demonstrating that stakeholders do not like the uncertainty associated with a project with no license. By looking at the interventions with A (the none choice) as target in Table 5, it is noteworthy that every time such change was made, the average project attractiveness decreased (a number smaller than one indicates the attractiveness ratio of after/before the change is on average pushed down). Additionally, when a project went from none to a restrictive and relicensable choice (A → E), this change was associated with an average change of 14% in attractiveness.

From a distinct perspective, interestingly, the intervention from none to non-restrictive and relicensable (e.g., MIT), and to restrictive, highly restrictive and relicensable (i.e., dual licensed) led to an attractiveness reduction (see from A → B and A → G in Table 5). At this moment, one can only wonder the actual reasons for such findings in a case-specific manner, but the general theoretical interpretation is that relevant stakeholders’ interests were harmed due to the project license change, affecting its consequent attractiveness.

Together, these findings related to the managerial decision of having no license specified can probably be interpreted in several ways, such as a sign of a not welcoming market to unregulated software, easier to suffer litigation, if you consider that a managerial change to not having a license specified is always detrimental. However, from another perspective, projects with no license can still be considered attractive, suggesting the possibility that the regular user does not take the license into account at all. Perhaps both explanations are valid and complementary, as the attractiveness measure adopted in this research groups the effects on developers and users together (downloads and membership numbers), and only future research can sort this out. Attractiveness is a cause that these variables have in common, but most likely it is not the only one (the first principal component extracted, for example, explains 64% of the variance, and so 36% is not due to this attractiveness measure). Future studies can dig into this line of inquiry, studying these indicators separately as well.

Back to the results interpretation, by focusing on the most popular choice, the GPL, or more generically, the most restrictive licensing (i.e., restrictive, highly restrictive and non-relicensable – the F-schema), it has been found beneficial to projects to abandon this scheme for source code regulation concerning attractiveness increase. Overall, a positive variation with such change in terms of attractiveness was detected, but such strategic move was detrimental to FSP attractiveness when projects went to “none” (A), or restrictive and non-relicensable (D), that is, normally, to the LGPL option (see changes involving F in Table 5). In support of these results, to become GPL was good to FSP attractiveness when the initial state was the absence of license (option A), the Academic Free License (C), or the LGPL one (D). These strategic interventions were detected 47, 7 and 67 times, respectively (Table 1). When taken together, these findings suggest that it is good to avoid the GPL, but it is better to adopt it when compared to having no license or the LGPL. The more challenging explanation for the findings of this type of change is the intervention from GPL to AFL (F → C) and the opposite (C → F), which are both positive. This means that it is good to change from GPL to Academic Free License, and it is also positive to change to GPL coming from the Academic Free License. This suggests that any change might be good to the project, depending on whether such change is aligned with FSP stakeholders’ demands. The (lack of) symmetry on the effects of interventions can be better observed by looking at the matrix shown in Table 5 (*the superscript letters*), a pattern of the findings dealt with in details later in this section.

Analyzing all interventions together, out of 35 types observed in the sample, 13 were positive to attractiveness, 21 were negative, and only one neutral. In total, 1012 intellectual property interventions were found (an average of more than one per project). When taking the initial state (involves F in Table 5) into account, the most common managerial intervention is F (detected 417 times), and it has a consistent positive impact on attractiveness. The least common origin is C (14), and it is associated with a negative change in attractiveness. The largest negative impact occurs for the abandonment of E (15%), which was found 49 times. The mixed results apparent in a visual inspection of Table 5’s coloring scheme suggests that interventions on types of licenses do not always come for good, and that there is always an impact, although only exploratory not statistical here, on attractiveness (the only exception is F to B). This reinforces the importance to carefully and strategically think through the decision, as its impacts do not seem to be irrelevant regarding associated changes in attractiveness.

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 type of license in terms of the restrictions analyzed in this research, it does impact attractiveness, suggesting that stakeholders prefer AFL to MIT, for instance, which makes sense as AFL was designed to improve MIT and that was the reason to include it separately in this study. However, the actual reasons for such finding should be an object of future research, as it suggests there is more to the licensing scheme as this quantitative research captures.

Finally, going from G to B led to a reduction of 15% on attractiveness. The dual-license option that G represents signals to projects' stakeholders that the software is suitable for a wider audience as this intellectual property model can accommodate the interests of various groups, being more market flexible (a generic strategy). Moving away from this management model appears to push attractiveness down, always, as mentioned before (a focused strategy).

6 Results and Findings: the asymmetry of effects and the statistical answer to RQ2

The lack of symmetry of effect is interesting and deserves further consideration. None of the types of licensing schemas analyzed in this research escapes from this. All the licensing schemas have asymmetric effects with at least one other type of license. The most contradictory type of license is B, which has symmetric effects only with E and G. The least contradictory scheme is A, having the opposite effect on attractiveness only when B is involved (see the superscript letters in Table 5). This finding suggests that a match between licensing scheme and projects' specific stakeholders might exist, or the direction of the effect of a given license would simply be reversed depending on whether it is the source or the destiny of the intervention. The suitability of one license schema is likely to rely on the context of its adoption, that is, on the momentary demands of stakeholders, and thus no combination of license should be treated as ideal in general, but only in specific according to stakeholders' expectations on a project-by-project basis.

Now towards the statistically based answer to RQ2, the results here reported were further analyzed. The reorganized dataset with mean month-projects attractiveness per licensing schema was subjected to analysis (see Table 4 for descriptive statistics). But, before getting into the mean difference comparisons (t-tests), the values for mean attractiveness for all the time were considered. These results taken together signal that less restrictive licenses are more attractive on average, as dual license beats the academic unrestrictive schema (e.g., MIT), which in turn is more attractive than the GPL highly restrictive choice. The conclusion is that the project attractiveness varies according to license schema

consistently. Of course, this analysis is basic in statistical terms, but what is clear is that variations on attractiveness indicators are associated with the licensing schema chosen by the FSP manager. The t-tests performed below give further confidence on the answer of RQ2.

As explained before, for the mean statistical comparisons, the monthly data was aggregated to increase sample size, as explained before, and the mean differences between each pair of licensing schema was calculated, along with the standard deviation of these differences and subsequent confidence intervals for statistical significance determination. The results are presented in Table 6 below, which considers if the mean difference is significant at 0.05 type I error with the Bonferroni correction procedure applied (marked with *), and the effect size of each pair of licensing schema based on Cohen's D^7 (marked with ^a).

According to the results shown in Table 6, one can see that 11 out of 21 pairs are statistically significantly different, using the most conventional statistical procedure to control for inflated alpha in the context of multiple comparisons (Bonferroni). Out of this 11, 4 have effect sizes between small and medium but significant according to Cohen's D famous suggested interpretations (higher than 0.2). This signals that the licensing schema is indeed associated with the average numbers of web hits, downloads and members a project can attract. These differences in absolute numbers and effect sizes between schemas peaks at the C-G pair, with a - 1.35 mean difference in favor of the dual license schema when a project moves away from the AFL license option. The rest of the results for each pair of licensing schemas can be found in Table 6.

Overall, these statistical results and analysis on the variations of attractiveness taken together allow for a solid answer to the second research question posed here in this paper of whether an intellectual property intervention (a managerial change in licensing schema). The licensing schema is indeed associated with variations attractiveness level, not in all, but in many cases, having a meaningful effect size in a few of them. In the next section, the general conclusions are discussed based on the answers found for both research questions, presenting directions for future research and guidelines for free and open source software managers.

7 Conclusions: implications to research and practice

This research focused on intellectual property rights interventions in free and open source software projects (FSP), on licensing schema changes that regulate the distribution allowances of the software source code under the hypothesis that such managerial interventions would affect stakeholders' perceptions of value and thus variations of FSP attractiveness before and after the managerial intervention could be observed.

Table 6 Statistical tests for attractiveness mean differences

one MINUS another licensing schema	Paired differences of attractiveness					t	df	p-value	Cohen's D	
	Mean	Std. Deviation	Std. Error Mean	99% confidence interval						
				Lower	Upper					
Pair 1	A - B	0.22	3.97	0.10	-0.02	0.47	2.33	1735	0.02	0.06
Pair 2	A - C ^{*a}	0.90	3.87	0.27	0.19	1.61	3.28	200	0.00	0.23
Pair 3	A - D	0.04	3.74	0.09	-0.19	0.27	0.50	1751	0.62	0.01
Pair 4	A - E	-0.31	4.08	0.14	-0.66	0.04	-2.26	895	0.02	-0.08
Pair 5	A - F	0.03	3.91	0.09	-0.21	0.27	0.34	1749	0.74	0.01
Pair 6	A - G*	-0.67	4.22	0.11	-0.95	-0.39	-6.24	1553	0.00	-0.16
Pair 7	B - C	0.39	3.44	0.25	-0.24	1.03	1.61	195	0.11	0.11
Pair 8	B - D*	-0.38	3.57	0.05	-0.51	-0.24	-6.95	4356	0.00	-0.11
Pair 9	B - E*	-0.57	3.84	0.13	-0.91	-0.22	-4.26	829	0.00	-0.15
Pair 10	B - F*	-0.32	3.95	0.06	-0.48	-0.17	-5.44	4419	0.00	-0.08
Pair 11	B - G ^{*a}	-0.90	4.19	0.11	-1.18	-0.62	-8.22	1458	0.00	-0.22
Pair 12	C - D	-0.06	3.44	0.25	-0.72	0.59	-0.26	186	0.80	-0.02
Pair 13	C - E ^{*a}	-0.90	3.60	0.24	-1.53	-0.28	-3.77	224	0.00	-0.25
Pair 14	C - F	0.09	3.23	0.23	-0.50	0.69	0.40	197	0.69	0.03
Pair 15	C - G ^{*a}	-1.35	3.79	0.25	-2.01	-0.68	-5.28	220	0.00	-0.36
Pair 16	D - E*	-0.60	3.67	0.13	-0.93	-0.27	-4.70	817	0.00	-0.16
Pair 17	D - F	0.07	3.66	0.05	-0.07	0.21	1.28	4614	0.20	0.02
Pair 18	D - G ^{*a}	-0.84	3.95	0.10	-1.10	-0.57	-8.17	1490	0.00	-0.21
Pair 19	E - F*	0.65	3.85	0.13	0.31	0.99	4.88	836	0.00	0.17
Pair 20	E - G	-0.31	4.07	0.13	-0.65	0.04	-2.29	927	0.02	-0.08
Pair 21	F - G*	-0.77	4.16	0.11	-1.05	-0.50	-7.20	1504	0.00	-0.19

*indicates significance at 0,05 with Bonferroni correction ($<0,0023 = 0,05/21$)

Cohen's D calculated as mean divided by std. deviation. Superscript letter A means effect size between small and medium

Source: Authors own

To validate such theoretical expectation, data on thousands of FSP over almost 4 years was filtered to identify a sample of 756 projects that changed their types of licenses, allowing then an empirical study of the various managerial interventions detected in a period of 44 months. These variations were cataloged and organized to allow for comparisons of attractiveness changes grouped by the intervention type, a finding so far missing from the free software literature. Moreover, further reorganization of these original datasets allowed the comparisons of projects' attractiveness to verify whether the licensing schema adopted by FSP managers were associated with the project performance concerning attraction of developers, users and visitors, represented by a liner combination of the numbers of members, downloads and web hits. The classification schema for the licenses adopted by FSP managers developed in this paper also represents a step forward in the literature, as up to now the reality of the adoption of various licenses with apparent contradictory allowances to the source code (with GPL and a public domain license, for example) was not captured in previous research. The result is a more complex but accurate classification, with of course pros and cons.

As for a general conclusion, the results indicate that the legal terms specified in the license are indeed associated with project attractiveness, as an aggregated measure. This is in line with previous research, which led to the expectation that the various business models possible with open source, expressed through their licensing schemas, are related to their success regarding the attraction of users and developers [10, 12, 26]. However, moving beyond the previously published literature, the findings suggest the specifics of such generic hypothesis are not well understood yet.

It has been found that changes in the software rights of distribution, to be fully understood, cannot be treated solely generically, as interventions vary in attractiveness variations associated with them, being beneficial or not depending on much more than what is known from published literature on free software. This research is the first to point that out, providing thus ground for future (case/qualitative) studies to follow the lead and explore the specific reasons for the license intervention and the consequent increase or reduction in attractiveness based on stakeholders' perceptions. Both projects

managers and stakeholders' perceptions should be considered in these future research endeavors.

This future line of inquiry based on case/qualitative studies would be able to shed light on the asymmetric effects detected in the sample as well. Quite often an intervention from one license to another did not have an opposite effect when a change from another to one was analyzed (a vice-versa comparison is not possible). Probably, FSP stakeholders have expectations related to an occasional change that might occur in the license terms of the free software they have the intention to adopt or contribute. This means that depending on the current license (the anchor), the effects of changing to one same license might be different; and that the specific interests of project stakeholders also matter (e.g., hardware production or service sale). Managers should take that into account when considering a license change.

FSP managers should be aware that the success of their projects is linked with their choice of license, as fewer market resources – the attention of users and the labor of developers – might flow in their direction depending on that. This means that managers must understand who are the relevant stakeholders of their application, what they want out of the software source code, and attempt to meet their expectations, carefully considering a change in the licensing only through a direct negotiation with these stakeholders to avoid unwanted consequences. This research indicates that there is no silver bullet concerning right licensing schema, or business model, signaling the general hypothesis here explored needs further elaboration.

Academically speaking, a contingent type of theory to explain the license schema impacts on attractiveness based on context, perhaps stakeholder-based, needs to be developed. To help guide future researchers in that direction, at this moment, it is possible to highlight that a general strategy (multiple licenses) appears to be superior to the specific license schema, as it perhaps accommodates stakeholders' conflicting interests better. This would explain the noticeable trend to adopt the "various licenses" strategy, and demonstrates how important it is to improve the classification schema previously adopted in the literature.

In conclusion, intellectual property interventions are not always beneficial for a free software project, but almost invariably are associated with attractiveness variations. Accordingly, FSP managers should be aware of the importance to carefully select and change the type of license for FSP to (continuously) succeed as a result of a growing market interest in the application and its source code. Nevertheless, such intervention decision should not occur unaware of the specific project under consideration and its stakeholders' intentions with the software in the future.

Nevertheless, methodologically speaking, future research must persist in pursuing the license-attractiveness

relationship, analyzing this longitudinal type of data with more advanced inferential statistical techniques, such as structural equation modeling, to explore and understand the causal relationships better and even more rigorously. The *t*-tests with the Bonferroni procedure applied here is a basic and reliable choice for the problem at hand, but analytical improvements are possible and welcome for a collective, scientific communication towards knowledge accumulation. Another downside of this research is its sample, which was restricted to Sourceforge.com projects. Nowadays there are many other free software repositories that could be considered. Nevertheless, the findings here reported are likely to be constant across these repositories, a hypothesis that future research can verify as well.

Finally, the measures of attractiveness here adopted are another point of improvement to be performed by future research. Only number of web hits, downloads and members were utilized, but other various measures are possible. For example, one could use market share as an alternative, or survey methods, to evaluate attractiveness subjectively. Moreover, attractiveness is probably the consequence of many things besides the license chosen by the project manager, and so other factors should be considered in future research. In this paper, this endogeneity issue was dealt with a sampling procedure that identified projects of various kinds and level of maturity, thereby controlling for some of those effects. Additionally, the results here discussed appear complex but seem to be a more accurate representation of FSP reality. As such, they are in themselves not fully understood, and so future research should use the same dataset, made available along with this paper, with different analytical and theoretical approaches to shed more light on these projects behaviour over time.

8 Endnotes

¹<http://www.gnu.org/gnu/initial-announcement.en.html>

²<http://www.nber.org/papers/w9363>

³<http://dl.acm.org/citation.cfm?id=2597116>

⁴<http://flossmole.org/>

⁵<http://www3.nd.edu/~oss/Data/data.html>

⁶<http://thestatsgeek.com/2013/09/28/the-t-test-and-ro-bustness-to-non-normality/>

⁷<http://nrs.harvard.edu/urn-3:HUL.InstRepos:11718205>

9 Additional file

Additional file 1: Dataset with the raw data used in the research. (CSV 1489 kb)

Abbreviations

AFL: Academic Free License; FSP: Free and open source software projects; GPL: General Public License; IPI: Intellectual property interventions; MIT: Massachusetts Institute of Technology (the license)

Acknowledgements

I appreciate the comments and guidance provided by Professors Julio Singer (statistics, USP) and Fabio Kon (computer science, USP). Their contributions on initial stages of this research were incredibly helpful. I also thank the Center for Technology Development (CDT) of the University of Brasilia (UnB) for the technical help provided in the work of Raphael Saigg. A previous version of this paper was presented at CSCW 2011.

Funding

I thank FAPESP (2009/02046-2) for funding.

Authors' contributions

I am the sole author.

Ethics approval and consent to participate

No need. Only secondary and public data used.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 2 March 2016 Accepted: 12 July 2017

Published online: 07 August 2017

References

- McCafferty D. Should code be released? *Commun ACM*. 2010;53:10.
- Stone R. Earth-Observation Summit Endorses Global Data Sharing. *Science*. 2010;330:6006.
- Sojer M, Henkel J. Code reuse in open source software development: Quantitative evidence, drivers, and impediments. *J Assoc Inf Syst*. 2010; 11(12):868–901.
- Allen RC. Collective invention. *J Econ Behav Organ*. 1983;4:1–24.
- von Hippel E. Cooperation between rivals: Informal know-how trading. *Res Policy*. 1987;16:291–302.
- Colazo J, Fang Y. Impact of license choice on Open Source Software development activity. *J of the Am Society for Inf Sci Tech*. 2009;60:5.
- Rosen, L. *Open Source Licensing: Software Freedom and Intellectual Property Law*. Prentice Hall; 2004.
- Stewart KJ, et al. Impacts of License Choice and Organizational Sponsorship on User Interest and Development Activity in Open Source Software Projects. *Inf Syst Res*. 2006;17:2.
- Raymond, Eric S. *The Cathedral & the Bazaar: Musings on linux and open source by an accidental revolutionary*. O'Reilly; 2001.
- Fitzgerald, Brian. *The Transformation of Open Source Software*, MIS Quarterly (30: 3). 2006.
- Agerfalk P, Fitzgerald B. Outsourcing to an unknown workforce, Exploring Open sourcing as a global sourcing strategy. *MIS Q*. 2008;32:2.
- Santos C, Kuk G, Kon F, Pearson J. The attraction of contributors in free and open source software projects. *J Strateg Inf Syst*. 2013;22(1):26–45.
- Maillart, T., Sornette, D., Spaeth, S., von Krogh, G. Empirical tests of Zipf's law mechanism in open source Linux distribution. *Phys Rev Lett*. 2008;101.
- Wiggins, A., Howison, J., Crowston, K. Heartbeat: measuring active user base and potential user interest in FLOSS projects. In: *Proceedings of the Fifth International Conference on Open Source Systems (OSS)*. 2009. p. 94–104.
- Crowston K, Howison J, Annabi H. Information systems success in Free and Open Source Software development: Theory and measures. *Soft Proc Improv Pract*. 2006;11(2):123–48.
- Vendome C, Linares-Vásquez M, Bavota G, Di Penta M, Daniel M, German DM, Poshyvanyk D. 2015. When and why developers adopt and change software licenses. In *Proceedings of the 2015 IEEE International Conference on Software Maintenance and Evolution (ICSM '15)*. IEEE Computer Society, Washington, DC, USA, 31–40. <http://dx.doi.org/10.1109/ICSM.2015.7332449>.
- Stewart K, Gosain S. The impact of ideology on effectiveness in open source software development teams. *MIS Q*. 2006;30(2):291–314.
- Sen, R. et al. Determinants of the Choice of Open Source Software License. *JMIS*. 25;(2008).
- Lerner J, Tirole J. The Scope of Open Source Licensing. *J Law Econ Org*. 2005;21:1.
- Raymond E. The cathedral and the bazaar. *Knowledge Technol Policy*. 1999; 12(3):23–49.
- Sing P, Phelps C. Networks, social influence, and the choice among competing innovations: Insights from open source software licenses. *Inf Syst Res*. 2009;24(3):539–60.
- Y. Wu, Y. Manabe, T. Kanda, D. M. German, and K Inoue. A method to detect license inconsistencies in large-scale open source projects. In *The 12th Working Conference on Mining Software Repositories MSR 2015, Florence, Italy, May 16-17, 2015*. IEEE; 2015.
- Howison J, Conklin M, Crowston K. FLOSSmole: A collaborative repository for FLOSS research data and analyses. *Int J Inform Technol Web Engr*. 2006; 1(3):17–26.
- Mardia, K. et al. *Multivariate Analysis (Probability and Mathematical Stats)*. Academic Press; 1980.
- Hannan M, Freeman J. Structural Inertia and Organizational Change. *Am Sociol Rev*. 1984;49(2):149–64. Retrieved from http://www.jstor.org/stable/2095567?seq=1#page_scan_tab_contents.
- Watson RT, et al. The business of open source. *Commun ACM*. 2008; 51(4):41–6.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com