PROTECTING OR SHARING INTELLECTUAL PROPERTY: LICENSES ANALYSES IN OPEN SOURCE SOFTWARE

Andrea Pontiggia
Università degli Studi di Cassino
ORGLab -Dipartimento Impresa Ambiente e Management-
via S.Angelo
03043 Cassino - Italia
andrea.pontiggia@unicas.it

Francesco Bolici
Università degli Studi di Cassino
ORGLab -Dipartimento Impresa Ambiente e Management-
via S.Angelo
03043 Cassino - Italia
francesco.bolici@eco.unicas.it

Università LUISS Guido Carli
CeRSI
via O. Tommasini, 1
00162 Roma - Italia
fbolici@luiss.it

ABSTRACT

In this paper we characterize the intellectual property rights (IPR) framework in open source (OS) context as an innovative distribution mechanism. We explain its main characteristics, which consists to oblige innovators to disclose the source code and any further improvement if they redistributed or resell it. This "weaker" or "opener" intellectual property protection avoids most of the established IPR mechanisms and offers a framework to support innovation diffusion and adoption. We analyze the licenses characteristics and their adoption using data from SourceForge.net, one of the most important OS community, it collects information of more than 100,000 projects. We discuss the consequences of such IPR model in institutional and economical policy and incentives. These results are part of a larger research program on IPR and Organizational Forms launches two years ago.

INTRODUCTION

This paper focuses on the analysis of the intellectual property rights (IPR) providing an empirical investigation of some fundamental characteristics of the open source licenses. This study represents a further step in the research program on IPR theme started almost one year ago by the ORGLab researchers.

Open source software (OSS) is a software whose source code is distributed with the object code under terms of licenses which allow the users to use, copy, modify and distribute the source code. The licenses present some interesting differences in the clauses content. These clauses show different regulation and protection mechanisms of property rights.

The OSS development process is similar to the "user-driven innovation" analyzed in other sectors where the role and contribution of the users are crucial for developing technological innovation (e.g. Rosenberg 1982, von Hippel 1988). Differently from traditional industry, where software is developed in-house by firms and then sold out on the market (Lakhani and von Hippel, 2000; von Hippel and von Krogh 2003), in OSS projects thousands of skilled programmers and users cooperatively develop the software online in a decentralized, geographically distributed, dynamic, knowledge intensive, ICT mediated and asynchronous way (Raymond, 2001; Lerner and Tirole, 2001; Kogut and Metiu, 2001).

In our previous studies (Pontiggia and Bolici 2005, 2006) we have identified three different steps in the OS development process: innovation, coordination and diffusion. We have also pointed out that the different forms of code's distribution --thus the different licenses-- can be considered as a crucial point to reach a deep understanding of the OS phenomenon. Lessig's thoughts explain why the property issue is important for the OS community from an ethical and historical point of you: "*If the code of cyberspaces is owned [], it can be controlled; if it is not owned, control is much more difficult. The lack of ownership, the absence of property, the inability to direct how ideas will be used-in a word, the presence of commons- is key to limiting, or checking, certain forms of governmental control.*" (Lessig, 1999 p.7). Secondly, the different OS licenses have introduced an anti-traditional IPR model that, for the first time, links the innovation not to a defensive-restrictive mechanism, but to an open-sharing model.

The OS phenomenon has been considered by many researchers as indicative of the need for a radical rethinking of the traditional IP protection for software (Kogut and Methiu, 2001; Benkler, 2002; Bessen, 2002). This is commonly explained in two different way. On one side, it is emphasized that the success of the voluntary and distributed model of innovation puts into question the traditional "market failure approach" to innovation, according to which exclusive rights to newly created knowledge are necessary to ensure the return of the rents from innovative investments. On the other side, attention has been directed on the possibility that the current system of intellectual property protection may pose a threat to the survival and further development of OS software. This is the case in countries, most notably the United States, where intellectual property protection has been stretched to encompass the use of software patents in addition to copyright and trade secrecy. The issue of the appropriate form of legal protection of software programs is, to some extent, an unresolved issue.

This crisis raises expectations, concerns and uncertainty for intellectual creators. Intellectual property management-based businesses feel endangered. They try to compensate for the undermining of the property-based approach by asking for ever-increased protection. Old forms of businesses like music publishers or broadcasters, or recent dominant players in biotechnology or software, call for longer, stricter monopolies, to be embedded in the access devices themselves, to be completed by regulation outlawing circumvention, etc.

However, a growing number of firms began to "open" part of their code aiming to draw benefits from the potential of development of the free software community or to facilitate a large diffusion of a core-product imposing it as a "free" standard and looking for benefits coming form sale of proprietary and complementary products. Firms introduce a totally different approach of intellectual property within their industrial strategies: a large number of "hybrid" licenses have been designed in order to control the degree of their openness. So, OS approach does not represent a denial of IPR but a new way to manage intellectual property. Through the OS licenses, intellectual property is not rejected, authors do not renounce to their rights but to the sole monopoly rent, such rights would authorize in a copyright regime.

Our research activity has focused on different contractual arrangements --license models-- that are adopted by open source software producers to distribute products and services. We analyzed the content of the 58 different licenses listed in the Open Source Initiative web site (July, 2005). After that, we have compared and classified them in order to have a deeper understanding of the similarities and differences among the licenses typologies. Then we analyzed their distribution among the Source Forge.net projects (an on-line public environment where more than 100.000 OS projects are stored). We collected data in a three years period (2002-2005) identifying trends and characteristics of the OS projects. Some preliminary data seem to confirm the emerging need for reducing the number and the tendency to simplify the contents of the licenses.

The main contribution of the article is focusing the analysis on the issue of defining the factors and the context which promote the use of OS model and of investigating the reasons for which different projects have different degree of licenses "openness". Providing preliminary answers to those questions could also improve the chance to understand under which conditions the innovative OS-IPR model could spread its advantage to other sectors different from the software. Looking at the underlying dynamics of software markets, where producers of software platforms and major applications (such as operating systems or internet communication servers) are able to choose different degrees of disclosure we can gather crucial information about the potential diffusion to other sectors of innovative IPR models. A further step in the research activity could be to analyse how different degrees of source-code disclosure affect the performance of (software) products and technologies and the profitability of their producers.

INTELLECTUAL PROPERTY RIGHTS: INNOVATION MODELS AND ECONOMIC PERSPECTIVES

The extension of legal protection to intellectual property in the software sector has been justified by reference to both moral and economic arguments, but in countries with a British legal heritage the latter have been the most influential, especially in relation to patents. Intellectual property protection is regarded by policymakers at both national and international levels primarily as a means of stimulating technological innovation. Thus, "[t]hough software intellectual property could not satisfactorily fall into any existing legal framework, all

countries have taken the decision to range it under the category of copyright. Then the double objective of intellectual property protection is not satisfied, which consists, on the one hand, to grant to the inventor a provisional monopoly for exploiting his invention and, on the other hand, to oblige him to disclose the principles of his invention" (Jullien and Zimmerman, 2005, p.2).

The open source software, with its principles of disclosing the source code of the program and the condition to maintain its "open" characteristic, provides an alternative model of intellectual property protection. The literature dealing with economic justifications for intellectual property protection is massive. Nelson and Mazzoleni (1998) identify four economic theories claiming to explain how patent protection promotes technological innovation:

- Invention-inducement theory: The anticipation of receiving patents provides motivation for useful invention.
- Disclosure theory: Patents facilitate wide knowledge about and use of inventions by inducing inventors to disclose their inventions when otherwise they would rely on secrecy.
- Development and commercialisation theory: Patents induce the investment needed to develop and commercialise inventions.
- Prospect development theory: patents enable the orderly exploration of broad prospects for derivative inventions.

These four theories seem to respond to the broad concept of the "tragedy of the commons" (Hardin, 1968). The tragedy of the commons suggests that not dividing the common into properties may lead to overuse and destruction of goods or innovations. Thus an intellectual property protection mechanism should be created in order to manage the common resources in an effective way.

In contrast, Mandeville (1996) observes that although conventional economic theories of the patent system do not give clear policy guidance, they seem to suggest that a strong patent system is desirable because strong property rights enable firms to control their technology and appropriate returns from it, thus providing incentives for the allocation of resources to innovative activity. He proposes an alternative, "information" perspective on innovation which points to a more complex but basically contrary view of appropriate patent scope in which

the degree of codification of technological information affects the ease, speed and mode of its diffusion, transfer or imitation. Moreover, Mandeville argues that supporting property rights on technological information may be not just unnecessary, but counter-productive to overall technological improvement. This perspective is based on Mandeville's assumption of innovation as cumulative and collective. Innovation is cumulative in the sense that the existing stock of technology is a crucial input in the production of new technology, and collective since it relies on the interaction of many participants. Because innovation is cumulative, it depends on information flow between present and future innovators. Because of collective process affecting the diffusion and innovation process, this latter depends on information flow among current participants, including rivals. From this viewpoint, even unauthorised copying among competing firms is beneficial to overall technological innovation because it is part of a process of transfer and learning.

In Mendeville's analysis, we can assume that easier is the communication and knowledge flow between different actors involved in the innovation process, higher are the probability to find a new effective solution. But, where property rights on multiple components of a single technology are owned by a number of separate entities, the development and commercialisation of new products requires co-ordination among many different actors. Thus, increasing the number and complexity of negotiations, transaction cost of the innovation process increase. Heller and Eisenberg (1998) have described the situation where multiple owners each have a right to exclude others from using a scarce resource as a "tragedy of the anti-commons". If owners are unable to negotiate successfully for the bundling of rights so that someone has an effective privilege of use, the resource may be underused and the total potential value of the rights (private and social) may not be realised.

We recognize that a growing number of enterprises[1] implement a strategy of "openness" freeing part of their software product and aiming to draw benefits from the potential of development of the OS community and to sustain the diffusion of a specific product as a "de facto standard" in order to take profit from

[1] Also firms as IBM and Windows have starter to open part of their products (e.g. Linux on the IBM server) in order to have benefits from their participation in the OS community.

complementary services and products. These companies have introduced a different approach to intellectual property rights management. A deeper understanding of this new way to manage intellectual property rights can be obtained also through the analysis of the OS phenomenon.

ANALYSIS

The research activity focus on different contractual arrangements --license models-- that are adopted by open source software producers to distribute products and services. The analysis is based on SoureForge.net community where more than 100.000 OS software projects are stored and classified. We considered the 58 different licenses listed in the Open Source Initiative web site (July, 2005) and we have evaluated their distribution among the Source Forge.net projects. The analysis considers a four years period (2002-2006) during which the data have been collected.

SourceForge is a highly available dataset, but it provides only a limited number of easily available variables, that is, variables that are pre-calculated and available from each project's homepage or in full lists (examples include: number of developers, project status, activity, downloads, page views and numbers of tracker items). Crowston (2004) presents a description of possible problems in collecting and analyzing data from SourceForge.net community, identifying spidering, parsing and summarizing as the most problematic research activities with this dataset. We have tried to avoid these problems having a direct access to the SourceForge database since Sept. 2005. However, we recognize that some of the limitations explicitated by Crowston (2004) have influenced also our analysis, in particular: i) there is a significant amount of 'dirty' data and it is hard to be sure of the extent of these problems without time consuming and costly manual checking; ii) there is a large amount of anonymous data in the SourceForge system that can not be attributed to any individual participant; iii) SourceForge has become the 'repository of record' for the FLOSS community, yet for important projects it is not the 'repository of use'; iiii) the projects of SourceForge.net dataset are highly differentiated as far as sizes and structures are concerning.

Preliminary Result: Licenses Distribution

Licenses, as mechanisms for distributing innovation, represent a key factor for understanding the intellectual property protection in the OS context. This analysis is focused on the 58 different types of licenses recognized till 2005 by OSI -Open Source Initiative-. The Open Source Definition and the OSI Certification Mark are the formal and official mechanisms that collect the characteristic of the different licenses, explicitating the conditions to be respected in order to have an OS license.

Exploring and analyzing the projects stored in the SourceForge.net community until 2005 (see Fig.1) we learn that most of the projects registered are adopting a General Public License (70%). The LGPL (Library/Lesser General Public License) and the BSD (Berkeley Software Distribution) follow the GPL with a percentage of adoption extremely lower (11% and 7%). The first five types of licenses (GPL, LGPL, BSD, Artistic and MIT) represent the 93% of all the projects stored at SourceForge.net. Thus the other 53 types of licenses recognized by OSI are used only by the 7% of the projects.

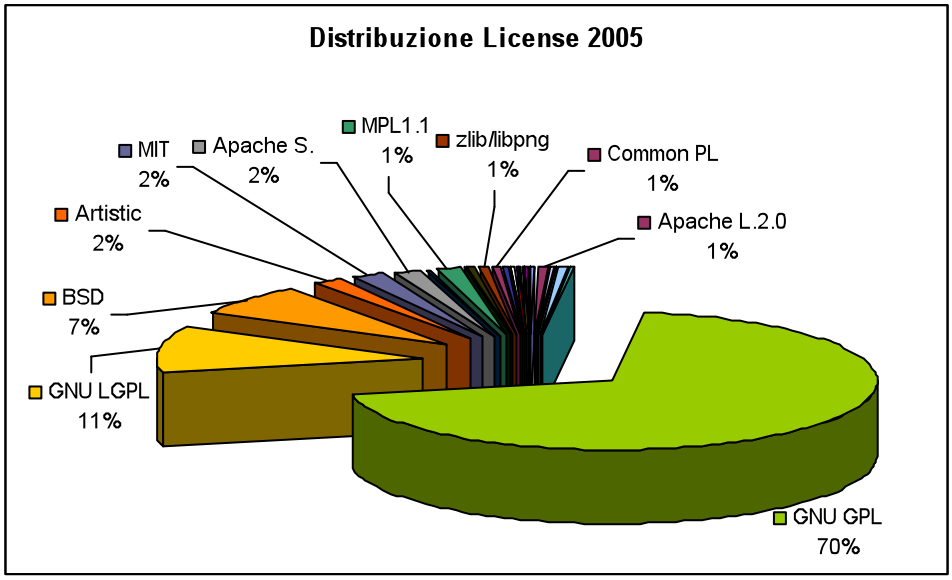
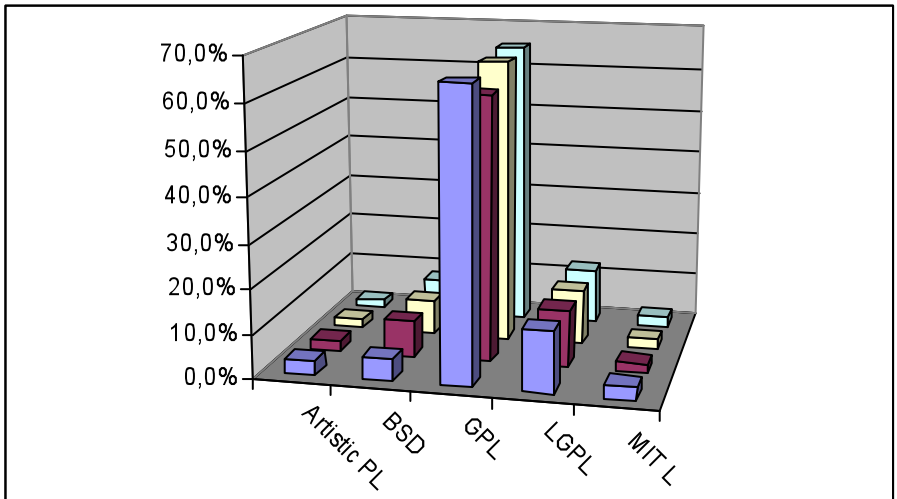


Figura 1 *Licenses Distribution in SourceForge.net, July 2005.*

Considering a six years time scale for our analysis, we can observe the evolution in the distribution of licenses inside the SourceForge community. As we can see in Fig.2, the sum of the first five types of licenses is always higher than the 85% of the total projects with file released, and the GPL with the LGPL never represent

less than 73% of the total amount. Moreover, in the last six years, only 12 types of licenses have covered at least 1% of the total amount of projects (9 for the last year). Thus we can identify a very high concentration in the adoption of different kinds of licenses. It is particularly important to recognize that also in 2002, when the variety of licenses extremely increased, no significant change regard the concentration of the top licenses.



	Artistic PL	BSD	GPL	LGPL	MIT L
Concentrazione 2000	3,2%	5,0%	65,2%	13,8%	3,0%
Concentrazione 2002	2,5%	8,2%	60,0%	13,0%	2,2%
Concentrazione 2004	2,1%	7,9%	64,5%	12,4%	2,2%
Concentrazione 2006	1,8%	7,7%	65,3%	12,4%	2,1%

Figura 2 *Top five licenses in the last 6 years (considering the percentage of projects with file released under a specific license).*

Instead some licenses (e.g. Pyton L., Artisitic PL., PHP L.) show an overall trend of reduction in their adoption. This consideration seems to confirm again the tendency in narrowing the adoption of a restrict and very specific kind of licenses for the projects in the OS community.

Considering the relationship between types of licenses and projects' main topic, we can also recognize that GPL maintains the leadership as license adopted for every topic (in a range between 47% and 75%) in 2006. The GPL is adopted by less than the half of the project in a single topic only in two case: the "formats and

protocols" (49,3%) and the "software development" (47,0%), while for the other topics GPL represents at least the 60% of the total amount. However it is interesting to underline that those two topics are exactly the two category in which the LGPL presents its maximum relative value (19,3% for formats and protocols and 20,8% for software development).

	GPL+LGPL	GPL+LGPL+BSD
Communications	78,2%	85,2%
Database	75,2%	82,6%
Desktop Environment	87,0%	87,0%
Education	77,4%	82,8%
Formats and Protocols	68,6%	78,6%
Games/Entertainment	82,0%	87,2%
Internet	76,1%	83,5%
Multimedia	80,5%	86,9%
Office/Business	76,9%	83,3%
Other/Nonlisted Topic	78,8%	86,4%
Printing	76,5%	85,4%
Religion and Philosophy	78,3%	84,1%
Scientific/Engineering	77,0%	84,1%
Security	75,4%	87,0%
Sociology	74,4%	81,7%
Software Development	67,8%	77,4%
System	77,8%	87,1%
Terminals	79,6%	86,0%
Text Editors	73,7%	80,4%

Tabella 1 *Percentage of projects adopting GPL, LGPL, BSD for each topic, 2006.*

From Fig.3 we can also understand that "communications", "internet", "software development" and "system" are the topic in which are focused most of the active projects in SourceForge.net.

By those considerations, we can conclude that GPL, LGPL and BSD present some difference in their adoption according to each specific topic but they are always the most adopted licenses, with more than 80% of the projects for almost all the topics.

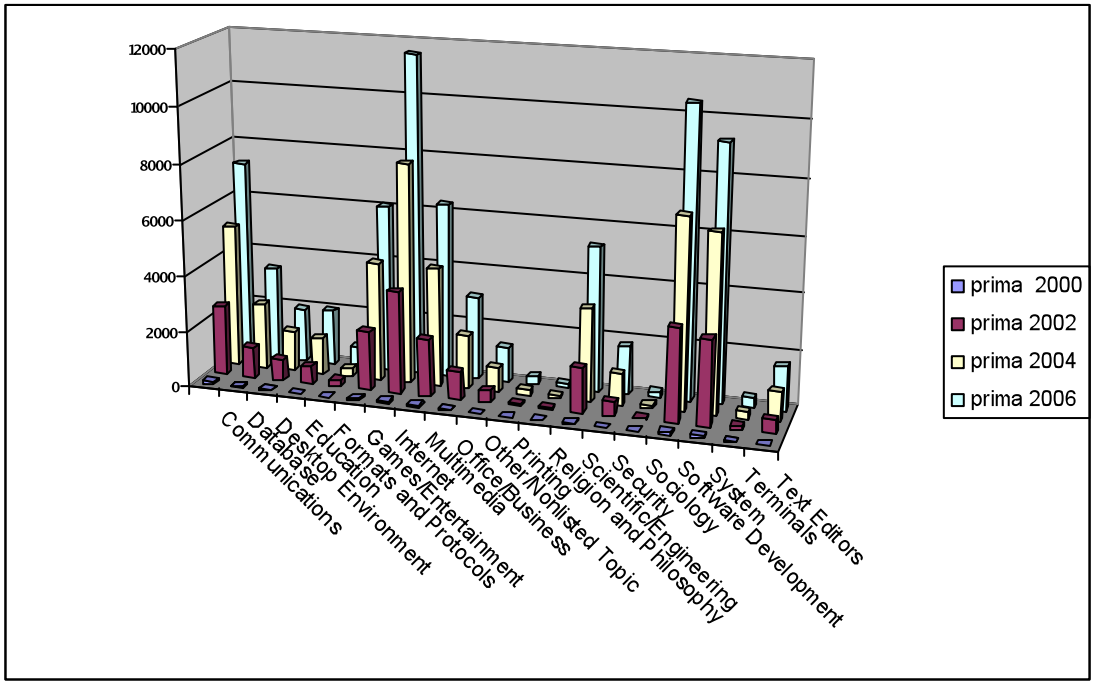


Figura 3 *Number of projects for each topic for four different periods.*

Excluding the main three licenses (GPL, LGPL, BSD), only three licenses have been adopted by more than 3% of the projects at least for one topic: Artistic Free License (3,4% of "education"), Apache 2.0 (4,2% of "formats and protocols", 3,6% of "software development") and MIT License (3,1% of "software development" and 3,4% of "terminals").

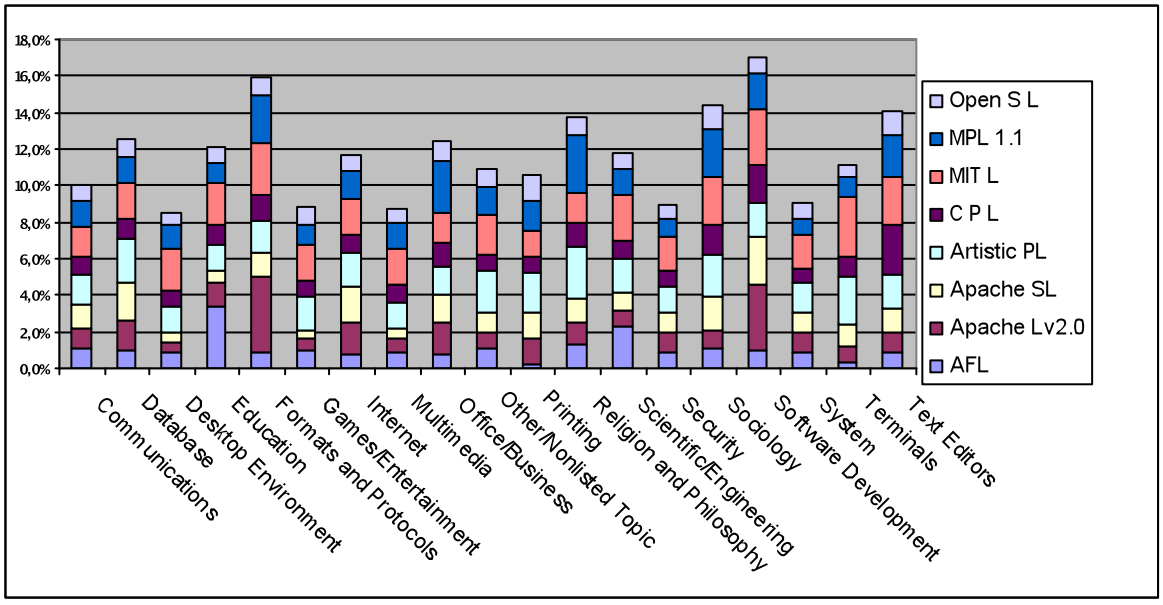


Figura 4 *License distribution for each topic (excluding GPL, LGPL, BSD), considering only the licenses that represent a minimum of 1% of the projects for at least a single topic, 2006.*

Analyzing the different characteristics of the OS licenses we have also classified them along a scale of "openness". We have assigned a high value of "protection" to those licenses that better represent the copy-left paradigm, guaranteeing a property right mechanism that preserve the common use of code (e.g. GPL has the higher value of this index). Considering the licenses recognized by OSI between 2002 and 2005 we have recognized that most part of licenses are characterized by a low or medium-law "protection" degree, instead only few licenses present a medium-high or high value of protection.

Licenze 2002 e 2005 - Indice di "protezione"

		Indice di "protezione"				
		Basso	Medio-basso	Alto-medio	Alto	Totale
Licenze 2002	N.	2	19	6	2	29
	Percentuale riga	6,9%	65,5%	20,7%	6,9%	100,0%
	Percentuale colonna	22,2%	61,3%	60,0%	50,0%	53,7%
Licenze 2005 (nuove)	N.	7	12	4	2	25
	Percentuale riga	28,0%	48,0%	16,0%	8,0%	100,0%
	Percentuale colonna	77,8%	38,7%	40,0%	50,0%	46,3%
Total	N.	9	31	10	4	54
	Percentuale riga	16,7%	57,4%	18,5%	7,4%	100,0%
	Percentuale colonna	100,0%	100,0%	100,0%	100,0%	100,0%

Figura 5 *License distribution for each topic (excluding GPL, LGPL, BSD), considering only the licenses.*

Moreover we have noticed that almost 75% of licenses' typologies allow to combine OS code within proprietary software, permitting to distribute the program though the traditional commercial model. However, more than 30% of the licenses introduced after 2002 provide different limitations in the use of OS code mixed with proprietary software. The most known license that avoids the combination of OS code with proprietary software is the GPL with its virus clause. But also other licenses with their characteristics cannot be easily distributed through the traditional business model, like the Educational License that obliges to distribute the software without any kind of cost or profit.

CONCLUSION

This paper aims at dispelling the issue of intellectual property rights in open source projects. There are three main contributions that can be drawn from the analysis presented above. First, from a policy point of view, we have underlined

how a good balance between individual incentives to innovate and the maximization of the social utility generated by the facilitated diffusion and use of the innovation is influenced by the socio-technical environment and its evolution over the time.

Second, open source movement shows that, for some knowledge intensive innovation, the traditional intellectual protection framework could be counter-productive. Avoiding a monopoly protection during the diffusion phase (to finance the initial phase of the innovation process), the OS projects are supported by the feedbacks and the needs produced by new users and community members. This consideration is also confirmed by the predominant presence of the most restrictive licenses (GPL) in the SoruceForge.net community.

Third, we have pointed out that most part of the projects, independently by the topic of the project, have been released under a GPL, LGPL or BSD license. All the other licenses cover a very small amount of projects. The licenses developed for a specific purpose never diffuse over a larger number of projects. This is coherent with the intuition of different author hypothesizing that in the long period only few different kind of licenses will survive and spread. The other licenses will disappear or will be limited to very specific area, topic or software functionalities.

In the future, we intend to understand if and how the open source model can exploit in a broader range of activities and industry beyond the sole software industry. The key issue to this development actually lies in the possibility to guarantee the long term sustainability of an open source model applied to a different context. New business models based on revised Intellectual Property Right regime (or system) clearly emerge and receive a great deal of attention of corporate strategy and policy makers.

REFERENCES

Bies R., T. Tripp e M. Neale (1993). Procedural Fairness and Profit Seeking: The Perceived Legitimacy of Market Exploitation. *Journal of Behavioral Decision Making*, 6, 243-56.

Bonaccorsi, A. e C. Rossi (2002). Why Open source Software can Succeed?, *LEM working paper series*, Sant'Anna Pisa.

Cusumano M.A., 2004, "Reflections on Free and Open Software", Communication of the ACM, Ottobre, vol.47, n.10.

Falk A., E. Fehr e U. Fischbacherm 1999, "On the Nature of Fair Behavior", working Paper n.17, University of Zurich.

Fehr W. e K. Schmidt, 1999, "A Theory of Fairness, competition and cooperation", Quaterly Journal of Economics, pp.817-868.

Frey B., 1997, "A Constitution for knavers crowds out civic virtues", Economic Journal, pp.1043-1053.

Fulk J., Heino R., Flanagin A.J. e P.R. Monge, (2004) "A Test of the Individual Action Model for Organizational Information Commons", Organization Science, Vol.15, n.5, settembre – ottobre.

Gatcher S. e E. Fehr, (1999), "Collective Action as Social Exchange", Journal of Economic Behavior and Organization, 39, pp.341-369.

Giuri, P. e S. Torrisi (2004) Economia e Management dell'Open source Software. alcune Note di Ricerca. LEM working paper series.

Giuri, P., G. Rocchetti e S. Torrisi (2002) Open source Software: From Open Science to New Marketing Models. LEM working paper series, n.23.

Hardin, G., 1968, "The tragedy of the commons", Science, 162, pp. 1243–1248.

Heller, M. and R. Eisenberg (1998). "Can Patents Deter Innovation? The Anticommons in Biomedical Research." Science 280, 698-701.

Jullien, N. and J.B. Zimmermann (2005). New Approaches to intellectual property: from open software to knowledge based industrial activities. *Cahier de Recherche,* n. 8.

Kahneman D., J.L. Knetsch e R.H. Thaler, 1986, "Fairness as a Constraint on Profit Seeking: Entitlement in the Market", American Economic Review, 76, 728—41.

Kogut B. e A. Metiu, 2001, "Open-Source Software Development and Distributed Innovation", Oxford Review of Economic Policy, vol.17, n.2, pp.248-264.

Lakhani, K.R. e E. von Hippel, 2003, "How Open source Software Works: Free User-to-user Assistance", Research Policy, n.32, pp.923-943.

Lee Gwendolyn e R.E., Cole, 2003, "From a Firm-Based to a Community-Based model of Knowledge Creation: The Case of Linux Kernel Development", Organization Science, Vol.14m n.6.

Lerner, J. e J. Tirole (2001) Some Simple Economicsof Open source. Journal of Industrial Economics, n. 50, pp.197-234.

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

Mandeville, T. (1996). *Understanding Novelty: Information, Technological Change, and the Patent System*. Ablex, Norwood, New Jersey.

Mazzoleni, R. e R.R. Nelson (1998). Economic Theories about the Benefits and the Costs of the Patents, Journal of Economics Issues, n.4, pp.1031-1052.

Muffatto, M. e M. Caldani, 2004, Open source. Strategie, organizzazione, prospettive, Il Mulino, Bologna.

Ostrom, E., 1990, Governing the Commons: The Evolution of Institutions for Collective Action, Cambridge University Press, Cambridge, MA.

Pontiggia, A. e F. Bolici (2005). Open Source e Licenze: il codice come innovazione, coordinamento e diffusione, negli *atti del II Workshop ITAIS*, Verona, 1-2 Dicembre.

Pontiggia, A. e F. Bolici (2006). Codice, Open Source e Diritti di Proprietà: Innovazione, Coordinamento e Diffusione, negli *atti del VII Workshop dei Docenti e Ricercatori di Organizzazione Aziendale*, Salerno, 2-3 Febbraio.

Raymond, E.S., (1999), The Cathedral and the Bazaar: Musings on Linux and Open source by an Accidental revolutionary, O'Reilly & Associates, Sebastopol, CA.

Raymond, E.S., 1998, The Cathedral and the Bazaar. First Monday, 3, 3.

Rosenberg, N. (1982), Inside the Black Box: Technology and Economics. Cambridge University Press, New York.

Titmuss, R. M. (1971), The Gift Relationship: From Human Blood to Social Policy, New York, Pantheon.

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