# Contributing to OS Projects. A Comparison between Individual and Firms

Andrea Bonaccorsi
*University of Pisa*
*bonaccorsi@sssup.it*

Cristina Rossi
*Sant'Anna School, IIT-CNR*
*cristina.rossi@iit.cnr.it*

## Abstract

*This paper studies the contributions software firms make to Open Source (OS) projects. Our goal is to ascertain whether they follow the same regularity of pattern seen for individual programmer An exhaustive empirical analysis was carried out using data on project membership[1] , project coordination and the contributions made by 146 Italian firms that do business with OS software. We compare our findings with the results of the surveys taken on OS programmers. The availability of the data gathered by Hertel et al. ([10]) on 141 developers of the Linux kernel allowed a direct comparison to be carried out between the two sets[2] .*

## 1. Introduction

The decentralized structure of the Open Source production mode makes it impossible to ascertain how many developers contribute to the movement. Agents can join the community of developers of an Open Source project freely, admission to the group only being subject to making a demonstrable contribution of some valuable code ([1]). At the same time, they can resign from the project whenever they want and for whatever reason they wish ([2]).

Open Source projects have a number of features in common with the pattern of contribution followed by individual programmers. Several surveys ([3], [4]) identified empirical regularities in the distribution of the main metrics in project contributions, i.e. project size (number of developers per project), project membership (number of projects each developer joined) and the contribution effort of each participant in terms of man/hours.Three main factors emerge which point to a highly skewed distribution of these variables. First of all, most of the development activity is concentrated on a small number of projects. Secondly, few agents take part in a large number of projects. Finally, the distribution of individual contributions within projects is spectacularly skewed. Most of the programming effort is carried out by very few programmers. This challenges the commonly held image of the Open Source community as a relatively flat network of interacting peers. Several researchers have examined this issue referring to the theory of power law distributions ([5][3] . In particular Zipf distributions were identified in project size, membership and contribution, as well as downloads, mailing list posts and CVS commits.

When a project grows in popularity it becomes more and more attractive for developers to join it. The opposite happens with unpopular projects that do not attract a large base of programmers. Individuals are more likely to expect to enhance their reputation ([7]) if they write valuable code seen by a large community, including large software firms. As a result, the dynamics of participation may share the properties of winner-take-all processes. Virtually all this evidence has been collected by looking at individual programmers. However, Open Source Software (OSS) is increasingly developed by firms that base their business on the legal and social rules of the community. They supply products under OS licenses and develop software according to the OS disclosure model. Little is known about the contribution of firms to OS projects. This paper is a contribution to fill this gap.

Following the approach used in previous surveys on Open Source developers, we prepared an on-line questionnaire; a Website was set up containing all the information on our study and the link to the data-gathering system. The relative novelty of the OSS phenomenon means that there is not yet any complete directory or list of firms involved in it and new firms are getting involved in it each year. So we approached an initial short-list of firms, asked for their help in referring to other firms active in Open Source and stopped when no new referral was generated (snowball procedure). Therefore our sample is not statistically representative of the Open Source firms' universe. However, given the

[1] *Project membership* refers to the number of projects a given firm is now joining or has joined in the past.

[2] We thank Professor Hertel for making his data available and Alessandro Scateni who built up the database

[3] Empirical regularities of this sort shape a wide range of natural and social phenomena [6].

exploratory nature of this study, this approach was considered methodologically appropriate[4] . We managed to contact 275 firms, obtaining 146 valid answers, which represents a good cross-section of the Italian firms operating in the supply-side of the Open Source market.

These firms produce and/or sell software under one or more Open Source licenses recognized by the Open Source definition. Most respondents offer services such as installation (80.1%), support (82.9%), maintenance (76%), consultancy (84.9%) and training (64.4%). 64.5% of firms surveyed began adopting the Open Source technology after 1999, while 35.5% used it before that date. Their average size is 17.3 employees (median value: 8.0), of which about 70% are software developers. Most firms are very young, 6.9% were born after 2002, 47.9% between 1998 and 2001. In general, their founders are entrepreneurs with technical background (28.8%) or with mixed technical and commercial background (65.1%). Their size, origin and business models makes them representative of a new wave of software firms that entered the market after the wording of the OS definition. Interestingly, they grow very rapidly (turnover rate of growth +121.3% over the 1999-2002 period) and have positive expectations about the potential market for Open Source solutions. On the average OSS represents 46.5% of the turnover. Although they share one of the typical features of Italian firms across all industries (namely the attitude to stay small), we do not believe their behaviour is heavily influenced by idiosyncratic national factors.

## 2. The activity of firms in OS projects

Table 1 shows various measures of firms' involvement in developing OS projects. Firms were asked how many projects they joined by registering on line and how many they coordinate by assuming the role of initiator or manager, in the last twelve months (year 2002) and since their initial involvement in the OS community. Further, they were asked to estimate how many Lines of Code (LOCs) they contribute as a percentage of the total size of projects, and how many patches and modules of their were accepted as official contributions. The data show that the activity of firms is quite limited. On average, they took part in fewer than 4 projects since taking their first steps in the Open Source community while the median value of projects participated in is 1. Similarly to the results of the surveys made on individual developers, all the variables observed display highly skewed distributions (table 1).

Almost half of the sample (49.6%) is not currently engaged in Open Source projects.

[4] Details on the data gathering procedures may be requested from the authors.

| *Variable* | *Acronym* | *Min* | *Max* | *Mean* | *Std. Dev.* | *Median* | *Skewness* |
|---|---|---|---|---|---|---|---|
| No. of projects the firms joined since the very start of their OS activity | *ALL_A_PM* | 0 | 50 | 3.8 | 7.8 | 1 | 3.5 |
| No. of projects the firms joined last year | *C_PM* | 0 | 20 | 1.6 | 2,8 | 1 | 3.7 |
| No. of projects the firms coordinated since the very start of their OS activity | *ALL_A_CP* | 0 | 28 | 1.1 | 3.4 | 0 | 5.9 |
| No. of projects the firms coordinated during 2002 | *C_CP* | 0 | 7 | 0.5 | 1.2 | 0 | 3.5 |
| % of Line of Codes (LOCs) the firms contributed to each project on average | *%_LOCs* | 0 | 99 | 10.56 | 23.5 | 0 | 2.5 |
| Contributions by the firms incorporated in the official versions of the projects | *N_C_OV* | 0 | 300 | 6.9 | 36.9 | 0 | 6.7 |

**Table 1. Descriptive statistics.**

About 68% of the firms have participated in no more than 2 projects and many of them (46.2%) have never joined one; only 7.7% have been involved in the development of more than 10 programs (figure 1).

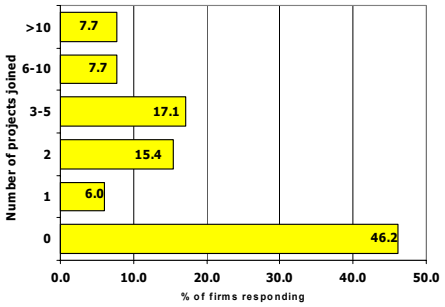


**Figure 1. No. of projects the firms joined since the very start of their OSS activity**

The large majority of respondents (72.9%) have never carried out coordination tasks (figure 2).

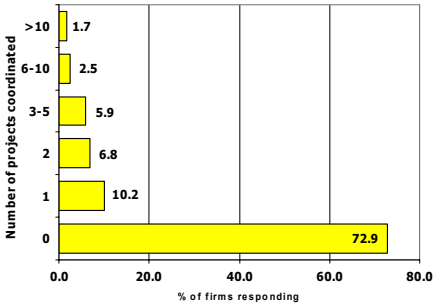


**Figure 2. No. of projects the firms coordinated since the very start of their OSS activity**

Only 26 firms (21.5%) are currently coordinating a project, with the majority of these (53.8%) just one. The limited involvement of respondents in Open Source projects is seen also in projects' contributions. If leading authors are classified as firms who have supplied more than 50% of a project's overall LOCs, then only 7.6% of them can be so classified (figure 3).

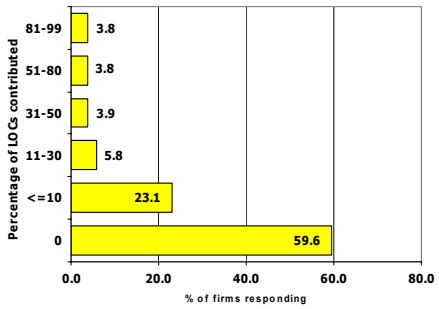


**Figure 3. %_LOCs: % of LOCs the firms contributed to each project on average**

Few have written pieces of code that have been accepted for the official version of a project. The mean value of variable N_C_OV is 7 but it is strongly influenced by three outliers who provided hundreds of accepted contributions[5] . This apart, the mean value becomes 1 (table 1).

In short it is possible to conclude that firms' development activities are more likely to be limited to merely adapting Open Source programs to suit their customers' requirements, giving less importance to circulating these ad hoc solutions within the community. Moreover like individual developers, firms probably prefer to join large, successful projects with a leadership already consolidated. These projects provide widely available software products, released in official and stable versions, which enable firms to provide their users with better services. Contributing to successful projects improves corporate image, giving a favourable impression to clients and potential venture capitalists. Finally, since most firms entered the OS arena just few years ago, they have not yet gained enough reputation among Open Source developers to be appointed to the coordination of a project.

We also collected data on why firms have decided to adopt a business model based on Open Source software. Each incentive is measured on a Likert scale ranging from 1 (not at all important) to 5 (very important)[6].

The number of contributions made by firms and accepted for official versions is positively correlated with the learning incentive and all the social motivations. The Open Source production mode places a huge volume of source code at the disposal of whoever wants to modify, debug or just study it. Firms that exploit this immense learning opportunity have greater chances of improving their programming skills and consequently are more likely to write pieces of code that the community will consider valuable enough to be accepted into the official versions of the programs. Moreover, interaction within projects is enhanced by positive social motivations. Firms that gift code and conform to the Open Source values have stronger linkages with the community than those who are there only to exploit new business opportunities. The former devote much more effort to OS activities and this has positive impact on the number of contributions accepted. The negative correlation between the economic motivation and the number of projects that firms are currently coordinating is also interesting. It seems that some firms exploit the innovative opportunities given by OSS without actively contributing to it. Insofar as their licensing behaviour conforms to the rules of the community, these firms seem to be accepted as legitimate members.

In order to examine closely the link between the motivations behind and level of OS activity, firms have been divided into two groups on the basis of the scores they assigned to the motivational variables, i.e. low score (1 or 2) or high score (4 or 5). In this way we may correct for a central tendency bias in reply to Likert-type scales. Table 2 summarizes the results. Only mean values displaying statistically significant differences between the two groups (Mann-Whitney test) are reported.

The various metrics of the level of activity show different patterns in the two groups, depending on the incentive type. Firms that attach much importance to purely economic considerations, have less experience in coordination activities. They are acting mainly out of extrinsic motivations such as promoting innovation or gaining freedom from the licensing and price policies of the large software companies and want to keep the

[5] Respectively 300, 200 and 100. The agent ranking fourth has only 12 accepted contributions.

[6] The taxonomy of motivations is as follows [Feller]: Economic motivations: because Open Source software allows small enterprises to afford innovation; because we want to be independent from the price and licensing policies of large software companies; because in the field of Open Source good IT specialists are easy to find; because opening our source code allows us to gain a reputation among our customers and competitors. Social motivations: because we agree with the values of the Open Source movement; because we want to place our source code and skills at the disposal of the Open Source community and hope that others will do the same; because we think that software should not to be a proprietary commodity. Technological motivations: because contributions and feedback from the Open Source community are very useful to fix bugs and improve our software; because of the reliability and quality of the Open Source software; because we want to study the code written by other programmers and use it for developing new programs and product; to obtain products not available on the proprietary software market.

cooperation link with Open Source developers active in order to obtain the feedback and contributions that allow them to lower developing costs.

| *Motivations* | *Metrics* | *N* | *Mean* | *Dev. Std.* | *N* | *Mean* | *Dev. Std.* |
|---|---|---|---|---|---|---|---|
| | | *LOW SCORE* | | | *HIGH SCORE* | | |
| We want to be independent of the price and licensing policies of the large software companies | *C_CP** | 20 | 0,7 | 1,1 | 76 | 0,4 | 1,4 |
| Open Source software allows small enterprises to afford innovation | *ALL_A_CP* | 13 | 2,0 | 2,4 | 85 | 0,9 | 3,3 |
| We want to place our code and skills at the disposal of the OS community and hope that others do the same | *ALL_A_PM** | *28* | *3,4* | *8,0* | *61* | *4,9* | *9,2* |
| | *N_C_OV+* | *24* | *0,6* | *2,1* | *51* | *1,5* | *3,0* |
| We think that software should not to be a proprietary commodity | *%LOC** | *38* | *8,4* | *23,0* | *38* | *14,2* | *26,3* |
| | *N_C_OV-* | *38* | *0,4* | *1,8* | *38* | *1,8* | *3,1* |
| Contributions and feedback from the OS community are very useful to fix bugs and improve our software | *%LOC+* | *14* | *6,6* | *24,0* | *70* | *10,1* | *22,8* |
| | *N_C_OV+* | *15* | *0,1* | *0,3* | *65* | *1,2* | *2,7* |
| We want to study the code written by other programmers and use it for developing new programs | *ALL_A_PM* | 28 | 4,1 | 10,9 | 59 | 4,5 | 7,9 |

**Table 2. Motivations: Mann Whitney tests. +: p value< 0.1; *p value<0.05; - p value< 0.01.**

However, these firms can win the trust of developers merely by gifting their code and avoiding hijacking that written by other programmers without devoting resources to coordination tasks. The percentage of LOCs contributed and the number of accepted patches are higher for firms that attach much importance to the feedback from the community. These results corroborate the findings of the correlation analysis about the role played by social motivations in shaping the level of activity of Open Source firms. Firms that assign high scores to social incentives are more likely to win the trust of the community. As a consequence more contributions by them are accepted into official releases. Firms that value the learning opportunities provided by the Open Source mode of production very highly clearly behave in the same way. At the same time, firms that want to fight for software freedom, make great efforts to contribute LOCs to Open Source projects. In this way they increase the code base released under Open Source license schemes.

## 3. Project activity of firms and individual developers. A comparison.

We compare our data with the findings of the surveys on individual developers ([8], [9], [10], [11], [12], [13]). The samples of some of these empirical analyses are larger than ours. However taking into account every single person working in the firms we surveyed, it can be estimated that data account for more than 1,500 developers[7].

All the studies show that on average the membership of firms in projects is lower than that of individual developers. This holds for both the current and the overall OS activity. In general firms devote less programming effort to the projects in which they take part. Case studies of single successful projects ([9], [10]) display similarities in programming effort as measured by the percentage of LOCs contributed on average by each developer[8] . A much larger value has been instead obtained in our study. However, such comparison is not correct from a methodological viewpoint. The strictly hierarchical organization of successful projects very often makes it difficult for developers to add LOCs to their official versions. Most programmers contribute no line and this significantly affects the mean value of this metric. In the case of the firms we examined, the percentage of LOCs does not deal with a single successful project but refers to project participation in general. Together with a large group of firms contributing no LOCs (59.6%) we found firms providing almost the whole code base of the projects in which they were involved. This is likely to happen for small projects that have been started by the firms themselves and have not been able to attract a large base of developers. The comparison between our surveys and the ones taken on individual developers ([8],[12],[13]), show that firms classified as leading authors contribute a lower percentage of LOCs with respect to similarly classified developers[9] . It could be thought that this is not due to the presence of fewer firms devoting large programming efforts but to the lower concentration of their contributions. However this is not the case, given that more than 88% of the agents contribute no more than 30% of the LOCs of the projects in which they are involved. Another metric, the number of contributions included in the projects' official releases, shows a poor performance by firms. Excluding the outlier values, it appears that firms manage to get fewer contributions into the official versions of projects than individual developers do.

Proceeding in our analysis we refer to the data collected by Hertel et al. ([10]) on 141 developers of the Linux kernel. Two of the metrics collected by the authors are comparable to ours, namely the percentage of LOCs contributed and the number of contributions incorporated into official releases. However, as we explained above, methodological reasons allow to use only the latter.

[7] Our sample includes total staff at around 2,388. The ratio between programmers and total staff is in the range 60-80%, according to a telephone survey on a sub-sample carried out during the follow-up.

[8] In these three studies the percentage of LOCs contributed on average by each developer are respectively 0.1, 0.33 and 0.26.

[9] Differences in the definitions of leading author need to be taken into account.

Table 3 shows the distribution of contributions in the two groups. In comparison to individual developers, there is a higher percentage of firms in the three lowest classes (92.1% vs. 86.4%). However no significant difference emerges in the percentage of agents that do not succeed in placing contributions. The two highest classes display an interesting pattern. About 5% of the firms have a number of contributions ranging from 7 to 10, while only 1.6% of the programmers performed likewise. Nevertheless, 12% of the developers had more than 10 contributions accepted versus 4% by the firms. This sheds light on the presence of an elite of individual programmers performing very well in Open Source developing activities. Firms do not seem to include of such a group. On average firms place fewer contributions into official versions than individual developers do (4.1 vs. 8.7, Mann Whitney test, p value = 0.005).

| | *Linux Kernel developers (Hertel et al. 2003)* | | *Firms (our survey, 2003)* | |
|---|---|---|---|---|
| *No. of contributions* | *%* | *Cum. %* | *%* | *Cum. %* |
| *0* | 73,6 | 73,6 | 72,3 | 72,3 |
| *1-3* | 8,8 | 82,4 | 14,9 | 87,2 |
| *4-6* | 4 | 86,4 | 4,9 | 92,1 |
| *7-10* | 1,6 | 88 | 4,0 | 96,0 |
| *>10* | 12 | 100 | 4,0 | 100,0 |

**Table 3. Distributions of the accepted contributions of individual developers and firms.**

## 3. Conclusions

These findings raise an interesting issue, namely that the level of contribution to Open Source projects seems to depend on social and technological motivations over purely economic ones.

There are firms that exploit the low cost, broad availability and good quality of Open Source software to build up a sustainable business model without contributing in the same proportion. More generally, these findings shed light on an interesting evolutionary property of the Open Source communities: robustness. The behaviour of contributing to a common pool resource does not have to be equally shared among contributors in order to be self-sustaining. Open Source communities permit some members to take more than they give, provided they do not violate minimal membership rules. By exploiting existing code more than they contribute, they still enlarge the bases of the Open Source users, indirectly enhancing the motivations of active producers. The literature on public good provision and free riding has probably overestimated the potentially damaging role of a few non-contributors, by assuming that their behaviour would inevitably self-propagate. This is not necessarily true. Nevertheless, because these firms are accepted in the Open Source community as legitimate partners, it is almost certain that, in general, they comply with the rules of membership and legal obligations for Open Source licensing schemes. That is to say, for example, that they do not hijack the code, but rather adapt and redistribute it under an appropriate license scheme. It is clear that these firms take more than they give. It seems that the new organisational mode of software production is strong enough to withstand a sharing of contributing behaviour that includes input effort no higher than the minimum entry level.

## References


[1] Bonaccorsi A. and Rossi C. (2003) "Why Open Source software can succeed," Research Policy, vol. 32, No. 7, 2003, pp. 1243-1258.

[2] S. O'Mahony, "Guarding the commons: how do community managed software projects protect their work", Research Policy, vol. 32, No. 7, 2003, pp. 1179-1198.

[3] K. Healy and Schssman A., "The ecology of Open Source software development", MIT Working paper.

[4] F. Hunt and P. Johnson "On the Pareto distribution of Sourceforge projects, Proc. The Open Source Software Development Workshop, Newcastle, UK, 2002, pp. 122-129.

[5] V. Freeh, G. Madey, and R. Tynan, "Understanding OSS as a self-organizing process", Proc. 2nd Workshop on Open Source Software Engineering, IEEE CS Press, Orlando, FL, May 2002

[6] P. Krugman The self organizing economy, Backwell, New York, N.Y., 1996.

[7] J. Lerner, J. Tirole, "Some simple economics of the Open Source", The Journal of Industrial Economics, vol. 2, pp. 197-234.

[8] R. A. Ghosh, R. Glott, B. Krieger, and G. Robles, "Survey of developers, "Free/Libre and Open Source Software: Survey and Study, FLOSS Final Report, International Institute of Infonomics,,

[9] S. Koch, G. Schneider, "Effort, co-operation and co-ordination in an Open Source software project: GNOME", Information System Journal, vol. 12, 2002, pp. 27-42.

[10] G. Hertel, S. Niedner, and S. Hermann, "Motivation of software developers in the Open Source projects: an Internet-based survey of contributors to the Linux kernel" Research Policy, vol. 32, No. 7, 2003, pp. 1159-1177.

[11] J. Feller and B. Fitzgerald, Understanding Open Source Software Development, Addison Wesley, Boston, M.A., 2002.

[12] G. Robles, H. Scheider, I. Tretkowski, and N. Weber, "Who is doing it? A research on Libre Software developers", Fachgebiet für Informatik und Gesellschaft TU-Berlin

[13] J. Bates, C. Di Bona, K. Lakhani, and B. Wolf, "The Boston Consulting Group hacker survey