

December 2002

An Exploratory Study of Factors Influencing the Level of Vitality and Popularity of Open Source Projects

Katherine Stewart
University of Maryland

Tony Ammeter
University of Mississippi

Follow this and additional works at: <http://aisel.aisnet.org/icis2002>

Recommended Citation

Stewart, Katherine and Ammeter, Tony, "An Exploratory Study of Factors Influencing the Level of Vitality and Popularity of Open Source Projects" (2002). *ICIS 2002 Proceedings*. 88.
<http://aisel.aisnet.org/icis2002/88>

This material is brought to you by the International Conference on Information Systems (ICIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ICIS 2002 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

AN EXPLORATORY STUDY OF FACTORS INFLUENCING THE LEVEL OF VITALITY AND POPULARITY OF OPEN SOURCE PROJECTS

Katherine J. Stewart
R. H. Smith School of Business
University of Maryland
College Park, MD USA
kstewart@rhsmith.umd.edu

Tony Ammeter
University of Mississippi
University, MS USA
tammer@bus.olemiss.edu

Abstract

In this research, we ask the question: What differentiates successful from unsuccessful open source software projects? Using a sample of 240 open source projects, we examine how organizational sponsorship, target audience (developer versus end user), license choice, and development status interact over time to influence the extent to which open source software projects attract user attention and developer activity.

1 RESEARCH OBJECTIVES AND QUESTIONS

There is growing interest in open source software (OSS) development in industry, government, and academia. OSS development is different from traditional software development in that it depends on volunteer labor pools coordinating their effort without a common organizational superstructure, and the end product is generally provided freely (Feller and Fitzgerald 2002). There are many ways to define success in the context of OSS development. Market penetration is often cited when discussing well-known open source software, such as Linux and Apache. A similar indicator for less well-known projects might be the number of subscribers to the project—i.e., the number of people who monitor announcements and new releases is one indicator of public interest. Another metric of interest is the number of people who contribute to a project, a figure that may be hard to determine, as evidenced by the widely varying numbers reported for Linux, ranging from about 1,200 to hundreds of thousands (Feller and Fitzgerald 2002). The output of those contributors (e.g., in terms of new software releases) is also a potential indicator of success.

As is evident from Feller and Fitzgerald's (2002) discussion, there is an accumulated body of common wisdom about what makes OSS development successful, but there is also a need for rigorous research and empirical examination of the phenomenon to confirm or refute this common wisdom. Our goal in this study is to provide an initial empirical investigation into factors that influence project success. Are some kinds of projects more successful than others?

We analyze projects listed on the www.freshmeat.net developer forum, and we operationalize success in terms of two indicators available on that forum. These are the *vitality* on the project, an indicator of how much developer effort and attention is expended on it, and the *popularity* of the project, an indicator of how much user attention is focused on it. Given that these metrics are tracked within the OSS community, we believe they are of practical relevance. Vitality is calculated using the number of announcements about a project and the time since its last release. Popularity is based on the number of people who subscribe to the project as well as hits to the Website.

We hope the study will contribute to the research literature by helping to develop an initial understanding of the context of open source development and factors that may distinguish it from more traditional development. Some have argued that the organizational processes employed in open source development are typical of virtual organizations and may be applicable in other domains (e.g., Gallivan 2001; Markus et al. 2000). Hence, increasing our understanding of what leads to success in OSS may help to increase our understanding of other kinds of virtual organizations.

2 STUDY DESCRIPTION

Since our question concerns why some open source projects are successful while others falter, we decided to track a sample of projects over time and analyze differences between those that exhibit increases in the success indicators versus those that do not. There are many choices made in the early stages of an open source project: what is the intended audience (i.e., what kind of application is being developed), what license will be used, at what stage should the project be opened to the community, should the project be affiliated with an organizational sponsor? We discuss each of these choices, and we look at how they influence the success of projects in terms of how much vitality and popularity the projects generate. We also examine the relationship between vitality and popularity, with the assumption that more active projects may become more popular over time. Similarly, popularity may encourage greater activity among developers and thereby enhance vitality.

Intended audience. Although the use of the term *open source* is relatively new, open source development is not a new phenomenon. Early programming efforts were often open source (see summaries in DiBona et al. 1999; Feller and Fitzgerald 2002). Historically, open source has been most successful in back-end types of applications such as operating systems (e.g., Linux), Web servers (e.g., Apache), and e-mail handling (e.g., sendmail). A second area of popularity has been in the development of games (e.g., Quake2). As Feller and Fitzgerald (2002) point out, these are areas of application useful to developers for their own work and entertainment.

More recently, the phenomenon of open source has spread into domains where the product may be targeted more toward a general end-user population; there are open source word processors, spreadsheets, and other tools (e.g., OpenOffice). It is unclear whether such projects will be as successful as those aimed at back-end uses. Projects targeted at developers may see higher levels of vitality and popularity because those audiences have the skills to contribute to the project (so it will have more activity) and they have the knowledge to assess quality (so, assuming quality is high, it will be more popular). Feller and Fitzgerald note that developers may be more likely than end users to choose an open source product based on their own quality assessment rather than marketing campaigns or FUD¹ tactics. Also, developers may have more discretion in choosing what products they employ than end users because they are able to install and run the products themselves, whereas end users may be less able to manage their own software.

License. A question of interest has been: Why do people contribute to open source projects? What motivates them to “give away” their efforts for free? Answers have focused on factors such as reputation and ego concerns, the need to obtain feedback to develop professional skills, and producing something of personal use (DiBona et al. 1999; Feller and Fitzgerald 2002). To maintain these motivations, licensing may be an important issue. The Open Source Initiative has created criteria to define what an open source license should contain.² The provisions in an OSI approved license tend to provide some guarantees to a contributor that he will be able to benefit from his own efforts (e.g., the code he contributes cannot later be closed), that others will not usurp his contributions, and that his name will travel with the project (if it were closed, names embedded within it would not be visible and therefore reputation gains would be less likely). For these reasons, the extent to which a license meets OSI standards may have an impact on the vitality and popularity of the product.

Development Status. Feller and Fitzgerald discuss the question of when a project should be opened versus closed. Following Raymond (2001), they suggest one needs to have some kernel of a working program to get people interested in contributing and to provide a structure into which those contributions can be incorporated. For this reason, projects that are posted before there is a working version, e.g., those posted in the planning or pre-alpha stage, may be less likely to garner activity and attention than those put up with at least a working alpha version. However, once a project has reached a mature, stable release, vitality may be expected to drop off because there are fewer “big” reputation-making contributions to be made and the “personal itch” that spurs some to contribute may be in large part satisfied. Based on the discussion of reputation-based motivations, it is intuitively logical that once the project is mature, with less glamorous work remaining to be done, we might expect vitality to drop off.

¹FUD—fear, uncertainty, doubt—is an abbreviation used in a Microsoft memorandum describing ways that users may be discouraged from adopting non-Microsoft products.

²See www.opensource.org for a detailed description of the licensing requirements.

The relationship between development status and popularity may continue to increase as a project matures because the higher the status, the more useful the product is to a wider range of people. Also, the longer a project has been in existence (a correlate of development status), the larger the base of users available to help spread its use through positive word-of-mouth.

Sponsorship. As open source development has entered the mainstream and garnered increasing attention in the media, there has been discussion over whether it is a viable alternative for businesses (e.g., Behlendorf 1999). Many open source projects are now associated with organizations. Affiliation with an organization may have both benefits and drawbacks in terms of project success. An organization may provide resources, such as paid development, thereby possibly enhancing vitality. Yet an association with a for-profit entity could dampen enthusiasm among some users and developers because certain tenets of the open source culture seem to value independence from organizational constraints and in some cases disdain of profit motives.

3 RESEARCH METHODOLOGY, CURRENT STATUS, AND FUTURE PLANS

We are currently tracking publicly available data on 240 open source projects registered on the freshmeat Website. As Feller and Fitzgerald (2002) point out, freshmeat is a gathering place for the “grassroots” OSS community, thus we expected that most of the projects we observe there are of a more “traditional” nature rather than those that are sponsored by large companies (e.g., Netscape) or that are already established and successful (e.g., Linux). We selected these projects using two sampling frames. First, we randomly selected a total of 120 projects from the utilities, software development, and games and entertainment areas. We then selected 120 projects from these forums that had been registered on the site during the two weeks prior to the start of our data collection effort. We selected this second set of projects so that we would be able to conduct analyses across a group of projects that had a wide variance in terms of development stage and team member experience, and so we can track some projects from their initial appearance on the site. Once each week, we record the vitality, popularity, number of subscribers to a project, and development status of the project (self-reported on the site and coded 1 = planning, 2 = pre-alpha, 3 = alpha, 4 = beta, 5 = mature, 6 = stable). We also record whether or not the project has posted a new release since the last observation.

During the first data collection period, we recorded data that were presumed to be relatively constant, including the license and target audience. These are currently being revisited to assess our assumption that they would be unchanging. Licenses are currently being analyzed to classify their compliance with OSI standards. Because this process is not yet complete, none of the analyses below include license data. We determine affiliation with an organization (i.e., sponsorship) by reading the description of the project and examining the URL of the project home page. If either of these indicates an affiliation, a sponsorship dummy variable is coded as 1, otherwise it is coded as 0. For 68 projects, the coder was uncertain as to the affiliation, hence we are currently having a second coder examine all projects to determine sponsorship. Of the projects for which an initial determination was made, 43 had an organizational sponsor and 125 did not.

Table 1 displays descriptive statistics about the projects during the period March 7, 2002, to June 30, 2002. For some projects, development status was not listed and for some releases were posted on another site, thus there is missing data. Table 2 shows how many projects displayed net increases and decreases in variables of interest over the observation period.

Table 1. Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Dev.
Change in Vitality	236	-0.85	1.08	0.02	0.14
Change in Popularity	236	-0.36	2.72	0.11	0.28
Change in Subscribers	236	-1.00	29.00	1.66	4.11
Number of Releases	182	0.00	11.00	1.08	1.90
Starting Development Status	165	1.00	6.00	4.16	1.06
Days registered at end of observation period	236	92.00	1579.00	435.58	408.62

Table 2. Project Changes over the Observation Period

	Net Increase	No Change	Net Decrease
Subscribers	87	146	5
Vitality	41	159	38
Popularity	140	37	61
Development Status	6	232	0
New Releases	77	161	0

Table 3. ANCOVA Predicting Change in Popularity

Source	Type III Sum of Squares	df	Mean Square	F	p
Change in vitality	.264	1	.264	3.313	.071
Development status	1.011	5	.202	2.537	.032
Sponsorship	.529	1	.529	6.637	.011
Project category	.581	2	.291	3.650	.029
End user target audience	.009	1	.009	0.108	.743

R Squared = .216 (Adjusted R Squared = .149)

We have conducted several preliminary analyses examining overall popularity and vitality as well as changes in popularity and vitality over the observation period. Space constraints prevent us from presenting the details of all of those analyses; hence we present only a few representative results. Table 3 shows the results of an ANCOVA predicting overall change in popularity over the observation period based on overall change in vitality, sponsorship, starting development status, project category (games, software, or utilities), and whether the target audience was limited to end users or not. The analysis explained approximately 22 percent of the variance in popularity changes.

Increases in vitality had a significant ($p < .10$) positive effect on popularity. Means comparisons indicated that sponsored projects showed significantly greater increases in popularity ($p < .05$) than non-sponsored projects; projects in the utilities category showed greater increases than those in games ($p < .10$) or software ($p < .05$). Projects in the stable development category showed greater increases than those in the alpha ($p < .10$), beta ($p < .01$), or mature ($p < .01$) categories. However it is worth noting that only seven projects were in the stable category (planning and pre-alpha had five and three projects respectively; 113 projects were in the other three categories). Projects in the alpha stage showed greater increases ($p < .10$) than those in the mature category. A similar ANCOVA predicting increases in vitality showed no significant effects.

4 CONCLUSION

The study is still at an early stage, and the preliminary results described above must be interpreted with caution. Based on the data and analyses to date, it appears that vitality has a significant effect on popularity over time, indicating that the more active a project is in terms of posting new releases and making announcements, the more attention it receives from the community. Sponsored projects achieved greater success in terms of popularity than non-sponsored projects. We found this result interesting, and as a next step we plan to investigate how and why sponsorship affects project outcomes. We were surprised that vitality was not affected by sponsorship, development status, project category, or target audience, and we plan to examine other possible antecedents.

5 REFERENCES

- Behlendorf, B. "Open Source as a Business Strategy," in C. DiBona, S. Ockman, and M. Stone (eds.), *Open Sources: Voices from the Open Source Revolution*. Sebastopol, CA: O'Reilly, 1999.
- DiBona, C., Ockman, S., and Stone, M. *Open Sources: Voices from the Open Source Revolution*. Sebastopol, CA: O'Reilly, 1999.
- Feller, J., and Fitzgerald, B. *Understanding Open Source Software Development*. London: Addison-Wesley, 2002.
- Gallivan, M. J. "Striking a Balance between Trust and Control in a Virtual Organization: A Content Analysis of Open Source Software Case Studies," *Information Systems Journal* (11), 2001, pp. 277-304.
- Markus, M. L., Manville, B., and Agres, C. E. "What Makes a Virtual Organization Work?," *Sloan Management Review* (42:1), 2000, pp. 13-26.
- Raymond, E. S. *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*. Sebastopol, CA: O'Reilly, 2001.