25 THOUGHTS ON STUDYING OPEN SOURCE SOFTWARE COMMUNITIES

Joseph Feller
Business Information Systems
University College Cork
O'Rahilly Building
Cork
Ireland

Abstract

This speculative paper offers a preliminary discussion of the academic study of open source software (OSS) communities. The paper includes (1) a description of the general growth of OSS research, (2) the argument that an understanding of OSS communities is critical to understanding OSS, (3) a review of the research to date which approaches OSS from a social or ethnographic viewpoint, and (4) an examination of one of the key issues to be addressed in future research on the subject.

1. INTRODUCTION

This speculative paper offers a preliminary discussion of the academic study of open source software (OSS) communities. In section 2, I summarize the massive increase in academic attention to OSS, and the wider context of industry and public acceptance of OSS. In section 3, the growing body of OSS research is examined, and the trend toward studying the social aspects of OSS is described. Furthermore, I argue that this trend is to be encouraged, and that an understanding of OSS communities is critical to the understanding of OSS in general. Finally, section 4 outlines one of the key issues to be addressed in the study of OSS communities.

2. THE GROWING SURGE OF OSS RESEARCH

In the last two years, there has been an extraordinary increase in the academic investigation of OSS. While earlier investigations of OSS tended to be either journalistic (Moody 1997) or testimonial (DiBona et al. 1999) in nature, more recently OSS has been rigorously examined from a variety of academic disciplines, including software engineering, systems development, and information systems. Research has appeared in a number of major journals and conferences, including *Communications of the ACM* (e.g., Fielding 1999), *First Monday* (e.g., Kuwabara 2000), *IEEE Internet Computing* (e.g., Lawrence 1998), *IEEE Software* (e.g., Hecker 1999), the *Americas Conference on Information Systems* (e.g., Hars and Shaosong 2000), the *European Conference on Information System* (e.g., Ljungberg 2000), the *International Conference on Information Systems* (e.g., Feller and Fitzgerald 2000) and the *International Conference on Software Engineering* (e.g., van der Hoek 2000). At the same time, several book-length treatments of OSS have been, or are soon to be, published (e.g., Feller and Fitzgerald, 2001; Pavlicek, 2000; Rosenberg 2000).

The surge of research evidenced above is highly encouraging, although not particularly surprising. In the past few years, open source software (OSS) has established itself in the public eye in a number of ways. One factor has been the extraordinary adoption rate and market share of the Linux operating system and the Apache web server. Even more significant (given the fact that the low cost of these products tends to color the statistics) is that OSS has become a part of the long term strategic plans of major companies like IBM, Sun, and Netscape. Also, although OSS companies are currently suffering along with the rest of the high tech stock market, these same companies (such as VA Linux) boast record breaking investment histories. Even the U.S. government has taken OSS under it's wing; former President Bill Clinton's *Information Technology Advisory Committee* advised the federal government to back "open source software as an alternate path for software development" (Lohr 2000).

Even without the enormous mainstream growth calling attention to it, OSS is an intriguing phenomenon in its own right. To address the most obvious, OSS development methods openly challenge much of the conventional wisdom surrounding software development methodologies, life-cycle models, and project management. Likewise, the distribution and licensing of OSS causes us to rethink the economic underpinnings (and consequent business models) of the software industry (Feller and Fitzgerald 2000; Ljungberg 2000). Ultimately, both of these aspects of OSS (developmental and economic) depend on the phenomenon of OSS community building and maintenance. The OSS development method (loosely characterized as massively parallel distributed collaboration) is directly dependent of OSS development communities. Economically,

OSS companies like Red Hat are succeeding largely because software users are responding positively to a software business model based on customer support communities. Thus, OSS success would seem to be twice dependent on the concept of community—be it a user support community or development community.

3. THE STUDY OF OPEN SOURCE SOFTWARE COMMUNITIES

In his much cited trilogy of papers, "The Cathedral and the Bazaar," "Homesteading the Noosphere," and "The Magic Cauldron," Raymond (1999) set a useful precedent for approaching open source software communities first and foremost as *communities*, not as "software development teams" or "demographic sectors of users" or anything else. Once we see the OSS communities in this way—as groups of people with a shared sense of identity, ethos, and world view—we are able to approach these communities utilizing the analytical tools and methods of ethnography, sociology, and related disciplines. In doing so, we can gain tremendous insight into the structures and processes which make these communities, and ultimately OSS, work.

To understand OSS, I would argue we must first understand OSS communities. Physicist Richard Feynman tells the story of being teased (as a child) by a friend because Feynman did not know the name of a bird which they saw in a nearby tree. Feynman explained his father's stance on the subject: knowing the name of the bird did not tell one anything about the bird, it was merely a label. If you wanted to understand the bird, there was only one option: to watch how it behaved (Gleick 1993). I would argue the same holds true for OSS. Technically (read legally), a piece of software is defined as open source purely by the terms of its license. "Open Source" is a certification mark held by the Open Source Initiative (OSI) (www.opensource.org) and is awarded by the OSI to software that is distributed under a license complying to the Open Source Definition (OSD) (www.opensource.org/osd.html). Although completely accurate, this definition of OSS is somewhat limited. While the issue of licensing might serve to label a particular piece of software as Open Source, it does not in any way describe the unique dynamics and characteristics of software that is developed under such a licensing scheme (Feller and Fitzgerald 2001). For example, while an open source project might be *defined* as open source because it is distributed using the GPL (GNU General Public License) or similar license, it is *characterized* by the ongoing collaboration of developers and users made possible by this license feature (and this is only one characteristic of an OSS project).

Thus, the rather circular statement "OSS is defined as software whose license conforms to the Open Source Definition" is not a useful tool for understanding OSS. For that, we require observation of behavior. For example, nowhere in the OSD does it say that, to qualify as open source, a software project must be nourished by an active community of developers, users, and advocates. However, the observable reality is that the most successful OSS projects (for example Linux, Apache, and Perl) are all sustained by just such communities, and OSS in general is spoken of as a cultural and social movement (Raymond 1999).   The investigation into how these communities are created and maintained is, therefore, essential to the understanding of the character, rather than the definition, of open source software. Furthermore, this investigation holds considerable promise for researchers focused on the related subjects of programming and computer use as a social and psychological activity, computer-mediated communication (CMC), computer supported cooperative work (CSCW), and virtual community.

By approaching the phenomenon of OSS development through the communities that support it, we can build upon the work of two disciplines which have made considerable use of social scientific methods in the field of Information and Communication Technology (ICT): computer-mediated communication (CMC) and computer supported cooperative work (CSCW). One of the driving motivations of open source is the pragmatism (sometimes affectionately termed the virtue of laziness) of programmers. While proprietary software models require developers to reinvent the wheel in isolation, the OSS model allows for programmers to build upon the work of others. It is important that OSS community researchers embrace this ethos of pragmatic laziness, and take advantage of the considerable amount of research already conducted on online communities.

CMC is highly germane to the study of OSS communities because the use of ICT (particularly e-mail, newsgroups, and the WWW) for communication and collaboration is perhaps the most common characteristic of such communities (Feller and Fitzgerald 2000; Ljunberg 2000). The substantial amount of research done on CMC (e.g., Jones 1998) should be taken into account when trying to make sense of OSS communities if our understanding is to move beyond the mere description of the media used to facilitate communication between group members. In approaching OSS, we must ask the questions of how facilitating media shape the modes of communication which they support, and how complex personal identities and group networks are able to grow within the virtual environment of cyberspace. In addressing the self-organizing nature of virtual communities (for example, Baym 1998; Rheingold 1994), the role of humor and play in community maintenance (for example, Marvin 1995), and the importance of trust in online groups (for example, Jarvenpaa and Leidner 1998), CMC

researchers have, for years, been laying the ground work for understanding the inner workings of online communities.

Although many OSS communities are interest-based (e.g., a local Linux user group) others are task-based (e.g., the network of Linux kernel hackers). OSS *development* communities thus differ most sharply from many other computer-mediated social groups (for example, the community of investors at www.fool.com) in that they are actively collaborating in building a product (or products). For this reason, CSCW, which seeks to understand both the way groups of people work and the technologies which enable them to do so (Applied Informatics and Distributed Systems Group 1997), is highly comple-mentary to the work that is to be done regarding OSS. A good deal of the OSS research to date has focused on the technology that supports OSS development, such as configuration management software (e.g., van der Hoek 2000). Much more remains to be done in addressing the non-technical aspects of CSCW, such as promotion of group awareness, group coordination, information space management, support for heterogeneous environments capable of integrating single-user applications, etc., in the context of OSS development.

Quite a bit of work has already been done on OSS communities. As referenced above, Raymond offers a substantial analysis of both the psycho-logical (individual) and societal (group) motivations for participating in OSS (development) communities, as well as observations on the dynamic cultural process within such groups. The work is significant as a piece of indigenous anthropology, since it is derived from an active community member turning the lens of participant-observation inward. Other useful pieces of analysis coming from within the OSS community include Fielding's (1999) essay on leadership structures within the Apache Project and Torvalds' (1998) interview with *First Monday,* which offers his personal perspective on the motivations of OSS development communities.

Raymond's work has also been significant in stimulating both derivative and contrapuntal analysis. In his critical reaction to Raymond's work, Bezroukov (1999a, 1999b) offers a completely alternative analysis of OSS community dynamics, likening it to that found within academic and scientific communities. Yet another perspective (more Raymond-friendly than Bezroukov's) can be found in Kuwabara (2000). Kuwabara's research is based on ethnographic (e-mail) interviews, and notably applies the concepts of complex self-organizing systems to OSS communities with great effect.

Coming from a more mainstream Information Systems (IS) perspective, Hars and Shaosong (2000) have proposed a useful taxonomy of motivations for participation in OSS communities, and Sawyer (2000) has compared the social aspects of OSS development with other developmental models. Finally, Ljung-berg (2000) begs the question of what impact the OSS community model will have on other forms of knowledge organizations.

In all of this research, a common perspective is apparent: OSS, to be understood, must be understood on a socio-cultural level. This perspective has led to a reliance on ethnographic data, whether coming from emic accounts like Fielding or Torvalds, or etic accounts like Kuwabara or Sawyer. I predict that this should (and most likely will) continue to be the case. If our understanding of OSS is tied to our understanding of the communities which support it, then the toolkit of ethnography, particularly as it has been used in CMC, CSCW, and the study of virtual community, will play a critical role.

4. KEY ISSUES

The possible agenda for researching OSS communities is quite broad, as is any investigation of an instance of human cultural process. In this section, I briefly discuss one research question which I consider to be both timely and critical: By what means will OSS communities survive, and adapt to, the growing commercial reality of OSS? The shift in label from "Free Software" to "Open Source Software," which has been the source of considerable contestation (Stallman 1999), has had the desired effect. The idea that open/free vs. pro-prietary software somehow equaled non-commercial vs. commercial (which was never Stallman's intention) has been shattered. The various business and techni-cal aspects of OSS, and OSS's enormous commercial success, represents a large, complex research agenda in itself. For present purposes, it is important simply to question how the commercialization (and commercial success) of open source will effect communities.

The nature of OSS user communities has been profoundly changed by commercial interest in OSS. While early OSS communities were generally self-organizing groups, there are now a growing number of commercial entities playing the role of community organizer and host (such as Red Hat or O'Reilly and Associates). This change does not make such groups less of a "community," but it does mean that they more closely resemble the virtual community business model articulated by Hagel and Armstrong (1997), rather than "in the wild" online communities, like the much studied WELL (Rheingold, 1994).

Likewise, the emergence of pure-play OSS companies like Open Sales (http://www.opensales.com/) and developer auction sites like SourceXchange (http://www.sourcexchange.com/) are likely to have an impact on the dynamics of development communities. Within these contexts, developers are asked to scratch an employer's/client's itch, not their own, calling into question one of the most cited motivations for personal participation in OSS development (Raymond 1999). Above I argued that OSS success seems to have stemmed from the robust activity of OSS community groups. There is no reason why this

should not continue to be the case. Therefore, to understand the viability and longevity of the OSS development and business models, I would argue that it is critical to understand whether, and in what form, these communities will survive the success they created.

5. REFERENCES

Applied Informatics and Distributed Systems Group. *CSCW: Computer Supported Cooperative Work*, 1997 (http://www.telekooperation.de/cscw/).

Baym, N. K. "The Emergence of Online Community" in *Cybersociety 2.0: Revisiting Computer-Mediated Communication and Community*, S. G. Jones (ed.), Thousand Oaks, CA: Sage, 1998.

Bezroukov, N. "A Second Look at the Cathedral and the Bazaar," *First Monday* (4:12), 1999a (http://firstmonday.org/issues/issue4_12/bezroukov).

Bezroukov, N. "Open Source Software as a Special Type of Academic Research (A Critique of Vulgar Raymondism)," *First Monday* (4:10), 1999b (http://firstmonday.org/issues/issue4_10/bezroukov/991012_index.html).

DiBona, C., Ockman, S., and Stone, M. (Eds.). *Open Sources*, Cambridge, MA: O'Reilly, 1999.

Feller, J., and Fitzgerald, B. *Understanding Open Source Software: Developmental and Business Models*, London: Pearson Education, 2001.

Feller, J., and Fitzgerald, B. "A Framework Analysis of the Open Source Software Development Paradigm" in *Proceedings of the Twenty-First International Conference on Information Systems*, W. J. Orlikowski, S. Ang, P. Weill, H. C. Krcmar, and J. I. DeGross (eds.), Brisbane, Australia, December 2000, pp. 58-69.

Fielding, R. T. "Shared Leadership in the Apache Project," *Communications of the ACM* (42:4), 1999, pp. 42-43.

Gleick, J. Geni*us: The Life and Science of Richard Feynman*, New York: Vintage Books, 1993.

Hagel, J., and Armstrong, A. G. *Net Gain: Expanding Markets Through Virtual Communities,* Cambridge, MA: Harvard Business School Press, 1997.

Hars, A., and Shaosong, O. "Why Is Open Source Software Viable? A Study of Intrinsic Motivation, Personal Needs, and Future Returns," in *Proceedings of the Sixth Americas Conference on Information Systems,* H. M. Chung (ed.), Long Beach, CA, August 10-13, 2000.

Hecker, F. "Setting Up Shop: The Business of Open-Source Software," *IEEE Software*, January/February 1999.

Jarvenpaa, S. L., and Leidner, D. E. "Communication and Trust in Global Virtual Teams," *Journal of Computer-Mediated Communication* (3:4), 1998, (http://www.ascusc.org/jcmc/vol3/issue4/jarvenpaa.html).

Jones, S. G. (ed.). *Cybersociety 2.0: Revisiting Computer-Mediated Communication and Community*, Thousand Oaks, CA: Sage, 1998.

Kuwabara, K. "Linux: A Bazaar at the Edge of Chaos," *First Monday* (5:3), 2000 (http://firstmonday.org/issues/issue5_3/kuwabara/index.html).

Lawrence, D. C. "INTERNETNEWS SERVER: Inside an Open-Source Project," *IEEE Internet Computing*, September/October 1998.

Lohr, S. "Code Name: Mainstream," *New York Times*, August 28, 2000 (http://www.nytimes.com/library/tech/00/08/biztech/articles/28code.html).

Ljungberg, J. "Open Source Movements as a Model for Organizing" in *Proceedings of the Eighth European Conference on Information Systems*, H. Hansen, M. Bichler, and H. Mahrer (eds.), Vienna, Austria, August 2-6, 2000.

Marvin, L. "Spoof, Spam, Lurk and Lag: The Aesthetics of Text-based Virtual Realities," *Journal of Computer-Mediated Communication* (1:2), 1995 (http://www.ascusc.org/jcmc/vol1/issue2/marvin.html).

Moody, G. "The Greatest OS That (N)ever Was," *Wired*, May 8, 1997.

Pavlicek, R. *Open Source Software Development : Embracing Insanity*, Indianapolis, IN: Sams Publishing, 2000.

Raymond, E. S. *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*, Cambridge, MA: O'Reilly, 1999.

Rheingold, H. *Virtual Community,* London: Minerva, 1999.

Rosenberg, D. K. *Open Source: The Unauthorized White Papers,* New York: IDG Books, 2000.

Sawyer, S. "A Social Analysis of Software Development Teams: Three Models and their Differences," in *Proceedings of the Sixth Americas Conference on Information Systems,* H. M. Chung (ed.), Long Beach, CA, August 10-13, 2000.

Stallman, R. "Why 'Free Software' is better than 'Open Source'," 1999 (http://www.gnu.org/philosophy/free-software-for-freedom.html).

Torvalds, L. "What Motivates Free Software Developers?" *First Monday* (3:3), 1998 (http://firstmonday.org/issues/issue3_3/torvalds/index.html).

van der Hoek, A. "Configuration Management and Open Source Projects," Third Workshop on Software Engineering over the Internet (at the Twenty-Second International Conference on Software Engineering), Limerick, Ireland, June 6, 2000 (http://sern.ucalgary.ca/~maurer/icse2000ws/submissions/Hoek.pdf).

About the Author

Joseph Feller is a College Lecturer with the Business Information Systems group at University College Cork, Ireland. Previously a lecturer at the Ringling School of Art and Design (Florida), Joe primarily teaches eBusiness strategy and economics, web design and development, web-database integration, and programming for Internet environments. He is the author of *Customer-Friendly: Design Guidelines for E-Commerce* (Blackhall, October 2000) and the editor of *Inside XML Solutions* (ElementK Journals). Joe earned his Ph.D. from University College Cork and his BA from The American University (Washington, DC). His current research focuses on the developmental, social, and business models of open source software (OSS), and he is writing a book on the topic with Brian Fitzgerald, to be published by Pearson Education in late 2001. Joe can be reached by e-mail at jfeller@afis.ucc.ie.