Understanding the Social, Technological, and Public Policy Implications of Open Source Software Development

Position paper

Walt Scacchi (wscacchi@uci.edu),

Institute for Software Research,

University of California, Irvine.

August 2002

Interest in open source software has emerged in many different communities. Much of this interest has focused attention primarily onto the products of open software development (source code), and secondarily onto the processes and productive units that facilitate such development. My research is focused to understanding the processes, practices, and communities that give rise to open source software. My research group is studying (a) the role of software informalisms (vs. formalisms and standards found in software engineering), (b) the emergence and articulation of open software requirements, (c) the forms and constituencies of the social worlds of open software, and (d) other processes and practices across multiple open software development communities [Scacchi 2001b, 2002a,b,c]. I am prepared to discuss our results, work in progress, and the need for further research on all of these topics. However, the remainder of this position paper identifies what I believe are areas, topics, or basic questions requiring further research in the arena of open source software development and how it may impinge on government policies. These follow in an unordered manner.

Understanding the *quality* of open source software from a socio-technical perspective

What is the best, most effective way to determine the quality of open source software products and development processes? Open source software is developed in a highly social online environment where developers are frequently dispersed in space and time, but may rarely if ever meet for face-to-face interaction. If so, how is the distributed asynchronous collaboration among developers brought into being and sustained over time? Does this mode of computer-supported collaborative work increase, decrease, or have no significant effect on the quality of the open source software? Elsewhere, it seems that the quality of open source software has been called into question, indicating that open source software can sometimes by very problematic and plagued with many critical bugs/errors. Public repositories of bug reports, errors, and other related problems/issues hosted on the Web, like Bugzilla and IssueZilla, contain hundreds to hundreds of thousands of reports for different, popular open source software systems (e.g., the Mozilla Web browser). How might these repositories be cultivated and their data systematically analyzed to see what kinds of first, second, or higher-order patterns might exist? How might such patterns better reveal the intertwined social and technical features that shape and evolve open software quality?

Understanding the occupational cultures and career contingencies of open software developers

Why people want to allocate a portion of their available time, skill, and sentiment to develop open source software is unclear. Many social constructed conditions or variables are often mentioned, including building trust and reputation; "geek fame"; being generous with one's time, expertise, and source code; and creating and contributing software as public goods or gifts [Pavlicek 2000]. However, many of the more vocal proponents of open source software [e.g., DiBona, *et al.* 2000] have financially profited or accumulated significant amounts of social capital based on the legacy of their experience and open software contributions. Do such benefits accumulate only to "early adopters" or advocates of open source software development?

Many developers of open source software do so as part of their paid job, or as a way of demonstrating their technical, communication, or social interaction skills as a way to get a (better) job. Are these conditions unique to software developers in general, or open source software developers in particular? Elsewhere, what is the role of women in the open software development community? Does open source software development, as a mode or style of technical work, tend to encourage, discourage, or have little effect in encouraging women to develop a career in the software R&D area? Last, though open source software projects engage developers who may be globally distributed, does such distribution cross all, some, or few ethnic boundaries?

Understanding the role of open source software in advancing (or inhibiting) research in the natural and physical sciences

It appears in the U.S. that many Federal research agencies will face the issue of whether or not to adopt a policy that mandates or encourages that all software developed with agency funding must be open source. However, what does such a policy imply for the advancement of scientific research and development in disciplines inside or outside of the computing research community? For example, the development of the national computation grid [Foster *et al.,* 2001] seems to rely on the use of open source software. But this community involves researchers with long track records in computer science or computational science research. In the Astrophysics community, there is growing enthusiasm for research and development of a computational "astrogrid" that can integrate dispersed astrophysical sensors and software systems to act as a *national virtual observatory* [NVO 2002]. NASA and the National Science Foundation are principal agencies funding such a grid. What principles, guidelines, or best practices should guide the development of a national research infrastructure like a virtual observatory that is to be built by astrophysicists using open source software development techniques? Elsewhere, the international medical informatics, bioinformatics (cf. http://bioinformatics.org/about and http://open-bio.org/bosc2002/), and ecological systems (http://www.open-research.org/) communities have already begun discussing why future medical, biological, and ecological research systems should be open source. However, the genomics and proteinomics communities seem reluctant to embrace open source systems or open databases, since there is still a great rush of patents being filed pertaining to intellectual property of a microbiological origin that can primarily be accessed and manipulated via software-intensive systems and instrumentation. Should a separate pool of Federal biotechnology research investments be targeted to those who develop open source

software or open databases as part of their research (cf.
http://www.openinformatics.org/petition.html)?

Understanding the role of open source software in national and international science policy

What could we learn from by comparing and contrasting efforts in the U.S., European Union, and High-Tech Asia (Japan, Korea, Taiwan, Singapore), as well as to other parts of the World in the encouragement of open source software development? For example, the European Commission's Information Society Technologies Programme [IST 2002], currently budgeted at 3.6B Euros over five years, stipulates all software developed with programme funding *should* be open source whenever possible. Does the IST programme represent a competitive advantage, disadvantage, or simply an alternative compared to U.S. Federal R&D investment which does not require open source software to be developed with its funds? Will it advantage or disadvantage the U.S. software industry, only certain firms in that industry (e.g., Microsoft, Oracle, Sun, or Adobe), or have little/no consequence? Will the IST programme policy towards open source software give rise to new products or services that are superior, inferior, or just different than those arising from U.S. funded R&D? Is a "hands-off" open source policy (i.e., the current U.S. position) or a "hands-on" policy better or worse for High-Tech Asia economies, or for emerging nation economies?

We do note that the President's Information Technology Advisory Committee [2000] report strongly advocates that all high-end computing software funding by the U.S. research agencies adopt a hands-on, open source technology policy. Should the U.S. research agencies follow the PITAC recommendation, and align or distance themselves with the policy objectives of the IST programme? Would such an alignment or distancing help, hurt, neutralize monopoly (or market dominance) positions of the U.S. software industry, or certain firms within it?

The emergence of "open government", arising from the integration of concepts and practices from open source and E-Government

How can open source development concepts be brought to bear in the realm of digital government? Digital government (or E-Government) encourages the adoption of modern IT business practices that exploit the World-Wide Web and Electronic Commerce capabilities to improve the government operations and public services. NSF's Digital Government initiative supports research in IT security and privacy [DG.O 2002], as well as the collection, statistical analysis, public access, and visual display of very large data bases of public data [DGRC 2002]. Other parts of the U.S. government are investigating E-Government strategies for procurement and acquisition [Scacchi and Boehm 1998, Scacchi 2001a], data storage and data entry (e.g., electronic filings of tax forms by individuals, and SEC forms by businesses), E-catalog based retail product sales (U.S. Mint), and smart cards [Steyaert 2001]. Outside of the U.S., nearly a dozen countries are already formulating legislation that requires or expresses preference for the adoption and use of open source software systems by national or regional government agencies [Anonymous 2002, Cortiana 2002, Nunez 2002].

In contrast, *open government* seeks to open for public sharing, discussion, review, ongoing development and refinement, and unrestricted reproduction (replication and redistribution) the "source code" of the products and processes of the business of government. Open government represents a concept that seeks more than just the adoption and use of open source software systems by government agencies. This concept seeks to explore the potential and opportunities that can emerge when one views the purpose of digital government as also including how to empower and engage an interested public in better understanding how government processes and practices can be made better, cheaper, and faster through the development of open source processes, practices, and communities of practice for government operations. Would open government allow for the establishment of community Web portals or other open testbeds (e.g., a national virtual government observatory or computational grid) where alternative government processes or practices might be (re)designed, prototyped, and evaluated via collaborative experimentation and engagement [Scacchi and Boehm 1998, Scacchi 2001a, 2002c]? Would open government systems provide new modes of access and participation in an open democracy through the development, use, and collaborative evolution by interested government system developers, industry, and citizens? Would open government represent a new avenue to explore how government operations might be made accessible for educational purposes in high school (grades 9-12) and college settings? Would open government enable more complete assessment of the financial and infrastructural costs/benefits of new legislation that is created and imposed, but otherwise be unfunded?

Overall research needs

The research community needs a better articulation and understanding of "critical mass" issues in open source software development projects. Are characteristics of large software projects working on "Internet time" [Cusumano and Yoffie 1999, MacCormack, Verganti and Iansiti 2001] fundamentally different than those projects which lack critical mass features? What may work well or be true about specific open source software development projects like the GNU/Linux operating system, Apache Web server, Mozilla Web browser, SendMail, and BIND, may not necessarily be indicative of the characteristics and critical success/failure factors of other open source software projects. For example, the SourceForge Web portal (http://www.sourceforge.net) currently hosts more than 49K open source software projects (with 70 new projects added per day) and more than 450K registered users (with 700 new users registering per day). None of the major open source software projects like GNU/Linux, Apache, or Mozilla are found on SourceForge or Freshmeat.org. Furthermore, of the 49K projects at least 10% are identified by their developers as *production quality and stable*, thus suitable for routine use by end-users who primarily want to use, rather than develop, such software. Similarly, at least 10% of the total set of projects is no further than the *interesting idea* stage of development. Many of these projects will wither due to an inability to realize their effort as a successful open source community motivated to develop and sustain an open software system. Thus, studies indicative of a single open source software project, or multiple projects of a similar kind, may produce results that cannot be generalized to other/smaller open source projects. Similarly, the results of such myopic studies may lead or mis-lead new open source projects in R&D communities with little prior experience with open source software products, processes, productive units, or information infrastructures. Thus, premature generalization of the desirable

features or characteristics of open source projects like GNU/Linux and Apache Web Server, or their adoption be particular government agencies, may give rise to inappropriate conclusions about what the best practices of open source software development really are for projects that lack critical mass and Internet time characteristics.

The scientific and policy research communities need to encourage comparative empirical study of open source software development products, processes, productive units, and information infrastructures that span different R&D communities and government agencies within and across nations. This is especially true for communities whose research or system usage interests are not primarily rooted in software system development (e.g., medicine, astrophysics, genomics/proteinomics, environmental ecology, national research ministries, and other government agencies) but who may believe that open source software is a better or best way to develop their software systems. We need to better understand how communities are similar or different in their patterns of software system development and use, and how open source processes and practices align within these communities and patterns.

Last, the scientific and policy research community needs to encourage interdisciplinary study of open source software development, especially studies rooted in understanding and advancing economic analyses, innovation theory, and collaborative software development processes and practices.

Acknowledgements

References

(Anonymous), Legislation on the use of Free Software in the Government, http://www.lugcos.org.ar/serv/mirrors/proposicion/doc/referencias/#ref.#1, August 2002.

F. Cortiana, "Norms in matter of computer science pluralismo on the adoption and the spread of the free software and on the portabilità of documents informed to us in Public Administration" (XIV Legislatura Action Senate n. 1188) (in Italian) April 2002. http://www.softwarelibero.it/altri/cortiana.shtml (English translation, http://translate.google.com/translate?hl=en&sl=it&u=http://www.softwarelibero.it/altri/cortiana.shtml&prev=/search%3Fq%3Dwww.softwarelibero.it/altri/cortiana.shtml%26num%3D100%26hl%3Den%26lr%3D%26ie%3DUTF-8)

M. Cusumano and D.B. Yoffie, Software Development on Internet Time, *Computer*, 32(10), 60-69, October 1999.

C. DiBona, S. Ockman and M. Stone, *Open Sources: Voices from the Open Source Revolution*,

O'Reilly Press, Sebastopol, CA, 1999.

(DG.O) Digital Government.Org, http://www.diggov.org, 2002.

(DGRC) Digitial Government Research Center, http://www.isi.edu/dgrc/, 2002.

FastLane, National Science Foundation, https://www.fastlane.nsf.gov/, 2002.

I. Foster, C. Kesselman, and S. Tuecke, The Anatomy of the Grid: Enabling Scalable Virtual Organizations, *Intern. J. High Performance Computing Applications*, 15(3), 200-222, 2001.

(IST) Information Society Technologies Programme, The European Commission, http://www.cordis.lu/ist/, 2002.

A.MacCormack, R. Verganti, and M. Iansiti, Developing Products on Internet Time: The Anatomy of a Flexible Development Process, *Management Science*, 47(1), January 2001.

(NVO) National Virtual Observatory, http://www.us-vo.org/, 2002.

E.D.V. Nunez, *Peruvian Congressman Refutes Microsoft "Fear, Uncertainty, and Doubt" (F.U.D.) Concerning Free and Open Source Software*, 8 April 2002, (original in Spanish), http://www.opensource.org/docs/peru_and_ms.html.

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

(PITAC) President's Information Technology Advisory Committee, *Developing Open Source Software to Advance High-End Computing*, September 2000. http://www.ccic.gov/pubs/pitac/pres-oss-11sep00.pdf

W. Scacchi, Redesigning Contracted Service Procurement for Internet-based Electronic Commerce: A Case Study, *J. Information Technology and Management*, 2(3), 313-334, 2001a.

W. Scacchi, Software Development Practices in Open Software Development Communities: A Comparative Case Study, presented at The 1st Workshop on Open Source Software Engineering, Toronto, Ontario, May 2001b.

W. Scacchi, Understanding the Requirements for Developing Open Source Software Systems, *IEE Proceedings - Software*, 149(1), 24-39, 2002a.

W. Scacchi, Is Open Source Software Development Faster, Better, and Cheaper than Software Engineering?, *Proc. 2nd. Workshop on Open Source Software Engineering*, Orlando, FL, May 2002b.

W. Scacchi, Exploring Open Software System Acquisition Processes and Architectures, Final Report, Institute for Software Research, University of California, Irvine, July 2002c.

W. Scacchi and B.E. Boehm, Virtual System Acquisition: Approach and Transitions, *Acquisition Review Quarterly*, 5(2), 185-216, Spring 1998. Also see,

http://www.dsmc.dsm.mil/pubs/arq/98arq/scacchi.pdf

J. Steyaert, (Deputy associate administrator, Office of Information Technology, GSA), View From the Top: What Government is Doing, *Solution Series Conf. On E-Government*, 24 April 2001. Webcast version at http://www.gcn.com/webcast/070201gcn.html