

Free/Libre and Open Source Software: Survey and Study

FLOSS

**Workshop on
Advancing the Research Agenda on Free / Open Source Software**

**14 October 2002, Brussels
European Commission**

Workshop report by Rishab Aiyer Ghosh

International Institute of Infonomics
University of Maastricht, The Netherlands

October 2002

<http://www.infonomics.nl/FLOSS/workshop/report/>

© 2002 International Institute of Infonomics

The workshop summary is based on notes taken by Philippe Aigrain, Rishab Ghosh, Rüdiger Glott and Bernhard Krieger during the 14 October 2002 workshop on "Advancing the Research Agenda on Free / Open Source Software". This is to be read in conjunction with the statements submitted to the workshop, available at www.infonomics.nl/FLOSS/workshop/papers/ (see also the slide presentations at the workshop on the same web page). More information on the workshop, including the participants list, is available at <http://floss.infonomics.nl/workshop/>

Introduction / Agenda	3
Format of the workshop	4
Comments on the FLOSS Final Report	4
Research priorities	5
Creativity, Community	6
Modeling, prediction and empirical research.....	7
Open source and the world: dynamics, learning, social barriers	8
Innovation, incentives, organisation and structure	9
Standards and interoperability	9
Funding open source	10
Incentives & IPR, licensing schemes & policy.....	11
Summary list of research questions	12
Creativity, Community	12
Measurement, productivity, efficiency and security.....	12
Modeling, prediction and empirical research.....	12
Open source and the world: dynamics, learning, social barriers	13
Innovation, incentives, organisation and structure	13
Standards and interoperability	13
Funding open source	14
Incentives & IPR, licensing schemes & policy.....	14

Introduction / Agenda

(by Brian Kahin and Rishab Ghosh)

Open source software is one of the unique phenomena of the digital economy. Enabled by the Internet, it has grown dramatically over the past decade. It includes a major operating system (Linux), the dominant web server software (Apache), and thousands of individual projects. It has been embraced by large computer companies and has been hailed a paradigm shift in software development.

These successes, the unusual nature of open source development, and the benefits claimed for open source software – security, reliability, adaptability, as well as economy and openness – create opportunities for complementary businesses and pose important policy questions. While there is still little published research on the open source phenomenon, there is a large amount of work in progress. Open source has already inspired far-ranging speculation and debate about institutional economics, information architectures, intangible assets, innovation processes, standards, ethics, contracts, and intellectual property policy. At the same time, there are divergent views and practices within the open source movement reflected in differences in license, motivation, and business orientation.

The International Institute of Infonomics at the University of Maastricht and the Center for Information Policy at the University of Maryland are organizing a one-day workshop at the offices of the Directorate General for the Information Society in Brussels. This workshop, to be held on October 14, 2002, follows a workshop held January 28 at the National Science Foundation in Arlington, Virginia. The workshops explore the issues raised by open source development, its relationship to other forms of enterprise and community, and the implications for institutions and public policies in a digital society. They are intended to help develop a vigorous global research community around open source with connections to the open source development community and industry. They seek to enhance the visibility of open source studies within the social sciences and professional disciplines and to help policy makers better understand the enterprise-transforming nature of the Internet and the special characteristics of Internet-enabled innovation.

31 invited participants (21 from the EU, 10 from the US) submitted their views on research priorities in advance; these statements were used to structure the substantive discussion and are made available on the Web. The workshop was open to 19 pre-registered observers. A list of participants is included in this report.

The workshop was organized and supported through the FLOSS project. The participation of US speakers was made possible through a US National Science Foundation (NSF) grant from the Program on Societal Dimensions of Engineering, Science, and Technology and the Program on Digital Society and Technologies (to the University of Maryland Center for Information Policy).

Format of the workshop

The workshop was in two parts: the first part was a brief presentation of findings from the FLOSS project, with the assumption that participants were familiar with the previously published version of the FLOSS final report. This was followed by a constructive discussion on the FLOSS findings, including suggestions that are incorporated in the current (and final) version of the report. A summary of comments to the FLOSS report is in the next section.

The main part of the workshop was structured to enable the maximum possible interaction and discussion among researchers and practitioners at the workshop venue itself, with a strong focus on the development of research priorities. To facilitate this, participants were required to submit short background statements prior to the event, in order that they be read by all the participants before the workshop itself.

The statements were intended to provide participants' individual answers to the question: What in your opinion are the most important questions for further research on the free/open source phenomena?

The statements were elaborated upon through short presentations at the workshop limited to 7 minutes each, grouped by related topic, followed by discussion. The intention was to get a sense of a combined answer to the question above representing the views of the research and practitioner community, which is further described in a later section of this report.

Observers were given an opportunity to interact and comment throughout the event.

Comments on the FLOSS Final Report

At the workshop summary presentations were made of the findings of the FLOSS project, specifically the user survey (Final Report part I), developer survey (part IV) and source code analysis (part V).

What follows is a summary of comments arranged by the section of the report to which they refer, and any response if provided by the FLOSS authors.

The majority of comments referred to part V of the FLOSS report.

Ilkka Tuomi expressed skepticism as to the possibility of analysing source code for author names especially considering the variety of methods of claiming credit in source code. Rishab Ghosh responded that the purpose of the FLOSS source code scan was not to identify detailed developer rankings, but patterns of authorship rather than specific authors, for which the methodology seems to be reliable even when compared with manual analysis on a smaller scale. Naturally, further analysis and drawing more detailed conclusions that what were presented in the FLOSS study requires greater care in terms of ensuring reliability of data extracted from source code.

Ben Laurie noted that the Apache Software Foundation discourages people to claim credit in source code directly, and for Apache it would be better to study the CVS (version control) records. Rishab Ghosh agreed that methods have to be adapted to the way projects use tools and claim credit, in order to conduct more detailed analysis, and referred to further analysis being conducted of the Linux kernel.

Alan Cox noted that in terms of validating automatically extracted data, one useful thing would be an explicit manual comparison on a variety of projects by people who are well placed to know how to analyse it for that project. It is dangerous to assume that one algorithm could work for all projects, due to their great variety of ways of working of projects.

James Herbsleb noted that a very small percentage of developers contribute the vast majority of code and that the FLOSS survey does not distinguish between major and marginal contributors (the number of projects one contributes to is not the same thing), leading to the survey results as a whole being dominated by the marginal participants, who may hold different views from those doing most of the work. Ghosh concurs but notes the FLOSS developer survey does allow some differentiation between major and marginal contributors, using variables such as hours spent per week on development, which have been used to aggregate views of different categories of developers. Herbsleb also suggested the correlation of data from the developer survey with that of the source code scan; this is being conducted (outside the scope of the FLOSS project) but care has to be taken as the data sets are different.

James Leach notes that “authorship” is not necessarily the right concept for developers who claim credit of source code. Authorship has a specific history in Europe. The problem is the implication of ownership. Rishab Ghosh agrees, and clarifies that the FLOSS project uses several terms, but sees authorship not only as the source of copyright, but more on being the source of work, the retainer of moral rights; besides, FLOSS uses the term “author” interchangeably with “developer” for want of a better term.

With reference to the developer survey (part IV of the FLOSS report), Nicolas Pettiaux asked about the comparison with the BCG/OSDN survey, in particular on contributors being paid. This is somewhat addressed (with regards to methodological issues) in FLOSS Final Report Part IVa: Survey of Developers - Annexure on validation and methodology.

Research priorities

Individual research priorities were best expressed by workshop participants in their background statements. What follows is a distillation of the main questions posed at the workshop with an attempt to organize them by category and/or discipline. References are made where relevant to the presentation at the workshop or a background statement (at the end of a section, in square brackets, “[Leach]”), and to specific comments from participants during discussion (following the comment, in braces, “{Cremer}”). After

reading a section, the statements of the participants in brackets should be read to know more about the topics summarized there. Participants' names may be listed under more than one section, and are usually listed following the sequence of their contribution's appearance (i.e. the first name may be more relevant to the beginning of the section).

References to the FLOSS research are also listed, as FLOSS5 for part 5 of the FLOSS final report.

As the purpose of this document is to highlight research priorities, phrases that summarise questions or issues of particular importance have been *italicized*. Reading the sequence of italicized phrases should therefore provide a concise synthesis of the workshop's outcome. Keywords (methodologies and or disciplines) are listed at the end of each section.

As the order of workshop sessions was based on category and discipline, this list tends to follow the order of presentations at the workshop.

Creativity, Community

The question of collaborative creativity and motivation, and how this impacts the process of organization and development of open source projects was a research priority with a clear anthropological and socio-psychological disciplinary basis. The relevant issues raised here were methodological: a preference for an anthropological, even ethnographic approach; as well as theoretical: the idea of the "gift" referring to Marcel Mauss. *What is the impact of ideology, personal trust, and a common set of beliefs on modes of organization and development?* Research into these issues can shed light on *what policies and ownership regimes could foster creativity and inclusion*, and also on whether different (possibly new) forms of organization used by open source developers can lead to greater or lesser success. Open source is certainly not the only arena of massive collaborative, semi-anonymous authorship, other examples include for instance the original Oxford English Dictionary {Herbsleb}. As a corollary, *what can we learn from open source modes of organization and production that could perhaps be applied to other areas as well?*

There is a need for ethnographic study of open source by researchers who actually go in the field, as it were, seeing how open source community participants live, work and interact {Iacono}.

[Leach, Stewart, Cox, Crowston]

Keywords: anthropology, sociology/psychology; ethnographic research

Measurement, productivity, efficiency and security

To some extent related to the line of research developed in Part V of the FLOSS report, *there is a whole research area in the measurement and quantification of specific aspects of the open source organization and development process*. Starting with the software

engineering approach of productivity cycles, code reuse and the density of bugs, to the analysis of success in terms of complexity of code or frequency of release, usage, adoption, *it is apparent that a number of tools are required to quantify metrics for the further analysis of the open source process*. This is one way of trying to answer the question: *What is efficiency in the context of open source development? What is the concentration of (and what are the methods of) contribution in development?* Contribution needs to be measured through several sources – code itself together with the version control and management data that come with the development process, but also the discussion groups, documentation and other processes that go into collaborative authorship.

Furthermore it is important to quantify efficiency and productivity in order to determine the benefits of the open source development model, especially in the area of product security (the many-eyes hypothesis about open source bug fixing). Measurement is not something limited to the developer end of the product chain – it also concerns users and deployment. In the area of security, for instance, how (and how efficiently) do bug fixes and innovations propagate from developer to end-user? *How does the development model scale? How much modularity is there in the production model, and how much duplication?* (Possibly very little duplication {Cox} and possibly irrelevant in the sense that developers aren't primarily driven by efficiency-related incentives {Glott}.)

What is the relationship between incentives and efficiency – especially in terms of improving security? Given the nature of bugs (which are not necessarily apparent) careful reviewers spend most of their time not actually making changes, thereby not getting rewarded {Laurie}. Is this a mismatch between incentives/motivations and desirable results? Do the incentives of reputation work against “mundane” but still very valuable tasks, such as reviewing and testing code, proper documentation etc?

[Stewart, Daffara, Jones, Laurie, Tuomi, Herbsleb, Fitzgerald, FLOSS5].

Keywords: software engineering analysis techniques, analysed from the perspectives of economics, sociology, software engineering.

Modeling, prediction and empirical research

Models of open source production need to be solidly grounded in empirical data, and in the way open source actually functions (i.e. this relates to the previous topic, measurement). This necessarily means looking for models without price-based markets, which aren't present in open source. However, there are other modes of collaborative production that are not based on priced markets, such as the academic (“open science”) community. The question then is: *are there other types of mechanisms, not based on markets, prices and their invisible hand, that govern production and allocation?* Clearly the answer is yes, and the open source is exploring quite a few of them.

A key issue is to *identify ways of predicting how open source allocates resources* – there are several models *based on multiple motivations driving participants*, including that of user innovation, personal motivations, reputation and product-focus. How do these

motivations and models combine to determine the trajectory of open source development, and indeed what will and what will not be developed? *To what extent does the open source model provide economic efficiency, and how does that relate to the fact that economic motivations are not necessarily important to participants themselves* {Glott}?

Is it possible to learn about or predict the behaviour of open source communities by running simulations of them (agent-based model)? Simulations can be used, among other things, to identify the impact of different information exchange models or different policy – especially IPR policy – regimes. It may be possible to simulate the interaction between rights-based appropriation of information and an information commons, to see how policy decisions may affect the future development of open source communities.

[Cremer, David, Dalle, Herbsleb, Jones, Tuomi, Crowston, Burt]

Keywords: economics, economic modeling, AI/agent-based simulation, empirical data collection and analysis

Open source and the world: dynamics, learning, social barriers

Many people are studying open source as an example of something else - we need a concerted effort on open source in itself, for itself and what it means to the world {Iacono}. What do processes of open source mean for the global economy, for the software industry, for the information industry, for developing countries?

Dynamics: what are the aspects of open source organization that evolve over time, and how do they do so? There is a need to examine the importance of learning/training in the open source process – open source communities are nurseries for the next generation of software developers, and an approach to solve the IT workforce shortage issue {Iacono}; many developers join at school age and quickly become leaders {Cox}.

How poorly are women represented in open source communities? Are there areas – design, documentation, discussion {Cox} *that have a higher representation of women? Are there specific attributes of the open source development method – aggressive, ego-driven or competitive leadership, say – that add to a bias against women? In corollary, are patterns of interaction and organization between women already active in open source (and between women and men) different as compared to men alone? What can be done to bring more women into the open source world?*

Shouldn't open source be compared as a developmental model with closed source software development? Modularity and distribution are considered emblematic properties of open source, but proprietary software development is also based on these principles. However, modularity is not purely technical, but a division of labour, based on a global political organization of countries that may reinforce economic or political superiority in closed-source approaches. *Is the transparency in organizational and developmental structure of open source more conducive to a meritocracy that ignores national, economic, socio-political hierarchy?*

How do language barriers (especially east Asian) create divisions within the open source source community? {Cox}

[Iacono, Metiu, Tuomi]

Keywords: sociology, policy, organizational studies

Innovation, incentives, organisation and structure

What is the form of organization in open source, and how does it fit with traditional classifications of organization systems? Both input and output in open source are non-rival intangibles, making it very difficult to even see what is the input and what is the output.

How does one measure (and predict) innovation in the open source model? What is the link between motivation, incentives and open source innovation? How do non-profit and profit-driven motivations affect open source development and software quality? The developers-are-users or “user needs” model implies that open source will not innovate in “grandmother-friendly” applications, requiring either closed-source or company-sponsored open source development {Kuan}. On the other hand, other motivations (originally ideology in the case of GNOME) can lead to “grandmother-friendly” applications {Cox}.

Can incentives and motivations among participants in projects/communities be correlated to their organizational structure, and structure correlated to quality, efficiency or other measures of “success”? Different projects have different distributions of contribution, stratified into core groups, or a periphery with lots of small contributions and a varying degrees of cooperation and competition. People at the bottom of hierarchy do not struggle against each other, but may compete for the attention of the core group. In organisational terms, there is much freedom of movement and competition at every level. There is often more competition in F/LOSS projects than in proprietary software projects {Ghosh}. Is it possible to measure the impact of intra- and inter-project competition and cooperation in open source communities, to correlate with and predict “success”?

[Garzarelli, Kuan, Crowston, FLOSS4]

Keywords: economics, sociology, innovation and organizational studies

Standards and interoperability

What is the impact of open source on standardization and interoperability in software and ICT? Open source can contribute to solving this problem of standardization, since in itself the availability of the source code increases the transparency of software and eases the development of compatible complementary software even where no formal standard is defined or adhered to. Open source may be more important, though, as a solution to the related problem of vendor-dependence and lock-in.

Although a distinction should be made between formal standardization and de facto standardization, (the latter is nothing more than software development, whether proprietary or open source) what matters is the extent of real interoperability achieved - not whether a formal standard has been developed. *What is the impact of different open source licenses on standards? Do some licences (e.g. GPL) increase the likelihood of retaining interoperability while making improvements to de facto standards?*

Open source provides some degree of de facto interoperability, an open source reference implementation could reasonably make anything a candidate for being treated as a standard {Ghosh}. *What is the impact on interoperability of requiring open source reference implementations for any “open” standard? How does open source, or a reference implementation, affect barriers to entry?* {Ghosh}

What is the impact of the ability to freely change development trajectories in open source (code-forking) have on interoperability and on innovation?

[Egyedi]

Keywords: economics, law, standards policy

Funding open source

What are the current ways that open source development is funded (publicly or privately)? Are there better ways to fund development, and to reward and recognise success and innovation through the recognition of contribution to open source? Even within companies, open source software gets recognized and used quicker than proprietary software, where strong barriers exist between development teams. More research is required on interoperation between open source and proprietary software, and on reward systems for open source.

Is dual licensing – GPL and proprietary commercial in parallel – useful as a viable business model? A model of licensing source as GPL for open source use while licensing separately for clients is used by businesses such as MySQL {Ärno}. It is also a possible solution to the public funding / corporate use debate, in that publicly funded or academic software could be released under the GPL for non-commercial use, where modifications must be retained in the “commons”; and released separately under commercial licence for incorporation or modification in proprietary software.

What are the ways in which public funds are being (or could/should be) used for the development of open source and what is their impact – on open source, public infrastructure and the software market? Public funds can be used to support open source in several ways – indirectly, through the acquisition of open source software for use in the public sector; through the requirement that software developed in the public sector (or customized and developed for the public sector) should be released as open source; or directly through the funding (or initiation) of open source software development projects. In some areas competition issues need to be explored in the use of public funds; in other

areas, such as academic science and engineering, much software is simply released as open source by default.

[Cathcart, Schmitz, Strawn]

Keywords: economics, business studies, policy

Incentives & IPR, licensing schemes & policy

What role does IPR play in supporting open source? Copyright law has been used to enforce the GPL and ensure that open source code is not re-appropriated to the exclusive advantage of any one party, as could happen with public domain software.

How do the incentives of IPR protection relate to innovation in open source? Do IPR regimes, especially software patents, act as disincentives for open source, and create entry barriers to software innovation? Surveys (Frauenhofer/Blind) show the software sector in general has a limited use of patents, with heavier use in large companies, but the motivation behind this use is claimed to be defensive against infringement cases – rather than as an incentive to innovate, and even proprietary software developers (except for very large firms) feel threatened by IPR in the software sector. For open source developers, any royalty is too high an entry barrier – individual developers may be liable for royalties or simply threatened with legal action, making open source development impossible.

What is the relationship between IPR and open standards? What are the implications on standardization and innovation of different licensing regimes, including different choices of open source licence? Tension between IPR and standardization exists in software and also telecommunications, especially in areas that could be termed “essential public infrastructure”. Different sorts of licensing regimes have a significant impact on the openness of standards, and especially on interoperability with open source software. Reasonable and Non-Discriminatory (RAND) licensing turns out to be discriminatory against open source developers unless it is completely royalty-free, for the entry-barrier reasons described above. Open source standards risk being “embraced and extended” into proprietary versions and evolving into de facto, but proprietary standards, unless the “openness” is protected (e.g. by GPL/commercial dual-licensing) {Ghosh}.

What IPR issues are faced by potential users/deployers of open source? Industrial deployment of open source may require careful examination of potential patent violations embedded in the source that may not have attracted legal attention simply when individual developers were responsible, but may well attract lawsuits when there is a large user that can be held liable. Uncertainty over possible third-party IPR claims in open source code may prevent large scale industrial deployment {Jaaksi}. Similarly, companies considering releasing their code as open source may not do so as any hidden patent infringements in their code will become apparent as soon as source code is available {Cox}.

[Van Alstyne, Blind, Kahin]

Keywords: economics, innovation, standards, IPR law, policy

Summary list of research questions

This is a sorted summary list of priority research questions highlighted from the overview of research priorities in the previous section. As with the previous section, references to background statements of workshop participants are made in [brackets] at the end of each section followed by keywords: discipline/methodology/area of study discussed.

Creativity, Community

What impact does ideology, personal trust, and a common set of beliefs have on modes of organization and development?

What policies and ownership regimes could foster creativity and inclusion?

What can we learn from open source modes of organization and production that could be applied to other areas as well?

There is a need for ethnographic study of open source by researchers who actually go in the field to see how open source community participants live, work and interact.

[Leach, Stewart, Cox, Crowston]

Keywords: anthropology, sociology/psychology; ethnographic research

Measurement, productivity, efficiency and security

A whole research area in the measurement and quantification of specific aspects of the open source organization and development process. A number of tools are required to quantify metrics for the further analysis of the open source process.

What is efficiency in the context of open source development? What is the concentration of (and what are the methods of) contribution in development?

How (and how efficiently) do bug fixes and innovations propagate from developer to end-user? How does the development model scale? How much modularity is there in the production model, and how much duplication?

What is the relationship between incentives and efficiency – especially in terms of improving security, or performing “mundane” tasks (as against actual coding)?

[Stewart, Daffara, Jones, Laurie, Tuomi, Herbsleb, Fitzgerald, FLOSS5].

Keywords: software engineering analysis techniques, analysed from the perspectives of economics, sociology, software engineering.

Modeling, prediction and empirical research

What are the other types of mechanisms, not based on markets, prices and their invisible hand, that govern production and allocation?

How does open source allocate resources and how can we predict this, based on multiple motivations driving participants?

To what extent does the open source model provide economic efficiency, and how does that relate to the fact that economic motivations are not necessarily important to participants themselves?

Is it possible to learn about or predict the behaviour of open source communities by running simulations of them (agent-based model)?

[Cremer, David, Dalle, Herbsleb, Jones, Tuomi, Crowston, Burt]

Keywords: economics, economic modeling, AI/agent-based simulation, empirical data collection and analysis

Open source and the world: dynamics, learning, social barriers

Many people are studying open source as an example of something else - we need a concerted effort on open source in itself, for itself and what it means to the world - What do processes of open source mean for the global economy, for the software industry, for the information industry, for developing countries?

What is the importance of learning/training in the open source process and how does it work?

How poorly are women represented in open source communities? What can be done to bring more women into the open source world?

Is the transparency in organizational and developmental structure of open source more conducive to a meritocracy that ignores national, economic, socio-political hierarchy?

How do language barriers create divisions within the open source source community?

[Iacono, Metiu, Tuomi]

Keywords: organizational studies, sociology, gender, policy,

Innovation, incentives, organisation and structure

What is the form of organization in open source, and how does it fit with traditional classifications of organization systems?

How does one measure (and predict) innovation in the open source model? What is the link between motivation, incentives and open source innovation?

Can incentives and motivations among participants in projects/communities be correlated to their organizational structure, and structure correlated to quality, efficiency or other measures of "success"?

Is it possible to measure the impact of intra- and inter-project competition and cooperation in open source communities, to correlate with and predict "success"?

[Garzarelli, Kuan, Crowston, FLOSS4]

Keywords: economics, sociology, innovation and organizational studies

Standards and interoperability

What is the impact of open source on standardization and interoperability in software and ICT?

What is the impact of different open source licenses on standards? Do some licences (e.g. GPL) increase the likelihood of retaining interoperability while making improvements to de facto standards?

What is the impact of the ability to freely change development trajectories in open source (code-forking) have on interoperability and on innovation?

What is the impact on interoperability of requiring open source reference implementations for any “open” standard? How does open source, or a reference implementation, affect barriers to entry?

[Egyedi]

Keywords: economics, law, standards policy

Funding open source

What are the current ways that open source development is funded (publicly or privately)? Are there better ways to fund development, and to reward and recognise success and innovation through the recognition of contribution to open source?

Is dual licensing – GPL and proprietary commercial in parallel – useful as a viable business model, or as an effective way of licensing publicly funded software?

What are the ways in which public funds are being (or could/should be) used for the development of open source and what is their impact – on open source, public infrastructure and the software market?

[Cathcart, Schmitz, Strawn]

Keywords: economics, business studies, policy

Incentives & IPR, licensing schemes & policy

What role does IPR play in supporting open source? Copyright law has been used to enforce the GPL and ensure that open source code is not re-appropriated to the exclusive advantage of any one party, as could happen with public domain software.

How do the incentives of IPR protection relate to innovation in open source? Do IPR regimes, especially software patents, act as disincentives for open source, and create entry barriers to software innovation?

What is the relationship between IPR and open standards? What are the implications on standardization and innovation of different licensing regimes, including different choices of open source licence

What IPR issues – e.g. uncertainty over IPR ownership, third-party infringement claims – are faced by potential users/deployers of open source?

[Van Alstyne, Blind, Kahin]

Keywords: economics, innovation, standards, IPR law, policy