journal homepage: www.intl.elsevierhealth.com/journals/ijmi

The State and Profile of Open Source Software Projects in health and medical informatics

Balaji Janamanchi[a], *Evangelos Katsamakas*[b], *Wullianallur Raghupathi*[b,*], *Wei Gao*[b]

[a] *Texas A&M International University, Laredo, Texas, United States*
[b] *Fordham University, New York, NY 10023, United States*

ARTICLE INFO

Article history:
Received 3 July 2008
Received in revised form
23 September 2008
Accepted 19 February 2009

Keywords:
Health and medical informatics
Open source software (OSS)
Sponsorship
License type

ABSTRACT

Purpose: Little has been published about the application profiles and development patterns of open source software (OSS) in health and medical informatics. This study explores these issues with an analysis of health and medical informatics related OSS projects on SourceForge, a large repository of open source projects.
Methodology: A search was conducted on the SourceForge website during the period from May 1 to 15, 2007, to identify health and medical informatics OSS projects. This search resulted in a sample of 174 projects. A Java-based parser was written to extract data for several of the key variables of each project. Several visually descriptive statistics were generated to analyze the profiles of the OSS projects.
Results: Many of the projects have sponsors, implying a growing interest in OSS among organizations. Sponsorship, we discovered, has a significant impact on project success metrics. Nearly two-thirds of the projects have a restrictive license type. Restrictive licensing may indicate tighter control over the development process. Our sample includes a wide range of projects that are at various stages of development (status). Projects targeted towards the advanced end user are primarily focused on bio-informatics, data formats, database and medical science applications.
Conclusion: We conclude that there exists an active and thriving OSS development community that is focusing on health and medical informatics. A wide range of OSS applications are in development, from bio-informatics to hospital information systems. A profile of OSS in health and medical informatics emerges that is distinct and unique to the health care field. Future research can focus on OSS acceptance and diffusion and impact on cost, efficiency and quality of health care.

1. Introduction

With the renewed urgency to adopt health and medical informatics applications, open source approaches [31] are gaining attention in the health care industry [9,12,17,19]. For example, an open source project called Care2X, with four components: hospital information system, practice management, a central data server, and a health exchange protocol is under development in Europe. The software is distributed under the GPL (General Public License). Another effort, *open*EHR, is sponsored by the *open*EHR Foundation (http://www.openEHR.org) and promotes the "development of an open, interoperable health

* *Corresponding author at*: Graduate School of Business, Fordham University, 113 W. 60th Street, New York, NY 10023, United States. Tel.: +1 212 636 7230; fax: +1 212 765 5573.
E-mail addresses: Balaji.janamanchi@gmail.com (B. Janamanchi), Katsamakas@fordham.edu (E. Katsamakas), raghupathi@fordham.edu (W. Raghupathi), Weigao@fordham.edu (W. Gao).
1386-5056/$ – see front matter © 2009 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.ijmedinf.2009.02.006

computing platform, of which a major component is clinically effective and interoperable electronic health records (EHRs)." Additionally, the Journal of Digital Imaging recently published a special issue on open source applications in imaging informatics [19]. Oyri and Murray [20] discuss the potential of open source in nursing informatics. These and similar initiatives have the potential to create low cost tools for physicians and other health care providers.

On a grander scale, government agencies (the predominant payers of health care bills) are looking to open source as a vehicle for health care transformation. Their primary objectives are lowering costs and enabling connectivity [22]. Canada Health InfoWay, funded by federal and provincial grants, started an open source initiative in 2005 to develop software that hospitals and developers could use to ensure the reliable exchange of patient health records [32] among various entities (http://www.infoway-inforoute.ca/en/home/home.aspx). The U.S. government has placed its Vista an integrated hospital software package in the public domain, enabling such projects as OpenVista to provide adopters with open source software [9]. These steps suggest that the open source development approach is a viable means of developing health care applications.

OSS adoption and diffusion in health care have garnered proponents across the spectrum [6,9,12,17]. Indeed, a search of MEDLINE reveals many papers are being published about OSS applications in medicine (primarily bio-informatics). However, while studies have addressed several research issues in open source [10,14,21,29], few have systematically analyzed the nature of OSS in health and medical informatics. It is difficult to gather information about what OSS is actually used for in health and medical informatics, by whom it is used, and how. What features characterize the OSS applications? What is their functionality? What are the profiles of OSS applications in health and medical informatics?

This exploratory study aims to shed additional light on OSS projects in health and medical informatics. In particular we are interested in examining the general profiles and patterns of OSS development. We performed detailed descriptive statistical analysis of a large sample of OSS projects in health and medical informatics publicly available at SourceForge (http://www.sourceforge.net). Specifically, it reveals the nature and magnitude of projects from across the world. We differentiated the projects according to several criteria, including but not limited to 'sponsorship', 'license type', and 'downloads'. Benefits are envisaged (limits, too), yet to be able to comprehensively assess and determine the profile of OSS, it is essential to study a large sample of projects, critically synthesize the data, and from this synthesis gain solid insight into the OSS development process. While OSS is expected to accelerate the diffusion of health and medical informatics applications into health care delivery organizations, achieving this goal will largely depend on functional capabilities, sponsorship status, license type and other features.

Therefore, our research questions were twofold: what types of health and medical informatics applications are being implemented as OSS? And what are the characteristics and descriptors of these applications?

2. Review of OSS

OSS is widely viewed as a way to accelerate the diffusion of health information systems [1] and lower development costs. Supporters suggest OSS prevents vendor lock-in, encourages innovation, and increases application usage [2,20,31]. Open source is an "umbrella term"—at once a noun and adjective—that describes a development method that allows researchers to exchange algorithms and IT professionals to share tools (http://en.wikipedia.org/wiki/Open_source) [19]. Open source software differs from proprietary software in terms of development process [26] and software licensing. Open source is developed in a collaborative and distributed way leveraging the Internet for coordination. The collaboration of individuals and organizations gives rise to a community, often referred to as open source community [13]. Participation in projects is voluntary and project teams are self-organizing [5] and loosely coordinated by a core team of developers who have the final saying on what piece of software developed by developers is incorporated into the code base and the latest release of the open source software [24,28]. Quality is ensured by a peer-review process facilitated by the access to the source code.

All open source software applications are licensed using an open source license. There are many open source licenses, all certified by the open source initiative (OSI—http://opensource.org/). Examples include GPL, used by Linux, the Mozilla Public License, used by the Mozilla foundation for its products (e.g., Firefox) and the Apache license used by Apache foundation to license Apache the dominant web server available. An open source license provides the user with the right to use the software, access and modify its source code, and redistribute the software. There are more and less restrictive licenses [16]. The choice of license for OSS in health and medical informatics is important because it determines the users' rights and affects the incentives of developers to participate in a project, the quality of development and the incentives of users to adopt a software application as we explore in our research.

Since open source refers to software developed using the open source development approach and is licensed under open source terms, it should not be confused with other open technologies and concepts, such as *open systems*, which refers to technologies with open interfaces, or *open standards*, which are specifications that can be used by vendors that wish to do so when they develop their products. In essence, open source incorporates open standards, but it goes beyond the standard specification, since it involves the *actual* development of a software application through collaborative projects that produce an output ready to be tried and deployed by users.

Traditionally, innovation has been the product of teams working within organizational boundaries under management supervision and control [30]. Open source innovation is the result of self-organizing geographically dispersed communities of individuals with shared interests and goals that leverage the Internet, or the result of open collaboration among organizations, dissolving innovation-restricting boundaries. The open source model leverages ideas, creativity and contributions from diverse developers, users and compa-

nies that have an interest in supporting a project. Moreover, open platforms such as Linux encourage innovation because they strengthen the incentives of independent firms or communities to develop more applications for the platform [7,8]. Open source promotes innovation because it makes easier for the users to modify, customize and reuse the source code of open source software to come up with new solutions.

OSS promises several other benefits in health care. It helps users avoid vendor lock-in and dependence on proprietary technologies and it increases choice, flexibility and interoperability through openness. Users can try and experiment with OSS before they are confident they would like to adopt it. Individuals can develop their skills and reputation by participating in open source projects [23]. Organizations can participate in open source projects to develop their internal expertise and influence the direction of the project so that the software addresses their needs better. In terms of costs, organizations save on software licensing fees, and often reduce their hardware expense because they avoid proprietary hardware. However, organizations need to develop their internal expertise in adopting and managing open source. The lack of such expertise increases implementation, maintenance and support costs and it may lead to higher total cost of ownership (TCO) than proprietary solutions. Despite the phenomenal success and high quality of Linux, Apache and other open source solutions, open source does not necessarily mean a superior quality solution. Organizations need to carefully evaluate the specific open source software they are interested in using. They should look at things such as the availability of support and vendors' services and compatible software and hardware, the maturity and activity of the project, the status and quality of the output and the level of adoption by other companies. Our research explores several of these latter crucial factors in the context of health and medical informatics projects.

3. Methodology

SourceForge has grown steadily and become one of the largest web-based repositories of OSS. We searched for projects on SourceForge, using *keywords* pertaining to health, medical and bio-informatics projects and within the period from May 1 to 15, 2007. The search was conducted using SourceForge's 'topics' menu starting with the topic 'scientific engineering' and further selecting the sub-topics of 'bio-informatics' and 'medical sciences applications'. Cross searches using other 'topics' as keywords were also performed to cross check the results. Our initial search identified 607 projects related to health and medical informatics. We then excluded all indirectly related projects (for example, those pertaining to pure medical sciences and medical devices), and 258 projects remained. More investigation of project descriptions' resulted in the exclusion of 79 projects not considered typical of health and medical informatics. An additional 5 projects were rejected because of duplication. The final sample of 174 projects related to mainstream health and medical informatics, such as health record systems, health office support, imaging, clinical decision support and utilities (e.g., interoperability).

We also gathered sponsorship information, if available, on each project from the sponsored project website. A Java program was written to extract the data for the key variables from the web pages of each of the open source projects pertaining to health and medical informatics. We replaced missing values with "0" or the median of the population (which, incidentally, was also "0"). Data was appropriately transformed to facilitate statistical analysis [18,25]. Note that the project web pages on SourceForge provide the data for the variables, that is, the variables already existed, with the exception of the 'sponsorship' variable which was determined using publicly available Web data for the projects.

Descriptions of some of the key variables follow.

3.1. *Sponsorship*

A project is sponsored when it is initiated and/or actively supported by a health care organization or a firm providing health related software. The organization may provide the technology infrastructure for the project; engineers that actively work and contribute to the project may contribute to the project code that was developed internally, and, may actively support user adoption of the project output through public announcements, marketing, etc. Economic theory suggests that 'sponsorship' may increase the likelihood of a project's success [3] because of the provision of such resources as non-volunteer developers, code [4], and/or tools. The introduction of commitment to a process that is otherwise self-organizing as well as the signal given by sponsorship that attracts other developers and users implies that sponsorship improves odds for success [27]. Additional research found that innovative users who contribute to firm-hosted communities are either hobbyists or responsive to firm recognition [11]. 'Sponsorship' is a categorical variable.

3.2. *Project rank*

Project rank measures the 'rank' of a project within the SourceForge database. The variable captures information about traffic, communication and development of each project.[1] Traffic reflects downloads and visits to project page. The value of this variable is to facilitate direct project to project activity comparison. Development reflects commits to Concurrent Versions System (CVS) repository and age of last release. Communication reflects tracker, mailing list and discussion forum activity. We used the twelve months' mean of the project 'rank' for this study.

3.3. *License type*

A software license defines the use, modification and distribution rights assigned to users [27]. The development of General Public License (GPL) by the Free Software Foundation was followed by a large number of other open source licenses, among them "Lesser" General Public License (LGPL), Berkeley Software Distribution (BSD), MIT and the Mozilla Public License.

[1] For more information see: http://sourceforge.net/forum/forum.php?forum_id=465092.

Compared to closed (proprietary) licenses, GPL provides users with the right to use or modify and redistribute software. There are three main types of licenses [15,16], namely, strong copyleft (highly restrictive), such as GPL which permits free use of the software but requires that any modifications be contributed back to the public domain; weak copyleft (restrictive), such as LGPL; and non-copyleft (non-restrictive), such as BSD which provide the language and structure needed to enable users to redistribute software and even sell it commercially. Highly restrictive licenses are less likely to be usurped by an organization that takes the open source code, modifies it, and then commercializes the result. Prior research found that restrictive licenses are used for projects targeted to end users rather than developers, and for projects attractive to consumers, such as games [16]. Earlier research also argued that projects with restrictive licenses should attract more contributors [26], but fewer users because of the restrictions and license uncertainty [26].

3.4. *Downloads*

Downloads indicate how popular a project is with users. This variable measures the number of 'downloads' of a project's code from its SourceForge page. Download is also a proxy of usage. Note that 'downloads' are also captured in the project 'rank' variable. Here too we used the twelve months' mean of downloads.

3.5. *Development status*

This variable captures the software development status (e.g., pre-planning, alpha, beta, etc.). The 'development status' characterizes the lifecycle of software development and would impact the success metrics of the project. Data captured as a means to track project team work during software development includes source code commits to the project's CVS repository or sub-version repository, age of the last file release made via the file release system, and how recently the administrators of a project have logged-in to the SourceForge site.

3.6. *Technological environment*

We also examined various technological environment variables such as programming language (PL), operating system (OS) and database (DB) environment as well as their relationship to the project.

We selected 'sponsorship', 'rank', 'license type', 'downloads', 'development status', and 'technological environment' variables among others, because of their availability and their significance. They provide a comprehensive profile of the most important aspects of a project.

The extracted data was input into an Excel worksheet and imported into SPSS statistical software for analysis. We then performed several descriptive analyses of the data, resulting in various types of visual diagrams and charts that were then interpreted. These are discussed below.

4. Results and analysis

Key insights from the analysis of the data for the 174 projects are discussed next. Several observations stand out from the descriptive statistics. Consider, for example, the relationship between license restrictiveness and project sponsorship in Fig. 1.

Approximately 110 projects (60%) did not have a primary sponsor such as an organization. However, 64 projects (40%) did have a sponsor indicating the growing interest in OSS among health care organizations. As OSS in health care evolves, continued 'sponsorship' by stakeholders (users, developers and managers) may lead to wider adoption of OSS in health and medical informatics. Also to be noted is the proportion of the restrictive vis-a-vis the non-restrictive license patterns. Nearly two-thirds of projects were of the restrictive license type. (Note that highly restrictive is plotted along with restrictive in Fig. 1). As discussed previously, the restrictive

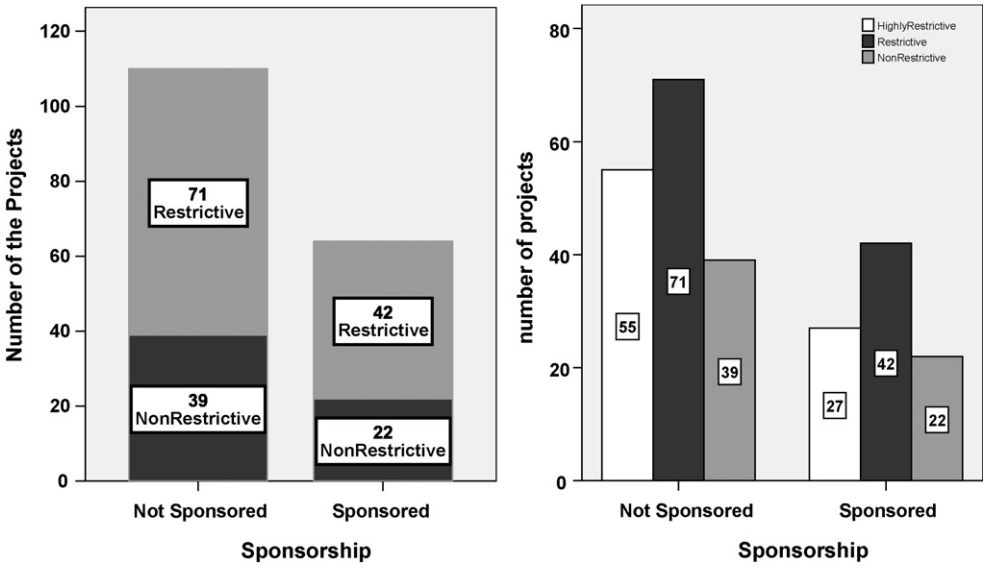

Fig. 1 – Project license restrictiveness by sponsorship.

type such as GPL is commonly used because the community wants the developers and users to contribute back to the project community.

A Pareto analysis of the 'development status' data revealed that only 4 of the 174 projects were in a mature state while one project was inactive. On the other hand, there is a good count of projects in the other development modes implying a thriving community that worked actively on various projects. Note though the 'development status' data was unspecified for approximately 20% of the projects. Another Pareto analysis of the 'operating system' with regard to the 'technological environment' used showed 25% of projects did not disclose a primary operating system. However, consistent with the objective of OSS, several projects ($N = 56$) were described as 'operating system' independent, and atleast eleven projects were described as using open source operating systems such as Linux. Several projects ($N = 17$) indicated the use of Windows implying a vendor-driven approach to OSS development. The repeated use of a specific 'operating system' over a period signals the likelihood of better maintenance and troubleshooting support for the project. It also provides stability over the various versions and releases of the project.

The Pareto analysis of the 'programming language' used showed that 20% of the projects did not indicate the use of a specific programming language for implementation. Among the projects that indicated a language, Java was the clear leader with over 33% of projects ($N = 58$) using it. A combination approach using multiple languages was the second most popular alternative. This seemed to imply a growing success of the .NET platform usage among OSS developers, particularly in providing a framework for the integration of modules developed in multiple languages. Although Java has been in use since 1995, the more recent .NET platform supports the Windows operating system. The combination of Windows and .NET offered a robust alternative environment for OSS development.

Next, a Pareto analysis of the project 'downloads' was conducted. This variable is considered as a metric of OSS success. The 'downloads' of projects with multiple programming languages was very high, with a download rate three times (>75%) that of the download rate of all other programming languages combined (<25%). This finding confirmed the earlier observation that developers used a combination of multiple languages. It is likely that projects were being developed to meet the requirements of the users of the various languages; alternatively, there may be an inherent objective to mitigate risks of development in a particular language only.

The relationship between 'sponsorship' and 'downloads' of a project vis-a-vis the 'database environment' (code) used was next examined. The ability to attract developers to an open source project is important and is a crucial variable of OSS success [5,19]. Identifying the key variables that influence that ability would help organizations attract good developers. Our analysis revealed that the 'downloads' of sponsored projects created in mixed database environments were very high when compared to those that were not sponsored. The high volume of 'downloads' displayed in Fig. 2A, in all likelihood suggests users' tendency to gravitate towards sponsored projects. Also, the number of 'downloads' for sponsored projects that used a mixed database environment of open source and non-open source (or other) was higher compared to the database environment in either open source or non-open source alone. This finding indicated a conservative approach to the underlying database environment with the implication that although the project itself is open source, its database is a combination of both. Overall though we interpret these findings with caution since a high number of 'downloads' may indeed be indicative of numerous developers downloading versions of the project during the development process.

Fig. 2B shows the 'downloads per developer' variable instead of the 'downloads'. The data indicated that the 'down-

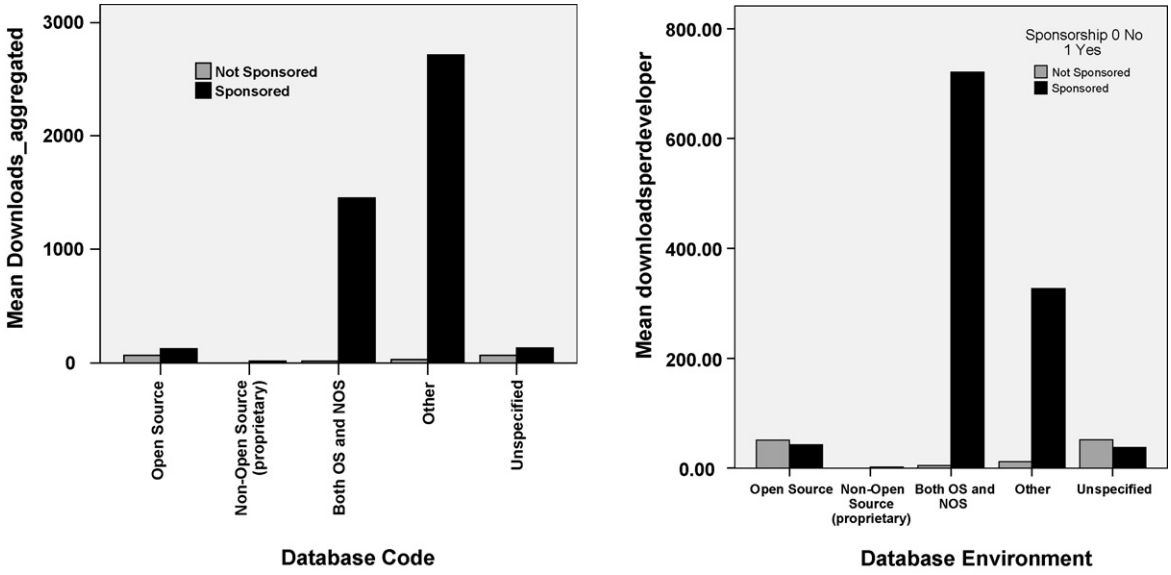

Fig. 2 – (A) Downloads under projects sponsored and non-sponsored vis-a-vis database environment. (B) "Downloads per developer" under projects sponsored and non-sponsored vis-à-vis database environment.

Table 1 – Application projects and intended audience.

Topic	Intended audience								
	Advanced end user	Industry, scientific organizations and research	Customer service	Developers	Education	End users	Government and non-profit	Unspecified	Total
	Count and % within intended audience code								
Artificial intelligence		1 2.00%		1 3.22%			1 25.00%		3 2.85%
Bio-informatics	1 33.33%	11 22.00%		3 9.67%				1 100.00%	16 15.23%
Content		1 2.00%	1 100.00%	6 19.35%		1 9.09%			9 8.57%
Customer relationship management/ scheduling/finance and resource booking		2 4.00%				2 18.18%			4 3.81%
Database	1 33.33%	4 8.00%		6 19.35%	1 25.00%				12 11.43%
Enterprise		6 12.00%		1 3.22%		1 9.09%			8 7.62%
Front-end applications		3 6.00%		1 3.22%			1 25.00%		5 4.76%
Information analysis					1 25.00%	1 9.09%			2 1.90%
Medical science applications	1 33.33%	17 34.00%		10 32.26%	2 50.00%	3 27.27%	2 50.00%		35 33.33%
Office/business		5 10.00%		2 6.44%		3 27.27%			10 9.52%
Project management				1 3.22%					1 95.00%
Total	3 100.00%	50 100.00%	1 100.00%	31 100.00%	4 100.00%	11 100.00%	4 100.00%	1 100.00%	105 100.00%

Table 2 – Utility tools' projects and intended audience.

Topic	Intended audience								
	Advanced end user	Industry, scientific organizations and research	Customer service	Developers	Education	End users	Government and non-profit	Unspecified	Total
	Count and % within intended audience code								
communications		2 18.18%		4 33.33%		2 66.67%			8 28.57%
Conferencing		1 9.09%							1 3.57%
Data formats	1 100.00%	1 9.09%		1 8.33%					3 10.71%
Graphics				1 8.33%					1 3.57%
Interface				1 8.33%					1 3.57%
Internet		1 9.09%							1 3.57%
Security		1 9.09%							1 3.57%
Simulation		1 9.09%							1 3.57%
Site management				1 8.33%					1 3.57%
Software development		1 9.09%		1 8.33%					2
Testing				1 8.33%					1 3.57%
XML				2 16.67%					2 7.14%
Other/non-listed		2 18.18%				1 33.33%	1 100.00%		4 14.28%
Desktop environment		1 9.09%							1 3.57%
Total	1 100.00%	11 100.00%		12 100.00%		3 100.00%	1 100.00%		28 100.00%

Table 3 – Programming language versus database environment.

Programming language	Database environment					
	Open source Count and % within database environment	Non-open source (proprietary)	Both open source and non-open source	Other	Unspecified	Total
Multiple (.NET, JavaScript, C++, etc)	14 28.60%	1 25.00%	7 63.60%	3 33.30%	23 22.80%	48 27.60%
C++, C, MUMPS, VB.NET		1 25.00%		1 11.10%	8 8.00%	10 5.70%
Java	25 51.00%	2 50.00%	4 36.40%	3 33.30%	24 23.80%	58 33.30%
Perl, Python, PHP, TcL, XSL	9 18.40%			1 11.10%	12 11.90%	22 12.60%
Unspecified	1 2.00%			1 11.10%	34 33.70%	36 20.70%
Total	49	4	11	9	101	174

loads per developer' was slightly higher for non-sponsored projects than for sponsored projects in two categories, namely "open source" and "unspecified" database environment. However, in two other categories, namely, "others" and "both OS and Non-OS" database environments—the 'downloads per developer' of sponsored project was much higher than the non-sponsored projects.

Projects in our sample were then analyzed by the project type ('project topic') and the target audience ('intended audience'). For this purpose, each project type was categorized either as an 'application' or a utility 'tool'.

Table 1 shows the cross tabulation of 'intended audience' and 'application project', and reveals some very interesting insights. As noticed, 105 of the 174 projects under study belonged to the 'application' category. Fifty projects out of these 105 targeted 'industry, scientific organizations and research' audience imply a growing interest in open source community in industry. The next major group of 'intended audience' for the application projects is 'developers' indicating the interdependence of developers in OSS. Further, projects targeting the advanced end user were primarily limited to three topics, namely, 'bio-informatics', 'databases', and 'medical science applications'. Interestingly, the number of projects divides equally among these three topics. Similarly, the projects whose 'intended audience' was 'educational institutions' were concentrated in the three topics of

Table 4 – Language vis-a-vis database environment.

Language (translation)	Database environment					
	Open source Count and % within database environment	Non-open source (proprietary)	Both open source and non-open source	Other	Unspecified	Total
English only	16 32.7%	3 75.0%	2 18.2%	3 33.3%	27 26.7%	51 29.3%
Non-english	5 10.2%		1 9.1%	1 11.1%	4 4.0%	11 6.3%
Multiple (english + other)	4 8.2%		5 45.5%	2 22.2%	12 11.9%	23 13.2%
Multiple (non-english)	2 4.1%			1 11.1%		3 1.7%
Unspecified	22 44.9%	1 25.0%	3 27.3%	2 22.2%	58 57.4%	86 49.4%
Total	49 100.0%	4 100.0%	11 100.0%	9 100.0%	101 100.0%	174 100.0%

Table 5 – License versus operating system cross tabulation.

License	operating system							
	Windows (proprietary)	OS independent	Mixed (prop +OS)	Portable OS	Open source (e.g., Linux)	OSX and OS independent + Win CE	Unspecified	Total
	Count and % within operating system category							
General public license (GPL)	11 64.70%	34 60.70%	17 54.80%	5 55.60%	8 72.70%	2 100.00%	5 10.40%	82 47.10%
Apache		3 5.40%	1 3.20%				3 6.30%	7 4.00%
LGPL "Lesser GPL)	3 17.60%	6 10.70%	5 16.10%	2 22.20%	1 9.10%		4 8.30%	21 12.10%
Berkeley software distribution (BSD)	1 5.90%	5 8.90%	1 3.20%				2 4.20%	9 5.20%
AFL, CFL, Eclipse, IBM, Mozilla, proprietary, public domain, Python, SISSL, CDDl[a]	1 5.90%	5 9.00%	4 13.00%	2 22.20%	2 18.20%		4 8.40%	19 10.90%
Multiple-public	1 5.90%		2 6.50%					3 1.70%
GPL or LGPL		1 1.80%	1 3.20%					2 1.10%
Unspecified		1 1.80%					30 62.50%	31 17.80%
Total	17 100.00%	56 100.00%	31 100.00%	9 100.00%	11 100.00%	2 100.00%	48 100.00%	174 100.00%

[a] Academic Free License (AFL), CFL, Eclipse (open source development tool–Eclipse Foundation), IBM (allows IBM public license), Mozilla, proprietary license, public domain, Python, SUN industry standards source license (SISSL), and common development and distribution license (CDDL).

Table 6 – Profiles of top fifteen projects based on # of downloads (source: SourceForge.net).

No	Project description	Topic	Downloads aggregated	Development status	Sponsored	License
1	**OsiriX** – 3D DICOM medical viewer for MacOS X. Complete DICOM viewer with complete DICOM network support! DICOM DICOM.	Bio-Informatics, medical science applications, Visualization.	10234	6 – Mature	Yes	GNU General Public License (GPL)
2	**HOSxP** is client/server hospital information system using in 150 hospitals in Thailand. HOSxP has many modules which keep data of patient image, symptoms, physical condition, investigation, diagnosis, treatment including procedure/medication, etc.	Enterprise, medical science applications.	3509	5 – production/stable	Yes	GNU General Public License (GPL)
3	**Integrated hospital information system**. PHP, MySQL, PostgreSQL. Surgery, nursing, outpatient, wards, labs, pharmacy, security, admission, schedulers, repair, communication and more. Multilanguage, WYSIWYG forms, user config, embedded work bots. Modular and scalable.	Bio-Informatics, human machine interfaces, medical science applications.	2633	4 – Beta, 5 – production/stable, 6–mature	Yes	GNU General Public License (GPL), GNU library or lesser GENERAL Public License (LGPL)
4	**The DICOM validation tool (DVT)** is a software utility and a set of .NET components that will assist in testing the medical/healthcare protocol DICOM. DVT provides you the methods to transfer and validate DICOM objects.	Testing	1325	4 – Beta	Yes	GNU library or Lesser General Public License (LGPL)

5	**GPL-licensed electronic medical record and practice management system** for medical providers that runs in any web browser in multiple languages. It provides an XML–RPC backend and multiple import and export formats, as well as reporting and other features	Resource booking, bio-informatics	1113	5 – Production/stable	Yes	GNU General Public License (GPL)
6	**Amide – a Medical image data examiner** Amide is a tool for viewing, registering, and analyzing anatomical and functional volumetric medical imaging data sets.	Medical science applications, visualization.	1092	4 – Beta, 5 – production/stable	No	GNU General Public License (GPL)
7	**Patient runner** is a mental health medical records system. Windows client software written in Delphi 7 connects to a MySQL database that contains progress notes, diagnoses, medications, and scales. User customized templates make note creation easy.	Front-ends, medical science applications.	1025	4 – Beta	No	GNU General Public License (GPL)
8	**Mirth** is an open source cross platform HL7 interface engine that enables bi-directional sending of HL7 messages between systems and applications over multiple transports.	Data formats, interface engine/protocol translator, medical science applications.	965	5 – Production/stable	Yes	Mozilla Public License 1.1 (MPL 1.1)
9	**A medical image conversion utility and library**; hereby hoping to lower atleast one barrier in medical research projects.	Graphics conversion, medical science applications.	893	5 – Production/stable	No	GNU General Public License (GPL), GNU library or lesser General Public License (LGPL)

Table 6 (*Continued*)

No	Project description	Topic	Downloads aggregated	Development status	Sponsored	License
10	**BIOSIG** is an open source software library for biomedical signal processing. \r\nLibrary works well with Octave and Matlab. Special emphasis is put on EEG/MEG/ECoG, but also on other types of biosignals like ECG, EMG, etc. are supported.	Data formats, human machine interfaces, medical science applications.	679	4 – Beta, 5 – Production/stable	No	GNU General Public License (GPL)
11	**OpenVista** is the open source version of Vista, which is an enterprise grade health care information system developed by the U.S. Department of Veterans Affairs (VA) and deployed at nearly 1500 facilities worldwide.	Database engines/servers, bio-informatics, medical science applications.	617	4 – Beta, 5 – production/stable	Yes	GNU General Public License (GPL), other/proprietary license, public domain
12	**Full featured free PACS based** on ctn, dcmtk and Mysql, with remote administration using Apache mod Perl and imaging processing capabilities using ImageMagick, Grevera\'s dcm2pgm DICOM converter and AFNI, available in Debian packaging format for i386	Medical science applications.	593	5 – Production/stable	No	GNU General Public License (GPL)
13	Our mission is furthering the cause of affordable healthcare information technology worldwide by advocating, championing, and employing the open source paradigm to expand the use and collaborative improvement of the **Vista** electronic health record.	Other/non-listed topic	523	5 – Production/stable	Yes	GNU General Public License (GPL), public domain

14	**ezDICOM** is a medical viewer for MRI, CT and ultrasound images. It can read images from Analyze, DICOM, GE Genesis, Interfile, Siemens Magnetom, Siemens Somatom and NEMA formats. It also includes tools for converting medical images from proprietary format	Medical science applications.	397	6 – Mature	Yes	BSD license
15	**Hospital OS** is a hospital information system for small-sized hospitals (100 beds or less, 200 out-patients per day or less). Hospital OS was developed for the healthcare system of Thailand targeting its small community hospitals nation-wide.	Front-ends	391	5 – Production/stable	Yes	GNU General Public License (GPL)

'database', 'information analysis', and 'medical science applications'. Projects targeting the intended audience groups of 'industry', 'scientific organizations', and 'research' ($N = 50$) and 'developers' ($N = 31$) appeared to be widespread across all the topics followed by projects whose 'intended audience' was 'end users' ($N = 11$).

Table 2 shows the cross tabulation of 'intended audience' and 'utility tools' projects, and provides interesting insights. Projects targeting the 'intended audience' groups of 'industry, scientific organizations and research' ($N = 11$) and 'developers' ($N = 12$) appeared to be fairly well represented across all the topics. This observation clearly indicates that the target audiences of these two groups appear to be the driving force behind the growth of OSS.

In terms of the number of projects by topic, a large percentage of projects (23.6%) did not specify the topic ($N = 41$). Of these, a large number of projects had an unspecified 'intended audience' ($N = 28$). This was followed 'by industry, scientific organizations and research' ($N = 6$) and 'developers' ($N = 5$), asserting the dominance of these two groups of 'intended audience'.

Another dimension of the project profile correlates the 'programming language' to the 'database environment' (see Table 3). Whereas 101 projects did not provide the details of their database environment, 60 of the 73 that did, employed an open source database environment (49 open source only, 11 both), while 15 projects employed a non-open source database environment (4 non-open source, 11 both). Nine projects employed other database environments. It is conceivable that the projects that did not specify the database environment are indeed using an open source environment. With regard to programming languages, Java emerged as the preferred language with 58 (33%) of the 174 projects.

From a global perspective of open source in health and medical informatics, the project 'language' (spoken language) was examined. While a majority of projects were written in English, several are being developed in other languages, such as French, German, and Spanish. This trend indicated the global pervasiveness of OSS in health and medical informatics. However, projects written in a language other than English were relatively few. Furthermore, the projects written simultaneously in multiple languages other than English, regardless of the database environment, were even fewer. Table 4 below shows the cross tabulation of 'language' (project translation) correlated to the 'database environment'.

Next, the important variable of project 'license type' was evaluated. Table 5 shows the cross tabulation of 'license type' correlated to 'operating system'. The GPL license type tallied the largest number with 47% ($N = 82$) of the projects utilizing it. This license is used widely because development groups generally would like developers and users to contribute back to the open source community with the modifications and upgrades. This type of restrictive license gives key developers better control over the project.

We also extracted data to identify the profiles of the top 15 projects sorted by the number of 'downloads'. Table 6 describes the profiles of these 15 projects. We examined

the monthly average of 'downloads' over the twelve months immediately preceding data collection. The number of times a project is downloaded may indicate the popularity or success of the project. These averages provide a unique perspective on understanding OSS projects in health and medical informatics. The table shows that OSS applications are being developed in a wide range of health and medical informatics domains including bio-informatics, digital imaging and medical record systems.

We note that the top ranked project (in terms of downloads), with a monthly average of 10,234 downloads in the immediate past 12 months, was downloaded nearly three times as often as the second ranked project which was downloaded an average of 3509 times monthly. While 10 of the top 15 projects were sponsored, 11 were in the 'production stable' 'development status' mode. The top project was in a "mature" stage of project development. Thirteen of the top 15 projects specified a GNU General Public License. Lastly, Open Vista project, a widely publicized OSS, was ranked 11th with respect to downloads over the past twelve months.

5. Scope and limitations

With regard to scope and limitations of this research, data for this exploratory study was gathered from SourceForge on a specific date. Therefore, this study is a snapshot in time. Two of the variables 'downloads' and 'activity percentile', for example, used twelve-month averages that create a picture at the end of a twelve-month period. Projects are added and deleted at these web sites frequently, precluding long-term conclusions.

We also recognize the fact that not all open source projects are registered with SourceForge. Many are registered at other websites, including Freshmeat (http://freshmeat.net/). Our findings, therefore, may underrepresent the total number of health and medical informatics OSS applications. Also, many high profile projects, maintain their own developer sites (e.g., Apache, Perl, Sendmail, and Linux). But other large projects, including OpenVista, *are* listed on SourceForge. Furthermore, projects may have outdated or erroneous data in their listings in addition to the missing data problem. An additional assumption we made is that the projects on SourceForge are the representatives of the open source movement, in large part because of its popularity and the large number of projects and developers registered there. Not all of the available variables were analyzed in this study. To keep the research within scope, we chose the ones that were most likely to contribute to the insight into OSS projects in health and medical informatics. The data and results appear to also reflect the developers' perspective more than the users' perspective because the primary objective of SourceForge is to provide an infrastructure to facilitate the development process.

6. Conclusions

This exploratory study examined the breadth and diversity of OSS applications in health and medical informatics. Several profiles and patterns of OSS development emerge. Many projects are sponsored, indicating a growing interest in OSS by health organizations. Project sponsorship by organizations signals a commitment to the project, attracts developers and other users, and provides crucial resources and support for the success of a project. Healthcare organizations that have an interest in the output of a project should consider becoming active project sponsors, rather than mere users who stay on the sidelines of the development process. Such organizations would stand to benefit from the continuous improvement of the open source code. They can also acquire skills that help them in open source implementation and management. Additionally, they can influence design decisions to include functionality and features that they need.

Nearly two-thirds of the projects have a restrictive license scheme. The type of license is an important decision of the core development team of an open source project. More restrictive licenses make it difficult for a for-profit entity to "hijack" an open source project and commercialize a "fork" (version) of the project code. Our data suggests that this is valued by healthcare organizations, since it provides a guarantee that the output will remain open source, with all the associated user benefits. Moreover, because of that protection, developers have stronger incentives to contribute to projects with highly restrictive licenses with the expectation that their voluntary contribution will not be exploited unscrupulously in the future.

A thriving OSS development community is indicated by the spread of a multitude of projects at various stages of development (status). Projects targeted towards the advanced end user primarily focused on Bio-informatics, data formats, database and medical science applications. The vast majority of applications are in English.

Our research study was conceived to identify the average profile of OSS development. While OSS adoption has potential, more research is needed to understand the nuances of OSS development and the cultural and technological barriers to adoption, particularly in the typical U.S. health care practice. The impact of introducing OSS into complex workflow is not well understood and cannot be automatically equated to improved clinical practice, quality of care or lower cost. Additional research is also needed to study the impact on quality and cost over a longer time frame. Further progress towards adoption of OSS in health care would require the embrace of standards for interoperability. It was impossible to gather this type of information for this study. Also, the compliance status with laws such as HIPAA and standards such as HL 7 was not available. Further research may also be needed to identify the variables that predict success of OSS projects. Another study may survey user organizations to get the users' perspective. So far, despite the short history of OSS in health care, a significant body of knowledge is developing and is available to developers and users for assessing the viability of OSS. Additional studies will help to build a deep corpus of knowledge related to OSS, including but not limited to impact (on clinical practice), cost effectiveness, acceptance and compliance with standards.

Summary points

What was known before the study:

- The potential for OSS in health and medical informatics has been discussed in the informatics literature.
- Numerous OSS projects, primarily in bio-informatics, have been reported in the literature.
- A large number of OSS projects in health and medical informatics are reported as under development at SourceForge website.

What this study has added to the knowledge base:

- This is perhaps the first aggregate level study that examines a large number of OSS projects in health and medical informatics.
- Our findings suggest that many projects are sponsored, indicating a growing interest in OSS by health organizations. Sponsorship may play a key role in ultimate project success.
- Nearly two-thirds of the projects have a restrictive license scheme.
- A thriving OSS development community is indicated by the spread of a multitude of projects at various stages of development (status).
- Projects targeted towards the advanced end user primarily focused on bio-informatics, data formats, database and medical science applications. The vast majority of applications are also in English.

REFERENCES

[1] D.J. Brailer, Economic perspectives on health information technology, Business Economics 40 (3) (2005) 6–14.
[2] D. Carnall, Medical software's free future, British Medical Journal 321 (October 21, 2000) 976.
[3] K. Crowston, J. Howison, Assessing the health of open source communities, Computer 39 (5) (2006) 89–91.
[4] K. Crowston, J. Howison, H. Annabi, Information systems success in free and open source software development: theory and measures, Software Process Improvement and Practice 11 (2) (2006) 123–148.
[5] K. Crowston, Q. Li, K.N. Wei, U.Y. Eseryel, J. Howison, Self-organization of teams for free/libre open source software development, Information and Software Technology 49 (6) (2007) 564–575.
[6] J. Dedrick, J. West, System Sciences. in: Proceedings of the 37th Annual Hawaii International Conference on 5–8 January, 2004, 10 pp., doi:10.1109/HICSS.2004.1265633.
[7] N. Economides, E. Katsamakas, Linux vs. Windows: a comparison of application and platform innovation incentives, in: J. Bitzer, P. Schroder (Eds.), The Economics of Open Source Software Development, Elsevier Press, 2006, pp. 207–218.
[8] N. Economides, E. Katsamakas, Two-sided competition of proprietary vs. open source technology platforms and the implications for the software industry, Management Science 52 (7) (2006) 1057–1071.
[9] M. Goulde, E. Brown, "Open source software: a primer for health care leaders", Forrester Consulting, http://www.chcf.org/documents/ihealth/OpenSourcePrimer.pdf, 2006.
[10] C.L. Huntley, Organizational learning in open-source software projects: an analysis of debugging data, IEEE Transactions on Engineering Management 50 (4) (2003) 485–493.
[11] L.B. Jeppesen, L. Frederiksen, Why do users contribute to firm-hosted user communities? The case of computer-controlled music instruments, Organization Science 17 (1) (Jan-Feb 2006) 45–63.
[12] G.S. Kantor, W.D. Wilson, A. Midgley, Open-source software and the primary care EMR, Journal of the American Medical Informatics Association 10 (6) (2003) 616.
[13] S. Krishnamurthy, "Cave or community? an empirical examination of 100 mature open source projects", First Monday [Online], vol. 7 (6) (3 June 2002).
[14] K.R. Lakhani, R.G. Wolf, Why hackers do what they do: understanding motivation and effort in free/open source software projects, in: J. et al. Feller (Ed.), Perspectives on Free and Open Source Software, MIT Press, 2005, pp. 3–22.
[15] J. Lerner, J. Tirole, Some simple economics of open source, Journal of Industrial Economics 50 (2) (2002) 197–234.
[16] J. Lerner, J. Tirole, The scope of open source licensing, Journal of Law Economics & Organization 21 (1) (2005) 20.
[17] C.J. McDonald, G. Schadow, M. Barnes, P. Dexter, J.M. Overhage, B. Mamlin, J.M. McCoy, Open source software in medical informatics-why, how and what, International Journal of Medical Informatics 69 (2003) 175–184.
[18] C.A. Mertler, R.A. Vannatta, Advanced and Multivariate Statistical Methods, second ed., Pyrczak Publishing, Los Angeles, California, 2002, p. 313.
[19] P. Nagy, Open source in imaging informatics, Journal of Digital Imaging 20 (Suppl. 1) (2007) 1–10.
[20] K. Oyri, P.J. Murray, Osni.info-Using free/libre/open source software to build a virtual international community for open source nursing informatics, International Journal of Medical Informatics vol. 74 (2005) 937–945.
[21] J.W. Paulson, G. Succi, A. Eberlein, An empirical study of open-source and closed-source software products, IEEE Transactions on Software Engineering 30 (4) (2004) 246–256.
[22] W. Raghupathi, W. Gao, An eclipse-based development approach to health information technology, International Journal of Electronic Healthcare 3 (4) (2007) 433–452.
[23] D. Riehle, The economic motivation of open source software: stakeholder perspectives, IEEE Computer (April 2007) 25–32.
[24] J.A. Roberts, I.H. Hann, S.A. Slaughter, Understanding the motivations, participation and performance of open source software developers: a longitudinal study of the Apache projects, Management Science 52 (7) (2006) 984–999.
[25] J. Stevens, Applied Multivariate Statistics for the Social Sciences, second ed., Lawrence Erlbaum Associates, Hillsdale, NJ, 1992.
[26] K. Stewart, S. Gosain, The moderating role of development stage in free/open source software project performance, Software Process Improvement and Practice 11 (2) (2006) 177–191.
[27] K.J. Stewart, A.P. Ammeter, L.M. Maruping, Impacts of license choice and organizational sponsorship on user interest and development activity in open source software projects, Information Systems Research 17 (2) (2006) 126–144.
[28] K.J. Stewart, S. Gosain, The impact of ideology on effectiveness in open source software development teams, MIS Quarterly 30 (2) (2006) 291–314.
[29] K.J. Stewart, D.P. Darcy, S.L. Daniel, Observations on patterns of development in open source software projects, in: Proceedings of the Fifth Workshop on Open Source Software Engineering (5-WOSSE), St. Louis, MO, USA, May 17, 2005.

[30] G. von Krogh, S. Spaeth, K.R. Lakhani, Community, joining, and specialization in open source software innovation: a case study, Research Policy 32 (7) (2003) 1217–1241.
[31] G. von Krogh, E. von Hippel, The promise of research on open source software, Management Science 52 (7) (2006) 975–983.
[32] W.P. Welch, D. Bazarko, K. Ritten, Y. Burgess, R. Harmon, L.G. Sandy, Electronic health records in four community physician practices: impact on quality and cost of care, Journal of the American Medical Informatics Association 14 (3) (2007) 32–328.