Open Source Software for Multicenter Image Management: ImTK Consortium

Seong K. Mun, Member IEEE, Mary Lou Ingeholm Walid Tohme, and Kevin Cleary, Member IEEE

***Abstract*- Development of software through open source approach has been gaining popularity in the information technology community. Recently a workshop was organized to review the role of open source can take in healthcare informatics. Successful open source efforts require a number of key elements such as free licensing, presence of active participants and engineering discipline for generate robust high quality software with necessary documentation. Open source approach however requires different business model as the code itself is made available freely. In healthcare the role that FDA plays in software engineering has to be addressed. The IT capabilities in healthcare is maturing rapidly for many types of patient care settings, there is a significant gaps in ability to share biomedical data in multi-center applications and research. A new consortium is being launched to promote the development of software tools for information and image exchanges in multicenter environment. Open source coupled with open architecture is seen as a critical component in open science. Furthermore US government agencies are promoting the open source approach as a means to transfer research software technology to greater commercial applications.**

1.0 INTRODUCTION

A recent conference on Multicenter Image Management (MCIM) held in Las Vegas in Feb 2006 explored open source strategies to support flexible access to various biomedical data for research. The group was nearly unanimous in recognizing the technology gaps between commercial information systems that focus on efficient clinical operations within a single institution and the urgent needs of the research community that require flexible access to multimedia data from multiple institutions that use variety of vendor products. It was further noted that the currents gaps are not likely be addressed by the commercial community any time soon as the market for such capability in current biomedical environment is very limited. The conference participants concurred that open source, open standards, and open architecture can be efficient methods of supporting open science and improved interoperability. Examples of robust open source projects and software methodologies were presented and there was broad agreement that adequate rigor must be incorporated into an open source process that would meet the highest standards of software quality. Several examples of successful business models for maintaining the development effort were presented and the importance of long term sustainability beyond initial government funding was also discussed. An open source approach is also seen as a new model for collaboration involving academia, industry and government. The conference concluded that an open source effort by the research community to develop robust, freely available tools that meet the information management needs of basic, clinical and translational research is essential to mend the gap between the research and clinical communities.

This work was supported by the Project Sentinel of National Library of Medicine, the National Institutes of Medicine, USA, and Telemedicine and Advance Technology Center of the US Army, Ft. Detrick, MD.

S.K. Mun, M. L. Ingeholm, W. Tohme and K. Cleary are with the ISIS Center, Georgetown University, Washington, DC 20057 USA (Corresponding author phone: 202-784-3483; fax: 202-784-3479; e-mail: mun@isis.georgetown.edu)

2.0 PROBLEMS TO BE ADDRESSED

The information requirements for a biomedical research environment are markedly different from the clinical environment. Commercial medical information and imaging systems are designed to support efficient clinical operations within a single organization whereas researchers need to be able to integrate research data with clinical data often residing in multiple distributed information repositories. The information management components for research must be able to handle more complex queries, data mining and a broad spectrum of data types beyond routine clinical data [1]. This gap between clinical and research requirements prevents the efficient exchange, sharing, management, and analysis of multimedia medical information such as clinical information, images, and bioinformatics data as well as proteomics data sets, significantly impacting the capability to translate research into clinical outcomes. Thus, while hospitals and research communities are collecting unprecedented amounts of clinical data, and research data the ability to data mine these rich collections to support research is limited within an institution and is essentially nonexistent across institutions. Bioinformatics and proteomics data have become increasingly important in clinical research but there are not efficient ways to incorporate these data with clinical information. Multi-center clinical trials are common activities yet many of the trials are still managed manually and cannot optimize the value that a multi-center model represents. Each of these issues is

a direct result of the inability to exchange multimedia clinical data and research information across different organizations and functional environments and impedes the ultimate goal of improving patient outcomes.

The current situation calls for innovative solutions that engage a broad community of users. Using an open source and open architecture framework would allow rapid implementation of scalable and robust software development in a cost effective manner by a community of users from academia, industry and government.

3.0 An Open solution: open source software development

Adopting an approach that includes open source software and an open architecture is essential to a solution that can bridge the information management gap between functional environments within an institution and across multiple institutions. An open source framework supports rapid software development while open architecture encourages interoperability across different environments. An open methodology for this effort will encourage development and implementation of software applications that can expedite translational research in a multi-center setting.

Open source software (OSS) development has become a cultural as well as an economic phenomenon within the information technology (IT) community. It efficiently harnesses global skills and resources, resulting in accelerated research and development. Open source initiatives encourage high-level technical communication, provide conventions for interoperable software development, establish a baseline for improvement, open the field to "beginners", and create common ground for product development [1]. There is also a growing body of evidence that OSS produces more robust code with fewer bugs. From a government perspective, the demand for open access for taxpayer funded projects and the need for quality and performance in mission critical applications is leading to an increased demand for open source solutions. Within the National Institutes of Health (NIH) specifically, the requirements for accelerating discovery include promoting team science, lowering barriers and entry costs, enabling (enforcing) repeatable results and eliminating oversight through transparency [1]. An OSS tactic reduces redundancy of research, enforces good research practices, and enables sharing of ideas. Overall, the OSS concept has the greatest potential for success in developing tools that can bridge the clinical information management gap between the research and clinical communities.

4.0 An Open solution in Biomedical Applications

There has been remarkable penetration of OSS in medical imaging research software. The Visualization Toolkit (VTK) and the Insight Toolkit (ITK), supported by the National Library of Medicine of the NIH represent two large, mature, and globally utilized open source toolkits that provide state-of-the-art imaging architectures and algorithms to application developers. VTK provides a wide range of advanced multi-dimensional visualization algorithms including volumetric reformat, volume rendering, and geometric surface rendering algorithms. ITK provides advanced image processing algorithms, with a particular emphasis on medical image segmentation and image registration algorithms. VTK and ITK were developed with a strong emphasis on advanced computing technologies and software quality. The C++ software architecture of these toolkits has evolved over the years to support a wide range of advanced algorithms and computing technologies including parallel computing. In addition, several computational tools and utilities have been developed that facilitate the global development of a high quality toolkit including a cross-platform build tool called CMake and a software quality dashboard called DART. These open source imaging toolkits, and their supporting tools and utilities, represent a large and growing resource for future open source technology solutions. [1]

The Image-Guided Surgery Toolkit (IGSTK), another project supported by National Institute of Biomedical Imaging and Bioengineering at the NIH, is an open source, cross platform, software toolkit. IGSTK integrates the basic components needed in surgical guidance applications and provides a common platform for fast prototyping and development of robust image-guided applications.

In recent years, OSS has gained visibility in the healthcare community. Several lead projects include OpenVistA, a patient information system based on the Veteran Administration's system, Care2X, an integrated practice management solution in Europe and Health Infoway, a patient data-exchange venture in Canada [1].

5.0 Requirements for Successful OSS framework

While a successful OSS effort can produce rapid, innovative and cost-effective software development, making it successful requires not only an understanding of the technical and business requirements of an OSS framework but the cultivation of a community of users who can contribute and benefit from the endeavor.

An OSS approach must be coupled with an open architecture to be sustainable in the long run. "Open" refers to the process used to develop standards that achieve interoperability where "architecture" defines the components, their organizations and interactions, and the design philosophy used [2]. Standardization is critical for creating interoperable, portable, and reusable components and systems; it also contributes to the development of secure, robust, and scalable systems. Grid technologies have emerged as a component of the national cyber infrastructure supporting effective healthcare information. The underlying open grid services architecture (OGSA) represents a growing

trend in systems architecture. The key to the realization of this Grid vision is standardization, so that the diverse components that make up a modern computing environment can be discovered, accessed, allocated, monitored, accounted for, billed for, etc…, and in general managed as a single virtual system—even when provided by different vendors and/or operated by different organizations [3].

Grid applications in biomedical environments enable the creation and operation of distributed communities across organizational boundaries. Enhanced collaboration environments, visualization tools, computational resources and storage capabilities are all grid services upon which Virtual Organization can build information infrastructure. This emerging information technology infrastructure enables the creation, administration and management of image based biomedical information. [1]

5.1 Technical Requirements for OSS framework

Open-source evangelist Eric S. Raymond suggests a model for developing OSS known as the Bazaar model. He advocates that all software should be developed using the bazaar style, described as "a great babbling bazaar of differing agendas and approaches" [4]. In order to make this model effective, Gregorio Robles suggests the following principles [5]:

(1) Users should be given access to the source code of the software and be encouraged to submit additions, code fixes, bug reports, documentation etc…. Having more co-developers increases the rate at which the software evolves. (2) The first version of the software should be released as early as possible so as to increase one's chances of finding co-developers early. (3) New code should be integrated as often as possible so as to avoid the overhead of fixing a large number of bugs at the end of the project life cycle. (4)
There should be at least two versions of the software - a development version with more features and a more stable version with fewer features. The development version is for users who want the immediate use of the latest features, and are willing to accept the risk of using code that is not yet thoroughly tested. The users can then act as co-developers. The stable version offers the users fewer bugs but fewer features. (5) The general structure of the software should be modular allowing for parallel development. (6)There is a need for a decision making structure, whether formal or informal, that makes strategic decisions depending on changing user requirements and other factors.

5.2 Distribution Scheme for Successful OSS framework

As with proprietary software, open source software is distributed under a license. To help establish some degree of uniformity, the Open Source Initiative (OSI) created the Open Source Definition which is a specification of what must and must not appear in a license in order for the software to be considered open source. To meet the open source definition, a license must provide the following features [6]: (1) The license shall not restrict any party from selling or giving away the software as a component of an aggregate software distribution containing programs from several different sources. (2) The program must include source code, and must allow distribution in source code as well as compiled form. (3) The license must allow modifications and derived works, and must allow them to be distributed under the same terms as the license of the original software. (4) The license must not discriminate against any person or group of persons. (5) The license must not restrict anyone from making use of the program in a specific field of endeavor. For example, it may not restrict the program from being used in a business, or from being used for genetic research.

6.0 Sustainability and Business Models

Although an OSS framework is cost effective, it is not free. There are costs associated with the process. To maintain and grow the effort requires a sustainability plan that goes beyond the initial funding period. Money will not come in through traditional licensing fees, thus other business models need to be considered. As OSS development has matured, a number of business models for sustainability have emerged.

In the service/maintenance model companies sell support and services around the OS software, for example, Red Hat (Linux) or Medsphere (OpenVista). In this approach, users pay for support of the software although they may choose to support the software themselves. In another approach, the vendor provides an open source code base with proprietary add-ons. Examples of this model include Sourcefire (security) and SugarCRM (customer relationship mgt). In a dual license approach, a company offers free use of its software with some limitations, or alternatively offers commercial distribution rights and a larger set of features for a fee. Both the MySQL and Sleepycat databases are examples of a dual license model. In the Aggregation Model also known as the “Lego” strategy, companies act as middlemen to assemble various open source packages into easy-to-use integrated units. SourceLabs and SpikeSource have adopted this model. [1]

7.0 New Business Models for Academia, Industry and Government

The National Library of Medicine (NLM) in the United States (US) has been one of the champions of open source software development. As the imaging data from the Visible Human Project were released for public use, the NLM set out to “create a dynamic, self-sustaining, public domain and extensible toolkit that will empower researchers throughout the world to develop new segmentation and registration algorithms and create new applications that leverage the NLM’s investment in the Visible Human Male and Female data sets” [7]. The project produced the Insight Tool Kit (ITK) after four years and seven million dollars of government funding. This experience made it clear to the government that while open source developed by government grants may promote open science and empower researchers, it is not free. There are costs associated with the

effort such as distribution of the software, quality control of the software, and user support. In order to cross the "valley of death" between research and successful technology transfer, it is imperative that an open source effort can be converted to a financially sustaining activity.

An OSS approach offers a unique way for academia, industry, and government to work in partnership to facilitate rapid dissemination of knowledge into the commercial sector for wider applications. Software developed by the academic research community, under government sponsorship can be offered to the open source community for further testing and development and eventual adoption by the commercial industry.

The US Army Medical Research and Materiel Command (USAMRMC), Telemedicine and Advanced Research Center (TATRC) is responsible for life cycle management of over 500 medical research and development programs, with a 2005 budget of approximately $300 million. The Center's research h responsibilities extend to execution of academic, government and industry programs in biomedical research. TATRC is currently developing a program to improve the productivity in technology transfer from research community to the commercial sector. This program uses Triple Helix strategies involving academia, industry and government to accelerate technology implementation. The open source approach is seen as a potentially effective means of making research results available for greater dissemination through timely commercialization. [1]

7.0 Establishment of a Consortium: ImTK

A new consortium has been formed to launch an open source/open architecture effort that narrows the gap between clinical and research needs by focusing on the development of software tools that enable the efficient exchange, sharing, management, and analysis of multimedia medical information. Imaging and informatics experts at Georgetown University, Washington University in St. Louis, the Northwestern University Feinberg School of Medicine and University of Geneva, Switzerland have agreed to form the Image Management Toolkit (ImTK) Consortium. Collectively this consortium represents demonstrated expertise in technology, clinical operations, technology development, and technology management within the academic, government and industrial environment.

The mission of the ImTK™ Consortium is to expedite translational biomedical research through the development of software tools that enable efficient exchanging, sharing, management, and analysis of multimedia medical information such as clinical information, images, and bioinformatics data. The ImTK Consortium, together with partners in academia, industry and government, will organize itself around four cores: 1) software tool development, 2) open architecture and data model implementation, 3) knowledge dissemination, and 4) management and sustainability. A well managed open source development process has been proven to produce high quality products in a cost efficient manner while simultaneously developing a collaborative user/developer community. The ImTK technology initiative will not only provide open source software tools and components but also an open architecture in which they may be configured and deployed. The tools will comply with existing standards such as DICOM and HL7 and build on the technical frameworks and workflow defined by the IHE initiative. The open architecture will draw on the best practices of the grid computing community and service oriented architecture. This new effort will build on the expertise, processes and development tools used to create ITK and VTK. These processes will ensure the robustness of the software and extend the family of toolkits from image analysis and visualization to multi-media information management, information fusion and data mining.

The consortium will start by developing a collaborative environment for a community of developers and users to work together to define use cases and application scenarios, design and develop new tools and components, and maintain a test bed on which components may be validated and training programs developed and conducted. It will draw on existing successful programs and activities for best practices and insights. The goal is to establish a dynamic, self-sustaining, public domain and extensible toolkit that empowers scientists, engineers and physicians throughout the world to improve the outcome of biomedical research and leverage the government's investment in open source initiatives. The consortium will support the development of robust software for research applications and commercial products through conferences, training sessions, and tutorials.

Acknowledgements

The Authors are grateful to Fred Prior of Washington University, David Channin of Northwestern University and Osman Ratib of University of Geneva for this leadership and expertise in launching the consortium.

References

[1] Multi-Center Image Management Workshop conference activities (abstracts and presentations), March 2006 at http://www.mcim.georgetown.edu/MCIM/DesktopDefault.aspx?tabindex=3&tabid=17, accessed on June 13, 2006.

[2] Open Grid Services Architecture Working Group (OGSA-WG), "Group Purpose" at http://forge.gridforum.org/projects/ogsa-wg, accessed on July 2, 2006.

[3] Linesch, M., "Future of OGSA, Web Services in Hands of IT Community" *Grid Today*, June 13, 2005 at http://news.taborcommunications.com/msgget.jsp?mid=396791&xsl=story.xsl, accessed on June 18, 2006.

[4] E.S. Raymond, "The Cathedral and the Bazaar," *First Monday*, 1998 at http://firstmonday.org/issues/issue3_3/raymond/, accessed 30 June 2006.
[5] G. Robles, "A Software Engineering Approach to Libre Software," *Open Source Jahrbuch 2004* edited von R.A. Gehring and B.Lutterbeck, Berlin (Lehmanns Media) 2004 at http://www.opensourcejahrbuch.de/2004/pdfs/III-3-Robles.pdf, accessed 30 June 2006.
[6] The Open Source Definition at http://www.opensource.org/docs/definition.php, accessed 28 June, 2006.
[7] Ackerman, M.J. and Yoo, T., "The Visible Human Data Sets (VHD) and Insight Toolkit (ITk):Experiments in Open Source Software,", *AMIA Symposium Proceedings*, 2003 at http://lhncbc.nlm.nih.gov/lhc/docs/published/2003/pub2003050.pdf accessed on July 16, 2006.