

The Evolving Role of Open Source Software in Medicine and Health Services

David Ingram and Sevket Seref Arikan

“In attempting to arrive at the truth, I have applied everywhere for information but in scarcely an instance have I been able to obtain hospital records fit for any purpose of comparison. If they could be obtained, they would enable us to decide many other questions besides the one alluded to. They would show subscribers how their money was being spent, what amount of good was really being done with it or whether the money was not doing mischief rather than good.”

Florence Nightingale (1820–1910)
Founder of modern nursing

The past five decades have witnessed immense coevolution of methods and tools of information technology, and their practical and experimental application within the medical and healthcare domain. Healthcare itself continues to evolve in response to change in healthcare needs, progress in the scientific foundations of treatments, and in professional and managerial organization of affordable and effective services, in which patients and their families and carers increasingly participate.

Taken together, these trends impose highly complex underlying challenges for the design, development, and sustainability of the quality of supporting information services and software infrastructure that are needed. The challenges are multidisciplinary and multiprofessional in scope, and they require deeper study and learning to inform policy and promote public awareness of the problems health services have faced in this area for many years. The repeating pattern of failure to live up to expectations of policy-driven national health IT initiatives has proved very costly and remains frustrating and unproductive for all involved.

In this article, we highlight the barriers to progress and discuss the dangers of pursuing a standardization framework devoid of empirical testing and iterative development. We give the example of the openEHR Foundation, which was established at University College London (UCL) in London, England, with members in 80 countries. The Foundation is a not-for-profit company providing open specifications and working for generic standards for electronic records, informed directly by a wide range of implementation experience. We also introduce the Opereffa open source framework, which was developed at UCL based on these specifications and which has been downloaded in some 70 countries. We argue that such an approach is now essential to support good discipline, innovation, and governance at the heart of medicine and health services, in line with the new mandate for health commissioning in the United Kingdom's National Health Service (NHS), which emphasizes patient participation, innovation, transparency, and accountability.

The Evolving Role of Open Source Software in Medicine and Health Services

David Ingram and Sevket Seref Arikan

Introduction

The quotation from Florence Nightingale, a pioneering advocate for a statistical approach to the study of healthcare services, serves two purposes in focusing this article. First, although from a time long past, it resonates tellingly with contemporary critical concerns about the efficiency and effectiveness of health services. Second, it makes a connection between health records that exist to document and make accountable the care given by a professional team to a patient, and the external overview and governance of those services for populations of patients. There is an unavoidable and continuous tension between the focus on the care of a particular patient, in a particular context over time, and a smoothed out picture of the care of a population of patients in a diversity of contexts. A burden of data capture is always imposed, which too often adds cost and pressure to professional practice but fails to deliver value in terms of efficiency of workload and improvement in outcomes achieved for patients. Data are extracted from medical records or collected again, mapped and codified, analysed, and represented in different contexts. At each stage, there is potential for clinical requirement to become blurred and provenance of data to become obscured. This may then reflect in poor design and implementation of software, and in the integrity of systems, creating potential for undue confusion, cost, and error.

In the 60-plus years of the National Health Service (NHS) in the United Kingdom, medicine has moved from a small-volume, low-profile, trusted, and generally accepted service to a high-volume, high-profile, critically scrutinized service. The breadth of services now provided within the healthcare system is very considerable, and their focus ever changing. From an information viewpoint, taking a snapshot as things are now, one might observe that:

1. Biomedical science is being transformed.
2. Healthcare and research are increasingly technology and information intensive.
3. Multiple legacy information systems are in use to support and link healthcare, research, and industry.
4. Governments now want pervasive and standardized ICT infrastructure for healthcare.
5. Many other initiatives in the commercial, public, and social enterprise and voluntary sector domains are creating relevant infrastructures and de facto standards.

Taking another snapshot of the experiences of patients in their encounters with advanced healthcare systems, Blendon and colleagues (2003; tinyurl.com/bbm8z76) conducted a large, multinational survey of patients with chronic conditions. In the United Kingdom, they found that:

1. Two-thirds of patients surveyed were not engaged in discussion about their own treatment and care, 40% did not have goals of treatment made clear, and 20% received conflicting information from different professionals.
2. Twenty percent of patients were victims of medical error in the past two years (9% with serious consequences).
3. Thirteen percent of patients were sent for duplicate tests and half had to repeat health history for different professionals due to medical records not reaching consultations on time.

The challenge of achieving an electronic health care record capable of representing and sharing the content and meaning of the treatment and care of a patient, or groups of patients, within their family and social contexts, has been tackled by many luminary clinical champions and pioneers, starting with Octo Barnett and Howard Bleich in Boston (Massachusetts General Hospital and Beth Israel Hospital) and John Anderson at King's College Hospital, London, in the 1960s.

The King's College Hospital project was the first attempt in the United Kingdom to computerize patient records in the NHS. The project, under the auspices of a brave and innovative clinician, then the Professor of Medicine at King's, was funded as an experiment in meeting what were likely, at the outset, to have been considered well-understood clinical requirements, using well-established computer technology. It was commissioned by the then NHS Supplies Division. In that era, there was almost nothing by way of digital imaging, fast international computer networks, nor even standard database methods – a 5 megabyte disk cartridge was a bulky item. We have learned a lot since then about the problems of clinical data management, and there have been huge technological advances, rendering much of everyday practice today highly dependent on information technology, with considerable convenience and benefit to all.

As the quotation from Florence Nightingale highlights, the quest to compile records that capture the essence

The Evolving Role of Open Source Software in Medicine and Health Services

David Ingram and Sevet Seref Arıkan

of what medicine does, and how well, goes back much further. But, the capture and communication of clinical notes and histories is apparently still not yet computable, at scale, within the NHS, except within communities that share systems. And, of course, clinical and other requirements for managing such data have moved on a very long way.

We need to pause to reflect on the underlying reasons for the repeating failure and also on the implications of the upcoming changes towards what is being called personalized medicine, in which patients and their families will expect to be informed and to participate much more fully. Past failures must be owned at all levels of policy, profession, and practice. The underlying causes are uncomfortable because they highlight large and unrecognized gaps between the aspirations and hype that policy documents sign up to, and the knowledge, capacity, and new learning that are required to achieve useful ends. This is not just about practical expertise in design and implementation of computer systems. It is also about professional capacity to frame, deliver, and sustain realistic information systems and services that meet requirements for integrated and shared health and social care services, and that can and are made to work, for and by professionals, the patients themselves, and the industry.

The rapid pace of advance in the biosciences has required major research focus on new numerical methods, software tools, and data repositories, which has just about enabled that community to keep pace with the rapidly escalating demands in the context of services that explore and exploit genomics science. And, this is mainly in world-leading specialist centres, working and sharing approaches closely together, with their highly capable and expensive research and development teams.

The challenges implicit in achieving coherence of records within wider health services are also immense, but receive less academic and research profile than they merit because they have to involve heavy engagement with more typical everyday healthcare. In addition, when designing to meet data management requirements across diverse healthcare services, an industrial and scalable model for delivery of the needed IT services, employing suitably standardized and generic data architecture, is inescapable. To its credit, the most recent NHS National Programme for IT initiative (NPfIT; tinyurl.com/36sgmdp) took on that challenge, but for complex reasons, it substantially failed.

In the following sections, we introduce two key drivers for a more open approach to software within core systems supporting medicine and healthcare services, arising from the requirements for both technical and semantic interoperability. These requirements are the need for a practically informed approach to common data standards and the need for greater rigour and transparency in the way that systems are designed, tested, and held to account within wider clinical and information governance. We argue that these requirements can best be served within the context of open source communities of practice. We then introduce the openEHR Foundation and an open source implementation of its specifications, called Opereffa (OPENEHR REFERENCE Framework and Application), as examples of initiatives pursuing these principles that have gained wide currency and adoption worldwide.

Data Standards

A fundamental challenge for healthcare IT is in marrying innovation led by scientific research and technological advance with innovation led by requirements for effective and efficient delivery of services that are valued by patients. The standardization of data and computer systems that work and can be sustained over time in these twin contexts is a considerable challenge. However, the challenge can and will, over time and with appropriate combinations of skills and resources, be overcome if it is properly framed. The quest for this standardization has been both albatross and Achilles' heel of NHS IT programmes for too long, the more so because, in some circles, a standard is seen as an enforced technical conformity of systems, as opposed to its primary purpose, which is to serve as a working language that enables and facilitates communication about meaning. This language will evolve over time, as clinical practice, science, and technology move forward, but therein lies the rub. For industry, control of such standards is an important insurance policy guaranteeing marketability of products. Standards are fought over and defended, becoming like tablets of stone, because software costs money and change in underlying standards can render systems rapidly unmarketable, unsustainable, and obsolete.

In recent decades, the difficulty of keeping pace with changing technology and infrastructure, both hardware and software, at scale, within large health systems, has proved unmanageable. So much so that the field has, in the main, been characterized by local successes (local, that is, to a particular clinical domain or institution de-

The Evolving Role of Open Source Software in Medicine and Health Services

David Ingram and Sevet Seref Arıkan

livering healthcare) and costly global failures (global in terms of sharing records across domains in which the data about a particular patient needs to be communicated and worked on, wherever they are cared for nationally or increasingly, internationally).

New approaches are needed to the challenge of defining practical and deliverable scope for standardization across health systems. These approaches need to focus on working more effectively and practically with professions, research, and industry, to learn, through practical implementation experience, about standards that work. The evidence of past endeavour is that the need and urgency of this goal is fully recognized, but the means of achieving it has, to date, been beyond health service, profession, and industry capabilities. As a result, our options when buying IT are too often locked down in inflexible designs, which cannot be changed because their technological underpinnings are already effectively obsolete, and where there has been too great a past investment committed in them to consider replacing them with further untried and untested approaches.

Towards Open Methods, Design, and Governance

In common with other developed health economies, the NHS is now embarking on the fourth major attempt to conquer this moving frontier. Other countries and large scale health providers have lost their billions, too. It seems a good time for the debate to focus differently, and we would suggest three general areas of concern.

1. **Discipline:** the computational science and professional skills needed for collecting, managing, and communicating high-quality clinical and biomedical data. This is a central but relatively underdeveloped discipline, as yet.
2. **Design for adaptability and change:** the principles underlying computer systems where data management can be locally customized and capable of evolving over time to meet new requirements and maintain lifetimes of analyzable clinical data, confidentially.
3. **Governance:** bringing a citizen focus to the overseeing of choices made about how personal health data is collected, shared and used in a manner that is more open and responsible about what we all, as citizens, may reasonably expect and over which we can, collectively, feel ownership and control.

These areas of concern relate directly to the rationale for requiring transparency of systems and for enhancing the prospects for progress in the domain through promotion of an open source approach to software, notably in key areas of shared information infrastructure where interoperability of systems is essential.

The Rationale for Open Source Implementation

This is not the place to analyze or present the breakdown that has occurred between clinical requirement, technological innovation, and institutional adoption and standardization of health records. Key barriers to progress have been and remain:

1. Lack of fit-for-purpose data standards
2. Failure to differentiate primary data requirements (for supporting the delivery of services at the coalface of care) from secondary data requirements (for supporting public health services, audit of health care delivery, and research)
3. Divergence of global and local requirements
4. Governance and confidentiality/privacy concerns
5. Sustainability over patient lifetimes
6. Multi-level, competing initiatives that lack a common strategy
7. Restrictive intellectual property protection; much that needs to be openly shared, debated, and learned from is hidden from view

The essential learning from past failure is threefold:

1. Overly centralized approaches are inevitably remote from the everyday realities of designing and delivering clinical services. These approaches rely too heavily on industry-derived solutions and have failed to conquer the problems of the domain.
2. Most, if not all, successful innovation in the field has rested on the shoulders of local clinical pioneers, who were ambitious, capable, and committed to the hard work of: i) learning by doing; ii) balancing the different drivers and constraints of medicine, technology, and organizational management; iii) plugging gaps in supporting infrastructure; and iv) keeping focused on clinical need.

The Evolving Role of Open Source Software in Medicine and Health Services

David Ingram and Sevket Seref Arikan

3. Local communities of practice and industry working in partnership need shared methods and recognized discipline. These groups are charged with providing and sustaining information services, and they must be enabled and supported to learn and adapt.

In earliest times, many local health providers built local team capability in the domain. This was costly and placed a huge pressure on these teams because the requirements of the domain as a whole were only slowly being clarified as they did their work. It is in fact a classic "wicked problem", as characterized by Rittel and Webber (1984; tinyurl.com/a48syez) because:

1. There is no clear ownership of the problem, permission to experiment, or right to judge.
2. Working towards a solution clarifies and revises understanding of the problem.
3. The problem does not have right answers; it requires "good enough" approaches.
4. The problem is never completely solved.
5. Solutions require changes in behaviour of stakeholder groups.
6. The manner in which the problem is tackled is as important as how it is tackled.

Open source commends itself as an approach, not because in some magical way it makes it cost-free – that is an almost laughable simplification only meaningful to those who wish to exploit other peoples' efforts rather than join in with and learn from them, adding value back into the commons. Key features of an open source ecosystem that could contribute most towards progressive solution of the wicked problem are:

1. The promotion of effective and efficient innovation within software development communities. Clinical pioneers have traditionally had to build whole local infrastructures, and unnecessary efforts to "reinvent the wheel" are still common.
2. The pooling of development and maintenance costs for essential infrastructure
3. The enablement of the research interface. Discipline grows through sharing, testing, and reviewing methods.

4. Increased awareness and understanding of the inner workings of systems to improve procurement

5. The integration of systems and services and continued efforts to combat fragmentation

6. The enablement of more patient-focused services that can be managed and regulated more effectively and transparently. The goal is to create a more trusting environment of professional practice, patient participation, and public engagement.

But, there must be a business case for this approach. Support from government and industry, as well as health professionals, is needed for the transition to a viable open source framework and community.

openEHR

The challenge of the electronic healthcare record has been progressively understood and refined, in information terms, to embrace both the narrative and quantitative representations of knowledge useful to describe patients and their clinical and healthcare problems, the interventions through which they are cared for, and the outcomes achieved over time. To address this challenge, the openEHR Foundation (openEHR.org) was founded as a not-for-profit company. It builds upon earlier attempts to formalize information architecture for an electronic patient record infrastructure. For full details on the origins of openEHR, see: openehr.org/about/origins.html.

In simplest terms, openEHR provides several levels of design to help enable better more sustainable electronic records:

1. A domain reference model comprising a formally rigorous, openly change-managed set of building blocks that are needed to capture, organize and communicate the content of all health records, and the structures needed such that they can meet the wider ethico-legal requirements for aggregation and sharing of records.
2. An archetype model, open source software tooling, and a growing repository of published artefacts, supporting the ways in which these building blocks can be combined to build a clinical record. Each concept, such as a laboratory test result or an instruction to initiate a treatment or document its application, is

The Evolving Role of Open Source Software in Medicine and Health Services

David Ingram and Sevket Seref Arikan

modelled by clinicians in ways that represent its generic sense and meaning (as an openEHR archetype – defined as a formal constraint on the reference model), for everyone working with it.

3. Customized local use of archetypes through templates (defined as a further constraint of the archetype), universal querying of the resultant record content, and a growing body of open source software "plumbing", to enable integration within the principal generic platforms on which software infrastructures are built.

The quest for a universal electronic health record has collapsed several times, about once every decade, in large part due to overly ambitious global and top-down programmes and initiatives, which lost touch with or ended up inhibiting or suppressing coal-face innovation by local champions. openEHR has sought to champion open specifications and an inclusive community of "doers" – people motivated principally to help necessary change happen. Individuals in this community are somewhat free-spirited, and thus sometimes a bit quarrelsome and disruptive, but none the worse for that, because the community has also sustained a largely united common vision and mission. It is now working internationally to evolve a business plan that can sustain its fundamental openness, empiricism, and inclusivity.

The architecture has been experimented with and adopted by individuals, small companies, universities, whole institutions (e.g., the Marand/Ocean openEHR-based paediatric centre patient record in Slovenia), and countries (e.g., Brazil). Spanish and Japanese language versions are available, and openEHR now has members in more than 80 countries, with progress underway towards regional chapters. It is gradually attracting the support of larger and more influential stakeholders. The Clinical Information Modelling Initiative (CIMI), which is centred in the USA and which dominates present markets for health IT, recently voted to base its work towards standardization of EHR content on the reference model, archetype formalism, and current shared CKM archetype repository of openEHR.

Opereffa

There are many gaps to be filled in making an open source framework for health information systems and services a reality. At the same time, as judged by the several hundreds of clinical database systems, collecting and analyzing clinical data for research in a typical

teaching hospital trust, there is a huge unmet need for workbenches and tooling to help a new generation of innovators create the systems required for tomorrow.

The rationale for the Opereffa open source framework initiative (tinyurl.com/b3t38m7), currently being undertaken at UCL by its spinout company Charing Systems, is as follows:

1. Data mining tools and frameworks focus on system designs that are traditional now for information storage and retrieval, and that embody relational databases with normalized tables. Electronic health records are rarely implemented with such normalized designs, though the relational database is still the most widely used underlying technology.
2. These tools and frameworks are expensive, especially if they must be scaled.
3. Open source development frameworks help with the cost, but they are still "tools for the few". And, there is still very little open source tooling available to implementers to help them build standards-driven clinical information systems.
4. The Eclipse-based Opereffa framework is an open source test bed developed in an attempt to deliver a flexible information repository with rich methods of connectivity to other platforms.

Opereffa has been downloaded 1000 times in 70 countries and has featured in many local systems developments, such as a national surveillance system for tuberculosis in Cambodia.

Opereffa is working to:

- leverage high-performance, open source development frameworks
- explore requirements for simultaneous support of clinical service delivery and machine learning
- explore design approaches that can simultaneously improve information retrieval and machine intelligence performance
- develop richer, smarter methods of access to clinical data to better support machine intelligence, especially by delivering alternatives to SQL-based access to relational databases

The Evolving Role of Open Source Software in Medicine and Health Services

David Ingram and Sevket Seref Arikan

- eliminate complicated bridges between clinical information systems and research data repositories
- explore new design and implementation patterns for the openEHR specifications that enable these goals

This all boils down to standards-driven persistence and data access: Opereffa and the openEHR Archetype Query Language (AQL). The importance of AQL lies in its capability to extend the usability of domain concepts used for modelling clinical data, into coherent querying of ensembles of clinical data. This approach goes beyond avoidance of impedance mismatch between object oriented and relational views of data; it converts the underlying domain models into a platform that supports all data access and processing requirements, with one coherent method.

The Opereffa architecture is illustrated in Figure 1. It uses:

1. Proven open source persistence stacks, aligned along the scale axis from PostgreSQL to Hadoop
2. High-performance open source parallel processing frameworks for scaling upwards: Akka, Hadoop

3. Open source tooling to eliminate complicated technology and infrastructure management processes: the Eclipse framework

The goal is to bring all these technologies together within the strongly model-driven approach of openEHR, to achieve outcomes that are portable to other domains.

Conclusion

Building something that must, by its very nature, be a shared discipline, requires development of tried and tested methods. This cannot happen if those methods remain hidden away in systems, out of reach of critical appraisal and opportunities for shared learning. Openness and transparency about these methods is thus crucial. Intellectual property rights must be respected but there can be no discipline for others to build on if the methods underpinning the design and standardization of systems is not clearly expressed, validated, and made accessible and open to improvement in wider communities of adoption and practice.

Adoption of the openEHR specifications, and associated open source implementations such as Opereffa,

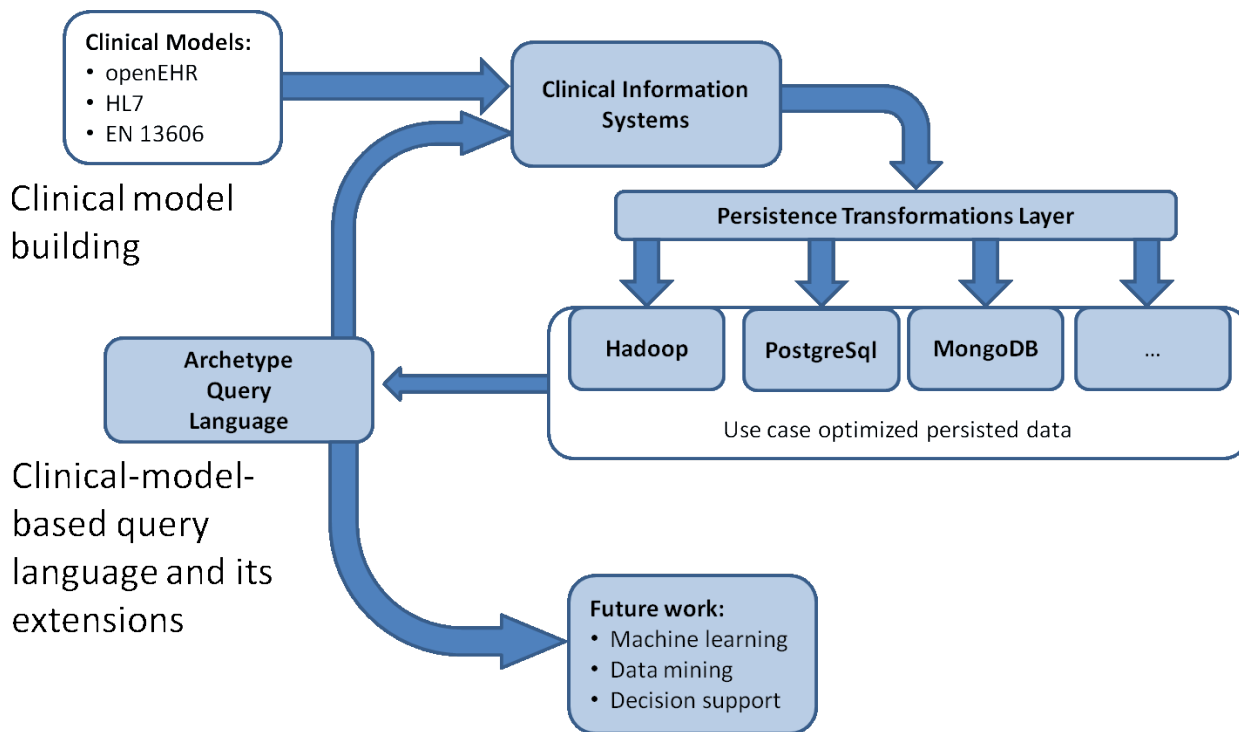


Figure 1. The Opereffa architecture

The Evolving Role of Open Source Software in Medicine and Health Services

David Ingram and Sevket Seref Arikan

have succeeded as exemplars of good practice in a number of ways:

1. They are vital steps for model-driven life and clinical sciences data management; i) analysis of domain data and processes, ii) building of computable information models, iii) implementation of model-driven functionality, and iv) avoidance of the tendency to "reinvent the wheel".
2. The archetype methodology (two-level modelling) of openEHR is reusable across a considerable variety of domains.
3. Tooling and computation is driven by one common infrastructure, which is, in turn, built on the reference model (RM).
4. The openEHR two-level modelling approach provides access to 20 years of research and implementation practice.
5. There has been continuous and increasing adoption of this approach at national and international levels.
6. Key factors helping towards a better return on investment in the domain include: i) it establishes an implementable, incremental route towards adoption and ii) it broadens industry's opportunities for integration with and investment into customer and partner technology stacks.

Without deeper underpinnings of patient and clinical engagement, engineering discipline, and trust, any new policy will likely lead us along the now familiar path: up hill and down dale, towards yet another policy. It is time for the task to be reformulated. In the spirit of Fred Brooks and his concept of "the mythical man month" (tinyurl.com/znl98), one might suggest, controversially, that a tenth of the money being spent today would, over time, if better focused, deliver far greater benefit for health care than has been achieved to date by the expensive systems and programmes currently deployed. Too much money is being spent on not solving the problem. The requirements and form of an open source ecosystem for health information systems and services deserves realistic and informed consideration. It must be soundly grounded in organizations and methods that can be demonstrated to work, and be sustained.

About the Authors

David Ingram has held posts in industry, the National Health Service, and university medical schools. After undergraduate physics at Oxford and several years in the medical engineering industry, he studied computer science and completed doctoral research on the mathematical modelling of biological systems at University College London. He was appointed to the first UK Chair in Medical Informatics in 1989 and participated, as partner and reviewer, of numerous EU and UK Research Council Health Informatics programmes and projects from 1985-2011, including coordinating the EU GEHR Project, which laid the foundations for a standard health record architecture and the openEHR Foundation and community, internationally. In 2011, he established Charing Systems as a spinout company of UCL, to provide services to developers and users of clinical systems, to support their integration within open-source platforms, utilizing the specifications and methods pioneered and made freely available under open license by the Foundation. He is an elected Honorary Member of the Royal College of Physicians, in recognition of his services to medical science.

Seref Arikan has worked in the software industry for 15 years and in the medical informatics domain for 10 years. He is strongly focused on research and development tasks and has wide experience of information technologies and architectures for projects ranging from workflow based systems to national e-health repositories. He has been studying at UCL, pursuing research for a PhD under the supervision of Professor David Ingram, since 2008 and is currently working at Ocean Informatics UK. His research interests are in innovative, high-performance architectures to enable and support computable machine intelligence in healthcare, supported by open source tools and frameworks. He has already released much of this work as open source – the Opereffa framework described here being the most significant item. The clinical context of his PhD programme is the ophthalmology record at Moorfields Eye Hospital in London, where he works alongside Dr Bill Aylward, leader of the openEyes initiative for an open source electronic patient record for eye care.

Citation: Ingram, D. and S. S. Arikan. 2013. The Evolving Role of Open Source Software in Medicine and Health Services. *Technology Innovation Management Review*. January 2013: 32-39.

