eHealth Beyond the Horizon – Get IT There
S.K. Andersen et al. (Eds.)
IOS Press, 2008

Knowledge Management for the Protection of Information in Electronic Medical Records

Nathan LEA [a,1], Stephen HAILES [b], Tony AUSTIN [a] and Dipak KALRA [a]
[a] *The Centre for Health Informatics and Multiprofessional Education, University College London, United Kingdom*
[b] *Department of Computer Science, University College London*

Abstract: This paper describes foundational work investigating the protection requirements of sensitive medical information, which is being stored more routinely in repository systems for electronic medical records. These systems have increasingly powerful sharing capabilities at the point of clinical care, in medical research and for clinical and managerial audit. The potential for sharing raises concerns about the protection of individual patient privacy and challenges the duty of confidentiality by which medical practitioners are ethically and legally bound. By analysing the protection requirements and discussing the need to apply policy-based controls to discrete items of medical information in a record, this paper suggests that this is a problem for which existing privacy management solutions are not sufficient or appropriate to the protection requirements. It proposes that a knowledge management approach is required and it introduces a new framework based on the knowledge management techniques now being used to manage electronic medical record data. The background, existing work in this area, initial investigation methods, results to date and discussion are presented, and the paper is concluded with the authors' comments on the ramifications of the work.

Keywords: Security, legal issues, compliance, repository system for medical records, knowledge management, ethics, patient identification, distributed systems, sharing personal data, personal privacy, confidentiality

Introduction

Research into establishing shareable, clinically meaningful and accurate Electronic Medical Records (EMR) has been continuing for about fifteen years to improve the provision of information resources for effective medical care. This research has produced several examples of repository systems for medical records across the world [1] [2], which facilitate disease management, decision support and patient monitoring for millions of individuals. These repositories can also support medical research, where information about large numbers of patients is made available to help improve care provision. The sharing of larger quantities of data is made easier by the

[1] Corresponding Author: Nathan Lea, CHIME, University College London, 4th Floor, Holborn Union Building, Highgate Hill, London, N19 5LW, United Kingdom; E-mail: n.lea@chime.ucl.ac.uk

international standardisation and use of information models (from Health Level 7 [3], EN / ISO 13606 and *open*EHR [4], amongst others), and the use of Archetypes to “provide the common basis for ubiquitous presence of meaningful and computer-processable knowledge and information” [5].

There are legal and ethical responsibilities to protect this information: the rights of the subject of information are considered in law to be paramount, and this gives rise to constraints on how clinicians and researchers may behave with respect to the information so that no harm comes to individuals as a result of the increased accessibility. The constraints exist in the form of legislation and standards (summarised in Section 2), which inform institutional and research governance in the form of policies [6] [7] at the level of both research institutions and healthcare authorities: these policies may have different human interpretations and are often difficult to implement.

The security literature recommends the use of policy-based controls to manage the protection requirements (as per ISO 17799, discussed in section 2), but this requires interpretation in the same way as legislative controls. The use of roles-based access control (RBAC) is a keenly researched methodology for computable policy implementation, but the process of interpreting and operationalising each policy remains manual, loses much of the semantic sense when put into computable form (as Becker illustrates [8]), and relies on the use and configuration of generic software solutions. These are not holistic in scope, have insufficient structure, syntax and weak semantics, and are under-specified for individual classes of EMRs.

This paper proposes a knowledge management framework to meet the comprehensive needs of managing policy-based controls governing a repository system for medical records, and advocates that this framework uses similar techniques to those that manage EMR knowledge (i.e. – Archetypes). This approach is expected to complement existing solutions so that their manual configuration and risks of human fallibility in policy specification can be alleviated.

1. Methods Used for Initial Investigations

The methods for investigating this problem space have included literary reviews of: legislation (including the UK Data Protection Act 1998, Human Rights Act 1998, Freedom of Information Act 2000 and National Health Service Act 2006 (sections 251 and 252) as well as European Equivalents); informatics knowledge management techniques (including the standards work for information models, construction of EMR repositories [9] [10]); commentaries on the sharing of information and associated risks of centralising information [11]; and existing solutions (including the Ponder Policy Specification [12] and Cassandra [8] amongst others). Standards reviewed include Information Security Management ISO 17799; the draft ISO 27799 for healthcare information security management; and the draft ISO 22271 for constructing EMR repositories for research. The academic literature is being continuously reviewed for indications of issues that have arisen and lessons learned in the deployment of EMR repositories.

Formal and informal interviews have been conducted to support the literature reviews. Medical practitioners and researchers are particularly useful in helping to explore current working practice, levels of knowledge and awareness of information protection issues and difficulties raised by the legislative and governance requirements.

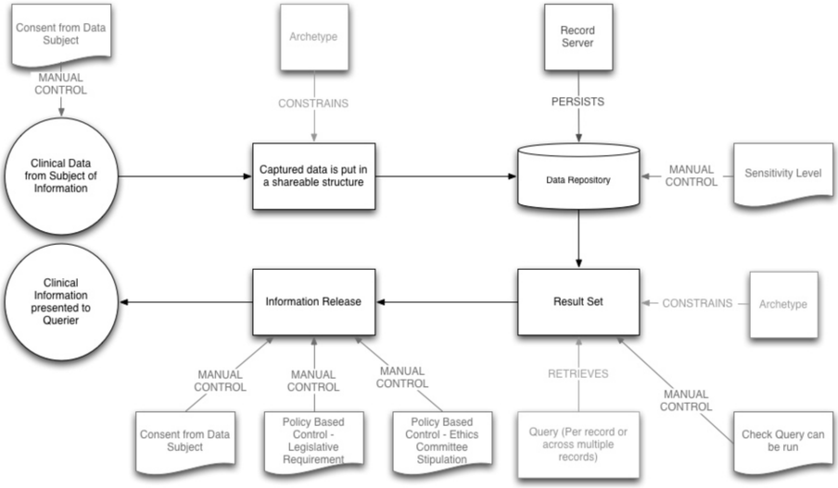
Consent from Data Subject
MANUAL CONTROL
Archetype
CONSTRAINS
Record Server
PERSISTS
Clinical Data from Subject of Information
Captured data is put in a shareable structure
Data Repository
MANUAL CONTROL
Sensitivity Level
Clinical Information presented to Querier
Information Release
Result Set
CONSTRAINS
Archetype
MANUAL CONTROL
MANUAL CONTROL
MANUAL CONTROL
RETRIEVES
MANUAL CONTROL
Consent from Data Subject
Policy Based Control - Legislative Requirement
Policy Based Control - Ethics Committee Stipulation
Query (Per record or across multiple records)
Check Query can be run

Figure 1. Information Flow and Management in EHR Supported Services

Wider discussions with legal experts, legislators and security experts have helped to refine the requirements of protection and policy setting.

There are often examples of the needs of the informatics community at more general conferences (particularly MedInfo 2007) and meetings that illustrate concepts for tooling and working practices. Further investigation of the protection requirements has been undertaken through the practical experience offered by the Clinical eScience Framework project [13]. The results of these reviews and discussions have been used to establish a set of preliminary requirements for the protection of sensitive data in EMR repositories. These requirements have been refined to model current working best practice in managing EMRs and proposed best practice.

2. Results

The reviews and discussions to date have yielded some preliminary conclusions about the process of protecting information stored in and shared by EMR repositories. Figure 1 illustrates current good practice in terms of managing the storage and retrieval of EMRs by Archetypes, assuming an EN 13606 or related scenario; the protection measures are included in the diagram and are labelled as 'MANUAL CONTROL'.

The requirements for the protection measures show that there is a need for individual patient consent to use the information, but gaining such consent is not a scaleable activity, especially in the case of medical research where thousands or millions of patients' records can now be accessed. The de-identification of patient information, where identifying attributes such as names, dates of birth, hospital identity numbers and postcodes are removed, is considered by medical ethics practitioners and in law to be a reasonable means of assuring some level of confidentiality where consent cannot be (or is not) gained [14], but it is very hard to facilitate in practice [11]. Furthermore, complex patterns of roles and data-uses have started to emerge as the sharing of medical information in different contexts of use becomes easier.

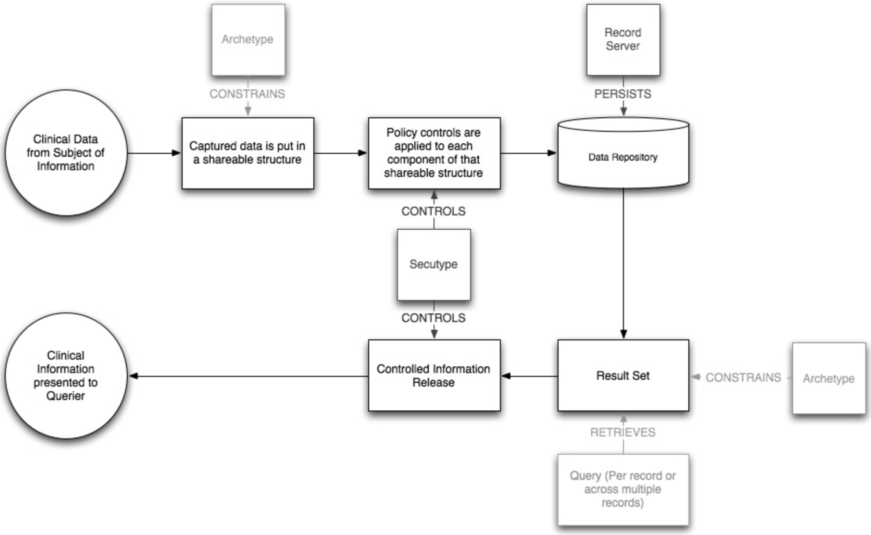
Archetype
CONSTRAINS
Record Server
PERSISTS
Clinical Data from Subject of Information
Captured data is put in a shareable structure
Policy controls are applied to each component of that shareable structure
Data Repository
CONTROLS
Secutype
CONTROLS
Clinical Information presented to Querier
Controlled Information Release
Result Set
CONSTRAINS
Archetype
RETRIEVES
Query (Per record or across multiple records)

Figure 2. Application of knowledge management component for information security control - Secutypes

It is clear that security controls need to be applied to individual data items, and those controls need to be asserted based on details about who is accessing the information, for what purpose and under what circumstances: when accessing EMR repositories for point-of-care uses, there may be consent and governance requirements that allow only named clinicians directly responsible for care to access details, whereas in the case of research queries an ethics committee may only allow access to data about deceased patients where the data has been anonymised or partly anonymised (pseudonymised).

Archetypes allow for the modelling of knowledge about discrete classes of data and may also provide a suitable foundation for applying security controls. There is a need to capture the policy specifications for the protection of those discrete data items, and it is proposed that this can be managed by the construction of a knowledge management framework. Part of the requirements analysis has led to the initial design of a new formalism, based on the theory that reusing the information model and Archetype approach will meet the protection requirements and policy based controls. This formalism is known as the *Secutype*. The function of the Secutype will be to bind policy-based control information to Archetypes, and allow that information to be used to assert those controls to software tools, some of which will require construction for the purpose (a de-identification component, for example). Figure 2 illustrates where Secutypes operate in the management of an EMR. On the next page, Figure 3 shows the structure of a Date of Birth Archetype with the encapsulated data value for the date of birth; Figure 4 shows a depersonalisation Secutype with the controlled data values that will be released, based on examples of policy that exist in the CLEF project.

An overhaul of existing editing tools for Archetypes will be required to support the new Secutype capabilities, and this is being conducted as part of a doctoral work by the first author. The results of analysing the working patterns of the informatics community

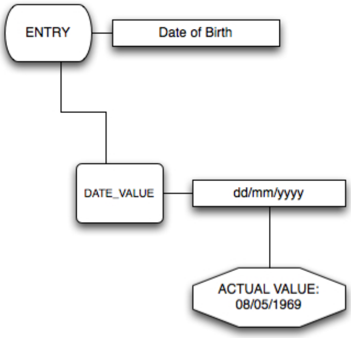

Figure 3. Date of Birth Archetype

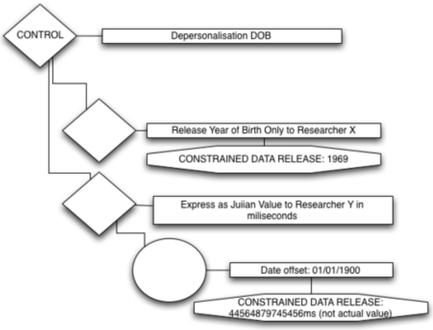
CONTROL
Depersonalisation DOB
Release Year of Birth Only to Researcher X
CONSTRAINED DATA RELEASE: 1969
Express as Juiian Value to Researcher Y in miliseconds
Date offset: 01/01/1900
CONSTRAINED DATA RELEASE: 44564879745456ms (not actual value)

Figure 4. Depersonalisation Secutype

are significant: international collaboration on building information models and Archetypes appears to be popular and essential based on the work of communities such as the *open*EHR Foundation (www.openehr.org) where commentary by interested parties on Archetypes occurs via email. Another community developing models for the United Kingdom National Health Service has adopted a Wiki approach to collaboration (see http://www.ehr.chime.ucl.ac.uk/display/nhsmodels/Home), where any party (even outside the editing team) is permitted to leave comments in support or criticism of the current specification. However the Wiki used is a standard one that does not provide editing support for Secutypes in particular. The planned new Secutype editor uses a highly generic model for the capture of definition statements, and a Web application that provides a more focussed collaborative, editing environment is being developed for this. The model and application is being created based on the protection requirements and the needs of the informatics community already outlined.

After the model and application are completed, it is anticipated that there will be issues of practicality when they are applied to EMR systems, particularly when scaled to thousands of patients' records. These issues and possible solutions will be investigated as another part of the doctoral work, where Secutypes will be added to an advanced implementation of the EN / ISO 13606 record standard. Comparisons of system performance will be run with the Secutype components enabled and disabled across increasing numbers of patient records to discover the impact that Secutypes will have: once the practical issues have been established, potential solutions can be proposed, implemented and evaluated. Additional future research will explore how large, diverse and granular a set of Secutypes is needed within one healthcare organization or region.

3. Conclusions

This paper has proposed that a knowledge management approach is reasonable for the assertion of the required policy-based controls, and that a new formalism called Secutypes, based on the design principles of information models and Archetypes, can be used to capture the required details of policies and facilitate control assertions where needed. The paper has also identified anticipated issues of scalability and performance once the Secutypes are applied to a running EMR, and indicated the proposals of further doctoral work on this subject.

Further research will provide a means to share details about security requirements for EMRs in the different use environments in which they can operate. A contribution of this work is anticipated to be the foundation of a library of Secutypes that might in the future facilitate consistent good practice across medical repositories within a nation or health system in which most policies ought to be the same. This will also support the widespread reuse of controls that already exist for managing access control, audit and data integrity, which currently require manual configuration based on a human-readable policy. It will allow for modularisation of the security in existing EMR servers so that policy can be automatically applied to fine-grained discrete data items, a requirement of modern data protection controls. Finally, there will be investigation into the performance impact of applying this control mechanism, and whether it is scaleable to the millions of patients' EMRs in the distributed computing environments used for large scale, national information technology projects.

References

[1] D. Kalra, J. Milan, T. Austin, D. Ingram, D. Lloyd, D. Patterson, J. Grimson, W. Grimson, Synapses in Use: Supporting Cardiac Care at the Whittington Hospital. *Towards an Electronic Health Record Europe '98*. Medical Records Institute for the Centre for Advancement of Electronic Records Ltd. (1998), 306-312.

[2] *openEHR Usage*: The *open*EHR Foundation. http://www.openehr.org/shared-resources/usage/governments.html (last accessed 15th November 2007).

[3] *Health Level 7 Record Information Models*: Health Level Seven. www.hl7.org (last accessed 15th November 2007)

[4] *openEHR release 1.0.1 Specification Road Map*: www.openehr.org/svn/specification/TAGS/Release-1.0.1/publishing/roadmap.html (last accessed 16th November 2007).

[5] S. Garde, E. Hovenga, J. Buck, P. Knaup. Expressing Clinical Data Sets with openEHR Archetypes: A Solid Basis for Ubiquitous Computing. *International Journal of Medical Informatics* **76** (2007), 334-341.

[6] *New Operational Procedures for NHS RECs Guidance for applicants to Research Ethics Committees*: Central Office for Research Ethics Committees (COREC), www.sehd.scot.nhs.uk/cso/AboutCSO/Ethics/COREC%20flyer.pdf 1st March 2004. (last accessed 15th November 2007).

[7] *Information Governance – Connecting for Health*: www.connectingforhealth.nhs.uk/systemsandservices/infogov (last accessed 16th November 2007).

[8] M.Y.Becker, Information Governance in NHS's NPfIT: A Case for Policy Specification. *International Journal of Medical Informatics* **76** (5-6) (2006), 432-437.

[9] T. Austin, The Development and Comparative Evaluation of Middleware and Database Architectures for the Implementation of an Electronic Healthcare Record. www.ehr.chime.ucl.ac.uk/download/attachments/71/Austin%2C+Tony+%28PhD+2004%29.pdf?version=1 (last accessed 16th November 2007).

[10] D. Kalra, Clinical Foundations and Information Architecture for the Implementation of a Federated Health Record Service: www.ehr.chime.ucl.ac.uk/download/attachments/71/Kalra%2C+Dipak+%28PhD+2002%29.pdf?version=1 (last accessed 16th November 2007).

[11] R. Anderson, Under Threat: Patient Confidentiality and NHS Computing: http://www.cl.cam.ac.uk/~rja14/Papers/drugsandalcohol.pdf (last accessed 16th November 2007).

[12] M. Sloman, E. Lupu. Security and Management Policy Specification. *IEEE Network* **16** (2002), 10–19.

[13] D. Kalra, P. Singleton, D.Ingram, J.Milan, J. MacKay, D. Detmer, A. Rector, Security and Confidentiality Approach for the Clinical eScience Framework (CLEF). *Methods of Information in Medicine* **44** (2) (2005), 193-197.

[14] C. Haynes, G. Cook, M. Jones, Legal and Ethical Considerations in Processing Patient-Identifiable data without Patient Consent: Lessons Learnt from Developing a Disease Register. *Journal of Medical Ethics* **33** (2007), 302-307.