INTERNATIONAL JOURNAL OF MEDICAL INFORMATICS 76 (2007) 449–453

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

A notary archive model for secure preservation and distribution of electrically signed patient documents

Pekka Ruotsalainen[a,*], *Bryan Manning*[b]

[a] *National Research and Development Centre for Welfare and Health (Stakes) Centre of Excellence for ICT, P.O. Box 220, 00531 Helsinki, Finland*

[b] *European Federation for Medical Informatics, Information Planning and Modelling Working Group, Centre for Business Information, Organisation and Process Management, Westminster Business School, University of Westminster, London, UK*

ARTICLE INFO

Keywords:
Long-term archiving
Notary archive
Fine grain data access
Time-stamp
Event record
Metafile

ABSTRACT

The healthcare industry is moving from paper-based documentation into the digital era. Electronic health records (EHR) are playing a major role in this development. Electronic health records will not only to be shared among a growing number of healthcare providers but they have also to be archived over long periods of time. The required life cycle depends of national regulations, but typically the preservation time of patient data varies between 20 and 100 years. Availability, integrity, confidentiality and non-repudiation of stored data over these lengthy preservation periods needs to be fully proven, both to preclude loss and also ensure the ability to read and understand content is maintained.

This document describes a co-operative trusted notary archive (TNA) which receives granular health data from different EHR-systems, stores data together with associated meta-information for long periods and distributes granular EHR-data objects. TNA communicates with EHR-systems and external users via archive request and distribution messages. TNA can store objects in XML-format and prove the non-repudiation and integrity of stored data with the help of event records, Time-stamps and archive e-signatures.

1. Introduction

New models of healthcare delivery emphasise the need for patient information to be shared among a growing number of healthcare service providers together with the patients themselves. As a result, more and more communication is occurring across traditional organisational boundaries. Electronic health records (EHR) are playing a major role in this development.

When patient data is recorded electronically, it has to be preserved over long periods either in a local database or in an external archive. During the past few years, various electronic archives, such as PACS—an archival system for medical images, have been created, and remain independent of local patient information systems. However, electronic archives are becoming a core information storage service shared between different groups of users. These electronic archives can also receive information from different organisations, and data to be stored can contain information that was created in different contexts and for different purposes [1].

Typically, a patient's health record should be signed after the care episode by the responsible doctor. In a digital environment, electronic signatures are also necessary to prove the integrity and originality of patient data.

Electronic health records have to be archived over long periods of time. The required life cycle depends of national regulations but typically the preservation time of patient data

* *Corresponding author.*
E-mail addresses: pekka.ruotsalainen@stakes.fi (P. Ruotsalainen), bryan.manning@binternet.com (B. Manning).
1386-5056/$ – see front matter © 2006 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.ijmedinf.2006.09.011

varies between 20 and 100 years. Availability, integrity, confidentiality and non-repudiation of stored data over the whole preservation period need to be fully proven [2].

Electronic data storage is threatened by the same basic hazards as paper storage. Data can disappear, integrity can be lost, together with the ability to read and understand its content. The useful lifetime of stored health information in many cases exceeds the life span of formats and technical tools used to preserve data. It is also possible, that during the storage period the validity of some digital signatures may become weakened, and PKI-certificates might be revoked or expire [3].

In the case of long-term preservation of electronic health records confirmation of the availability of stored information will be a demanding task. Data structures and formats can change during preservation time, which will make it difficult to locate and use the data.

A notary archive is one possible solution, as it enables the availability and integrity of digitally signed data to be proven together with precluding repudiation of data stored over lengthy preservation periods. This is particularly important when structural conversions are necessary and data has to be transferred to a new storage medium.

2. Definitions

An *archive* is an organisation providing health record preservation services that allow strictly controlled access to an identified group of consumers for a regulated time period [5].

An *electronic archive* (eArchive) preserves information in digital format. A passive eArchive stores fixed data content with associated metadata and policies. This fixed data content should be fully defined and structured atomically within an application before it can be sent to the archive. After the data is archived, it cannot be modified or deleted before its preservation time expires [4]. On the other hand, an active eArchive allows both random access and update of any data element during preservation time.

An *archiving system* is an organisation required to deliver available information in a correct and independently understandable form over a protracted period, within an appropriate set of access and security constraints.

An archive data package is a collection of archived data objects with associated information, an associated evidence record and an archive metafile.

A notary archive is a trusted organisation providing both services for long-term preservation of health records and services ensuring the integrity and non-repudiation of the original data by extending this to include the periodic renewal of Time-stamps and collection of supporting evidence. A notary archive performs the following functions: authentication of electronic transactions, non-repudiation and confirmation of the data integrity. After a notary archive has signed documents electronically, they have the same status as documents signed by a person [3].

A Time-stamp itself is an attestation generated by a Time-stamp Authority that a certain data item existed from a certain time [3].

An archive Time-stamp is an attribute which contains Time-stamps and additional information needed to verify the existence of a data object/object group from the time certified by a Time-stamp [3].

An evidence record is a collection of sets of evidence connected to one data object or group of objects. It can include Time-stamps, verification data, PKI-certificates and security policy information. It may be used to preserve descriptive information so that trust can be established in the certificates after they have expired. An evidence record should be signed in such a way that any modification of data objects or evidence record can be detected [3].

A policy is a collection of rules describing which actions are allowed under certain specified circumstances. A security policy and an archiving policy are basic policies needed for the conduct of any archive.

3. Requirements for secure archiving of health records

The primary role of the archiving system is to make information available in a correct and independently understandable form over the whole regulated preservation period. It also has to guarantee the long-term availability, integrity and confidentiality of all of its stored data throughout this period. Digital archiving of health records must also meet regulatory requirements set by legislation, for example the preservation time set by national legislation [4].

An eArchive service also has responsibility [5] for:

- managing changes in the legal status of the EHR during preservation time;
- managing overriding access conditions;
- managing patient's consent;
- protecting EHR-data based on its purpose and context;
- proving the originality of stored EHR-data.

In order for the preserved information to be both accessible and understandable to those who need access to it, related transactional and other associated historical data gathered over the preservation period must be stored together with it. This data consists of descriptive, representation, content and preservation description information which must be archived together with the actual data as one information unit. The archive must also manage data migrations or translations during preservation time [6].

4. Security requirements for a trusted notary archive

A trusted notary archive (TNA) preserves signed health records for long periods and must meet the general security requirements set out in Section 3 above. The additional security task of a TNA is to prove the integrity and origin of data as well as to prove the non-repudiation of data throughout the whole preservation period. Integrity has to be proven after any data migration, which requires changes of formats. A TNA must also include services for verifying signatures and integrity of data, as well as proving that a specific event has occurred.

To meet those requirements, a TNA has to:

- collect evidence;
- create evidence records;
- generate and renew archive Time-stamps;
- e-sign documents;
- create audit-logs

and then reliably preserve this information throughout the whole archiving period [3].

5. A notary archive model for preservation and distribution of digital health records

There are different combinations of EHR-systems and archives. The e-archiving systems can be classified as centralised, co-operative or federated. A centralised eArchive can be considered as an embedded part of an enterprise EHR-system. In this case, the EHR-system and the archive have the same centralised security policy, access control and privilege management services.

The archive and the EHR-system can also form a co-operative information system. In this case, the eArchive communicates with an EHR-system typically in the form of data requests and answers. The EHR-system sends data to the archive and receives data from it. A typical example of this kind of combination is a PACS which communicates with an EHR-system using DICOM messages.

A federated eArchive is an extension of the co-operative approach involving a number of separate and often widely dispersed organisations combining their resources as a "Virtual Enterprise" [7] and operating within a trusted third party agreement [8,9].

As a result, the EHR-systems and the archive itself form a distributed system of varying complexity dependent on the number of parties involved. Whilst this calls for rigorous overall service management, it has the benefit of providing greater resilience and business continuity capabilities in the face of potential operational risks or malicious interventions.

The eArchive model for preserving digital health records described here and illustrated in Fig. 1 outlines the co-operative approach inherent within a trusted notary archive (TNA), which may be communicating with various different EHR-systems, at local, regional and national levels. Whilst a single eArchive is shown for simplicity, both its potential size as well as back-up and disaster recovery requirements dictate that in reality this will be a complex interconnected multi-site based data storage facility.

The TNA supports multi-source data collection. It stores signed single objects (an EHR) or documents having a multi-grained data structure (e.g. groups of objects) received from different EHR-systems for a period specified in the metafile associated to the data object.

The TNA collects data objects having the same identification code (e.g. objects belonging to same patient) and stores them in a single virtual patient folder. The TNA has evidence record management services to prove non-repudiation of objects and a secure audit-log service. The TNA has also

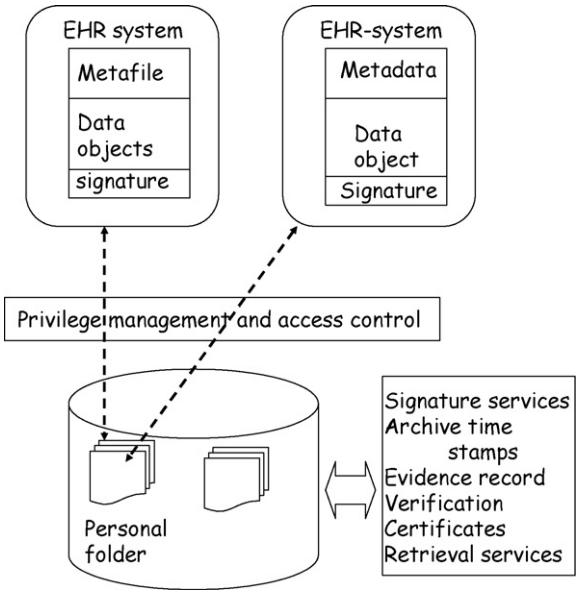

Fig. 1 – Co-operative TNA and EHR-systems.

a signature management system to produce archive Time-stamps and archival e-signatures.

Fig. 2 shows the conceptual data model used both in communications between an EHR-system and the TNA and for long-term preservation. Every data object storing raw-data includes also a definition file (a metafile).

This file is comprised of the following information:

- identification codes of participants (e.g. of the EHR-system and of the patient);
- archive policy under which submitted objects should be handled including data preservation time;
- used classifications, codes and formats;
- security policy associated with this data object;
- context and purpose associated with this data object.

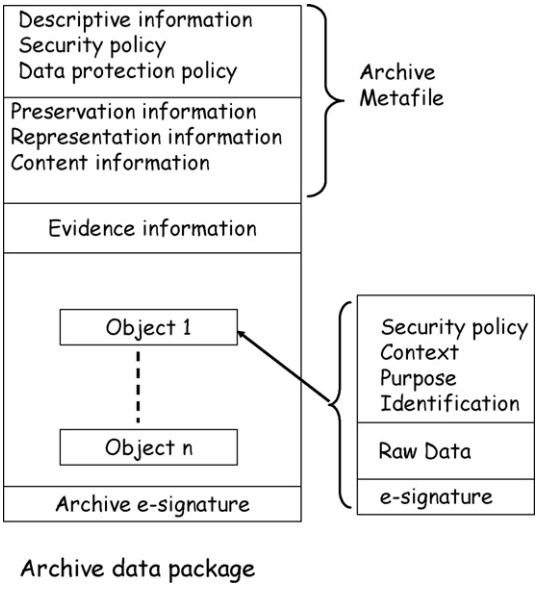

Fig. 2 – Model of archive data package.

In this model, a description file is included in the content of each data object. This file includes also information on how different objects of an object's group are associated with each other.

Before sending a data object to the TNA for preservation, the EHR-system (a data submitter) must:

- collect data objects;
- calculate the container identification code (e.g. hash-code);
- write the object description file associated with that data object;
- put them all together into a single data container.

Once the stages described above have been completed, the EHR-system must then sign the whole container [3]. On completion, the EHR-system then sends the container to the TNA as a part of an archiving request message.

5.1. *Basic archiving process*

Before technical preservation, the TNA receives an archiving request message, which includes the data container, and transforms data objects and associated definition files into an archive data package. The next step is to build an archiving metafile for received objects. This file is a combination of archiving policy, preservation, data representation and data content information [2,6]. The TNA also collects events, produces archive-Time-stamps, formulates the evidence record, and stores objects, metafile and evidence record into the archive package and finally signs the package [3]. Archive data packages can be stored in XML-format [10].

During the preservation period, the TNA needs to periodically renew archiving Time-stamps to prove the non-repudiation of data [3]. The TNA should also manage a secure audit-log, which includes information of all events associated with storage and data disclosure activities. It is also necessary to send an acknowledgement to the EHR-system to confirm that the correct data was accepted for preservation.

5.2. *Secure data distribution*

The TNA can distribute both single data objects and groups of data objects. Objects will be distributed in data distribution packages. Such a package is similar to an archive data package; however the TNA has to add archive identification information and a Time-stamp to this package to confirm the distribution time. This fine grain data distribution mechanism makes it possible to collect data objects sent by different sources and distribute them at the same time.

For data distribution, the TNA has two basic service models. The first model is used where an EHR-system is accessing data objects it has previously sent to the TNA (i.e. the EHR-system is trying to get back objects it has previously sent). The second model is used in cases where an external user is trying to access data objects stored on behalf of another EHR-system.

In the first case, the TNA and the EHR-system must have the same level of security policy. In this case, the EHR-system is using data received from the TNA in the same way that it is using its own local data and there is no need to ask patient's consent for data access. To initiate the data transmission, the EHR-system sends a data distribution request to the TNA. This message includes an identification of the EHR-system and necessary information for retrieval of needed data objects. On receipt the TNA collects the required copies of the data objects and returns them back to the EHR-system for further use.

In the second case, an external user or process is trying to access data objects stored in the TNA. In this case, a more elaborated privilege management and access control mechanism is needed. After identification of the requester, the TNA must first check the security policy of the requester. The next step is to analyse the access request message to ensure that it contains all the information necessary for the TNA to either allow or deny data access. The access request message should contain the following information [11]:

- patient identification;
- identification of data requester and his/her functional and structural roles;
- purpose for and context in which the data will be used;
- which data objects or object categories are planned to be accessed;
- patient's consent.

To allow data access for any external user, the TNA must have a privilege management and access control service that is capable of making decisions on security policy, purpose and context of data, role, consent and rule based access [12,13]. Access rules should also cover special conditions where above mentioned access conditions can be overruled (e.g. where access is allowed by legislation).

5.3. *Partial data delivery*

The TNA can distribute either a single data object or a collection of data objects. Unfortunately, many present EHRs do not have a well defined granular data structure and they can only send the entire EHR as a single object to the TNA. Partial delivery happens in situations where the data user wants to access only some part of the EHR, which has been stored as a single object. To make data access possible in this kind of situation, the TNA must include a partial delivery service capability.

In case of partial delivery, the TNA must collect the requested data items from the archived EHR and transform these into a new data object. The TNA must also write a metafile for this new object and add a "Partial_delivery" tag to it. The next step is to construct an evidence record to be associated with the new data object and add a Time-stamp. Finally, the TNA has to put all this information into a data container, sign it using archive signature and send the container to the user.

The TNA must also archive all distributed partial data objects having the tag "Partial_delivery" into a separate sub-archive. These data objects should have the same preservation period assigned as the original EHR from which they were extracted.

5.4. *Data updating process in a TNA*

The EHR-system uses data objects retrieved from a TNA for clinical or other legal purposes. During a patient's care

episode, some objects are modified and new objects will also be created as well. After the end of a care episode, the patient's EHR will be updated with those modified and new data objects and the updated EHR will be sent to the TNA. From TNA point of view, an EHR looks like a dynamic document [14].

In this case, the TNA receives the updated patient's record and marks it as the newest version. This most recent version will be signed and stored by the archive. The newest version is used for future data distribution purposes and older versions are stored into a history archive for verification purposes. The benefit of this method is that the TNA has only to manage cascading signatures instead of nested signatures.

6. Discussion

Secure long-term archiving of personal health information in digital format is an essential issue that needs to be solved over the next few years. The two key problems that have to be solved are:

- how to archive the non-repudiation of archived data;
- how to prove the integrity of stored information after digital migrations.

The trusted notary archive (TNA) is a promising answer to those problems.

The TNA model described in this paper is targeted to receive granular EHR-data from different EHR-systems; store data objects for long periods and distribute them as legitimately requested. The TNA communicates with EHR-systems and external users via archive request and distribution messages. The TNA can store data objects in XML-format and prove the validity of any events that have taken place over the regulated preservation period. This TNA model has also an additional benefit in that only the keys used by TNA have to be stored for this period. As a result, it is not necessary to store personal signature keys of health care professionals [3].

As described in the TNA model, a patient's health information received from different sources is stored in a single personal data folder, which in the future can form the patient's life-long personal health record.

In the long run communication based purely on these types of messages will not be sufficient. In future, cross-organisational seamless care, mobile health professionals and patients will require dynamic access to granular data. This could be facilitated by a future expansion of this TNA model; however, it will require a dynamic privilege distribution mechanism that has to be developed [15].

The ability to store data objects that may have differing purposes, contexts and security policies makes it possible not only to store official health records, but also other information, such as documents or data that is produced and owned by the patient himself (e.g. home measurements and personal health status monitoring reports).

REFERENCES

[1] P. Ruotsalainen, Security Requirements in EHR-systems and Archives, Medical Care and Compunetics 1, Studies in Health Technology and Informatics, vol. 103, IOS Press, 2004, pp. 453–458.
[2] P. Ruotsalainen, Archiving data: how to do it? Who can access? Tutorial SP6, in: Sixth Annual European Health IT Conference and Exhibition (TEHRE 2001), 11–14 November, 2001.
[3] L. Wallace, Long-term Archive and Notary Services, LTANS, LTAP, IETF (www.ietf.org/internet-drafts/drafts-ietf-ltans).
[4] Digital Archiving Strategies for Regulatory Compliance in Health Care, White Paper, Archiva, Inc. (www.archivas.com).
[5] ISO/TC 215, Health Informatics—Security Requirements for Archiving of Electronic Health Records, ISO/PDTS part1 Principles and Requirements, 15-09-2006.
[6] Consultative Committee for Space Data Systems (CCDS), Reference Model for Open Archival Information System (OAIS), RED BOOK, Washington, DC, U.S.A, June 2001.
[7] M. McKeon Stosuy, B. Manning, "Joining Up" e-Health and e-Care Services, Medical Care and Compunetics 2, Studies in Health Technology and Informatics, vol. 114, IOS Press, 2005, pp. 65–81.
[8] The *tScheme* Guide to Securing Electronic Transactions, *tScheme* Ltd., September 2002 (www.tScheme.org).
[9] Trusted Third Party Management: *tScheme* and Confidence in Online Identity, *tScheme* Ltd., September 2004 (www.tScheme.org).
[10] H. Peterson, Long-term storage of electronic healthcare information in XML format, The PARK Project, Sweden, 2000.
[11] E. Coiera, R. Clarge, e-Consent: the design and implementation of consumer consent mechanism in an electronic environment, J. Am. Med. Assoc. 11 (March/April (2)) (2004).
[12] B. Blobel, Analysis, designs and implementation of secure and interoperable distributed health information systems Studies in Health Technology and Informatics, vol. 89, IOS Press, 2002.
[13] Enterprise Privacy Authorization Language (EPAL 1.1), IBM Research Report, IBM 2000–2003.
[14] R. Ruusalepp, RIKSARKIVET, digital preservation in archives: an overview of current research and practices, Estonian Business Arch. (2005).
[15] J.M. Gardier, Identity federation – introduction, value and evaluation, ISSE 2005 – securing electronic business processes, in: Highlights of the Information Security Solution Europe 2005 Conference, Vierweg, 2005.