Evaluation of Healthcare Institutions for Long-Term Preservation of Electronic Health Records

Juanjo Bote[1], Miquel Termens[1], and Gemma Gelabert[2]

[1] Universitat de Barcelona, 08014-Barcelona, Spain
[2] Hospital Sant Joan de Deu, 08950-Esplugues de Llobregat, Spain
{juanjo.botev,termens}@ub.edu, gemma@hsjdbcn.org

Abstract. An evaluation of health institutions using the Trustworthy Repositories Audit and Certification (TRAC) is presented. TRAC is an audit methodology for information systems to evaluate its ability to preserve digital information securely over the medium and long term. With this methodology, different healthcare organizations in the metropolitan area of Barcelona (Spain) have been analyzed to determine their capacity for long-term preservation of the Electronic Health Records (EHR). From these results it is expected to propose a model of long-term preservation of the EHR. This paper concludes with lessons learned regarding the implementation of TRAC in healthcare organizations.

Keywords: digital preservation, audit, health organization, EHR, TRAC.

1 Introduction

This paper presents an ongoing work of the situation of different Healthcare Organizations in order to preserve their Electronic Health Records (EHR) on the long term. A survey was conducted in seven healthcare organizations of the metropolitan area of Barcelona (Mútua de Terrassa, Consorci Sanitari de Terrassa, Corporació Sanitària del Maresme, Consorci Sanitari del Maresme, Institut Universitari Dexeus, Grup Pere Mata and Grup SAGESSA). Since all of these organizations are private entities, they facilitate services to the public health network. All of them have a total of 25 hospitals covering a wide range of different specialties and population.

All entities analyzed were doing diverse initial processes or they are finalizing them. Processes are relative to the use of fully integrated digital EHR management within their Information and Communications Technologies (ICT) system. The first and most basic process is the conversion of analog data and paper health records into a digital format by digitizing processes. The second process is the development of an ICT system to manage EHR which are completely digitally born. The last strategy is the integration of the previously digitized health records with those created digitally on the computer system. Thus, all data will be digital objects to be managed by healthcare organizations, with different challenges such as electronic health records management or long-term digital preservation.

Due to this reason the minimum conditions to preserve digital information have been analyzed over these seven institutions. To carry out the study, it was applied the Trustworthy Repositories Audit and Certification (TRAC) methodology. TRAC [8] is

M.M. Cruz-Cunha et al. (Eds.): CENTERIS 2011, Part III, CCIS 221, pp. 136–145, 2011.

a qualitative methodology for information systems audit, designed to assess a repository to be trustful to retain information over the long term. To analyze the information system, TRAC analysis is divided in three areas as Organizational Infrastructure, Digital Object Management and Technology assessment. The TRAC analysis was performed by checking that there are appropriate processes running on to the organization or processes are clearly documented. TRAC result is a report that can detect improvement areas to get a reliable information system in order to preserve digital objects over time.

Once report advertences have been improved according TRAC guidelines it is possible to start designing a digital archive for long-term preservation or apply report enhancements onto the information system in the organization. Improvement applications report can also be carried out whether the repository exists. Open Archive Information System (OAIS), ISO 14721:2003 is a reference model to develop an archive to preserve the long-term records. OAIS reference model [11] does not define a particular technology, but emphasizes on both the information model definition and the Designated Community definition. The information model indicates how data to be kept is going to be represented onto the OAIS archive. The Designated Community is a system, person or agent responsible to receive and understand information to be retrieved.

EHR can be retained over the long-term by health care organizations through the use of TRAC audit and lately developing an OAIS model on its medical archive. Electronic health record is the health data representation of a patient to the course of his life. This data are represented by different expressions of data such as text, audio [1], graphics or video. According to Spanish law regulations, EHR must be preserved on the long term by an indefinite period of time. To accomplish this challenge, hospitals will have to organize processes in a near future in order to maintain long-term digital medical and research information. Institution analysis has been carried out through process assessment or documentation belonging to their clinical services documentation and archiving unit or their ICT unit. In some cases it was necessary to evaluate both units.

The document is organized into the following sections. Section 2 explains how TRAC can be performed, its relationship with digital preservation and EHR digital preservation implications. It also explains the methodology applied to healthcare organizations. Section 3 explains the results over the analyzed healthcare institutions that participated in the assessment. Finally, conclusions are exposed.

2 TRAC Methodology

2.1 Related Work

TRAC methodology is fairly new compared with other audit methodologies and there is a lack of published reports. However, some previous work on trusted digital repositories has been proved. Thus, the Network of Expertise in Long-Term Storage of Digital Resources (NESTOR) developed a catalogue for trusted digital repositories for long-term preservation, addressed to German cultural heritage organizations [3].

The Centre of Research Libraries has published two reports where two preservation audits have been performed. The first one is related to Portico Archive[1] a scholarly preservation service where its main objects ingested are journals, books and scholarly content. The second one is related with HathtiTrust[2] a large digital repository with more than 7.5 million digital objects ingested. Its primary objects ingested are digitized books. Both reports agree that Portico [9] and HathtiTrust [10] are trustworthy repositories to the general need of the CRL community.

It is also possible to find cases where TRAC is not only used for digital preservation but for other scenarios. TRAC is then applied to a repository that is not dedicated itself to digital preservation. In this case all the standards relating to audit of preservation [13] cannot be applied.

2.2 Performing TRAC Methodology

TRAC methodology was conducted to these entities through a survey and later interviews to different staff involved on the seven entities. Personnel involved in EHR management and archiving had different responsibilities and professional profiles, like ICT managers or medical archivists.

Trusted Repositories Audit & Certification: Criteria & Checklist (TRAC) is an audit qualitative methodology to be applied over a digital repository or an information system to retain information on the long-term. To do this, TRAC has indicators to be assessed onto the three sections mentioned above. All indicators use the vocabulary of the Open Archival Information System (OAIS) reference model, ISO 14721:2003.

This methodology is open enough about in reference to other rules that may help a repository to be trustful. This can occur in cases of special communities, as medical communities would be. Relevant standards such as ISO 9000, ISO / TS 21547:2010 or ISO 15489-2:2001, additional audits are tools that facilitate the audit with the TRAC methodology.

TRAC assumes that the main object to analyze is a repository where information is preserved or is going to be retained on the long term. It is also possible to apply TRAC methodology even whether a repository does not exist into the organization to be assessed. Therefore, TRAC result assessment is the conformance of a set of criterions to be analyzed. The conformance of theses criterions is verified by the existence of evidences, running processes or documentation. Repository has functions such ingest, management or access functions which prepare information for the Designated Community.

The Designated Community is a repository element responsible for receiving the information being retrieved to understand it without the need of technical support.

A digital repository is an information system that allows the preservation of long-term data through appropriate policies. Operations that can be performed in a repository can be started by data ingestion and can be finished on the retrieved information certifying that it is an exact copy of the original ingested. A series of intermediate steps will permit digital object management processes. There are several initiatives in repositories like DSpace[3] or Fedora[4] [7] where OAIS model has been implemented.

[1] http://www.portico.org
[2] http://www.hathiturst.org
[3] http://www.dspace.org
[4] http://www.fedora-commons.org

Such repositories have been widely used in digital libraries [6], but it is possible that a healthcare institution needs to apply another kind of technology also valid to long-term preservation. The reason is because it is possible that EHR management systems are technologically more advanced and possibly safer while DSpace or Fedora does not offer a complete solution preserving digital information. To carry out a repository analysis, TRAC analysis divides the information system in three sections. These sections are Organization Infrastructure, Digital Object Management and finally Technologies, Technical Infrastructure and Security sections.

All of three sections have a sum of 84 criterions to assess an organization. According to sections, Organization Infrastructure has 16 indicators, Digital Object Management has 44 indicators and Technical Infrastructure and Security with 16 respectively.

Organization infrastructure section emphasizes over legal or legislative mandates, regulatory requirements, structure and staffing organization, repository records retention strategy and financial sustainability. Although TRAC is in the process of becoming an international standard (ISO/TC20/SC13), many indicators in this section may be subjected to national law or Regulations.

Indicators belonging to the Digital Object Management section point on the repository ingest process, preservation strategies and the accurate information production and dissemination of the authentic versions of the digital objects. Assessment of this section is particularly relevant, because in order to be analyzed the structure and design of a preservation plan should exist. In case of a healthcare organization a preservation plan is the design of the processes involving the EHR long-term digital preservation. These processes range from the definition of the information model to be preserved to the design of the Designated Community or archivable information packets generation.

In the third section Technologies, Technical Infrastructures and Security indicators assess the audit over the technology adequacy, accuracy such as to detect bit corruption or loss. Security issues are also assessed verifying service continuity plan, risk assessment [12].

These indicators are verified by testing processes, analysis of documentation or analysis of evidences of the information system that belongs to the institution. After analyzing requirements for indicators information about the certification status and auditing of the repository is reported.

TRAC has also a minimum set of compliance indicators. These indicators are the minimum documentation or processes available that an institution who wants to keep long-term data should have. These minimum set of indicators, 6 belong to the Organization Infrastructure, 6 to Digital Object Management and 7 to Technologies, Technical Infrastructure and Security. Some of these indicators like those belonging to Organization Infrastructure section may be determined by the legal regulations in some countries and its verification can be more accessible. As mentioned earlier, according to TRAC guidelines it is not necessary to carry out the assessment to an organization without a repository for long-term preservation. However, there must be minimum conditions necessary to preserve information.

If the repository concept is applied to a healthcare institution, the repository would be part of the archive from a hospital. A hospital archive has active health records and passive health records. Active health records are those that their data are being updated

frequently and therefore correspond to patients receiving care at the center. Passive records are not updated records over a reasonable period of time. In all the assessed organizations, an analog health record becomes part of the passive archive within 3 to 5 years on average provided they have completed the specific patient care. This situation suffers changes when the system becomes digital because information workflows and treatment is different. Currently, EHR are lying in the same computerized system all together being an active system constantly, but not necessary running under the same software. This means that interoperability through the standard HL7 is implemented to access EHR data [4]. This policy would have to be necessary changed as soon as the volume of information would be measurable in terms of Petabytes and information should have been kept in different silos separated from the active system. This situation may occur because the EHR have information attached such as DICOM radiographic images [5], documents in Portable Document Format (PDF), test information, audio, video and other digital elements. Therefore an EHR is a complex unit of information that requires different treatment in the long term, not as a single unit of information.

In any case, the concept of passive digital archive should not change whether a proper digital preservation planning [14] is done. However, changes in the concepts of ingest or retrieval information into the repository has to come about. EHR information to be preserved over the long term would be information that is already ingested and audited by the internal hospital processes. Therefore, this information already exists, is audited and should have the necessary elements for its preservation. These elements can be a metadata wrapper or another structure to identify information, according to the information model defined by the institution that retains the data.

3 Results

Due to the length of the data obtained, the analysis of the necessary minimum indicators will be presented. Analysis is reflected in three tables where rows are the assessed TRAC indicators and the columns (HC1 through HC7), the data belonging to the healthcare organizations processed to ensure anonymity for reasons of confidentiality. Table 1 shows the data of Organizational Infrastructure. In Table 2, data are corresponding to the Digital Object Management and data in Table 3 corresponds to the section on Technology and Security structure. As mentioned before, TRAC methodology consists on checking the evidences and documented processes of an organization. Thus, the result of TRAC assessment is the conformance of the correspondent indicator. This conformance is reflected in Table 1, 2 and 3 by a "+" sign. When criterion is not conformed or it was not possible to be assessed the result is reflected by a "-" symbol.

3.1 Organization Infrastructure Analysis

This section is responsible for analyzing the repository attributes affecting their performance. Thus, issues such as financial sustainability, preservation policies and strategies, were part of the objectives to be tested. Other aspects can be transparency in the documentation for the repository, regulations or international standards that also meet the information system, even without a direct connection with digital preservation, but whether it affects and organizational issues.

Table 1. Organizational Infrastructure minimum assessed indicators

Criterion	HC1	HC2	HC3	HC4	HC5	HC6	HC7
A1.2 Contingency plans, succession plans, escrow arrangements	+	+	+	+	+	+	+
A3.1 Definition of designated community(ies), and policy relating to service levels	-	-	-	-	-	-	-
A3.3 Policies relating to legal permissions	+	+	+	+	+	+	+
A3.5 Policies and procedures relating to feedback	+	+	+	+	+	+	+
A4.3 Financial procedures	+	+	+	+	+	+	+
A5.5 Policies / procedures relating to challenges to rights	-	-	-	-	-	-	-

Evidences examined in this section have been ICT strategic plans, training plans, staff development plans, missions and goals of the organizations and legislative requirements.

In Table 1 indicators A3.1 and A5.5 were not assessed. A3.1 indicator is associated to the definition of Designated Community and policies in place to dictate how its preservation requirements will meet. The reason for this result is that although the Designated Community by default will be the professionals who work for the healthcare institutions, there are no policies for long-term digital preservation planned.

A5.5 is an indicator that is not necessary compulsory to be accomplished by a healthcare institution as points over unclear ownerships / rights on digital content. It is clear that all information is generated by the healthcare institution. It is the institution who has the data ownerships. These data has been introduced on the Electronic Health Records, but patient has the right to access information to its own data.

The rest of indicators who are accomplished by all of them are related with ensuring the continuity of the information system (A1.2), acquisition of legal permissions required to preserve digital content over time (A3.3), quality assurance records (A3.5) and evidences of financial audits already taking in place (A4.3).

3.2 Digital Object Management Section Analysis

Digital Object Management section is responsible for analyzing all processes related to data retention processes within the repository. This means to evaluate the consistency of the digital data stored with the information model defined by the repository.

Ingest information management and recovery processes are also discussed in this section. The major part of indicators to be evaluated are all related with the functional entities of the OAIS model, ingest, data management, archival storage, administration, preservation planning and access. Concerning the minimum indicators, this section is the one where there are the most indicators to be evaluated. This section has a total of 13 indicators. Indicators to be evaluated are all indicators of the B1 section (B1.1 through B1.8). None of the entities met the requirements because preservation of passive records is not still planned.

Table 2. Digital Object Management minimum indicators rated

Criterion	HC1	HC2	HC3	HC4	HC5	HC6	HC7
B1 Procedures related to ingest	-	-	-	-	-	-	-
B2.10 Process for testing understandability	+	+	+	+	+	+	+
B4.1 Preservation strategies	-	-	-	-	-	-	-
B4.2 Storage/migration strategies	-	-	-	-	-	-	-
B6.2 Policy for recording access actions	+	+	+	+	+	+	+
B6.4 Policy for access	+	+	+	+	+	-	+

The main evidences conformed in this section have been processes involved on the ICT department like security processes or access policies.

In Table 2 there are just three indicators accomplished by the assessed institutions. As mentioned earlier, B1.1 to B1.8 are indicators related to the digital object workflows, preservation properties identification of each object or the use of the appropriate technology to correctly identify the digital objects to preserve. B2.10 is the indicator that has been experienced by all entities it is relative with process for testing understandability of the information content. The reason to be clearly accomplished is because current information is retained by individuals with discipline expertise. The Designated Community is also clearly well defined. Physicians or medical archivist are in charge of introducing or managing information. Indicator B4.1 is relative to the existence of documented preservation strategies or the evidence of its application. It is not professed because as metadata are generated in most cases by these healthcare entities, none of them generates preservation metadata, such as PREMIS [15]. B4.2 indicator is relative over the demonstration that a preservation strategy has been performed or the use of the appropriate metadata is done. B6.2 indicator is relative on the recorded actions on the access on the repository. The evaluation of this indicator has to be constant, because it depends on the institution to track user actions over the information accessed. All of these institutions track their users as a preventive action to avoid mistaken usage on their systems. According to this question, maybe some policies over tracking user would have to be modified.

The other accomplished point B6.4 is related to the access validation mechanism within the system. This means that, information stored should be accessed and being protected against deliberate or accidental damage. But in specific communities as health care communities are, the use of user credentials is important on accessed information to avoid access or personal data unprotected.

Since digital preservation of EHR is not yet planned in these institutions, most of these points are difficult to be accomplished. Another reason is that on their ICT systems there are active records while passive digital records no longer exist.

3.3 Technologies, Technical Infrastructure and Security Analysis

This section is responsible for assessing the safety and information integrity. It is also in charge of analyzing the technical structure that guarantees the security. Safety criteria used for digital preservation are similar to ISO 17799 and must be therefore considered. Evaluation of trusted and secure infrastructure is a common point to the security of an ICT system that a specific evaluation applied to digital preservation.

As it is noted on Table 3, the major parts of criterion are fairly accomplished by healthcare institutions. This result is partly to the legal requirement to these entities to have secure systems on managing EHR. Evidences examined in this section were technical specifications, security processes and ICT documented protocols.

Security structure form is similar to all of them because none of them had access outside the institution electronically. C1.7 indicator evaluates evidences or documentation such as hardware manufacturers, polices related to hardware support. Indicator C1.8 is relative to the accomplishment of having documented changes management processes. This means having documented process related to the six functional entities of the OAIS model as mentioned on the Digital Object Management section. Indicator C1.9 is relative to have documented processes for testing critical changes to the system. This means the replacement of software or hardware and its later monitorization. As this point just 5 entities get the result.

Four of the healthcare entities are succeeded over C1.10 indicator which is relative to react to software security updates based on a risk-benefit assessment. This means to have carefully documented any updates of the security software.

Relative to indicator from C2 section, appropriate technologies which are relative to technology watch, hardware or software inventory, most of the healthcare entities have these processes perfectly coordinated or are getting them.

Table 3. Technologies, Technical Infrastructure and Security evaluated indicators

Criterion	HC1	HC2	HC3	HC4	HC5	HC6	HC7
C1.7 Processes for media change	-	+	+	+	-	-	+
C1.8 Change management process	-	-	-	-	-	-	-
C1.9 Critical change test process	-	+	+	+	-	+	+
C1.10 Security update process	-	+	+	-	+	-	+
C2.1 Process to monitor required changes to hardware	+	+	+	+	-	+	+
C2.2 Process to monitor required changes to software	+	+	+	+	+	+	+
C3.4 Disaster plans	+	+	+	+	+	-	+

The last indicator, C3.4 belongs to the security section is related to the evidence of ISO 17799 certification, disaster and recovery plans of having a backup of the

preserved information with a copy of the recovery plan. All organizations accomplished the requirements unless two who are doing modifications and by the time they were assessed, most modifications were being done.

4 Conclusions and Further Development

Our conclusions to this paper are lessons learned from assessing these healthcare entities. Analysis results are not very optimistic about the degree of development or current implementation of digital preservation processes since digital preservation is not yet a priority. Health care organizations will be affected by legal obligations in the near future. This fact will force them to develop new policies on data curation and making changes into their structures. These changes will suppose them a huge effort especially in hardware and software acquisition or staff training. Digital preservation is not yet a worrying problem while most institutions are still developing complete digitally born ICT systems using electronic health records. To have good results on assessment is essential to count on most people organization help. Understanding on the importance of digital preservation techniques and its implication over organization is also important. Organizational structure assessment in most cases comes by mandatory obligations or legal regulations. On evaluating the Digital Object section it is important to have at least a documented preservation plan to be assessed accurately. Relative to technologies, technical infrastructure and security section is the most probable section to be well assessed. The reason is that the importance of information to be kept in healthcare organizations is essential to their core business while information is active. When information is no longer active, it must be well preserved according to a preservation plan.

Further work resides on the application of a risk analysis methodology based on a recognized international standard emphasizing on digital preservation issues. Since there is an existing methodology [2], all aspects are not covered.

References

1. Amano, S., Kondo, T., Kato, K., Nakatani, T.: Development of japanese infant speech database from longitudinal recordings. Speech Communication 51(6), 510–520 (2009)
2. Digital Curation Centre and DigitalPreservationEurope: Digital Repository Audit method Based on Risk Assessment, `http://repositoryaudit.eu` (retrieved April 20, 2011)
3. Dobratz, S., Schoger, A.: Trustworthy Digital Long-Term Repositories: The Nestor Approach in the Context of International Developments. In: Kovács, L., Fuhr, N., Meghini, C. (eds.) ECDL 2007. LNCS, vol. 4675, pp. 210–222. Springer, Heidelberg (2007)
4. Hammond, W.: eHealth Interoperability. Paper presented at CeHR International Conference 2007, Regensburg, Germany (2007)
5. International Standard Organization: Health informatics – Digital Imaging and communication in medicine (DICOM) including workflow and data management. ISO 12052 (2006)
6. Misra, D., Mao, S., Rees, J., Thoma, G.: Archiving a historic Medico-legal collection: Automation and workflow customization. Paper presented at the 4th IS&T Archiving Conference, Arlington, VA, USA (2007)

7. Ramahlo, J., Ferreira, M., et al.: RODA and Crib. A Service-Oriented Digital Repository. In: The Fifth International Conference on Preservation of Digital Objects, London, UK (2008)
8. The Center for Research Libraries, Online Computer Library Center, Inc.: Trustworthy Repositories Audit and Certification: Criteria and Checklist (2007), http://www.crl.edu/PDF/trac.pdf (retrieved April 20, 2011)
9. The Center for Research Libraries: Report on Portico Audit Findings (2010), http://www.crl.edu/archiving-preservation/digital-archives/certification-and-assessment-digital-repositories/portico (retrieved June 22, 2011)
10. The Center for Research Libraries. Certification: Report on the HathiTrust Digital Repository (2011), http://www.crl.edu/archiving-preservation/digital-archives/certification-and-assessment-digital-repositories/hathi (retrieved June 22, 2011)
11. The Consultative Comittee for Space Data Systems. Reference Model for an Open Archival Information System (2002), http://public.ccsds.org/publications/archive/650x0b1.PDF
12. Smith, A., Eloff, J.: Security in health-care information systems—current trends. International Journal of Medical Informatics 54, 39–54 (1999)
13. Steinhart, G., Dietrich D., Green A.: Establishing Trust in a Chain of Preservation. The TRAC Checklist Applied to a Data Staging Repository (DataStaR). D-Lib Magazine 15(9) (2009)
14. Strodl, S., Becker, C., Neumayer, R., Rauber, A.: How to choose a digital preservation strategy: Evaluating a preservation planning procedure. In: Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries, Vancouver, BC, Canada, pp. 29–38 (2007)
15. Preservation Metadata: Implementation Strategies (2005), http://www.loc.gov/standards/premis/ (retrieved April 20, 2011)