ELECTRONIC HEALTH RECORDS STORAGE: A SYSTEMATIC REVIEW

José Ferreira de Souza Filho[1], Sergio T. Carvalho[1], Juliana Pereira de Souza-Zinader[1]

[1]Instituto de Informática (INF), Universidade Federal de Goiás (UFG), Brasil

Abstract: Objective: This systematic review aims to identify how data storage has been achieved in Electronic Health Record (EHR) systems. **Methods:** Following meta-analysis systematic review methodology, the authors reviewed papers published between 2000 and 2015 in PubMed, IEEE and ScienceDirect databases, which describes the storage of EHRs. Additionally, an inductive content analysis was performed to summarize the steps and methodologies followed in order to build EHR systems. **Results:** 633 articles were screened, 79 were selected for the full review and 32 have been elected for final review. These articles elected were analyzed to extract relevant information about EHR storage. It has been noticed through the reviewed articles that there is no standard and common way adopted for EHR storage. **Conclusion:** There is little available information about EHR storage, suggesting opportunities in the sense of design a methodology for best practices for EHR storage and retrieval.

Keywords: Eletronic Health Record, Information Storage and Retrieval.

***Resumo**: **Objetivo:** Esta revisão sistemática tem como objetivo identificar como o armazenamento de dados tem sido realizado em sistemas de RES (Registro Eletrônico de Saúde). **Métodos:** Seguindo a metodologia de revisão sistemática com meta-análise, os autores revisaram artigos publicados entre 2000 e 2015 nas bases de dados PubMed, IEEE e ScienceDirect, que descrevem o armazenamento de RES. Além disso, uma análise de conteúdo indutivo foi realizada para sumarizar os passos e metodologias seguidos a fim de construir sistemas de RES. **Resultados:** 633 artigos foram selecionados inicialmente, 79 foram selecionados para a análise completa e 32 foram eleitos para revisão final. Estes artigos eleitos foram analisados para extrair informações relevantes sobre armazenamento de RES. Foi observado, por meio dos artigos revisados, de que não há nenhuma maneira padrão e comum adotada para o armazenamento de RES. **Conclusão:** Há pouca informação disponível sobre armazenamento de RES, sugerindo oportunidades no sentido de se conceber uma metodologia de boas práticas para o armazenamento e a recuperação de RES.*

***Palavras-chave:** Registros Eletrônicos de Saúde, Armazenamento e Recuperação da Informação.*

Introduction

The Electronic Health Record (EHR) is an evolving concept, which has been the subject of many researches, models and specifications, for example, openEHR (www.openehr.org) and ISO 13606[1].

From the technical perspective, the two-level modeling approach is gaining relevance to develop EHR systems[2]. This architecture proposes a separate definition on two modeling levels: information and knowledge. The information level is provided by the reference model and represents the generic clinical data structures. The knowledge level is provided in the form of archetypes and templates, which defines specific representations and meanings of those data structures. Archetypes define clinical concepts and are usually built by domain experts. They are deployed at runtime via templates that specify particular groups of archetypes to use for a particular purpose, often corresponding to a screen form. Figure 1 presents the structure of such model.

Figure 1- Two-level modeling structure.

The first level mostly referred as information level or reference model (RM) defines the entities and properties that are not likely to change over time. This model must be generic enough to avoid modifications for supporting new characteristics or requirements from a clinical domain perspective.

The RM entities are the basic building blocks for the conceptual level. The archetype model (AM) allows defining clinical information models by constraining specific data structures of the RM, to support specific clinical use cases. Such definitions are called archetypes. Archetypes define the maximum data schema of a clinical concept. To attach a formal specification of the meaning of archetypes, they can be linked to clinical terminologies. This makes archetypes a powerful mechanism to define information structures with attached meaning that support semantic interoperability among systems. This fact changes the way health information systems are developed. Domain experts define the structure and element types of the domain concepts (making it possible to create new concepts or update the current ones), while the system developers are just concerned about creating instances that represent the data according to the RM and the archetypes and creating user interfaces for the templates[3]. The ISO 13606 series of standards follows a similar approach, but it is based on a different RM[1,4].

EHR systems using either openEHR or ISO 13606 need to efficiently store and retrieve archetype-based patient information, which it is not straightforward[5,6,7]. Although, archetype storage and retrieval is a challenge, there is very little available information of how the storage issue is addressed.

This systematic review aims to identify the trends used for EHR storage (also known as EHR persistence) in the researches published between January 2000 and September 2015. In particular, a systematic review and an inductive content analysis have been performed in order to learn about EHR storage and experiences in building medical systems. The question being addressed in this study is whether an emergent consensus (good practice) strategy for EHR storage exists, taking into account approaches that use ISO 13606 or openEHR approach, and, therefore, if it is possible to propose a common or unified storage approach.

The remaining sections of this paper are organized as follows. The second section describes the materials and methods used to define the systematic review protocol. The third section shows the results of our systematic review. The fourth section, in turn, presents a discussion about the results. Finally, the last section presents final considerations.

Methods and Materials

In order to perform this systematic review we defined a systematic review protocol containing the eligibility criteria, the exclusion criteria and terms that we would be observing following a systematic review methodology[8]. The eligibility criteria were:

- Papers with any of the following terms in their title or abstract: "electronic health records","Health Information Systems", "EHR", "electronic medical records", "Health Information Systems".
- Papers with the terms "storage" or "persistence".

- Papers published between "2000/01/01" and "2015/09/30".

When deciding on the search criteria, we preferred to have a broad scope focused on storage of EHRs, rather than searching for each of the specific EHR mechanisms that could be applied for storage, such as relational database, NoSQL, XML databases. We have performed the search in PubMed[1], IEEE Xplore[2] and ScienceDirect[3] databases.

Figure 2 shows the structure of the search performed according to the previously defined search criteria.

```
(
  "electronic health records"  OR "Health Information Systems" OR "EHR" OR
  "electronic medical records" OR "Health Information Systems"
)
AND
(
  "storage" OR "persistence"
)
AND ("2000/01/01":"2015/09/30")
```

Figure 2- Search terms used to perform the systematic review.

With the results of the search performed on the selected databases, a two-phase procedure was adopted for the systematic review. In phase 1, also known as screening, the title and abstract have been reviewed. The following exclusion criteria were adopted:

- Short Papers.
- Any paper that does not follow neither openEHR nor the closely related ISO 13606 standard.
- Articles not related with storage somehow (Interoperability, Security, Privacy).

In cases where limited information was available in the titles and abstracts, the papers were accepted for full review. In phase 2 (full review), we reviewed the paper thoroughly. The objective in this step was two-fold: to reject those papers that did not fit the purpose of this systematic review and, from only those papers that were finally accepted, to extract a set of data items and indicators to perform further analysis.

Results

As a result of the search, 633 papers were found, 20 of which were duplicates. In total 613 papers titles and abstracts were screened, and 79 of them were selected for full review, after which it was determined that only 32 papers contained relevant data for the objectives of this research. The summary of this systematic review process is presented in Figure 3.

Figure 4 shows the publication date distribution of the selected papers. Note that 2015 only included the period between January and September of that year.

1 http://www.pubmed.gov

2 http://ieeexplore.ieee.org

3 http://www.sciencedirect.com

None of the papers addressed the storage of EHR solutions directly and we tried to identify how storage is being achieved in each one of the papers. Some of them are using relational data model approach[9,10,11,12], although most of them do not detail the storage mechanism[13,6,14,15,16,17,18,19,20,21,22,23,24,25]. Other storage data model are XML[26], Data Warehouse[27] and NoSQL databases[7,28,29].

Most of the researches and EHR systems developed today use some kind of standard, mostly openEHR or ISO 13606. With these standards come a lot of architecture definitions, such as Clinical Document Architecture (CDA)[30], which is an exchange standard and does not define the persistence storage requirements for CDA documents. There is not yet any finalized openEHR specification of service interface to aid application developers in creating, accessing, and the storing the EHR content[29]. Although storage of EHR records has not being the subject of many researches, effective storage of electronic healthcare record is the key for statistics, analysis and further use[28].

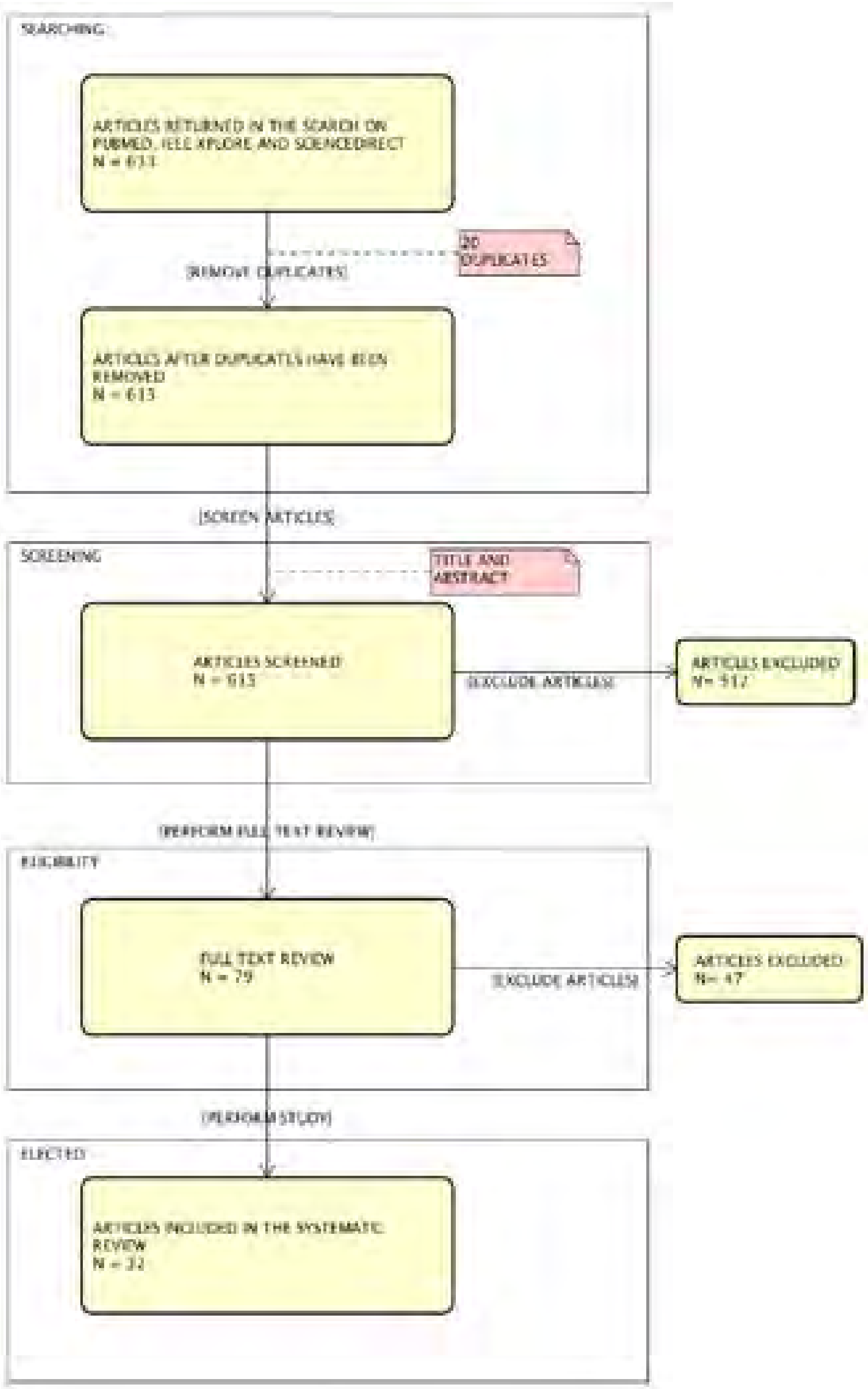

Figure 3- Systematic review process.

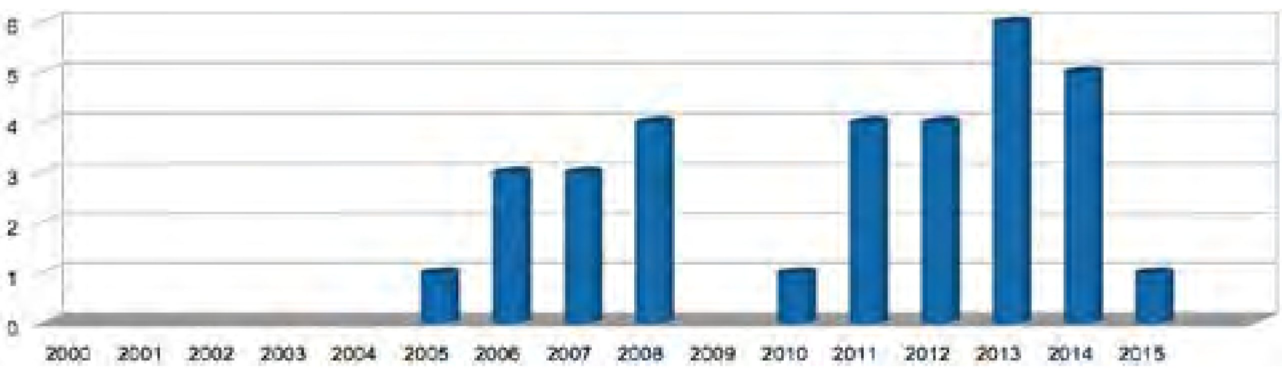

Figure 4- Publication date distribution of papers.

Discussion

This systematic review analyzed the selected papers searching for the storage mechanisms used. The focus was to determine if a standard or best practice exists for EHR archetype storage and watch closely the adoption of NoSQL or a non-conventional database approach for EHR storage (e.g., XML, object databases). Our reflection on the results of our publication searches confirmed that the decision not to include more specific search criteria was appropriate. Using a generic search without including specific terms for the types of storage proved to be successful, because it allowed the inclusion of an extensive range of storage experiences, using different technologies and standards. As we suspected before conducting this systematic review, storage or persistence is not the main concern of research in EHR. Yet, the focus of specifications such as openEHR and ISO 13606 is semantic interoperability, however, data interoperability is a challenging effort as well.

We have seen many papers addressing semantic interoperability, integration, communication, however, there is not yet much effort in how an EHR system based either on openEHR or ISO 13606 should do to store its records with their archetype tree-like structure. Once EHR relates to all the health-related data about the whole life of a person, it will contain a huge amount of data. Medical data must be available for large periods of time (for at least the entire lifespan of the patient), so there is a need for a lifelong solution. Otherwise there will be data loss because of the conversion from one system to another. Health information systems must be accurate, preventing data mixing, data loss and must be available at all times for authorized viewers.

NoSQL Databases – Archetype based EHR data is logically tree-structured[5], and different tree-structures appear every time a new archetype is used. Tree structures can of course be mapped to relational models[31], but you would in that case want to avoid methods that involve too many joins at retrieval time and avoid methods that need manual creation of new tables when new archetypes appear. In relational database models, it is needed to design many tables and a lot of fields to store EHR. When the whole EHR of a person is queried, it will involve the connection among the tables, and this will weak the system performance, in addition, there will be many empty values.

Although storage of archetypes using NoSQL solutions were not specified in the search string (Figure 2), we were interested in non-conventional approaches of storing archetype records, such as NoSQL or XML Databases. Some surveys have been published showing the increasing of NoSQL or XML databases[6,32] investigates NoSQL and XML databases for clinical models considering the performance of them over relational databases[33,34]. Data models created using NoSQL or other storage data model seems promising since the nature of archetype based systems do not fit directly into a relational database without transformation and conversion, what decreases more when the information needs to be retrieved, therefore store the archetype directly without conversions seems appealing, once the schema-less feature of non-conventional databases offers great flexibility when a new archetype arises, without the need to change the whole structure. Additionally, standards for archetype query have been developed (e.g., AQL[4]) to help retrieve records stored as archetype directly without any transformation prior to storage. Besides, a system must be designed to be scalable, resilient, open for extension, but closed for modification (open/closed principle)[35].

Conclusion

Clinical data are dynamic in nature, often arranged hierarchically and stored as free text and numbers. Effective management of clinical data and the transformation of the data into structured format for data analysis are therefore challenging issues in electronic health records development. Different standards and technical approaches exist (e.g., ISO 13606 and openEHR, using archetypes, or HL7 v3, using templates), but there is no formal definition of how store and retrieve archetype-based patient information. Thus, it is important clear and well-defined standards or best practices for archetype-based EHR storage.

This research characterized published experiences related to EHR storage between 2000 and 2015, in order to obtain a better understanding of the different approaches used for EHR storage. It was found that there is no consensus on EHR storage, specially for archetype-based systems. This suggests that it should be possible to create a common or unified methodology for future clinical information data modeling for systems that use archetypes. This conclusion is, however, limited, due to the selected papers' lack of detail on the storage details used and how the storage issue is addressed. Some measurements and implementation suggestions have been published[35]. Yet, there is no standard for archetype storage. Despite of the popularity of relational databases, the scalability of the NoSQL database model and the document-centric data structure of XML databases appear to be promising features for effective clinical data management over relational databases.

As part of our reflection on the set of selected papers, we also observed the cloud computing adoption[14,22,7,34]. Cloud computing is a promising platform for health information systems in order to reduce costs and improve accessibility. The cloud can provide several benefits to all the stakeholders in the health care ecosystem through systems such as health information management system, laboratory information system, radiology information system, pharmacy information system, etc. With public cloud-based EHR systems hospitals do not need to spend a significant portion of their budgets on IT infrastructure. Public cloud service providers provide on-demand provisioning of hardware resources with pay-per-use pricing models. Thus, hospitals using public cloud-based EHR systems can save on upfront capital investments in hardware and data center infrastructure and pay only for the operational expenses of the cloud resources used. The biggest obstacle, however, in the widespread adoption of cloud computing technology for EHR systems is security and privacy issues of health care data stored in the cloud, due to its out-sourced nature[7]. Therefore, we envision healthcare systems being created in the cloud, storing archetypes in non-conventional databases (NoSQL or XML databases) and providing semantic interoperability through services offered via cloud service providers.

References

[1] ISO 13606: Health Informatics Electronic Health Care Record Communication Part 1: Reference Model. International Organization for Standardization, p. 28, 2008.

[2] Beale, T. Archetypes constraint-based domain models for futureproof, information systems, In: OOPSLA, Workshop on Behavioural Semantics, , p. 1-18, 2002.

4 https://openehr.atlassian.net/wiki/display/spec/Archetype+Query+Language+Description

[3] Heard, S., Beale, T. openEHR Architecture Overview. OpenEHR Foundation, 2008.

[4] ISO 13606: Health Informatics Electronic Health Care Record Communication Part 1: Reference Model. International Organization for Standardization", p. 124, 2008.

[5] Frade, S., Freire, S., Sundvall, E., Almeida, J., Correia, R. Survey of openEHR storage implementation. In: Int. Simposium on Computer-Based Medical Systems. 2013; Porto, Portugal.

[6] Almeida, J., Frade, S., Correia, R. Exporting Data from an openEHR Repository to Standard Formats. Procedia Technology. 2014; 16:1391-1396.

[7] Bahga, A., Madisetti, V. K. A cloud-based approach for interoperable electronic health records (EHRs). IEEE Journal of Biomedical and Health Informatics. 2013;17(5):894-906.

[8] Moher, D., Liberatti, A., Tetzlaff, J., Altman, D. Preferred reporting items for systematic reviews and meta- analyses: the PRISMA statement. BMJ. 2009; 339:b2535.

[9] Alabbasi, S., Ahmed, A., Kaneko, K., Rebeiro-Hagrave, A., Fukuda, A. Data types managed database design for dynamic content: A database design for personal health book system. In: IEEE TENCON. 2014; Bangkok, Thailand.

[10] Kim, H., Tran, T., Cho, H. A Clinical Document Architecture (CDA) to Generate Clinical Documents within a Hospital Information System for E- Healthcare Services. In: *The Sixth IEEE Int. Conf. on Comp and Inf Tech. 2016; S*eoul.

[11] Muñoz, A., Somolinos, R., Pascual, M., Fragua, J., Gonzalez, M., Monteagudo, J., et al. Proof-of-concept Design and Development of an EN13606-based Electronic Health Care Record Service. Journal Amer Med Inf Association. 2007;14(1):118-129.

[12] Späth, M., Grimson, J. Applying the archetype approach to the database of a biobank information management system. Journal of Medical Informatics. 2011;80(3):205-226.

[13] Ghani, M., Wen, L., The design of flexible Pervasive Electronic Health Record (PEHR). In: IEEE Colloquium on Humanities, Science and Engineering. 2011; Penang.

[14] Almutiry, O., Wills, G., Alwabel, A. Toward a framework for data quality in cloud-based health information system. In: *Information Society. 2013;* Toronto.

[15] Anciaux, N., Berthelot, M., Braconnier, L., Bouganim, L., De la Blache, M., Gardarin, G. A tamper-resistant and portable healthcare folder. Journal of Telemedicine and Applications. 2008;763534.

[16] Andriuskevicius, E., Dobravolskas, E., Punys, V., Sinciene, V., Valentinavicius, A. Architecture for national e-health infrastructure in Lithuania. Studies in Health Technology and Informatics. 2006;124:421-6.

[17] Bamiah, M., Brohi, S., Chuprat, S., Ab Manan, J. A study on significance of adopting cloud computing paradigm in healthcare sector. In: Cloud Computing Technologies, Applications and Management. 2012; Dubai.

[18] Basu, S., Karp, A. H., Li, J., Pruyne, J., Rolia, J., Singhal, et. al. Fusion: managing healthcare records at cloud scale. *Computer. 2012;* 11:42-49.

[19] Dehelean, S., Sauciuc, D. G., Miclea, L., Hóka, I. EN13606 overview and compatibility with the Romanian healthcare system. In: Automation, Quality and Testing, Robotics. 2008; Romania.

[20] Grasso, G. M., Cuzzocrea, A., Nucita, A. A Patient-Centric Distributed Architecture for Electronic Health Record Systems. In: Network-Based Information Systems. 2014; Salerno, Italy.

[21] Hanzlicek, P., Spidlen, J., Heroutova, H., & Nagy, M. User interface of MUDR electronic health record. Journal of medical informatics. 2005; 74(2):221-227.

[22] Laohakangvalvit, T., Achalakul, T. Cloud-based data exchange framework for healthcare services. In: Computer Science and Software Engineering. 2014; Shenzhen, China.

[23] Prados-Suárez, B., Molina, C., Yañez, C. P., Reyes, M. P. Improving electronic health records retrieval using contexts. Expert Systems with Applications. 2012; 39(10):8522-8536.

[24] Sartipi, K., Yarmand, M. H., Down, D. G. Mined-knowledge and decision support services in electronic health. In: *International Workshop on Systems Development in SOA Environments. 2007; Washington, DC, USA*.

[25] Špidlen, J., Hanzlíček, P., Říha, A., Zvárová, J. Flexible information storage in MUDR II EHR. *journal of medical informatics. 2006; 75*(3):201-208.

[26] Haomin, L., Huilong, D., Xudong, L., Zhengxing, H. A clinical document repository for CDA documents. In: Bioinformatics and Biomedical Engineering. 2007; Beijing, China.

[27] Marco-Ruiz, L., Moner, D., Maldonado, J. A., Kolstrup, N., Bellika, J. G. Archetype- based data warehouse environment to enable the reuse of electronic health record data. Journal of medical informatics. 2015; 84(9):702-714.

[28] Jin, Y., Deyu, T., Yi, Z. A distributed storage model for EHR based on HBase. In: Information Management, Innovation Management and Industrial Engineering (ICIII). 2011; Hong Kong.

[29] Sundvall, E., Nyström, M., Karlsson, D., Eneling, M., Chen, R., Örman, H. Applying representational state transfer (REST) architecture to archetype-based electronic health record systems. BMC medical inf. and decision making. 2013;13(1):1.

[30] Dolin, R. H., Alschuler, L., Boyer, S., Beebe, C., Behlen, F. M., Biron, P. V. Shabo, A. HL7 clinical document architecture, release 2. Journal Amer Med Inf Association. 2006; 13(1):30-39.

[31] Wang, L., Min, L., Wang, R., Lu, X., Duan, H. Archetype relational mapping - a practical openEHR persistence solution. BMC Med Inform Decis Mak. 2015; 15(1):1.

[32] Lee, K. K. Y., Tang, W. C., Choi, K. S. Alternatives to relational database: comparison of NoSQL and XML approaches for clinical data storage. Computer methods and programs in biomedicine. 2013; 110(1):99-109.

[33] Freire S.M., Teodoro, D., Wei-Kleiner, F., Sundvall, E., Karlsson, D., Lambrix, P. Comparing the Performance of NoSQL Approaches for Managing Archetype-Based Electronic Health Record Data. PLoS ONE. 2016; 11(3): e0150069.

[33] Abbot, M., Fisher, M. The Art of Scalability. Addison-Wesley, 2009.

[34] Freire, S. M., Sundvall, E., Karlsson, D., Lambrix, P. Performance of XML Databases for Epidemiological Queries in Archetype-Based EHRs. In: Scandinavian Conference on Health Informatics. 2012; Linköping, Sweden.

[35] Tancer, J., Varde, A. S. Cloud technology and EHR data management. In: Information and Automation for Sustainability. 2012; Beijing, China.

Contato

Sergio T. Carvalho
sergio@inf.ufg.br