Professional practice article

Corresponding author:

Antti Vikström, Department of Future Technologies, FI-20014 UNIVERSITY OF TURKU, Finland. Email: anelvi@utu.fi

Secondary use of electronic health records – availability aspects in two Nordic countries

Antti Vikström[1], M.Sc. (Tech.)

Hans Moen[1], PhD

Sanaz Rahimi Moosavi[1], M.Sc. (Tech.)

Tapio Salakoski[1], PhD

Sanna Salanterä[2,3], PhD, RN

[1]Department of Future Technologies, University of Turku, Finland

[2]Department of Nursing Science, University of Turku, Finland

[3]Turku University Hospital, Finland

Abstract

Background

The potential for the secondary use of electronic health records (EHRs) is underused due to restrictions in national legislation. For privacy purposes, legislative restrictions limit the availability and content of EHR data provided to secondary users. These limitations do not encourage health care organisations to develop procedures to promote the secondary use of EHRs.

Objective

The objective of this study is to identify factors that restrict the secondary use of unstructured EHRs in academic research in Finland and Sweden.

Method

A study was conducted to identify these availability-restricting issues that pertain to the academic secondary use of unstructured EHRs. Using semi-structured interviews, 14 domain experts in science, hospital management, and business were interviewed to evaluate the efficiency of procedures and technologies that are implemented in secondary use processes.

Results and Conclusion

The results demonstrate three aspects that restrict the availability of unstructured EHRs for secondary purposes: (i) the management and (ii) privacy preservation of such data, as well as (iii) potential secondary users. Based on these categories, two approaches for the secondary use of unstructured EHRs are identified: the protected processing environment and altered data.

Implications

The protected processing environment ensures patient privacy by providing unstructured EHRs for exclusive user groups that have preferred use intentions. Compared to the use of such processing environments, data alteration enables the secondary use of unstructured EHRs for a larger user group with various use intentions, but that yield less valuable content.

Keywords

Electronic Health Records, Health Information Management, Patient Data Privacy, Secondary Use, Open Access

Introduction

Electronic health records (EHRs) are primarily maintained to support patients' health and care. Health care professionals use these records to document the patients' care, while maintaining the high quality of the whole care process. Proper documentation offers health care professionals the ~~a~~ possibility to evaluate care results and eases the process of setting new objectives for future treatment (Häyrinen et al., 2008). Storing this information in an electronic format enables the simultaneous use of EHRs throughout the entire hospital. The motivation to implement EHRs also includes the ability to conduct scientific studies and other forms of secondary use (Häyrinen et al., 2008).

Such an intention to provide original health information for secondary users can be justified by the shared benefits between the data provider and the corresponding EHR user. Secondary EHR users, such as researchers and business operators, gain authentic health information to produce and develop new business ideas and research opportunities. The results of successful secondary use can provide valuable information to the data provider, thus adding value to health information.

Due to the sensitive, patient-specific information that EHRs contain, these documents are considered confidential. Therefore, the variety of authorised users and use cases is restricted to direct use in the national health care setting. As a consequence, the possibilities of using health documents for secondary purposes are limited. These limitations are in conflict with the concept of secondary use, which refers to data utilisation for initially undefined purposes.

The lack of relevant data management procedures, together with a publicly justified motivation to provide personal data for secondary purposes, has been identified in European governments. In the European Union (EU), the General Data Protection Regulation (GDPR) has been updated and developed to unify national legislation and regulations concerning the protection of personal data in member countries (Regulation 2016/679, 2016). In the GDPR, the secondary use of personal data is allowed for historical, statistical, and scientific research, as well as to fulfill archiving purposes in the public interest. Finland and Sweden are both members of the EU~~,~~ and, therefore, they need to implement the contents of the GDPR within their own national legislation.

In Finland and Sweden, the national legislation sets requirements for the covered content in health records. The purpose of such requirements is to ensure good care for patients by storing all necessary information and documenting the entire care process, including conditions, treatment, and recovery (Allvin et al., 2011). The design of an EHR generally includes both structured and unstructured information. Structured data refer to systematically coded content in EHRs, including laboratory test results and diagnosis codes. Systematic coding enables the efficient utilisation of such records by health care professionals and clinical researchers (Häyrinen et al., 2008). The unstructured content, the so-called clinical text, is primarily narrative text that is created to describe patients' condition and outline their care by health care professionals (Allvin et al., 2011). The challenge of using unstructured narratives for secondary purposes considers the mismatch between natural language~~s~~ and structured health information in terms of systematic expression~~s~~. The lack of systematic coding in unstructured free text complicates the development of automatic analysis methods. Therefore, manual labor is generally required to analyse unstructured EHRs.

In our previous work (Vikström et al., 2016), a study was conducted to explore the availability affecting factors of EHR data used in Finnish health care for secondary purposes. In this article, the study is extended to include both Finnish and Swedish health care settings. The aim is to identify factors that restrict the secondary use of unstructured EHRs in academic research. The issues restricting secondary use are identified by conducting semi-structured interviews that discuss EHR-related procedures and technologies. The interviewees included experts in medicine, information technology, and business. The interview content is analysed using a qualitative content analysis process.

Background

The successful secondary use of EHRs is dependent on the quality of the documents that are included. Factors affecting the perceived quality of unstructured health information include data content and structure. Both of these factors can be compromised by modifying the characteristics of EHR data to preserve patients' privacy. The process of securing EHRs and the privacy of the patients does not exclusively include only technological elements. Furthermore, two other levels can be used: human and procedural (Merkow and Breithaupt, 2006). Protection of all three levels is a necessity to implement a secure system. However, additional perspectives can be employed to enhance privacy, especially in the health context. Furnell's proposal (2008) for a comprehensive security solution includes five essential perspectives to be considered: technology, physical environment, people, organisational procedures, and legislation.

In this article, the concept of privacy preservation considers the process of protecting patient-sensitive information for the secondary use of EHRs. The implementation of privacy preservation mechanisms depends on the recognition of sensitive and identifying information stored in EHRs (Neamatullah et al., 2008). Such information is typically defined as protected health information (PHI) (Meystre et al., 2014). A manually conducted identification process is considered laborious: therefore, automatic solutions are favored. Developing an automatic system to identify PHI from unstructured health information can be challenging. As a consequence, manual actions are generally required in the overall implementation (Meystre et al., 2010).

In the United States, the Health Insurance Portability and Accountability Act (HIPAA) (1996) defines 18 types of PHI, including names, dates, and telephone numbers. If all of the 18 types of PHI are removed or masked from EHRs, then the secondary use of these documents is possible without the individual consent of each patient. A comparable set of PHI types is not explicitly defined in Finnish or Swedish legislation.

Privacy preservation mechanisms that are based on the implementation of static sets of PHI identifiers, such as those stated in the HIPAA legislation, are considered more valid in the context of structured information rather than unstructured data. Structured information primarily consists of a finite number of fields and potential expressions, thus generating predictable data variations. For example, a structured EHR consisting of a patient's ID and diagnosis code has limited data variations due to the systematically coded content. Therefore, identifying sensitive content from a document using pre-defined PHIs is trivial. The identification process for unstructured clinical text is more complex. The complexity is caused by unpredictable variations in natural languages, which require context expertise to be identified. As a result, understanding data semantics is a necessity when developing automatic privacy preservation mechanisms for EHRs. The right to privacy also applies to clinicians and other medical professionals who record entries within the EHR. The name of the author is typically coded in a systematic manner in these entries for both structured and unstructured EHRs. The author names can then be conveniently protected via systematic coding. In the following, we present two privacy preservation mechanisms: (i) anonymisation and (ii) de-identification.

Anonymisation refers to actions that ensure the non-identifiability of information within the corresponding data set. Non-identifiability prevents the possibility of linking the information to explicit entities (Meystre et al., 2010). This process breaks the connection between health records and individual patients.

De-identification enables the possibility of protecting sensible health information without disconnecting the data from the corresponding patient. In this article, de-identification refers to the reversible process of deleting, suppressing, generalising, or masking explicit identifiers in the processed data collection (Neamatullah et al., 2008). Reversibility ensures that the information can be linked back to the original entity, unlike in anonymisation.

Methods

The premise for this study was based on analysing academic institutions that process unstructured EHR data for research purposes. These institutions should have collaborating business partners to support this study by providing a product development perspective. The diversity between studied entities was pursued with analysing institutions from two separate countries: Finland and Sweden. Both of these countries have long traditions of implementing EHRs in health care. This practice results in a significant amount of available data, experience, and conducted research in the field.

Case 1: A Finnish research consortium at the University of Turku. In Case 1, the consortium's stakeholders, consisting of an EHR data provider and business operators, were analysed. A total of 9 interviewees included professors and chief officers in medicine, information technology, and business.

Case 2: A Swedish research group at Stockholm University. Case 2 only considers the research group and the corresponding EHR provider. A total of 5 interviewees included academic researchers and officials in information technology and medicine.

The data collection process was conducted by using semi-structured, in-person interviews. The purpose of these interviews was to evaluate the effectiveness of the protection of unstructured EHRs by analysing implemented procedures and technologies. In addition, the motivation to enhance the availability of secondary used EHR data was also discussed. The questions and conversation topics focused on data availability and privacy preservation measures in the context of secondary used EHR data. The data collection was performed from December 2015 to March 2016. The interviewees were divided into three groups: scientific researchers, product developers, and hospital management experts working for data owners.

Informed consent to participate in the study was received from each interviewee via email or verbally. Participating in the study was voluntary and the interviewees had the possibility to refuse their participation. The study was conducted according to good scientific practice.

The study material was analysed by using a qualitative content analysis process (Elo and Kyngäs, 2008). First, the content was prepared by selecting the actual analysis unit. Second, the study material was sorted by using both inductive and deductive approaches. Starting with the deductive phase, an analysis matrix was developed using technology and procedures perspectives, and the corresponding content was put into the matrix. Next, the inductive phase progressed from data grouping to categorisation and abstraction. The interview material was divided into specific sub-categories, which were merged into a few broader categories. Third, the study results were reported using these categories. Cross-case conclusions were formed to identify the case-specific differences. In this study, Case 2 was considered as a supporting example because of its smaller interviewee population.

Results and Discussion

The results are presented in the following three subsections: data management, privacy preservation, and secondary users. An overview of the study results is presented in Figure 1 and case-specific differences in Table 1.

Figure 1. Overview of study results

[Insert Figure 1.]

Table 1. Case-specific differences

[Insert Table 1.]

Data management

Request procedure

In Finland, the clarity of the EHR data request procedure was criticised by the representatives of scientific research. The request procedure was considered to be extensive and unstructured, thus the lack of simplicity requires the data applicants to perform excess work. The data provider offers potential secondary users with generic instructions to support the data request process. However, the basic nature of such instructions causes interpretation issues among potential users.

The requirements set by the data provider also include the demand to form explicitly specified research purposes. This demand encourages the users to narrow the scope of the study to fit some specific health-related category: ~~,~~ for instance, a symptom or medical specialty. This limits the amount and variety of EHR data requested by potential users. However, study-specific EHR categorisations and related data sets do not guarantee that the quality requirements set for research data will be met. The concept of EHR completeness for secondary use is highly context related and depends on understanding study-specific information needs (Weiskopf et al., 2013). Irrelevant data limitations set by the data provider were found to serve medical research intentions better than those in the context of formal sciences.

The data provider in the case involving Finland stated that relevant difficulties have been identified in the process of data requests. These difficulties include incoherence and irrationality factors, which have affected the clarity of previous information request procedures. The lack of relevant procedures considering the secondary use of EHRs can be rationalised with the organisational motivation to define use procedures for the data provider instance itself. However, the data provider in the case involving Finland underlines the desire to support the use of health information for secondary purposes.

The practicalities involved in data request procedures are primarily divided into two classes: (i) ethical permission required for the research intention and (ii) technological solutions for actual data delivery. Ethical permission for research use is granted by the corresponding ethical review board in Finland and Sweden. In addition, an agreement is required from the supervising physician in the corresponding health organisation. No notable variance in the ethical process was identified during the data collection interviews, hence the practicalities in this domain are settled. The data provider of case Sweden emphasised the possibility of declining the use of EHRs for legitimate secondary purposes. Therefore, the purpose of doing academic research does not ensure use permission. Use intentions, research resources, and storage methods must be provided to clarify the motivation to gain EHRs for secondary purposes.

Information control

In the case involving Finland, an external service provider operates the information system that is implemented in the hospital district. Therefore, the management of technical EHR data delivery from the provider organisation to potential users is managed by the same service provider. This results in the data owner's responsibility to control and evaluate the submitted data inquiries while leaving the technical process to external operators. However, this affects potential secondary users by adding excess costs charged by the service provider for managing the actual data delivery. Working with third–party operators also requires more time in comparison to a one-to-one process between the secondary user and the data owner.

Privacy preservation

The preservation of a patient's privacy is both essential and challenging in the context of unstructured EHRs. Primarily, such information is marginally used for secondary purposes if compared to structured health information, such as lab results and diagnosis codes. The disparity in the number of potential users favors processing structured data. For example, a higher volume of users correlates with more sophisticated procedures and mechanisms. However, the data provider requires secondary users to implement the same privacy preservation principles for both types of health information. These principles include protective actions to ensure patients' privacy and the requirement for ethical responsibility of individual researchers or users. In the context of secondary use, ethical responsibility is a consequence of data provider's deficient resources to monitor secondary users and procedures.

The results suggest controversy concerning the privacy preservation of unstructured EHRs in the case involving Finland. The identified issue considers the privacy preservation mechanisms that are used for non-trivial personal identifiers in health information. According to the data provider, there is a lack of motivation considering the use of such mechanisms in the first place instead of optimising the level and extent of privacy-preservation-related actions. The lack of motivation was justified with two factors: (i) the effect on data value and (ii) the challenge of the automated process. (i) With regard to the effect of privacy preservation on data value, the collected interview material discusses the inequality concerning the data alteration process of EHRs. The interview participants preferred to leave the information content unaltered to maximise the data value. This preference is justified with the possibility of generating semantically incoherent data sets by implementing extensive privacy preservation mechanisms. Broken data sets radically affect the use of unstructured EHRs for secondary purposes. Thus, the purpose of such data should direct the extent of connected privacy preservation actions. (ii) The process of automatically conducting privacy preservation is found to be a challenge. The challenge is caused by the demanding need to automatically detect PHI in EHRs. In turn, implementing a manual approach is considered laborious and cost inefficient. Therefore, both of these implementations were criticised by the data provider.

In the cases involving Finland and Sweden, the scientific representatives discussed the possibility of connecting the extent of privacy preservation actions to the corresponding use purpose. This approach results in a dynamic process in which the level of such actions is linked to the desired application. Therefore, there is no need to define an optimal and static level for the privacy preservation mechanisms that are implemented. In the case involving Sweden, privacy preservation actions related to secondary use are performed by the data provider. However, the provider organisation argued that the implementation of privacy preservation mechanisms does not ensure permission for use involving secondary purposes.

Secondary users

Open data

In Finland, the motivation to enable open access use of health information exists within research and business user groups. The collected interview data discuss the empowering effect of open access use on generating new research approaches and business ideas. Correspondingly, required limitations and regulations connected to the secondary use of EHRs would be moderated through reduced exclusivity of such information. However, these mitigations require the proper protection of patients' privacy.

Potential users representing scientific research justified the implementation of an open access process with enhanced study transparency and generated common good. Providing public access to corresponding research data enhances result reliability by limiting fabrication possibilities. This results in improved transparency and implementation of the American Medical Informatics Association's recommendation (Safran et al., 2007) for open secondary use processes. In this case, the achieved common good refers to improved efficiency of use for publicly funded EHRs in the Finnish health care setting.

The Finnish data provider instance has recognised the possibility of providing open access to EHR content. The motivation to provide health data using such implementation is found to be justified in the EHR provider organisation. Still, corresponding implementations or relevant procedures have not been defined. The lack of process definitions considering the open access use of EHRs is a consequence of inadequate national legislation in this context. Therefore, a conservative approach has been implemented in the data provider organisation.

A conservative approach refers to the process of offering coarse EHR data sets for publicly available open-access use. In this context, coarseness considers the intense aggregation of health data by limiting information diversity. As a result of data aggregation, the value of EHR data is reduced, especially in the domain of unstructured clinical text. Nevertheless, the process of sharing aggregated health data has provided pharmaceutical companies with a possibility to explore promising patient groups within the corresponding health organisation. For this purpose, data aggregation does not exclude successful secondary use.

In Sweden, the data provider instance has also implemented the conservative approach introduced for the case involving Finland. Therefore, coarse data sets are set available for open-access users in both Finland and Sweden. However, in the Swedish health care setting, the Swedish Healthcare Quality Registries offer individual health information regarding relevant patients for secondary purposes. Respective registries record the disease-specific information of corresponding patients by focusing on a single specialty. Such quality registries mainly record aggregated data, thus unstructured content is not provided for secondary users (Emilsson et al., 2015).

The results of the study process the relation between a patient's informed consent and open-access use of health information. It was identified that informed consent from an individual patient grants the open-access use of their health data. In this context, such consent must not be confused with the secondary use defined in the EU and national-level legislation that considers potential researchers and a data provider. Individual patients may use health-specific data-sharing platforms that are developed especially for secondary users and use intentions. In the context of unstructured health information, the successful use of such platforms is difficult. EHRs are created by a health organisation, which possesses the creator's right to unstructured content within health documents. Hence, patients face limitations considering the amount and variety of unstructured data to be shared using these platforms. Data availability is also restricted by the voluntary nature of data-sharing platforms, which are dependent on active users. In contrast to relevant health organisations, the difference in the amount of processable EHR data is evident, preferring operators in the health care domain.

Product development

The Finnish data provider finds the process of sharing EHR data for product development purposes to be valid and mutually beneficial in an information systems context. These shared benefits refer to successful product development, which results in improved usability and quality factors, further enhancing the cost efficiency of relevant information systems. Hence, the secondary use of health information is possible for product developers considering the identified benefits of such actions. According to the data provider, none of the data requests performed by secondary users are automatically rejected.

In Sweden, the data provider argued against providing EHR data for any purpose other than scientific research, which is stated in the national legislation. Such data can be provided for other use intentions only if informed consent is received from individual patients. The Swedish data provider criticised the competence of relevant secondary users for processing unstructured EHR content. The lack of potential users was identified as another justification for limited data availability.

The results of the study indicate a conflict between the Finnish data provider and potential business operators in terms of potential product development. The interviewees representing business users criticised that relevant EHR data are not generally available for such purposes. In addition, the amount and variety of potentially available EHR content was found to be limited. As a result, business operators emphasised the purpose of national encouragement by considering the further use of health information for product development. The lack of relevant legislation and procedures reduce EHR availability, which further damages related actions that support business.

The Finnish data provider did not find it necessary to process genuine unstructured EHRs for product development purposes. This view was motivated by widely available artificial patient scenarios, which can be used for business-driven use cases. These patient scenarios deal with artificially made patients and connected documentation of their care, which are both generated by health care personnel. The advantage of developing these documents includes the preserved privacy of relevant patients while sharing fairly genuine health information for secondary users. However, instead of receiving artificial patient scenarios for product development, there is a chance to use genuine unstructured data if the content is comprehensively anonymised and mixed. This approach results in wide diversity between potential secondary users (e.g., scientific researchers vs. business operators) in terms of further use of EHRs.

Limitations of the study

The limitations of this study include a disproportion between the number of interviewees for Cases 1 and 2. This is a result of lacking business representatives in Case 2. Therefore, the interviewee population was limited to 14 people. In addition, the affiliations of the interviewees may bring bias to the research data. The potential secondary users of EHRs are more likely than data owners to support the use of such data for secondary purposes.

These results are not generalisable, but they provide information for designing studies for data protection.

Conclusions

This article discussed various EHR availability–affecting factors confronted by secondary users of such health information. Based on the study results, two approaches for implementing the secondary use of EHRs were identified. First, protected processing refers to the use of a secure processing environment that is established by the data provider organisation to ensure patients' privacy with an exclusively used system. Second, data alteration ensures the confidentiality of sensitive information by implementing privacy preservation mechanisms on personal identifiers within the data set. As a result, both approaches are developed and implemented to primarily enhance the privacy of the secondary use of health information.

Declaration of conflicting interests

The authors declare that there is no conflict of interest.

Funding

This study was funded by the Finnish Funding Agency for Technology and Innovation (TEKES). According to the funding terms of TEKES, this project was co-funded by three companies: BCB Medical Oy, Fujitsu Finland Oy, and Lingsoft Oy.

Acknowledgements

References

Allvin H, Carlsson E, Dalianis H, Danielsson-Ojala R, Daudaravicius V, Hassel M, Kokkinakis D, Lundgren-Laine H, Nilsson GH, Nytrø Ø et al. (2011) Characteristics of Finnish and Swedish intensive care nursing narratives: a comparative analysis to support the development of clinical language technologies. *Journal of biomedical semantics* 2(S-3): S1.

Council Directive 95/46/EC (1995) Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data.

Elo S and Kyngas H (2008) The qualitative content analysis process. *Journal of advanced nursing* 62(1): 107–115.

Emilsson L, Lindahl B, Koster M, Lambe M and Ludvigsson JF (2015) Review of 103 swedish healthcare quality registries. *Journal of internal medicine* 277(1): 94–136.

Furnell S (2008) *Securing information and communications systems: principles, technologies, and applications.* Boston, MA: Artech House.

Hayrinen K, Saranto K and Nykänen P (2008) Definition, structure, content, use and impacts of electronic health records: a review of the research literature. *International journal of medical informatics* 77(5): 291–304.

Health Insurance Portability and Accountability Act (HIPAA) (1996) Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. http://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html.

Merkow M and Breithaupt J (2006) *Information Security: Principles and Practices.* Upper Saddle River, NJ: Pearson Education.

Meystre SM, Ferrández Ó, Friedlin FJ, South BR, Shen S and Samore MH (2014) Text de-identification for privacy protection: A study of its impact on clinical text information content. *Journal of biomedical informatics* 50: 142–150.

Meystre SM, Friedlin FJ, South BR, Shen S and Samore MH (2010) Automatic de-identification of textual documents in the electronic health record: a review of recent research. *BMC medical research methodology* 10(1): 70.

Neamatullah I, Douglass MM, Li-wei HL, Reisner A, Villarroel M, Long WJ, Szolovits P, Moody GB, Mark RG and Clifford GD (2008) Automated deidentification of free-text medical records. *BMC medical informatics and decision making* 8(1): 32.

Regulation 2016/679 (2016) Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC.

Safran C, Bloomrosen M, Hammond WE, Labkoff S, Markel-Fox S, Tang PC and Detmer DE (2007) Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. *Journal of the American Medical Informatics Association* 14(1): 1–9.

Vikström A, Moosavi SR, Moen H, Salakoski T, Salanterä S (2016) Factors Affecting the Availability of Electronic Patient Records for Secondary Purposes – A Case Study. In: International Conference on Well-Being in the Information Society: pp.47-56. Springer, Cham.

Weiskopf NG, Hripcsak G, Swaminathan S and Weng C (2013) Defining and measuring completeness of electronic health records for secondary use. *Journal of biomedical informatics* 46(5): 830–836.

Table 1. Case-specific differences

	Scientific researchers		Hospital management		Product developers
	Finland	Sweden	Finland	Sweden	Finland
Data management	The data request process is found to be extensive and unstructured.		Motivation to promote secondary use of health information. Difficulties in data request procedures.	Limited interest for processing unstructured EHRs for secondary purposes in Sweden. Structured EHRs are promoted.	
Privacy preservation	The extent of privacy preservation mechanisms should be connected to the use purpose.		Privacy preservation mechanisms affect data value and are challenging to automate.	The use of privacy preservation mechanisms does not guarantee secondary use.	
Secondary users	Open access enhances the transparency and reliability		Secondary use of EHRs is possible for product	EHRs can be provided for product developers if	EHR data is not generally available for

	of scientific studies.		developers if such use is identified to be beneficial.	informed consent is received from corresponding patients.	product development.