Recognition and Pseudonymization of Personal Data in Paper-Based Health Records

Stefan Fenz[1], Johannes Heurix[2], and Thomas Neubauer[1]

[1] Vienna University of Technology, Vienna, Austria
{stefan.fenz,thomas.neubauer}@tuwien.ac.at
[2] SBA Research, Vienna, Austria
jheurix@sba-research.org

Abstract. E-health requires the sharing of patient-related data when and where necessary. Electronic health records (EHR) allow the structured and expandable collection of medical data needed for clinical research studies and thereby not only enable the optimization of clinical studies, but also results in higher statistical significance due to a larger number of samples. While the digitization of medical data and the organization of this data within EHRs have been introduced in some areas, massive amounts of paper-based health records are still produced on a daily basis. This data has to be stored for decades due to legal reasons but is of no benefit for research organizations, as the unstructured medical data in paper-based health records cannot be efficiently used for clinical studies. Furthermore, legal regulations prohibit the use of documents containing both personal and medical data for clinical studies, which leads to expensive data acquisition phases and limited samples. This paper presents the MEDSEC system for the recognition and pseudonymization of personal data in paper-based health records. MEDSEC integrates unique methods for (i) automatically identifying personal and medical data, (ii) automatically annotating the optical character recognition (OCR) output data of paper-based health records with standard-compliant metadata, and (iii) automatically pseudonymizing the personal data. With MEDSEC, health care organizations profit by (i) strengthening clinical research resulting in faster and more reliable results and reduced costs, and (ii) providing an environment of trust for its patients and employees that guarantees privacy.

Keywords: EHR, privacy, annotation, HL7 CDA, pseudonymization, transformation, OCR.

1 Introduction

In today's health care system, the availability of sound information has tremendous impact on decisions regarding patients' care and, as a result, on the quality of treatment and patients' health. The digitization of medical data (e.g., by using electronic health records (EHR)) promises (i) the reduction of adverse drug events accounting for about US$175 billion a year in the US, (ii) the reduction of the very high number of more than 200,000 cases of deaths a year in the US [1] as it provides physicians and their health care teams [2] with decision support systems and guidelines for drug interactions, and (iii) massive savings that can be achieved by digitizing diagnostic tests and images.

W. Abramowicz et al. (Eds.): BIS 2012, LNBIP 117, pp. 153–164, 2012.

A study by the non-profit research organization Rand Corporation found out that adopting the EHR could result in more than US$81 billion in annual savings in the US if 90% of health care providers used it [1].

In addition to the direct benefits, the digital storage and analysis of medical data could mean a quantum leap in clinical research, because it allows the improvement of communication between health care providers and of access to data and documentation, leading to better clinical and service quality [3]. Today, the success of clinical trials heavily depends on the recruitment of enough eligible participants in a timely manner. Failing to meet recruitment goals can hamper the development and evaluation of new therapies and can not only increase drug development costs but also health care system costs (cf. [4] for estimates about the costs of clinical trials). Today, 86% of all trials fail to start on time because subjects cannot be recruited in time and because only 7% of eligible patients enroll in a clinical trial. One study, which looked at 4,000 clinical trials over five years, discovered that nearly half of the time spent on the trial process involved patient, site and investigator recruitment [5]. Clinical research is ending up in a vicious circle because clinical trial capacity does not meet the demand, and whereas the number and the duration of trials is increasing, the number of patients available for trials is decreasing. The structured organization of digitized medical data (e.g., within an EHR) allows (i) the more effective and efficient recruitment of clinical trial participants, (ii) the reduction of administrative overhead, (iii) the impact reduction of data errors due to larger samples and (iv) the faster identification of adverse outcomes. However, the vast majority of health records is still only available on paper and experts agree that the amount of paper-based health records will never be beat down below 20%, leaving enormous potential for improving clinical research. There are three major problems preventing the use of paper-based health records in clinical research:

- First, paper-based health records do not provide machine-interpretable metadata and circumvent the automatic identification of personal and medical data elements. Currently, no methods for the automatic identification of personal and medical data exist. Existing high-level privacy taxonomies (e.g., [6]), ontology-based trust negotiation approaches (e.g., [7]), and web standards, such as W3C P3P [8], provide a categorization of privacy-relevant data items but do not provide common synonyms and formal specifications of personal and medical data elements to enable their automatic detection in paper-based health records. Since it is not possible to use only the content of a data element to automatically determine its type, formal descriptions have to include potential identifiers used in paper-based health records. The further use of digitized and pseudonymized paper-based health records for clinical research highly depends on the complete identification of personal and medical data.
- Second, the sole digitization of paper-based health records is not sufficient for providing clinical research with suitable data. In addition to the actual digitization, it is of paramount importance for the distinction between different data elements to enrich the gathered data with appropriate standard-compliant metadata (e.g.,

according to the HL7 standard). While products for indexing optical character recognition (OCR) output data by self-defined profiles exist[1,2,3], no methods are available for the automatic annotation of personal and medical data. The lack of such methods prevents the further processing of the gathered OCR output for clinical research. Existing methods for enriching OCR output with standard-compliant and appropriate metadata do not meet clinical research requirements and do not guarantee the complete identification of personal data according to the Austrian Data Protection Act. Furthermore, the data complexity in the health care domain and the need for exchanging this data over existing system boundaries requires the usage of standardized data structures and communication protocols. While standards such as HL7 are already implemented in several health care information systems, no methods for the automatic transformation of semantically enriched OCR output data into a standard such as HL7 exist.

– Third, privacy is one of the fundamental issues in health care today. With informative and interconnected health-related data comes highly sensitive and personal information. Due to the high sensitivity of the data, there is increasing social and political pressure to prevent the misuse of personal health data. It is the fundamental right of every citizen to demand privacy (cf. HIPAA, EU Directives), and furthermore, the disclosure of medical data can cause serious problems for the patient. The increasing fear of data abuse as well as the adoption of laws lead to the development of a variety of techniques for protecting patients' identity and privacy. The concept of pseudonymization (cf. [9,10]) allows the data to be associated with a patient only under specified and controlled circumstances. Existing approaches can be differentiated into two groups: the first group of approaches has major security shortcomings (cf. [11,12,13,14,15,16]); the second group solves these shortcomings, but is not designed for the centralized (mass) pseudonymization of data (due to different requirements regarding architecture, security, and performance).

2 Background

The annotation of OCR output data with appropriate metadata (e.g., birth date, first name and gender) requires the formal specification of what personal and medical data actually is. The most mature approach for classifying personal data is the Platform for Privacy Preferences Project (P3P) [8]. P3P Specification 1.1 [8] defines a base data schema for personal data, including data elements such as first name, birth date, phone number, and email. The ICD-10 Standard is an international statistical classification of diseases and related health problems. Together with the HL7 standard it can be used as a foundation to classify medical-related data elements.

While the mentioned data schemes outline personal and medical data elements, they do not describe how concrete instances of these data elements could look like (e.g., that

[1] Dynamic Zone OCR: `http://www.simpleindex.com`
[2] docWorks: `http://www.content-conversion.com`
[3] ImageNet: `http://www.miteksystems.com`

a name does not include any numbers). Another shortcoming is that no synonyms are given for the different data elements (e.g., gender/sex, first name/given name). We use the HIPAA PHI schema and the ontology-based trust negotiation approach (cf. [7]) as the basis for the development of a personal data ontology that includes common synonyms in multiple languages and formal descriptions that enable the automatic identification of personal data elements in health records. On the medical side we will use the HL7 standard as the foundation for creating common multi-lingual synonyms and formal descriptions for relevant medical data elements.

Besides the mere identification of personal and medical data elements, it is crucial to annotate the identified data elements with metadata that corresponds to well-established health care standards (e.g., HL7). We plan to combine existing indexing tools with the developed formal data element descriptions to automatically annotate personal and medical data elements.

The use of open standards can considerably reduce the costs of electronic data capture in clinical research. The CDISC ODM [4] is oriented towards drug development and clinical research. CDISC, for example, allows the automatic setup of the EDC system, the creation and instantiation of the database, and the full automation of the creation of electronic case report forms. HL7 is a standards development organization dealing with data standards for all health care operations. Because of its broader scope, HL7 has not dealt much with the nuances of clinical trials, while CDISC has not dealt with health care applications important to HL7, such as reimbursements and order processing. The HL7 Clinical Document Architecture (CDA) is a document markup standard that specifies the structure and the semantics of clinical documents in Extensible Markup Language (XML). Persistence, stewardship, potential for authentication, wholeness and human readability are the main characteristics of the CDA [17]. One of its main characteristics, however, is also its main downside: Human readability allows the convenient use in health care environments but inhibits privacy.

In order to protect patients' privacy when using, transferring and storing medical records, a variety of privacy enhancing technologies (cf. [18]) have been proposed. However, existing approaches (i) do not comply with the current legal requirements (cf. [19,20,21,22,23]), (ii) do not fulfill basic security requirements (cf. [24,25]), and (iii) are not applicable for use with clinical studies. In 2006, the United States Department of Health & Human Services issued the Health Insurance Portability and Accountability Act (HIPAA) [26], which demands the protection of patients' data that is shared from its original source of collection. While no explicit European standards regarding the protection of PHI exist, HIPAA defines 17 PHI identifiers that have to be removed from the health record: (i) names, (ii) locations, (iii) dates, (iv) ages greater than 89, (v) telephone numbers, (vi) fax numbers, (vii) email addresses, (viii) social security numbers, (ix) medical record numbers, (x) health plan beneficiary numbers, (xi) account numbers, (xii) certificate numbers, (xiii) vehicle identifiers, (xiv) device identification numbers, (xv) URLs, (xvi) IP addresses, (xvii) biometric identifiers, and (xviii) any other unique identifying number, code, or characteristic such as full face photos.

[4] Clinical Data Interchange Standards Consortium: Specification for the Operational Data Model, `http://www.cdisc.org`

Since 2005, the processing and movement of personal data in the EU has been legally regulated by Directive 95/46/EC [19]. A citizen's right to privacy is also recognized in Article 8 [27] of the European Convention for the Protection of Human Rights and Fundamental Freedoms. Additionally, domestic acts in many EU member states contain strict regulations for the processing of personal data.

Please note that e-health and especially clinical studies demand the pseudonymization of data: (i) Anonymization - the removal of the identifier from the medical data - has the major drawback that patients cannot profit from the results gained in the clinical studies (e.g., patients cannot be informed about actual findings such as newly developed medical treatments or major changes in the healing progress). (ii) Encryption assures patients' privacy by encrypting the medical records with the patients' private key. However, encrypted data cannot be used for clinical research (secondary use) without the explicit permission of the patient who has to decrypt the data and in doing so, reveals her identity. Pseudonymization is a technique where identification data is transformed and then replaced by a specifier that cannot be associated with the identification data without knowing a certain secret [9,25,10]. Pseudonymization allows the data to be associated with a patient only under specified and controlled circumstances. A pseudonymized database must contain at least two tables, one where all the personal information is permanently stored, and one where the pseudonyms and the pseudonymized data are stored. The process of identifying and separating personal from other data is called depersonalization. After depersonalization and subsequent pseudonymization, a direct association between individuals and their data cannot be established. However, existing approaches and systems have a variety of shortcomings. The system developed by Thielscher et al. (cf. [12]) relies on a centralized patient pseudonym list which provides a fallback mechanism in case a patient looses her smart card, as otherwise there would be no way to recover the identifier. Thielscher et al. circumvent the security flaw of a centralized patient pseudonym list by operating it off-line. This organizational workaround seems to promise a higher level of security until a social engineering attack is conducted on a person inside the system [28,29] or an attacker gains physical access to the computer that holds the list. The approaches developed by Pommerening (cf. [15,16]) use a combination of a hashing and an encryption technique. The encryption itself is based on a centralized secret key, which opens a vulnerability, as an attacker who knows this single key might gain access to all patients' medical data. The approach developed by Peterson [11] comes with some serious drawbacks: As all keys needed for decrypting the medical data are stored in the database, an attacker gaining access to the database could decrypt all information. Even more importantly, as the password is also stored in the database as well as the keys, the attacker could change data stored in the database. The architectures proposed by Schmidt et al. [30] and the Fraunhofer Institute, supported by the German Federal Ministry of Health [31,32], are based on encryption. As a result the data is fully encrypted, which is not practicable for the use in clinical studies.

The PIPE framework is a new patented architecture (cf. [33,34,35,36,37,38,39,40] for details on our patent and previous work) that improves existing approaches by (i) allowing the authorization of health care providers or relatives of the patient to access

specified medical data at encryption level, (ii) providing a secure fallback mechanism in case the security token is lost or worn out, (iii) storing the data without the possibility of data profiling, and (iv) allowing secondary use without establishing a link between the data and the person it refers to. Patient-identifying details are separated from the actual health data, resulting in detached data records. The relation between the patient and her health data is established with pseudonyms that are accessible only under specifically defined conditions. In this way, only persons who know the pseudonyms are able to link the patient with the health data. Pseudonyms are also used for data access permissions, e.g., defining new pseudonyms for access authorizations or revoking access rights by deleting the pseudonyms.

Apart from the security shortcomings, existing pseudonymization approaches - including the PIPE approach - have a number of characteristics in common:

- They depend on the smart card's crypto chip for performing cryptographic operations. Although this technique, combined with a certified card reader and a PIN, can be considered secure [41], it is not usable if central and automatic pseudonymization (e.g., in the case of pseudonymizing large amounts of data) is needed.
- They do not provide high performance (e.g., 12 millions documents a year) solutions for central pseudonymization. The cryptographic chip on the smart card does not provide anything close to the performance needed for pseudonymizing such a number of documents.
- There is no access to the data owner's card at the moment of pseudonymization. It would be logistically impossible to gain synchronous access to the data owner's keys. Asynchronous options are not considered in current architectures.
- The architectures are designed for patient-centric scenarios (e.g., use in EHRs) but not for allowing central pseudonymization while at at the same time guaranteeing a high level of security and privacy.

What is required for the pseudonymization of data archives is a (i) central, (ii) high-performant, and (iii) automatic pseudonymization approach. In this proposal we define such an approach as 'mass pseudonymization'.

3 The MEDSEC System

The goal is to provide clinical studies with pseudonymized and structured medical data gained from existing paper-based health records. The proposed technical solution is divided into four main phases. Figure 1 shows an overview of the proposed solution.

OCR: The purpose of this phase is to digitize the content of paper-based health records. As the development of OCR engines was not the focus of this project, we use Google's open-source OCR engine Tesseract[5], which is one of the most accurate open source OCR engines available[6]. Besides digitizing the actual content of paper-based health records, we enrich the corresponding OCR output with metadata containing information

[5] Tesseract: `http://code.google.com/p/tesseract-ocr/`

[6] Willis, Nathan (2006). Google's Tesseract OCR engine is a quantum leap forward: `http://www.linux.com/articles/57222`

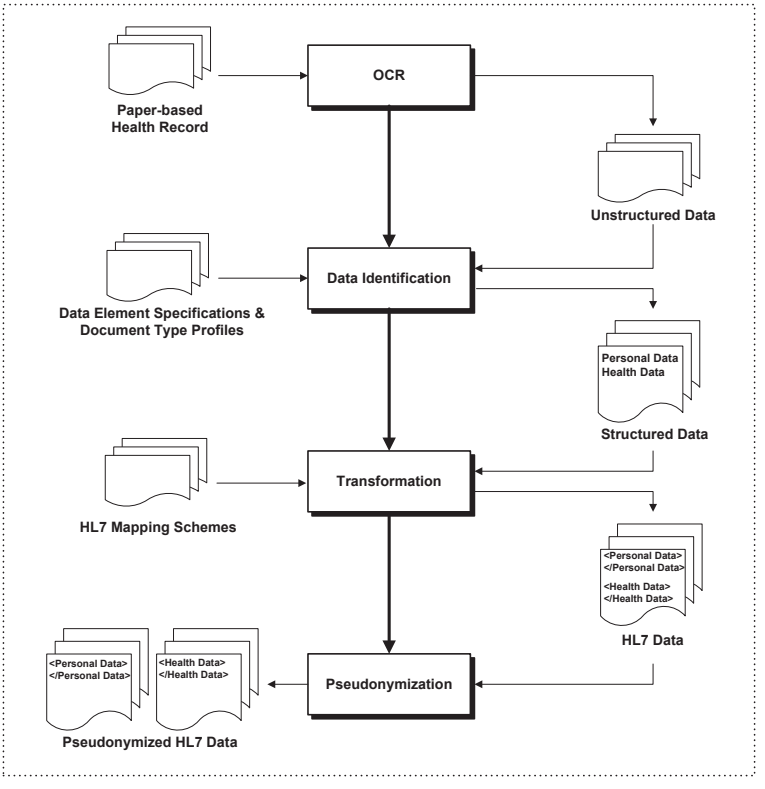

Fig. 1. System overview

about the document type (e.g., physician's letter, medical evidence, etc.). In our case we are using separate sheets with bar codes to identify the document type of each health record. As the assignment of these description sheets is a cumbersome manual process, we use a method for recognizing the document type automatically in the OCR phase. By matching the gathered layout data with the developed layout profiles we can improve the efficiency of the document type classification.

Data Identification: The data identification phase transforms the unstructured OCR data into a structured data format using the developed document type profiles. The document type profiles provide offsets for each data element/document type combination and enable us to identify data elements based on their position on the original health record. Additionally, we developed methods for identifying personal and medical data independently of an existing document type classification:

– Content-based identification: the data element is identified based on the content of the data item in question, e.g., checking each 10-digit number for its potential to be a social security number by calculating the check-digit or matching a string against a list of given names. We use the HIPAA PHI schema as the basis for categorizing privacy-relevant data elements and the HL7 data classification for categorizing

medical data elements. Based on that categorization we define synonyms and formal specifications of personal and medical data elements to enable their automatic detection in health records.
- Context-based identification: the data element is identified based on the given context, e.g., each string located next to the string 'Social Security Number' has a high potential of being a social security number. Together with the defined synonyms of personal and medical data elements we use state-of-the-art document analysis engines to automatically identify personal and medical data more reliable.

Transformation: Clinical research frequently utilizes proprietary data formats that are often incompatible with the data standards of other organizations. As a result, clinical data can rarely be exchanged between different organizations [42]. The purpose of the transformation phase is to convert the structured personal and health data into standardized data formats. Due to the complexity of standards such as HL7 or CDISC, we developed appropriate mapping schemes to ensure standard compliance of the generated output. Standard data formats, such as HL7 CDA, consist of a header and a body. The header includes the context in which the document was created, and the body contains the actual content of the document. The purpose of the header is to support communication across and within institutions, facilitate clinical document management, and facilitate the compilation of an individual patient's clinical documents into a lifetime electronic health record. MEDSEC guarantees that

- the body of a CDA document (either an unstructured blob or a structured markup) does not include any personal data, and that
- all information, needed for further processing of the data is included in the header without reducing privacy.

Pseudonymization: A server-side instance (e.g., HSM) acts as cryptographic module for executing the necessary cryptographic steps within a trusted secure environment. The cryptographic operations include all encryption and decryption operations required for functions, such as user authorization and authentication. The client-side cryptographic operations, required, e.g., for the challenge/response-style authentication procedure, are carried out with the user-owned security token doubling as secure keystore for the authentication credentials and a client-side cryptographic module. The architecture (see Figure 2) is realized as a multi-tier hull model with three different layers. Each layer is responsible for one step in the data access process. The user has to pass all layers in order to retrieve the actual health records. The outer hull, the authentication layer, is responsible for authenticating the user by requiring him to prove his identity. Technically, the outer hull is realized by the outer asymmetric keypair (outer public key OPuK and outer private key OPK) that is stored on the user's security token. The keys on the security token are only accessible when entering the correct PIN, thus providing two-factor authentication. Authentication involves the user's and the server's outer keypair, the user's internal user ID (IUID), and a random value. The user's outer private key is also used to decrypt his inner private key, which in turn is needed for decrypting the inner symmetric key. The inner symmetric key (ISK) and the inner private and public keys (IPK and IPuK) form the inner hull, the authorization layer. Without the inner symmetric key, the user cannot access the correct pseudonyms which are encrypted with his

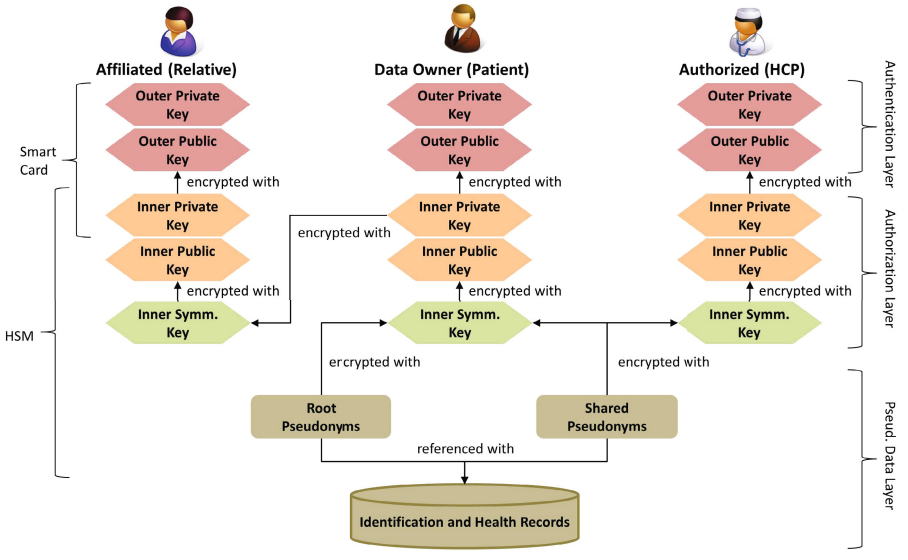

Fig. 2. Pseudonymization architecture ([36])

inner symmetric key. The pseudonyms could be directly encrypted with the inner public key and would still be secured against unauthorized access. However, defining an additional inner symmetric key has the following advantages: As symmetric encryptions are executed faster than the costly asymmetric cryptographic operations, reducing the number of encryptions/decryptions involving the asymmetric keys increases the overall execution speed. At the same time, it prevents the user from directly accessing the inner symmetric key, as it is only present in plaintext within the secure environment of the HSM where the pseudonyms are encrypted and decrypted. The plaintext pseudonyms are attached to the actual health records, and both together represent the innermost

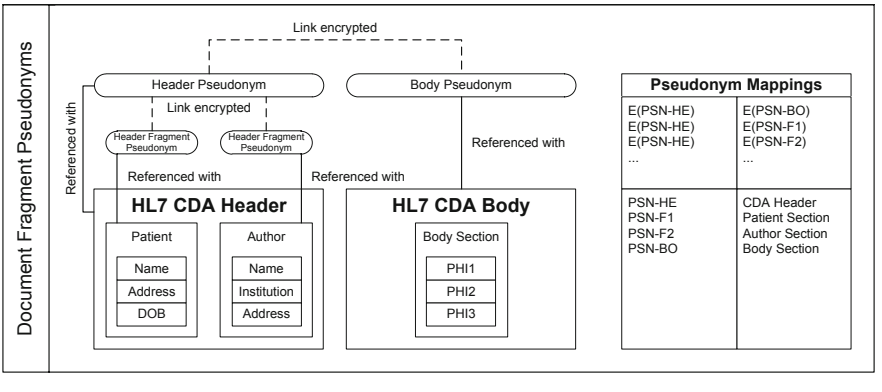

Fig. 3. Document Fragment Pseudonyms

layer, the concealed data layer. Figure 3 demonstrates the pseudonyms of HL7 CDA documents that are fragmented into several sections: While the CDA body section is assigned a single body pseudonym, the CDA header section is further fragmented (into header, patient, and author sections) and each fragment assigned an individual fragment pseudonym. As can be seen in the Pseudonym Mappings, the pseudonyms are attached to the CDA document fragments in plaintext, while the links between the fragments are effectively concealed by encryption. Thus, if in possession of the correct decryption key, the mappings can be decrypted and thus the links between the fragments restored.

4 Conclusion

MEDSEC was implemented into a software solution and tested within a national healthcare provider in Austria that treats about 250.000 inpatients and 600.000 outpatients annually. Initial test runs with a limited document base demonstrated the system's practicality, producing promising results. The system is currently undergoing a test run on a larger scale with minor modifications to further improve the system, especially concerning the quality of the OCR and data identification output. The results will be presented in detail in a future publication. The project results enable to strengthen clinical research and harbor considerable economic benefits for the society due to the decreased treatment costs and more efficient clinical trials: MEDSEC simplifies the analysis of medical data by providing more representative samples and, thus, reduces the time required for carrying out clinical research (including clinical trials). This has two major advantages: (i) Clinical research can be carried out in a fraction of the (original) time due to faster recruitment. This is a powerful argument, because research organizations rely on the fast publication of research results. (ii) A larger sample results in more reliable and significant outcomes and has a major influence on the research quality. Digitized health records reduce costs for hospitals and research organizations in the following ways: (i) They save expensive archive space of paper-based health records. (ii) Digitization has the side effect of allowing the categorization of data and, thus, the fast and efficient search for specific information, which results in improved treatment processes for the patient. (iii) The conversion of medical data into standard formats, such as HL7, allows the more efficient administration and use of this data in clinical environments.

Acknowledgments. The research was funded by BRIDGE (#824884) and by COMET K1, FFG - Austrian Research Promotion Agency.

References

1. Ernst, F.R., Grizzle, A.J.: Drug-related morbidity and mortality: Updating the cost-of-illness model. Journal of the American Pharmacists Association 41(2), 192–199 (2001)
2. Pope, J.: Implementing EHRs requires a shift in thinking. PHRs–the building blocks of EHRs–may be the quickest path to the fulfillment of disease management. Health Management Technology 27(6), 24 (2006)
3. Maerkle, S., Koechy, K., Tschirley, R., Lemke, H.U.: The PREPaRe system – Patient Oriented Access to the Personal Electronic Medical Record. In: Proceedings of Computer Assisted Radiology and Surgery, Netherlands, pp. 849–854 (2001)

4. Masi, J.D., Hansen, R., Grabowski, H.: The price of innovation: New estimates of drug development costs. Journal of Health Economics 22, 151–185 (2003)
5. 2000, C.I.: R&D Briefing: Benchmarking for Efficient Drug Development (2000)
6. Anton, A.I., Earp, J.B., Reese, A.: Analyzing website privacy requirements using a privacy goal taxonomy. In: Proceedings of the IEEE Joint International Conference on Requirements Engineering, pp. 23–31 (2002)
7. Squicciarini, A., Bertino, E., Ferrari, E., Ray, I.: Achieving privacy in trust negotiations with an ontology-based approach. IEEE Transactions on Dependable and Secure Computing 3(1), 13–30 (2006)
8. W3C: Platform for Privacy Preferences (P3P) Project (October 2007), `http://www.w3.org/P3P/`
9. Pfitzmann, A., Koehntopp, M.: Anonymity, Unlinkability, Unobservability, Pseudonymity, and Identity Management – A Consolidated Proposal for Terminology. LNCS. Springer, Heidelberg (2005)
10. Taipale, K.A.: Technology, Security and Privacy: The Fear of Frankenstein, the Mythology of Privacy and the Lessons of King Ludd. International Journal of Communications Law & Policy 9 (2004)
11. Peterson, R.L.: Patent: Encryption system for allowing immediate universal access to medical records while maintaining complete patient control over privacy. US Patent US 2003/0074564 A1 (2003)
12. Thielscher, C., Gottfried, M., Umbreit, S., Boegner, F., Haack, J., Schroeders, N.: Patent: Data processing system for patient data. Int. Patent, WO 03/034294 A2 (2005)
13. de Moor, G.J., Claerhout, B., de Meyer, F.: Privacy enhancing technologies: the key to secure communication and management of clinical and genomic data. Methods of Information in Medicine 42, 148–153 (2003)
14. Gulcher, J.R., Kristjánsson, K., Gudbjartsson, H., Stefánsson, K.: Protection of privacy by third-party encryption in genetic research. European Journal of Human Genetics 8(10), 739–742 (2000)
15. Pommerening, K.: Medical Requirements for Data Protection. In: Proceedings of IFIP Congress, vol. 2, pp. 533–540 (1994)
16. Pommerening, K., Reng, M.: Secondary use of the Electronic Health Record via Pseudonymisation. In: Medical and Care Compunetics 1, pp. 441–446. IOS Press (2004)
17. Dolin, R.H., Alschuler, L., Beebe, C.: The hl7 clinical document architecture. J. Am. Med. Inform. Assoc. 8(6), 552–569 (2001)
18. Fischer-Huebner, S.: IT-Security and Privacy: Design and Use of Privacy-Enhancing Security Mechanisms. Springer (2001)
19. European Union: 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. Official Journal of the European Communities L 281, 31–50 (1995)
20. Hinde, S.: Privacy legislation: a comparison of the US and European approaches. Computers and Security 22(5), 378–387 (2003)
21. Hornung, G., Goetz, C.F.J., Goldschmidt, A.J.W.: Die künftige Telematik-Rahmenarchitektur im Gesundheitswesen. Wirtschaftsinformatik 47, 171–179 (2005)
22. U.S. Department of Health & Human Services Office for Civil Rights: Summary of the HIPAA Privacy Rule (2003)
23. U.S. Congress: Health Insurance Portability and Accountability Act of 1996. 104th Congress (1996)
24. Schabetsberger, T., Ammenwerth, E., Göbel, G., Lechleitner, G., Penz, R., Vogl, R., Wozak, F.: What are functional requirements of future shared electronic health records? Connecting Medical Informatics and Bio-Informatics, 1070–1075 (2005)

25. Riedl, B., Neubauer, T., Goluch, G., Boehm, O., Reinauer, G., Krumboeck, A.: A secure architecture for the pseudonymization of medical data. In: Proceedings of the Second International Conference on Availability, Reliability and Security, pp. 318–324 (2007)
26. United States Department of Health & Human Service: HIPAA Administrative Simplification: Enforcement; Final Rule. Federal Register / Rules and Regulations 71(32) (2006)
27. Council of Europe: European Convention on Human Rights. Martinus Nijhoff Publishers (1987)
28. Maris, K.: The Human Factor. In: Proceedings of Hack.lu, Luxembourg (2005)
29. Thornburgh, T.: Social engineering: the "Dark Art". In: Proceedings of the First Annual ACM Conference on Information Security Curriculum Development, pp. 133–135. ACM Press (2004)
30. Schmidt, V., Striebel, W., Prihoda, H., Becker, M., Lijzer, G.D.: Patent: Verfahren zum Be- oder Verarbeiten von Daten. German Patent, DE 199 25 910 A1 (2001)
31. Fraunhofer Institut: Spezifikation der Lösungsarchitektur zur Umsetzung der Anwendungen der elektronischen Gesundheitskarte (2005)
32. Caumanns, J.: Der Patient bleibt Herr seiner Daten. Informatik-Spektrum, 321–331 (2006)
33. Heurix, J., Karlinger, M., Neubauer, T.: Pseudonymization with metadata encryption for privacy-preserving searchable documents. In: Proceedings of the 45th Hawaii International Conference on System Sciences, HICSS 45 (2012)
34. Heurix, J., Karlinger, M., Schrefl, M., Neubauer, T.: A Hybrid Approach integrating Encryption and Pseudonymization for Protecting Electronic Health Records. In: Proceedings of the Eighth IASTED International Conference on Biomedical Engineering, p. 117 (2011)
35. Heurix, J., Neubauer, T.: Privacy-Preserving Storage and Access of Medical Data through Pseudonymization and Encryption. In: Furnell, S., Lambrinoudakis, C., Pernul, G. (eds.) TrustBus 2011. LNCS, vol. 6863, pp. 186–197. Springer, Heidelberg (2011)
36. Neubauer, T., Heurix, J.: A methodology for the pseudonymization of medical data. International Journal of Medical Informatics 80(3), 190–204 (2011)
37. Neubauer, T., Kolb, M.: An Evaluation of Technologies for the Pseudonymization of Medical Data. In: Lee, R., Hu, G., Miao, H. (eds.) Computer and Information Science 2009. SCI, vol. 208, pp. 47–60. Springer, Heidelberg (2009)
38. Neubauer, T., Riedl, B.: Improving patients privacy with pseudonymization. In: Proceedings of the International Congress of the European Federation for Medical Informatics (2008)
39. Riedl, B., Grascher, V., Fenz, S., Neubauer, T.: Pseudonymization for improving the privacy in e-health applications. In: Proceedings of the Forty-First Hawai'i International Conference on System Sciences (2008)
40. Riedl, B., Grascher, V., Neubauer, T.: A secure e-health architecture based on the appliance of pseudonymization. Journal of Software (2008)
41. Hendry, M.: Smart Card Security and Applications, 2nd edn. Artech House, Inc., Norwood (2001)
42. Waegemann, C.: Status report 2002: Electronic health records. Medical Records Institute, Boston (2004)