An Approach towards the Fulfilment of Security Requirements for Decision Support Systems in the Field of Evidence-Based Healthcare*

Nevena Stolba[1] and A Min Tjoa[2]

[1] Women's Postgraduate College for Internet Technologies
Institute of Software Technology and Interactive Systems
Vienna University of Technology
Favoritenstr. 9-11/188, A-1040 Vienna, Austria
stolba@wit.tuwien.ac.at

[2] Institute of Software Technology and Interactive Systems
Vienna University of Technology
Favoritenstr. 9-11/188, A-1040 Vienna, Austria
amin@ifs.tuwien.ac.at

Abstract

Evidence-based medicine is a new healthcare scientific paradigm aiming at the prevention, diagnosis and treatment of diseases using medical evidence. Because of the complex nature of the clinician's task to gather all the necessary knowledge about given diseases, the practice of evidence-based medicine could not be envisioned without IT support. Integration of external evidence-based data sources into the existing clinical information system and finding of appropriate therapy alternatives for a given patient and a given disease is a major research challenge. With the rapid changes taking place in the field of health care, decision support systems play an increasingly important role. We propose a data warehouse solution, as an easy-to-use decision support platform, to enable the integration of a wide range of data sources.

Caused by the innovative character of the field, no explicit common security regulations and standards yet exist. Defining these measures is a process, where both the patient's individual rights (patient's privacy and data protection) and the collective, societal demands (scientific progress and development of new technologies) need to be considered. This paper discusses the security measures to be enforced in order to protect the extraordinary sensitive nature of health data while using clinical data sources for knowledge discovery and the development of evidence based guidelines.

In this paper, we first give a short introduction into decision support systems and data warehouses. Subsequently we briefly present the relevant characteristics of evidence-based medicine. In the

* This research has been partly funded by the Austrian Federal Ministry for Education, Science and Culture, and the European Social Fund (ESF) under grant 31.963/46-VII/9/2002.

sequel, the paper sketches how data warehouse can be used to support the practice of evidence-based medicine by the physicians at the point of care as well as how it can support clinical management in their decision making process. Consequently, we propose a security concept for protecting privacy of high-sensitive patient data. This security concept comprises depersonalisation, pseudonymisation and role-based access.

This paper shows the need of a high-secure decision support system in order to facilitate the practical use of evidence-based medicine with respect to the privacy regulations. We consider the proposed data warehouse approach as a suitable solution for this problem.

1 Introduction

1.1 Data Warehouse

A Data Warehouse (DWH), as defined by Inmon [5], is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process.

The business success of an organisation is highly dependable on the proactive use of information which is stored in its operational systems. A DWH integrates the relevant information, originating from the diverse internal and external data sources. Data in a DWH is prepared for users with different analytical and software skills, and consequently also different types of requirements. Apart of responding to pure reporting and data analysis requests, a data warehouse also supports high-level users to track business trends, improve strategic decisions and enhance forecasting.

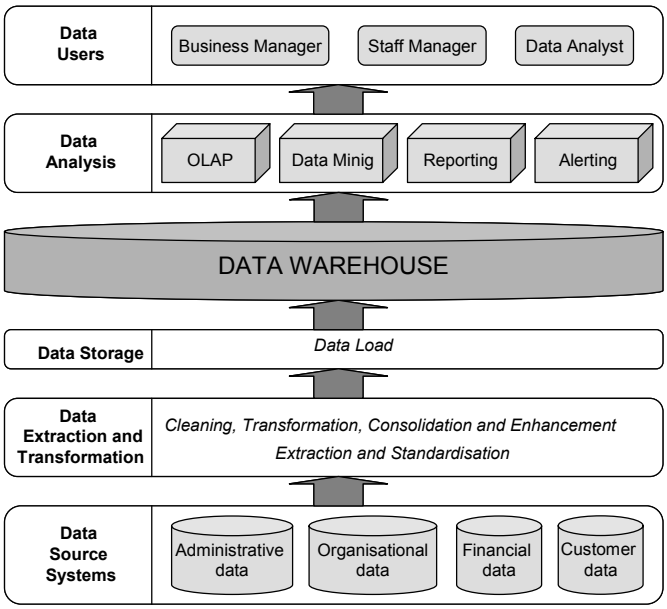

Figure 1: Data warehouse model

Figure 1 shows the entire data warehouse process, from collecting source data to the ultimate data delivery to the decision makers. Source data is usually originated from the very diverse operational systems, where it is stored in source-system-specific formats. In the *extraction phase,* data is directly accessed from the legacy systems or, indirectly in case of proprietary systems. In the *transformation phase*, selected data is cleaned and converted into the format and structure compatible to the one existing inside the data warehouse. Syntactic and semantic distinctions between operational sources are adjusted and local logical models are mapped and integrated into the global data warehouse data model. In the *data storage phase*, new data is loaded into the data warehouse and merged with the existing and historically stored data. After being integrated into the data warehouse, data is ready for querying and analysing by OLAP and data mining tools. Decision makers and staff managers are supported by predefined reports or can retrieve desired information in an ad-hoc manner. Furthermore, there is a trend to incorporate alerting mechanisms into the data warehouse for necessary proactive near real-time actions.

1.2 Evidence-Based Medicine

Most clinical practice is based on limited evidence, like textbook information, often defective research or case studies, unverified reviews and personal experiences.

Evidence based medicine is the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients. [8]. Evidence-based medicine aims at integrating the most recent research evidence into the existing clinical decision making process. Its task is not only to support the care givers in their daily practice on point of care, but also to support clinical management, administration, human resources and other departments in their decision making roles.

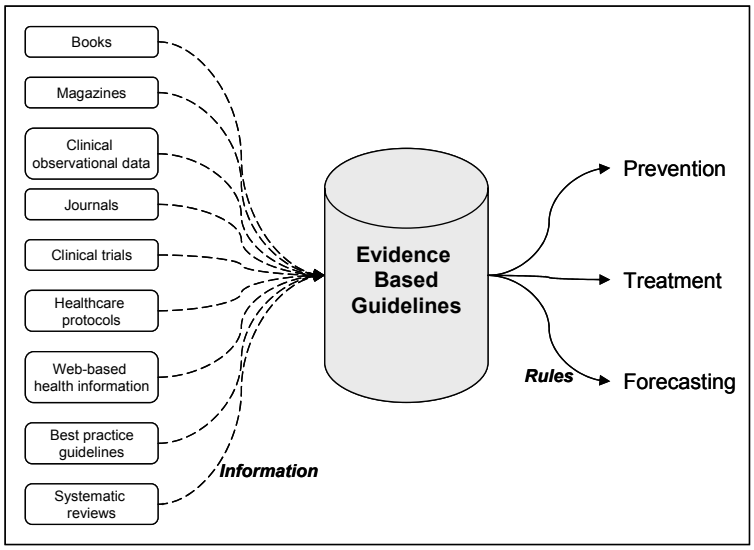

Figure 2: Data flow in evidence based medicine

In figure 2, the main information sources providing accurate medical evidence are presented. After the information is collected, it is analysed and used for the generation of evidence based guidelines.

Those guidelines are used for prevention and treatment of diseases including the forecast of relevant patient state parameters. The major goal of evidence-based medicine is that its scientific significant results will give a push on effective therapies measures to replace ineffective ones.

2 Data Warehouse Facilitating Evidence-Based Medicine

It is a complex task for clinicians to gather all the necessary knowledge about given diseases. Therefore the practice of evidence-based medicine would not be imaginable without IT support. The deduction of high-complex evidence data requires computer-based management systems, where decision support systems play an increasingly important role. Health care institutions are deploying data warehouse applications and decision support tools on top of them for their strategic decision making processes.

The use of data warehouses to facilitate the practice of evidence-based medicine promises to essentially improve health care quality. The main role of the clinical decision support systems is:

- To reduce medical errors
- To increase operating efficiency
- To reduce treatment costs (i.e. by avoiding duplicate examinations)
- To give advice about staffing plans etc.

Data warehousing enables the health care management to negotiate care contracts based on accurate data on resource utilization. Management can plan the use of their financial resources more efficiently through the implementation of clinical pathways by providing better resource utilization and by supporting physicians according to best practices based on evidence-based medicine.

A few representative examples of data warehouse applications in the area of evidence-based medicine should illustrate its usefulness:

- Data warehouses offer an appropriate platform for the *generation of evidence-based guidelines*.
Data originating from different sources related to health care (as shown in figure 1) is preprocessed and loaded into the data warehouse. Supported by complex data mining tools, knowledge workers are able to discover data patterns, which were unknown so far, to identify trends, and to recognise best practices for the different disease treatments. The gained hypotheses are then subject of thorough statistical analysis.

- The most common and important application field of data warehouse facilitating evidence-based medicine is the *support of clinicians on the point of care*. It means the support of the decision making process of a physician in his (her) clinical practice (figure 3), as well as deploying it for controlling of clinical treatment pathways [7].

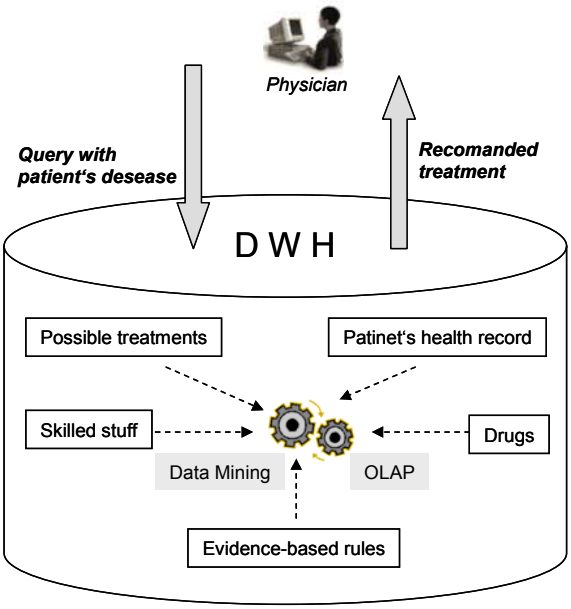

Figure 3: Decision support at the point of care

-Last but not least, data warehouse and its OLAP analytical functions are *used to support decision making processes of clinical management, human resources and clinical administration*. Evidence-based recommendations are used for creation of stuffing plans, treatment scheduling and for generation of business strategies that satisfy both patients' expectations and financial potential.

3 Security Concept for Healthcare Decision Support Systems

Over the past few years, we can observe an increase awareness of data privacy and protection in the healthcare sector. Healthcare decision support systems comprise large volumes of sensitive data and therefore must guaranty a high degree of data protection. Protection of high confidential patient data is subject of international regulations. Issues of privacy, software regulation and ethical and legal aspects of data processing in healthcare may build main sources of conflicts.

The usability of decision support systems is of extreme importance and therefore one has to prevent the danger, that the medical staff is overwhelmed by non-user-friendly security procedures.

Security risks need to be eliminated by implementing suitable technical and organisational measures [1]. In the following, we propose most important security measures, which need to be considered to protect data privacy in decision support systems in order to facilitate evidence based medicine:

1. All users of a healthcare decision support system must identify themselves through a password
2. Any data modification must bear a digital signature
3. Data access needs to be logged, and log files are preserved for a certain period of time, in order to enable later tracking of any data manipulation
4. Confidential health data should only be stored in a coded or encrypted form on a mobile medium

5. Transportation security must be assured through Public Key Infrastructure
6. Data used for evidence based medicine purposes needs to be depersonalised and pseudonymised
7. A role-based access model has to be implemented

We consider the last two security measures to be of a very special interest in the area of evidence-based medicine. The next section will therefore deal with these two measures.

3.1 Depersonalisation and Pseudonymisation

To obtain accurate patient data for the purposes of evidence-based medicine, security regulations and standards, which guaranty that the patient data will be handled confidentially, need to be designed. This kind of standards prevents that patients would presumably be alienated when disclosing their information to their physician. The lack of confidentiality bears the risk that patients would not allow the use of their medical records for scientific purposes. This would result an considerable effect of producing untrustworthy medical evidence. Currently, the Health Insurance Portability and Accountability Act (HIPAA) [4] and the European Commission's Directive on Data Protection [2] (which heavily take into consideration issues of confidentiality of patients' data) have created a great impact on the sharpness of security regulations.
The goal of evidence-based medicine is to recognise the symptoms, best treatments and prevention patterns for a given disease. Due to these data protection and secrecy objectives this goal can solely be accomplished by analyzing *unidentifiable* patient data. In this paper depersonalization and pseudonymisation procedures are used to prevent re-identification of personal data.

3.1.1 Depersonalisation

Taweel et al. [11] define depersonalisation as removal of any residual information that might risk identification – e.g. names of relatives, nick names, place names, unusual occupations, etc.
The goal of depersonalization is to assure the impossibility and impracticability of re-identification of a person by means of other personal data available. As stated by Stolba et al. [10], depersonalisation may be done by:

- *Grouping data* – hiding sensitive data through grouping (for example: patient's age is not shown precisely but in the age areas of 0-5, 5-10, 10-15, 15-20,…). The grouping of data is performed according to the research goals.
- *Hiding data* – all personal data interesting for detailed data mining (occupation, hobbies), which can potentially be used for patient identification are concealed.
- *Removing data* – key identifying data unnecessary for the research (e.g. name, exact birth day, precise address, nick names, name of relatives etc) that can be used for patient identification are removed.

The depersonalisation process begins by specifying sensitive data and its sensitivity levels. This is accomplished by the administrative users (most often: clinical management). Table 1 represents an example of specified sensitivity levels.

Entity	*Attribute*	*Sensitivity level*	*Depersonalization measure*
Patient	Name	Very High	Remove: very sensitive data, not supposed to be seen by anyone.
Patient	Date of birth	Medium	Group: Create new attribute "age" and group patients into following groups: 0-5, 5-10, 10-15, 15-20,…
Patient	Gender	Low	None, accessible by all users
Patient	Degree	High	Hide: highly sensitive data, may be seen only by authorized users.
Address	Street	Very High	Remove
Address	City	Medium	Group: Create new attribute "region" and group cities geographically (i.e. Steyr, St.Valentin, Linz = Upper Austria)

Table 1: Data sensitivity levels for depersonalisation

Attributes with the sensitivity level *very high* can easily identify the patient and are irrelevant for evidence based purposes. They are therefore not included into the decision support system. *High sensitive* data is interesting for sophisticated data mining activities, so it is hidden and can only be seen by authorized users. The attributes with a *medium* sensitivity level are transformed in order to protect their sensitiveness. The attributes with *low sensitivity* present no security risk, and can be accessed by all users.

3.1.2 Pseudonymisation

Pseudonymity is a state of disguised identity resulting from the use of a pseudonym. The pseudonym identifies a *holder*, that is, one or more human beings who possess but do not disclose their true names (legal identities) [12].

Pseudonymisation is especially suitable for the requirements of evidence-based medicine because it enables a consolidation of different patients' data without revealing patient identities.

Depending on the requirements, two kinds of pseudonymisation can be used:

1. one-way pseudonymisation (if no need for person re-identification exists)
2. reversible pseudonymisation (later re-identification by data owner is possible)

For purposes of evidence-based medicine *research*, no re-identification of a patient is needed, and one-way pseudonymisation can be applied. (More complex, reversible pseudonymisation forms have been described by GVG [3].) According to Austrian law [13], pseudonymisation must be performed by a trusted third-party organization.

A patient is uniquely identified by her (his) social security number, which is made secret by the means of pseudonymisation. Figure 4 represents privacy preserving measures during query processing in the data warehouse supporting evidence-based medicine. After the user has submitted the query to the data warehouse, and the answer has been retrieved, following three parts of the result can be recognised:
SSN – Social Security Number, which is used as unique patient identifier
PD – sensitive Personal Data, to which the user access is restricted
HCD – Health Care Data, which is non-sensitive medical data

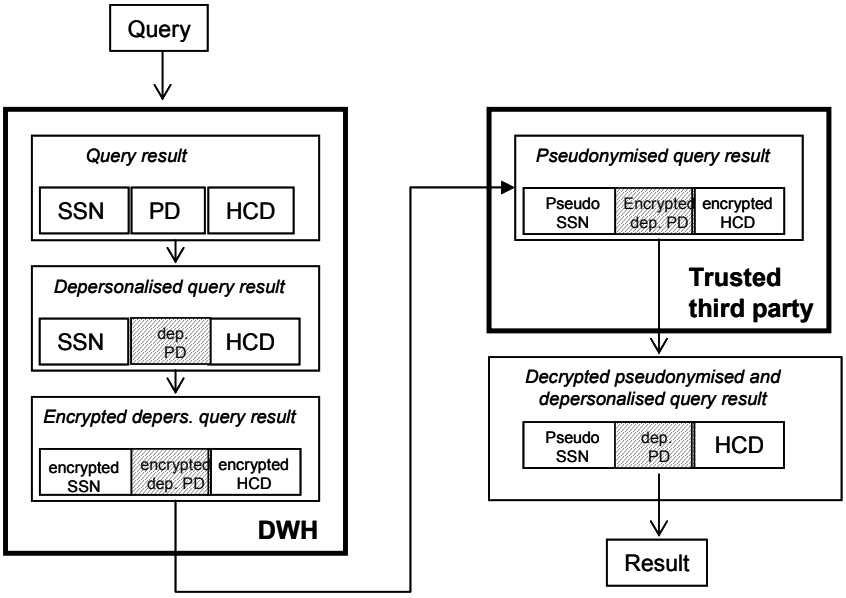

Figure 4: Data pseudonymisation process

The query result undergoes a data depersonalization process within the data warehouse system. Here, all sensitive personal data (PD) is grouped or hidden, so that it cannot be used to disclose the patient's identification. The depersonalized query result is encrypted for secured transportation to the pseudonymisation service. Pseudonymisation is performed by the trusted third party. After the SSN has been pseudonymised, the decrypted query result is delivered to the user.

3.2 Role-Based Access

The role based access model is used for decision support systems in order to ensure that in evidence-based medicine users can only access those data, which is granted to the role they have.

Stolba et al [9] state that the role should be regarded as a job description regardless of the actor performing it. Roles should exactly be assigned with those authorisations that are needed to fulfil the duties of the job. Each user in the data warehouse should be assigned to at least one role, though multiple roles are allowed. A user can play only one role at the time. This policy prevents authorisation conflicts among the roles of a user and it does not mean a limitation to real-life situations, as long as users can easily change their role due to the tasks needed to be fulfilled..
Since data warehouse offers read-only access, the role-based security model is also limited to it.

Figure 5 represents data access rights for five different roles, using evidence based rules to accomplish the specified duties. Each one of the roles (nurses, clinicians, human resources manager, administrative clerk and clinical management) has different data needs. Clinical management acts as a super-user and has access to the all data stored in the data warehouse. None of the users can identify the patient, since his (her) social security number has been pseudonymised and the rest of the data depersonalised.

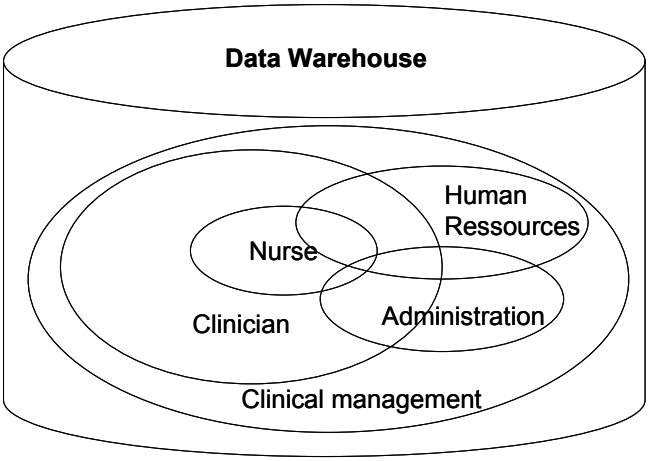

Figure 5: Role based access rights

4 Related Work

Stolba et al. [9] propose a security concept for OLAP, which is a role-based security model for data warehouses. The authors use a discretionary access model (DAC) to describe security rules as quadruple (s,a,o,p), where subject *s* has the access right *t* to access security object *o* within the range of predicate *p*. According to these security rules, a derived data cube is defined for each role. The dimensions and dimensions hierarchies of those data cubes are subsets of the un-restricted dimensions and dimensions hierarchies.
Priebe und Pernul [6] deals with the mapping of access schemes between relational data model in operational systems and non-traditional multidimensional models in OLAP systems. They approach the issue from the application side by introducing a methodology and a language for conceptual OLAP security design.
ATG [3] describe the pseudonymity and anonymity models for healthcare purposes. They distinguish between centralised and decentralised data store. For each alternative they propose different pseudonymity and anonymity models. The models presented by ATG are evaluated according to their security, administrational and organisational efforts.
The CLEF project described by Taweel et al. [11] aims to provide a pseudonymisation repository of histories of cancer patients that can be accessed by researchers. They outline that robust mechanisms and policies are needed to ensure that patient privacy and confidentiality are preserved while delivering a rich medical repository for the purposes of scientific research. They summarise the overall approach adopted by CLEF to meet data protection requirements..

5 Conclusion

Nowadays, not enough attention is paid to the protection of high sensitive patient data. Awareness in this issue needs to be widely promoted in the health sector. The application of the data warehouses in the area of evidence based medicine contributes to the improvement of healthcare quality and increasing efficiency. The system complexity, high amount of users and great data volumes residing in a medical decision support system are the main reasons for the security threats. In this paper, we

propose a security concept that needs to be implemented to eliminate these privacy risks. The proposed security approach allows the use of the valuable medical evidence data for knowledge discovery leading to the scientific progress in evidence based medicine.

6 References

[1] ECKHARDT A., Data Abuse and What Can be Done to Prevent it, The Patient in the Data Network, TA 36A/2000, http://www.ta-swiss.ch/www-support/reportlists/publicationsinfosoc_d.htm

[2] The European Commission's Directive on Data Protection , http://www.cdt.org/privacy/eudirective/

[3] GVG©, Gesellschaft fuer Versicherungswissenschaft und –gestaltung, Aktionsforum Telematik im Gesundheitswesen, Management-Papier „Pseudonymisierung/Anonymisierung“, Köln, 2004 []

[4] The Health Insurance Portability and Accountability Act (HIPAA), http://aspe.hhs.gov/admnsimp/pl104191.htm

[5] INMON W.H., Building the Data Warehouse, Second Edition, J.Wiley and Sons, New York, 1996

[6] PRIEBE T., PERNUL G., “A Pragmatic Approach to Conceptual Modeling of OLAP Security”, ER 2001: 20th International Conference on Conceptual Modeling, Yokohama, Japan, November 27-30, 2001, pp. 311-324, Springer Verlag

[7] ROEDER N. et al. Clinical Pathways, Medizincontrolling/ DRG Research Group, Universitätsklinikum Münster, http://drg.uni-muenster.de/de/behandlungspfade/cpathways/clinicalpathways_reisebericht.html

[8] SACKETT D.L., ROSENBERG W.M., GRAY J.A., HAYNES R.B., RICHARDSON W.S., Evidence-Based Medicine: what it is and what it isn’t [editorial]. BMJ. 196; 312 (7023) 71-72 [www.pubmed.com]

[9] STOLBA M., KIRKGÖZE R., KATIC N., TJOA A M., “A Security Concept for OLAP”, Proc. of the 8th International Workshop on Database and Expert System Application 1997, IEEE Computer Press, 1997

[10] STOLBA N., BANEK M., TJOA A M., The Security Issue of Federated Data Warehouses in the Area of Evidence Based Medicine, submitted to ARES2006, Vienna, Austria, February 20-22, 2006.

[11] TAWEEL A., RECTOR A., KALRA D., ROGERS J., et al, CLEF – Joining up Healthcare with Clinical and Post-Genomic Research, Healthcare Computing 2004: 203-211, BJHC Limited

[12] Wikipedia, http://en.wikipedia.org/

[13] 179. Bundesgestz, Gesundheitsreformgesetz 2005, 7. Unterabschnitt, § 84a. (5), Bundesgestztblatt für die Republik Österreich, 30.12.2004, Teil I, Seite 12.