Conference Track: IT Management in Healthcare

Authors:

Name:	**Nevena Stolba**, Dipl.-Ing.
Affiliation:	Woman's Postgraduate Collage for Internet Technologies Institute of Software Technology and Interactive Systems Vienna University of Technology
Address:	Favoritenstr. 9-11/188 A-1040 Vienna Austria
Phone:	+43 58801/18818
Fax:	+43 58801/18895
Email:	stolba@wit.tuwien.ac.at

Name:	**Tho Manh Nguyen**, Dr.techn.
Affiliation:	Institute of Software Technology and Interactive Systems Vienna University of Technology
Address:	Favoritenstr. 9-11/188 A-1040 Vienna Austria
Phone:	+43 58801/18801
Fax:	+43 58801/18899
Email:	tho@ifs.tuwien.ac.at

Name:	**A Min Tjoa**, O.Univ.Prof.Dipl.-Ing.Dr.techn.
Affiliation:	Institute of Software Technology and Interactive Systems Vienna University of Technology
Address:	Favoritenstr. 9-11/188 A-1040 Vienna Austria
Phone:	+43 58801/18800
Fax:	+43 58801/18899
Email:	amin@ifs.tuwien.ac.at

Towards a Data Warehouse Based Approach to Support Healthcare Knowledge Development and Sharing*

Abstract.

In the past, much effort of healthcare decision support systems were focused on the data acquisition and storage, in order to allow the use of this data at some later point in time. Medical data was used in static manner, for analytical purposes, in order to verify the undertaken decisions. Due to the massive economical impact of today's health system, great changes in medical treatments are notable. Apart of the humanitarian and healing nature of medicine, this industry is becoming more and more business like. The exploitation of evidence-based guidelines becomes a priority concern, as the awareness of the importance of knowledge management rises.

Consequently, interoperability between medical information systems is becoming a necessity in modern health care. Under strong security measures, health care organisations are striking to unite and share their (partly very high sensitive) data assets in order to achieve a wider knowledge base and to provide a matured decision support service for the decision makers. Ontological integration of the very complex and heterogeneous medical data structures is a challenging task. Our objective is to point out the advantages of the deployment of a federated data warehouse approach for the integration of the wide range of different medical data sources and for distribution of evidence-based clinical knowledge, to support clinical decision makers, primarily clinicians at the point of care.

Keywords: data warehouse, clinical decision support systems, evidence-based medicine, knowledge management, wrapper, mediator, ontological integration

1. Introduction

Despite the scientific and technological development progress over the recent years in the healthcare delivery, a significant portion of the decision-making information on the treatment of a patient's illness is still based on unstructured information or even hand-written notation. Furthermore, patient records are distributed over many different locations, instead of being available when required at the point of care.

* 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.

We propose a data warehouse (DWH) based approach, in order to enhance the process of clinical knowledge management, which is a prerequisite for a successful clinical decision support. The benefits of such a solution are:

- The most recent medical knowledge can be provided, in order to take treatment decisions, i.e. enabling the practice of evidence-based medicine. Due to the huge data volumes, a DWH approach is required for the production of evidence-based knowledge.
- Clinical business management can perform the desired knowledge-oriented strategic decision making, planning and management of the healthcare enterprise. The financial resources can be used more efficiently through the implementation of DWH-based clinical pathways. Henceforth, clinical business management is then enabled to negotiate appropriate care contracts based on accurate knowledge on resource utilization.
- Clinical DWH, which gathers all available data related to the patients, treatments and drugs they received, will represent an ideal knowledge base for research purposes.

The addressed issues are very demanding and comprehensive. In this paper we confine ourselves to handle one aspect – the goal of our paper is to point out the advantages of the deployment of DWH for integration of wide range of different data sources and for distribution of evidence-based clinical knowledge to support clinical decision makers.

The contribution of our work is to propose a DWH based, ontologically focused integration model for integration of heterogeneous health care data sources, which could even include external data suppliers or some third-party stakeholders' web services.

The remainder of the paper is organized as follows: in section 2, we present a short introduction into the concepts of DWH facilitating evidence-based medicine. An integrated health care information system, built upon a federated DWH is described in details in section 3. Section 4 outlines the advantages of the application of DWH technology for the support of care givers at the point of care via a case study. Section 5 deals with related work and finally, we conclude the paper in section 6.

2. Data Warehouse Facilitating Evidence-Based Medicine

As discussed in related work [5], 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, having different types of requirements. A DWH should not only respond to pure reporting and data analysis requests but also support high-level users to track business trends, improve strategic decisions and enhance forecasting.

David Sackett [11] defines evidence-based medicine (EBM) as the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients. Published clinical evidence is not necessarily sufficient for providing integrated support to care givers. EBM complements existing clinical decision making process with the most accurate and most efficient research evidence. Application of EBM concepts speeds up the transfer of clinical research findings into practice, leading to cost reduction and to the improvement of the healthcare process as the whole.

Both medical institutions and health insurance companies are primarily interested in increasing the patient healing rate and in reducing treatment costs. The right use of data warehousing in the area of EBM could prove as economically highly advantageous (in the long term) by the avoidance of duplications of examinations, by the automation of routine tasks and by the simplification of accounting and administrative procedures.

3. Federated DWH Model for Integration of Medical Information Systems

During the last few years, healthcare organizations are confronted with massive knowledge processing challenges primarily caused by the increasing amount and complexity of medical data. DWH offer a comprehensive support for gathering, analyzing and presenting medical data.

The system complexity, the heterogeneity of health care data sources, massive volumes of medical data and high number of concurrent users are the main reasons for the use of federated DWH integration model. As shown in Figure 1, the central point of such an integrated system is a federated DWH where all participants have access to and where the user's conceptual view is preserved.

In related work [10], we discussed the advantages of the federated approach against the centralized approach, in the aspects of cost-effectiveness, security and usability. A federated approach [1], [3] is required, when several independent organisations share their data for mutual purposes, but do not allow any physical copy of their data to be created in any external system. Most often, they demand to maintain full control of the access to their data. Highly confidential healthcare records are a typical example of such data. In case of interoperability in the health care domain, a federated DWH is therefore always a preferred solution.

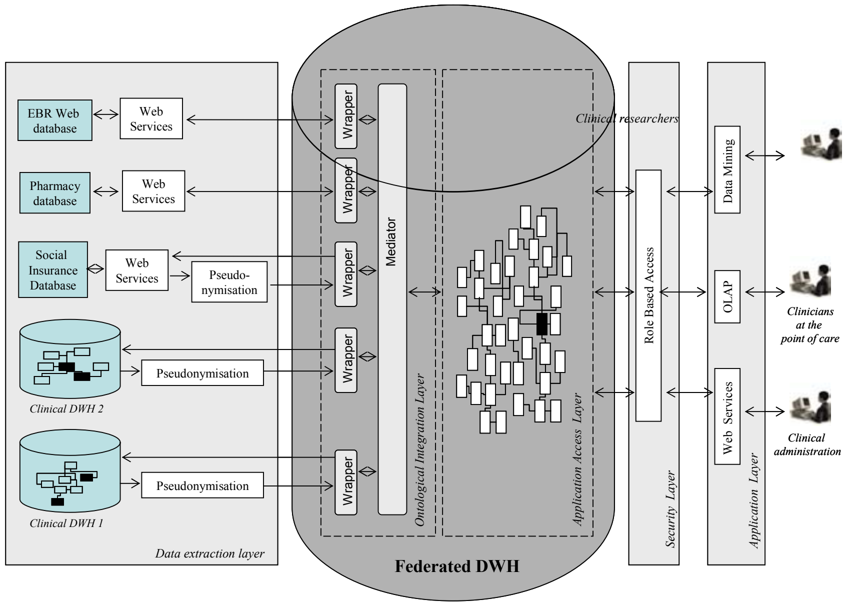

Figure 1: Medical federated DWH Integration Workflow

3.1. Participators of DWH Federation

In our DWH model, the different medical treatment domains, i.e. the social insurance domain, the pharmaceutical domain and the evidence-based guidelines repository are participating in one federation. The existence of the federation is invisible to the users of the source systems.

The clinical DWH contains data which originates from a wide variety of data sources, such as:

- clinical data (patient, pharmaceutical, medical treatment, length of stay),
- administrative data (staff skills, overtime, nursing care hours),
- financial data (i.e. treatment costs, staff salaries, accounting)
- organizational data (facilities, equipment, room occupation)

The health insurance company stores information about patient encounters, treatments, therapies and drug prescriptions it supports. It communicates with the federation via web services.

Both clinical DWHs and the health insurance transfer their sensitive data to the federation, in case of a federated query. We assume that depersonalisation and pseudonymisation techniques are used to protect the confidentiality of patient data, as described in related work [2],[4],[10],[15],[16].

The pharmaceutical sector provides the federation with drug information such as medication description, packaging size, pharma-id number, indication group, pharmaceutical form (pills, juice etc.), and medication fee. In an analogous way to the insurance company, it uses web services to provide the federation with the necessary data.

The evidence-based rules repository is a collection of most accurate and most efficient research evidence. Through the web services, it provides the federation "on demand" with the best fitting guidelines for a given patient and given disease.

Since only a "single version of evidence" and a unique interpretation of the joined data should exist, it is necessary to have a unique singular common federated schema. For building the conceptual model of the federated DWH, just data relevant for further analyses and reporting are considered. In this phase, business users (from the domain of clinical/social insurance management) have to specify the respective sensitivity levels of data. The data modeller incorporates the specified privacy restrictions into the resulting logical data model.

3.2. Ontological Integration

The essential part of the integration of logical schemas of the underlying DWHs as well as of the data structures originating on the diverse participating legacy systems (such as relational or XML databases), is the ontological integration layer, as shown in the model depicted in Figure 1. Our model includes wrappers and mediator, which are two main architectural components of a mediated query system.

Wrappers encapsulate local data sources and export their functionalities and the metadata stored therein. They accept queries in a certain language and return metadata in a united form [12],[13]. The wrapper keeps locally the data schema for the specific data source it deals with. By integrating wrappers, we can cope with technical heterogeneities among local systems, without having to modify them.

The mediator [12], [13] handles the global queries from the application layer, unfold them into sub-queries and disperse these sub-queries to the relevant local data sources via their wrappers. The local results will be returned from wrappers; the mediator finally combines and presents the result to the client. Hence, the mediator will keep the global data schema and the mapping between global and local schemas. To maintain the dynamic mapping between local and global schemas, an ontological-based mediator/wrapper is one of the interesting problem solving approaches [13].

Figure 2 shows an example of how a user's query is handled by the wrappers and the mediator, from the query submission to the presentation of the result. On receiving the SQL query, the system performs the following:

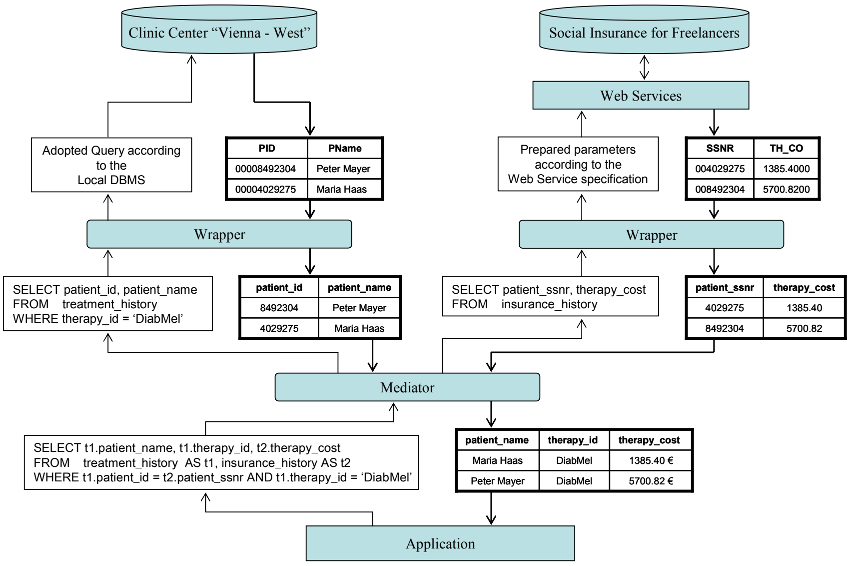

Figure 2: Mediated query system in health care environment

1. DWH-Application invokes the mediator.

Query unfolding:

2. The mediator resolves the submitted query into partial queries, according to the exported schemas previously exposed by the wrappers. It determines which wrappers are relevant to the correspondent sub-queries.

3. The mediator passes the sub-queries to the affected wrappers.

4. A wrapper receives its sub-query and translates it into the format so that it can be understood by the underlying data source (database, web service etc.).

5. The wrapper forwards the adapted sub-query to the local DBMS or to the responsible Web Service for execution.

Query answering:

6. A wrapper retrieves answer data set, translates it into its exported schema.

7. The wrapper passes its answer data set to the mediator.

8. The mediator integrates partial query results into one answer set, transforms and formats it, so that it can be processed by the application.

9. The mediator passes the answer set to the application, so that it can be presented to the user.

The users of a federated DWH are not aware of the fact, that the data they are querying may be distributed across the network. Through data mining tools, web services, ad-hoc queries and predefined reports (OLAP tools), users are able to analyse data, as if they were physically stored in a centralised DWH. A role-based access model [14] guaranties that each user get access only to those data, which are necessary for the performing of his (her) tasks.

4. Case Study: DWH Supporting Clinician at the Point of Care

To illustrate the described DWH-based approach we present a case study, in which we suppose that the emergency room clinician is querying the federated DWH (as described in chapter 3) while examining the patient. The following scenario, which has not yet been implemented, builds the starting point of our case study: The diabetes patient, suffering from a progressive liver disease is complaining about itchy rash on her hands. Since the attending physician is not familiar with the patient's medical history and needs to handle quickly, he(she) is using federated DWH facilitating EBM, to find the most efficient therapy, which does not conflict with the patient's ongoing diabetes and liver disease treatment.

The clinician is using an OLAP tool, which is set up on the federation logical data model (LDM). One small but representative part of this model is shown in Figure 3. (In this paper, we are not handling data mining or web services based data retrieval. This is a part of an ongoing project and will be published in our future work.) Dashed arrows show the data flow between underlying data sources and resulting tables. Since the federation exists only on a virtual level, data are physically stored in their originating sources but queried and presented through the federation model.

The clinician is querying the patient's healthcare record, containing all the existing anamnesis/diagnostic data and all the patient's treatments in the past. Further, he(she) is interested in the patient's overall health condition (allergies and medication incompatibilities) as well as the personal data (age, weight, family predisposition to some diseases etc.). In addition, the clinician has to find out what kind of therapies will be covered by the patient's social insurance. Finally, he(she) is aiming at finding the treatment which proved to be the most effective under the given conditions.

The answering procedure takes place in two phases:

1. The user (the clinician) is querying the federated DWH by providing only the patient's name and the corresponding social insurance number. In the first step, the mediator sends queries containing these two parameters to all relevant participating sources (in this case study, these are: clinical DWHs,

pharmacy and social insurance company). The aim of this step is to retrieve all patient data, which might be interesting as input for the second step, namely querying EBG database.
Responsible wrappers return the corresponding data to the mediator, which in sequel joins them and produces a result data set.

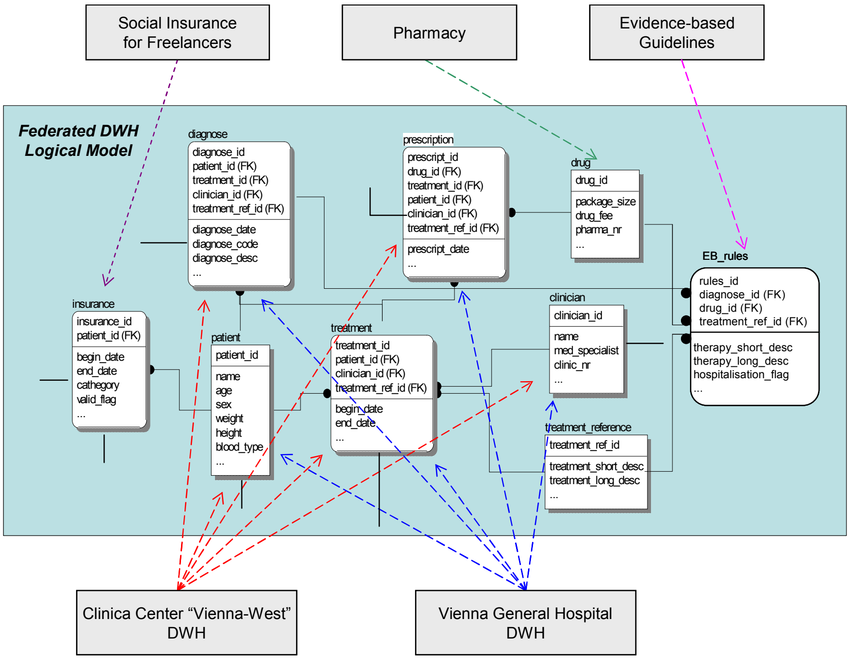

Figure 3: Extract from federated DWH LDM with belonging data sources

2. In the second step, the resulting data set (containing patient's age, blood pressure, blood type, diseases history, list of received treatments and medications, social insurance categorisation etc.) is used as input parameter for the querying of the evidence-based guidelines database. The mediator forwards a new query to the EBG- database wrapper and retrieves the final data set.

As shown in Figure 4, some of the existing treatment rules are disqualified due to medication or treatment incompatibilities. Since the proposed treatments must be adjusted to the patient's health risks, parameters like blood pressure may play the determining role in the treatment verification process. Nevertheless, the scope of a patient's social insurance contract (refers to treatment categorization) is significant for the determination of applicable treatment.

Patient

patient_id	name	age	sex	blood_type	blood_press_h	blood_press_l	...
4029275	Maria Haas	65	F	AB	140	80	...

Diagnoses

diagnose_id	patient_id	treatment_id	diagnose_date	diagnose_code	diagnose_desc	...
DM114	4029275	178756	10.04.1998	DiabMel	Diabetes Mellitus	
LD323	4029275	556802	22.10.2004	LeavDis	Leaver Disease	

Treatment

treatment_id	patient_id	treat_ref_id	treat_code	begin_date	end_date	cathegory	...
178756	4029275	0770	339025	12.01.2000	12.01.2000	4	
321321	4029275	1035	**920572**	04.02.2000	03.08.2000	3	

EBrule

EBG_id	diag_id	probl_treat_code	drug_incomp	bp_high_lim	bl_low_lim	reco_treat_code	...
4450241	DM114	**920572**	857215	NULL	NULL	222435	...
4450242	DM114	592688	NULL	150	60	823044	...

Figure 4: A part of tables involved into finding of the best fitting evidence-based rule for a given patient and given diseases.

In our example, the evidence-based guideline with identification: EBG_id = '4450241' is not applicable, for the reasons of treatment incompatibility. Namely, in the past, Mrs. Haas has received a treatment with treatment_code = '920572', which is listed as problem treatment for this guideline.

The evidence-based guideline with identification: EBG_id = '4450242' is the best fitting guideline for Mrs. Haas's medical condition, so it is forwarded to the federated DWH.

In the last step, an OLAP tool presents the result to the clinician in an understandable and illustrative way.

5. Related Work

Abidi et al. describe in [6] an Integrated Clinical Evidence System designed to augment the typical literature–based clinical evidence with additional technology-mediated clinical evidence. They propose a technology-enriched strategy to exploit advance computer technologies – knowledge management, data mining, case-based reasoning strategies and internet technology – within traditional evidence based medicine systems to derive all-encompassing clinical evidence derived from heterogeneous clinical evidence modalities.

The four steps in incorporating the best available research evidence in decision making is the subject of the research project in [9]. The authors formulate the following steps: asking answerable questions; accessing the best information; appraising the information for validity and relevance; and applying the information to patient care. Further, they state that applying evidence-based medicine to individual patients requires drawing up a balance sheet of benefits and harms based on research and individual patient data. Wu et al. state in [8] that growing evidence indicates that the integration of clinical decision support into the computer-based patient record can decrease medical errors, enhance patient safety, decrease unwanted practice variation and improve patient outcomes. Clinical Pathways are the subject of research of Roeder et al., at the DRG Research Group at the Universitätsklinikum Münster [7]. They investigated 8 different international DRG-systems on the basis of data from cardiac surgery and concluded that the Australian AR-DRG-system excellently matches levels of complexity. Thus it provides a good basis for the German R-DRG-system, which will serve for the reimbursement of all in-patient cases, according to the German Ministry of Health.

A self-medication information system, which proposes to patients information and services on mild clinical signs and associated treatments, is illustrated in [17]. Given the simplified patient's electronic health record as input, an ontology is used to infer the right treatment proposal out of the self-medication knowledge base.

Integrating the Healthcare Enterprise (IHE) [18] is an initiative designed to stimulate the integration of the information systems that support modern healthcare institutions. It defines a technical framework for the implementation of established messaging standards to achieve specific clinical goals. The Cross-Enterprise Document Sharing (XDS) provides a general mechanism for sharing of documents between different healthcare enterprises. The main difference between IHE and our approach is that it does not facilitate statistical and in-depth analysis of clinical data, nor it offers pattern recognition capabilities.

6. Conclusion

The collaboration project for evidence-based medicine described in this paper merges data originating in a pharmacy database, a social insurance company database and diverse clinical DWHs. Our universal, simple and flexible common conceptual model enables potential future integrations of other health care organisations to be done seamlessly and with a minimum effort.

In this paper we showed, that a federated clinical DWH that facilitates evidence-based medicine is a reliable and powerful platform for production and dissemination of clinical knowledge. Knowledge originating from evidence-based medicine represents a valuable resource for healthcare policy makers. Integration of knowledge management into clinical decision support system enables strategic decision making for both clinical business management and for the caregivers at the point of care, which results in a better service for the patient, the medical personal and administrative staff.

Acknowledgement

- 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.

- This research has been supported by the Austrian National Bank (ZELESSA: An Enabler for Real-time Business Intelligence -Projectno. 11806)

- We gratefully acknowledge the contributions and support of SVA (Sozialversicherungsanstalt der gewerblichen Wirtschaft, Vienna, Austria) team members, in particular Dr. Thomas Mueck, for supplying us with heterogeneous data models and for their cooperativeness at advising us with their health insurance expertise.

References

[1] Sheth A.P., Larson J.A.: "Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases", *ACM Computing Surveys*, Vol. 22 (3), ACM Press, New York, USA, 1990, pp. 183-236.

[2] Clifton C., Doan A.H., Elmagarmid A., Kantarcioğlu M., Schadow G., Suciu D., Vaidya J.: "Privacy-Preserving Data Integration and Sharing", *Proceedings of the 9th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD 2004)*, ACM Press, New York, USA, 2004, pp. 19-26

[3] Jindal R., Acharya A.: "Federated Data warehouse Architecture", Wipro Technologies white paper, 2004, http://hosteddocs.ittoolbox.com/Federated%20data%20Warehouse%20Architecture.pdf

[4] Kalra D., Singleton P., Ingram D., Milan J., MacKay J., Detmer D., Rector A., "Security and Confidentiality Approach for the Clinical E-Science Framework (CLEF)", HealthGrid 2004 Conference, *Methods of Information in Medicine*, Vol. 44 (2), 2005, pp. 193-197

[5] Stolba N., Tjoa A M.: An Approach towards the Fulfilment of Security Requirements for Decision Support Systems in the Field of Evidence-Based Healthcare, Knowledge Rights - Legal, Societal and Related Technological Aspects (KnowRight2006), Vienna, Austria, Austrian Computer Society, pp. 51-59, February 2006.

[6] Abidi S.S.R., Abidi S.R., "A Case of Supplementing Evidence Base Medicine with Inductive Clinical Knowledge: Towards Technology Enriched Integrated Clinical Evidence System", 14th IEEE Symposium on Computer-Based Medical Systems, (CBMS'2001), 26-27 July 2001, Bethesda (USA).

[7] Roeden N. et al., "Clinical Pathways", Medizincontrolling/DRG Research Group, Universitätsklinikum Münster, http://drg.unimuenster.de/de/behandlungspfade/cpathways_reisebericht.html

[8] Wu R., Peters W., Morgan M.W., "The next generation of clinical decision support: linking evidence to best practice", J Healthc Inf Manag 2002 Fall;16(4):50-5.

[9] Craig J.C., Irwing L.M., Stockler M.R., "Evidence-based medicine: useful tools for decision making", The Medical Journal of Australia [http://www.mja.com.au], MJA 2001; 174:248-253M. Young, The Techincal Writers Handbook. Mill Valley, CA: University Science, 1989.

[10] Stolba N., Banek M. and Tjoa A M., “The Security Issue of Federated Data Warehouse in the Area of Evidence-Based Medicine”, ARES 2006, The First International Conference on Availability, Reliability and Security, April 20-22 2006, Vienna

[11] Sackett D.L., Rosenberg W.M.C., Gray J.A.M., Haynes R.B., Richardson W.S., “Evidence-Based Medicine: what it is and what it isn’t”. *British Medical Journal (BMJ)*, Vol. 312 (7032), 1996, pp. 71-72

[12] Beneventano D., Bergamaschi S., Guerra F., Vincini M.: "Synthesizing an Integrated Ontology", IEEE Internet Computing Magazine, September-October 2003,42-51.

[13] Beneventano D., Bergamaschi S.: “Semantic Search Engines based on Data Integration Systems”, In Semantic Web: Theory, Tools and Applications (Ed. Jorge Cardoso), Idea Group Publishing, 2006.

[14] Stolba N., Tjoa A M.: An Approach towards the Fulfilment of Security Requirements for Decision Support Systems in the Field of Evidence-Based Healthcare, Knowledge Rights - Legal, Societal and Related Technological Aspects (KnowRight2006), Vienna, Austria, Austrian Computer Sociaty, pp. 51-59, February 2006

[15] Pommerening K., Reng M., “Secondary Use of the Electronic Health Record via Pseudonymisation”, *Medical and Care Compunetics 1*, Studies in Health Technology and Informatics, Vol. 103, IOS Press, Amsterdam, The Netherlands, 2004; pp. 441 - 446.

[16] Priebe T., Pernul G., “A Pragmatic Approach to Conceptual Modeling of OLAP Security”, *Proceedings of the 20th International Conference on Conceptual Modeling (ER 2001)*, Lecture Notes in Computer Science, Vol. 2224, Springer Verlag, Berlin-Heidelberg, Germany, pp. 311-324.

[17] Curé, O.: *Ontology Interaction with a Patient Electronic Health Record*. In: Proc. Symp. Computer-Based Medical Systems (CBMS’05). IEEE Computer Society (2005) 323-32.

[18] IHE, http://www.ihe.net/Technical_Framework/, published on June 9, 2006, last checked on December 28, 2006.