International Journal of Medical Informatics (2006) 75, 201—208

International Journal of Medical Informatics

www.intl.elsevierhealth.com/journals/ijmi

Flexible information storage in MUDR^{II} EHR

Josef Špidlen*, Petr Hanzlíček, Antonin Říha, Jana Zvárová

European Centre of Medical Informatics, Statistics and Epidemiology, Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod Vodarenskou vezi 2, 182 07 Prague 8, Czech Republic

KEYWORDS
Electronic health record;
Healthcare;
Software architecture;
Graph theory

Summary An important research task of the EuroMISE Centre is the applied research in the field of electronic health record (EHR) design including electronic medical guidelines and intelligent systems for data mining and decision support. The research in this field was inspired by several European projects. We have proposed a mathematical meta-description of a flexible information storage model based on the experience gathered in cooperation in those projects. In this model, we use two basic structures called a knowledge base and data files. We describe those two structures using the graph theory concepts. Furthermore, we use logical formulas to express conditions that should be valid. Additionally, we present a description of a global system architecture of a 3-tier EHR application with interfaces based on the latest technologies; predominately on Web Services, SOAP, XML, HTTP, CORBA, etc. According to our experience and test results gained from the MUDR EHR usage, we describe an open universal solution, which can be applied as the EHR kernel of hospital information systems. To realize this approach in a daily practice for health professionals we have started a co-operative project with clinical information systems developers. Within that project we are developing a new system for continual shared health care.

1. Introduction

The European Centre for Medical Informatics, Statistics and Epidemiology—Cardio (EuroMISE Centre—Cardio) is focused on new approaches to the electronic health record (EHR) design, including electronic medical guidelines and intelligent systems for data mining and decision support [1]. The research in the field of data storage, data acquisition and data mining [2] was inspired by several European projects [3], mostly by the I4C and TripleC projects [4,5]. Study of the CEN TC 251 standards and cooperation with physicians resulted in a list of 15 requirements on EHR [6]. The proposed model was mostly influenced by two of them:

- A structured way of data storage combined with a free text.
- A possibility of dynamic extension and modification of the set of collected attributes without any change of the database structure.

The main goal of our work was to suggest common general principles to increase the quality of EHR systems, to simplify data sharing and data

* Corresponding author.
E-mail address: spidlen@euromise.cz (J. Spidlen).
URL: http://www.euromise.org.

1386-5056/$ — see front matter © 2005 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.ijmedinf.2005.07.044

migration among various EHR systems and to help to overcome the classical free-text based health record. We did not want to choose a particular database or an operating system; we tried to propose an open information storage meta-model with various implementation possibilities as inspirations and recommendations for EHR software vendors. To test the functionality of our solutions, we have developed a pilot EHR application called ''MUltimedia Distributed Record'' (MUDR) [7].

2. Flexible information storage model

Because of the requirement of a dynamically extensible and modifiable set of collected attributes, it is complicated to use as the basis of the information storage a classical relational database structure with columns corresponding to the gathered features. In our solution, two main structures described by the graph theory expressions are used instead. The collected attributes and relations among them are stored in a so-called knowledge base. Another graph structure named ''data-files'' is used to store the patient data itself.

2.1. The knowledge base

The main function of the knowledge base is to store the hierarchy of the collected attributes and relations among them. We define the *MUDR KB* knowledge base as an oriented graph:

$$MUDR\,KB = (\mathbf{V}_{kb}, \mathbf{E}_{kb}).$$

The graph vertices, representing symptoms, elements $n \in \mathbf{V}_{kb}$, are called knowledge nodes. Each node is a quaternion:

$$n = (\gamma, \varphi, \omega, \varepsilon_{nd})$$

where γ is a unique identifier among all knowledge nodes, φ is a name constructed as a mnemonic string identifier, ω is a data type of the node and ε_{nd} contains administrative data about the node such as identification of the user who has entered or deleted the node. Later in this text we use the square brackets [ ] to index the items of ordered sets.

Edges $e \in \mathbf{E}_{kb}$, representing relations between symptoms, are also defined as quaternions:

$$e = (\alpha, \beta, \tau, \varepsilon_{ed})$$

where $\alpha, \beta \in \mathbf{V}_{kb}$ determine starting and ending vertices of the edge, τ is the edge type determining the type of the relation between α and β and ε_{ed} contains administrative information about the edge e (e.g. the creator of the edge).

A special edge type called inferior (inf) can be distinguished. The edge of this type leads from a parent vertex to a child vertex in the graph. Thereby a hierarchical relation on knowledge nodes is set. Using other edge types we can add additional medical knowledge into the knowledge base like equivalence, contraindications, usable scales, etc. Their detailed description is out of the scope of this text. The knowledge nodes and the edges of the type inferior create an oriented forest with a few trees. These trees are called *knowledge base domains*. Each domain associates nodes used to a similar aim. The domain $\mathbf{V}_s$, $\mathbf{V}_s \subseteq \mathbf{V}_{kb}$, is used to store collected attributes of a patient. Nodes in this domain are called *semantic types*. Other domains can be used for example to store the International Classification of Diseases and Related Health Problems (ICD10) or the Anatomical Therapeutic Chemical Classification of drugs (ATC).

Nodes with the same parent are called siblings — $n_1 \lozenge n_2$.

$$n_1 \lozenge n_2 \overset{\text{def}}{\Leftrightarrow} \exists e_1,\ e_2 \in \mathbf{E}_{kb},\ e_1[\alpha] = e_2[\alpha],\ e_1[\beta] = n_1,\ e_2[\beta] = n_2,\ e_1[\tau] = e_2[\tau] = \text{''inf''}$$

Nodes without any parents are called knowledge domain roots. For $n \in \mathbf{V}_{kb}$ we define:

$$n \in \mathbf{R}_{kb} \overset{\text{def}}{\Leftrightarrow} (\forall e \in \mathbf{E}_{kb},\ e[\tau] = \text{''inf''} \Rightarrow e[\beta] \neq n)$$

We require that the name φ is unique among siblings. The same we require among domain roots:

$$(n_1 \lozenge n_2) \vee (n_1, n_2 \in \mathbf{R}_{kb}) \Rightarrow (n_1 = n_2) \vee (n_1[\varphi] \neq n_2[\varphi])$$

This uniqueness enables constructing a dot-separated node full name, which can be used for identification of a knowledge node among other knowledge nodes.

The node data type is especially important for the semantic types. We distinguish basic data types like numbers, strings or Boolean variables, multimedia data types like pictures, audios and videos and reference data types. In the latest MUDR^{II} version we have added some useful data types like enumerations and extended the group of multimedia data types. Now the Multipurpose Internet Mail Extension Type (RFC 2048) is used to determine the multimedia type.

2.2. The data files

Patient data themselves are stored in a graph structure *MUDR DF*, realized as an oriented forest:

$$MUDR\ DF = (\mathbf{D}_{df}, \mathbf{E}_{df})$$

Information about one patient corresponds exactly to one tree in the forest. A graph vertex $d \in \mathbf{D}_{df}$ is a quaternion:

$$d = (\delta, \sigma, \lambda, \varepsilon_{df})$$

In this case δ is a unique identifier among all data files, $\sigma \in \mathbf{V}_s$ is a semantic type of the data file, λ is a data file value and ε_{df} contains administrative information like the identification of the person who has entered or confirmed the stored value. The domain of the value is implicitly determined by the semantic type, including the data type within.

The edges $e \in \mathbf{E}_{df}$ do not have any types, they express the ''parent—child'' hierarchical relation among stored data and $\mathbf{E}_{df} \subseteq \mathbf{D}_{df} \times \mathbf{D}_{df}$. For the *MUDR DF* we require two expressions to be valid:

- $\forall d \in \mathbf{D}_{df} (\exists d' \in \mathbf{D}_{df}, (d', d) \in \mathbf{E}_{df} \vee d[\sigma] \in \mathbf{R}_{kb}$
- $d, d' \in \mathbf{D}_{df}, (d', d) \in \mathbf{E}_{df} \Rightarrow \exists e \in \mathbf{E}_{kb} e[\alpha] = d[\sigma],$ $e[\beta] = d'[\sigma], e[\tau] =$ '' inf ''

The first condition means that each data file has a parent or its semantic type is a domain root. The second one expresses that if a data file d has a parent $d\check{\imath}$, then the semantic type of d is the child of the semantic type of $d\check{\imath}$. It simply means that the *MUDR DF* corresponds to the *MUDR KB*. A little example can be seen in Fig. 1.

The equivalent information presented in a user interface called MUDRc can be seen in Fig. 2. The lines connect the data files on the right side with knowledge nodes on the left. At the bottom an example of attributes creating the ε_{nd} and ε_{ed} administrative data about the node and data file can be seen.

3. MUDR architecture

3.1. Classical MUDR architecture

The MUDR EHR is based on a three-layer architecture with a data layer, an application layer and a user interface. This decomposition enables separating different system modules to small functional parts, which makes the system more flexible. In our system, we define the global architecture with communication interfaces based predominately on XML and HTTP. We also define an interface for connecting medical guidelines formalized in the form of dynamic libraries. These libraries extend the capabilities of the MUDR EHR.

3.2. MUDR^{II} extensions

MUDR^{II} architecture extends the basic three-layer architecture in the way shown in Fig. 3. The

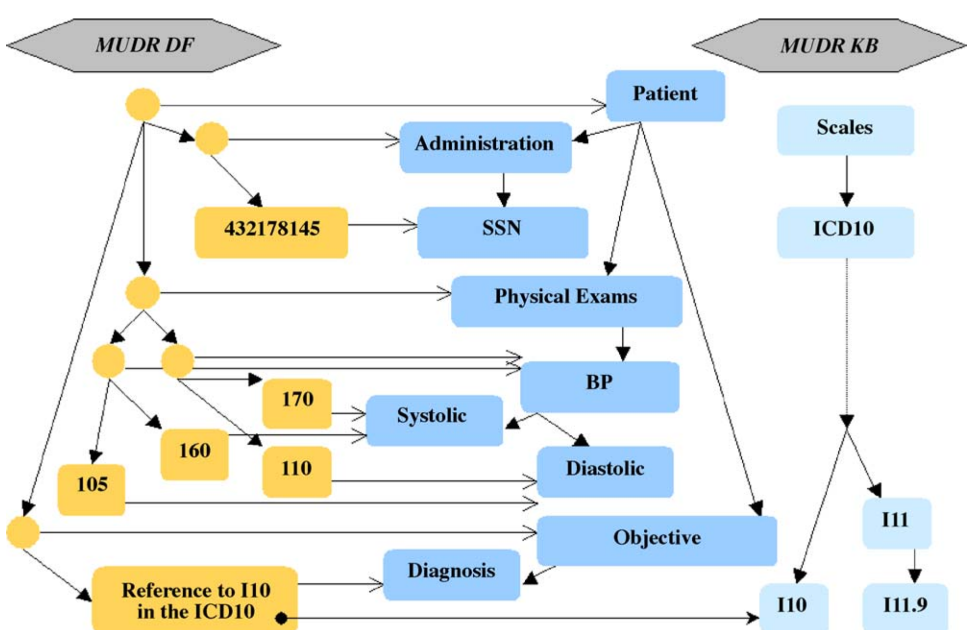

Fig. 1 Fragment of the MUDR information storage Model.

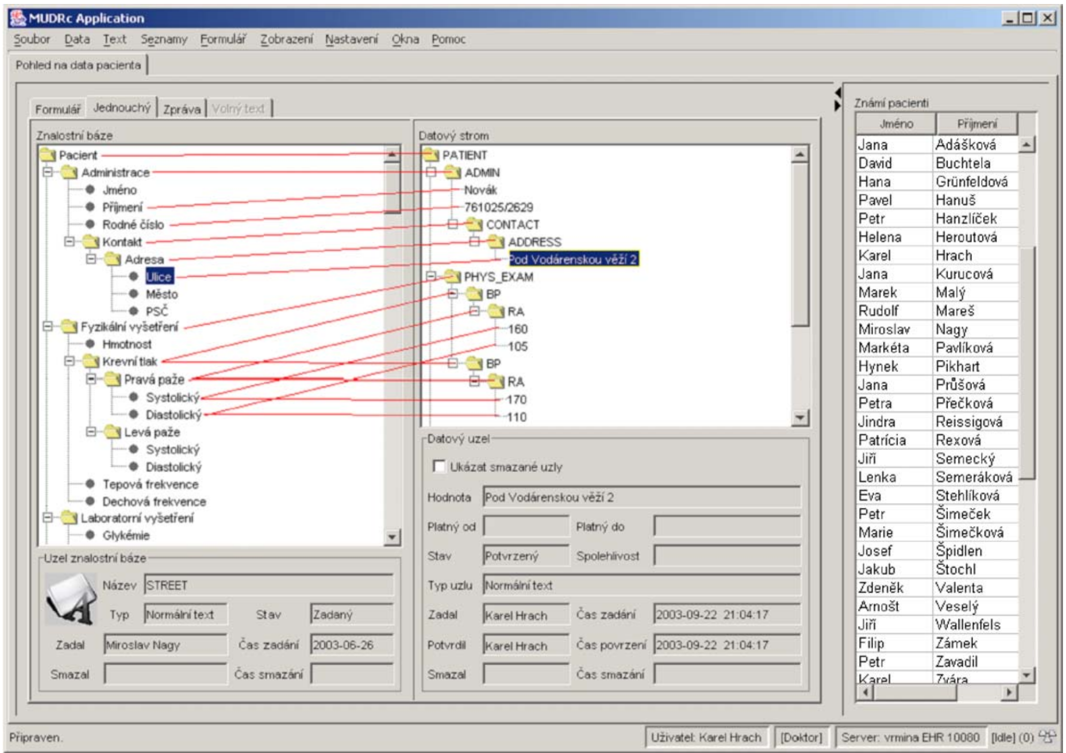

Fig. 2 MUDR information model viewed in the MUDRc user interface.

communication between the MUDR DB Server and the MUDR Application Layer Service uses the MUDR DB Connection Module, which enables implementing the data layer on various database platforms. The application layer chooses the right module for the database currently used. Using this scheme, the application layer communicates transparently with the data layer using always the identical interface.

The main difference lies in the application layer interface, used to provide the functionality of the application layer. The application layer integrates communication modules to communicate with various client types. The main communication module is called MUDR WS. This module provides objects using the MUDR .NET Remoting API (MUDRNRAPI). Using .NET Remoting, the remote call of methods of the shared objects is possible. This functionality is utilized by another application layer component called MUDR Web Service. This web service makes the MUDR Web Service Application Interface (MUDRWSAPI) accessible to common clients. A typical general practitioner (GP) uses a PC on his table to work with the MUDR EHR remotely. The communication is enabled by the HTTP Server installed on the application server. The commands and parameters are encoded using the SOAP standard.

For the possible usage of MUDR thin clients in the form of HTML and WAP browsers we use a MUDR WS Proxy Service. This service is implemented as a Common Gateway Interface (CGI) program; it provides classical HTML or WAP pages to thin clients and appears in the role of a classical MUDR Web Service client at the other side.

3.3. CORBA version

The described Web Service version is not the only way we have extended the classical MUDR architecture. In the pilot version we have made more implementations and we have tested more middleware technologies, mainly to be able to compare them. An example of another implementation is a CORBA implementation where the EHR client communicates with the application layer using the CORBA technology. In this version, CORBA is also used for communication purposes within the application layer, e.g. the communication with the decision support module, which interprets medical guidelines formalized with the GLIF Model [8]. One reason for using CORBA was for example its seamless integration to many programming languages, which provides better interoperability among products from different suppliers.

4. Results

4.1. Pilot implementation

To test our ideas of a flexible information storage model and the functionality of the presented EHR architecture, we have implemented a pilot EHR application. Our implementation consists of many different modules.

The Oracle 9*i* database is used as the MUDR database layer platform. The Oracle database enables using features like nested tables or object references to implement the *MUDR KB* and *MUDR DF* structures as described above. Simultaneously, we are proposing a way of implementation of these structures using a classical relational database.

To implement the application layer we have chosen the Microsoft Windows platform. The application layer runs as a win32 service developed using the C++ language with the Microsoft XML Parser. The communication is established using the Apache web server together with small proper CGI scripts. At the application layer we have connected two formalized medical guideline libraries in the form of win32 DLL libraries. One implements the 1999 WHO/ISH Guidelines for the management of hypertension [9] and the other one implements support for genetic data evaluation [10].

Two thick clients were implemented [11] as the user interfaces. We have also developed two CGI scripts to enable the usage of thin clients. One enables working with the MUDR EHR using

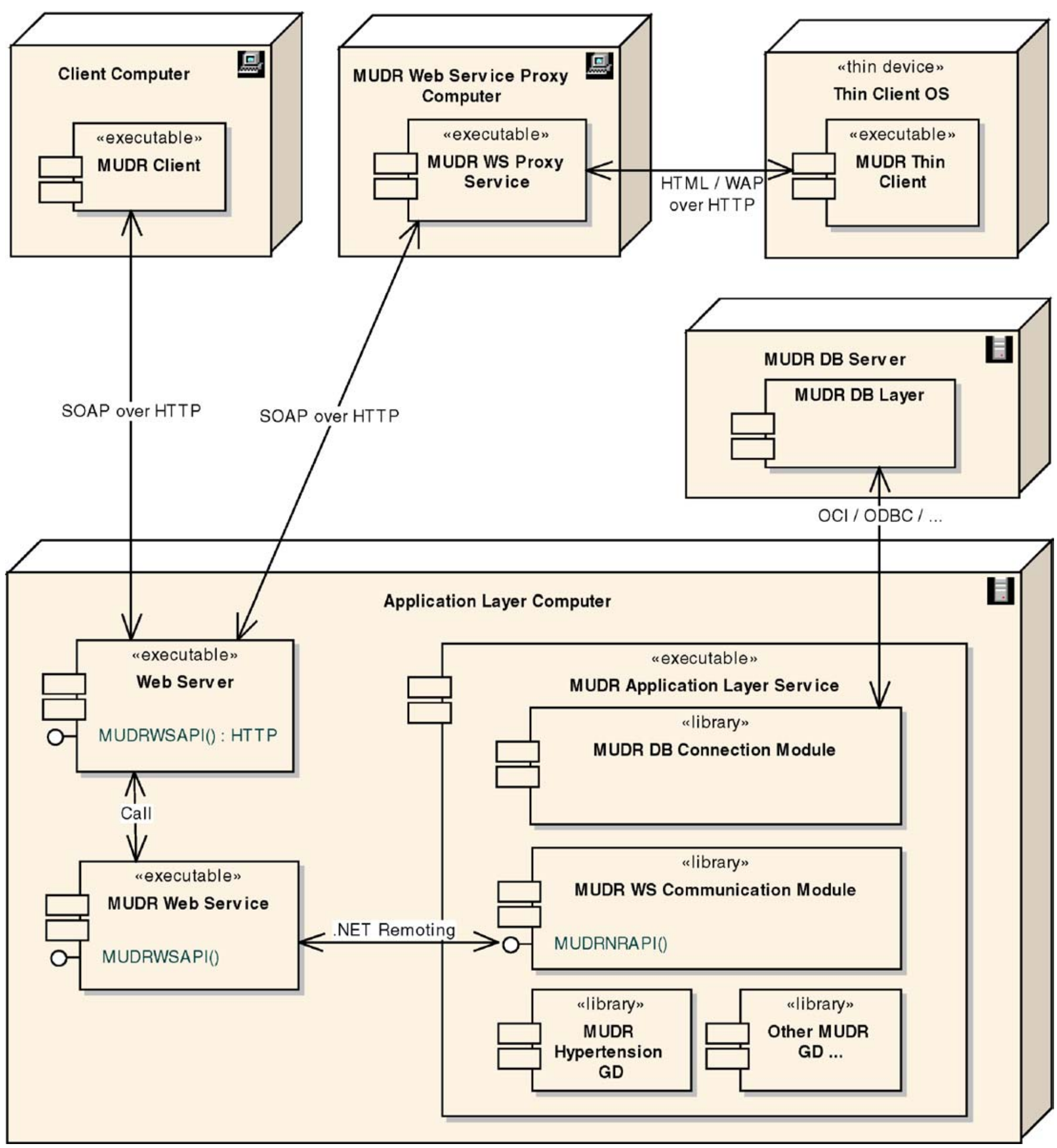

Fig. 3 MUDR^{II} extended architecture.

the Internet Explorer browser and the other one emulates the MUDR EHR client on the Nokia 9110i browser [12].

4.2. Security issues

For a long time, privacy as a social, legal, organisational and political issue has been a concern of social scientists, philosophers and lawyers. To protect personal data and thus to maintain the privacy, IT security mechanisms are needed as the technical side of privacy. According to those issues, there are many standards and recommendations concerned with the security management [13] and security evaluation criteria [14]. We are aware that the MUDR EHR is (as each information system) exposed to threats influencing its behaviour and functionality, which carries a great risk. To minimize the probability of a successful attack paralysing the EHR system, a model summarizing possible threats was built. To safeguard the MUDR EHR we have had to deal with communication as well as with the application security [15]. To ensure the communication security we use secure connections based on SSL (e.g. HTTPS). As for the application security, we mainly focused on two particular services: authorisation and access control, both depending on granted roles. Each operation begins with the identification and authentication of the user. The user can typically play more roles and it depends on the user and his granted role, which parts of a particular health record are accessible and how they are accessible (e.g. the same physician acting one time as an ordinary internist and another time as a member of an emergency staff can access different data). In MUDR EHR, the access control rules are partially encoded in the knowledge base. Each request coming from the client layer must have a user signature attached to it. Digital signing is also used to authenticate the origin of the health record and to verify its integrity. Currently, we are upgrading this part within a new project in cooperation with clinical information systems vendors. To provide the cryptographic functionality, we use the PKCS standard [16] for digital signing and other purposes. At the present time, we are using the iKey 3000 cryptographic device supporting both the PKCS (#11, #12 and #15) as well as the MS CAPI standard.

4.3. Ensuring interoperability

EHR MUDR was mainly inspired by CEN TC 251 standards. According to the ENV 13606 pre-standard, we have created a mapping between the ''Extended Architecture: Folder, Composition, Headed Section, Cluster, Data Item'' and our Knowledge Base modelled for the cardiology domain. The European standards can support the interoperability among various electronic health records, but, unfortunately, they cannot ensure it alone. To ensure the interoperability at least within the Czech Republic, there is a common widely used national standard called DASTA [17]. However, this standard was mainly developed to transmit laboratory results and they are nearly the only structured items that it contains. This and the fact that the applicability of DASTA is restricted to the Czech Republic only lead us to the decision not to use it. Currently, we have started a co-operative project together with clinical information systems vendors and within it we are developing a new system for continual shared health care, which is based on the presented approach. Within this project, we are focusing on the HL7 version 3. Nowadays, HL7 RIM covers any information in the healthcare domain in a generic and comprehensive form and, in our opinion, this standard will soon become the only way to create a real shared health record. However, many fundamental changes accompanying the new HL7 version 3 (e.g. the change of HL7 RIM from an entity centred to an act centred view) contributed to the fact that the integration of HL7 version 3 is quite time-consuming.

4.4. Achievements

We have tested our implementation with physicians in the ambulance of the EuroMISE Centre where service is provided by cardiologists from two cooperating Czech hospitals. We have confirmed that our information storage model is really flexible; it allows a dynamical change of the knowledge base and of the set of collected attributes. We have also verified the functionality of MUDR/MUDR^{II} architecture, which enables an easy decomposition of the system modules, an easy way of data sharing and the remote data access.

But we have found out that the bottleneck of these systems lies in the user interface. It is a very difficult problem to design an EHR user interface, which would be comfortable enough [18]. In our case, the problem is more difficult because of the dynamically changing knowledge base content. To ensure a large use of an EHR system we should provide a user interface, which would be easy and fast to control.

Physicians and GPs in the Czech Republic are used to write their reports in the form of free texts. The structured way of storing information is very important for the additional information processing using for example data mining or knowledge discovery techniques. Unfortunately, not all information can be structured, so our system enables storing both structured and unstructured information.

5. Discussion

Nowadays, most hospital and clinical information systems vendors provide systems with a small possibility of structuring health record, very often only the administrative information and laboratory examinations are well structured but it is important to structure also other parts of the health record. We provide a flexible solution but it cannot be a stand-alone EHR application. It is unacceptable to force the users to enter the information twice, first into the EHR and then into the information system. It is necessary to implement the EHR application as a part of a complex information system, which would bring benefit to the users, save their time and offer them additional information using data mining and knowledge discovery techniques. A detailed financial evaluation is out of the scope of this paper, but it is also obvious that a reduction of costs comes with the improvement of efficiency. Unnecessary examinations are often performed because of the data unavailability. With our approach, there is a potential of costs reduction, because the health record can easily be shared and the data included in the health record can always be structured according to any needs.

6. Conclusion

Our interest is to increase the quality of EHR systems, to simplify data sharing and data migration among various EHR systems and to help in overcoming the classical free-text based health record. This is the way, which would increase the quality of healthcare, which brings benefit for the patient first of all. To do this, we have suggested common general principles of EHR system architecture and a model of flexible information storage without the implementation details. This enables to use various ways of implementation.

We also present an EHR solution MUDR^{II}. It is an open universal solution, which can be applied as the EHR kernel of a HIS. We have not implemented special modules needed for a large commercial use in a particular environment, e.g. modules for the communication with health insurance companies, laboratories, pharmacies or various modalities and public administration offices. There are many standards used in different countries, so it is impossible to give a worldwide universal solution.

To realize this approach in a daily practice for health professionals we negotiated about the implementation of our EHR kernel in hospital and clinical information systems of two big Czech companies. This negotiation resulted in a co-operative project *Information Technology for the Development of Continual Shared Health Care*. Within that project we are developing a new system for continual shared health care, which should bring advantages for the health care in the Czech Republic.

Acknowledgments

The research is partially supported by the projects 1ET200300413 and AV0Z10300504 of the Academy of Sciences of the Czech Republic.

References

[1] J. Rauch. Mining for Statistical Association Rules. In: J. Fong, M.K. Ng (Eds.), The Fifth Pacific/Asia Conference on Knowledge Discovery and Data Mining, University of Hong Kong, 2001, pp. 149—158.
[2] A. Riha, V. Svatek, P. Nemec, J. Zvarova, Medical guideline as prior knowledge in electronic healthcare record mining, in: A. Zanasi, C.A. Brebbia, N.F.F.E. Ebecken, P. Melli (Eds.), Data Mining III, WIT Press, 2002, pp. 809—818.
[3] I. Iakovidis, Towards personal health record: current situation, obstacles and trends in implementation of electronic healthcare record in Europe, Int. J. Med. Inform. 52 (Nos. 1—3) (1998) 105—115.
[4] F.H. Pierik, A.M. Ginneken, T. Timmers, H. Stam, R.F. Weber. Restructuring routinely collected patient data: ORCA applied to andrology. In: J.H. Bemmel, A.T. McCray (Eds.), Yearbook of Medical Informatics 98, Health Informatics and the Internet, Schattauer, Stuttgart, 1998, pp. 257—263.
[5] A.M. Ginneken, The computerized patient record: balancing efort and benefit, Int. J. Med. Inform. 65 (2002) 97—119.
[6] P. Hanzlíček, Development of universal electronic health record in cardiology, in: G. Surján, R. Engelbrecht, P. McNair (Eds.), Health Data in the Information Society, IOS Press, Amsterdam, 2002, pp. 356—360.
[7] J. Zvarova, P. Hanzlíček, J. Spidlen, Electronic health record in cardiology: pilot application in the Czech Republic, in: MIST2002 Proceedings, Wiley, Taiwan, 2002, pp. 10—13.
[8] P. Mor, B. Aziz, T. Samson, W. Dongwen, O. Omolola, Z. Quing. Guideline Interchange Format 3.5 Technical Specification. InterMed Collaboratory, 2002, http://smi-web.stanford.edu/projects/intermed-web/guidelines/GLIF_TECH_SPEC.pdf.
[9] WHO—International Society of Hypertension: Hypertension Guidelines, 1999, http://new.euromise.org/mgt/who1999/who1999.html.
[10] J. Spidlen, J. Adaskova, H. Heroutova, J. Zvarova, I. Mazura, in: J. Vignerova, J. Riedlova, P. Blaha (Eds.), Sta-

tistical Processing of Genetic Information in the MUDR Health Record, Anthropology and Society, Charles University, Prague, 2003, p. 190.
[11] P. Hanzlíček, J. Spidlen, H. Heroutova, User interface of MUDR electronic health record, in: Medical Informatics Europe Proceedings CD., 2003.
[12] P. Hanzlíček, J. Spidlen, J. Zvarova, The Internet in Connecting Electronic Health Record Mobile Clients, Technology and Health Care, vol. 10, Num 6, IOS Press, 2002, pp. 502—503.
[13] ISO/IEC 17799, International Standards Organisation: Information Technology—Code of Practice for Information Security Management, 2000.
[14] ISO/IEC 15408, International Standards Organisation: Information Technology—Security Techniques—Evaluation Criteria for IT Security, parts 1—3, 1999.
[15] B. Blobel, Analysis Design and Implementation of Secure and Interoperable Distributed Health Information Systems, IOS Press, Amsterdam, 2002, ISBN: 1-58603-277-1.
[16] RSA Security Inc., RSA Laboratories: PKCS #11: Cryptographic Token Interface Standard, 2004 ftp://ftp.rsasecurity.com/pub/pkcs/pkcs-11/v2-20/pkcs-11v2-20.pdf.
[17] J. Lipka, Z. Mukensnabl, F. Horacek, V. Bures, Nowadays Communication Standard of the Czech Healthcare: DASTA. In: J. Zvarova, P. Preckova (Eds.), Information Technology in Health Care, Prague, EuroMISE, 2004, pp. 84—92 (in Czech).
[18] A.M. Ginneken, M.J. Verkoien, in: V.L. Patel, R. Rogers, R. Haux (Eds.), A Multi-disciplinary Approach to a User Interface for Structured Data Entry, IOS Press, 2001, pp. 693—697, MEDINFO 2001.