2007

Federated Data Warehouse Approach to Support the National and International Interoperability of Healthcare Information Systems

N. Stolba
stolba@wit.tuwien.ac.at

A. Tjoa
amin@ifs.tuwien.ac.at

T. Mueck
thomas.mueck@sva.sozvers.at

M. Banek
marko.banek@fer.hr

Follow this and additional works at: http://aisel.aisnet.org/ecis2007

Recommended Citation
http://aisel.aisnet.org/ecis2007/121

This material is brought to you by the European Conference on Information Systems (ECIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ECIS 2007 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
FEDERATED DATA WAREHOUSE APPROACH TO SUPPORT
THE NATIONAL AND INTERNATIONAL INTEROPERABILITY
OF HEALTHCARE INFORMATION SYSTEMS

Stolba, Nevena, Women’s Postgraduate College for Internet Technologies (WIT), Institute of
Software Technology and Interactive Systems, Vienna University of Technology, Austria,
stolba@wit.tuwien.ac.at

Tjoa, A Min, Institute of Software Technology and Interactive Systems, Vienna University of
Technology, Austria, amin@ifs.tuwien.ac.at

Mueck, Thomas, Sozialversicherungsanstalt der gewerblichen Wirtschaft, Wiedner
Hauptstrasse 84-86, 1051 Vienna, Austria, thomas.mueck@sva.sozvers.at

Banek, Marko, Department of Fundamentals of Electrical Engineering and Measurements,
Faculty of Electrical Engineering and Computing (FER), University of Zagreb, Croatia,
marko.banek@fer.hr

Abstract

The interoperability of medical information systems is gaining more attention as the need for the
collaboration in healthcare domain increases. Consolidation of distributed patient data and building a
sustainable e-health system is the major research challenge. In order to overcome institutional
barriers and competences for changes across sectors, standardized data structures and corresponding
underlying infrastructure are needed. A concisely defined information network for the exchange of
sensitive patient’s data is a significant step towards enabling national and international cooperation.
In this paper, we discuss the challenges and the advantages of the healthcare information integration
into a federated data warehouse (DWH). We present technical and organizational aspects for creation
of Austrian healthcare information network, and describe a scenario-based federation of Austrian
health-insurance DWHs in compliance with the national governing framework for electronic exchange
of patient-related data. This example will illustrate the benefits of standardized, identifiable and
secure consolidation of the local DWHs and promote innovation in the public sector.

Keywords: federated data warehouse, interoperability of medical information systems, HL7,
ontological integration

1 This research has been partly funded by the Austrian Federal Ministry for Education, Science and Culture, and the
# INTRODUCTION

During the last few years, healthcare organizations are facing a massive problem of information processing, caused by increasing amount and complexity of medical data. Information gathering, analyzing and presenting is not possible without comprehensive IT support. Public sector must prepare for the demography challenge in health and welfare, so research based innovations are needed for a major shift in efficiency and quality. In order to improve the quality of patient care, to reduce the treatment costs and to create a wide base for medical research, healthcare organizations are willing to share their data assets, under the strong security measures. Decision support systems are deployed to support decision making process of care givers and business management, and so to meet rising expectations of quality and quantity of public service production and deliverance.

Both commercial institutions and governmental bodies are interested in building effective and sustainable healthcare system. Increasing patient expectations and rising budgetary pressure are forcing governmental health authorities to rapidly adopt research based innovations into healthcare system. These innovations are needed for a major shift in efficiency and quality of communication between single health system participants. Information aggregation and processing of presently available paper based patient health record has major limitations. Consolidation of diverse health IS can meet the pressure governmental authorities are coping with. Since proper management of public health on both national and international level can be committed only on the basis of comprehensive clinical data, building of exhaustive, standardized electronic patient lifetime records is a necessary step. More comparable health data can help national health authorities to collaborate, for example to gain productivity by supporting reimbursement procedures or to tackle communicable diseases.

This paper is a result of wider research into eHealth in Austria including national policy and local implementation settings.

The goal of our paper is to point out the advantages of virtual integration of heterogeneous medical information systems using federated approach abandoning the building of huge physical data repositories.

The contribution of our work is to present a secure federation model, considering widely adopted international standards for the exchange of healthcare data. We show the benefits of such an approach in the area of usability, security, cost-reduction and for supporting of e-government projects in the social sector.

The paper is structured as follows: in the section 2, we discuss the need for the integration of heterogeneous medical information systems. The legal regulations for the creation of a healthcare information network in Austria (MAGDA-LENA) are described in section 3. In section 4, we show a scenario, representing MAGDA-LENA guided electronic data exchange between the Austrian social insurance DWHs on one side and the DWH federation on the other side. An outline of the related work is given in section 5. Conclusions are drawn in section 6.

## 2 INTEGRATION OF HETEROGENEOUS MEDICAL INFORMATION SYSTEMS

In the past few years, the awareness of the necessity for the integrated healthcare system has risen in both the provider and consumer sector. Sharing of medical data enables hospitals and single work-site physicians to provide more efficient care to the patients, which results in cost reduction for social insurance institutions. Additionally, patients are in many cases liberated to consume healthcare from different providers.
The Commission of the European Communities (2004) announced in their Action plan for a European e-Health Area that achieving a seamless exchange of health information across Europe requires common structures and ontologies of the information transferred between health information systems. Member states, as well as the other countries concerned with raising the quality of healthcare thorough integration of medical information systems, support initiatives aiming the building of electronic healthcare networks.

Austrian social insurance service providers are well aware of the benefits they can gain from an integrated healthcare information system. A huge data base, resulting from this integration, would represent a precious knowledge source. A decision support system and the deployment of evidence-based medicine would enable social insurance institutions to locate the efficient ways for enhancing the quality of healthcare provided to the insurants with the possibilities of reducing the treatment costs at the same time.

Stolba, Banek and Tjoa (2006) were examining the issues of data consolidation under the security restrictions for the confidential patient data transmission. Most of the health insurance organisations have been developing their local information systems independently (using different, mutually incompatible data formats) which let the integration of the heterogeneous warehouses to be a major challenge. Since business decisions must be made over large, statistically significant data patterns, social insurance institutions need to join their data into a single DWH, which becomes the foundation of the knowledge discovery system. Due to the high confidentiality of healthcare data and the privacy policy of participating organisations, the proposed warehouse will not be created physically but as a federated system. Unlike integration into a centralized, physical decision support system, where sensitive data is duplicated, federated DWH virtually ties together a network of social insurance providers so that fundamental operational, analytical and research activities can be efficiently realized.

The participants of the designated collaboration project described in this paper are different health insurance providers. The project is guided by the regulations provided by the commission designated by the Austrian ministry of health. A short overview of the major recommendations from this framework is given in the next section.

Other organisations in the healthcare domain are expected to join the DWH federation and contribute to the development of the integrated healthcare network, which would be the foundation for e-prescription, drug cost control, patient’s electronic lifetime record and further e-Health projects for enhancing the quality of medical care.

3 TECHNICAL AND ORGANIZATIONAL ASPECTS FOR CREATION OF AUSTRIAN HEALTHCARE INFORMATION NETWORK

The Austrian healthcare system is highly fragmented. The vast majority of patients receive their healthcare from multiple healthcare providers. Hospitals, physicians, recovery centres, laboratories, pharmacies and health insurance institutions have each their own, isolated patient records and therefore fragmental knowledge about the patient’s health condition. Absence of integrated healthcare bears the risks of medical treatment errors, duplicate examinations, lack of coordination and increased therapy costs. In order to make information exchange between healthcare service providers possible, the reformation of the information system is a necessary step.

In 1995, the Austrian Ministry of Health and Woman Affairs created the STRING commission to advise the minister on all strategic healthcare issues. Three years later, this commission developed the MAGDA-LENA framework as the governing framework for electronic exchange of patient-related data in Austria. Unlike some other frameworks, which are formulated as legally binding rules (for example HIPAA (2006) in USA, MAGDA-LENA has solely a recommendatory status. MAGDA-LENA (STRING-Kommission, 2000) outlines the technical and organizational aspects for the creation
of an Austrian healthcare information network, which facilitates the exchange of patient-related data between healthcare providers and social security institutions.

The MAGDA-LENA framework contains the following four sections:

1. Message contents, models, standards
2. Identification variables
3. Data privacy and security
4. Network providers and nodes

Duftschmied et al. (2003) give a concise overview of the main recommendations.

3.1 Message Contents, Models and Standards

The standard formats for the exchange of health information are specified. Messages exchanged between communicating parties should not be defined by parties themselves, but are based on internationally adopted standard message formats. The development of a healthcare information model, as a basis for standard messages, remains as a subject of international efforts. Since for the time being such a system is still in a planning stadium, MAGDA-LENA prescribes the use of existing international and national standards related to medical informatics, which satisfy the particular requirements of a communication process. In case that such standards do not exist for a given area of application, messages are developed along the procedural model defined in MAGDA-LENA, to optimize their uniformity.

3.2 Identification Variables

In order to eliminate any kind of data abuse during the transmission process, the communicating parties as well as the transmitted data must be uniquely identified. In the following, the authentication procedure for the communicating parties is briefly described. MAGDA-LENA requires communication parties to be identifiable via registered directories within the organization they are assigned to. Furthermore, the roles of the communication parties as well as the identification of transmitted data is approached here.

3.3 Data Privacy and Security

Communicating parties are required to implement certain security measures in order to guard data privacy and security in their internal domain as well as during electronic communication with others. For this purpose, a list of obligatory implementation specifications (e.g. content encryption algorithms and protocols, requirements for electronic signature, security tokens and password systems) and guidelines for their proper implementation is provided.

3.4 Network Providers and Nodes

Minimum standards that have to be obeyed by network providers participating in the Austrian healthcare network are defined. They include guidelines regarding contracts with clients and guidelines concerning network interfaces.
Our approach is build upon our previous work about Austrian DWH federation in the area of evidence-based medicine (Stolba et al., 2006), where we handled the relevant security issues for federated DWHs introducing healthcare standards (HL7, openEHR, ENV 13606) and in particular HL7 RIM Foundation Classes. The goal of our investigations was focused on the technical issues of gathering local DWHs into a federation. In our current work, we intend to adopt the proposed federated DWH model corresponding to the MAGDA-LENA recommendations, which means to broaden our focus according to the necessary limitations of legal issues.

The nature of our problem is founded by the existence of health insurance DWHs participating at the federation storage of sensitive data about insurants, diagnoses, therapies and drugs prescriptions, which are covered by the various insurance institutions. In order to deliver best medical treatment to the insurants by applying the most effective therapies as well as to reduce administrative and treatment costs, business management intends to participate in DWH federation under strict security measures. Although the advantages of healthcare information integration are clear, divergence still exists about how such integration should be achieved.

Figure 1: Multidimensional Clinical LDM

4.1 Building a Federated DWH According to the MAGDA-LENA Recommendations

Austrian health insurance institutions operate heterogeneous decision support systems, which are functionally integrated into a single unit. Each local DWH is providing the federation with the designated data, according to the MAGDA-LENA recommendations. In the following sections, we...
will handle these recommendations in more detail. Although this scenario illustrates the DWH federation for social insurance institutions, it can be used as much general model for any other institutions in healthcare sector. Figure 1 presents an extract from the multidimensional clinical logical data model, as it is stored in federated DWH. This model contains Facts, Time, Diagnose, Social Insurance, Location, Drug and Patient Dimension and is created according to ADAPT™ (ADAPT, 2006) notation.

4.2 HL 7 RIM for Standardized Data Transmission

In our approach each local DWH operates independently from the others and from the federation. Generally, each DWH stores its data in its own system specific format; obviously a large portion of the different data contents are semantically closely related. Data stored in the social insurance DWH refer to the insurant, who is uniquely identified by his (her) social security number. Information about insurant’s status (self-insured, co-insured), scope of insurance as well as various descriptive fields are stored here. Additionally, these DWHs hold data about consumed services (treatments, therapies, drugs) covered by the social insurance. For the purposes of compensation calculation for the insurant, the whole range of accounting data is loaded into the DWH. Data about drug dosage, as well as charging data for pharmacy contribution calculation is another important part of it.

According to the MAGDA-LENA recommendation about message content, models and standards, underlying DWHs have to use internationally adopted standard message formats for transmission of their health information to the federation. Messaging standards are particularly important as they define how information is packaged and communicated from one party to another. Such standards set the structure and data types required for seamless integration from one system to another. Currently, a lot of effort is being invested into creation of standardized message formats for healthcare sector (e.g. HL7, ENV 13606, openEHR). In our approach we choose the HL7 V3 communication standard since it is general enough to allow the necessary level of flexibility.

![Figure 2: HL7 RIM class diagram for treatment and charging process](image-url)

HL7 (Health Level Seven, 2006) is generally accepted as an ANSI-accredited standard by the Standard Developing Organization (SDO) for healthcare data. HL7 specifies the communication contents and messaging formats for the communication between healthcare institutions on the application level. It is widely independent of underlying hardware and network infrastructure as well as database and applications used. All HL7 Version 3 models are based on the HL7 Reference Information Model (HL7 RIM, 2006). RIM introduces six backbone foundation classes: Act, Entity,
Role, Participation, RoleRelationship and ActRelationship. It generically describes processes in healthcare environment.

Figure 2 (Stolba et al., 2007) shows an example representing HL7 RIM model for the treatment charging process taking place in a social insurance company. On one side, we see two person entities in the roles of clinician and patient. They participate in the activity treatment as, respectively, performer and subject. On the other side is the entity person, in the role of a payer, subject of the billing activity. Another person entity participates the billing activity as a performer in the role of an accountant. This role is scoped by the entity controlling. These two activities (treatment and billing) are related to each other, which is represented through the act relationship treatment charging.

Using HL7 RIM message formatting, as shown in Figure 2, for transmission of healthcare data between local DWHs and the federation fulfils the first requirement of the MAGDA-LENA framework, as described in section 3.1.

4.3 Ensuring the Uniqueness of Instance-Identifiers

According to the MAGDA-LENA framework, communicating parties as well as the transmitted data must be uniquely identified, in order to prevent any kind of data abuse during transmission process. Each participant of the healthcare network must apply for access at the designated authority. Additionally, a participant is assigned a predefined role. He (she) is then accountable for addressing the messages to authorized receivers only. Whenever patient’s data is transmitted, it needs to be identified by a unique patient-id. Finally, transmitted documents have to bear a unique document identifier.

Table 1 represents the items which need to be identified and corresponding identification method. In our scenario, communication parties are the social insurance institutions participating the federation. For the time being, their identification method is still undefined, but we will assume that there is a specific authority responsible for assignment of the communicating party’s ID. The role of the communicating party is specified by the Ministry of Health. At present, social insurance number (SV-Nr) is the only suitable candidate for the purpose of the unique patient’s identifier. Unique identification of transmitted documents still needs to be specified. We will explain this issue in details in sequence.

<table>
<thead>
<tr>
<th>Object of identification</th>
<th>Identification method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physician</td>
<td>Contracting party number provided by Hauptverband or physician number provided by the Medical Association (Ärztekammer)</td>
</tr>
<tr>
<td>Pharmacy</td>
<td>Pharmacy operating number</td>
</tr>
<tr>
<td>Hospital</td>
<td>Hospital number</td>
</tr>
<tr>
<td>Social Insurance Institution</td>
<td>Unknown</td>
</tr>
<tr>
<td>Role of Communicating Party</td>
<td>Assigned by Ministry of health</td>
</tr>
<tr>
<td>Patient</td>
<td>Social insurance number</td>
</tr>
<tr>
<td>Document</td>
<td>Document ID</td>
</tr>
</tbody>
</table>

Table 1: Identification list

The structure and the semantic of clinical documents for the purposes of their exchange is specified by the HL7 Clinical Document Architecture (CDA), which is an XML-based document markup standard. A CDA document consists of a body, containing medical data and a header, containing data about patient, clinician and hospital. In our scenario, we follow the HL7 V3 RIM approach of identification and classification as it has been described by Heitmann (2006). Heitmann illustrates these procedures through an example, which presents the Observation class. (Figure 3 represents the XML statement of the observation.) Here, a code is used to state the art of the examination and value is used for storing
the actual result. **Effective time** contains the exact time of the observation and the observation is uniquely identified through *id*. All attributes have preassigned and well defined data types.

In this example, the specific observation (e.g. systolic blood pressure) is characterized as an instance of an observation (*id* is an instance identifier). For these purposes, the application system must create a globally unique instance-identification. This can be assured by a HL7 procedure. Element *code* bears the classification of the observation. HL7 uses common medicine coding system LOINC. The actual code is stored in the XML-attribute *code* and *codeSystem* contains the unique identifier of the coding system.

When electronic data exchange takes place, the concept of HL7 requires a world-wide unique identification of all medical observations. HL7 provides a procedure to ensure that, but the hosting DWHs must be upgraded in order to support the production of these references.

### 4.4 Securing Patient Data Confidentiality during Message Exchange

The following security precautions are recommended by MAGDA-LENA:
- Password identification for the user of the federated DWH (authentication)
- Any data modification must bear a digital signature
- Tracking of data manipulation through log files
- Confidential health data must be transmitted in encrypted form
- Transmission confirmation upon receiving of confidential data
- A role-based access model has to be implemented

The last section of MAGDA-LENA recommendations (as described in section 3.4), concerns network providers and nodes. Since this subject is not DWH specific and for the reasons of paper length restriction, we will not handle it in this article.

### 4.5 Overall Federation Model

Figure 4 represents the model of the DWH federation process according to MAGDA-LENA recommendations. Underlying social insurance DWHs are providing the federation with requested data in their native XML format. Using XSLT transformation, fetched data is transformed from native into HL7 V3 XML schema. Further, identification and coding procedures round off the desired HL7 V3 XML structure. For security reasons, confidential patient data is encrypted for the transmission.

The acceptance of the transmitted data is acknowledged by a confirmation message. The essential part of the integration of logical schemas of the underlying DWHs is the ontological integration layer, as shown in the model depicted in Figure 4. Our model includes wrappers and mediator, which are two main architectural components of a mediated query system.

Classic wrapper/mediator architecture is based on a Global Virtual Schema (Global Virtual View – GVV) and a set of data sources (Beneventano, 2006). The data sources provide the applications with real data while the GVV represents an integrated, virtual view of local data sources. There are two basic approaches for specifying mappings between GVV and underlying sources described in the literature: Local-As-View (Halevy, 2001) and Global-As-View (Ullman, 1997).
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 unified form (Beneventano, 2003), (Beneventano, 2006). 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 (Beneventano, 2003), (Beneventano, 2006) 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.

---

Here is 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:

1. DWH-Application invokes the mediator.

*Query unfolding:*

---

**Figure 4: DWH federation according to MAGDA-LENA recommendations**
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 result (consolidated HL7 V3 message) is subsequently converted into XML or directly into SQL.

Only authorized users may access the federated DWH. Ministry of health is responsible for access and role assignment to all participants of this healthcare network. According to the role-based access model, each user may see only those portions of data necessary for performing of the tasks of his (her) role. Authentication procedures are put in place and all data manipulation is logged for later analysis. Users are querying federated DWH through data mining and OLAP-Tools, by applying predefined reports or performing ad-hoc analysis.

5 RELATED WORK

Countries concerned with the enhancing of healthcare quality through the cooperation of healthcare organizations are starting initiatives aiming at building of electronic healthcare networks.

Sjunet (2006) is the Swedish Health Care Network incorporating an infrastructure for communication between hospitals, primary care centres and home care. It is hosting a wide range of governmental services and healthcare service providers and vendors. Sjunet is used for telemedicine, for secure administrative communication and distribution of confidential patient data and for educational and training purposes. It is a cooperative network as well as a technical communicative platform for Swedish healthcare.

In Belgium, the Crossroad Bank for Social Security project (CBSS, 2006), an electronic network for the collaborations between the social security institutions, has been created. The CBSS has been created in order to improve the service delivery to the socially insured people and the companies and its mission is to be the motor of e-Government in the social sector. The focus of the project is to create a nation wide electronic network for improving predefined administrative tasks for the insured people and companies.

Further initiatives to which we refer but will not discuss in more detail are MedCom in Denmark (2006), HYGEINANET in Greece (2006), NPfIT in the UK (2006), ONCHIT in United States (ONCH, 2006), Canada Health Infoway in Canada (CaHI, 2006) and HealthConnect in Australia (HeCo, 2006).

Dorda et al. (2005) present the current situation in healthcare Telematics in Austria and they discuss similar activities in other countries. They state the necessity of EHR standardization, in order to avoid an isolated EHR solution in Austria. The focus of their work is mainly on the standardization and
confidentiality of EHR contents, but they do not discuss the issue of technical standardization (federated DWHs). In another article (Gall et al., 2004), the authors are investigating the approaches to standardize EHR architecture in more detail, with the regard to the specification of temporal components.

GRID computing paradigm as a possible solution to the problem of integration of medical information has been proposed by Bilykh et al. (2003). The authors have developed a prototype for the Health Information Grid, a middleware technology that supports mediation of medical information among rapidly evolving, heterogeneous medical information sources. Health-InfoGRID promotes the separation of three main concerns: inter-organizational workflow, secure Grid service configuration and federation of legacy systems internal to organizations. This paper discusses issues and requirements arising from constructing a global health information network in Canada, with a special concern about security and privacy issues.

6 CONCLUSION

In order to support the interoperability of e-Health organizations, to offer easier access to healthcare information and to disseminate best practices, national institutions concerned with improving the quality of healthcare are supporting the integration of healthcare information systems. In our work, we described the benefits of integrated healthcare information systems and we addressed the major organizational and technical challenges for its realization. The proposed federated DWH model is developed in compliance with legal recommendations, as they are in force in Austria. The proposed approach capacitates the integration of wide range of heterogeneous data sources and also supports decision making process through its OLAP and data-mining facilities. It builds a suitable solution for integration of high volumes of complex healthcare data. Since the data of participating healthcare organizations is solely virtually tied in a network and not physically duplicated into a central data store, the introduced solution incorporates possible security threats for the highly sensitive patient data. This approach is a significant step towards decentralization of security assurance and towards strengthening of interoperability in healthcare sector, which results in a better service for the patient, the medical personnel and the administrative staff.

ACKNOWLEDGMENT

- 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 gratefully acknowledge the contributions and support of SVA (Sozialversicherungsanstalt der gewerblichen Wirtschaft, Vienna, Austria) team members, in particular Mr. Martin Strasser, Mr. Raimund Lustig and Mr. Johann Redl for supplying us with heterogeneous data models and for their cooperativeness at advising us with their health insurance expertise.
- We are very indebted for the project support of the Competence Center “Secure Business Austria” (funded by the Austrian Federal Ministry of Economics and Labour)

References

Bilykh I., Bychkov Y., Dahlem D., Jahnke J.H., McCallum G., Obry C., Onabajo A. and Kuziemsky C. (2003). Can GRID services provide answers to the challenges of national health information


