Towards shared patient records: An architecture for using routine data for nationwide research

Petra Knaupa, Sebastian Gardeb, Angela Merzweilerc, Norbert Grafd, Freimut Schillinge, Ralf Weberg, Reinhold Hauxg

a University of Heidelberg, Department of Medical Informatics, Im Neuenheimer Feld 400, D-69120 Heidelberg, Germany
b Central Queensland University, Faculty of Informatics and Communication, Rockhampton, Qld. 4702, Australia
c SRH Learnlife AG, Lempenseite 55, D-69168 Wiesloch, Germany
d University Hospital of the Saarland, Pediatric Hematology and Oncology, Homburg, Germany
e Olghahospital Stuttgart, Pediatrics 5 (Oncology, Hematology, and Immunology), Bismarckstr. 8, D-70176 Stuttgart, Germany
f University of Heidelberg, Coordination Center for Clinical Trials, Im Neuenheimer Feld 221, D-69120 Heidelberg, Germany
g Technical University of Braunschweig, Institute for Medical Informatics, Mühlenpfordtstraße 23, D-38106 Braunschweig, Germany

KEYWORDS
Electronic patient record; Architecture; Multiple use of data; Medical informatics

Summary Ubiquitous information is currently one of the most challenging slogans in medical informatics research. An adequate architecture for shared electronic patient records is needed which can use data for multiple purposes and which is extensible for new research questions.

We introduce eardap as architecture for using routine data for nationwide clinical research in a multihospital environment. eardap can be characterized as terminology-based. Main advantage of our approach is the extensibility by new items and new research questions. Once the definition of items for a research question is finished, a consistent, corresponding database can be created without any informatics skills.

Our experiences in pediatric oncology in Germany have shown the applicability of eardap. The functions of our core system were in routine clinical use in several hospitals. We validated the terminology management system (TMS) and the module generation tool with the basic data set of pediatric oncology. The multiple usability depends mainly on the quality of item planning in the TMS. High quality harmonization will lead to a higher amount of multiply used data.

When using eardap, special emphasis is to be placed on interfaces to local hospital information systems and data security issues.
1. Introduction

1.1. Subject and motivation

Ubiquitous information is nowadays one of the most challenging slogans in medical informatics research. Due to the continuing enormous progress in medical research, health care is becoming more and more specialized. Often, a patient has to consult a variety of physicians for the treatment of one medical problem. 'Systematic care in partnership' [1] can efficiently be supported by information and communication technology [2]. A substantial part of this is extensively sharing data and expertise [3]. Sharing patient data means that data items are entered only once and can be used for multiple purposes in various locations. The objective is to avoid multiple documentation, measures and examinations for the sake of the patient and to save costs. Therefore, shared patient records are currently heavily researched in medical informatics [4,5].

We identify four stages of multiple use and shared entry in health care:

1. Document-based: exchange of (electronic) documents without the possibility to further process the data contained in the document.
2. Item-based: access to single items of shared data allowing their processing in another institution.
3. Terminology-based: a standard terminology is used, so that there is agreement on the naming and semantic meaning of items. This is a prerequisite for using the data, for example for comprehensive analyses for research purposes.
4. Extensible: additionally, data for new questions under investigation can be added, for example if legal requirements change or new research questions arise.

The higher the stage, the more efficiently shared care can be supported by information and communication technology. Therefore, an adequate architecture for a shared electronic patient record is needed which supports the use of data for multiple purposes and which is extensible for new questions under investigation.

1.2. Objectives of the paper

The objectives of this paper are:

- to introduce an extensible architecture for using routine data for additional purposes (eardap) like clinical research based on a standard terminology,
- to report on our experiences in establishing eardap in pediatric oncology in Germany,
- to discuss the applicability of eardap for shared patient records.

2. Methods

In this chapter we introduce eardap—the architecture for using routine data for additional purposes. To build up an eardap-based architecture for a particular documentation environment, methods and tools of the fields:

- systematic requirement analysis,
- systematic planning of clinical documentations,
- software engineering, and
- communication interfaces.

have to be applied. Particular methods and tools can be chosen according to circumstances, resources and individual preferences.

2.1. Prerequisites for building up an eardap

Since the main objective of eardap is to achieve multiple use and shared entry, all the purposes for which the data should be used have to be identified and specified for all locations. It has to be carefully distinguished between:

- General functions of an electronic patient record like patient administration or electronic reports. These functions are intended to be used by a variety of locations. General functions should be permanent over time.
- Research questions like questions on therapy optimization or epidemiologic questions. These can be tailor-made for a particular institution and can change over time.

For each purpose, the necessary items to fulfill this purpose, respectively to answer a research question, have to be defined. This process can be supported by a method for systematic planning of clinical documentations (for example [6]).

When specifying items and corresponding answering options, it should be regarded right from the beginning that they will become part of a standard terminology. Items have to be clearly defined and structured. Overlapping items and homonyms should be avoided and it must be adequately dealt with synonyms.

Finally, the detailed requirements for the layout and functionality can be worked out systematically in cooperation with future users.
2.2. Components of eardap

The architecture of eardap consists of the basic components: terminology management system (TMS), core system for data recording and management, research-specific non-autonomous modules for data recording and a module generation tool. We briefly introduce these components of eardap before we describe in more detail how they work together.

2.2.1. Terminology management system (TMS)

The objective of the terminology management system (TMS) is to formally represent all items and corresponding answering options so that a reference terminology is available. Requirements are:

- Management of a concept-oriented shared terminology.
- Derivation of research-specific terminologies by formally representing the items in the context of a particular research question. For this it could become necessary to refine answering options.
- Management of integrity constraints on items.
- Version management of items and their assignment to research questions.
- Adequate user interface for the definition and administration of items, answering options and integrity constraints.

Fig. 1 summarizes the components of the TMS. To build up the content of the TMS is a great challenge. All partners have to agree on the terminology and it has to be guaranteed that it is used exclusively in the way it is represented in the TMS. An authorization is needed to change and maintain the terminology.

2.2.2. Core system

The objective of the core system is to implement a minimum basic data set for routine data and to provide the functionality for data management, recording and analysis according to the user’s needs. Examples for the functions of the core system are given in Fig. 2. Requirements are:

- The core system is an autonomous computer-based application system.
- All items have to be consistent with the reference terminology in the TMS.

Fig. 2 The core system as a component of eardap and its cooperation with the terminology management system. The core system consists of a variety of functions that operate on patient data. The patient data is consistent with the reference terminology specified in the terminology management system.
Functions of the core system have to be linked dynamically to the data.
Among others it should offer functions for:
- user administration,
- communication with the user when integrity constraints are violated,
- data exchange and provision.

2.2.3. Modules
The objective of modules is to extend the data in the core system by data necessary for a specific research question. Normally, this data is not relevant for all patients of an institution but only for patients, which fulfill certain selection criteria. Requirements are:
- Modules are non-autonomous application components for data recording. Besides offering data entry they do not inhere any functions for interaction with the user.
- All items of the module have to be consistent with the standard terminology in the TMS.
- Modules can have its own database for storing the recorded data.

Cooperation with other components:
- A core system can be extended by several modules.
- A module inhere an identification of the research question it answers and of the items it provides.
The core system integrates this information so that functions can operate on data of the core system as well as on data of the module. An example could be the anonymisation of the data to prepare them for research purposes.

The modules and their cooperation with the core system and the terminology management system are illustrated in Fig. 3.

2.2.4. Module generation tool
The objective of the module generation tool is to automatically generate modules on the basis of the reference terminology in the TMS. Requirements are:
- user interface for accessing the terminology in the TMS and structuring the data;
- automatic generation of a relational database for the research-specific module in accordance with the standard terminology;
- interactive design and generation of (electronic) case report forms in accordance to the standard terminology;
- transformation of integrity constraints into check routines;
- generation of databases for the site that will analyze the research data.

These requirements are fulfilled by the tools demonstrated in Fig. 4.

Cooperation with other components:
- The module generation tool operates on data provided by the TMS. It accesses the items, answering options and integrity constraints, which are assigned to a particular research question.
- The module generation tool provides the module with the identification of the research question and the corresponding items.
- New modules are registered in the core system.

2.3. eardap—the resulting architecture
Fig. 5 illustrates how the eardap-components work together: a core system for recording and processing routine data is implemented and established in a medical center where patient data originate. When
Towards shared patient records: An architecture for using routine data for nationwide research

Fig. 4 The module generation tool as a component of eardap and its cooperation with the research-specific modules. Main components of the module generation tool are a database generator, which automatically generates the databases for the research-specific modules as well as for the trial centers and a form generator, which automatically generates the case report forms. The module generation tool can access the terminology management system so that the user can specify the research-specific terminology that is used in the modules.

When a new research question comes up, the responsible research institution has to specify and harmonize the necessary terminology in the TMS. Based on this, the module generation tool accesses the TMS and automatically generates a database for the research center and for the participating clinical sites. Afterwards, electronic case report forms are generated. The module is installed in the medical center and integrated in the core system. The core system offers functionality for data communication so that routine patient data can be transmitted to the relevant research institution. This research institution then analyzes the data to answer the research question. When new research questions come up or modifications are necessary, modules must be exchanged. Of course, this exchange has to be consistent with the standard terminology in the TMS.

Fig. 5 Cooperation of all eardap components and its users (represented here as medical center, terminology management authority and research centers). The data of the core system is transferred to the research center for scientific analysis. The staff of the research center cooperates with the terminology management authority to specify the research-specific terminology in the modules.
2.4. Additional remarks

This workflow and its underlying architecture are independent of particular decisions concerning the employed software development environment, standards and technologies. The core system for example can be realized in accordance to harmonization efforts like openEHR or standards like EHRCA (ISO CEN ENV 13606, http://www.chime.ucl.ac.uk/work-areas/ehrs/EHCR-SupA/overview.htm). Technologies for data entry can be manual recording, speech recognition or automatic transfer from technical devices. Solutions for data privacy and protection have to be integrated: models, methods and tools which allow formal and structured policy definition, policy agreements, role definition, authorization and access control [7] should become part of the eardap components. These considerations are necessary extensions to eardap, which have to be fulfilled considering the latest state of research and legislation.

3. Results

We implemented the architecture described above in the field of pediatric oncology in Germany. Patient care in pediatric oncology in Germany is characterized by a network of about 20 nationwide multi-center clinical trials. Each trial is dedicated to a particular diagnosis and the objective is therapy optimization [8]. Thus, the trial centers release therapy protocols with treatment schemata according to the newest research results. About 90% of patients are treated according to these schemata [9]. Typically, the treatment’s basic component is chemotherapy. For ongoing research the trial centers require data on the results of the treatment of each patient. Therapy in pediatric oncology can last for years, therefore the amount of data, which has to be transmitted to the trial centers is enormous and the communication process is complex. The data is recorded manually in addition to the routine data in the local patient record and sent by conventional mail. The idea was to use routine data gathered in the medical centers of pediatric oncology also for research purposes in trial centers and to transmit routine and research-specific data electronically.

3.1. Requirement analysis

We started in pediatric oncology with a systematic analyses of the user requirements towards documentation purposes [10]. The results were rather heterogeneous because the future users were spread over various hospitals all over Germany. Thus, a consortium of members of the German Society for Pediatric Oncology and Hematology (GPOH) was involved in all decisions. An agreement on basic routine data was already available [11] but had to be adapted according to state-of-the-art requirements for formally represented terminologies [12]. For the requirements of the trial centers special analyses were performed [13]. Additional requirement analysis became necessary for the function of chemotherapy planning [14]. Requirements changed over time and a systematic reengineering became necessary [15]. With respect to data privacy the research network for pediatric oncology has established a central pseudonymisation service, which can be used for our implementation. Additional research resulted in a generic model for data privacy, which is provided for several medical research networks in Germany (http://www.uni-mainz.de/~pommeren/).

Results of the above mentioned analyses are part of the following descriptions of the implemented components.

3.2. Implemented components

We implemented the components with Borland Delphi. They run in all common Microsoft Windows environments (Windows 95 and higher) and are usually shipped with a Borland Interbase database. The core system can be operated in a client–server environment.

3.2.1. Terminology management system

The aims of the terminology management system were to formally represent the items of the basic data set in pediatric oncology and to harmonize trial specific data. We implemented a concept-oriented terminology management system [16] and a user interface for the management of its terminology. Together with a consortium of members of GPOH we reworked the basic data set of pediatric oncology substantially. A terminology management authority was constituted within the GPOH and items resulting from the standardization process were entered in the TMS. We started the harmonization and representation process of trial specific data for particular trials.

3.2.2. DOSPO core system

The aim of our core system is documentation and therapy planning in pediatric oncology (DOSPO). Computer-based therapy planning takes place in adherence to protocol schemata. We implemented a separate tool for acquiring protocol knowledge
Towards shared patient records: An architecture for using routine data for nationwide research 197

and a DOSPO-function for calculating particular therapy plans for a patient [17]. Further DOSPO-functions are:

- Report writing (based on the standard terminology).
- Documentation of diagnosis and procedures for administrative purposes.
- Data exchange with local hospital information systems.
- Flexible, analyses of local data based on the standard terminology.

DOSPO was used in clinical routine by approximately 10 medical centers across Germany and was introduced in about 20 further medical centers during our investigation. In order to be able to exchange data with existing information systems in the various centers, we developed a flexible HL7-interface.

3.2.3. Trial specific modules
The aim of our trial specific modules was to enable data recording of those data which are specific for a particular therapy trial and which were not part of the reworked basic data set for pediatric oncology. Each module represented exactly one trial. The modules contained their own database for data storage, since children in pediatric oncology participate in only one trial at one point of time. Therefore, there was no risk of overlapping items and therefore inconsistency. The case report forms in the modules were linked to case report forms in the core system, a button was placed on each of the respective forms and the user was led to the trial specific data by clicking on the button.

We manually implemented one module for acute leukemia and one for brain tumors and tested the linking mechanisms to the core system.

3.2.4. Module generation tool
The aim of our module generation tool is the terminology-based automatic generation of trial databases and interactive design of electronic case report forms. Therefore, it consists of two components:

- A component for the automatic transformation of items of a concept-oriented TMS into a schema of a relational database. A SQL-DDL-script is generated automatically and can be executed by a SQL-server.
- A component for building terminological consistent and organizational reasonable case report forms. It provides a tree view of all contexts and the associated items as they are specified in the TMS. The user then can create empty forms and, by drag and drop, fill them with the required contexts and items. After finishing the construction of a new form is finished the SQL-functions for the database access are automatically generated.

We validated both components by processing the basic dataset of the GPOH. To test the applicability of the database generation component we generated a database for the basic dataset of the GPOH. It consists of 118 relation schemes with 73 being relation schemes for sets of options.

The form-building component is still under development. Basic concepts like the mapping of the trial specific forms to an integrated database could already been proved as well as the correct interaction between the TMS and the development components of the module generation tool.

4. Discussion

4.1. Architecture
We introduced eardap as architecture for using routine data for nationwide clinical research in a multi-hospital environment. eardap can be characterized as a terminology-based and component-based architecture. In contrast to other component-based approaches [18–20] eardap is not functionality-centered but data-centered: the functionality remains the same when the data changes on which the functions operate. eardap consists of three main components: core system, TMS and module generation tool. In an established eardap environment, only the core system and the modules generated for the core system by the module generation tool are used in clinical routine. Main advantage is the comfortable extensibility of any implemented architecture by new items and new research questions. The module generation tool is used each time a new module has to be generated or an existing one has to be adapted. If the underlying terminology has to be changed, the TMS is used: further modules will then be built upon the changed terminology. Once the definition of a terminology for a trial in the TMS is finished, a consistent, corresponding database can be created within short time and without any informatics skills. The process of building forms takes place under strict terminological control.

As experiences in various medical fields have shown, the terminology-based approach is limited to specialized fields [21,22]. A comprehensive terminology has proven to be too complex and too difficult to maintain [23]. The quality of the multiple use of data depends mainly on the quality of the
planning of the items in the terminology management system. High quality harmonization will lead to a higher amount of multiply used data.

There are two further aspects, which will considerably influence the success of a documentation environment based on eardap:

- The architecture does not provide a universal solution for interfaces to local hospital information systems. According to state-of-the-art research in multihospital environments these interfaces have to be implemented for each participating hospital [24]. The effort for this can be minimized by employing existing standards like HL7.
- The architecture does not provide a universal solution for data protection, security issues [25,26] or long time archiving aspects [27,28]. To ensure certainty of law, legislation efforts on state, national level have to be fulfilled. For example, European legislation or standardizing efforts like CEN TC 215 (http://www.centc251.org/, e.g. ENV 13608 (2000), Health Informatics—security for healthcare communication) have to be regarded. For each implementation solutions have to be applied according to local circumstances and legislation. An example is the decision, whether the use of clinical data for research is legalized by informed consent of the patient or if the data is provided using pseudonyms or even anonymously.

4.2. Experiences in pediatric oncology

Our experiences in pediatric oncology have shown the applicability of the architecture. The functions of our core system were in routine use in several hospitals all over Germany. We validated the TMS and the module generation tool by entering and processing the basic data set of pediatric oncology in Germany. The algorithm we implemented in the data base generation component is universally valid, that means, with an appropriate interface it can also be used with other concept-oriented terminology management systems as well. The whole process of building the documentation systems is under strict terminological control.

Our experiences confirmed that terminology harmonization is a key factor for success and a challenging task in a nationwide project.

In contrast to other terminological approaches [29] we added textual definitions for concepts in order to prepare future formal definitions, but also to build glossaries for staff of clinical trial centers and sites. Furthermore, we offer the possibility to document free text and distinguish between standardized and not standardized items. In addition, we store which items are recorded in different contexts. Thus it is known which items have to be standardized and which clinical trials collect comparable data.

In contrast to second-generation systems [30] our TMS is more concept-oriented. Further, the TMS provides the possibility to manage common and trial specific terminologies, to store textual definitions of concepts and to manage integrity constraints and database terms. This structure is not only used to structure concepts but also as a basis for case report forms.

4.3. Conclusion: using the architecture for shared patient records

We have mainly regarded eardap in the context of multiple use of patient data. Continuing specialization in medicine strengthens the efforts for adequate support of documentation purposes in a shared care environment. That means that data is not only used at different locations for different purpose but in a shared care environment they originate from various sources. When using eardap as an underlying architecture for shared entry, special emphasis has to be laid on:

- Reduction of efforts for smooth integration of an eardap in existing health information systems. Further research is needed here: we pin great hopes on HL7 version 3 and document standards like the clinical document architecture (CDA).
- Reduction of efforts for exchanging electronic patient records or parts of electronic patient records between partners. For this, we currently analyze if eardap could employ openEHR methodology.

Thus, eardap is a proposal for an architecture, which is not in competition but in accordance with new concepts and approaches for electronic patient records [31—33]. It can be applied additionally in order to strive for higher quality and efficiency in future medical care for the sake of the patient.

We extensively tested the architecture in pediatric oncology where we were confronted with a variety of medical centers and research centers. We could prove the usefulness in heterogeneous environments. A next step would be to test the practicability in a new application area.
Towards shared patient records: An architecture for using routine data for nationwide research

Acknowledgments

The architecture was developed in the framework of the DOSPO-project at the Department of Medical Informatics of the Institute for Medical Biometry and Informatics of the University of Heidelberg. From 1995 until 1999 the DOSPO-project was supported by the German Leukemia Research Aid. From 2000 until 2002 the DOSPO-project was part of the Competence Network Pediatric Oncology and Hematology (GPOH) who supported this work and all colleagues who participated in the DOSPO-project.

References


[27] R. Brandner, U. Pordesch, Long-term Conservation of Prov-
ability of Electronically Signed Documents, Information
[28] A. Bess, R. Brandner, Deduction of principles on long-term
conservation of electronically signed documents by multi
professional analyses, in: The New Navigators: From Profes-
sionals to Patients, Proceedings of the Medical Informatics
Zanstra, W. Claassen, A terminology server for medical lan-
guage and medical information systems, Meth. Inform. Med.
[30] A. Rossi Mori, F. Consorti, E. Galeazzi, Standards to sup-
port development of terminological systems for health-
563.
[31] P. Ramsaroop, M.J. Ball, A Model for More Useful Patient
[32] E. De Clercq, The index as a new concept towards an inte-
grated framework for the electronic patient record, Meth.
generic model of clinical practice, Meth. Inform. Med. 42