MEDINFO 2010
C. Safran et al. (Eds.)
IOS Press, 2010

doi:10.3233/978-1-60750-588-4-1117

Facilitating Secondary Use of Medical Data by Using *open*EHR Archetypes

Christian D. Kohl[a,b], Sebastian Garde[c], Petra Knaup[a]

[a] *Department of Medical Informatics, Institute of Medical Biometry and Informatics, University of Heidelberg, Heidelberg, Germany*
[b] *Coordination Center for Clinical Trials Heidelberg, University of Heidelberg, Heidelberg, Germany*
[c] *Ocean Informatics, Düsseldorf, Germany*

Abstract

Clinical trials are of high importance for medical progress. But even though more and more clinical data is available in electronic patient records (EPRs) and more and more electronic data capture (EDC) systems are used in trials, there is still a gap which makes EPR / EDC interoperability difficult and hampers secondary use of medical routine data. The openEHR architecture for Electronic Health Records is based on a two level modeling approach which makes use of 'archetypes'. We want to analyze whether archetypes can help to bridge this gap by building an integrated EPR / EDC system based on openEHR archetypes. We used the 'openEHR Reference Framework and Application' (Opereffa) and existing archetypes for medical data. Furthermore, we developed dedicated archetypes to document study meta data. We developed a first prototype implementation of an archetype based integrated EPR / EDC system. Next steps will be the evaluation of an extended prototype in a real clinical trial scenario. Opereffa was a good starting point for our work. OpenEHR archetypes proved useful for secondary use of health data.

Keywords:

Computerized medical record systems, Clinical trials, Interoperability, Multiple use of data, Single source, *open*EHR archetypes.

Introduction

Clinical trials are the major element in clinical research and of high importance for medical progress. By now, more and more clinical data is available in electronic patient records (EPRs). Similarly, there is a trend to collect data in clinical trials by using electronic data capture (EDC) systems. Numerous efforts in research and industry have led to improvements of EPR and EDC systems in the past years. However, there is still a gap which hampers a straightforward data exchange between EPR and EDC systems and which makes it difficult to use routinely collected medical data in clinical research [1-3] – especially in multicenter studies [4].

The lacking interoperability between EPR and EDC systems not only hampers translational biomedical research but also results in the necessity to manually enter data for a clinical trial in an EDC system which is already available electronically in an EPR: A time-consuming task which beyond that may also cause transcription errors [5].

Another obstacle for EPR / EDC interoperability is that standards for electronic data exchange in clinical care (for instance Health Level 7 version 2) and in clinical research (for example standards of the Clinical Data Interchange Standards Consortium – CDISC) are still quite different [2-4].

Facilitating EPR / EDC system interoperability would not only help clinical research by saving time and money and avoiding transcription errors but also support translational research by reusing clinical care data for medical research and vice versa. Therefore, different efforts have been made in the last years to bridge this gap and there are some examples of successful approaches (for instance [5]). These approaches often integrate a particular EPR system with a particular EDC system (for example [6]). Even though this is a good starting point, many times multicenter trials involve trial centers which have different EPR systems in place.

Bridging the gap with *open*EHR

The *open*EHR architecture for Electronic Health Records (EHRs) is based on a two level modeling approach which makes use of 'archetypes' [7, 8]. Archetypes "... are reusable, structured models of clinical information concepts that appear in EHR, such as 'test result', 'physical examination' and 'medication order'..." [9]. We think that archetypes could not only facilitate the documentation of medical data and interoperability between different EPR and EDC systems but also support data recording in clinical trials and help to integrate study data management systems [10].

The overall aim of our study is to explore how a system for recording data in clinical trials can be built based on the *open*EHR approach so that it is able to reuse the data of an EPR system. We further want to investigate whether such an archetype-based approach enables smooth integration of existing medical data at different sites.

Therefore, we built the **open S**tudy **D**ata **M**anagement **S**ystem (openSDMS). OpenSDMS is a prototypical system based on

*open*EHR concepts which allows not only recording of medical routine data but also electronic data capture for clinical trials. OpenSDMS facilitates both entering medical data for clinical trials manually and also (in a 'single source approach') reusing medical routine data which were previously stored in the EPR module. Furthermore, the system shall be able to integrate data of different (also non-*open*EHR) EPR systems for one trial. This paper describes our concept and first experiences with implementing the openSDMS.

Materials and Methods

Opereffa

As a basis for openSDMS we use the '***open*EHR REF**erence **F**ramework and **A**pplication' (Opereffa – http://opereffa.chime.ucl.ac.uk) which is developed at the University College London. Opereffa is a web based application which uses Java Server Faces as web layer technology. In addition, the Dojo Toolkit is used to provide Asynchronous JavaScript and XML (AJaX) and extended user interface capabilities. To physically store data Opereffa uses Hibernate with an underlying PostgreSQL database as persistence layer. In an entity-attribute-value (EAV) model approach all archetype nodes are saved as attribute/value pairs in a single generic table [11]. That means the system uses only one table to store all medical data. Every row of this table contains one data element recorded for a certain patient at a certain time. Expressed in simplified terms, one dataset respectively one row consists of the following elements: Patient id (identifying a certain patient), session id (identifying a certain documentation context), archetype name, archetype path (together with archetype name unambiguously identifying one data element) and the actual value of the data element.

Opereffa itself makes use of the *open*EHR Java Reference Implementation [12].

Archetypes for clinical trials

As we wanted to document not only medical data which has been recorded for a clinical trial but also the structure and a description of the trial itself, we started to assemble meta data which describe typical concepts of clinical trials. We identified these data elements in a systematic review of the feature categories which are recorded in the German Clinical Trials Register (German CTR – http://www.germanctr.de) respectively in the International Clinical Trials Registry Platform of the WHO (ICTRP – http://www.who.int/ictrp/en/). Additionally we evaluated feature categories which are used in the Coordination Center for Clinical Trials Heidelberg (http://www.kks-hd.de) to register information about the trials conducted or supported by the coordination center. This list of feature categories was developed in an iterative process involving its end users.

In a next step, we defined archetypes for the study meta data in addition to existing archetypes for medical data. For this, we designed a model of concepts based on the previously assembled meta data which are necessary to describe the structure and nature of a clinical trial. The model is described in detail in the results section (Figure 1).

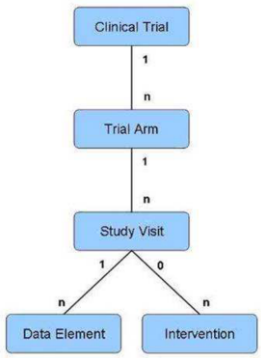

Figure 1 - Generic trial structure

Even though, these concepts are not directly related to clinical data but to clinical trials at a meta level, we used the *open*EHR Clinical Knowledge Manager (http://www.openehr.org/knowledge/) to check whether corresponding archetypes were already available or not. We used 'clinical trial', 'clinical study', 'trial', and 'study' as search terms. As suitable archetypes were not available, we started to define new archetypes using the 'Ocean Archetype Editor'. We created one archetype per concept of our model. Each of these archetypes defines the data elements which are necessary to describe an instance of the underlying concept.

openSDMS

In a next step, we built the openSDMS by extending Opereffa according to our requirements. For this, we added a 'trial view' to the user interface of Opereffa and implemented the underlying program logic. There, we took advantage of existing Opereffa code and used our newly defined archetypes.

Results

Archetypes for clinical trials

Based on our model of concepts for describing the structure and nature of clinical trials (Figure 1) we defined the archetypes 'clinical trial', 'trial arm', and 'study visit' which allow describing clinical trials by recording the particular meta data. For instance, the archetype 'clinical trial' defines 47 data elements like 'trial name' or 'phase' (Figure 2).

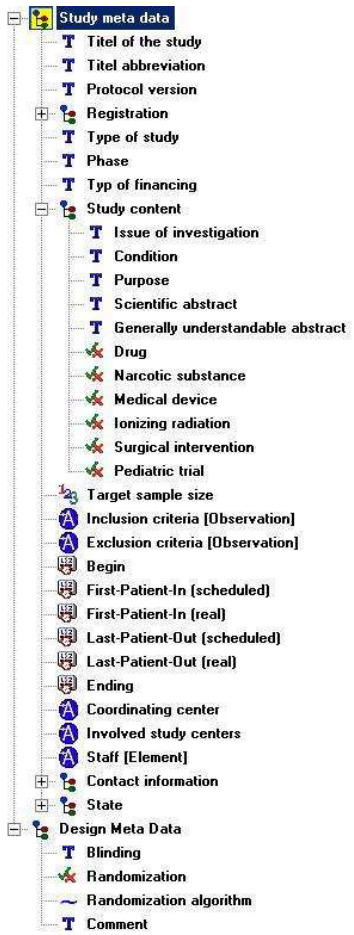

Figure 2 - Definition of the archetype 'clinical trial'

The concept 'clinical trial' is the root element of our model. One trial consists of at least one 'trial arm'. One trial arm consists of at least one 'study visit'. Within one visit at least one data item is captured or one intervention (medication, surgery...) is undertaken and recorded. While an observational study typically consists of one arm maybe with only one visit, a controlled trial consists of at least two arms.

Using the archetypes which reflect our model enables easy storage of data which describe a clinical trial in a system based on archetypes. It thus becomes possible to collect and store medical data and to assign these data to a certain trial arm, for example.

openSDMS

Our overall aim was to build a prototypical system based on *open*EHR concepts which allows electronic data capture for clinical trials not only by entering the data manually but also by reusing data which were previously stored in an EPR module in a single source approach. With openSDMS we now have a prototype for such a system. OpenSDMS is completely based on archetypes which means, that almost all data processing and storage within the system is based on archetypes.

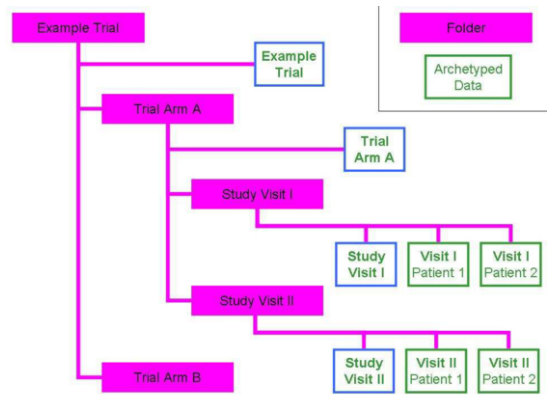

Figure 3 - Folder structure to model clinical trials

To facilitate a smooth presentation of all data items which logically belong together (for instance all data items which were captured for a certain study visit in a certain trial), we used the *open*EHR concept of 'folders' [13]. *Open*EHR folders enable logically grouping of recorded data and thus presenting the same data compositions in different folders. We used a similar mechanism in openSDMS to pool medical data not only of one patient but also of a certain clinical trial. For this, we use a virtual folder structure (Figure 3) according to our previously introduced generic trial structure (Figure 1). As a result, every folder contains – beside the medical data –some kind of meta data which describes the particular element and which is based on our newly defined archetypes. The concrete mapping of our generic trial structure to the folders is described in the following.

The 1:n relation between clinical trial and trial arm is modelled by having one folder per clinical trial which contains one dataset describing the trial itself (based on the clinical trial archetype) and a number of subfolders where each subfolder represents one trial arm. The 1:n relation between trial arm and study visit is represented analogously. The folder of each visit contains a visit description (based on the study visit archetype) and all medical datasets of different patients which have been recorded in the context of that particular visit. The single medical datasets will be visible in respective folders of the single patients together with further medical data. This means that medical data – independent of its use in a medical record or a clinical trial record – is physically stored only once in a generic way in openSDMS which complies with a 'single source approach'.

That way, it is possible to bring together all data of one study visit (of a particular trial arm in a particular trial) in one folder. The respective medical data of all patients at whom that visit took place is mapped to the respective directory. Supplemented by a description of the trial structure and other study

meta data, that results in a complete data set of one clinical trial.

Discussion

Archetypes for clinical trials

Per definition, *open*EHR archetypes are geared to comprehensively describe and structure a clinical concept in any context (cf. [14], pp. 10). Therefore, we believe that archetypes used to record medical data in EPR systems are also suitable to record medical data in the context of clinical trials. Furthermore, if an archetype was not found to be sufficient for the requirements of a clinical trial, it can be extended. It seems to be of advantage especially for clinical research that data stored in an *open*EHR based system have a 'built in' degree of validity as they have to conform to both a generic and stable information model and the models of clinical data as defined by the used archetypes.

We developed a dedicated set of archetypes to enable documentation of study meta data and to describe concepts associated with clinical trials like 'trial arm' or 'study visit' in the openSDMS. It would have been also possible to store these data just by creating additional tables in the database used by openSDMS. However, we decided to make use of archetypes because of two reasons:

1. Archetypes originally allow to systematically structure and describe the data which characterize the underlying concepts and by that, built a helpful discussion basis for the description of these concepts.
2. The archetype based representation enabled us using the same (already existing) program code which is used to process, store and transmit medical data, also for the processing of study meta data.

The archetypes we have developed for clinical trials are still draft versions. To enable their reusability in clinical trials nationally and internationally, we will bring these archetypes into a critical, public review process.

openSDMS

With our system we are able to explore technical issues associated with secondary use of medical data in EPR systems for clinical trials and to evaluate possible solutions. However, our system is still a prototype in the sense that just core functionality has been implemented. In contrast to a real study data management system further functionality is missing. For instance no plausibility checks or query management functionality have been implemented, yet.

Furthermore, Opereffa proved as a good starting point for the development of openSDMS. However, Opereffa itself is still under development and has to be improved step by step to fully implement the openEHR specifications. Therefore, we tried to implement openSDMS based on Opereffa but by observing information hiding principles as good as possible to allow for an update of Opereffa without having to adjust internal parts of the openSDMS implementation.

One relevant aspect of reusing data in EPR systems for clinical trials is data privacy protection. The underlying database of openSDMS consists of mainly one entity-attribute-value model based data table which only contains a patient id for relating the single entries to the particular patients. This seems to be an advantage, as it facilitates a separation of identifying data (like patient name) and medical data. According to the respective view and user (for instance medical record/physician or trial record/data manager) the system could display the patient name respectively only a pseudonymized patient id. We will explore whether this idea can be supported by integrating a pseudonymization software which has been developed by the German ‚Telematikplattform – Verbund zur Förderung vernetzter Medizinischer Forschung' (TMF – http://www.tmf-ev.de) specifically for scenarios like this one. The TMF crosslinks medical research in Germany as umbrella organization.

Perspective

In a next step, openSDMS will be extended by an export function for the study data based on the Operational Data Model (ODM – [15]) which has been developed by the Clinical Data Interchange Standards Consortium (CDISC). This shall enable openSDMS to export the captured trial data to other study data management systems.

However, the system shall not only facilitate storing data in an integrated EPR module and using that data for clinical trials but also to integrate data of different EPR systems for one trial. Therefore, in a further step, interoperability with other EPR systems will be established to enable importing of existing medical data into openSDMS. It is planned to make use of integration archetypes for that task.

Having openSDMS acting as a bridge between EPR and EDC systems, in a third step, the strengths and weaknesses of openSDMS will be evaluated. We plan, to evaluate the openSDMS in a real but manageable clinical trial setting.

Our approach and aims have similarities with other existing approaches – for instance the STARBRITE Proof-of-Concept Study [5] which uses the CDISC Operational Data Model and HL7 Clinical Document Architecture to integrate clinical routine and trial data in a single source approach. In contrast to these approaches we use openEHR archetypes for integrating clinical routine and trial data. Archetypes as structured models of clinical information concepts pledge to be a good basis for the future integration of different data sources which typically occur in multicenter trials. Furthermore, in contrast to other approaches, systems based on *open*EHR archetypes promise high sustainability as they strictly separate the modeling of clinical concepts (archetypes) and the technical implementation. If a concept has to be adapted to new requirements this can be done easily and not only by technicians but also by domain experts – for instance physicians.

Conclusion

Archetypes for clinical trials

*Open*EHR archetypes which have been defined to document medical data in EPR systems seem to be also suitable to document medical data in the context of clinical trials and to support single source approaches. Furthermore, the archetypes we developed make it possible to document trial metadata in an *open*EHR based system in the same way as medical data.

openSDMS

The core functionality of openSDMS has been developed. However, some fine tuning is required – for instance with regard to the graphical user interface – until the system can be evaluated in a real clinical trial scenario. This will be a further step of our research. We assume that one necessary key factor for the success of openSDMS in addition to its flexibility will be its ease of use. Especially the generation of case report forms has to be supported by easy-to-use tools.

Acknowledgments

The authors would like to thank all those who have contributed with commitment and enthusiasm to the *open*EHR project, the *open*EHR Java Reference Implementation, and the *open*EHR Reference Framework and Application. We especially thank Heather Leslie and Ian McNicoll of Ocean Informatics for their feedback in terms of the modeling of clinical trials.

References

[1] Embi PJ, Payne PR. Clinical research informatics: challenges, opportunities and definition for an emerging domain. J Am Med Inform Assoc 2009: 16 (3): 316-27.

[2] Ohmann C, Kuchinke W. Future developments of medical informatics from the viewpoint of networked clinical research. Interoperability and integration. Methods Inf Med 2009: 48 (1): 45-54.

[3] Prokosch HU, Ganslandt T. Perspectives for medical informatics. Reusing the electronic medical record for clinical research. Methods Inf Med 2009: 48 (1): 38-44.

[4] Winter A, Funkat G, Haeber A, Mauz-Koerholz C, Pommerening K, Smers S, Stausberg J. Integrated information systems for translational medicine. Methods Inf Med 2007: 46 (5): 601-7.

[5] Kush R, Alschuler L, Ruggeri R, Cassells S, Gupta N, Bain L, Claise K, Shah M, Nahm M. Implementing Single Source: the STARBRITE proof-of-concept study. J Am Med Inform Assoc 2007: 14 (5): 662-73.

[6] Katzer C, Röhrig R, Weismüller K, Little S, Langefeld T, Meier M, Menges T, Chakraborty T, Hempelmann G. Erfassung, Verarbeitung und Qualitätssicherung von Studiendaten mit einem klinischen Informationssystem – Ein Erfahrungsbericht. In: Löffler M, Winter A, editors. 51 Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (gmds) Leipzig: Düsseldorf, Köln: German Medical Science; 2006; pp. 344-45.

[7] Leslie H, Heard S. Archetypes 101. In: Westbrook J, Callen J, editors. HIC 2006 Bridging the Digital Divide: Clinician, consumer and computer: Health Informatics Society of Australia Ltd (HISA); 2006.

[8] Michelsen L, Pedersen SS, Tilma HB, Andersen SK. Comparing different approaches to two-level modelling of electronic health records. Stud Health Technol Inform 2005: 116: pp. 113-18.

[9] The openEHR Foundation. Introducing openEHR. 2007; Available from: http://www.openehr.org/releases/1.0.1/openEHR/introducing_openEHR.pdf - last access: September 23, 2009.

[10] Garde S, Knaup P, Schuler T, Hovenga E. Can openEHR Archetypes Empower Multi-Centre Clinical Research? Stud Health Technol Inform 2005: 116: pp. 971-76.

[11] Arikan S, Ingram D. Web Technologies used in Opereffa. 2009; Available from: http://opereffa.chime.ucl.ac.uk/technologies.jsf - last access: September 23, 2009.

[12] Chen R, Klein G. The openEHR Java reference implementation project. Stud Health Technol Inform 2007: 129 (Pt 1): 58-62.

[13] Beale T, Heard S, Kalra D, Lloyd D. EHR Information Model. The openEHR Reference Model 2008; Available from: http://www.openehr.org/releases/1.0.2/architecture/rm/ehr_im.pdf - last access: September 23, 2009.

[14] Beale T, Heard S. Archetype Definitions and Principles. 2007; Available from: http://www.openehr.org/releases/1.0.2/architeture/am/archetype_principles.pdf - last access: September 23, 2009.

[15] Iberson-Hurst D. The CDISC Operational Data Model: Ready to Roll? Appl Clin Trial 2004: 13 (7): 48-53.

Address for correspondence

Christian D. Kohl
Coordination Center for Clinical Trials (KKS)
Voßstr. 2 / Building 4410
D-69115 Heidelberg
Phone: +49 (0) 6221 / 56 - 34510 Fax: - 33508
Email: Christian.Kohl@med.uni-heidelberg.de