Domain Modeling and Application Development of an Archetype- and XML-based EHRS

Practical Experiences and Lessons Learnt

Stefan Kropf[1,2,*]; Claire Chalopin[1,*]; Dirk Lindner[3]; Kerstin Denecke[4]

[1] Innovation Center Computer Assisted Surgery (ICCAS), Leipzig University;
[2] Institute for Medical Informatics, Statistics and Epidemiology (IMISE), Leipzig University;
[3] University Hospital, Department of Neurosurgery, Leipzig University;
[4] Institute for Medical Informatics, Bern University of Applied Sciences;
[*] Contributed equally

Keywords

Electronic health record, EHR standards, medical informatics applications, health information exchange, openEHR, XML

Summary

Background: Access to patient data within the hospital or between hospitals is still problematic since a variety of information systems is in use applying different vendor specific terminologies and underlying knowledge models. Beyond, the development of electronic health record systems (EHRSs) is time and resource consuming. Thus, there is a substantial need for a development strategy of standardized EHRSs. We are applying a reuse-oriented process model and demonstrate its feasibility and realization on a practical medical use case, which is an EHRS holding all relevant data arising in the context of treatment of tumors of the sella region. In this paper, we describe the development process and our practical experiences.

Methods: Requirements towards the development of the EHRS were collected by interviews with a neurosurgeon and patient data analysis. For modelling of patient data, we selected openEHR as standard and exploited the software tools provided by the openEHR foundation. The patient information model forms the core of the development process, which comprises the EHR generation and the implementation of an EHRS architecture. Moreover, a reuse-oriented process model from the business domain was adapted to the development of the EHRS.

Results: The reuse-oriented process model is a model for a suitable abstraction of both, modeling and development of an EHR centralized EHRS. The information modeling process resulted in 18 archetypes that were aggregated in a template and built the boilerplate of the model driven development. The EHRs and the EHRS were developed by openEHR and W3C standards, tightly supported by well-established XML techniques. The GUI of the final EHRS integrates and visualizes information from various examinations, medical reports, findings and laboratory test results.

Conclusion: We conclude that the development of a standardized overarching EHR and an EHRS is feasible using openEHR and W3C standards, enabling a high degree of semantic interoperability. The standardized representation visualizes data and can in this way support the decision process of clinicians.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

© Schattauer 2017

Correspondence to:
Stefan Kropf
Innovation Center Computer Assisted Surgery (ICCAS)
Leipzig University
Semmelweisstraße 14
04103 Leipzig
Germany
Email: stefan.kropf@imise.uni-leipzig.de

Appl Clin Inform 2017; 8: 660–679
https://doi.org/10.4338/ACI-2017-01-RA-0009
received: January 14, 2017
accepted: April 20, 2017
published: June 28, 2017
Citation: Kropf S, Chalopin C, Lindner D, Denecke K. Domain modeling and application development of an archetype- and XML-based EHRS: Practical experiences and lessons learnt. Appl Clin Inform 2017; 8: 660–679
https://doi.org/10.4338/ACI-2017-01-RA-0009
Funding
This work was part of The Digital Patient Model Project (ICCAS), granted by the BMBF (03Z1LN11).

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

1. Introduction

The number of patient data generated during the diagnosis, treatment and monitoring of diseases is continuously increasing due to the availability of more and more complex examination methods. Their formats are very heterogonous: Image datasets, laboratory test results, histopathological findings, various reports and medical letters are provided as results from different examinations. In medicine, it is relevant to access latest research results, past patient cases and the medical knowledge as fast as possible as basis for clinical decision making. This requires easy access to the relevant data, which is currently often problematic due to information distributed in various information systems.

In hospitals, patient data are digitalized in form of electronic health records (EHRs) and stored within distributed databases in hospital information systems (HIS) [1]. Image data acquired within the hospital are stored in a Picture Archiving and Communication System (PACS) using the DICOM (Digital Imaging and Communications in Medicine) standard and visualized to the clinician by a DICOM-viewer [2]. Results of laboratory tests are also commonly managed within specific software, specially designed for such particular data. External medical letters and findings are scanned and stored in specific data bases, accessible through their respective user interfaces. It becomes clear that it can be very complex for clinicians to find the relevant information among the available data distributed in the different software components of the HIS.

In brief, different vocabularies and terminologies, different formats and languages yield in a diversity of data [3, 4] and a lack of data integration [1]. This implies a need for a centralized EHR system (EHRS) that integrates the data and that is capable of handling EHRs which are represented in a structured way according to standards. Therefore, the data should be fast accessible for users and applications which work on EHRs. ►Fig. 1 illustrates the latter described scenario.

In this paper, we introduce a reuse oriented process for developing an overarching EHRS and report on our experiences in the development process. This process is implemented for a specific medical use case which is the surgical treatment of patients with tumors of the sella region. This use case was selected because different patient data of various formats are involved in the treatment of this disease and in the follow-up monitoring. The remainder of this paper is structured as follows: After introducing the medical use case, requirements for the EHRS development are summarized. Subsequently, we describe the standard selection process, and the reuse oriented modeling and development process. In the result section, the obtained information model and architecture of the EHRS will be presented. The paper finishes with a discussion on the approach and conclusions.

2. Materials

2.1 Medical use case: treatment of tumors of the sella region

Tumors of the sella region are located in a saddle-shaped cavity of the sphenoid bone of the human skull whose deepest part includes the pituitary gland (►Fig. 2).

Usually, the tumor is benign, but it can affect the normal production of hormones, such as the sexual or growth hormone. These dysfunctions are responsible for example for infertility, erectile problems or enlargements of the extremities (hands, nose, ears) and of the internal organs (heart, kidneys) with higher risk causing premature death. Because of the proximity of the sella region with the optic chiasm, which is the area where the optic nerves cross, large tumors can moreover affect the vision of patients. The diagnosis of the presence of a tumor of the sella region is primarily performed based on magnetic resonance (MR) imaging. The position of the tumor and its size are estimated. Additive examinations, especially eye examinations and blood tests which measure the quantity of hormones, provide complementary information about the kind of tumor. All these features strongly influence the choice of therapy. After initial treatment, the patients are returning to the hospital for a regular check-up (for example once a year). This requires the availability of the updated data to be able to monitor possible progress of the disease.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

2.2 Requirements for the development of an improved EHRS

By interviewing a leading surgeon working at the department of neurosurgery at the university hospital of Leipzig, we collected requirements for an improved patient information access.

1. The main expectation of neurosurgeons is to *facilitate and speed up the access of patient data* of various formats.
2. A second objective is to offer the possibility of *exchanging patient data* between hospitals.
3. *Optimal visualization* of data in the EHRS is important so that physicians can get quickly an overview on the patient health status. For example, labor data are usually reported with text and tables of values.
4. Finally, new kinds of patient data are continuously generated at hospitals. This requires continuous adaptation and flexibility of the information system.

From these requirements, we derived the following tasks and technical requirements for our EHRS development:

1. Development of a centralized EHRS integrating *various multimedia data* like image datasets (DICOM images, screenshots and videos) and text data to speed up the access to patient data.
2. Representation and storage of treatment relevant patient data according to *standards* to support the exchange of patient data between hospitals.
3. Integration of visualization tools to facilitate the interpretation of patient data. *Laboratory test values*, and especially hormone concentrations in blood, are considered as an example. In general, the concentration of 12 endocrine values is measured for the diagnosis of tumors of the sella region: thyroid hormones (TSH, fT3, fT4), sex hormones (LH, FSH, testosterone, prolactin), adrenal cortical hormones (ACTH, cortisol) and growth hormones (hGH, IGF-1, IGF-1 SDS). These different values are currently represented by tables with numbers which complicates the quick identification of clinical problems. The visualization of such data still has potentials for improvements.
4. Use of software development models and available software tools to facilitate and support the implementation and update of the EHRS.

2.3 Related standards

In the healthcare domain, several interoperability standards have been specified. Interoperability describes the extent to which systems and devices can exchange data, and interpret that shared data. For two systems to be interoperable, they must be able to exchange data and subsequently present that data in a user-understandable manner. However, choosing the right standard is challenging, because a comparison of EHR standards reveals no real winner [5] and their application to real world problems is difficult [6]. By means of a literature search, we identified diverse European Norms that are relevant for our use case. For a pre-selection of suitable standards, we specified the following requirements towards the standard. The first requirement concerns self-evidence: a standard should be published by an authority. Another requirement results from the fact that the standard should be practically applicable. In order to be able to develop a practical solution, the standard should have a level of specificity of *logical* or *physical* and a view of perspective of *what* or *how* regarding ISO 17119 [7]. Therefore, the standard should deliver an applicable reference information model (RIM), including data type definitions or ideally development instructions for the mapping of healthcare data in data structures. Further, modularization and reuse of EHR components and support of the EHR development by using available software tools is required, enabling separation of the patient information modeling from the development of the application.

From 21 European Norms (ENs) that were in principle relevant, 19 could be excluded quickly, since they are too specialized to concrete use cases, like audit trails (EN 27789), electrocardiography (EN 21091) or device communication (EN 11073). Other standards were considered as too generic for an implementation, since concrete data types or development instructions in the service architecture were not specified (EN 12967). In particular, the norm for the patient health card data (EN 21549) defines only the wrapping envelope, not the content. In addition to ENs, CDISC ODM (Operational Data Model) [8] was identified as a suitable standard for clinical EDC (Electronic Data

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

Capture). The data model is excellent adapted to entry forms, but too rigid for a flexible modeling of clinical information and DICOM integration is not possible.

Hence, after the initial standard review, only two standards could be selected as suitable for our medical use case, namely EN 13606 [9], which is a subset of openEHR [10] and EN 14822 aka Health Level 7 (HL7) v3 Reference Information Model (RIM) [11]. HL7 RIM is used for structuring Clinical Document Architecture (CDA) [12] and Fast Healthcare Interoperability Resources (FHIR). Though CDA is based strictly on the HL7 RIM, FHIR is not [13] and even though there is comprehensive development ongoing around FHIR, it is still a draft version [14], hence it could be not stable enough for long term persistence.

By reading papers [5, 6], the standards itself [9, 11] and related websites [12, 13, 15] EN 13606, openEHR and CDA have been examined in more detail and compared with respect to the requirements of our use case. In brief, we decided for the usage of openEHR because free available modeling tools are available as well as reusable components.

2.4 OpenEHR modelling tools

We decided for the usage of openEHR because free available modeling tools are available as well as reusable components. The reuse of clinical models, which are developed by domain experts and are available for other users, promises an accelerated system development process. A large database of patient information models, called *archetypes*, is available in the *Clinical Knowledge Manager (CKM)* repository [15]. The *Archetype Editor* [16] tool aims at supplementing existing archetypes, at building new ones and at assembling them. The *Template Designer* is a tool to build the patient information model by arranging the archetypes together and by defining constrains on the archetypes.

The archetype structure offers the possibility of modularization to include child-archetypes into a parent-archetype. In this way, additional and more specific information can be provided. For example, the existing archetype *Imaging examination result* which reports radiological findings can be linked with the archetypes *Anatomical location* and *Organisation* which indicates the examined body part and the hospital where the examination was performed. Another property of openEHR is information coding. For example, the status of an examination result is coded using the codes "*Registered*", "*Interim*", "*Final*", "*Amended*" or "*Canceled / Aborted*". Beyond, it is possible to link archetypes with international standards, like the Logical Observation Identifiers Names and Codes database (LOINC) for identifying medical laboratory observations.

2.5 Archetype-based EHR-development frameworks

Ten frameworks supporting the development of openEHR-based application could be identified by a literature and web search. To further decide for our development methodology, we tested existing openEHR-frameworks in an explorative way. In the first step, three frameworks were excluded since neither an English nor a German documentation was available. Another one had to be excluded because the framework was not sufficient for the persistence of the EHRs. In the second step frameworks were installed based on their documentation. Two frameworks could not be properly installed and were excluded of the evaluation. After that, the persistence of one simple EHR was tested on the four remaining frameworks. GastrOS [17] was a promising candidate but another architecture was discovered; in the frameworks EHRflex [18], LiU EEE [19] and MEDrecord [20] the persistence of the EHRs was realized on an XML database, implying that the EHRs are accessible by a REST interface [21].

Our EHRS approach consists of the same XML database architecture, comparable to another archetype-based data warehouse environment [22]. Because we want to build up the EHRS on W3C standards, we extended the REST architecture using the W3C XForm method for a binding of EHR elements to web form items, like already done in the Single Source Architecture of x4T [23, 24] for CDISC ODM. In addition to that we discovered in our prototype how several multimedia elements like PDFs, images or movies can be integrated on EHR XForms and how laboratory values can be visualized.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

3. Methods

3.1 Separation of patient information modeling and application development

OpenEHR has been specifically designed to allow a division of work: patient information modeling [25] can be clearly separated from the application development (▶Fig. 3). In this way, both tasks can be performed within an interdisciplinary working group; domain modeling can be realized in close collaboration by physicians and scientists. The result is a patient information model of a given medical context. The patient information model is a precondition for the application implementation which is realized by a computer scientist. This separation of domain modeling and application development is also reflected in the structure of this paper; domain modeling and application development are presented in different subsections.

Therefore, domain modeling of an EHRS development (upper part of ▶Fig. 3) can be considered as a cyclic process. In reality, not only domain modeling is cyclic, there are circles during the application development as well. This consideration reminds on software developing methods and well explored contemplations about reusing software components, which will be re-introduced in the following subsection.

3.2 Domain modeling

3.2.1 Reuse-oriented process model

We selected a reuse-oriented process model from the business domain devised by Caldiera and Basili [26] and adapted it to the modeling of patient data underpinned with openEHR tools. In brief, Caldiera and Basili identified an overarching project organization circle in the business domain and refined the component generation by cycles in the component factory domain. We substantiated this process model for modeling reusable components like Archetypes or Templates, which are defined by the information model [9]; moreover, Compositions are defined as "set of information committed to one EHR by one agent, as a result of a single clinical encounter or record documentation session" [9].

The business domain circle (left circle in ▶Fig. 4) represents a generic cyclic software development method. Different approaches are available, for example a waterfall model, a spiral model or even scrum or rapid prototyping. We used rapid prototyping for our development. The model factory domain (▶Fig. 4 right) refines the modeling process and consists of two circles. The upper one is traversed for each queried component, which are archetypes or templates. If no suitable component is found, a new one has to be developed (lower right circle).

Our process of developing an overarching EHRS started with the upper left cycle in ▶Fig. 4. In the first days the requirement analysis was performed based on interviews with an experienced neurosurgeon; then, the clinical workflow of the treatment of patients with tumor of the sella region was described, and relevant associated patient data was selected. After the specification of the system, modeling took place in the factory domain. During the modeling process, archetypes and templates, which are reusable components, were queried using the *CKM*. If no reusable component was available, a generation process was triggered: new archetypes were modeled (*Archetype Editor*) and integrated into templates by exploiting the *Template Designer*. The modeled archetypes and templates were stored in a local repository. After evaluation, the Compositions were released into the business domain. During the integration process, EHRs were instantiated out of the Compositions. The EHRs were then integrated into the EHRS. At the end of the business domain cycle, an integration and system test was carried out before the EHRS was finally released. The reuse-oriented process of modeling patient data was retriggered when new requirements arrived.

3.2.2 Patient information modeling using openEHR archetypes

From the relevant patient data that is generated in the treatment workflow, we selected information items and parameters included in the raw patient data which are relevant for clinicians, and particularly for the neurosurgeon. Beside medical parameters like hormone concentrations, tumor size or

© Schattauer 2017

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

histopathological findings, the place of examination and the time point in the clinical workflow and in the absolute time are relevant information items. Moreover, conclusions that physicians deduce from the raw patient data, for example the tumor classification based on radiological, laboratory and histopathological examinations and the current patient state in comparison to previous examinations were modeled. The individual archetypes were integrated in a common patient information model, called *Template* in openEHR, using the *Template Designer* tool. Elements can be set *prohibited*, *optional* or *mandatory*. This allows adjusting the archetypes to a given medical context and this avoids the overload with irrelevant information.

3.3 Application development

The development of the EHRS is based on the standard selection and patient information modeling. We described this combination of standardized EHRs and XForms already in previous work [27]. In brief, an XML based EHRs template is generated out of the XSDs, which are generated after patient information modeling by the *Template Designer*. Afterwards the XML based EHR template is bound to an XForm. After that step, the EHRs can be populated with data and persisted on the XML database eXist-db [28], accessible for EHR applications by a REST interface. Besides the EHR itself, a management mechanism around the EHR is necessary, e.g. lists of EHRs or the instantiation of a new EHR out of the template. In our prototype, this was realized by XQuery [29], which can be hosted and executed on eXist-db.

Multimedia elements or DICOM images could not be represented by XForms. This issue was solved by introducing links in multimedia elements and a simple XQuery expression. Hence, the integration method of the required multimedia formats was a URI [30] reference in the DV_MULTIMEDIA Class [31]. DICOM data was hosted by the free and open-source PACS orthanc [32]. For the storage of the other data any kind of content management system (CMS) can be used, which offers direct links and can manage different kinds of resources, like images or PDFs.

Moreover, this multimedia integration method offers the visualization of a preview of the multimedia elements directly on the EHR. In addition to strict HTML tags, many different languages, frameworks and quasi methods are available for the development of visualization styles. Some examples are hGraph [33], Google Charts [34], HTML5 Canvas [35] and SVG [36]. SVG has one big advantage, it is XML based, because of that it is producible out of XML based EHRs by XQuery or XSLT, which is applicable in our XML based EHRS environment. Consequently, we used SVG as method for the visualization of hormone laboratory tests.

4. Results

In this section, we firstly describe the selected patient data associated to the treatment of tumors of the sella region, the information modeling based on archetypes and the resulting patient information model. Secondly, we present the developed EHRS.

4.1 Domain modeling

4.1.1 Clinical workflow of treatment of tumors of the sella region and associated patient data

The analyzed clinical workflow includes three phases: the *diagnosis*, the *therapy* of the disease and the *monitoring* of the patient. Patient data involved in the surgical treatment of tumors of the sella region are reported in ►Tab. 1.

The diagnosis is mainly performed based on patient symptoms, on the examination of MR or CT image data, on the measurement of hormone concentration in blood and on eye examination. These examinations allow correlating the possible presence of the tumor with patient symptoms and disturbances in hormone production and in the visual field. Findings are documented in outpatient notes, radiological, laboratory and medical reports. An accurate tumor classification influences the therapy choice.

© Schattauer 2017

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

Surgical therapy starts with operation planning, which is performed based on CT and MR image data, acquired preoperatively. An ENT examination is realized prior to the operation. Both, CT and MR image data, are used within a navigation system to guide in a first step the ENT surgeon to reach the tumor through the nose ways and in a second step, guide the neurosurgeon to remove the tumor. Video data and screenshots of the endoscopic views are recorded during the operation and are intended for documentation. Tumor removal is examined and histopathological findings are reported. After the operation, postoperative CT or MR image data are acquired in order to control the operation outcome. Finally, the surgeons write a report describing the operation steps and outcome.

During the monitoring phase, the same patient data is generated as those obtained in the diagnosis phase, hence archetypes of the *diagnosis* phase could be reused. Clinicians document the current patient status in comparison to the one reported in the last consultation.

4.1.2 Patient information model based on openEHR archetypes

For the selected parameters and information items, openEHR archetypes have been specified or – wherever possible – reused from existing archetypes available in the CKM. 18 archetypes were required; ten could be reused from available archetypes (►Tab. 2), eight have been developed for our purposes.

For example, the archetype *Sella region tumor imaging classification* which documents the tumor classification deduced from radiological examinations has been defined by us (►Fig. 5). It includes geometrical parameters of the tumor (sizes, localization, symmetry), the expected kind of tumor, and if relevant, a comparison with previous examinations. Possible items for the expected kind of a tumor and the Hardy classification (classification based on tumor geometry) were coded.

After the archetype development, constraints on the elements of the existing archetypes were added by the tool *template designer*. In total, around 20% of the items were retained in the reused archetypes included in our application.

The core of the reuse-oriented process model devised by Caldiera and Basili is supported by the openEHR tools CKM, Archetype Editor and Template Designer (►Fig. 5). Especially the CKM turns out to be very important for the following processes: (1) the querying of components, (2) storage in a repository, (3) evaluation of the archetype (a collaborative review by domain experts), and finally (4) the release of the model as boilerplate for application development.

4.2 Application development

4.2.1 Generated XML based EHR

In the center of the EHRS architecture is the EHR, which is stored on an XML database.

Listing 1 (►supplementary online Appendix) shows a part of the imaging examination report model which is related to the first two items of the entry form in ►Fig. 7. The first item is simple text (DV_TEXT), the second one is coded text (DV_CODED_TEXT). The code string in the correspondent tag (at0061) is defined as CT with contrast agent in the Imaging examination result archetype and consequently in the resulting XSD. This XML based EHR snippet was stored as part of an overarching EHR on the resulting EHRS.

4.2.2 EHRS architecture

►Fig. 6 illustrates the architecture of the developed EHRS. The core is the persisted EHR, which is connected to the PACS and the CMS by URIs. Inside eXist-db, XForms was used for the generation of entry forms, SVG for EHR visualization and XQuery for business logic, respectively EHR administration. EHR applications (cf. ►Fig. 1) can access the EHR by a REST interface, e.g. for the purpose of natural language processing (NLP) in certain parts of an imaging report. Because we persisted the EHR based on the openEHR standard, other systems can rely on the structure of the data, which is defined by the reference model and the archetypes; another EHRS can interpret the elements of the EHR, therefore semantic interoperability is ensured.

The GUI provides the user with an easy access to one overarching EHR, which contains all relevant patient data. ►Fig. 7 shows the resulting GUI visualization of the archetype *Imaging examination result*. To keep the cognitive load of the clinician as low as possible, colors were used for a better overview, and each level of the EHR is foldable.

© Schattauer 2017

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

For visualizing hormone concentrations in an intuitive, graphical manner (►Fig. 8), absolute values are noted in the table cells. Columns include laboratory values measured at a given time point, while rows show value variations in time. Graphics on the top provide an overview of normal and pathological values. Bar charts on the right side depict the concentration variations in time for a given hormone. Graphics and bar charts can be zoomed by a click.

5. Discussion

In this paper, we introduced the development of an overarching EHR using openEHR. OpenEHR supports the modeling of EHRs by archetypes, delivers software tools for modeling EHRs, and separates domain modeling from application development. The visualization of laboratory test data was done by XQuery, SVG and JavaScript.

In the following, we discuss advantages and limitations of openEHR, the XML based architecture and information modeling standards.

5.1 Experiences using openEHR for domain modeling

We demonstrated that it is possible to develop a customized EHRS based on standards and on a patient information model. One elaborated advantage is that the complete modeling and development process is closely related to well explored reuse oriented software development cycles. OpenEHR tools, especially the repository *CKM,* support the processes of the factory domain.

Altogether the approach has the following advantages:

- Patient information modeling and application development are separated; both can be performed by different persons with appropriate background knowledge.
- The availability of free software tools and repositories allows scientists and physicians to structure the patient information model by archetypes. The reuse of existing archetypes leads to a significant time gain in the implementation and update process.
- When different systems from different vendors recycle the same archetypes, system boundaries can be overcome because of the increase in semantic interoperability.

The documentation on the openEHR standard is very comprehensive [37]. Anyway, it was sometimes difficult to use the openEHR tools in practice. Much effort had to be done to understand the domain modeling process and also the persistence process of the EHR instances, because different successful solutions are conceivable [38]. A documented recommendation about the standardized persistence of openEHR based EHRs would be helpful for developers.

5.2 Application development

5.2.1 Experiences using an XML based EHRS architecture

An XML based EHRS architecture comes along with advantages, compared to a relational database architecture:

- The development of a relational database can be skipped; costs can be avoided.
- Adaptions of the information model can be translated quickly in XML.
- XML is more accessible and can be semantically enriched by semantic web techniques.
- Access to DICOM data can be gained by a conversion to XML [3].

On the other hand, there are some limitations of the approach. All media data were linked by URIs to external systems. Interoperability between two EHRS will only work, when these links are accessible by both EHRS. Multimedia data could be stored directly in the EHR, but this would cause an increased load for the suggested XML data base. Instead, it has to be considered, whether the multimedia data should be stored in a relative directory or directly in the EHR. Such an overarching EHR, including all kind of media data linked relatively, could be stored and packed physically in one file.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

This would solve many interoperability and security issues since overarching EHRs containing all kind of media data could be transmitted like DICOM images and stored on health cards.

The reference information model [9] allows the commendable definition of repeating nodes, e.g. for the definition of multiple events or even multiple leaf elements like report or image instances. This multiplicity was not considered in the GUI of the developed prototype yet, but is required.

The XML has much overhead, resulting in large XML files, costing disc space and performance as well. The reason is that tags are often mandatory defined in the XSDs. An easy technical solution would be to keep away empty tags. The decision which tags are mandatory has to be done by the standard publishers in order to stay conform to the standard definition.

To avoid bottlenecks in the performance, it is recommended to estimate the size of a productive XML based EHRS during the specification step. Based on this estimation an early performance test should proof practicability and scalability, especially for big data scenarios. The performance test for the medical use case presented in this work was published in previous work: after processing 10'000 files (64 KB), the XForm GUI generation takes in average less than 2s [27], which is acceptable.

5.2.2 Information modeling standards

►Fig. 9 illustrates present and future standards, which are relevant in this paper.

Using standards and terminologies enables a high degree of semantic interoperability and is largely determined by the architecture [39]. In this work, we have chosen openEHR to be able to reuse archetypes. Further, for openEHR development powerful tools are available which simplifies the development.

XPath, XQuery and XSLT are used for retrieval and transformation of the EHRs. Though none of the latter languages is at present popular [40], it is possible to develop rapid EHRS prototypes, especially customized for semantic querying [41]. The management of the EHRs with XQuery results in functional code but in a violence of the model view controller (MVC) pattern and a resulting mix of business logic and visualization, because of that the usage of an additional higher-order programming language like Java should be considered. However, XQuery and XSLT strongly depend on XPath expressions; the development of XPath expressions can be supported by the search ontology XML extension [42].

In addition to persistence, retrieval and visualization, a semantic enrichment of the EHRs by the ontology standards RDF/RDFS and OWL is necessary in future. Accordingly, the resulting XML can be extended by additional semantic information directly in XML. Classes and properties of the Health and Lifesciences Extension of schema.org [43] or concepts of existing ontologies could be bound to terms or whole sections in the EHR. Listing 2 (►supplementary online Appendix) exemplifies the semantic enrichment of the item *Clinical information provided* by external semantic information.

In summary, openEHR could be successfully applied and implemented. We shared our knowledge and our strategy, but anyway, there is still a need for effective strategies for secondary use of EHR data [44]. Beside of such strategies a paradigm shift in thinking is required, based on the requirement for standardized concept representation [45].

5.3 Future Work

In a first rapid prototype, we proofed the concept of developing an EHRS based on openEHR for research purposes only. We tested the system with anonymized patient data without person-identifying data (see section Human Subject Research Approval). We excluded at this point of the project the use of the EHRS for clinical decision making. In future work, before we can roll out the EHRS in a real productive clinical environment, we have to respect audit trails and norms like quality management (EN 13485), risk management (EN 14791) as well as several other norms and check for regulations for medical devices.

The implementation of a standard based EHRS implementation enables an unambiguous representation and exchange of clinical meaning [46]. EHR applications (►Fig. 1) can access and process the persisted data by a REST interface. In future work, NLP methods could be applied for deriving structured information about patient phenotypes [46], machine-learning for predicting patient characteristics [47], and maybe the linkage to molecular and genetic data [46,48]. The inclusion of

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

image processing and text mining algorithms within the EHRS could automatically extract the relevant information from the raw patient data. An expanded visualization by visual analytics could help the surgeons by monitoring EHR alerts or dashbords [49].

6. Conclusion

The primary goal of this work was to investigate whether it is feasible to develop a custom EHRS for the treatment of tumors of the sella region based on a patient information model generated using openEHR. The reuse of archetypes of a repository accelerates the EHRS development and enables the application of a reuse oriented process model, which separates cyclic EHR modeling processes from cyclic EHRS development processes. The system approach is applicable for other medical use cases and showed valuable advantages in the area of structuring, standardization and semantic interoperability of health information. It became apparent that the realization of an interoperable and context free overarching EHR requires that any kind of data has to be integrated directly in the EHR, instead of using links to a distributed system environment. Beside of openEHR, W3C standards can increase semantic interoperability.

7. Multiple Choice Question

Are German hospitals constrained to archive their EHR based on EHR-standards?
A) Yes, EN 13606 is the unique authorized standard.
B) Yes, EN 14822 is the unique authorized standard.
C) Yes, and any of the following standards can be used: EN 14822, EN 13606, HL7 CDA, HL7 FHIR, openEHR.
D) No, there is no legislation on this issue yet.

Human Subject Research Approval

The development was a proof of concept project. During development, the system was tested with randomized patient data. After finishing, we tested the system with anonymized real patient data. There was no person-identifying data included on any EHR at any time.

Conflict of Interest

The authors confirm that they have no involvement in any organization or entity with any financial or non-financial interest which was discussed in this paper.

Acknowledgement

Thanks to Heinrich Herre, Jürgen Meixensberger, Alexandr Uciteli, Lars Voitel and Yixin Wu for their support and the openEHR community, especially Sebastian Garde. The authors are grateful to the anonymous reviewers for their comments, which greatly improved the quality of the paper.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

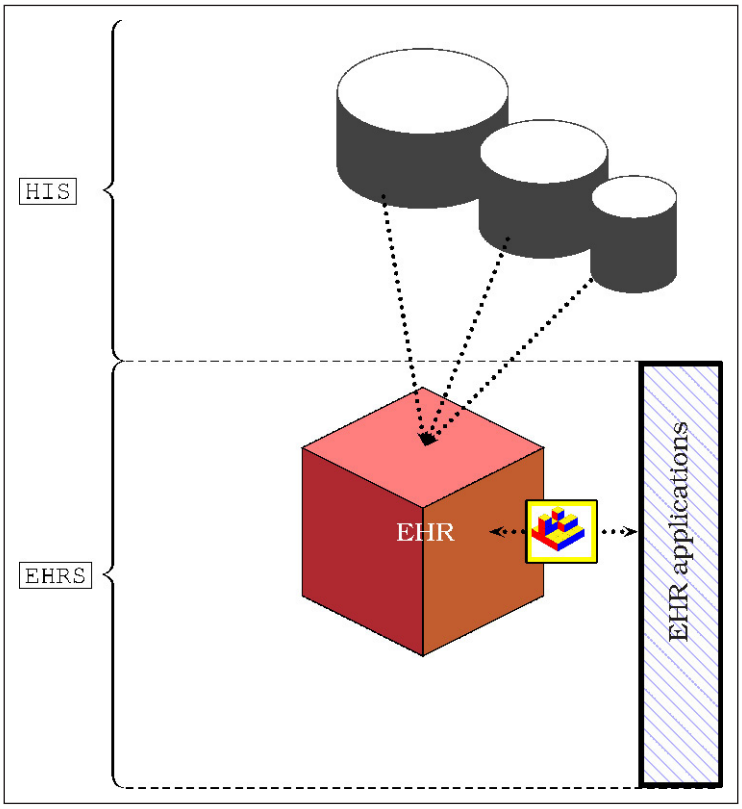


Fig. 1 Merging of distributed HIS data stored in different databases into one overarching EHR centralized EHRS. EHR applications work on the EHRs, visualize the data and exchange healthcare data standardized by an accessible interface.

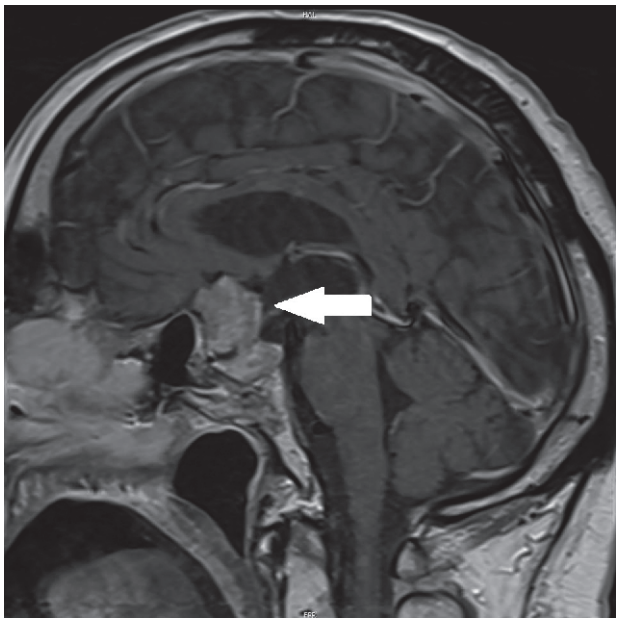


Fig. 2 MR data of patient revealing a large tumor (white arrow) in the sella region.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

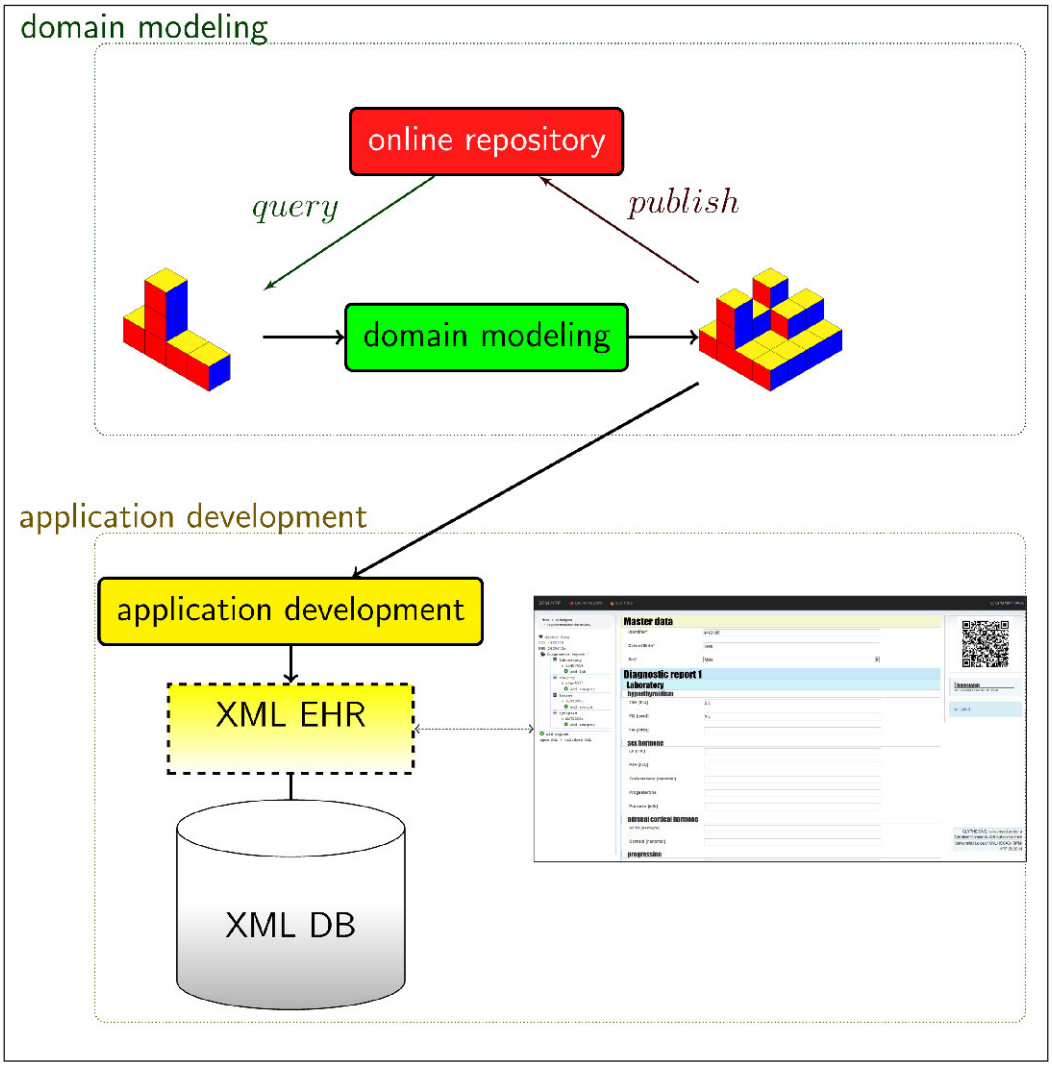


Fig. 3 Modeling as cyclic process: The upper circle starts with a query in an online repository. The small building brick is representing an archetype. During domain modeling archetypes can be developed and reassembled as templates, represented by the bigger building brick, which can be published in the repository and is used as boiler plate for application development. During application development XML EHRs are generated which are persisted on the XML DB and can be accessed by XForms.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

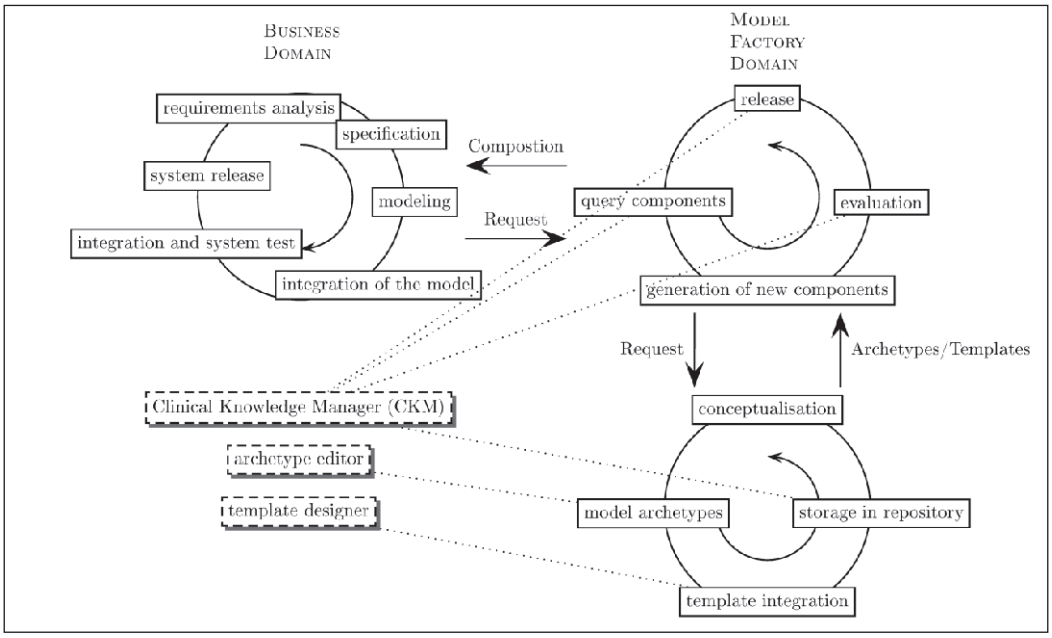


Fig. 4 Reuse-oriented process model developed by Caldiera and Basili [26] adapted to the modeling of patient data with the aid of openEHR tools. The left circle represents cyclic processes in the business domain. During the modeling process new models were requested, which triggers cyclic modeling processes in the model factory domain.

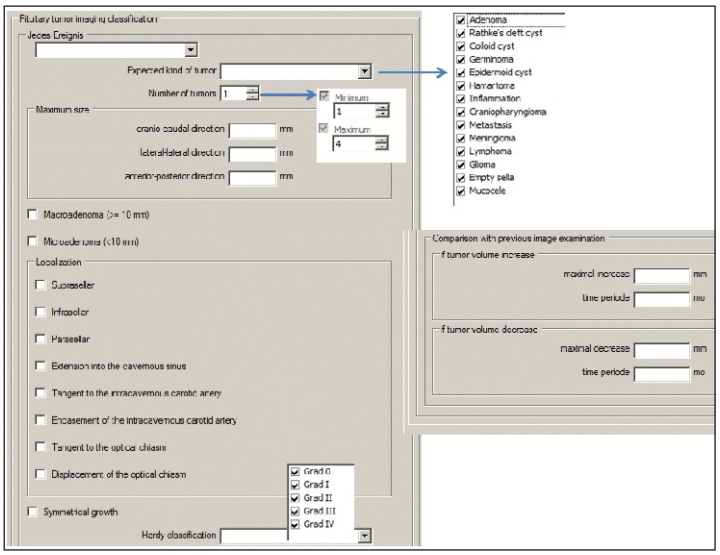


Fig. 5 The developed archetype sella region tumor imaging classification which documents the tumor classification deduced from radiological examinations (Combination of Template Designer previews).

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

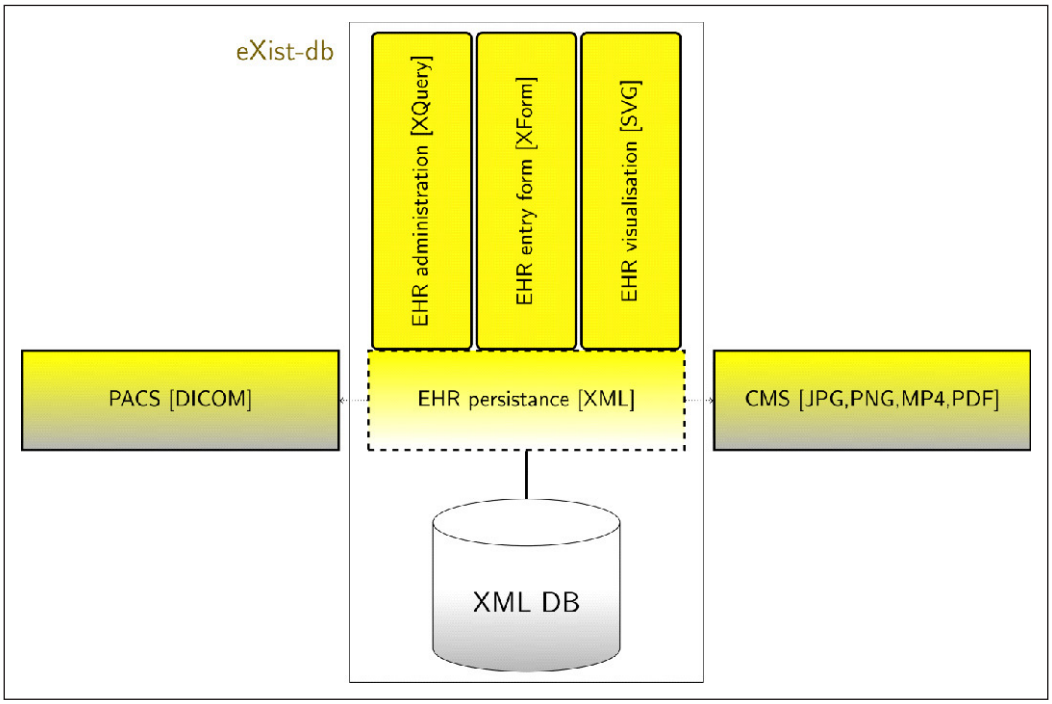


Fig. 6 Abstraction of the EHRS architecture stack. EHRs are persisted on the XML DB. On the EHRs different W3C based languages are applied for administration, data entry and visualization. Multimedia data and DICOM data was linked to the persisted EHR by URIs to external systems.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

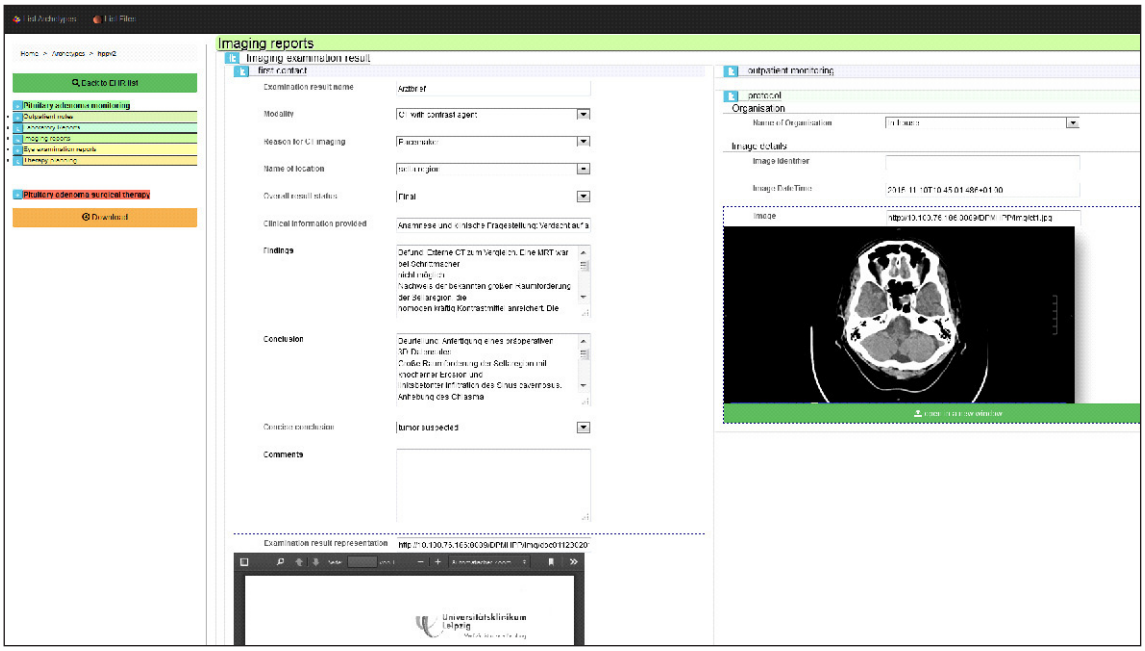
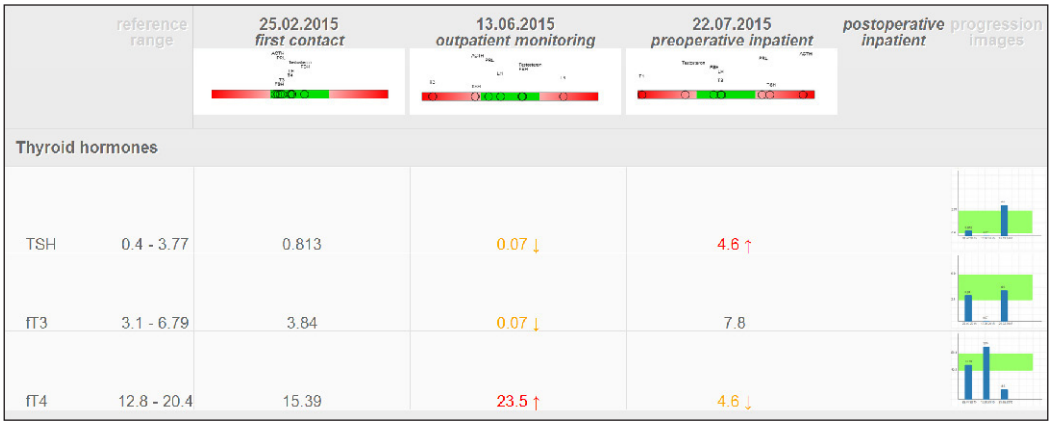


Fig. 7 Part of the entire entry form; the color in the left navigation menu changes from green during the monitoring (unfolded) into red in the therapy (folded); the entry form in the middle is based on the structure of the archetype imaging examination result. The unfolded section (first contact) describes the content of a physician letter in a structured form, containing the modality type (CT), the location (sella region), and the status of the examination (Final). In this case, a CT was necessary because the patient carries a pacemaker, mapped in the reason for CT imaging item. The German section "Anamnese und klinische Fragestellung" was mapped into the archetype item "Clinical information provided", the German section "Befund" was mapped into "Findings" and the German section "Beurteilung" was mapped into "Conclusion". The structured examination report concludes that a tumor is suspected. The original imaging examination report (PDF) was linked in the "Examination result representation" item. On the right side, the linked CT image in the protocol section is shown. Both, the original medical report (PDF) and the image data (JPG) are stored in an external CMS, which was running on the same machine, linked by URIs. A preview of these elements is provided directly in the EHR, accessible without circumstantial navigation effort.



	reference range	25.02.2015 first contact	13.06.2015 outpatient monitoring	22.07.2015 preoperative inpatient	postoperative inpatient	progression images
Thyroid hormones						
TSH	0.4 - 3.77	0.813	0.07 ↓	4.6 ↑		
fT3	3.1 - 6.79	3.84	0.07 ↓	7.8		
fT4	12.8 - 20.4	15.39	23.5 ↑	4.6 ↓		

Fig. 8 Visualization of laboratory tests. Concentrations of thyroid hormones in blood, which were measured at different time points in the clinical workflow, are depicted.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

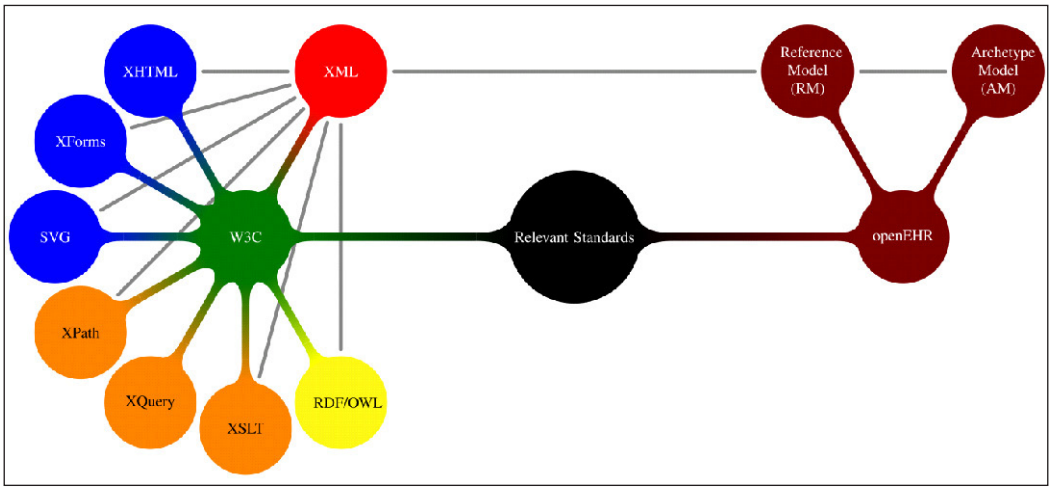


Fig. 9 Bouquet of relevant standards, blue indicates visualization, orange query and transformation, yellow ontologies and red data persistence. On the left side W3C standards are illustrated. XML plays a central role because it is interconnected to all other used W3C standards and the RIM of openEHR, because we persisted the EHRs in XML.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

Table 1 Patient data and data format involved in the surgical treatment of tumors of the sella region.

Treatment phase	Medical task	Data generated	Format
Diagnosis /monitoring	General patient examination	Outpatient notes	Text
	Blood analysis	Laboratory report	Text
	Image acquisition	MR or CT image data	DICOM
		radiological report	Text
	Eye examination	Medical report	Text
	Therapy decision	Medical report	Text
Surgical therapy	Image acquisition	Preoperative MR and CT	DICOM
		Intraoperative endoscopic videos and screenshots	MP4, JPG, PNG
		Postoperative MR or CT	DICOM
		Radiological reports	Text
	ENT examination	Medical report	Text
	Tumor resection	Operation report	Text
	histopathology of resected tissue	Histopathology report	Text

Table 2 List of reused and new developed archetypes used in the patient information modeling.

Patient data	Archetypes	openEHR	new
Master patient data	Research study		x
Outpatient note	Sella region tumor outpatient notes		x
	Multimedia Resource	x	
Laboratory report	Pituitary hormones		x
Imaging report	Imaging examination result	x	
	Anatomical location	x	
	Organisation	x	
	Sella region tumor imaging classification		x
Eye examination report	Sella region tumor eye examination report		x
Therapy planning	Procedure request	x	
ENT examination report	Sella region tumor ENT examination report		x
	Multimedia Resource	x	
Surgery documentation	Procedure undertaken	x	
	Person name	x	
Histology reports	Neurohistopathology		x
	Specimen	x	
	Physical properties of an object	x	
	Sella region tumor histopathological classification		x

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

References

1. Bisbal J, Berry D. An analysis framework for electronic health record systems. Methods Inf Med 2011; 50(2): 180-189.
2. Boochever SS. HIS/RIS/PACS integration: getting to the gold standard. Radiol Manage 2004 May-Jun; 26(3): 16–24; quiz 25–7.
3. Cohan S, Gilboa F, Shani U. PACS and electronic health records. Medical Imaging 2002; 288-298.
4. Weitschek E, Felici G, Bertolazzi P. Clinical data mining: problems, pitfalls and solutions. Database and Expert Systems Applications (DEXA), IEEE Computer Society 2013; 90-94.
5. Eichelberg M, Aden T, Riesmeier J, Dogac A, Laleci G. Electronic health record standards-a brief overview. Proceedings of the 4th IEEE International Conference on Information and Communications Technology 2006.
6. Begoyan A. An overview of interoperability standards for electronic health records. Proceedings of the 10th International Conference on Integrated Design and Process Technology, IDPT 2007, 2007; 3–8.
7. Health informatics – Health informatics profiling framework 2004 ISO/TR 17119:2004(E).
8. Operational Data Model (ODM)-XML [cited 2017 Feb 20]. Available from: http://www.cdisc.org/standards/foundational/odm
9. Health informatics – Electronic health record communication 2012. EN ISO 13606 standard family.
10. Schloeffel P, Beale T, Hayworth G, Heard S, Leslie H. The relationship between CEN 13606, HL7, and openEHR. HIC 2006 and HINZ 2006: Proceedings, 2006.
11. Health informatics – General purpose information components 2006. DIN EN 14822 standard family.
12. Dolin RH, Alschuler L, Boyer S, Beebe C, Behlen FM, Biron PV, Shabo Shvo A. HL7 Clinical Document Architecture, Release 2. J Am Med Inform Assoc 2006; 13(1): 30-39.
13. Comparison-v3 – FHIR v1.0.2 [cited 2017 Feb 22]. Available from: http://www.hl7.org/fhir/comparison-v3.html
14. Introducing HL7 FHIR [cited 2016 Dec 22]. Available from: http://www.hl7.org/implement/standards/fhir/summary.html
15. openEHR Clinical Knowlede Manager [cited 2017 Feb 21]. Available from: http://www.openehr.org/ckm/
16. openEHR Modelling Tools [cited 2017 Feb 21]. Available from: http://www.openehr.org/downloads/modellingtools
17. GasroOS [cited 2017 Feb 20]. Available from: http://sourceforge.net/projects/gastros/
18. EHRflex: An archetype-driven EHR system [cited 2017 Feb 21]. Available from: http://ehrflex.sourceforge.net/
19. LiU-IMT/EEE: A REST based Educational EHR Environment based on openEHR [cited 2017 Feb 21]. Available from: http://github.com/LiU-IMT/EEE
20. MedRecord: The fastes way to medical grade apps [cited 2017 Feb 21] Available from: http://medrecord.nl/
21. Sundvall E, Nyström M, Karlsson D, Eneling M, Chen R, Örman H. Applying representational state transfer (REST) architecture to archetype-based electronic health record systems. BMC Med Inform Decis Mak 2013. DOI:10.1186/1472–6947–13–57
22. Marco-Ruiz L, Moner D, Maldonado JA, Kolstrup N, Bellika JG. Archetype-based data warehouse environment to enable the reuse of electronic health record data. Int J Med Inform 2015; 84(9): 702-714.
23. Dziuballe P, Forster C, Breil B, Thiemann V, Fritz F, Lechtenbörger J, Vossen G, Duglas M. The single source architecture x4T to connect medical documentation and clinical research. Stud Health Technol Inform 2011; 169: 902-906.
24. Bruland P, Forster C, Breil B, Ständer S, Dugas M, Fritz F. Does single-source create an added value? Evaluating the impact of introducing x4T into the clinical routine on workflow modifications, data quality and cost-benefit. Int J Med Inform 2014; 83(12): 915-928.
25. Schloeffel P. openEHR archetypes: Putting the clinician back in the driver's seat. HIC 2003 RACGP12CC. 2003.
26. Caldiera G, and Basili VR. Identifying and qualifying reusable software components. Computer 1991; 24(2): 61–70.
27. Kropf S, Chalopin C, Denecke K. Template and model driven development of standardized electronic health records. Stud Health Technol Inform 2015; 216: 30-34.
28. Meier W. eXist: An open source native XML database. Web, Web-Services, and Database Systems 2003 169–183.
29. XQuery 3.0: An XML Query Language [cited 2016 Jun 29]. Available from: http://www.w3.org/TR/xquery-30/

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

30. RFC 3986: Uniform Resource Identifier (URI): Generic Syntax [cited 2016 Jun 27]. Available from: http://tools.ietf.org/html/rfc3986
31. openEHR Data Types Information Model [cited 2016 Jun 29]. Available from: http://www.openehr.org/releases/RM/latest/docs/data_types/data_types.html#_dv_multimedia_class
32. The official homepage of *Orthanc*, the DICOM server for medical imaging from Belgium. www.orthanc-server.com/ [cited 2016 Jun 29]
33. hGraph: Your health in one picture [cited 2016 Jun 29]. Available from: http://www.hgraph.org
34. Charts – Google Developers [cited 2016 Jun 29]. Available from: http://developers.google.com/chart
35. HTML Canvas 2D Context [cited 2016 Jun 29]. Available from: http://www.w3.org/TR/2dcontext
36. Scalable Vector Graphics (SVG) 1.1 (Second Edition) [cited 2016 Jun 29]. Available from: http://www.w3.org/TR/SVG/
37. openEHR specifications [cited 2016 Jun 29]. Available from: http://github.com/openEHR/specifications
38. openEHR Persistence FAQs [cited 2016 Jun 07]. Available from: http://openehr.atlassian.net/wiki/display/resources/Persistence+FAQs
39. Atalag K, Kingsford D, Paton C, Warren J. Putting Health Record Interoperability Standards to Work. Electron J Heal Informatics 2010; 5 (1): 1–17.
40. PopularitY of Programming Language [cited 2017 Feb 20]. Available from: http://pypl.github.io/PYPL.html
41. Kropf S, Krücken P, Mueller W, Denecke K. Structuring Legacy Pathology Reports by openEHR Archetypes to Enable Semantic Querying. Methods Inf Med 2017 Feb 28. DOI:10.3414/ME16–01–0073.
42. Kropf S, Uciteli A, Krücken P, Denecke K, Herre H. Querying standardized EHRs by a Search Ontology XML extension (SOX). Ontologies and Data in Live Sciences. Proceedings of the 7th Workshop of the GI Workgroup OBML. 2016.
43. Health and Lifesciences Extension): terms relating to healthcare, medicine and the life sciences [cited 2016 Nov 03]. Available from: http://health-lifesci.schema.org/
44. Botsis, Hartvigsen G, Chen F, Weng C. Secondary Use of EHR: Data Quality Issues and Informatics Opportunities. AMIA Jt Summits Transl Sci Proc 2010; 2010: 1–5.
45. Hovenga E, Grain H. An Information Paradigm Shift is Required to realise EHR Benefits. Stud Health Technol Inform 2015; 216: 26-29.
46. Jensen PB1, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. Nat Rev Genet 2012; 13(6): 395–405.
47. Lee EK, Atallah HY, Wright MD, Post ET, Thomas IVC, Wu DT, Haley Jr LL. Transforming hospital emergency department workflow and patient care. Journal Interfaces 2015: 45(1): 58–82.
48. Warner JL, Jain SK, Levy MA. Integrating cancer genomic data into electronic health records. Genome Med 2016; 8(1): 113.
49. Simpao AF, Ahumada LM, Gálvez JA, Rehman MA. A review of analytics and clinical informatics in health care. J Med Syst 2014; 38(4): 45.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.