EHR-ARCHE
Clinical information needs
and archetype-based electronic health records

Final report
January 2010 – December 2012

Version 1.0 from May 30th, 2012

Prof. Dr. Elske Ammenwerth
Institute of Health Informatics
UMIT – University for Health Informatics, Medical Informatics and Technology
A – 6060 Hall i.T., Austria
Mail: elsk.e.ammenwerth@umit.at

Prof. Dr. Georg Duftschmid
Center for Medical Statistics, Informatics and Intelligent Systems
Section of Medical Information Management and Imaging
MedUni Wien – Medical University of Vienna
A – 1090 Vienna, Austria
Mail: georg.duftschmid@meduniwien.ac.at

Project WebSite: http://www.meduniwien.ac.at/msi/arche/

This project was funded by FWF – Der Wissenschaftsfonds –
under grant agreement no. P21396.
Executive summary

The electronic health record that is currently being introduced in Austria is discussed controversially. Health care professionals often fear that the electronic health record will contain too much information (information overload), thus making retrieval of specific patient information, such as the history or allergies of a patient, too cumbersome or even impossible for them. The motivation of this project is to address this challenge. Our basic assumption is that the information needs of physicians can better be fulfilled when they are able to search for distinct information items in structured (archetype-based) EHR documents instead of using meta-data-based querying, which yields sets of PDF documents only.

The objectives of this project were (a) to identify the information needs of health care professionals when accessing the EHR, (b) to develop concepts to fulfill these information needs by combining document meta-data and dualmodel EHR architectures, and (c) to evaluate the developed concepts in a trial implementation.

We identified 446 information needs of health care professionals, using a combination of literature review, guideline analysis, observations of patient encounters and interviews with clinicians. Based on these data, ten clinical situations and the related information items could be described.

We then prepared an IHE-XDS based architecture that comprised, in addition to standard IHE-XDS actors such as Document Repository and Document Registry, the following actors: (a) Document Consumer Actor that allows to query and retrieve both unstructured (PDF) and structured archetype-based documents using a standard XDS document query, or to search both for individual information items and for combinations of items using a content-based query; (b) Document Crawler Actor that processes the content-based queries sent by the Document Consumer and then uses a standard XDS document query to retrieve relevant structured archetype-based to retrieve distinct information items from archetype-based documents; (c) Archetype Repository that contains archetypes corresponding to the identified information items. We tested our approach with 128 ISO/EN 13606 archetypes that represented the 446 information needs. The prototype is available at: http://ehrarchenew.dyndns.org:8080/ehrArcheConsumer.

All components were implemented and tested with two complex Diabetes mellitus patient cases containing approx. 80 documents each (both in structured and unstructured form). We conducted a controlled evaluation study, in which eight Diabetes physicians compared the standard-based query and the content-based query. During the evaluation, the participants were asked to complete pre-defined clinical scenarios that required different types of EHR data to be retrieved (e.g. search for progression of HbA1c measurements). They were asked to use the standard search and the content-based search. The evaluation results are promising: The EHR Arche Consumer was stable and performant. All participants appreciated the concept of content-based search options for specific information items (or combinations of them) based on structured documents. When using the extended search options, the participants were much more successful and much quicker in solving the given tasks, compared to a standard-based search for PDF documents.

The project has produced useful solutions for better fulfillment of information needs of physicians. While we have focused on the setting “diabetes mellitus treatment”, the overall approach and architecture are generalizable to other clinical settings. All applications are developed in a way that they can handle any list of information items and queries, and any list of related archetypes.
# Table of contents

1. Background and objective of the project ................................................................. 1  
2. Project duration and team members ....................................................................... 1  
3. Working plan and organization of the project ......................................................... 2  
4. Results of the project ............................................................................................... 3  
4.1 Information needs of clinicians when treating Diabetes patients (WP1) ................. 3  
   4.1.1 Methods ........................................................................................................... 3  
   4.1.2 Results ........................................................................................................... 3  
   4.1.3 Project documentation ..................................................................................... 5  
4.2 Technical preparation and EHRArche concept (WP2.1, WP2.2) ............................ 6  
   4.2.1 Methods ........................................................................................................ 6  
   4.2.2 Results ........................................................................................................... 6  
   4.2.3 Project documentation ..................................................................................... 8  
4.3 Preparation of archetypes (WP 2.3) ....................................................................... 9  
   4.3.1 Methods ........................................................................................................ 9  
   4.3.2 Results ........................................................................................................... 10  
   4.3.3 Project documentation .................................................................................... 10  
4.4 Development of the EHRArche Document Crawler (WP 3) .................................... 11  
   4.4.1 Methods ....................................................................................................... 11  
   4.4.2 Results ........................................................................................................... 11  
   4.4.3 Project documentation .................................................................................... 12  
4.5 Development of EHRArche Extended Document Consumer (WP 4) ..................... 13  
   4.5.1 Methods ....................................................................................................... 13  
   4.5.2 Results ........................................................................................................... 13  
   4.5.3 Project documentation .................................................................................... 20  
4.6 Evaluation and Demonstration (WP5) .................................................................... 20  
   4.6.1 Methods ....................................................................................................... 20  
   4.6.2 Results ........................................................................................................... 23  
   4.6.3 Project documentation .................................................................................... 23  
5. Discussion and Conclusion ....................................................................................... 24  
5.1 What has been achieved? ....................................................................................... 24  
5.2 What are further research questions? ..................................................................... 24  
6. Publications of the project team .............................................................................. 25  
7. References .............................................................................................................. 26
1 Background and objective of the project

Today’s health care professionals can access an ever increasing amount of patient-related information and clinical knowledge, one of the most-promising applications being the Electronic Health Record (EHR). While several clinical and economic benefits are expected by the EHR, it may also lead to information overload, as the relevant information item that an EHR user searches for may be hidden in vast amounts of information within the lifelong EHR.

Therefore, we see an urgent need for supporting EHR users in selectively retrieving information that is relevant in their respective search context. Besides document meta-data, archetype-based dual-model EHR architectures propose a corresponding support, but it is unclear whether they fulfill this promise. Overall, it is unclear what the information needs of health care professionals are and how to support them adequately by dual-model EHR architectures.

The objectives of this project were therefore

1. to identify the information needs of health care professionals when accessing the EHR, considering the respective search context;
2. to develop concepts to fulfill these information needs by combining document meta-data and dual model EHR architectures;
3. to evaluate the developed concepts in a trial implementation.

2 Project duration and team members

The project started on January 1st, 2010 and will run until December 31st, 2012.

It was conducted as a cooperation project between the Institute of Health Informatics at UMIT in Hall in Tyrol and the Section of Medical Information and Retrieval Systems of the Medical University of Vienna.

The team members were:

- Elske Ammenwerth, PhD, UMIT, project manager
- Georg Duftschmid, PhD, MUW, vice project manager
- Samrend Saboor, PhD, UMIT, responsible for WP2 (technical preparation and concept) and WP4 (Extended Document Consumer)
- Gudrun Hübner-Bloder, PhD, UMIT, responsible for WP1 (information needs analysis) and for WP5 (evaluation and demonstration)
- Christoph Rinner, MSc, MUW, responsible for WP3 (document crawler)
- Michael Kohler, MSc, MUW
3 Working plan and organization of the project

Figure 1 shows the original time planning of the project. Overall, the project followed well this time plan, and despite some minor delays in some tasks, the overall duration of the work packages was as planned.

During the project, we conducted six face-to-face project meetings of all team members and ten telephone conferences of all team members. In addition, numerous face-to-face and telephone meetings of smaller groups took place to work on the specified working packages.

To support team communication, a SVN server was prepared and used. It served the project team as a central repository for all documents that were of relevance for the project (e.g. specification documents) and managed the different versions of those documents that were created during the collaborative work.

The joint software development was organized as follows: The software development team consisted of members of both project partners. The development team held weekly face-to-face meeting (at least one per week) in order to tightly coordinate the development efforts. Further, well established development tools and methods were agreed on and used (e.g., Eclipse IDE, Java). A main prerequisite for the success of the collaborative work was the usage of a dedicated SVN server that stored and managed all software sources.

For communication of the project with the public, a website ([http://www.meduniwien.ac.at/msi/arche/](http://www.meduniwien.ac.at/msi/arche/)) was created and regularly maintained.
4 Results of the project

4.1 Information needs of clinicians when treating Diabetes patients (WP1)

The goal of this WP was to identify clinical information needs of physicians during the treatment of Diabetes Mellitus (DM) patients.

4.1.1 Methods

We used a triangulation of methods (1) by combining literature search, interviews, observations and documentation analysis.

Literature research: We analyzed five international evidence-based Diabetes Mellitus (DM) guidelines for clinical diagnostics and medical treatment of DM (2-6) to identify information items needed during DM treatment.

Expert Interviews: We performed oral, partial-standardized expert interviews. The general objective of expert interviews is the investigation of function-specific know-how(9). We conducted these expert interviews with six internists with specialization DM in the Diabetes outpatient clinics of the University Hospital of Innsbruck, the Regional Hospital of Hall in Tyrol, the Hospital St. Vinzenz in Zams and with one internal physician in private practice. To analyze the results we used qualitative content analysis (7).

Observation: Additional to the expert interviews, we decided to make unstructured participant observations of clinical encounters (10), to gain additional insight and to validate information from the interviews. The observation of 22 DM patient encounters took place in the DM outpatient clinic of the internal medicine of the university hospital of Innsbruck. We decided to analyze the data based on grounded theory (11-12). This analysis allows inductive concept and theory development during the data collection.

Clinical documentation analysis: We analyzed the clinical information documented in the electronic records of three Diabetes outpatient clinics of internal medicine (University Hospital of Innsbruck, Regional Hospital of Hall in Tyrol and the Vienna General Hospital). For analysis of these recorded data we used qualitative content analysis.

All sources were analysed using qualitative content analysis with inductive creation of categories (7), using MAXQDA 2007) (8) as software tool. The analysis started with the DM guidelines and was then extended with the information items from the expert interviews, the observations, and the clinical documentation analysis.

4.1.2 Results

Information items: An information item reflects a clinical information need. We identified 446 distinct information items, for example, type of DM classification (Typ 1 DM, Typ 2 DM, gestational diabetes etc.), onset of DM, weight-height status like body mass index, weight gain, or weight loss. These information items were structured in the following nine main categories and corresponding subcategories:

I. Patient master data
II. Self-monitoring of the patient
III. Diabetes mellitus classification
IV. General medical history
V. Diagnosis
VI. Recent surgery  
VII. Recent check-ups  
VIII. Laboratory findings  
IX. Medication/Therapy

Table 1 shows an example of identified information items.

<table>
<thead>
<tr>
<th>Time windows:</th>
<th>Typical time windows</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.</td>
<td>0-3 months</td>
</tr>
<tr>
<td>II.</td>
<td>0-6 months</td>
</tr>
<tr>
<td>III.</td>
<td>0-12 months</td>
</tr>
<tr>
<td>IV.</td>
<td>0-36 months</td>
</tr>
<tr>
<td>V.</td>
<td>0-60 months</td>
</tr>
<tr>
<td>VI.</td>
<td>all available data</td>
</tr>
</tbody>
</table>

**Clinical situations and brief queries:** Both the expert interviews and also the observations showed that information needs are different according to the clinical situations. Overall, we identified ten most important clinical situations and the related information items (compare Table 2):

1. Initial clinical interview  
2. Routine check - brief data set  
3. Routine check - extended data set (exemplary)  
   - Routine check by patients with lipometabolism problems  
   - Routine check by patients with cardiovascular problems  
   - Routine check by patients with glycaemic incidents  
   - Routine check by patients with Neuropathy  
   - Routine check by patients with Nephropathy  
   - Routine check by patients with Gestation diabetes  
   - Routine check by patients with ophthalmological problems (e.g. retinopathy)  
   - Routine check by patients with dermatological problems (e.g. diabetic foot)

Table 2: Overview of identified clinical situations and the number of corresponding information items.
Table 3 presents the clinical situation “routine control – short dataset” in more detail.

<table>
<thead>
<tr>
<th>Kategorie</th>
<th>Unterkategorie</th>
<th>Information Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patientenstammstaten</td>
<td>Name, Geburtsdatum etc.</td>
<td>Allergieanamnese</td>
</tr>
<tr>
<td>Diabetes Klassifikation</td>
<td>Diabetes Klassifikation</td>
<td>DM Erstdiagnose</td>
</tr>
<tr>
<td>Selbstmessungen des Patienten</td>
<td>Blutdruck</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Blutglukose nüchtern</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Blutglukose pp (2h)</td>
</tr>
<tr>
<td>Allg. Anamneseparameter</td>
<td>Gewicht und Größe-Status</td>
<td>Größe</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Körpergewicht</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kommentar (zu Körpergewicht)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BMI</td>
</tr>
<tr>
<td>Vitalparameter</td>
<td>Herzfrequenz/Pulssrate</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Blutdruck</td>
</tr>
<tr>
<td>Lebensstil</td>
<td>Rauchkonsum</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Alkoholkonsum</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Essgewohnheiten &amp; Sport</td>
</tr>
<tr>
<td>Laborparameter</td>
<td>Blutbild</td>
<td>HbA1c</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BG nüchtern</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BG postprandial</td>
</tr>
<tr>
<td>Harnstatus</td>
<td>Glucosurie</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mikroalbuminurie</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Makroalbuminurie</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Leukozyten</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nitrit</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Aceton</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Urobilinogen</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bilirubin</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Erythrozyten</td>
</tr>
<tr>
<td>Therapie/Medikation</td>
<td>Therapie/Medikation</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Clinical situation “routine control – short dataset”, showing the related information items.

In addition, we identified 68 additional brief queries that consist of queries that give additional short information such as general anamnesis items, short list of lab results, glucose values (see Table 4) or earlier surgical treatments.

<table>
<thead>
<tr>
<th>Fasting plasma glucose (FPG)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Postprandial plasma glucose (PG pp)</td>
</tr>
<tr>
<td>HbA1c</td>
</tr>
<tr>
<td>Oral glucose tolerance test (OGTT)</td>
</tr>
</tbody>
</table>

Table 4: Example of the brief query on “glucose status”, comprising four information items.

4.1.3 Project documentation

More detailed documentation is available in the following project documents:

4.2 Technical preparation and EHRArche concept (WP2.1, WP2.2)

The objective of these WPs was to prepare the general IHE-XDS-based architecture based on the SENSE solution of the company ITH-icoserve and to develop a concept for the overall EHR Arche architecture.

4.2.1 Methods

Based on an analysis of IHE-XDS and of the ITH-icoserve sense architecture used for EHR systems, several workshops of the team were conducted to develop an overall general EHR Arche architecture.

4.2.2 Results

The technical architecture concept describes the overall technical architecture, the various use cases of document query and result presentation, and mock-ups of the planned user interface.

The general architecture was based on IHE-XDS and used the SENSE solution of the company ITH-icoserve. SENSE (http://www.ith-icoserve.com/loesungen/healthnet/uebersicht.html) was chosen because it is the basis for the Tyrolean electronic health record system and because it follows the Austrian standard for electronic health record systems. Additionally, the Document Consumer Adapter offered by SENSE enabled us to easily integrate the IHE-XDS queries with the EHR Arche components (“Document Consumer Adaptor”, see Figure 2).

Figure 2 depicts an overview of the standard IHE XDS actors (white colored) as well as those actors/parts that had to be developed for achieving the goals of the EHR Arche project (blue colored). Figure 2 also shows how all actors interact.

The standard XDS actors such as the Patient Identity Source, Document Registry and Document Repository are based on the sense solution of the company ITH-icoserve GmbH. These actors were available as web services.

Using the Document Consumer Adapter from SENSE, the Document Consumer Actor and Document Source Actor (both standard XDS actors) were developed and implemented completely from scratch by the project team.

Document Crawler and the Archetype Repository were completely newly developed actors.
Document Registry Actor

The Document Registry Actor is a standard actor of the IHE XDS.b profile. According to this profile (please see the IHE Technical Framework Vol. 1, p.71), the Document Registry Actor has the task to manage a standardized set of metadata for each registered medical document. Amongst others, this set of metadata contains a link to the actual document that is stored by the Document Repository Actor. When being queried by a Document Consumer Actor, the Document Registry Actor returns this link and also the metadata for those documents that match the query criteria.

Document Repository Actor

The Document Repository Actor is a standard actor of the IHE XDS.b profile. According to this profile (please see the IHE Technical Framework Vol. 1, p.71), the main task of this actor is to persistently and securely store medical documents. In case of new documents, the Document Repository Actor contacts the Document Registry Actor to register these new documents appropriately. Here, the Repository Actor transmits – amongst other standardized details – a specific link with which the respective document can be retrieved.

Document Consumer Actor

The Document Consumer Actor is a standard actor of the IHE XDS.b profile. According to this profile (please see the IHE Technical Framework Vol. 1, p.70), the Document Consumer Actor provides possibilities for querying and retrieving medical documents from a trans-institutional EHR. To fulfill this task, the Document Consumer Actor offers a graphical user interface (GUI) that offers search criteria for querying documents. Furthermore, the Document Consumer Actor presents the search results and allows the user to select a document for retrieval. The selected document is then retrieved from the Document Repository Actor and displayed by the Document Consumer Actor.
This basic functionality and user interface design of The Document Consumer had to be extended to also allow content-based document search. Thus, the “extended” Document Consumer Actor developed in this project offers a content-related document query, based on single information items or combinations of those items. These content-related queries cannot be answered by a standard XDS Document Registry Actor. Instead, the query is thus sent to the newly developed Document Crawler Actor. This actor analyzes the content-related query, coordinates all necessary communication with the standard XDS actors and sends a prepared result set back to the Extended Document Consumer Actor.

Details of the Document Consumer Actor are presented in chapter 4.5.

**Document Crawler Actor**

The Document Crawler Actor is a specialized non-standard actor. Its main function is to support and execute content-based document queries. The Document Crawler receives queries by the Extended Document Consumer Actor and decides whether it has to be answered using a standard XDS document query (i.e., using standard metadata as filtering criteria), or whether an extended content-based query (i.e. using archetype-based information items) is required. In the latter case, the Document Crawler Actor determines what standard XDS-queries can first be applied in order to find medical documents that potentially contain the relevant content. It then retrieves these candidate documents and analyzes their content. After this, the Document Crawler Actor returns the links, the metadata and the relevant content-related details to the Document Consumer Actor.

Details on the Document Crawler are presented in chapter 4.4.

**Archetype Repository**

The Archetype Repository contains all information entities that are contained in archetype-based medical documents. Within the Archetype Repository, these information entities are represented as a hierarchical tree independently of the underlying Reference Model. For this project, we used ISO/EN 13606 Archetypes. Thus the tree consists of information items that can be instances of the ISO/EN 13606 Reference Model classes *Element* (when the items cannot be further separated into other more elementary items), *Cluster* (that groups *Element*-objects), *Entry* (that groups *Cluster*-objects), or *Composition* (that groups *Entry*-objects). The information items are linked to the corresponding information entities.

As part of its initialization procedure, the Extended Document Consumer retrieves all available information items from the Archetype Repository. The retrieved information items can then be either used to support simple queries that contain only one information item or more complex queries that combine multiple information items.

**4.2.3 Project documentation**

More detailed documentation is available in the following project documents:

- Rinner C, Kohler M, Saboor S, Duftschmid G. WP2.2 Schnittstellenkonzept. Version 0.6 from July 14th, 2010.
4.3 Preparation of archetypes (WP 2.3)

The objective of this WP was to map the information items identified in WP1 to existing archetypes, and where necessary to develop new archetypes. In addition, this WP was responsible for preparing both unstructured and structured test documents.

4.3.1 Methods

Based on the required clinical information items existing archetypes were searched. For ISO/EN 13606 no archetypes are publicly available, hence we could only rely on openEHR archetypes. Available resources included the openEHR Clinical Knowledge Manager\(^1\), the openEHR SVN repository\(^2\) and the NHS SVN repository\(^3\).

The archetypes were created using the linkEHR archetype editor (14) and translated to German and English.

The initial planning was based on the identified 446 clinical information items (see chapter 4.1). These information items were structured hierarchically, i.e. higher level groupings of related finer-grained items were already identified in this step. Using the results of the planning step, archetypes were created and adapted.

Originally we planned to convert the existing openEHR reference archetypes into ISO/EN 13606 archetypes using (13) and specialize these archetypes according to our requirements. However, we finally decided to refrain from this idea. The main reason was that the openEHR approach includes the concept of templates, which allows existing archetypes to be further constrained and thus customized for local purposes. As a result openEHR archetypes are designed to include the maximum set of potentially relevant items for a particular medical concept. The ISO/EN 13606 standard however does not support the template mechanism. Therefore we would have had to reduce the typically extensive openEHR archetypes to those few items relevant for our context by specializing the converted openEHR archetype and setting the cardinality of most items to zero. As this seemed as a laborious and also not really convincing approach, we decided to create our own set of archetypes and replicate those structures of existing archetypes that were relevant in our context. Still, the openEHR archetypes found were a valuable source during the archetype creation and complemented our clinical information items.

From the identified/developed archetypes, electronic case report forms (eCRFs) were automatically derived that were used to test the archetypes and create test EHR documents. Based on the evaluation of the results, the required archetype changes were planned and the cycle continued until the archetypes were finally released. The creation of the test documents and the evaluation using the eCRFs was done by a medical domain expert located in Hall/Tyrol. The results were documented using a text file. Computer scientists in Vienna created and adapted the archetypes for the next iteration cycle based on this information.

---

\(^1\) [http://openehr.org/knowledge/](http://openehr.org/knowledge/)

\(^2\) [http://www.openehr.org/svn/knowledge/archetypes/](http://www.openehr.org/svn/knowledge/archetypes/)

\(^3\) [https://svn.connectingforhealth.nhs.uk/svn/public/nhscontentmodels/TRUNK/cm/archetypes/adl/openehr/ehr/](https://svn.connectingforhealth.nhs.uk/svn/public/nhscontentmodels/TRUNK/cm/archetypes/adl/openehr/ehr/)
4.3.2 Results

For roughly one quarter of the clinical information items existing openEHR archetypes were found.

We developed an approach for an iterative design of archetypes which was tested with 128 ISO/EN 13606 archetypes that represent 446 information needs. In the design of our archetype set, we intensively used the Archetype Definition Language (ADL) slot and specialization mechanisms to allow a high level of Archetype reuse. In Figure 3 this is illustrated for the “family history” archetype. At the average each archetype underwent 2 – 3 development cycles.

![Figure 3: Example of “Family History” archetype. This archetype is included in four COMPOSITION archetypes via slots and includes four ENTRY archetypes via slots itself (black line and dark arrow). Two of these ENTRY archetypes are specializations of other ENTRY archetypes (dashed line and white arrow).]

4.3.3 Project documentation

More detailed documentation is available in the following project documents:

- Kohler M: Relevant existing archetypes for information needs. Version 0.1 from November 11st, 2010.
- Kohler M, Rinner C, Duftschmid G, Concept for mapping information needs to meta-data/archetypes. Version 0.5 from January 17th, 2011
4.4 Development of the EHRArche Document Crawler (WP3)

The objective of this WP was to specify, develop and test the EHR Arche document crawler. The aim of the Document Crawler is to support and execute content-based document queries in fully structured EHR documents conforming to the ISO/EN 13606 Reference Model and described by Archetypes.

4.4.1 Methods

The document crawler was implemented as a web-service in Java.

4.4.2 Results

The Document Crawler receives XML-based query containing the requested information items by the Extended Document Consumer Actor and decides whether it can be answered using a standard XDS document query (i.e., using standard metadata as filtering criteria), or whether an additional extended content-based query (i.e. using archetype-based information items) is required.

The Document Crawler Actor determines what standard XDS-queries can first be applied in order to find medical documents that potentially contain the relevant content. Using the Archetypes, the relevant “class codes”, “practice setting code”, the temporal information, the patient identification and the author identification, an IHE-XDS metadata-based search is conducted, and candidate documents are retrieved via Document Registry and Document Repository. Each condition in the query from the Extended Document Consumer is then translated to an XQuery expression using the structured information from the archetypes. By applying the XQuery to the retrieved candidate documents from the metadata-based search, the relevant information is extracted. If the query can be answered using just metadata, the whole document is selected with an XQuery. All XQuery results are combined to a single result. This result is further transformed and structured.

After this, the Document Crawler Actor returns the links, the metadata and– the relevant content-related details to the Document Consumer Actor.

Figure 4 gives an overview of the tasks of the Document Crawler.
Figure 4: Tasks of the document Crawler: A query representing the information needs is sent from the Extended Document Consumer to the document crawler. Using the query, a search in the IHE XDS registry is dispatched and an XQuery to extract the information needs from the documents is created. In the next step the relevant documents are fetched from the IHE XDS repository. In the final step the XQuery is applied to the documents. The result is sent back to the Extended Document Consumer and is visualized there.

4.4.3 Project documentation

More detailed documentation is available in the following project documents:

- Kohler M, Rinner G, Duftschmid G: Concept for mapping information needs to metadata/archetypes. Version 0.5 from January 17th, 2011.
- Rinner C, Kohler M, Saboor S, Duftschmid G: Schnittstellenkonzept. Version 0.6 from June 22nd 2010
4.5 Development of EHRArche Extended Document Consumer (WP4)

The objective of this WP was to specify, develop and test the Extended Document Consumer. The aim of the Document Consumer is to provide the user interface for querying and retrieving information in both structured as well as in unstructured documents.

The prototype is available for testing at http://ehrarchenew.dyndns.org:8080/ehrArcheConsumer.

4.5.1 Methods

The specification of the Extended Document Consumer was developed in various workshops. For the implementation, mainly the Ajax-based Framework ZKoss as well as standard JAVA EE was used.

4.5.2 Results

The Extended Document Consumer Actor provides all possibilities of the standard IHE XDS Document Consumer Actor and further offers the possibility to execute an archetype-based content-based search for information items.

The basic structure for the extended content-based document search is shown in Figure 5. The Front-End contains the graphical user interface (GUI) for definition of queries. The Back-End contains all functions that are required to process the queries and the resulting responses.

![Figure 5: Overview of the Extended Document Consumer Actor’s architecture with a focus on the content-based document search. It distinguishes the actual graphical user interface (in the Front-End) from the communication mechanisms that handle the required XML messages (in the Back-End).](image)

**Interrelationships of the search query categories in the Front-End**

The GUI comprises two main parts: The creation of metadata-based queries according to the standard XDS profile, and the creation of content-based queries. The latter further distinguishes two different categories of Archetype-based search queries: the first comprises queries that only contain a single Information Item (i.e., a single Element-object with a specified value or a Cluster-, Entry- or a Composition-object), while the second comprises combinations of the single information items.
**Interrelationships of the Back-End parts**

In order to achieve a flexible and re-usable solution, we implemented the extended content-based search as a “construction kit”. This construction kit is the basis for different Archetype-based search queries and provides easy-to-use possibilities for specifying search queries by combining single information items.

Each of these queries can be initialized individually. Here, the values for this initialization either are entered by the user (in case of single information search queries) or read from a persistent description (in case of the more complex search queries). In principle, the construction kit provides all possibilities to also implement a graphical query editor with which complex queries could be compiled at run time.

These query objects can be stored persistently in the Extended Document Consumer Actor (hard coded as part of the source code) or stored locally in proprietary files. During the initialization procedure of the Extended Document Consumer, these stored descriptions are either read from the proprietary files (i.e. de-serialization) or loaded from within the Consumer and are used to provide the content for the GUI (i.e., combo boxes that contain the single information items or lists of the predefined queries).

When the user decides to use a content-based query, the Document Consumer Actor has to forward the request to the Document Crawler Actor. For this purpose, an own XML structure was developed that allows the structured and scalable transmission of all required details. Thus, the Java-based search query objects are used to create adequate XML-messages that contain the main query. These XML-messages are sent to the Document Crawler Actor. After the Document Crawler queried, retrieved and analyzed all possibly relevant documents, the results are compiled in a response XML-message that is sent back to the Document Consumer Actor. The received response XML is then parsed in the Back-End and translated into the abstract Java-query objects which are easier to handle. The result content is displayed in an own result section in the GUI.

**Description of the main use cases**

Figure 6 shows all use cases that were anticipated for the Extended Document Consumer Actor. Those use cases that have been implemented in a fully functional prototype are then described in more detail. First, the initial cross-institutional patient identification procedure is described; then, the standard IHE metadata-based document search and the presentation of the results; finally, the content-based document query, together with the aggregated presentation of all content-based results.
Unique patient identification

The user has first to specify a patient. The IHE proposes the profile Patient Identifier Cross-Referencing (PIX) for the unique identification of a patient in a cross-institutional context. It allows a Document Consumer Actor to ensure the identity of a specific patient. Figure 7 shows how the user can fill in search criteria for a patient. For instance, the social security number is used together with the given and family name to query a patient of interest (in all cases a unique organization identifier must be provided as well in order to allow for a query at the specific site). The implemented prototype provides a preset of patient details (see “Registrierte Patienten” in Figure 7) in order facilitate the whole process (e.g., prevent typing errors). After the user selected one of the presets or filled in new criteria for the patient, the search a PIX-search is performed. When an appropriate patient is found the whole group box (see in the figure “Angaben zum Patienten”) is closed and retrieved details regarding the respective patient are displayed (shown in the next screenshots).
Standard IHE document search

After the PIX-search resulted in a unique patient identification, the user can to proceed to the actual document search. Figure 8 shows a tabbox that contains all GUI elements that allow a standard IHE XDS document search. The user can choose whether to perform a very sensitive query (i.e., performing a query without any filtering criteria) or a more specific query by using the filtering criteria that is based on the available IHE-XDS metadata. The GUI provides common filter criteria that are grouped according to their content. Thus, the user can narrow the search using chronological criteria (grouped in “Zeitliche Angaben”) or criteria that deal with the document type (grouped in “Dokumenteneigenschaften”) or the facility type (grouped in “Medizinische Organisation”).

Figure 8: Part of the GUI that shows the standard IHE XDS-document query. It provides common metadata-based search criteria. There are chronological search criteria (i.e., “Zeitliche Angaben”), criteria that are related to the document type (i.e., “Dokumenteninformation”) and those criteria that deal with the medical facility type (“Medizinische Organisation”)

Once the Document Registry Actor processed the query request of the Extended Document Consumer it sends back a list of relevant document entries (i.e., not the actual documents but relevant details such as a unique id that the Consumer Actor can use in order to retrieve a document). The Document Consumer shows these results as a list of all documents that are identified as relevant by the Document Registry. This result list shows the creation date and the document title for every found document. Additionally, a button at the end of every row allows for the user to retrieve the respective document and automatically display it in a separate popup-window (see Figure 9).
Figure 9: The result list contains the creation date and the title of every document entry that was returned by the Document Registry Actor. The button at the end of every row allows the user to retrieve and automatically display the respective document.

When the user uses the button at the end of a document entry row in the result list the Document Consumer Actor sends an appropriate request to the Document Repository Actor. When the document is found and returned by the requested Document Repository it is displayed in a separate popup-window once the Extended Document Consumer Actor received it completely (see Figure 10).

Figure 10: The Document Consumer displays a document that was returned by the Document Repository Actor in a separate popup-window.
Content-based document search

Figure 11 shows the definition of content-based queries in the extended Document Consumer. The user can choose between two search modes: The simple content-based query that only contains one single Information Item (without any further value restrictions), or the predefined queries that are combinations of several Information Items (also including value restrictions where necessary and useful from a medical point of view). In either of the both extended search modes, the user can additionally specify a time frame.

![Figure 11: Frontend of the EHR Arche extended search interface: The user can choose whether to search for clinical information on the basis of single information items (1). Any of the 446 identified information items (e.g., “medication”, “smoking”, “HbA1c” or “Parathormone”) can be queried individually. Alternatively, pre-defined queries can be executed that search for a combination of information items (2). For example, “Erstgespräch” (= first encounter) includes around 200 information items that are of interest during a first encounter with a diabetes patient, while “Routinekontrolle” (= routine check) contains 42 information items. Both kinds of searches further allow to constrain the timeframe in which the documents were created, such as “letzten 6 Monate” (= last six months).

Presentation of the results of content-based search

Results of a content-based search are presented by the Document Consumer presents in a table-based way (Figure 12). Every row represents the aforementioned hierarchy of Composition (highest level that contains all further details), Entry, Cluster and Element (most basic level that contains a concrete value). If the user wants to see the whole document he/she can click on the small symbol under a value entry (Figure 12: symbol showing a small folder with a sheet of paper) and then gets the full document.
Figure 13: Here, the viewer directly jumps to the value of interest (typically the value under which the button is located that the user used to call the document).

Figure 12: Frontend of the extended result presentation: On the left side, different rows represent different information items (e.g., 1a = regular defecation, 1b = diarrhea), which are organized in logically connected groups (e.g., 2 = defecation). On the right side, each column represents a single document. For each document, the value of the specific information item is presented (e.g., 3 = yes). It is also possible to retrieve and display the complete document together with the contained information items by clicking on the document name and date (e.g., 4 = discharge letter from Sept. 13th, 2011).
The objective of this WP was to evaluate the feasibility and usefulness of the EHR Arche concept in a simulation study with several clinicians.

4.6 Evaluation and Demonstration (WP5)

The objective of this WP was to evaluate the feasibility and usefulness of the EHR Arche concept in a simulation study with several clinicians.

4.6.1 Methods

We prepared test cases of two Diabetes mellitus patients, Mr. Adamec and Mrs. Buchsbaum. These test cases were based on realistic data and represented patients with a long-year history of Diabetes with related numerous complications and therapeutical interventions. For each patient, around 80 documents were prepared such as discharge letters (e.g. discharge letters of internist, neurologist, dermatologist, general practitioner etc.), lab results, patient self monitoring documents etc. Each test document was identically made available both in unstructured form (PDF document) as well as in highly-structured archetype-based document (ISO/EN 13606). The ISO/EN 13606 documents were based on the archetypes identified in WP2. All documents were made available in the EHR Arche application.

We then prepared two scenarios describing typical contact between a physician and each of the two test patients. The scenarios described routine checks where the patients reported on problems with e.g. insulin dosing etc. Each scenario also comprised a list of related clinical questions that the physician might need to respond. For each clinical question, the best answer based on the available EHR data was defined as gold standard (compare Table 5).
**Klinische Fragen zur Patienten-Fallstudie 1**

<table>
<thead>
<tr>
<th>Frage</th>
<th>Goldstandard</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1F2</td>
<td>17</td>
</tr>
</tbody>
</table>
| Sie wollen sich einen schnellen Überblick über die letzten 6 Monate des Patienten Adamec verschaffen: | DM Klassifikation DM Typ 2  
| - Diabetes Klassifikation | DM Erstdiagnose: 01-02-1995  
| - DM Erstdiagnose | Glukosestatus: (2 HbA1c, 2BG nü)  
| - Glukosestatus der letzten 6 Monate | Medikation  
| - Medikation aktuell | NovoMix 30 15/10/00  
| | Insulatard 00/00/00/12 (14)  
| | Aegopton 30 1-0-0  
| | Thyrax 0.1 3/4-0-0  
| | Tritazide 25 1-0-0  
| | Crestor 10 0-0-1  
| | Plavix 1-0-0  
| | Herz ASS 1-0-0 nach dem Essen  
| | Gastrosil Tr.10-10-10; Levitra 10ng; Lamisil Creme 2 x täglich |
| P1F3  | 4            |
| Wie verhält sich der HbA1c im Verlauf des letzten Jahres, da im heutigen Labor der HbA1c Wert sehr hoch war. | Anzahl der HbA1c Werte:  
| | Arztbrief/Laborbefund 13.09.11 HbA1c 10.0 %  
| | Arztbrief/Laborbefund 03.06.11 HbA1c 9.2 %  
| | Arztbrief/Laborbefund 09.03.11 HbA1c 11.2 %  
| | Arztbrief/Laborbefund 28.12.10 HbA1c 10.8 % |
| P1F4  | 1            |
| Sie möchten bei Hr. Adamec unter Umständen von Tritace auf ein anderes Hypertensivum wechseln. Der Patient weiß, dass er ein bestimmtes Medikament schlecht vertragen hat, kann sich aber nicht an den Namen erinnern. Welches Medikament war das? | Cenipres – Arztbrief vom 15-03-10  
| | NW: Cenipres wurde abgesetzt, da der Patient über Palpitationen und starken thorakalem Druckgefühl geklagt hat. |

We then recruited eight internists with specialization DM in the Diabetes outpatient clinics of the Vienna General Hospital, University Hospital of Innsbruck, the Regional Hospital of Hall in Tyrol, the Regional Hospital of Natters in Tyrol, the Hospital St. Vinzenz in Zams and internal physicians in private practice.

**Schedule of the evaluation:** First of all we gave a short introduction of the study and their goals. Then each physician got an individual introduction to the EHR consumer. After that the physician got the description of the scenario and the related list of questions. He/she was then asked to try to respond to each question by using the EHR Arche consumer application (see WP4 for a description of the consumer). Each physician had to work on both scenarios: for one scenario, he/she was asked just to use the standard search of the EHR Arche consumer that allows meta-data based search and retrieval of PDF documents. For the other scenario, he/she was asked to use the extended query interface that allowed using information items and predefined queries to retrieve individual data from the structured documents.

Summarizing, each physician used the standard search for one patient, and the extended content-based search for the other patient, with the order of patients changing for each physician. Overall, as eight physicians participated, we had eight cases where the standard query was used and eight cases where the extended query was used.

Additionally we observed the physicians while trying to solve the patient cases and documented whether the requested information described in the scenario were correctly found. The time for each scenario was limited to app. 20 minutes (compare Table 6).
<table>
<thead>
<tr>
<th>Step</th>
<th>Estimated time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction to the project and the objectives</td>
<td>app. 5 min.</td>
</tr>
<tr>
<td>Presentation of the functionality of EHR Arche Consumer</td>
<td>app. 15 min.</td>
</tr>
<tr>
<td>Presentation of the patient scenarios and than working on the patient scenarios</td>
<td>app. 20 min for each scenario</td>
</tr>
<tr>
<td>Short questionnaire of the physician</td>
<td>app. 5 min</td>
</tr>
<tr>
<td>Interview with the physician</td>
<td>app. 10 min</td>
</tr>
</tbody>
</table>

Table 6: Overall flow of an evaluation session.

After the scenarios, the physician was asked to fill in a short standardized survey (four-point Likert scale) asking for the overall benefit of the EHR Arche consumer concept. The survey addressed the following issues:

1. Support while searching for information
2. Support while using the found information
3. Usability and user friendliness
4. Miscellaneous

Finally we performed semi-structured expert debriefing interviews with each physician to discuss observed problems, usefulness of the tool and ideas for improvement. These interviews where audio-recorded and transcribed (compare Table 7).
III. Kategorie: Benutzerzufriedenheit, allgemeine Fragen

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>Wie finden Sie die <strong>Gestaltung der Suche</strong> nach med. Daten (Bildschirmaufbau, Funktionalität …)</td>
</tr>
<tr>
<td>(2)</td>
<td>Welche <strong>zusätzlichen Funktionalitäten</strong> wären noch wünschenswert, damit dieses Tool sinnvoll in der Praxis eingesetzt werden kann.</td>
</tr>
<tr>
<td>(3)</td>
<td>Was hat Ihnen besonders an der strukturierten Suche (erweiterte Suche - <strong>Einzelwertsuche</strong>) in Dokumenten mit Hilfe des EHR Arche Prototypen gefallen.</td>
</tr>
</tbody>
</table>

Table 7: Interview guideline for expert interviews.

4.6.2 Results

The EHR Arche Consumer prototype was found to be sufficient stable and performant for this study.

In the study group, all participants were able to correctly answer all clinical questions within the time limit of 20 minutes. In the control group, 20% of the expected information could not be found within the time limit.

In the study group, the time needed to answer the questions was 10 to 14 minutes for the first scenario, and 8 to 12 minutes for the second scenario. In the control group, only one participant was able to find the correct answers to all clinical questions within 20 minutes. The others had to stop the search after 20 minutes or aborted the search in face of the huge amount of documents.

In an debriefing interview and a written survey conducted after the study, all seven participants stated that the extended search was simple and self-explaining to use, that it was more intuitive than the standard search, that it was quicker than the standard search, that the predefined queries were useful to get an overview about a clinical situation, that information overload was better manageable with the extended search, and that it makes sense to further develop software tools to support searching in clinical documents. The participants appreciated the offered pre-defined queries as well as the offered brief queries.

The decision to make approx. 80 documents with realistic content regarding the additional diseases, medication etc proved to be very important, because in the evaluation the physician “treated” the patient in the scenario like a patient in the outpatient clinic.

4.6.3 Project documentation

More detailed documentation is available in the following project documents:

- Hübner-Bloder G: Instruction guideline of the evaluation (in german).
- Hübner-Bloder G: Scenario and related clinical questions of the patient Mr. Adamec (in german).
- Hübner-Bloder G: Guideline of the semi-structured expert interview (in german).
5 Discussion and Conclusion

5.1 What has been achieved?

We successfully implemented the original concept and evaluated it with clinicians. The implementation was done based on an IHE-XDS based EHR architecture and using ISO/EN 13606 archetypes. Based on a structured analysis of information needs, relevant information items were identified and - where possible - put together to predefined queries. The evaluation showed a high acceptance of this overall approach: Information searching is faster and more successful when information item-based queries analyzing structured documents are possible.

The project has produced the following outcome:

- An in-depth analysis of the information items, time windows, clinical scenarios and queries related to treatment of Diabetes mellitus patients.
- A list of ISO/EN 13606 archetypes that correspond to these information items.
- An application “Document Crawler” that enables mapping of queries to archetypes and that allows retrieving documents with related archetypes from IHE-XDS based architectures.
- An application “Document Consumer” that allows the clinician to search within structured and unstructured documents, either using meta-data based queries or information item-based queries.

While we have focused on the setting “diabetes mellitus treatment”, the overall approach and architecture can be generalized to other clinical settings. Both applications are developed in a way that they can handle any list of information items and queries, and any list of related archetypes.

5.2 What are further research questions?

- How much effort is needed to adapt the developed applications for other clinical settings, such as cardiovascular diseases?
- How can unstructured clinical documents developed in clinical routine be automatically converted into an archetype-based structured version?
- Can the benefits that were visible in our lab study be transferred to the real clinical setting?
- How can the benefit of structured clinical documents be exploited for others usage scenarios, such as clinical trials or clinical research?
6 Publications of the project team


7 References


4. American Diabetes Association. Standards of Medical Care in Diabetes - 2010. 2010 [cited 2010 02.02]; Available from: http://care.diabetesjournals.org/content/33/Supplement_1


8. MAXQDA. MAXQDA - The professional tool for qualitative data analysis. [cited 2010 15.09]; Available from: http://www.maxqda.com/


