Design and multicentric Implementation of a generic Software Architecture for Patient Recruitment Systems re-using existing HIS tools and Routine Patient Data

B. Trinczek[1]; F. Köpcke[2]; T. Leusch[3]; R.W. Majeed[4]; B. Schreiweis[5]; J. Wenk[6]; B. Bergh[5]; C. Ohmann[6]; R. Röhrig[4]; H.U. Prokosch[2]; M. Dugas[1]

[1]Institute of Medical Informatics, University of Münster, Germany; [2]Chair of Medical Informatics, Friedrich-Alexander-University Erlangen-Nuremberg, Germany; [3]Department of Information- and Communication-Technology, Düsseldorf University Hospital, Germany; [4]Department of Anesthesia and Intensive Care Medicine, Justus-Liebig University Gießen, Germany; [5]Center for Information Technology and Medical Engineering, Heidelberg University Hospital, Germany; [6]Coordination Centre for Clinical Trials, Faculty of Medicine, Heinrich Heine University Düsseldorf, Germany

Keywords

Clinical trials as topic, patient selection, software architecture, patient identification systems, hospital information systems

Summary

Objective: (1) To define features and data items of a Patient Recruitment System (PRS); (2) to design a generic software architecture of such a system covering the requirements; (3) to identify implementation options available within different Hospital Information System (HIS) environments; (4) to implement five PRS following the architecture and utilizing the implementation options as proof of concept.

Methods: Existing PRS were reviewed and interviews with users and developers conducted. All reported PRS features were collected and prioritized according to their published success and user's request. Common feature sets were combined into software modules of a generic software architecture. Data items to process and transfer were identified for each of the modules. Each site collected implementation options available within their respective HIS environment for each module, provided a prototypical implementation based on available implementation possibilities and supported the patient recruitment of a clinical trial as a proof of concept.

Results: 24 commonly reported and requested features of a PRS were identified, 13 of them prioritized as being mandatory. A UML version 2 based software architecture containing 5 software modules covering these features was developed. 13 data item groups processed by the modules, thus required to be available electronically, have been identified. Several implementation options could be identified for each module, most of them being available at multiple sites. Utilizing available tools, a PRS could be implemented in each of the five participating German university hospitals.

Conclusion: A set of required features and data items of a PRS has been described for the first time. The software architecture covers all features in a clear, well-defined way. The variety of implementation options and the prototypes show that it is possible to implement the given architecture in different HIS environments, thus enabling more sites to successfully support patient recruitment in clinical trials.

Correspondence to:

Benjamin Trinczek
Institut of Medical Informatics
Albert-Schweitzer-Campus 1
Building A11
48149 Münster
Germany
E-mail: Benjamin.Trinczek@uni-muenster.de

Background

Clinical trials are an important source to gain knowledge in medicine. Within clinical trials, new treatment options such as drugs or surgical techniques need to be evaluated in the well-defined setting of a clinical trial before they can be accepted as evidence based medicine. The validity of each clinical trial heavily depends on the number of trial participants.

According to ClinicalTrials.gov [1], more than 40,000 clinical trials are currently recruiting or active. Although this field of clinical research is highly active and standardized, it still struggles to recruit sufficient numbers of trial participants on time. One third of all clinical trials do not recruit enough patients within the estimated recruitment period, and every second trial has to extend its recruitment period [2]. According to Sahoo, only 18 per cent of European and 7 per cent of US trials complete enrolment on time [3]. This can result in termination of clinical trials and consequently in financial loss, lack of knowledge and less effective treatment of diseases. Obvious reasons might be too complex eligibility criteria and a lower disease prevalence than expected. However, according to Siminoff et al. [4], only 38 per cent of all eligible patients are offered to participate in a clinical trial. Thus, the problem of patient recruitment seems to be a problem of identifying all potential trial participants.

Several projects have shown that it is possible to support the process of patient recruitment with electronic systems. A recent literature review by Cuggia et al. [5] shows that at least 28 systems or methods have been published between 1998 and October 2009, each of them aiming at supporting patient recruitment. 20 out of 28 systems "address 'individual' barriers, at the level of the physician during the critical prescreening step". Additionally, clinical or research assistants can be involved as users. The users get notified about potentially eligible patients based on the comparison of computable eligibility criteria with documented patient data. As Cuggia et al. state, "this kind of system in general seems to be effective, since none of them failed in their goals". That is, why we will focus on this type of Patient Recruitment Systems (PRS).

To our knowledge, each of these systems/projects had focused on one particular or a small number of similar Hospital Information Systems (HISs). As HISs from different vendors vary in their functionality and parameterization, it is not guaranteed that these implementations can be adopted by other hospitals. This raises the challenge of finding a generic, scalable solution and implementation options for PRSs.

In the following, a HIS is defined as the integrated framework of all administrative or clinical information systems that are processing patient data in a hospital. A HIS thus contains the Electronic Health Record (EHR) System, which presents patient data in a graphical user interface to the clinicians and offers the possibility to search, insert and edit patient data.

Objectives

The overall objective of this project is to design and implement a generic, thus portable and scalable architecture for software-based PRS compatible with most of the currently available German HIS environments. The specific objectives are:

a) To identify and describe required features and data sources, trial metadata and patient data item groups that are necessary to implement a software-based PRS. Required trial metadata should already be available by the sponsor in an easily computable format, while patient data (elements) should be re-used from routine care to follow the concept of secondary use and thereby avoid redundant data entry.

b) To design a software architecture containing all necessary features, capsuled into modules with well-defined interfaces. To address common data privacy and security constraints, it should be possible to implement the architecture within a single HIS.

c) To identify existing components and tools of five different HISs that can be used to implement each of the architecture's modules.

d) To implement a PRS following the given architecture developed in objective b) utilizing the given implementation options from objective c) at each of the participating sites as a proof of concept.

Methods

Setting

To ensure representative results, the most commonly used HIS settings were analyzed regarding implementation options and architecture design. Given the central role of EHR Systems within HIS environments, a survey with Chief Information Officers (CIOs) of 33 German university hospitals and a telephone survey with 78 non-university hospitals were conducted regarding their EHR Systems. Based on these results, five university hospitals with different EHR Systems were chosen to participate in the project. This project focuses on university hospitals because clinical research is mainly conducted within academic sites. ▶Table 1 provides an overview on the participating hospitals.

Requirements engineering

A combination of methods was used to collect the necessary features of a software-based PRS. First, an analysis of already existing solutions according to Cuggia et al. [5] and referred systems was performed. It focused on common and successful features of PRS. These features were collected and separated into "must have" and "nice to have" by the authors. A feature was categorized as "must have" if it met at least one of the following criteria:

1. the process of querying for potentially eligible patients, notifying trial staff and listing potentially eligible patients cannot be performed without the respective feature;
2. an implementation of the feature has shown that a PRS is more successful/better accepted as without it (e.g. avoiding redundant notifications);
3. it is contained in several published PRS. Any feature not being categorized as "must have" was classified as "nice to have".

Second, a small set of potential PRS users was interviewed at the participating sites. Two members of the coordination centre for clinical trials of site #4 and #1, both being domain experts as study designers or study coordinators, listed necessary features from their point of view. Additionally, two study nurses were interviewed about the current process of patient recruitment and their feature requests. All interviews took one to two hours. These interviews were unstructured, giving the interviewees the possibility to detail any number of current hurdles and feature requests. A software developer either conducted the interview or was present, thus being able to directly react to feature requests and avoid misunderstandings.

Third, participating sites held a focus group meeting about the collected feature requests with regard to their prioritization, technical aspects of PRS and necessary trial metadata and patient data item groups. At least one software developer per site participated in the meeting.

Design of an architecture

Based on the lists of requested and prioritized features, trial metadata and patient data item groups, use cases were created according to the definition in the Unified Modeling Language (UML) version 2 [6]. As use cases addressing the same topic can be encapsulated into modules, software architects from two of the participating sites defined such modules. The draft has been discussed, refined and consented in the consortium as part of the focus group meetings. The required information for each of the modules was collected from its inherited use cases, together with its interfaces, connections to other modules and data exchange. A UML component diagram has been created using an open source modelling tool [7] to depict the architecture.

Analysis of implementation options

Each participating site analyzed implementation options within the local HIS for each of the modules of the architecture. A questionnaire, created by the participating sites, had to be answered for each of the sites' HIS tools. The questions were derived from the "must have"-feature list and mod-

ules contained in the architecture. These questions focus on technical aspects, such as the possibility to

1. access patient data,
2. implement eligibility criteria within a query,
3. send notifications and
4. present query results to the user together with the option to display a potentially eligible patient's EHR.

The sites had to provide feedback about the availability of the features and to go into detail with comments whenever possible. The complete questionnaire can be found in▶ Appendix 1.

By definition, a HIS contains a various number of subsystems. Due to the number of components, analyzing all available systems at each of the participating sites was not possible. Thus, each site analyzed at least their core systems (usually EHR System) since these are covering most of the patient data and are used by the majority of clinical staff. Implementation options involving a communication server and a data warehouse were analysed by site #5 and site #2, respectively.

A list of implementation options, containing each tool's advantages and potential drawbacks, was created.

Implementation of a PRS

Based on the developed architecture and the respective local HIS setting, a PRS was implemented as a proof of concept at each participating site. Each site implemented at least all "must have"-features of the architecture using available implementation options of the local HIS environment. An implementation option was prioritized if it

- implemented the features completely or almost completely electronically,
- was already in place and only had to be configured for the given use case,
- processed and transferred both trial and patient data electronically and
- was available within the standard license of each EHR system.

These criteria were chosen to ensure that structured, electronic data processing was applied for patient recruitment as soon and long as possible, users and developers could use tools they were already familiar with and make portability of the solution as likely as possible.

Additionally, each site tested its PRS in the context of an existing clinical trial[1]. The test included the following steps:

1. To process the trial's metadata with regard to notifications of the respective clinics.
2. To break down each eligibility criterion into single data items and map these items to patient data whenever possible.
3. To automatically query for potentially eligible patients.
4. To notify trial personnel about the results.
5. To list potentially eligible patients to authorized staff.

Results

HIS landscape in Germany

Thirty-three of 111 surveyed German hospitals currently apply Agfa ORBIS as EHR System. EHR Systems distributed by Siemens, i.s.h.med., Soarian (28) and medico (10) are installed at 38 of the surveyed hospitals. These two EHR vendors currently have a market share of 64 per cent regarding the survey. Other EHR Systems, such as Nexus/KIS, brightone iMedOne or CSC/iSoft ClinicCentre,

---

1 These trials were tested (ClinicalTrials.gov Identifier): NCT01252355, EudraCT-Nr. 2009-015740-42, NCT01409330, NCT01106820 and an investigator-initiated trial (IIT) about diabetes mellitus type-2.

are present at 1 to at most 5 hospitals, indicating that nearly a third of the EHR Systems' market is distributed broadly. Based on the survey, the participating sites, together with a locally developed system, represent the most commonly installed systems. The complete overview of EHR system distribution can be found in ▶Appendix 2.

Feature requirements and data set

Based on the description by Weng et al., a PRS can "operate in two modes: (a) patient-driven mode (rank or filter clinical protocols one patient a time) or (b) protocol-driven mode (rank or filter patients one protocol a time)" [8]. The first one corresponds to a decision support system, while the latter matches the current process of patient recruitment, where trial staff searches for eligible patients. The results focus on the "protocol-driven mode".

All of the interviewees described pre-screening as the most difficult step of patient recruitment. Identifying potential trial candidates, checking their eligibility by reviewing patient data and contacting the patients or treating physicians currently require a lot of manual work with high drop-out rates. A PRS is supposed to support or even supersede these steps. The potential users did not expect the PRS to assess all of the eligibility criteria automatically and thus prefer recommendations on potential candidates based upon key criteria.

Some of the existing PRS are able to detect and notify each patient's eligibility within the documentation process ("live recruitment"), which leads to direct interaction of the system with the physician and the patient. Unfortunately, patient data is not always documented immediately but in the end or even after a patient's stay, which makes live recruitment a "nice to have".

24 features could be found within the literature, from interviews and focus group meetings. 13 features were prioritized as "must have" based on the criteria listed in the methods. The complete list of required features can be found in ▶Appendix 3. ▶Table 2 exemplarily lists all features required with regard to querying for potentially eligible patients.

Two different types of data are necessary to successfully develop and implement a PRS: (1) trial metadata for each trial, containing eligibility criteria, trial title, recruitment period and participating clinics/departments, including trial staff. These data are usually available within a trial's protocol, which is used by the trial staff to identify patient candidates; (2) patient data that is queried to decide whether a patient is potentially eligible or not. Ideally patient data from all systems inside a HIS should be queried. Unfortunately, this would be complicated due to the heterogeneity of HIS. Thus, a set of patient data item groups has been chosen regarding eligibility criteria: demographics, diagnoses, treatment and procedures, laboratory findings and medications. These data item groups are usually available within the core systems, since they are necessary for billing purposes and routine care. Additionally, these data item groups are addressed by many eligibility criteria, as being observed by Ross et al. [16] and the authors previously [17, 18].

▶Table 3 lists trial metadata, ▶Table 4 contains item groups that are required to implement all of the above mentioned features of a PRS.

Modules and architecture of a generic PRS

Given the complete list, the focus group assigned the features to five different domains. Features within a domain cover similar functionalities. Each of the domains can be implemented by one particular software module, with different implementation options.

A generic PRS is composed of five modules covering the main topics of patient recruitment. ▶Figure 1 displays all five modules as a UML version 2 component diagram. Each module is represented by a rectangle with its name as label. Each module offers at least one interface that enables other modules of the PRS to exchange either trial metadata or patient data. The direction of data flow can be seen via the connection's representation: a circle defines an interface or service that is offered by the module it is connected to; another module makes use of the service by being connected with it using the semicircle notation.

In the following sections, each of the modules is described in detail, from the top left to the right side of ▶Figure 1. This description follows the common workflow of patient recruitment, which starts with input and administration of trial metadata, a (repeating) query for potentially eligible pa-

tients, notification of clinical staff, listing of eligible patients and documentation of their recruitment status.

Trial Administration Module

The Trial Administration module's function is to administrate clinical trials running at the study site. It enables the user to document the subset of trial metadata relevant for the PRS (▶Table 3). Besides the Screening List Module, it is the only one that requires a graphical user interface (GUI) for direct user interaction. Features regarding the administration and editing of these data have to be implemented. As three other modules need to receive trial metadata (e.g. trial title and list of staff to notify in the notification module), the Trial Administration module has to offer three interfaces, respectively. The purpose of this Module is to give an overview of trial metadata and pass relevant subsets of it on to the Query Module, Notification Module and Screening List Module.

Since the Trial Administration Module does not process patient data, one possible implementation option may be as a separate system within a HIS environment, for example as web application. Such types of implementations have already been reported, for example by Ainsworth and Buchan as well as Olasov and Sim [19,20]. The Trial Administration Module can also be implemented within the EHR system, if it offers non-patient-centric forms. This would probably lead to faster user adoption, since the users already know the EHR system's design and behaviour.

Finally, the Trial Administration Module does not even have to be implemented in an electronic system. In its most simple implementation option, the Trial Administration Module might be implemented by using a paper- or form-driven documentation and inserting necessary trial metadata into the other modules manually. This would lead to less implementation work load for the IT department, while waiving the option of a comfortable electronic trial administration environment.

Patient Data Module

The Patient Data Module offers access to patient data that is available in the HIS. Queries – for example in SQL, SPARQL or defined as Medical Logic Modules are received via the Module's Interface. The corresponding results always contain a list of patient IDs or pseudonyms of potentially eligible patients. Data security and mechanisms for both user authorization and authentication are essential for this module, since data contained and offered by the Patient Data Module can be used to identify patients. Patient IDs or pseudonyms have to be offered by this module (e.g. via a trusted third party), otherwise the trial staff is not able to identify patients, contact them/their treating physician and check the full eligibility.

Typically not all HIS components can be queried simultaneously; therefore a repository is needed to offer the functionalities of the Patient Data Module. In our analysis of 5 German HIS environments three potential systems with the required patient data were identified:

First, the EHR system's database can be re-used. As a core system, it usually contains most of the patient data and provides query capabilities. Additionally, patient data is relatively uptodate, since both clinical staff and connected systems document/import patient data directly. On the other hand, there might be technical limitations running queries for several trials, as this feature is typically not the core functionality of such a routine patient care system.

Second, a data warehouse (DWH) can be used. DWHs are designed to collect as much data as possible from different systems. In HIS environments DWHs are typically used for reporting purposes and quality management. Because DWHs are usually updated only on a daily or even weekly basis, there is a time delay regarding the ability to query patient data.

Third, the communication server can be used to query for potentially eligible patients. The main functionality of a communication server is to receive (administrative) patient data from any system inside a HIS, process it whenever necessary, and send it to another HIS component. A communication server usually knows only the specific snippets of patient data it receives, processes and sends via one particular interface message. Many communication servers however offer the functionality to also store those messages temporarily, thus allowing to create an electronic patient data repository on the fly. This data store can then be compared to the eligibility criteria from the Trial Administration Module. As soon as all necessary information is gathered and checked, the decision about inclusion or exclusion of a patient can be sent to the Query Module. This implementation option's development effort however is much higher than for other options, since communication servers are

usually not designed to store patient data and have those queried. In addition, not all patient data is sent to the communication server. On the other hand, it offers the possibility of a timely detection of potential trial candidates.

Although some patient data is available in structured, standardized formats – for example based on ICD10 or LOINC -, most EHR data items are stored with local, proprietary representations (e.g. local value lists). In addition, patient data is often documented as free-text. This makes pre-processing and transformations inevitable.

Query Module

The Query Module transforms eligibility criteria given by the Trial Administration Module into formal queries that can be processed by the Patient Data Module. It queries the Patient Data Module with the derived request after authentication and authorization. Queries are performed automatically with a specific frequency (defined in the corresponding trial metadata), relieving the clinical/trial staff from the responsibility to perform this task manually. Patients already found in previous queries are filtered in order to avoid overalerting. The Query Module passes a non-redundant list of patient IDs to the Notification and Screening List Modules for further processing.

Several implementation options with various advantages and disadvantages are available in the analyzed HISs:

Database triggers can be used to automatically check a patient's eligibility as soon as data is inserted or updated. This functionality is common in relational databases for EHR Systems and can be utilized easily. Though, a high usage of database triggers might lead to performance issues. Additionally, each eligibility criterion has to be hard-coded into the trigger list.

Some EHR Systems offer the functionality of triggering events or checks whenever certain data items within a form are changed or saved. This feature can then be used to trigger an eligibility test and would lead to timely detections of trial candidates. However, this would make the Query Module much more dependent on the EHR system's proprietary capabilities. Additionally, each form that might contain an important patient attribute has to be configured.

Cronjobs (time-based execution of commands) can be used to automatically run self-written scripts that concatenate a machine-readable representation of the eligibility criteria and query them against the Patient Data Module. This implementation option typically has the least development effort. Since it is a standard tool, cronjobs enable portability.

A Workflow Engine (WFE) is a GUI-supported trigger mechanism which is available in some EHR Systems. It offers the same functionality as database triggers, but on an application level. Like form triggers, eligibility criteria have to be implemented for each trial, while the tool's capabilities highly depend on the respective EHR system.

Finally, a communication server can be utilized to contain the Query Module, if the Patient Data Module is implemented within it. Checking a patient's eligibility while processing transferred data would enable a timely detection of trial candidates. Portability might be impacted, since every HIS contains a communication server, but modifications in terms of temporal storage are not common.

Notification Module

This module alerts clinical staff about the potential eligibility of trial candidates for a specific trial. It receives contact information of trial personnel for each of the trials running at the respective hospital from the Trial Administration Module. Together with a list of newly detected trial candidates, provided by the Query Module, it is able to alert the respective trial staff. In the literature some PRS implementations send notifications containing identifiable patient data [15,21–23], others only refer to another system to check new entries [24]. Since identifiable patient data must only be accessible to authorized clinical staff, we recommend notifying about new trial candidates and referring to the Screening List Module.

Depending on the existing infrastructure within a HIS and the need for timely notifications, several implementation options were identified:

Trial personnel can be notified via automatic emails, a feature already offered for instance by communication servers. As email addresses of the clinical staff are known/queryable, the creation of emails should not require too much development effort. However, this method does not guarantee a timely notification, since the receiver decides when to read an email.

Similar to emails, pagers or cell phones can be used to notify trial staff by displaying short messages. Sending short messages is a basic service commonly implemented in HISs. The clinical staff usually reads these messages immediately, therefore it is possible to notify trial personnel in a timely manner. On the other hand, this tool has to be utilized very carefully due to the well-known issue of alert fatigue [25].

An automated telephone call with text-to-speech software can be used to notify trial personnel in nearly the same way as using messages on cell phones. A duty telephone can be called, delivering basic information about both the trial and newly detected trial candidates.

Besides mobile implementation options, functionalities of the EHR Systems' GUI can be re-used to notify trial personnel. It commonly provides work lists with patient lists and related tasks. New entries can be created inside such work lists for each new trial candidate. As work lists are the common place for clinical staff to check for new tasks, adoption of this implementation option might occur easily. On the other hand, the implementation highly depends on the capabilities of the respective EHR system.

Pop-up windows/alerts show latest information, usually superimposing other GUI elements, to call the user's attention to this information. They could be used whenever a trial candidate is found and trial staff is logged in.

Screening List Module

The Screening List Module presents trial candidates, provided by the Query Module, to the clinical staff. A list of identifiable data such as patient IDs, patient names and birthdays can be presented to authorized clinical persons in accordance with data privacy regulations. If a patient is treated by another clinic of the same hospital, a pseudonym together with the treating physician's contact details can be shown.

To support clinical personnel, a direct link to each trial candidate's electronic health record should be provided, thus enabling the clinical staff to review full patient details and verify eligibility. Additionally, it should be possible to document a patient's recruitment status.

Similar to the Trial Administration Module, a non-patient-centric form of a core system, such as the EHR System, can be defined to implement this Module.

EHR work lists could be re-used for this module, too. Work lists usually delete an entry as soon as its task has been completed. Since screening lists contain completed entries, too, a work list might have to be enhanced.

Printouts, based on reports, could be used to implement the Screening List Module. Some features, such as links to a patient's EHR, won't be supported due to media disruption, but paper-based lists are easier to implement and can be signed.

PRS implementation

Each of the participating sites chose a combination of available implementation options to implement a PRS compliant with the proposed architecture.

Since there was no DWH available at site #1, the EHR System contained most of the patient data in the respective HIS. It offered a sufficient set of tools to implement the required features. Due to the customization options of the installed EHR System and experience in developing electronic forms locally, both the Administration Module and Screening List Module have been implemented in non-patient-centric electronic forms. Clinical staff is able to administrate trials on their own, e.g. by setting the recruitment status of the trial or editing the personnel to notify. Authentication and authorization were directly provided by the EHR system. The Query Module has been implemented with SQL queries within the EHR database. A cronjob executes active trial queries, checks for new trial candidates and addresses the communication server to create notification emails.

Site #2 utilized an i2b2 instance [26], which is updated daily. A separate DB schema within the i2b2 instance was developed. It lists each eligibility criterion separately, allowing to re-use them for several trials. A cronjob-triggered script concatenates the criteria for each trial and executes them. A dedicated work list for patient recruitment is populated with the respective results.

Site #3 implemented the Query Module with its WFE already in place. It observes pre-defined data item groups of the EHR System for new entries and checks the patients' eligibility every time a

patient attribute changes. Emails are created to notify trial personnel, together with an entry in the available work list. This entry reminds the trial personnel to check a patient's eligibility. Administration and Screening List were implemented in paper-based forms. These documents can be used for regulatory documentation, because they can be signed.

Site #4 developed a local trial register to keep track of currently running trials and cooperation. All necessary trial metadata is transferred to the other modules. Queries can be performed manually or automatically on a daily basis. Potentially eligible patients are tagged within the clinic's working list, from which the users can access the patient's EHR and document the recruitment status.

Site #5 developed a rule-based query engine listening to patient data processed by the communication server. Patient data is temporarily stored, until a patient's eligibility can be determined. If one or several new trial candidates are found, a telephone call with text-to-speech output and dual-tone multi-frequency-signal sensitive menu is placed. Basic information is given to the user at the phone, completed with a printout of the screening list.

▶ Table 5 lists the implementation options chosen at each site to implement a software-based PRS.

A key component of PRS is the representation of eligibility criteria. Since all of the sites and implementations receive trial metadata in a free-text format, a transformation of the data into a computable format has to be done. All EHR Systems use proprietary data structures, making it necessary to represent eligibility criteria in a local format. Whenever possible, structured data, such as ICD10, LOINC or OPS (the German version of the ICPM) have been used to query patient data, to avoid handling synonyms, abbreviations, misspelling or negations within free texts.

Each of the implemented PRS has been tested successfully over at least 6 months. At least one trial with active recruitment at the site has been chosen and implemented with the PRS. Automated queries have detected trial candidates, which were identified and contacted by trial personnel after notification. After review of patient eligibility and informed consent, participation status has been documented.

Discussion

By reviewing existing solutions and interviewing both users and software developers, it was possible to define a set of 24 features for PRS. They cover an electronic system's functionality to support patient recruitment from receiving a trial's eligibility criteria to the documentation of a patient's recruitment status. Out of these features, 13 were declared to be necessary for the successful usage of PRSs, because previous publications reported a benefit with this feature towards the system's effectiveness or users stated that they expect a system to offer this functionality (and by that reduce their workload).

Although several publications describe the workflow of an implemented PRS or selected parts of it [14,27–29], they often do not describe a generic software architecture in a structured, standardized way.

For example, Weiner et al. [21] picture several components together with connections indicating information flow. The components covered by them can be matched to this publication's architecture: The emergency department registration system and three other databases match the Patient Data Module. The real time recruiting component matches the Query and Notification Modules, as it queries these databases and sends out messages. The Intranet log corresponds to the Notification Module. Weiner et al. additionally list the Beeper, while our generic architecture abstracts from the medium and devices that transport and display the notifications. The Trial Administration Module cannot be matched. Furthermore, the setting described by Weiner et al. has only been installed and tested within an emergency department at one single site. These settings usually differ between hospital departments.

Patel et al. [30] list four different databases that are queried from an Ontology Reasoner to identify eligible patients. As the eligibility criteria are extracted from clinicaltrials.gov [1], a Trial Administration is not implemented. On the other hand, some useful metadata, such as the recruitment period or participating clinics of a hospital, might be missing when going without a Trial Adminis-

tration Module. Their case study only describes how to identify a cohort of potentially eligible patients, but misses the presentation of these results or notification of trial staff.

The FARSITE architecture [19] describes a system that spans over multiple general practitioners' patient records. Compared to our software architecture, it contains all of the modules, some of them implicitly. Their "Trial Recruitment Tool" runs queries, sends notifications and lists the eligible patients, which results in one bigger module that covers three of our proposed modules. Unfortunately, it depends on a central DB that can be queried by the Trial Recruitment Tool. To our knowledge, only a few projects or healthcare systems (such as the NHS), offer such a DB. The architecture described here is meant to be implemented within a HIS environment consisting of several systems.

For each of the developed modules a set of different implementation options could be identified, ranging from three to five different options. Nearly all of the implementation options are available at the participating sites. Thus, each site was able to choose and combine implementation options with best fit to the local infrastructure and requirements. PRS implementation was successful at each of the participating sites, therefore the concept can be realized with very different HIS. The decisions were made in favor of electronic options that were already available within standard installations to foster portability between hospitals. Nonetheless, EHR vendors still have to agree on standard definitions not only for data transfer, but also for data structures and modules to enable portability of implementations between different sites and systems. Until such standards are agreed on, the portability of the solutions is limited to sites using the same HIS. Our proof of concept covers the four biggest standard EHR products in Germany and a „home-grown" system – therefore we assume that our architecture is also portable to other sites.

The proposed PRS architecture was designed at five German hospital sites. However, we included very diverse HIS settings: from locally developed software to different standard-based systems of major software vendors. In addition, we included results from several international publications related to PRS. The basic HIS architecture, together with commonly implemented systems and tools, should be relatively similar in other settings. By a combination of interviews with domain experts at several sites and a non-formal review of published PRS we are confident that we cover the most important requirements of PRS. Therefore we expect that the proposed architecture can be applied in other countries as well.

The architecture needs to be implemented within the site-specific infrastructure. Some of the modules, especially the Trial Administration Module, might be implemented and deployed externally. If the Trial Administration Module is only accessible inside the respective HIS, trial coordinators or sponsors cannot enter, validate or update trial metadata. Hospital organizations with several sites and separated network infrastructures will have to implement the Trial Administration Module for each HIS.

Although several PRS have been implemented and published in the last years, their high-level functionality and data structures are quite similar. It is possible to develop a common feature set for a software-based PRS.

PRS can query a DWH or an EHR System. Each option has its advantages and disadvantages: On one side patient data in a DWH is usually not current (because of the delay in the ETL processes), which leads to a delayed detection of trial candidates and problems, if time critical patient recruitment is required for a particular trial. On the other side an EHR System usually does not contain all of the patient data documented within a hospital, thus one might not be able to query for all eligibility criteria of a trial directly in the EHR. Further, a communication server can also be used as a data source for a PRS, with the disadvantage, that not all clinical patient data is usually communicated via such a component (e.g. many assessment items directly documented within an EHR are never communicated to any other system).

Representation of eligibility criteria is a key component within PRSs. Due to the heterogeneity of criteria and data structures, we had to choose a subset of available patient data to query and to represent those criteria platform-dependently. Some additional patient data could be used, such as well-defined scores and classifications, but we observed that even this type of data is often documented only in free-text format, making it difficult to query appropriately.

Processing and presenting patient data is always a matter of data protection. With the given architecture and implementations, patient data is only presented to physicians within the treatment context of a given patient. The patient can then be approached by a physician who is already allowed

to see his/her data. Only after consenting, the study personnel are able to see patient details. Although all participating sites had to follow different regulations of different German states, it was possible to implement this methodology. Even more, we think that processing patient data within the boundaries of the treatment context leads to a higher security in comparison to the traditional method, where clinical staff reads paper records, approaches patients in waiting rooms and asks colleagues about potentially eligible trial participants.

The proof-of-concept implementations show that combinations of paper-based and electronically implemented modules can work together. For example, a Trial Administration Module can be paper-based while the other modules are implemented electronically. Although it has not yet been proven, we assume that a completely electronic PRS implementation can support patient recruitment best. Every time the users of such a system have to switch between paper- and electronic-based modules, a manual transfer (and, in worst case, conversion) of data has to be done, with all disadvantages like transcription errors, differences in interpretation or time-costly labor. Thus, we recommend implementing every module of a PRS electronically.

Conclusion

A generic PRS software architecture compatible with different HIS settings is feasible. Several implementation options for the modules of this architecture are available. The architecture and its implementation options can support software developers and decision makers to realize a PRS in their HIS with already existing tools. Each of the five participating sites successfully deployed a PRS.

Clinical Relevance Statement

With an increasing number of clinical trials conducted at hospitals and higher workload for clinical staff, there is a demand for software-based Patient Recruitment Systems. Required features and data items for a PRS were defined and a generic software architecture was developed, along with several implementation options. This PRS architecture was successfully implemented in 5 sites with heterogeneous HIS settings.

Conflicts of Interest

The authors declare that they have no conflicts of interest in the research.

Protection of Human and Animal Subjects

Neither human nor animal subjects were included in the project. PRS implementations were reviewed and approved by local data protection officers.

Author's contributions

BT coordinated the generic architecture design, designed the graphical representation of the architecture, conducted one of the PRS user interviews, checked already existing PRS implementations and wrote the manuscript. MD supervised the work of BT and reviewed the manuscript. BT, FK, TL, RWM, BS and JW designed the modules and architecture and contributed local implementation options and implementation of the PRSs.

All authors discussed and prioritized the requirements. All authors read and approved the final version of the manuscript.
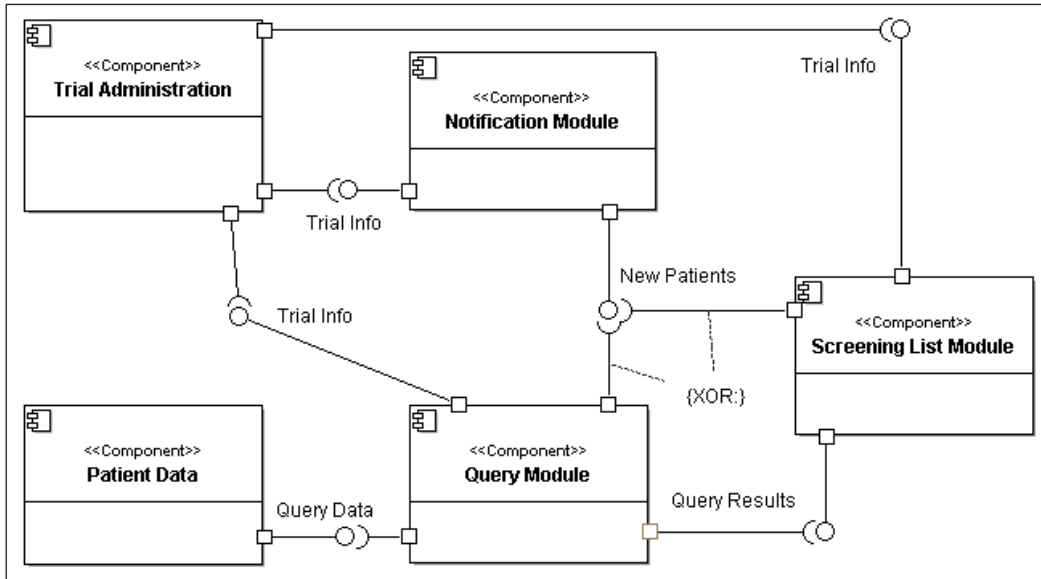
Fig. 1   UML component diagram of the generic software architecture for Patient Recruitment Systems. Each module is represented by a rectangle, with interfaces (lines) to offer (closed circle) or utilize (semicircle) data exchange.

Table 1   Key characteristics (number of inpatient beds, in- and outpatients per year, installed EHR System) of partici-
pating hospitals

Hospital	Beds	Inpatients/year	Outpatients/year	EHR System
1	1,400	47,000	376,000	Agfa Orbis
2	1,300	58,000	390,000	Siemens Soarian
3	1,100	46,000	160,000	Siemens medico//s
4	1,900	60,000	960,000	Siemens i.s.h.med
5	1,200	54,000	318,000	Locally developed

Table 2   List of features required for querying for potentially eligible patients; 2 out of 6 features prioritized as
"must have", while "nice to have" features mainly support the administrator's work

Feature	Priority	Mentioned in/by
Translate/compile human readable into computable criteria	Nice to have	Lonsdale et al. [9]; Borlawsky et al. [10]
Detect if patient has been identified previously	Must have	Afrin et al. [11]; Embi et al. [12]; Rollman et al. [13]; Embi et al. [14]
Query for potentially eligible patients automatically frequently	Must have	users; Butte et al. [15]
Concatenate computable criteria into single query	Nice to have	Software developers
Test eligibiltiy criteria/compiled query	Nice to have	Software developers
Trigger eligibility determination while documenting (live recruitment)	Nice to have	Embi et al. [12]

Trial metadata
Trial name
Trial identificator/register number
Eligibility criteria
Recruitment period
Participating clinic/department
Trial personnel to contact/notify

Table 3   Trial metadata required to implement all features and to
identify potentially eligible patients

Table 4   Patient data item groups required to implement all features of a PRS

Patient data item group
Name, birthday and patient ID to (1) link to EHR and (2) enter into screening list so that trial staff can identify patient, access additional information and contact patient
Demographics, such as age and gender
(Coded) Diagnoses
Treatment and procedures (Coded)
Laboratory data
Clinical findings
Medication
Patient's case data (clinic and treating physician)

Table 5   Implementation options chosen per participating site and module for implementing a PRS prototype.

Site	Administration	Patient Data	Query	Notification	Screening List
#1	Form in EHR System	EHR DB	Cronjob + SQL	Email	Form in EHR System
#2	Paper-based	DWH	Cronjob + SQL	Work List	Work List
#3	Paper-based	EHR DB	WFE	Work List	Paper-based
#4	Locally developed	EHR DB	Cronjob	Work List	Form in EHR System
#5	Paper-based	Comm. Server	Rule-Based	DECT call	Paper-based

References

1. ClinicalTrials.gov. Available from: http://clinicaltrials.gov/
2. McDonald A, Knight R, Campbell M, Entwistle V, Grant A, Cook J, Elbourne D, Francis D, Garcia J, Roberts I, Snowdon C. What influences recruitment to randomised controlled trials? A review of trials funded by two UK funding agencies. Trials. 2006 ; 7(1). Available from: http://dx.doi.org/10.1186/1745–6215–7–9
3. Sahoo A. Patient Recruitment and Retention in Clinical Trials; Emerging strategies in Europe, the US and Asia. London: Business Insights Ltd; 2007 Jun p. 149. Report No.: RBI00152.
4. Siminoff LA, Zhang A, Colabianchi N, Sturm CM, Shen Q. Factors that predict the referral of breast cancer patients onto clinical trials by their surgeons and medical oncologists. J Clin Oncol 2000; 18(6): 1203–1211.
5. Cuggia M, Besana P, Glasspool D. Comparing semi-automatic systems for recruitment of patients to clinical trials. Int J Med Inform 2011; Available from: http://www.ncbi.nlm.nih.gov/pubmed/21459664
6. UML 2.4.1. Available from: http://www.omg.org/spec/UML/2.4.1/
7. modelio open source modeling environment. Available from: http://www.modelio.org/downloads/download-modelio.html
8. Weng C, Tu SW, Sim I, Richesson R. Formal representation of eligibility criteria: a literature review. J Biomed Inform 2010; 43(3): 451–467.
9. Lonsdale DW, Tustison C, Parker CG, Embley DW. Assessing clinical trial eligibility with logic expression queries. Data & Knowledge Engineering 2008; 66(1): 3–17.
10. Borlawsky T, Payne PRO. Evaluating an NLP-based approach to modeling computable clinical trial eligibility criteria. AMIA Annu Symp Proc 2007; 878.
11. Afrin LB, Oates JC, Boyd CK, Daniels MS. Leveraging of Open EMR Architecture for Clinical Trial Accrual. AMIA Annu Symp Proc 2003; 2003: 16–20.
12. Embi PJ, Jain A, Clark J, Harris CM. Development of an Electronic Health Record-based Clinical Trial Alert System to Enhance Recruitment at the Point of Care. AMIA Annu Symp Proc 2005; 2005: 231–235.
13. Rollman BL, Fischer GS, Zhu F, Belnap BH. Comparison of electronic physician prompts versus waitroom case-finding on clinical trial enrollment. J Gen Intern Med 2008; 23(4): 447–450.
14. Embi PJ, Jain A, Clark J, Bizjack S, Hornung R, Harris CM. Effect of a Clinical Trial Alert System on Physician Participation in Trial Recruitment. Arch Intern Med 2005; 165(19): 2272–2277.
15. Butte AJ, Weinstein DA, Kohane IS. Enrolling patients into clinical trials faster using RealTime Recuiting. Proc AMIA Symp 2000; 111–115.
16. Ross J, Tu SW, Carini S, Sim I. Analysis of Eligibility Criteria Complexity in ClinicalTrials. San Francisco, CA.; 2010.
17. Trinczek B, Köpcke F, Leusch T, Majeed RW, Schreiweis B, Wenk J, Bergh B, Ganslandt T, Ohmann C. Röhrig R, Dugas M. Multizentrische Analyse von Routinedaten im Hinblick auf Ein-/Ausschlusskriterien klinischer Studien. Mainz: German Medical Science GMS Publishing House; 2011.
18. Köpcke F, Trinczek B, Majeed RW, Schreiweis B, Wenk J, Leusch T, Ganslandt T, Ohmann C, Bergh B, Röhrig R, Dugas M, Prokosch HU. Evaluation of data completeness in the electronic health record for the purpose of patient recruitment into clinical trials: a retrospective analysis of element presence. BMC Med Inform Decis Mak 2013; 13: 37.
19. Ainsworth J, Buchan I. Preserving consent-for-consent with feasibility-assessment and recruitment in clinical studies: FARSITE architecture. Stud Health Technol Inform 2009; 147: 137–148.
20. Olasov B, Sim I. RuleEd, a Web-based Semantic Network Interface for Constructing and Revising Computable Eligibility Rules. AMIA Annu Symp Proc 2006; 2006: 1051–1051.
21. Weiner DL, Butte AJ, Hibberd PL, Fleisher GR. Computerized recruiting for clinical trials in real time. Annals of Emergency Medicine 2003; 41(2): 242–246.
22. Chow E, Zuberi M, Seto R, Hota S, Fish EN, Morra D. Using real-time alerts for clinical trials. Applied Clinical Informatics 2011; 2: 472–480.
23. Ferranti JM, Gilbert W, McCall J, Shang H, Barros T, Horvath MM. The design and implementation of an open-source, data-driven cohort recruitment system: the Duke Integrated Subject Cohort and Enrollment Research Network (DISCERN). Journal of the American Medical Informatics Association: JAMIA 2012; 19(e1): e68–e75.
24. Dugas M, Lange M, Berdel W, Muller-Tidow C. Workflow to improve patient recruitment for clinical trials within hospital information systems – a case-study. Trials 2008; 9: 2.
25. Embi PJ, Leonard AC. Evaluating alert fatigue over time to EHR-based clinical trial alerts: findings from a randomized controlled study. Journal of the American Medical Informatics Association: JAMIA 2012; 19(e1): e145–e148.

26. Ganslandt T, Mate S, Helbing K, Sax U, Prokosch HU. Unlocking Data for Clinical Research – The German i2b2 Experience. Appl Clin Inform 2011; 2(1): 116–127.

27. Dugas M, Lange M, Muller-Tidow C, Kirchhof P, Prokosch H-U. Routine data from hospital information systems can support patient recruitment for clinical studies. Clin Trials 2010; 7(2): 183–189.

28. Fink E, Kokku PK, Nikiforou S, Hall LO, Goldgof DB, Krischer JP. Selection of patients for clinical trials: an interactive web-based system. Artif Intell Med 2004; 31(3): 241–254.

29. Thadani SR, Weng C, Bigger JT, Ennever JF, Wajngurt D. Electronic Screening Improves Efficiency in Clinical Trial Recruitment. Journal of the American Medical Informatics Association 2009; 16(6): 869–873.

30. Patel C, Cimino J, Dolby J, Fokoue A, Kalyanpur A, Kershenbaum A, Ma L, Schonberg E, Srinivas K. Matching patient records to clinical trials using ontologies. Proceedings of the 6th international The semantic web and 2nd Asian conference on Asian semantic web conference. Berlin, Heidelberg: Springer-Verlag; 2007. p. 816–29. Available from: http://portal.acm.org/citation.cfm?id=1785162.1785223

31. Ash N, Ogunyemi O, Zeng Q, Ohno-Machado L. Finding appropriate clinical trials: evaluating encoded eligibility criteria with incomplete data. Proc AMIA Symp 2001; 27–31.

32. Gennari JH, Sklar D, Silva J. Cross-tool communication: from protocol authoring to eligibility determination. Proc AMIA Symp 2001; 199–203.

33. Kamal J, Pasuparthi K, Rogers P, Buskirk J, Mekhjian H. Using an information warehouse to screen patients for clinical trials: A prototype. AMIA Annu Symp Proc 2005; 2005: 1004–1004.

34. Breitfeld PP, Ullrich F, Anderson J, Crist WM. Web-based decision support for clinical trial eligibility determination in an international clinical trials network. Controlled Clinical Trials 2003; 24(6): 702–710.

35. Grundmeier RW, Swietlik M, Bell LM. Research subject enrollment by primary care pediatricians using an electronic health record. AMIA Annu Symp Proc 2007; 2007: 289–293.

Appendix 1 List of questions used for site analysis

- Does the site have a software to administrate trials?
- For each implementation options:
    - Is it possible to add, update and delete trials?
    - Is it possible to import trial metadata? If yes, which data?
    - Is it possible to (implement and) show trial metadata in one of the core systems?
    - Is it possible to export trial metadata from one of the core systems?
    - Which data formats and transport protocols are part of the implementation possibility?
    - Is it possible to set a trial as being currently (in-)active?
    - Does the site have a (clinical) Data Warehouse?
    - How can the patient data be accessed?
    - Is the patient data up-to-date or is it updated frequently? If frequently, at what frequency does the update happen?
    - Which patient data item groups does the implementation possibility contain?
    - Is it possible to add patient data, such as recruitment status?
    - Is it possible to annotate patient data? Is it already annotated? If yes, which terminology is used?
    - Is it possible to run a query automatically, e.g. triggered by a cronjob?
    - Is it possible to derive queries from a specified eligibility format automatically?
    - Is it possible to check for patients that have already been identified as potentially eligible?
    - Is it possible to test a (derived) query manually?
    - Which device/medium can be used to notify trial personnel about potentially eligible patients?
    - Which existing systems offer a possibility/an interface to notify trial personnel?
    - Is it possible to list potentially eligible patients in the existing core systems?
    - Is it possible to restrict lists of patients to certain users/roles?
    - Is it possible to link and access a patient's EHR directly without re-login?
    - Is it possible to document a patient's recruitment status in the existing core systems?

Appendix 2   EHR Systems installed in German non-university hospitals and participating sites, sorted by number of appearance

EHR System	Number of installations in German hospitals	Represented by site #
ORBIS	33	1
i.s.h.med/Soarian	28	4; 2
Medico	10	3
NEXUS / KIS	5	
iMedOne	4	
Locally developed	3	5
ClinicCentre	3	
CLINICOM (Care Center)	3	
MCC	3	
Medicare	3	
fd klinika	2	
Lorenzo	2	
GPM	1	
IFU-KIS	1	
INTEGRA.Klinika	1	
ixserv	1	
Systema components	1	
No answer	7	
Total	111	

Appendix 3 Complete list of features

Feature	Priority	Mentioned in/by
Import trial metadata	Nice to have	Users; software developers
Define eligibility criteria from library	Nice to have	Users
Describe/design eligibility criteria in a machine- and human-readable way	Nice to have	Ash et al. [31], Gennari et al. [32]
Administrate notification staff	Must have	Software developers
Administrate participating clinics	Must have	Users; software developers; implicitly in existing solutions
Change trial's recruitment status	Must have	Users
Export trial metadata	Nice to have	Software developers
Administrate user access	Must have	Software developers; Users
Archive and lock closed trials	Must have	Software developers; Users; implicitly in existing solutions
Adjust frequency or type of trial	Nice to have	Software developers
Notify trial staff	Must have	Butte et al. [15], Afrin et al. [11]
Re-use routine patient data	Must have	Kamal et al. [33]; users; software developers; objectives
Pseudonymise patient id(entifying data)	Must have	Software developers; Users
Translate/compile human readable into computable criteria	Nice to have	Lonsdale et al. [9], Borlawsky et al. [10]
Detect if patient has been identified previously	Must have	Afrin et al. [11], Embi et al. [12], Rollman et al. [13], Embi et al. [14]
Query for potentially eligible patients automatically frequently	Must have	users; Butte et al. [15]
Concatenate computable criteria into single query	Nice to have	Software developers
Test eligibiltiy criteria/compiled query	Nice to have	Software developers
Trigger eligibility determination while documenting (live recruitment)	Nice to have	Embi et al. [12]
Display identified patients	Must have	Breitfeld et al. [34]
Link to patient's EHR	Nice to have	Kamal et al. [33]
Document patient's recruitment status	Nice to have	Embi et al. [12], Grundmeier et al. [35]
Print recruitment list	Must have	Users
Import candidate list into EHR System	Nice to Have	Breitfeld et al. [34]