Supporting Patient Screening to Identify Suitable Clinical Trials

Anca BUCUR[a,1], Jasper VAN LEEUWEN[a], Njin-Zu CHEN[a], Brecht CLAERHOUT[b], Kristof DE SCHEPPER[b], David PEREZ-REY[c], Raul ALONSO-CALVO[c], Lina PUGLIANO[d] and Kamal SAINI[d]

[a] *Philips Research Europe, The Netherlands*

[b] *Custodix NV, Belgium*

[c] *Grupo de Informática Biomédica, Universidad Politécnica de Madrid,Spain*

[d] *The Breast International Group, Belgium*

Abstract. To support the efficient execution of post-genomic multi-centric clinical trials in breast cancer we propose a solution that streamlines the assessment of the eligibility of patients for available trials. The assessment of the eligibility of a patient for a trial requires evaluating whether each eligibility criterion is satisfied and is often a time consuming and manual task. The main focus in the literature has been on proposing different methods for modelling and formalizing the eligibility criteria. However the current adoption of these approaches in clinical care is limited. Less effort has been dedicated to the automatic matching of criteria to the patient data managed in clinical care. We address both aspects and propose a scalable, efficient and pragmatic patient screening solution enabling automatic evaluation of eligibility of patients for a relevant set of trials. This covers the flexible formalization of criteria and of other relevant trial metadata and the efficient management of these representations.

Introduction

Clinical trials are key instruments in clinical research that enable the validation of research hypotheses, turning them into evidence that can be applied in standard clinical care. The population to be enrolled in a trial is usually described by a set of free-text eligibility criteria that are both syntactically and semantically complex, which makes their automatic evaluation on the patient data in order to assess the eligibility of that patient for a set of trials a challenging task.

To automate the evaluation of eligibility of patients for trials it is necessary to (1) extract and represent the semantics of the eligibility criteria in a machine-processable way and (2) automatically match each criterion with the relevant data elements available for each patient. The main focus in the literature has been on the first aspect and on proposing different methods for modelling and formalizing the eligibility criteria [1]. However the current adoption of these approaches in clinical care is limited and there is little evidence of the use of relevant healthcare standards or of evaluations

[1] Corresponding Author.

in real clinical settings. Due to the semantic complexity of criteria, these formalisms are either too complex and their modelling and implementation requires significant effort, or do not sufficiently capture the semantics of each criterion to enable its automatic evaluation. Less effort has been dedicated to the second issue, the automatic matching of eligibility criteria to the patient data managed in clinical care.

We address both aspects and propose a scalable, efficient and pragmatic Patient Screening solution enabling automatic evaluation of eligibility of patients for a relevant set of clinical trials. This covers the flexible formalization of criteria and of other relevant trial metadata and the efficient management of these representations. Additionally, we rely on our standards-based semantic interoperability solution to provide shared semantics between formalized trial information and care data, and to facilitate automatic linkage and matching of trial criteria to patient data. The solution has been built and evaluated with clinical trials of the Breast International Group (BIG)[2] and with anonymized patient datasets collected in the context of care[3].

1. Methods

In this section we describe our approach to formalization and representation of the trial criteria and introduce the key components of the Patient Screening solution. We also briefly describe the approach that is at the basis of our semantic solution as this provides key benefits leveraged by the application.

1.1. The formalization and standard representation of trial metadata

We have previously investigated the structure of the criteria in terms of re-occurring context patterns and of the core semantics encapsulated in each criterion, i.e. concepts that we have subsequently identified in standard terminologies [2]. Despite their syntactic semantic complexity, criteria have similarities in their structure and include one or more context patterns (e.g. expressing restrictions, conditions, etc.) and one or more concepts that express the core meaning of the criterion and give an indication of the patient data that is necessary for the evaluation of that particular criterion. Figure 1 gives examples of both types of entities co-occurring in criteria.

The concepts present in the criteria (e.g. cancer, Tamoxifen, radiotherapy) are linked in our solution to one of several standard terminologies/ontologies and become part of the core dataset used by the semantic solution. The context patterns such as "No concurrent()" or "No prior()"/"No history()" express what needs to be evaluated in the patient data in relation to the concept. We associate the patterns to templates that express the execution logic of the criteria. Complex criteria can include several patterns and therefore they will be represented by combining several templates.

Different context expressions with the same meaning are associated to the same template. E.g. "No prior()" and "No history()" require the same execution logic so they correspond to the same template. Templates can be linked to several formalisms and representations stored in our Trial Metadata Repository (described in Section 1.2).

[2] www.breastinternationalgroup.org.

[3] This work has been partially funded by the European Commission through the INTEGRATE project (FP7-ICT-2009-6-270253). www.fp7-integrate.eu.

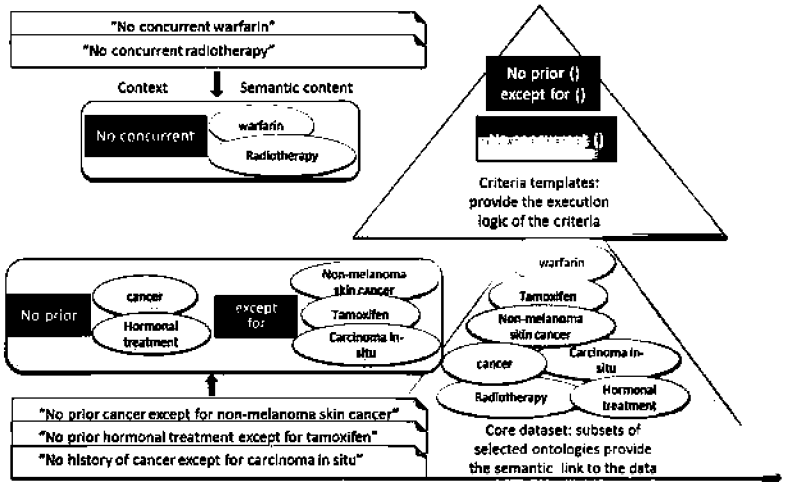

Figure 1. The structure of the eligibility criteria: elements describing context and core meaning

The placeholders in the patterns/templates, depicted above with "()", are filled in by concepts as Figure 2 shows. In the template these will be input parameters for an instantiation of that template. The formalisms and representations (and their chosen instantiations) can be specific to an application, a deployment, etc.

1.2. The Trial Metadata Repository

An important component in our solution is the Trial Metadata Repository (TMR) that manages all the metadata regarding clinical trials. This component flexibly addresses the needs of all applications in the environment that need access to trial metadata. For instance, for Patient Screening the TMR registers the following: (1) the trial acronym, (2) the trial description, (3) the target accrual range, (4) the number of enrolled subjects, (5) the start and (expected) end date of the trial, (6) the status of the trial (e.g. recruiting, closed, etc.) and (7) the eligibility criteria (text, executable logic, etc.). The TMR leverages the Biomedical Research Integrated Domain Group (BRIDG) model [3] to improve future interoperability of the repository and to gain the knowledge of the different perspectives ultimately encoded in the domain model. The BRIDG model is scoped and extended to address the needs of our solution. The TMR allows to store for each criterion several representations/formalisms and versions of execution logic.

1.3. The Core Dataset and the Semantic Interoperability Solution

At the centre of our semantic solution that links trial descriptions to the information model representing the patient data is the core dataset as in [4]: Soundly defined and agreed-upon clinical structures consisting of standard-based concepts, their relationships, quantification etc., that together sufficiently describe the clinical domain. To maximize reuse we chose to capture the semantics of the clinical terms by standard terminology systems such as SNOMED-CT[4] and LOINC[5], which are widely used in

[4] http://www.ihtsdo.org/SNOMED CT/

the clinical domain. Our data model leverages the HL7-RIM and the HL7 implementation guidelines[6]. The underlying semantic solution is described in [5]. In [6] we estimated the effort required for implementing mappings to the patient information model, the scalability of our solution with the number of trials and its extensibility to other clinical domains.

2. Results

For matching trial criteria to available patient information we developed an application which retrieves from the TMR the relevant template-instantiations of the criteria and applies them on the patient data retrieved from the patient data source. In the current implementation each template is instantiated to a representation that is a Groovy [7] script including SPARQL [8] queries to retrieve the necessary patient data (an example is provided in Figure 2). This solution is both efficient and flexible. It exploits the structure of the criteria enabling a high degree of reuse of templates across trials and criteria as the same patterns often re-occur across trials, and enables automatic matching to patient data leveraging the semantic solution.

We do not attempt full formalization of criteria as we believe that this is not necessary for efficient patient recruitment. The high overlap in templates across trials enables us to efficiently reuse the execution logic to address the real issue of automatic eligibility assessment: matching criteria of trials to the data elements of a large number of patients. Taking into account that the number of very complex criteria is rather small, we do not expect that further effort in formal representation of criteria and in pattern extraction yields high impact in real life scenarios for eligibility assessment.

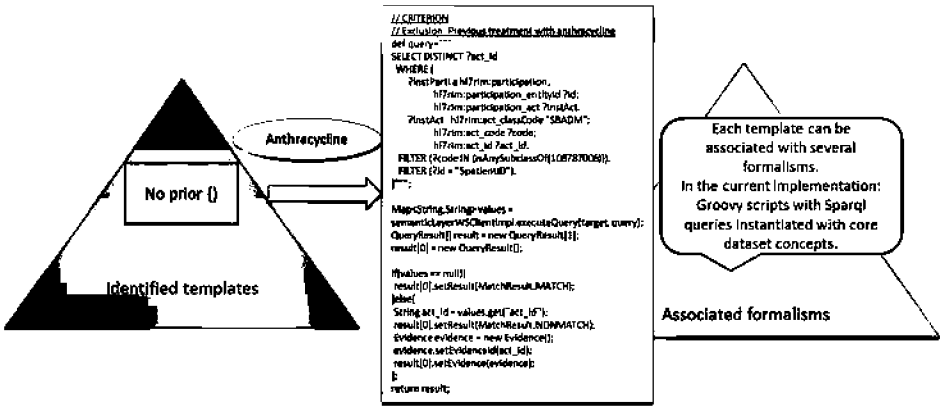

Figure 2. From templates to instantiated templates and their associated formalism

Based on the requirements of our clinical users we selected the relevant parts of the BRIDG model. We extended the subset of classes and class attributes of BRIDG with application-specific constructs, resulting in the information model underlying the trial metadata repository. The information model is subsequently exposed by web-services. It contains constructs to express inclusion and exclusion criteria and their

[5] http://loinc.org/

[6] https://www.hl7.org/implement/standards/

relation to trials, including specific content to allow the actual matching/verification of a trial criterion with a patient's data. When a new trial begins, its metadata is added to the TMR. In this step, criteria are linked to existing templates and corresponding formalisms, and the context variables (i.e. initial concepts) are specified. New templates can be added for criteria that are (partially) not covered by existing templates and in that case the corresponding execution logic is defined. As a trial has on average no more than 15-20 eligibility criteria and as our previous analysis showed that many of the patterns and concepts in criteria frequently repeat across trials, this step is not effort intensive. Once the trial metadata is uploaded to the TMR, it can be reused by the application in any other healthcare organization connected to the environment by updating the queries to the local information model.

3. Discussion

The approach described in this paper aims to automate certain tasks of the patient screening procedure to enhance modern clinical trial recruitment. Our novel solution automatically identifies the clinical trials for which a patient is potentially eligible by matching the trial criteria to the available patient information. Our approach is generic and flexible, integrating loosely coupled components with well-defined standard interfaces and making use of prominent standards in the healthcare domain, such as BRIDG for the Trial Metadata Repository and SNOMED-CT and LOINC to model semantics. We combine the use of formal representations of eligibility criteria with a pragmatic and efficient implementation in which templates are linked to execution logic and extensively reused. The semantics in our environment leverage standard ontologies and terminologies which supports efficient deployment in real healthcare environments.

The solution has been developed to suit the needs of a large clinical research network in the breast cancer domain. Extension to other clinical domains and deployment in new environments is straightforward due to the generic loosely-coupled architecture and the extensive use of standards. The solution can make use of different formalisms and data representations, and any of the components can be independently extended or replaced to fit the needs of new environments.

References

[1] Weng C, Tu SW, Sim I, Richesson R. Formal representation of eligibility criteria: A literature review. Journal of Biomedical Informatics. 2010; 43(3): pp. 451-67.

[2] Milian K, Bucur A, ten Teije A. formalization of clinical trial eligibility criteria: Evaluation of a pattern-based approach. Proc. of the BIBM. 2012; pp. 1-4.

[3] http://bridgmodel.nci.nih.gov/

[4] Vdovjak R, Claerhout B, Bucur A. Bridging the gap between clinical research and care - Approaches to semantic interoperability, security & privacy. Proc. of HEALTHINF. 2012; pp. 281-286.

[5] Paraiso-Medina S, et al. Semantic interoperability solution for multicentric breast cancer trials at the Integrate EU project, Proc. of HEALTHINF. 2013; pp. 34–42.

[6] Bucur A, et al. Identifying the semantics of eligibility criteria of clinical trials based on relevant medical ontologies. Proc. of the BIBE. 2012; pp. 413-421.

[7] http://groovy.codehaus.org/

[8] http://www.w3.org/TR/rdf-sparql-query/