dHealth 2019 – From eHealth to dHealth
D. Hayn et al. (Eds.)

doi:10.3233/978-1-61499-971-3-226

Expressing Patient Selection Criteria Based on HL7 V3 Templates Within the Open-Source Tool ART-DECOR

Simon OTT[a,1], Christoph RINNER[a] and Georg DUFTSCHMID[a]
[a]*Section for Medical Information Management, Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Austria*

Abstract. Background: Reuse of EHR data for selecting patients who are eligible for clinical research can substantially improve the recruitment process. ART-DECOR is an open-source tool that is commonly used to design and publish HL7 V3 templates of national (e.g. ELGA) and international EHR initiatives. Objectives: Extend ART-DECOR to allow the definition of criteria that may be used for patient selection. Methods: Using the native ART-DECOR development framework we extended existing ART-DECOR template associations by allowing conditions to be formulated. Results: An editor for the specification of conditions was implemented. The resulting criteria are internally translated to XPath expressions and can be immediately applied to CDA documents. As a prototypical application of our approach we implemented a "Trial Criteria Evaluator" tool that allows trial eligibility criteria to be composed of our ART-DECOR criteria and have them checked against a patient's CDA documents. Conclusion: Referring to HL7 templates, our criteria can be applied to documents of national EHR systems such as ELGA and hereby reach a broad patient cohort. Implementing our approach within ART-DECOR alleviates its reuse and enhancement by other researchers.

Keywords. electronic health records, patient selection, reference standards

1. Introduction

According to a recent WHO study, almost every second member state of the EU already has an operative national electronic health record (EHR) system in place [1]. In Austria, the ELGA system [2] was started in 2015 and aims to finalize the rollout phase in the outpatient sector this year.

Documents stored in EHR systems are frequently formatted according to the HL7 Clinical Document Architecture (CDA) standard [3]. Since the CDA model is very generic, HL7 V3 templates [4] are used to specify the structure and content of particular document types. ART-DECOR (https://art-decor.org) is an open-source tool and methodology that is commonly used to design and publish HL7 V3 templates of EHR systems. Templates are available for the ELGA CDA document types (http://elga.art-decor.org/) and of various other international EHR initiatives (https://art-decor.org/decor/services/Statistics?list=bbrs).

[1] Corresponding Author: Simon Ott, Section for Medical Information Management, Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Spitalgasse 23, A-1090 Vienna, Austria, E-Mail: simonott@gmx.at

Routine data recorded in EHRs have become an interesting source for clinical research [5]. A typical step in a clinical research project is the identification of a patient cohort with particular characteristics. This may for example be patients with a particular diagnosis and age, who would be eligible for a particular clinical trial or patients, who received a particular treatment and should be interviewed in the course of an outcome study. An automatic identification of patients based on their EHR data [6] could substantially reduce the high effort and rate of missed patients that typically characterize manual patient recruitment [7].

ART-DECOR allows the definition of high-level information needs, such as 'age' or 'diagnosis', and map them to a component of a CDA template that holds the corresponding data. These so-called ART-DECOR "concepts" and "template associations" can be stored in individual ART-DECOR project files. Currently they are not sufficient for the definition of fine grained criteria to select patients as they do not support the specification of particular conditions that the mapped template components would have to satisfy, such as "age ≥ 6" and "diagnosis = type 1 diabetes".

In this paper we therefore present an extension of ART-DECOR that allows the definition of criteria that may be used to identify patients with particular characteristics relevant for a clinical research project. They are based on ART-DECOR template associations and thus include a reference to those components of a CDA document that hold the source data of the criteria. The criteria are stored in the ART-DECOR project file and can then be applied to a patient's CDA documents to check whether they satisfy the criteria.

2. Methods

We will explain the requirements for our extension of ART-DECOR by means of an example. Assume that patients need to be checked for an elevated white blood cell count as a prerequisite to participate in a clinical trial. In the current version of ART-DECOR, we can define a concept "white blood cell count" (see fig. 1 top) of type Quantity and associate it with code "26464-8" of code system LOINC (OID = 2.16.840.1.113883.6.1).

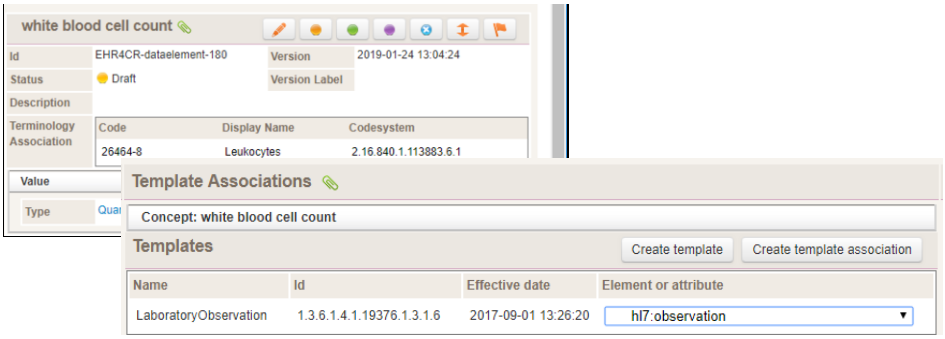

Figure 1. Definition of concept "white blood cell count" with associated LOINC code in ART-DECOR (top). Concept "white blood cell count" is mapped to element "hl7:observation" of template "Laboratory Observation" that is used in ELGA CDA laboratory reports for storing laboratory data (bottom).

We can further map concept "white blood cell count" to a suitable HL7 V3 template that refers to a particular component of a CDA document, which would hold the white blood cell count value of a patient (see fig. 1 bottom).

For the definition of our desired criterion "elevated white blood cell count" the following points are missing:

- Specification of conditions (e.g., we consider the white blood cell count to be elevated if the observation's "hl7:interpretationCode" element holds code "H" or "HH")
- Detailing of template associations if a concept is mapped to a generic template (e.g., in fig. 2 the generic template "Laboratory Observation" will only hold a white blood cell count value, if the observation's "hl7:code" element holds the LOINC code "26464-8" for Leucocytes)
- Executing the criterion against a CDA document to check whether the contained data satisfy the criterion

For implementing the first two points we decided to develop an editor that allows the specification of conditions for an ART-DECOR template association. For the third point, we added a testing component to ART-DECOR that allows the specified criteria to be immediately executed against an uploaded CDA document. All extensions should be implemented within the native ART-DECOR development frameworks (i.e. Orbeon Forms and eXist-db) to easily integrate them into an existing ART-DECOR environment using the ART-DECOR package manager.

In order to demonstrate a potential prototypical application of our extension, we planned to develop a "Trial Criteria Evaluator" tool. It should allow to evaluate whether a patient may be a potential candidate for a clinical trial based on his/her CDA documents.

3. Results

3.1. Adding conditions to ART-DECOR template associations

Originating from an existing template association, our extension allows the required "implicit" and "explicit" conditions to be added to define a criterion (see fig. 2).

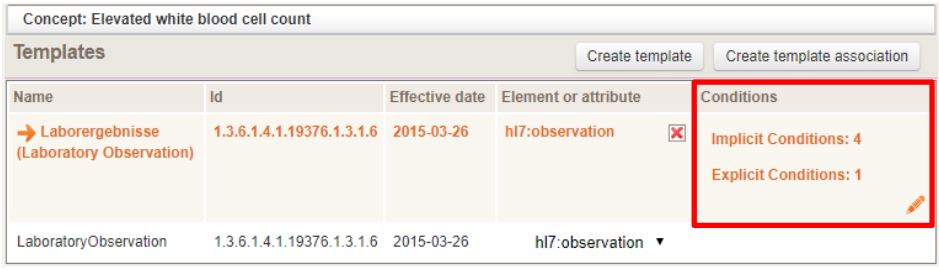

Figure 2. Conditions can be added to existing template associations.

For the definition of implicit conditions, existing ART-DECOR metadata specifications (i.e. fixed value constraints, terminology associations) are offered to the user (see fig. 3). As an example, the fixed value constraint for the template's element

"hl7:templateId" can be used to create the implicit condition that concept "elevated white blood cell count" should only refer to CDA observations that hold the OID "1.3.6.1.4.1.19376.1.3.1.6" in their "hl7:templateId" element. Further, the terminology association defined for concept "elevated white blood cell count" can be used to create the implicit condition that this concept should only refer to observations that hold code "26464-8" in their "hl7:code" element.

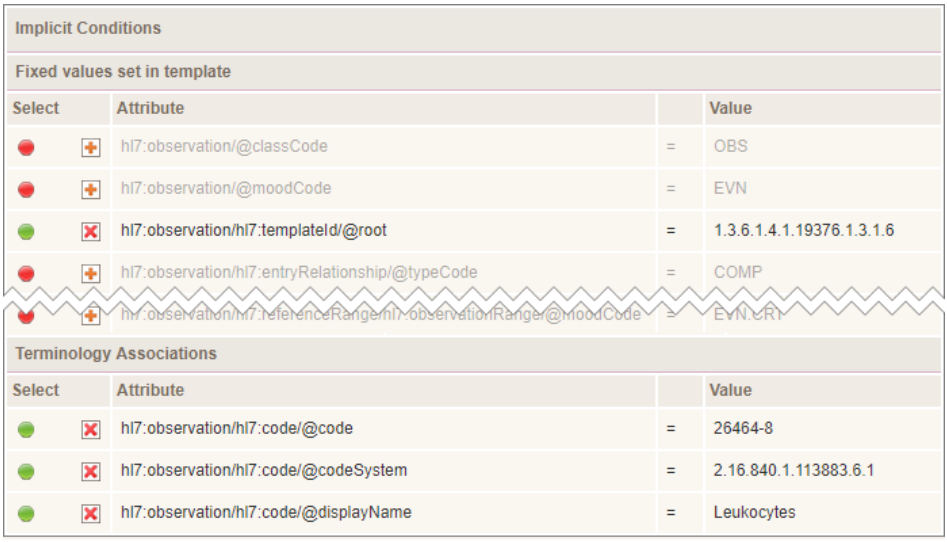

Figure 3. Specification of implicit conditions.

Explicit conditions are defined by the user by manually specifying one or more statements and linking them by Boolean operators (see fig. 4). Each statement is composed of an attribute of a template element, an operator and a comparison value. All attributes as predetermined by the template element's datatype and its child elements are offered for selection. As an example, element "hl7:code" is of datatype CE (Coded with Equivalents) and thus allows attributes "code", "codeSystem", "codeSystemName", "codeSystemVersion", "displayName", and "nullFlavor" to be selected. Further, operators "=", "≠", "<", "≤", ">", "≥" and "IS NULL"(i.e., value of the attribute is empty) are available to formulate a statement.

All data concerning the defined conditions are stored as additional elements of the corresponding <templateAssociation> component within the ART-DECOR project file.

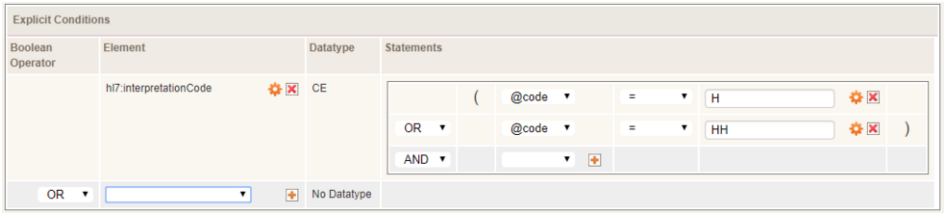

Figure 4. Specification of explicit conditions.

3.2. Checking criteria against CDA documents

In order to allow the specified criteria to be checked also independently of our tool, we automatically translate them to XPath expressions. Hereby, we logically link all implicit and explicit conditions of a criterion with a Boolean AND operator and generate the correct XML Schema Datatypes [8] (e.g., numerical comparison values are converted to xs:double, values of HL7 datatype TS or any flavor are converted to xs:dateTime). The expression derived from a criterion's conditions is used as the predicate of the XPath. The node-test of the XPath is the root element of the template or in case of multiple root elements of a template the parent of these elements as a wildcard. The generated XPath (see fig. 7) uses the HL7 defined XML namespace for v3 "urn:hl7-org:v3" [9].

For an immediate checking of the criteria we also implemented a testing component within our tool. It allows the user to upload a CDA document and execute the generated XPaths against the document. All components of the CDA document that are found by means of the XPath and thus satisfy the criteria are displayed.

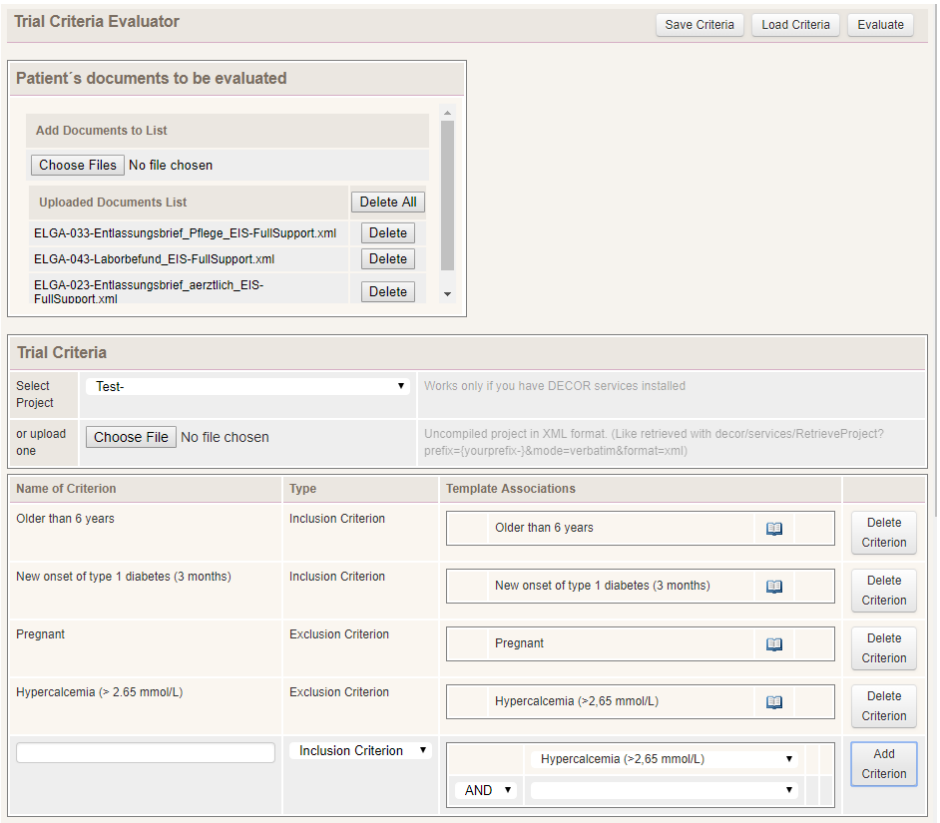

Figure 5. Trial Criteria Evaluator. Criteria defined in ART-DECOR project "Test" were used to compose four trial criteria of a diabetes trial (https://clinicaltrials.gov/ct2/show/NCT01390480). Three CDA documents were uploaded to be checked against the trial criteria.

3.3. Trial Criteria Evaluator

The "Trial Criteria Evaluator" (see fig. 5) demonstrates in a prototypical way, how our extension could be applied to check a patient's eligibility for a clinical trial.

Being an Orbeon forms application, it can be added to an existing ART-DECOR environment by installing the corresponding package in the eXist-db. As it is completely independent of ART-DECOR itself, it may also be installed as standalone version without ART-DECOR. It allows the user to load the criteria that he/she defined within ART-DECOR and combine them with Boolean operators to form complex trial inclusion and exclusion criteria.

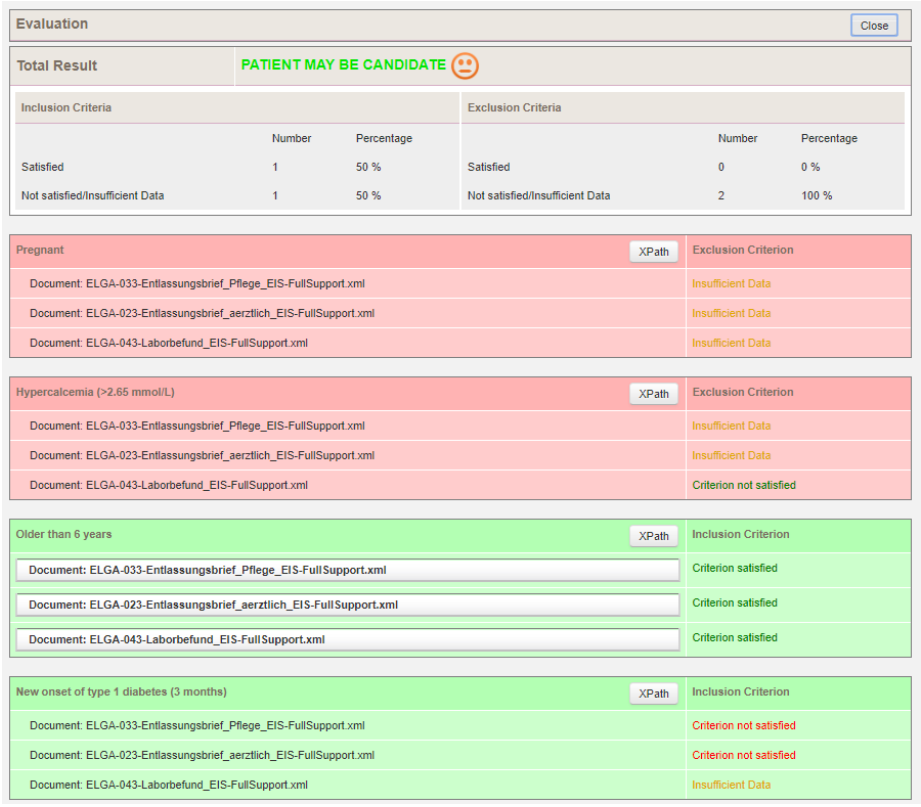

Figure 6. Result screen. At the top, the total result and an overview of satisfied inclusion/exclusion criteria is shown. Below, a detailed report is displayed that shows for each criterion and document, whether the criterion is satisfied/not satisfied by the document's data or whether the document does not contain the required data. Inclusion criteria are depicted in green background color, exclusion criteria in red.

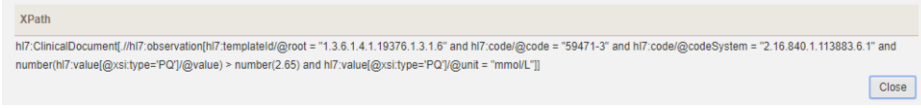

Figure 7. XPath for criterion "Hypercalcemia (>2.65 mmol/L) of fig. 6.

The user can then upload a set of documents of a particular patient and evaluate whether the patient could be a potential candidate for the trial. Hereby all uploaded documents are checked iteratively whether they satisfy one of the trial criteria. Our evaluation is conservative insofar, as we assume that the uploaded documents represent only a subset of the patient's complete medical history [10]. We thus have to expect that there may be additional data that we are not aware of but may nevertheless satisfy a trial criterion. Consequently, the only safe assessment that our trial criteria evaluator is capable of, is to exclude a patient if the uploaded documents satisfy one or more exclusion criteria. Otherwise it concludes that the patient may be eligible and displays a report of which criteria are satisfied by the uploaded documents, respectively for which criteria the uploaded documents do not contain the required data (see fig. 6).

For each criterion the corresponding generated XPath can be displayed (see fig. 7).

4. Discussion

The work presented in this paper is the result of an ongoing bachelor thesis of the first author and is thus of preliminary nature. The final results will be presented at the conference. We further plan to make our extensions of ART-DECOR publicly available as open-source code by then.

Various suggestions have been made to automate the identification of patient cohorts based on EHR data [11]. These approaches typically refer to institutional EHR systems with proprietary data models. This limits the number of patients that can be addressed and requires individual mappings of the selection criteria to the data model of each single EHR system. Fernandez-Breis et al. suggest to map the selection criteria to standardized EHR data defined by means of openEHR archetypes [12]. Our approach is similar but is based on the more prevalent HL7 CDA standard and associated HL7 v3 templates. We further implemented our method within the open-source tool ART-DECOR that is used within several national EHR system initiatives. The EHR4CR platform [13] uses a distributed architecture, where trial criteria can be defined at a central server and transmitted to clinical data warehouses (CDW) of participating hospitals. It requires individual mappings of the criteria expressed in the central ECLECTIC syntax [14] to the data models of each single CDW. A similar approach is pursued by SHRINE [15], which allows queries to be distributed to CDWs that are based on the i2b2 [16] model. As a prerequisite the participating CDWs have to support common i2b2 ontologies.

Compared to earlier projects, the first main contributions of our work is that our criteria directly refer to elements of HL7 CDA documents and hereby make use of the knowledge of the CDA structure as specified within HL7 v3 templates. They may thus be applied to the document types of national EHR systems such as ELGA and hereby reach a broad patient cohort. Our second main contribution is that we implemented our approach as an extension of the open-source tool ART-DECOR that is widely used in the course of (inter)national EHR initiatives. This alleviates reuse and further enhancement of our work by other researchers.

Several alternatives exist for the expression and execution of criteria, such as Arden Syntax or SNOMED CT. Applying ontology-based semantic knowledge [17] in the processing of criteria would also be a useful extension of our work, e.g. when searching for patients with different types of lung diseases. A comparison with these alternative methods is beyond the scope of the present paper but is planned for a more elaborate version of this work.

Our tool's functionality to generate XPath expressions for ART-DECOR template associations could also be useful in other domains, where data has to be retrieved from CDAs. If, for example, a particular lab value of a patient's CDA documents has to be processed in an Arden Syntax MLM, the required XPath to be applied within the MLM's curly braces clause could be easily generated by our tool.

When implementing an automated identification of patients that might be eligible for a research project, obviously existing judicature with respect to data protection would have to be strictly considered.

References

[1] World Health Organization, Global eHealth survey 2015, http://portal.euro.who.int/en/data-sources/ehealth-survey-2015/, last accessed January 10, 2019.
[2] S. Herbek, H.A. Eisl, M. Hurch, A. Schator, S. Sabutsch, G. Rauchegger, A. Kollmann, T. Philippi, P. Dragon, E. Seitz, and S. Repas, The Electronic Health Record in Austria: a strong network between health care and patients, *European Surgery* (2012), 155-163.
[3] R.H. Dolin, L. Alschuler, S. Boyer, C. Beebe, F.M. Beilen, P.V. Biron, and A. Shabo, HL7 Clinical Document Architecture, Release 2, *J Am Med Inform Assoc* **13** (2006), 30-39.
[4] K.U. Heitmann, J. Curry, and L. Nelson, HL7 Templates Standard: Specification and Use of Reusable Information Constraint Templates, Release 1, in, 2014.
[5] H.U. Prokosch and T. Ganslandt, Perspectives for medical informatics. Reusing the electronic medical record for clinical research, *Methods Inf Med* **48** (2009), 38-44.
[6] M. Cuggia, P. Besana, and D. Glasspool, Comparing semi-automatic systems for recruitment of patients to clinical trials, *Int J Med Inform* **80** (2011), 371-388.
[7] E. Fink, P.K. Kokku, S. Nikiforou, L.O. Hall, D.B. Goldgof, and J.P. Krischer, Selection of patients for clinical trials: an interactive web-based system, *Artif Intell Med* **31** (2004), 241-254.
[8] D. Peterson, S. Gao, A. Malhotra, C. Sperberg-McQuenn, and H. Thompson, W3C XML Schema Definition Language (XSD) 1.1 Part 2: Datatypes, https://www.w3.org/TR/2012/REC-xmlschema11-2-20120405/datatypes.html, last accessed January 3, 2019.
[9] C. McCay and M. Stephens, XML Implementation Technology Specification, Release 2, http://hl7.ihelse.net/hl7v3/infrastructure/its_r2/its_r2Spec.html, last accessed January 5, 2019.
[10] O. Dameron, P. Besana, O. Zekri, A. Bourde, A. Burgun, and M. Cuggia, OWL model of clinical trial eligibility criteria compatible with partially-known information, *J Biomed Semantics* **4** (2013), 17.
[11] F. Kopcke and H.U. Prokosch, Employing computers for the recruitment into clinical trials: a comprehensive systematic review, *J Med Internet Res* **16** (2014), e161.
[12] J.T. Fernandez-Breis, J.A. Maldonado, M. Marcos, C. Legaz-Garcia Mdel, D. Moner, J. Torres-Sospedra, A. Esteban-Gil, B. Martinez-Salvador, and M. Robles, Leveraging electronic healthcare record standards and semantic web technologies for the identification of patient cohorts, *J Am Med Inform Assoc* **20** (2013), e288-296.
[13] Y. Girardeau, J. Doods, E. Zapletal, G. Chatellier, C. Daniel, A. Burgun, M. Dugas, and B. Rance, Leveraging the EHR4CR platform to support patient inclusion in academic studies: challenges and lessons learned, *BMC Med Res Methodol* **17** (2017), 36.
[14] R. Bache, S. Miles, and A. Taweel, An adaptable architecture for patient cohort identification from diverse data sources, *J Am Med Inform Assoc* **20** (2013), e327-333.
[15] G.M. Weber, S.N. Murphy, A.J. McMurry, D. Macfadden, D.J. Nigrin, S. Churchill, and I.S. Kohane, The Shared Health Research Information Network (SHRINE): a prototype federated query tool for clinical data repositories, *J Am Med Inform Assoc* **16** (2009), 624-630.
[16] S.N. Murphy, G. Weber, M. Mendis, V. Gainer, H.C. Chueh, S. Churchill, and I. Kohane, Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2), *J Am Med Inform Assoc* **17** (2010), 124-130.
[17] S. Liu, Y. Ni, J. Mei, H. Li, G. Xie, G. Hu, H. Liu, X. Hou, and Y. Pan, iSMART: Ontology-based Semantic query of CDA documents, *AMIA Annu Symp Proc* **2009** (2009), 375-379.