URL: https://ieeexplore.ieee.org/abstract/document/8402022
DOI: 10.1109/ICDEW.2018.00019

Dynamic Service Detection for Automated Patient Selection for Study Recruitment Purposes

Efthymios Chondrogiannis, Vassiliki Andronikou, Efstathios Karanastasis and Theodora Varvarigou
National Technical University of Athens, NTUA
9 HeroonPolitechniouStr, 15773, Athens, Greece
{chondrog, vandro, ekaranas}@mail.ntua.gr, dora@telecom.ntua.gr

***Abstract*—ICT-enabled automated processes facilitating subject recruitment in clinical trials is a hot topic in the clinical research domain. However, the successful detection of candidate subjects across patient records of different healthcare providers requires dealing with a variety of issues, including but not limited to different terminologies and structures as well as missing patient data. In this work a novel system for automated patient selection for study recruitment purposes is being presented. The system is based on the dynamic service detection for eligibility criteria evaluation, while the patient data retrieved from each healthcare entity are further processed for clinical research purposes. The presented system is also compared with a different framework that we have already developed in the past that is based on query rewriting and translation for eligibility criteria evaluation purposes. The advantages and limitations of the two aforementioned approaches are being discussed as well as their potential combination for automated patient selection.**

Keywords—automated patient selection; service-oriented architecture; dynamic service detection; semantic web; clinical trials

I. INTRODUCTION

Clinical studies are tests or experiments performed on human volunteers that intend to discover new knowledge about existing interventions (e.g., drugs) or new ones. Eligibility Criteria (EC aka Inclusion/Exclusion criteria) have a distinctive role in every clinical study since they define the characteristics that the patients must meet for being eligible to participate in the study.

Natural Language is often used by clinical experts for the expression of EC. However, further processing of EC by software agents is quite difficult and prone to errors due to the EC complexity [1]. In order to allow for the well-structured, unambiguous and consistent specification of EC, the design of a graphical environment (GUI) for the formal expression of EC is necessary. The design of the GUI requires a considerable amount of time while it is often tailored to cover the specific needs of an institute. Hence, it cannot be used in different studies or organizations. The design of a GUI that can be used (or easily adapted) in many different clinical studies across different entities is of great value. Such an attempt is rather complicated given that the GUI should minimize the human effort, be addressed to non IT experts and should ensure the correct interpretation of the provided data.

Detecting potentially eligible patients for a clinical trial based on the EC specified is another challenging issue, since the EC are often expressed using different terminologies, semantics and overall wording than the ones used in healthcare entities. Additionally, the availability and accuracy of the patient data recorded within each healthcare entity often differ with direct impact in the evaluation of user-defined EC, as described in the following section. In fact, there are quite a few cases that it is difficult or even impossible to match patients' records with the EC, since the corresponding information is not available in the datasource.

In this work, we present the architecture of a system developed for detecting the eligible patients based on the EC of a clinical trial. Both the graphical environment that enables users to formally express EC and the system developed for patient selection purposes are being described. The document is structured as follows. Initially, section 2 describes the EC representation and formal expression as well as the challenges met when using the EC for detecting candidate subjects for recruitment. In section 3 we present the overall architecture of the system developed and in section 4 the approach followed and the mechanisms developed for dynamic (web) service detection for EC evaluation purposes. A discussion follows in section 5 in which we compare the system being presented with a similar work performed within the PONTE project [2]. Finally, in the last section we summarize our work.

II. RELATED WORK

A. Eligibility Criteria Representation and Formal Expression

Expressing the EC in a computer-understandable language is quite difficult due to their complexity and, especially, the variety of concepts mentioned within each one. For this purpose, a lot of relevant work has been published by international standards developing organizations such as the Clinical Data Interchange Standards Consortium (CDISC) [3] and Health Level 7 (HL7) [4]. Also, various works have been published so far in the literature, as we have already presented in our previous work [5] that they focus on a specific type of clinical studies while they can often support specific tasks. As a result, the existing approaches cannot adequately cover the EC representation across the different types of clinical studies, while the application of EC to patient records in Healthcare Entities for detecting candidate subjects for recruitment through automated processes is not supported.

For EC representation we have developed an XML schema that enables the specification of the Inclusion and/or Exclusion (I/E) criteria of a clinical study along with the main parameters of each one. According to the schema developed, for each criterion, a unique ID should be provided along with a human readable description and zero or more formal expressions of the same criterion in machine language. Concerning the formal description of a criterion, a model-based approach has been followed according to which the users can specify the conditions that the patients' data should satisfy based on the terminology within the EC model along with the international classification systems and codings that are currently linked with our model. For expressing the conditions (i.e., the set or range of values that the patient data should belong to) two options are offered, being XML and SPARQL [6] with the transition from one representation to the other one being feasible through an automated process.

In this work the same approach has been followed, with the main difference being that JSON [7] has been used for specifying the conditions that each parameter should meet in a way that facilitates the automatic construction of SPARQL queries, if being necessary. The main reason for using JSON instead of XML lies in the fact that we would like to separate the technology used for EC organization from their formal expression and simultaneously maintain their size as small as possible. Consequently, for the representation and formal expression of EC, we have used a variety of technologies including XML (overall structure of EC), JSON (formal expression of conditions) and OWL [8] (EC model representation) along with international classification systems and codifications (formally expressed using the aforementioned technologies).

B. Eligibility Criteria Consumption for Patients Selection

For the successful application of EC on patients' records in Healthcare Entities for estimating the size of potentially eligible patient population and detecting the specific patients (i.e., patients satisfying the EC) for potential participation in a clinical study, a variety of issues that are briefly described in the following paragraphs should be overcome.

Models and terminologies used for the formal expression of eligibility criteria have often significant differences with the corresponding ones used for healthcare purposes, as a result of their independent design and the different purposes they serve. For detecting the similarity among the terms of two models (including ontologies) there are many tools and algorithms that can automate this process [9]. However, bridging the gap between the EC and patient record model is much more complicated and existing algorithms and tools are not effective. Moreover, the Language used for expressing EC is often different from the one used for accessing the data stored within a healthcare entity. For instance, within PONTE project [2], the Eligibility Criteria specified during the design of a clinical study have been formally expressed with XML and/or SPARQL. However, the patients' data provided by two different healthcare entities have been stored within separate Relational Databases in which specific schemas and vocabularies are used. SPARQL endpoints such as D2R server [10] enable users to treat a relational database as a virtual RDF graph by automatically translating SPARQL queries to SQL. However, when there are significant differences among the models and terminologies used, this process is quite difficult, if not impossible.

Another important issue is the fact that healthcare patients' data are in general scattered across different data sources even within the same entity and they often do not necessarily follow the same format in terms of structure, terminology and interface, preventing their accessing in a uniform way. For instance, within a hospital department some data may be internally stored within a relational database while a considerable amount of data may be available in the form of images (e.g., scans), text documents (e.g., reports) or excel files. Hence, information needs to be combined from various heterogeneous sources.

Missing data in patients' records, either because the healthcare provider does not record such information (e.g., lifestyle) or no such information has been collected for some patients, also pose challenges in patient selection. In fact, the evaluation of Eligibility Criteria is much more complicated since in many cases the evaluation of a single criterion implies that several parameters, with only a part of them being available in the patients' records.

The aforementioned issues require the development of a system that is adjustable to the specific needs of the different clinical research attempts while also provides great flexibility for accessing highly heterogeneous patient data. Service-oriented Architectures (SOA) in combination with Semantic Web technologies and Automated Web Service composition methods and techniques [11] provide the means for the definition and discovery of healthcare specific services that can be accordingly combined for patient selection purposes.

III. APPROACH FOLLOWED

A. Layered System Architecture

For simplifying the access to the data source, several components have been developed, which are organized in layers, as presented in Fig. 1. The higher layers are data source independent and enable users to specify the EC through a graphical environment as well as further process the EC specified for the particular clinical research purpose. The lower layers are data source specific and enable users to access the data residing in a specific data source taking into account the data structure and terminology used.

The system interacts with the end users through a graphical environment developed (GUI), while the EC specified are internally stored in a machine processable manner using a combination of data interchange formats (i.e., XML and JSON), knowledge representation technologies (i.e., OWL) and standards (i.e., international classification systems such as ICD-10 [12] and ATC [13], and codifications such as HL7 administrative gender and/or CDISC age group). The two layers below the GUI (i.e., data source independent and data source specific layers) use the EC specified for detecting the eligible patients and providing answers to clinical research specific questions (e.g., the total amount of patients that meet criteria or the expected amount of eligible patients). The data source independent layer provides a description of the

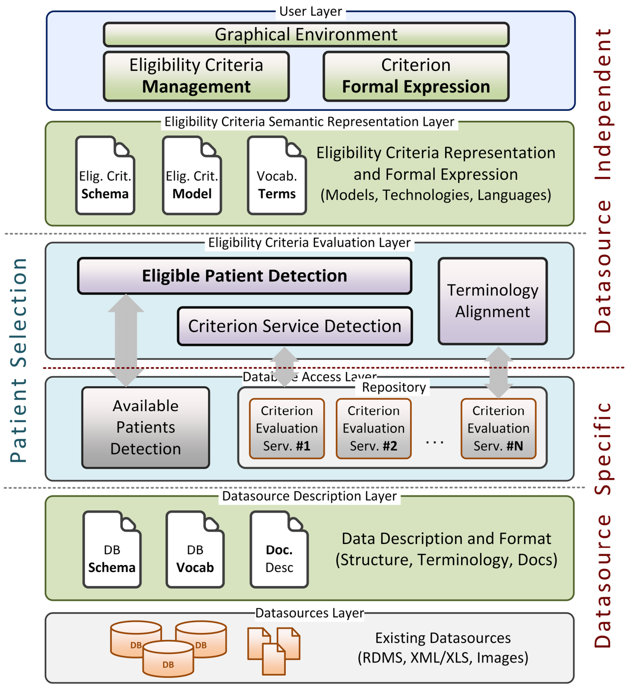

Fig. 1. System Layered Architecture

mechanism used in an abstract manner, while the data source specific layer provides an implementation of a series of services for accessing patient data within a specific healthcare entity. The patient data recorded are often structured, in which case they follow specific schemas and vocabularies to a great extent. Nevertheless, additional information may be available in semi-structured data or even unstructured documents including free text files (e.g., patient diagnosis reports) and images (e.g., MRI and CT). The aforementioned data would belong to the lowest layer of the suggested system architecture.

B. Graphical Environment

The graphical environment enables users to specify each EC in a human readable form along with its formal representation using the EC model. For this purpose, the user interface consists of two separate (but interlinked) components named EC Management and Criterion Formal Expression. The first component enables users to manage the specified I/E criteria as well as specify new ones providing a human readable description along with its formal expression. In the latter case, a second component is used that enables users to formally express the conditions that the data should satisfy, based on the terminology specified in the Reference Model and the International Classification Systems and Codifications (Vocabularies) selected.

Concerning the formal expression of EC, the graphical environment (Fig. 2) is totally based on the classes and properties specified in the EC model, the Controlled set of terms used (based on existing Vocabularies) and Natural Language Processing for expressing the conditions that each parameter should satisfy in human understandable form. Especially for the expression of temporal constrains, the implemented system enables users to type on their own the desired period of time, while it automatically detects the internal parameters of the expression (i.e., anchor, before/during/after operators and period of time duration).

C. Background Mechanisms

The system developed for accessing the data recorded within a specific healthcare entity for patient selection purposes consists of an abstract mechanism that uses the services provided by a specific datasource for the evaluation of the EC specified. The *workflow* for patient selection is to a great extent static in terms of internal steps followed, which have been specified in advance. In particular, for detecting either the number of patients or the actual individual patients (i.e., the distinctive IDs that uniquely determine a patient within a specific healthcare entity) satisfying a set of eligibility criteria, the next three steps should be followed.

- Step 1: Detect Patient Unique IDs

Initially, all available patients must be retrieved. For this purpose, a service has been implemented that provides a List with the Unique IDs of patients that have visited a specific healthcare entity (no input argument is necessary).

- Step 2: Examine Eligibility Criteria Specified

Accordingly, the data recorded for each of the available patients must be examined in order to find those who satisfy all the EC specified. For this purpose, a novel mechanism has been implemented which dynamically detects the service(s) that should be used for the evaluation of each of the criteria specified. All of the available services have exactly the same interface. More precisely, they all receive as input a patient ID along with a JSON message containing the parameters specified for a particular criterion, and return whether the patient satisfies this criterion or not along with any additional information that may be necessary. For instance, in case the terminology used for the EC is different in the particular healthcare entity, the system undertakes the required transformations and informs the end user about the relevant processes involved and changes made (e.g., replacement of a terms with another one with the same or similar meaning) for the application of the particular criterion on patient data from the specific healthcare entity.

For terminology alignment purposes, a service has been developed that uses as input an initial term along with its stemming and the target/desired coding systems, and searches for term/terms in the target coding system with the same or similar meaning. The service provides the relevant term (or even a combination of terms using “AND”/”OR” operators) along with the relation to the initial term (e.g., that they have exactly the same meaning or a narrower meaning) based on mapping files. Currently, limited mappings are supported by this service, but it is only a matter of system–configuration-related effort for the functionality of this service to be enriched.

- Step 3: Analyze Patient Data

Finally, the data recorded about each of the patients that meet the specified EC are further processed for query answering purposes. More precisely, the system enables users

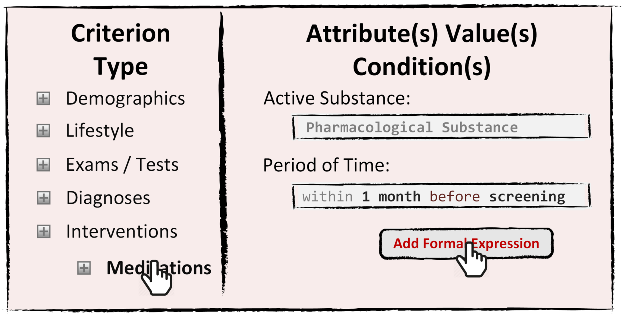

Fig. 2. Graphical Environment for the Formal Expression of a Criterion

to find the number of eligible patients that meet a set of EC (in case of missing data, patients are considered to be eligible) along with their IDs. For each eligible patient the system examines their recorded data for maximizing patient safety and drug efficacy.

IV. DYNAMIC SERVICE DETECTION

A. Overview

In this work, particular focus has been given in the application of the EC on patients' data and, especially, the approach followed and the mechanisms developed for automatic service detection based on the EC specified by the end users. For this purpose some services should be developed in advance for EC evaluation purposes (phase A). Also, for each service a description of the criteria that it can handle should be provided, so that it can be used by the system for the evaluation of a specific criterion (phase B).

Both EC and services should be expressed using the terminology specified in the EC model and, especially, the parameters recorded for each type of criterion. However, it should be noted that in both cases only a subset of the EC attributes available for each type of criterion is being used. Consequently, in case subset 2 (i.e., ontology terms for the service description) encompasses all the elements specified in subset 1 (i.e., ontology terms for the criterion formal expression), then the criterion will be used for filtering the patients. Alternatively, in case subset 2 does not contain all the terms existing in subset 1 (as in the example presented) then some parameters in the formal description of the criterion (in our example, symptoms duration) will be ignored at the criterion application, on condition that there is no other service that can handle all the given data recorded for the user-defined criterion.

B. Service Description and Development

Each of the services implemented for patient selection purposes should have the same signature (i.e., implement the same interface). More precisely, they should receive as input the ID of a patient along with a criterion they should examine. Accordingly, the service should inform the end user whether the patient satisfies this criterion (fully or even partially) or not along with additional information concerning assumptions and/or changes made and/or parameters been ignored (if any). For example, in case different terminologies are used in the patient records about diseases, then during the application of relevant criteria, the service should inform the end user whether each patient satisfies this criterion or not. This information should be accompanied by the transformations made in the terminology used for the formal expression of the criterion (e.g., replacement of an ICD-10 term as in the model with the most relevant ICD-9 term as in the patients' records). The formal description of the service and, especially, the type of criteria it can handle is also important. For this purpose, we have used the terminology employed in the EC model and, more specifically, the parameters recorded for the different types of criteria. In detail, we have specified not only the criterion type (i.e., OWL class) that a service can handle, but also the specific parameters that this service considers during the criterion application on the data.

C. Service Detection and Execution

For EC evaluation purposes, the system examines the user-defined EC and especially their formal expression for detecting the semantic category or categories of each criterion along with the internal parameters specified for each of them. Accordingly, the system searches among the registered EC services for detecting those that can be used for the evaluation of each criterion. During this process it compares the definition of each service with the service description in the ontology in order to find the one(s) that can be used for EC evaluation purposes. In case there is no registered service, then the criterion cannot be used for filtering the patients of the specific healthcare entity. Alternatively, the system uses the most appropriate service from those detected in order to examine whether the patient satisfies the criterion or not, based on the data residing in the data source.

For participating in a clinical study, the patient should satisfy all inclusion criteria and not satisfy any exclusion criteria from those specified, on condition that these criteria were examined. For this purpose, the system additionally informs the end users about the criteria used for filtering patients along with additional information for each one (e.g., changes that were made in the terminology, parameters ignored during the enforcement of a criterion, etc.) based on the responses retrieved from the services used for the evaluation of each criterion.

It should be noted that the system can detect in advance some criteria that would not be used for patients' selection purposes based on the formal description of registered EC evaluation services. For instance, in case there is no service suitable for processing the *amount of tobacco or cigarettes consumed by the patients*, the relevant criteria (if any) will be ignored during the patient selection process. Additionally, in case some of the parameters specified within a single criterion were not used (e.g., the symptoms of the disease in the example presented in Fig. 3) the system can detect the fact that one or more parameters have been ignored and hence inform the end user that the criterion was partially used for filtering patients.

Based on the definition of criteria and, especially, the description of services, the system knows in advance whether a criterion will be used (even partially) or not for all the patients of the specific healthcare entity. However, in some cases, a criterion cannot be evaluated for only a subset of the patients recorded within a specific healthcare entity. For

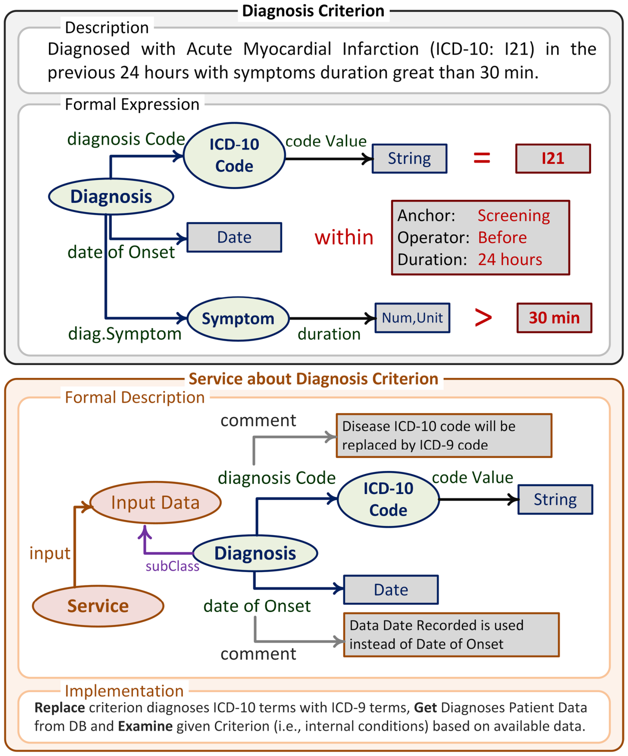

Fig. 3. Service Description and Criterion Formal Expression using EC model

example, let's think of a scenario in which we would like to find patients with the value of a laboratory examination being *greater than twice the upper normal limit* (this is our criterion). In case there is information in the patients' records about the value of the specific laboratory examination, we can quickly answer this question. However, in case there is only information about whether the value of the corresponding laboratory examination was within normal limits (WNL) or not, then we can only say that the patients with examination value "WNL" are within normal limits. In case the value is not "WNL" whether the value may be less than the lower normal limit (LNL), if applicable, or greater than upper normal limit (UNL), but we certainly cannot tell whether the value is more than twice the UNL. The service-oriented approach followed can efficiently deal with such cases and inform the end user about whether each patient satisfies the criterion or not.

V. Discussion and Next Steps

The system presented can efficiently deal with a variety of issues that may be encountered when using a common way for accessing data in patient health records of a specific healthcare entity. However, the development of the specific services requires a considerable amount of time and effort. Additionally, knowledge of Semantic Web technologies and, especially, OWL is necessary for the proper description of each service. On the other hand, the developers are free to use any procedural language for the implementation of each service. Additionally, they can combine data from a variety of sources (if necessary) for EC evaluation purposes.

In the approach presented, the system initially detects all available patients and accordingly examines whether each patient satisfies the EC specified. As a result, the detection of eligible patients often requires a considerable amount of time, as already mentioned. The service that has a crucial role in this process is the service used at the very beginning of the patient selection process for detecting all patients who have visited a specific healthcare entity. If some of the EC specified were used for reducing the amount of patients to be further processed, the total amount of time required for the detection of eligible patients should be significantly reduced. Time could further be reduced by examining criteria in parallel, instead of successively.

In our previous work [14], we have presented a different approach for patient selection purposes. More precisely, for detecting the eligible patients (either the number of them or their IDs) we initially generated a SPARQL query which specifies the information that we would like to retrieve from the data source along with all the EC specified (within the WHERE clause of the SPARQL query). This query was accordingly rewritten and finally translated to the corresponding SQL query taking into account the data structure and terminology used in the relational database. For this purpose, both the ontological representation of the database as well as the specification of mappings among the terms of EC and DB ontologies were needed. Data retrieved from the execution of the rewritten SPARQL query through a D2R server were further processed in order for the data retrieved to be aligned with the terminology used within PONTE project [2] for EC expression.

The approach followed enabled users to easily link a new data source with the system without the necessity to develop a new service. More precisely, for linking the data source the DB ontology had to be created and accordingly the mapping with the EC ontology terms had to be specified through a semi-automatic process, by using tools and components developed for this purpose [15, 16]. In many cases however, this approach cannot efficiently deal with missing data, since the evaluation of SQL queries is based on the information existing in the data source (closed world assumption). Additionally, in cases that the data are scattered among several data sources, the development (or at least use) of additional components is necessary, which perplexes the evaluation of rewritten SPARQL queries [17], especially when these data sources do not follow the same format, as e.g. a Relational Database and an XML documents.

The seamless combination of the two aforementioned systems could reduce the limitations of each approach while maintaining the advantages of each one. More precisely, a query rewriting approach could be used for the transformation of some criteria to queries towards eligible patients' selection in the first step. The query rewriting approach could also be used, where possible, for the implementation of one or more services for EC evaluation purposes, while these services could be defined in such a way that the simultaneous evaluation of more than one EC would be feasible. The combination of the two approaches for patient selection purposes, despite the fact that would increase the complexity of the overall system, would enable users to efficiently deal with a variety of issues that may be encountered during the patient selection process with the least human effort necessary.

Both systems make the necessary interventions in the EC specified and accordingly use them (either each criterion separately or all of them together) for detecting eligible patients. Another approach would be to process the patient data (stemming from various health entities, if this is the case) and create a new data source which would be quite close to the models and terminologies used for EC evaluation purposes. This approach reduces both the complexity and execution time of the background mechanisms. However, a lot of human effort would be required for the development of such a data source and especially the data transformations, which is a crucial step during the data Extraction Transformation and Load (ETL) process [18]. Nevertheless, even in the case of the design of a new repository within a specific healthcare entity (data warehouse), dynamic service detection is necessary in order to efficiently deal with both missing data as well as for further processing the information available for each patient.

Concerning the Graphical Environment used for EC representation and formal expression, in this work, we have clearly separated EC management from their Formal Expression. In the latter case, the highly interactive environment being developed for the formal description of each criterion (based on the terms specified in the EC model and selected vocabularies) can be used in many different clinical studies, since the EC model itself is quite abstract so it can support a wide range of clinical studies [5]. Even so, any required change, and especially the introduction of classes and parameters, is directly reflected in the web interface without any additional development needed.

Another feature that we highlighted in this work is the use of Natural Language Processing (NLP) for the formal expression of the internal conditions of a single criterion, rather than for the formal expression of the whole criterion. Following this approach we have avoided the erroneous interpretation of EC (due to their complexity) and at the same time accelerated the EC formal expression specification, since we have reduced the complexity of the web interface and the number of user clicks (or intervention) being necessary.

VI. Conclusion

Automated patient selection for facilitating clinical trial recruitment accelerates the process with direct impact in the study feasibility, duration and overall cost. In this work we have presented a novel approach based on dynamic service detection. For accessing the patients' data as recorded within a specific healthcare entity, the development of a series of services as well as their proper semantic description are essential. This process, as currently implemented, is to a great extent manual, while the evaluation of the EC often requires a considerable amount of time which is analogous to the amount of both available patients and EC. Nevertheless, using dynamic service detection the system can efficiently deal with a variety of issues that may be encountered in a real case scenario. The total amount of time required for linking and accordingly accessing patients' data can be significantly reduced by combining this new approach with the tools and mechanisms that we developed in the past, which enable the transformation of eligibility criteria to queries towards patient records as well as rewriting and translating of queries for overcoming a variety of heterogeneity issues. We believe that the combined system which emerges with the seamless integration of these features can efficiently and effectively deal with a variety of difficulties related with automated patient selection and have positive impact on clinical trial viability.

Acknowledgment

This work has been partially funded by the European Commission's activity of the Horizon 2020 project HarmonicSS [19] under contract number 731944.This paper expresses the opinions of the authors and not necessarily those of the European Commission. The European Commission is not liable for any use that may be made of the information contained in this paper.

References

[1] J. Ross, S. W. Tu, S. Carini, and I. Sim, "Analysis of Eligibility Criteria Complexity in Clinical Trials", in Summit on Trans. Bioinform., 2010, pp. 46–50.

[2] Efficient Patient Recruitment for Innovative Clinical Trials of Existing Drugs to other Indications, available at http://www.ponte-project.eu

[3] Clinical Data Interchange Standards Consortium (CDISC), available at https://www.cdisc.org/

[4] Health Level Seven International (HL7), available at http://www.hl7.org/

[5] E. Chondrogiannis, V. Andronikou, A. Tagaris, E. Karanastasis, T. Varvarigou, and M. Tsuji, "A novel semantic representation for eligibility criteria in clinical trials", in J Biomed Inform., vol. 69, 2017, pp. 10-23.

[6] SPARQL Query Language for RDF, available at https://www.w3.org/TR/rdf-sparql-query/

[7] JavaScript Object Notation (JSON), available at https://www.json.org/

[8] Web Ontology Language, available at https://www.w3.org/TR/owl-ref/

[9] P. Shvaiko, and J. Euzenat, "Ontology Matching: State of the Art and Future Challenges", in IEEE Trans. Kn. Dat. Eng., vol. 25, 2013, pp. 158-176.

[10] C. Bizer, and Cyganiak, R., "D2R Server – Publishing Relational Databases on the Semantic Web", Poster at ISWC conf., 2006.

[11] J. Rao, and X.Su, "A Survey of Automated Web Service Composition Methods", In Proc. of SWSWPC Conf., 2005. pp. 43-54.

[12] International Statistical Classification of Diseases and Related Health Problems (ICD), available at http://www.who.int/classifications/icd/en/

[13] Anatomical Therapeutic Chemical (ATC) Classification System, available at https://www.whocc.no/atc_ddd_index/

[14] E. Chondrogiannis, V. Andronikou, K. Mourtzoukos, A. Tagaris, and T. Varvarigou, "A novel query rewriting mechanism for semantically interlinking clinical research with electronic health records", in Proc. WIMS'12 Conf., Craiova, Romania, 2012, Article 48, pp. 12.

[15] E. Chondrogiannis, V. Andronikou, E. Karanastasis, and T. Varvarigou, "An intelligent ontology alignment tool dealing with complicated mismatches," in Proc. SWAT4LS 2014 Workshop, Berlin, 2014.

[16] E. Chondrogiannis, V. Andronikou, E. Karanastasis, and T. Varvarigou, "An Advanced Query and Result Rewriting Mechanism for Information Retrieval Purposes from RDF datasources," in Proc. KESW Conf., Moscow, Russia, 2015, pp. 32-47.

[17] O. Hartig, C. Bizer, and J.C. Freytag, "Executing SPARQL queries over the web of linked data", In Proc. of ISWC Conf., 2009, pp. 293-309.

[18] P. Vassiliadis, "A Survey of Extract–Transform–Load Technology", In IJDWM, Vol. 5(3), 2009, pp. 1-27.

[19] HARMONIzation and integrative analysis of regional, national and international Cohorts on primary Sjögren's Syndrome (pSS) towards improved stratification, treatment and health policy making, available at http://harmonicss.eu/