Developing A Semantic Web-based Framework for Executing the Clinical Quality Language Using FHIR

Guoqian Jiang[1], Eric Prud'Hommeax[2], and Harold R. Solbrig[1]

[1] Mayo Clinic, Rochester, MN, 55905, USA
[2] W3C/MIT, Cambridge, MA 02142 USA
jiang.guoqian@mayo.edu

Abstract. The Clinical Quality Language (CQL) is an HL7 specification, aiming to provide a human-readable language to define clinical quality measures and decision support rules while it expresses logic independent of any specific data models (e.g., Quality Data Model, HL7 Fast Healthcare Interoperability Resources – FHIR). FHIR adopts RDF as its third representation/interchange format in addition to XML and JSON, and uses Shape Expressions (ShEx) schema to standardize and validate FHIR RDF graphs. In this presentation, we propose a Semantic Web-based framework that enables the execution of CQL using FHIR RDF technologies. The framework comprises the following four modules: 1) a CQL2SPARQL transformation module; 2) a value set service module; 3) a FHIR RDF transformation module; and 4) a ShEx-based data validation module. We implemented a prototype to demonstrate the utility of the framework and discussed the challenges and ongoing tasks.

Keywords: Clinical Quality Language, FHIR, RDF.

1 Introduction

Future advances in translational and precision medicine research will be increasingly dependent on the creation of patient cohorts encompassing both highly detailed phenotypic and molecular data. The growth of electronic health records (EHRs) has been recognized as a viable and efficient model for enabling translational and precision medicine research [1]. The distinct advantages of EHRs include cost efficiency, large amounts of available clinical data, and the ability to analyze data over time. However, these data are highly complex and frequently inaccurate or missing. The healthcare informatics community faces huge methodological and computational challenges in repurposing EHRs for translational and precision medicine research, typically using standards-based data normalization for effective data integration and accurate phenotyping.

The Clinical Quality Language (CQL) is a HL7 specification, aiming to provide a human-readable language to define clinical quality measures and decision support rules. One of key feature is that it makes logic expressions independent of any specific data models. Such data models include the Quality Data Model (QDM) and HL7 Fast Healthcare Interoperability Resources (FHIR). As the creation of EHR-driven phenotype algorithms shares many common requirements with the definition of clinical quality measures/clinical decision support rules, there are emerging interests in the clinical research informatics communities to explore the CQL as a tool for the standard representation and execution of phenotype algorithms (i.e., structured selection criteria designed to produce research-quality phenotypes).

HL7 FHIR is an emerging next-generation standards framework for the exchange of electronic healthcare data. Many major EHR vendors (e.g., Epic, Cerner) and healthcare providers (e.g., Mayo Clinic, Intermountain Healthcare) are adopting the FHIR standard to enable expanded information sharing for electronic health records through a collaborative initiative known as the HL7 Argonaut project. FHIR also adopted RDF as its third representation/interchange format in addition to XML and JSON, and uses the Shape Expressions (ShEx) schema to validate FHIR RDF clinical data. This enables promising opportunities for the clinical research informatics communities to leverage existing semantic tools developed in the Semantic Web communities for standards-based data integration and discovery. The objective of this study is to propose a Semantic Web-based framework for executing the CQL using FHIR, with the ultimate goal of producing a next-generation phenotyping infrastructure of EHRs.

2 Related Work

CQL has been actively developed by the ONC's Standards and Interoperability (S&I) Clinical Quality Framework focusing on two main use cases: electronic Clinical Quality Measurement (eCQM) and Clinical Decision Support (CDS). As of September 2017, the Centers for Medicare & Medicaid Services (CMS) and the ONC are working on converting the QDM-based implementation of eCQMs to the CQL-based implementation. A CDS Connect project has been initiated to demonstrate how evidence-based care can be more rapidly incorporated into clinical practice through interoperable decision support, in which CQL is adopted for representing decision support rules. The phenotype execution and modeling architecture (PhEMA: http://projectphema.org) consortium has been looking into the standard representation and execution of phenotype algorithms using QDM and related standards [2].

The Yosemite Manifesto positions RDF as the "Universal Healthcare Exchange Language" and asserts that healthcare information should either have a "standard mapping to RDF" or should already be in an RDF format. The RDF representation of FHIR data follows the principle of fully "round trippable", meaning that an instance of a FHIR resource can be serialized and transformed between the XML, JSON or RDF formats. The HL7 ITS/W3C RDF Task Force has made initial decisions on the structure of FHIR RDF graphs, and our team at Mayo Clinic reviewed the existing FHIR RDF decision documents and created a minimal set of elements and defined the

FHIR ShEx transformation rules for the elements. We implemented the FHIR ShEx transformation tools and evaluated the utility of the FHIR ShEx schemas leveraging the ShEx validation tools developed in the ShEx community [3]. The ShEx schemas and tooling have been adopted by the FHIR for its STU3 release.

3 System Architecture

Fig. 1 shows a Semantic Web-based framework of executing the CQL using the FHIR RDF technologies. The framework comprises the following four modules: 1) a CQL2SPARQL transformation module; 2) a value set service module; 3) a FHIR RDF transformation module; and 4) a ShEx-based data validation module.

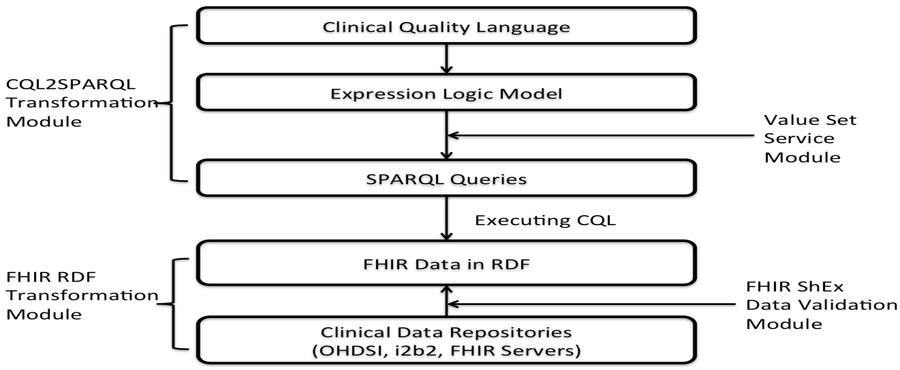

Fig. 1**.** The system architecture for executing CQL using the FHIR RDF technologies.

The CQL2SPARQL transformation module. CQL is designed to cover three levels of representations. In the conceptual level, CQL is defined for authors to produce libraries containing human-readable yet precise logic. In the logic level, the Expression Logic Model is used for machine-friendly rendering of the CQL logic. In the physical level, different implementation environments will be leveraged to execute the ELM, in which the translation from ELM to target environment language is needed. In this module, as our target environment language is the SPARQL query language, a CQL2SPARQL transformation tool is needed for the CQL execution. Although the ANTLR parse tree can be used as a direct representation of the input CQL, the representation in ELM, which is focused on implementation, can provide the CQL parsing supports for type verification, type inference and operator resolution, and processing higher level constructs like timing phrases. The CQF community has released a complete implementation of the CQL-to-ELM Translator, the APIs of which is utilized for parsing and processing CQL. The Jena SPARQL APIs is used for building target SPARQL queries.

The value set service module. Value set is one of key CQL constructs, specifying that logic within the CQL library may reference the specified value set by the give name. For instance, the statement [*valueset "Antibiotic Medications":*

'2.16.840.1.113883.3.464.1003.196.12.1001'] means the value set of antibiotic medications can by referenced by its name and resolved by its OID through external value set services. By default, the Value Set Authority Center (VSAC) value set services are recommended. In this module, a tool is developed to invoke the external VASC services and retrieve the coded values and associated metadata required for the SPARQL query construction.

The FHIR RDF transformation module. As mentioned above, FHIR RDF has become part of the official FHIR STU3 release and a FHIR RDF transformation tool has been incorporated into the FHIR building toolkit. Our team at Mayo Clinic is actively developing tools for transforming either XML or JSON representations to RDF. In addition, large volume of clinical data are stored in the relational databases and the tools (e.g., Ontop) that support the executing SPARQL queries over relational databases can be leveraged to transform the relational data into the FHIR-based RDF representation.

The ShEx-based data validation module. ShEx is a constraint language for formally describing RDF structures and can serve the same role with RDF as that of XML schema to XML. Validating an RDF node against a shape tests the adjacent nodes against the constraints in the shape. As mentioned above, the FHIR ShEx schemas have become part of the official FHIR STU3 release. The FHIR ShEx transformation and validation tools will be leveraged and extended to support the data validation needs in executing the CQL.

4 Prototype Implementation

We are actively implementing each module of the framework for executing the CQL-based phenotype algorithms using FHIR RDF technology. We first examined available FHIR-based CQL examples and identified a number of patterns for building the SPARQL queries. **Fig. 2** shows an example SPARQL pattern identified. The VSAC value set services were invoked to retrieve the codes for two value sets "Acute Pharyngitis" and "Acute Tonsillitis", in which 17 and 19 codes (from the code systems SNOMED CT and ICD-10-CM) were retrieved respectively. A Java-based program is developed for transforming identified patterns from the CQL to SPARQL while invoking the VSAC value set services to retrieve coded values in a value set.

For the FHIR RDF transformation tools, we have implemented a stand-alone tool to convert FHIR resources in the JSON format to their equivalent in the FHIR RDF format (https://github.com/BD2KOnFHIR/fhirtordf). We are also developing an FHIR-based data access framework to enable exposing the clinical data stored in the OHDSI CDM-based data repositories in the FHIR RDF format through leveraging an ontology of the FHIR metadata vocabulary and an open-source Ontology-based Data Access (OBDA) system known as Ontop. A test query generated from **Fig. 2** was run successfully against the FHIR-Ontop-OHDSI platform and 1640 patients were identified.

As the clinical data may come from heterogeneous data sources, the FHIR ShEx schemas play an important role on ensuring that the data sources are normalized as specified to support meaningful data retrieval.

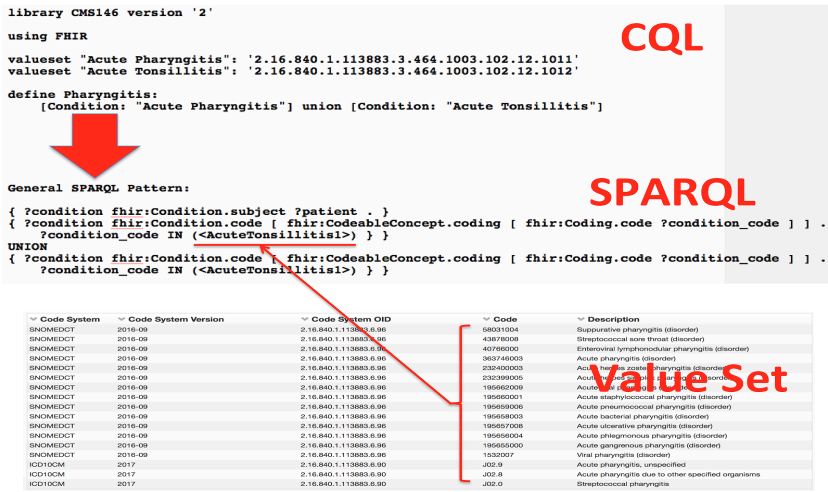

Fig. 2. An example pattern identified for building a SPARQL query.

5 Discussion and Conclusion

In this study, we proposed a framework of executing CQL-based phenotype algorithms using FHIR RDF technologies and demonstrated the feasibility of using CQL and FHIR in support of EHR-driven phenotype algorithm creation and execution. It turns out that temporal patterns (e.g., ages, intervals) are commonly used in CQL, and existing SPARQL functions may need to be extended to capture and execute such patterns effectively. While CQL uses the basic expression definition capabilities defined by the FHIRPath for its core expression terms, we noticed that there are some discrepancies between the FHIRPath expressions and the FHIR RDF model which need to be harmonized through a community-based collaboration in the future.

Acknowledgement: This study is supported in part by NIH grants U01 HG009450, U01 CA180940, and R01 GM105688.

References

1. Hripcsak G, Albers, DJ.: Next-generation phenotyping of electronic health records. J Am Med Inform Assoc. 2013 Jan 1;20(1):117-21.
2. Jiang G, Kiefer RC, Rasmussen LV, Solbrig HR, Mo H, Pacheco JA, Xu J, Montague E, Thompson WK, Denny JC, Chute CG, Pathak J.: Developing a data element repository to support EHR-driven phenotype algorithm authoring and execution. J Biomed Inform. 2016 Aug;62:232-42.
3. Solbrig HR, Prud'hommeaux E, Grieve G, McKenzie L, Mandel JC, Sharma DK, Jiang G.: Modeling and validating HL7 FHIR profiles using semantic web Shape Expressions (ShEx). J Biomed Inform. 2017 Mar;67:90-100.