

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

Guoqian Jiang¹, Eric Prud'Hommeaux², Guohui Xiao³, and Harold R. Solbrig¹

¹ Mayo Clinic, Rochester, MN, 55905, USA

² W3C/MIT, Cambridge, MA 02142, USA

³ Free University of Bozen-Bolzano, Bolzano, 39100, Italy
jiang.guoqian@mayo.edu

Abstract. 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 while it makes logic expressions independent of any specific data model (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 the Shape Expressions (ShEx) schema to standardize and validate the FHIR RDF graphs. In this presentation, we propose a Semantic Web-based framework that enables the execution of CQL using the FHIR RDF technologies. The framework comprises the following four modules: 1) a CQL-to-SPARQL 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 research¹. The distinct advantages of EHRs include cost efficiency, large amounts of available clinical data, and the ability to analyze data over time, however, the data are highly complex, inaccurate and frequently missing. The healthcare informatics community faces huge methodological and computational challenges in repurposing EHRs for translational and precision medicine research, with respect to standards-based data normalization, effective data integration and accurate phenotyping.

The Clinical Quality Language (CQL)^{2,3} is a HL7 specification providing a human-readable language to define clinical quality measures and decision support rules. 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) driven by EHRs. Notably, the phenotype execution and modeling architecture (PhEMA)⁴ consortium has been looking into the standard representation and execution of phenotype algorithms using CQL and related standards⁵.

HL7 Fast Healthcare Interoperability Resources (FHIR)⁶ is an emerging next-generation standards framework for the exchange of electronic healthcare data. FHIR adopted the Resource Description Framework (RDF) as its third representation/interchange format in addition to XML and JSON, and uses the Shape Expressions (ShEx) schema to standardize the structure of FHIR RDF graphs⁷. The HL7 ITS/W3C RDF Task Force has made initial decisions how the FHIR RDF graphs look like, and our team at Mayo Clinic reviewed the existing FHIR RDF decision documents and created a minimal set of elements, 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 W3C ShEx community⁷. 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 phenotype algorithm definition. The objective of this study is to propose a Semantic Web-based framework for executing the CQL using FHIR to enable a next-generation EHR-driven phenotyping infrastructure.

2 System Architecture

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

The CQL-to-SPARQL 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 (ELM) 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 CQL-to-SPARQL transformation tool is needed for the CQL execution. The representation in ELM can provide the CQL parsing supports for type verification, type inference and operator resolution, and process higher-level constructs like temporal patterns. There is a complete implementation of the CQL-to-

ELM Translator³, the API of which is utilized for parsing and processing CQL. The Jena SPARQL API⁸ is used for building target SPARQL queries.

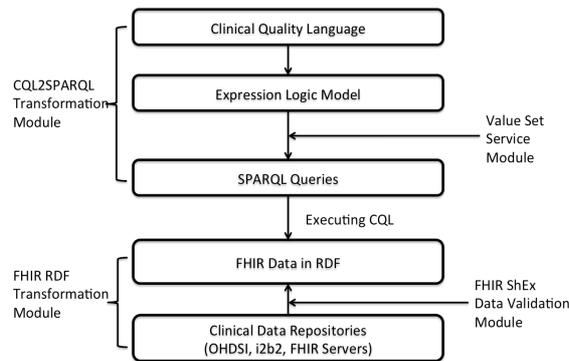


Fig. 1. The system architecture for executing CQL using the FHIR RDF technologies.

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 given 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 be referenced by its name and resolved by its OID through external value set services. By default, the Value Set Authority Center (VSAC)⁹ value set services are recommended. In this module, a tool is developed to invoke the external VSAC services and retrieve the coded values and associated metadata required for the SPARQL query construction.

The FHIR RDF transformation module. As mentioned above, the 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 that help the FHIR RDF transformation from either XML or JSON formats. In addition, large volumes of clinical data are stored in the relational databases and the tools (e.g., Ontop¹⁴) that support the execution of 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, FHIR ShEx schemas have become part of the official FHIR release. The FHIR ShEx transformation and validation tools will be leveraged and extended to support the data validation needs in executing the CQL.

3 Prototype Implementation

We are actively implementing each module of the framework for executing the CQL-based phenotype algorithms using the FHIR RDF technology. We have first examined available FHIR-based CQL examples and identified a number of patterns for building the SPARQL queries. **Fig. 2** shows such a SPARQL pattern. 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 was developed for transforming identified patterns from 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 format¹⁰. We are also developing a 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 Ontop^{11,14}, in which an OHDSI virtual machine¹² is used to provide sample clinical research data for the testing purpose. A test query generated from **Fig. 2** was run successfully against the FHIR-Ontop-OHDSI platform and 1640 patients were identified. The results were verified accurate using the OHDSI cohort identification tool.

```

library CMS146 version '2'

using FHIR

valueset "Acute Pharyngitis": '2.16.840.1.113883.3.464.1003.102.12.1011'
valueset "Acute Tonsillitis": '2.16.840.1.113883.3.464.1003.102.12.1012'

define Pharyngitis:
  [Condition: "Acute Pharyngitis"] union [Condition: "Acute Tonsillitis"]

```

↓

General SPARQL Pattern:

```

{ ?condition fhir:Condition.subject ?patient . }
{ ?condition fhir:Condition.code [ fhir:CodeableConcept.coding
  [ fhir:Coding.code ?condition_code ] ] .
  ?condition_code IN (<AcutePharyngitis> ) }
UNION
{ ?condition fhir:Condition.code [ fhir:CodeableConcept.coding
  [ fhir:Coding.code ?condition_code ] ] .
  ?condition_code IN (<AcuteTonsillitis> ) }

```

CQL

SPARQL

Value Set

Code System	Code System Version	Code System OID	Code	Description
SNOMEDCT	2016-09	2.16.840.1.113883.6.96	58231024	Suppurative pharyngitis (disorder)
SNOMEDCT	2016-09	2.16.840.1.113883.6.96	43878008	Streptococcal sore throat (disorder)
SNOMEDCT	2016-09	2.16.840.1.113883.6.96	40786000	Enteroviral lymphonodular pharyngitis (disorder)
SNOMEDCT	2016-09	2.16.840.1.113883.6.96	363746003	Acute pharyngitis (disorder)
SNOMEDCT	2016-09	2.16.840.1.113883.6.96	232400003	Acute streptococcal pharyngitis (disorder)
SNOMEDCT	2016-09	2.16.840.1.113883.6.96	232398005	Acute viral pharyngitis (disorder)
SNOMEDCT	2016-09	2.16.840.1.113883.6.96	195662009	Acute streptococcal pharyngitis (disorder)
SNOMEDCT	2016-09	2.16.840.1.113883.6.96	195660001	Acute staphylococcal pharyngitis (disorder)
SNOMEDCT	2016-09	2.16.840.1.113883.6.96	195659006	Acute pneumococcal pharyngitis (disorder)
SNOMEDCT	2016-09	2.16.840.1.113883.6.96	195658003	Acute bacterial pharyngitis (disorder)
SNOMEDCT	2016-09	2.16.840.1.113883.6.96	195657008	Acute ulcerative pharyngitis (disorder)
SNOMEDCT	2016-09	2.16.840.1.113883.6.96	195656004	Acute phlegmonous pharyngitis (disorder)
SNOMEDCT	2016-09	2.16.840.1.113883.6.96	195655000	Acute gangrenous pharyngitis (disorder)
SNOMEDCT	2016-09	2.16.840.1.113883.6.96	1525207	Viral pharyngitis (disorder)
ICD10CM	2017	2.16.840.1.113883.6.90	J02.9	Acute pharyngitis, unspecified
ICD10CM	2017	2.16.840.1.113883.6.90	J02.8	Acute pharyngitis due to other specified organisms
ICD10CM	2017	2.16.840.1.113883.6.90	J02.0	Streptococcal pharyngitis

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

4 Discussion and Conclusion

In this study, we proposed a framework for 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 FHIRPath¹³ for its core expression terms, we noticed that there are some discrepancies between FHIRPath expressions and FHIR RDF decisions, which needs 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;20(1):117-21.
2. HL7 Cross-Paradigm Specification: Clinical Quality Language, Release 1, 2017.
3. The Clinical Quality Language GitHub Site 2017 [Available from: https://github.com/cqframework/clinical_quality_language].
4. The phenotype execution and modeling architecture (PhEMA) 2017 [Available from: <http://projectphema.org>].
5. Jiang G, Kiefer RC, Rasmussen LV, Solbrig HR, Mo H, Pacheco JA, et al. Developing a data element repository to support EHR-driven phenotype algorithm authoring and execution. *J Biomed Inform.* 2016;62:232-42.
6. HL7 Fast Healthcare Interoperability Resources (FHIR) 2017 [Available from: <http://hl7.org/fhir/>].
7. Solbrig HR, Prud'hommeaux E, Grieve G, McKenzie L, Mandel JC, Sharma DK, et al. Modeling and validating HL7 FHIR profiles using semantic web Shape Expressions (ShEx). *J Biomed Inform.* 2017;67:90-100.
8. ARQ - A SPARQL Processor for Jena 2017 [Available from: <https://jena.apache.org/documentation/query/>].
9. The Value Set Authority Center (VSAC) 2017 [Available from: <https://vsac.nlm.nih.gov/>].
10. The FHIR-to-RDF Transformation Tool 2017 [Available from: <https://github.com/BD2KOnFHIR/fhirtordf>].
11. Jiang G, Xiao GH, Kiefer RC, Prud'hommeaux E, Solbrig HR, editors. Building an FHIR Ontology based Data Access Framework with the OHDSI Data Repositories. *AMIA 2017 Annual Symposium* (in press); 2017; Washington DC.
12. The OHDSI Virtual Machine 2017 [Available from: <http://www.ohdsi.org/web/wiki/doku.php?id=documentation:software:ohdsi-in-a-box>].
13. FHIRPath 2017 [Available from: <http://hl7.org/fhirpath/>].
14. Calvanese D, Cogrel B, Komla-Ebri S, Kontchakov R, Lanti D, Rezk M, Rodriguez-Muro M, and Xiao GH. Ontop: Answering SPARQL Queries over Relational Databases. In: *Semantic Web Journal.* 2017; 8.3:471-487