MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI230922

Structured Queries to AQL: Querying OpenEHR Data Leveraging a FHIR-Based Infrastructure for Federated Feasibility Queries

Lorenz ROSENAU[a,1] and Josef INGENERF[a]
[a] IT Center for Clinical Research, University of Lübeck, Lübeck, Germany
ORCiD ID: Lorenz Rosenau https://orcid.org/0000-0002-8614-2792

Abstract. In digital healthcare, data heterogeneity is a reoccurring issue caused by proprietary source systems. It is often overcome by utilizing ETL processes resulting in data warehouses, which ensure common data models for interoperability. Unfortunately, the achieved interoperability is usually limited to an institutional level. The broad solution space to achieve interoperability with different health data standards is part of the problem, resulting in different standards used at various institutions. For cross-institutional use cases like federated feasibility queries, the issue of heterogeneity is reintroduced. This work showcases how the existing German infrastructure for federated feasibility queries based on Hl7 FHIR can be extended to support openEHR without further data transformation. By utilizing an intermediate query format that can be transferred to FHIR Search, CQL, and AQL.

Keywords. Federated queries, feasibility study, openEHR, AQL, FHIR, CQL

1. Introduction

Clinical trials require a sufficient sample size to reach statistical significance. Feasibility queries allow researchers to estimate the potential patient count for their studies. For multicentric studies, executing federated feasibility requires the means to execute the same query in the heterogeneous environment of clinical data warehouses.

In Germany, the Medical Informatics Initiative (MII) currently develops the infrastructure and a central portal for federated feasibility queries over all German university hospitals [1]. Their work is based on the long-term development plan of the MII, the establishment of so-called data integration centers (DICs) at each university clinical site. Within DICs, data from clinical source systems are stored and made accessible based on harmonized standards. Four consortia (DIFUTURE, HiGHmed, MIRACUM, SMITH) have been established and are commissioning their DICs. Despite the common goal of making clinical data available across institutional boundaries, each consortium chose its common data model to represent the model-agnostic medical concepts of the MII core data set [2], causing syntactic heterogeneity.

[1] Corresponding Author: Lorenz Rosenau, Phone: +4945131015636, email: lorenz.rosenau@uni-luebeck.de.

The HiGHmed consortia chose the openEHR standard, while HL7 FHIR was chosen for cross-consortia use cases. OpenEHR utilizes standardized computable models of clinical information (archetypes) to enable interoperability and has its query language called Archetype Query Language (AQL). FHIR, on the other hand, defines Resources to express clinical data and makes them accessible through RESTFul API and its query languages FHIR Search and Clinical Query Language (CQL). Despite their differences, for the sake of this paper, we can consider both as common data models with different query formats, but both represent identical healthcare concepts.

With an established infrastructure for FHIR data [3], the cross-consortia use case of federated feasibility queries poses the challenge of making the openEHR data accessible.

The solution space for this problem offers two main approaches:

- To establish extract, transfer, and load (ETL) processes to convert the data available in the DIC to the FHIR standard.
- By converting the incoming requests to query languages supported by the DIC.

The first approach is technically feasible. The software toolkit FHIRBridge2 allows a rule-based translation from openEHR templates to FHIR resources. The commercial provider better also presents a similar approach for their openEHR CDR (Common Data Model) [4]. However, to make the data accessible, a secondary data warehouse for FHIR in the DIC needs to be established to allow feasibility queries based on FHIR Search or CQL. Besides the additional hardware costs and maintenance, this leads to data redundancy.

The second approach uses the existing data format and creates a translation, not for the data itself but for the incoming query. In this work, we analyze the feasibility of expanding the established federated feasibility platform of the MII with a translation component to convert its query format to AQL.

2. Methods

2.1. Adding AQL support to the existing infrastructure for federated feasibility queries

Based on the existing infrastructure (Figure 1) for federated feasibility queries, we analyzed the requirements to extend its capabilities to support AQL.

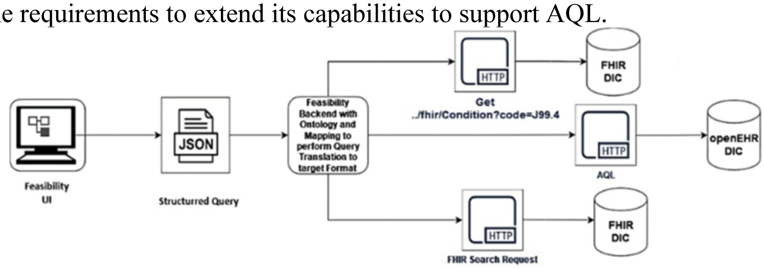

Figure 1. Existing feasibility architecture extended with AQL support.

The established architecture consists of a generic user interface, a backend responsible for providing the search ontology and translating the incoming intermediate query format (named Structured Query) from the user interface, and middleware to forward the queries to the DICs where they are executed. The user interface presents a set of simplified criteria representing the complex underlying instance data at the clinical sites. Each

[2] https://github.com/ehrbase/fhir-bridge

criterion has an identifying Coding and an optional value that further specifies the criterion (i.e., the LOINC code 8302-2 identifies the criterion body height with a value that allows the restriction of the height). In case no value is provided, the existence of the criterion is the value (i.e., the ICD10-GM code G47.3 specifies the presence of sleep apnea). Boolean algebra enables the user to create feasibility queries based on the criteria.

The backend uses a so-called Mapping to translate the Structured Query from the front end to the target query language. This Mapping contains all the necessary information to enrich the criteria with details about the underlying FHIR data model, enabling the backend to execute it as FHIR Search or CQL query. All the criteria available in the search ontology and their corresponding Mappings are based on the German core data set and automatically created using the approach described in [5]. As showcased in Figure 1, the criteria are independent of the data format. To utilize the criteria from the FHIR-based ontology for openEHR, we need to enrich the data for each identified criterion by mapping its Coding to its representation in the openEHR data model. Additionally, we must create a translation process that uses this information to generate the AQL query for the openEHR CDR. From a software development perspective, the existing software stack of the feasibility platform is easily extended by adding a Java library for the AQL translation in the backend component.

2.2. Mapping criteria to their openEHR representation

OpenEHR uses so-called archetypes as fundamental building blocks to represent medical concepts. By building upon the archetypes for specific use cases, templates combine and restrict existing archetypes by defining rules on top of them. The templates and the archetypes are shared using the clinical knowledge manager (CKM). For the German core dataset, openEHR templates exist and are available from the HiGHmed CKM[3]. Each archetype within a template is uniquely identifiable with its path and archetype id. To create the Mapping, we must identify the archetype path within the template whose rule restricts the possible Codings. This procedure is also applied to obtain the archetype path of the value. For example, the "KDS_Diagnosis" template defines the structure of a medical condition. Within the template, the archetype for a diagnosis is used and restricted to a value from the ICD-10-GM code system.

We implemented individual handling for each template based on our findings when analyzing the template and archetypes to create a Mapping that holds the information on the archetypes' path necessary for the AQL translation [6].

AQL is a declarative query language that directly uses the archetype path to access data elements in complex data models. OpenEHR provides an Antlr grammar definition for AQL expressions with valid syntax. The most relevant for our use case are the logical and comparison operators. Building upon the grammar definition, we implemented Java classes to represent parts of the grammar. Using a strong typed programming language can ensure the creation of valid queries. We implemented automated processing of the structured query format to AQL using the previously obtained mapping information.

3. Results

The established methodology allowed us to extend the existing software architecture of the central feasibility platform with a translation component for AQL. The Mapping lists

[3] https://ckm.highmed.org/ckm/

archetypes and their types for each criterion based on the composition, using path information to query the data type. Following this representation, if a value can further restrict the concept, valuePathElements, and valueCodePath elements contain the information needed for that value.

```
"key": {"code": "G47.3", "name": "Sleep apnea" "system": ".../bfarm/icd-10-gm", "version": "2022"},
"openEhrType": "DV_CODED_TEXT",
"termCodePath": "/data[at0001]/items[at0002]/value",
"termCodePathElements": [
    {
        "archetypeId": "openEHR-EHR-COMPOSITION.report.v1",
        "openEhrType": "COMPOSITION"
    }, {
        "archetypeId": "openEHR-EHR-EVALUATION.problem_diagnosis.v1",
        "openEhrType": "EVALUATION"
    }
]
```

Figure 2. Excerpt from the *Mapping* for sleep apnea.

Applied to the previous example, the "KDS_Diagnosis" template's composition archetype has an evaluation archetype that specifies the diagnosis with a coded text, resulting in the Mapping shown in Figure 2 for the criterion sleep apnea. The archetype ids provide unique identification for each archetype.

For single criteria, the translation component has all information needed to construct the AQL statement. We query the database for unique EHR IDs containing a specific composition with a diagnosis evaluation archetype. We filter the evaluations to find instances of sleep apnea. Each distinct EHR ID corresponds to a patient in the database.

```
SELECT DISTINCT
e/ehr_id/value FROM EHR ehr_alias
CONTAINS COMPOSITION Composition_alias[openEHR-EHR-COMPOSITION.report.v1]
CONTAINS EVALUATION diagnosis_alias[openEHR-EHR-EVALUATION.problem_diagnosis.v1]
WHERE (diagnosis_alias/data[at0001]/items[at0002]/value/defining_code/code_string MATCHES {'G47.3'} AND
diagnosis_alias/data[at0001]/items[at0002]/value/defining_code/terminology_id/value
MATCHES {'http://fhir.de/CodeSystem/bfarm/icd-10-gm'})
```

Figure 3. AQL Statement to query all patients with sleep apnea diagnosis.

Utilizing the UNION, MINUS, and INTERSECT operators allow the combination of these AQL snippets to create more complex combinations based on the structured query. Using ETL-transferred synthetic data to evaluate our work, we created a test set of 10 synthetic patients in an openEHR CDR. We performed equivalence class tests for all supported templates by performing queries with basic logical combinations and queries with and without exclusion criteria. In all cases, we could use an instance of the existing feasibility platform extended with the sq2aql translation component to identify the correct number of patients in the test openEHR. The translation component and guidance on the integration are available at https://github.com/itcr-uni-luebeck/sq2aql.

4. Discussion

We acknowledge that developing a feasibility query portal is an ill-structured problem. And the solution space is as manifold as the data formats available. Solutions exist for the i2b2 data schema [7], OMOP data [8], and proprietary data stored in SQL format [9]. Further, solutions exist specifically for the openEHR data format. However, due to the existing and established infrastructure, we believe in providing a solution that caters to the needs of the HiGHmed consortia in Germany and can also be adapted for other standards. Furthermore, we firmly believe that our solution of utilizing an intermediate query format can be applied independently of the concepts used in this project.

Besides ours, other solutions exist that could be applied. Fette et al. [10] present the translation of CQL queries to AQL. While [11] outlines the extension of CQL to apply it to an openEHR CDR. Unfortunately, the required software to utilize CQL for openEHR is not made public, and we did not get a response from the author. This work leverages the common concepts between the FHIR and openEHR representation of the core data set. A concept cannot be mapped if only available in either data model. We further require that all value restrictions applied to the concept are represented identically in both data models. I.e., the body height should be expressed using the same unit in both data models. The Operators UNION, MINUS, and INTERSECT are currently not supported by EHRbase. We implemented a client that supports these operations.

The intermediate query format "Structured Query" developed and utilized for this work is proprietary. We hope to adopt an upcoming version of the FHIR Resource EvidenceVariable, allowing the interoperable use of the intermediate query format. We also intend to expand the translation capabilities to other data models like OMOP and I2B2 and make other/more concepts available.

Acknowledgment

We thank Alexander Kiel, whose work on the Structured Query to CQL component greatly influenced the software implementation. The project was funded by the German Federal Ministry of Research (BMBF; Grant 01ZZ1802Z).

References

[1] Prokosch HU, Baber R, Bollmann P, Gebhardt M, Gruendner J, Hummel M. Aligning biobanks and data integration centers efficiently (ABIDE_MI). Stud Health Technol Inform. 2022 May;292:37-42, doi: 10.3233/SHTI220317.

[2] Semler SC, Wissing F, Heyder R. German Medical Informatics Initiative. Methods Inf Med. 2018 Jul;57(S 01):e50-6, doi: 10.3414/ME18-03-0003.

[3] Gruendner J, Deppenwiese N, Folz M, Köhler T, Kroll B, Prokosch HU, Rosenau L, Rühle M, Scheidl MA, Schüttler C, Sedlmayr B, Twrdik A, Kiel A, Majeed RW. The architecture of a feasibility query portal for distributed COVID-19 Fast Healthcare Interoperability Resources (FHIR) patient data repositories: design and implementation study. JMIR Med Inform. 2022 May;10(5):e36709, doi: 10.2196/36709.

[4] Allen A. Introducing FHIR Connect [Internet]. [cited 2022 Oct 26]. Available from: https://news.better.care/introducing-fhir-connect

[5] Rosenau L, Majeed RW, Ingenerf J, Kiel A, Kroll B, Köhler T, Prokosch HU, Gruendner J. Generation of a Fast Healthcare Interoperability Resources (FHIR)-based ontology for federated feasibility queries in the context of COVID-19: feasibility study. JMIR Med Inform. 2022 Apr;10(4):e35789, doi: 10.2196/35789.

[6] medizininformatik-initiative/fhir-ontology-generator [Internet]. [cited 2022 Oct 27]. Available from: https://github.com/medizininformatik-initiative/fhir-ontology-generator.

[7] Weber GM, Murphy SN, McMurry AJ, Macfadden D, Nigrin DJ, Churchill S, Kohane IS. The Shared Health Research Information Network (SHRINE): a prototype federated query tool for clinical data repositories. J Am Med Inform Assoc. 2009 Sep-Oct;16(5):624-30, doi: 10.1197/jamia.M3191.

[8] ATLAS – A unified interface for the OHDSI tools – OHDSI [Internet]. [cited 2022 Feb 10]. Available from: https://www.ohdsi.org/atlas-a-unified-interface-for-the-ohdsi-tools/

[9] Dobbins NJ, Spital CH, Black RA, Morrison JM, de Veer B, Zampino E, Harrington RD, Britt BD, Stephens KA, Wilcox AB, Tarczy-Hornoch P, Mooney SD. Leaf: an open-source, model-agnostic, data-driven web application for cohort discovery and translational biomedical research. J Am Med Inform Assoc. 2020 Jan;27(1):109-18, doi: 10.1093/jamia/ocz165.

[10] Fette G, Kaspar M, Liman L, Ertl M, Krebs J, Dietrich G, Störk S, Puppe F. Query translation between AQL and CQL. Stud Health Technol Inform. 2019 Aug;264:128-32, doi: 10.3233/SHTI190197.

[11] Li M, Zhi Y, Lu X, Cai H. Extending CQL with openEHR to express clinical quality indicators. Stud Health Technol Inform. 2019 Aug;264:1853-4, doi: 10.3233/SHTI190681.