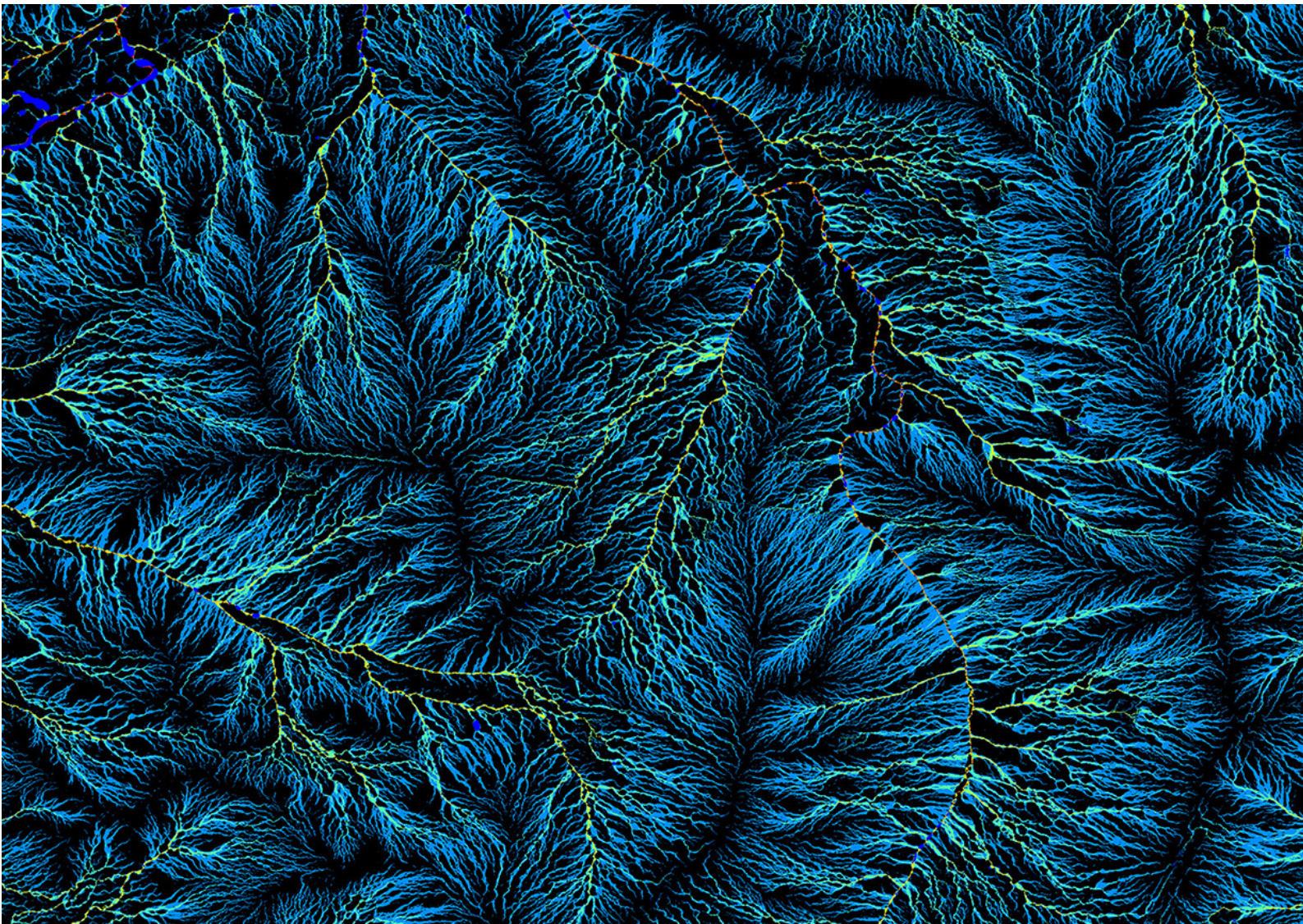
Australia's National
Science Agency

FHIR Terminology Services for OMOP – opportunities report

Michael Lawley, Jim Steel, John Grimes
10 September 2021

Brisbane Diamantina Health (BDHP) Partners



Health & Biosecurity

Citation

Lawley MJ, Steel J and Grimes J (2021) FHIR Terminology Services for OMOP – opportunities report. CSIRO, Australia.

Copyright

Important disclaimer

Contents

Figures

Tables

Acknowledgments

The work described in this report was supported by the Australian Government's Medical Research Future Fund (MRFF) Rapid Applied Research Translation program grant awarded to Brisbane Diamantina Health Partners.

1 Introduction

Health Level Seven International (HL7®) and the Observational Health Data Sciences and Informatics (OHDSI) recently announced a collaboration to address the sharing and tracking of data in the healthcare and research industries across the FHIR and OMOP data models. The organizations will integrate HL7 Fast Healthcare Interoperability Resources (FHIR®) and OHDSI's Observational Medical Outcomes Partnership (OMOP) common data model to achieve this goal.

The broad emerging goals of the work include:

1. Importing bulk data FHIR resources into an OMOP Database; acquisition of FHIR data to build an OMOP database for analysis.
2. Extracting OMOP data in FHIR format as FHIR bulk data (alternatively putting a FHIR server API in front of an OMOP database).
3. SMART-on-FHIR apps using OMOP source data; running FHIR servers with OMOP data (leveraging a FHIR API to an OMOP database)
4. Making FHIR Terminology services available for use by OMOP framework. Additionally, OMOP has many of its own terminology content which could be made accessible to the FHIR community.

There are several projects underway in both Qld and Nationally, to leverage the valuable data sets across healthcare system by transforming EMR data into both FHIR Data Models and OMOP data models to support both real time data analytics and population health research.

One of the components of the harmonised data model, is a standard approach to managing, referencing and accessing the controlled vocabularies which underpin these data models. FHIR Terminology Resources provide a useful approach to managing Code Systems, Value Sets and Concept Maps and FHIR Terminology Operations such as Expand, Validate and Translate provide a standard approach to more readily use and interact with clinical terminologies.

1.1 Goals

OMOPs implementation of clinical terminology support is relatively static and limited compared the terminology resources and capabilities that are supported within FHIR.

The objectives of the project are to:

1. Understand how both the OMOP model and tooling supports the use of standard terminologies,
2. Explore how FHIR Terminology Services can be used within the OMOP tooling – including prototype integration,
3. Identify opportunities within BDHP partners that can use the Qld Clinical Terminology Service to support OMOP related projects, and

4. Identify and scope opportunities for Australian uses of the AHRA Terminology service to support OMOP related projects.

This document focuses on the specific opportunities that exist to integrate FHIR with existing tools, which will enable FHIR-based terminology services to better support OMOP users.

1.2 Terminology in OMOP

OMOP's approach to vocabulary/terminology is based on the use of a standardised database tables for managing codes and code systems. This vocabulary sub-schema is part of the OHDSI Common Data Model (CDM) and can be referenced in queries that are defined for code sets or cohorts. One of the notable characteristics of the schema is that as vocabulary sets are imported into Athena (the platform for managing vocabularies), integer surrogate keys are assign to each code system / code pair. The rest of the CDM uses these surrogate keys in place of the original terminology codes. Because the allocation of these integers is non-deterministic, they are prone to change over time and across different deployments of the OMOP tooling. Additionally, a lookup table must be consulted whenever a query needs to convert from "real" terminology codes, to the internal surrogate keys.

Versioning of code systems in these tables is also fairly simplistic – each code system row in the vocabulary table can have a version associated with it, and each concept then references that code system row. This means that although it is possible to include, for example, multiple editions of SNOMED CT, there is no automatic equivalence between the same codes in each of them; the same code from two different versions of a code system will have distinct and unrelated surrogate keys.

To populate the schema, OMOP designates a number of "standard" code systems for the different domains, where domains encompass large-scale things like Condition, Observation, Procedure or Drugs, or more specific things such as units (of measurement), gender, or condition status. Each code in these code systems can be marked as a "standard" code, meaning that they represent the "official" encoding of the concept (these tend to be codes from US-centric code systems). The import of these standard code system also comes with mappings from non-standard codes to standard codes.

The population of relationships is also included in the OMOP tables. This means it would be possible to construct SQL queries based on some of the modelling in rich terminologies such as SNOMED CT. While this allows for expressing (in SQL) queries similar to those possible with SNOMED's Expression Constraint Language (ECL), it is distinctly less powerful because it does not retain information relating to "role grouping". Hence it may not be possible to determine which body structure a particular morphology applies to, or which substance has a specific strength in a multi-ingredient medication.

At a UI level, OMOP supports the definition of Concept Sets, representing sets of concepts that are of interest for some analytics task or tasks. These concept sets are defined as a set of inclusion and exclusion criteria, which can optionally choose to include subsumed concepts, or mapped concepts. Concept Sets can then be referenced in Cohort Definitions, and either exported or executed as SQL queries as part of the Cohort Definition, as part of the user's analytics tasks.

1.3 Relevant OMOP tooling

1.3.1 Atlas

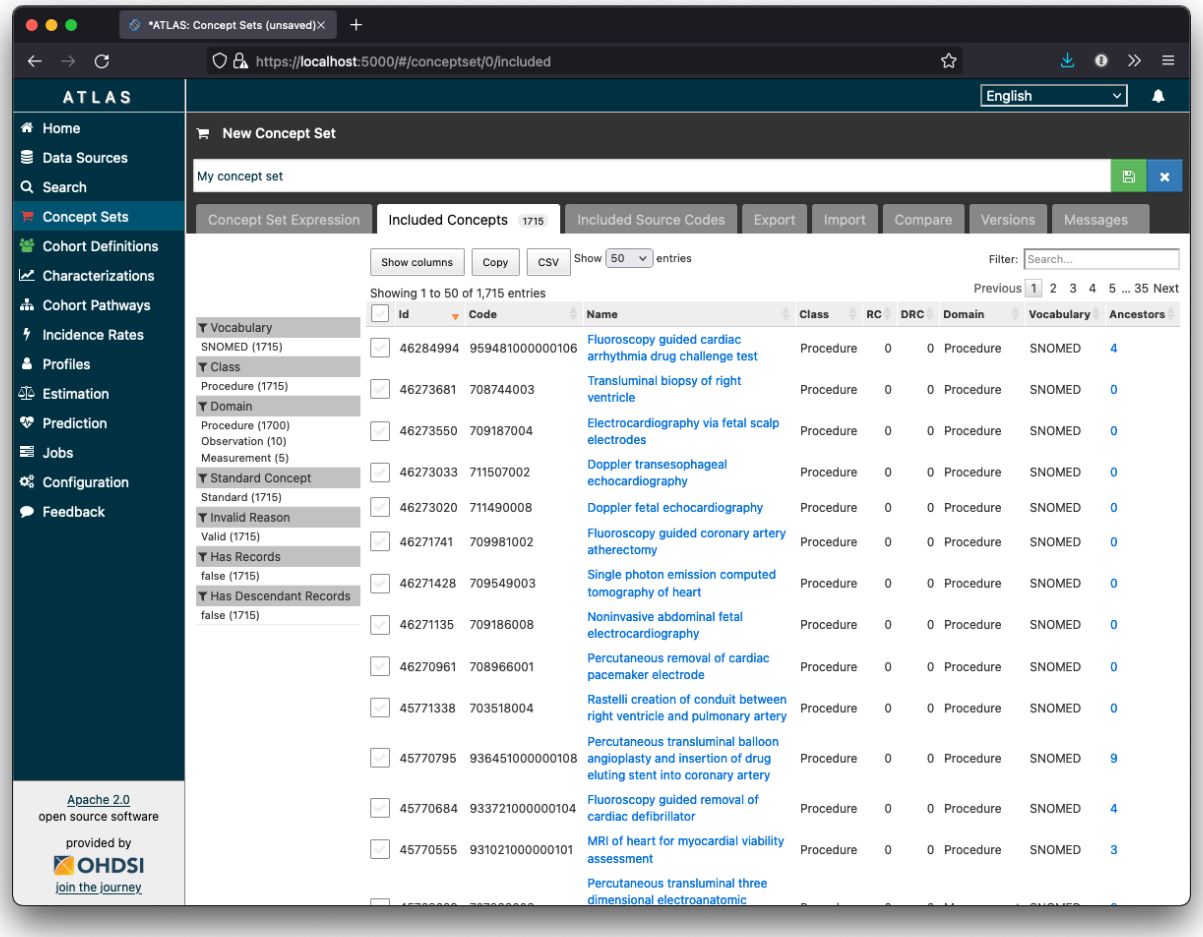


Figure 1 - Authoring a concept set within Atlas

Atlas is tool designed for use by researchers wishing to conduct exploratory data analysis over data structured using the OMOP Common Data Model (CDM). It allows users to create and compare cohorts, and has support for use of terminology within these cohort definitions.

To use terminology within a cohort definition, Atlas provides for the definition of a "concept set". Concepts from OMOP standard vocabularies can be searched for and included in a concept set, and simple compositional rules are available of the form:

- "Descendants" (include all concepts that exist below a code within the hierarchy of its vocabulary);
- "Mapped" (include all concepts that a code maps to within OMOP), and;
- "Exclude" (for inverting the effect of the other rules).

Once authored, concept sets can then be used within the definition of cohorts. For example, a cohort for a study of chronic obstructive pulmonary disease (COPD) might have the following inclusion criteria:

- Aged 40-75 years
- Chronic respiratory symptoms (based upon a concept set)
- High risk factors (based upon a concept set)

Atlas is aimed at providing a single, graphical entry-point into interacting with the OMOP CDM and is heavily used by the OHDSI research community.

1.3.2 Usagi

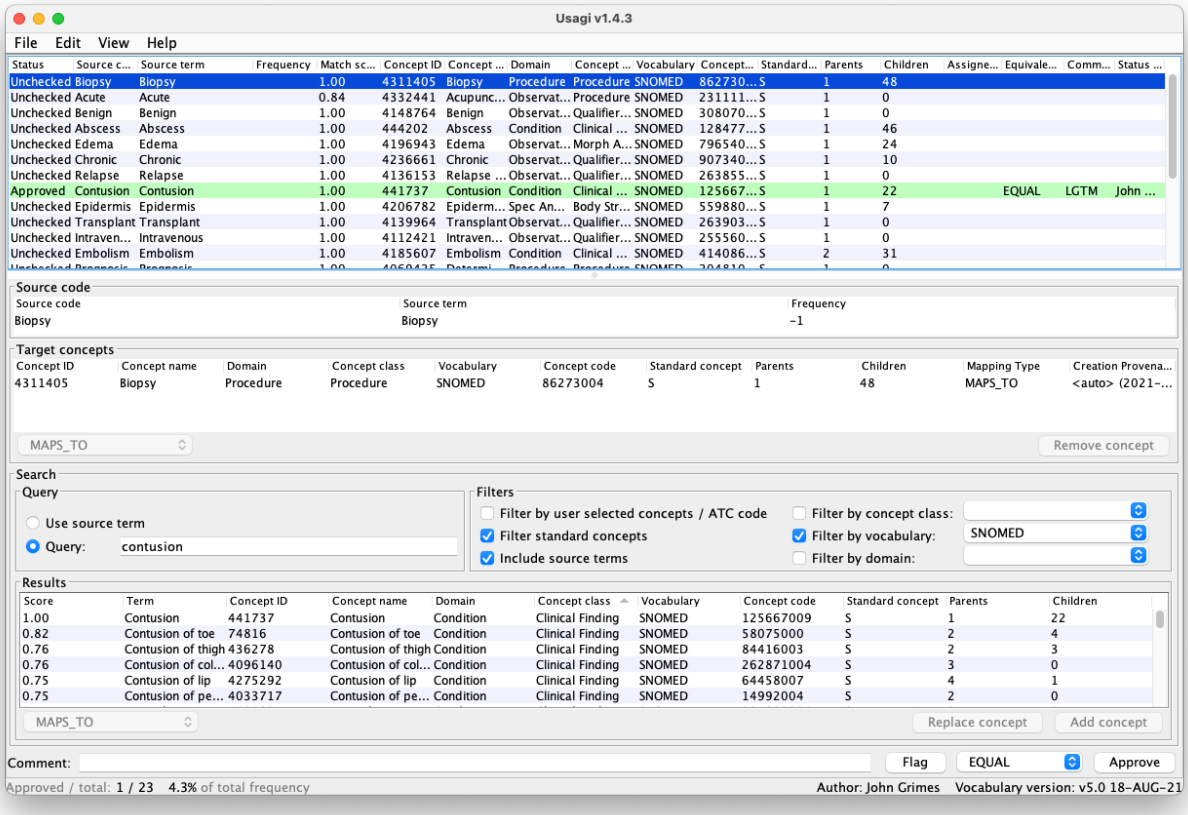


Figure 2 - Mapping source codes to OMOP concepts within Usagi

Most users of the OMOP CDM need to design some sort of extract, transformation and load (ETL) process in order to make source data available within this model. Part of this process is the harmonisation of terminology within the data.

Usagi is a desktop application designed to assist with the mapping of source terminology to concepts within the OMOP standard vocabularies. Source codes can be imported into Usagi, and mappings are automatically identified from within the standard vocabularies based on textual matching. Once the mappings are reviewed, they can be exported within the "source to concept map" format, which is a table within the CDM that allows non-standard source codes to be included within concept sets and cohort definitions.

1.4 Glossary

Quite naturally both OHDSI/OMOP and FHIR use an overlapping set of terms to describe similar but distinct things. This can be the source of much confusion. Here we document some of these terms and attempt to capture the relevant meanings in the two contexts.

The basis for much of the following was drawn from the FHIR R4 specification (https://hl7.org/fhir/R4) and the OMOP Common Data Model documentation (https://ohdsi.github.io/CommonDataModel/vocabulary.html).

Table 1 Glossary of key terms used in the OMOP and FHIR domains

Term	FHIR definition	OHDSI definition
Code	A sequence of characters that identifies a concept from a CodeSystem.	The sequence of characters that identifies a concept in a source code system.
Concept	A representation of an idea.	The OMOP representation of a concept from a code system. Every concept is also given an OMOP identifier and is assigned to a Domain.
Standard Concept	FHIR does not distinguish any specific source of concepts as primary.	A concept from a Standard code system that is considered the preferred representation of the idea. Belongs to a single Domain. Only Standard Concepts can appear in _concept_id fields of CDM tables and the ANCESTORS table.
Concept Set	N/A - see ValueSet	A collection of concepts based on a concept set expression.
Concept Set Expression	N/A - analogous to ValueSet.compose	A set of inclusion and exclusion criteria.
CodeSystem	*The CodeSystem resource is used to declare the existence of and describe a code system or code system supplement and its key properties, and optionally define a part or all*	N/A - see

Term	FHIR definition	OHDSI definition
	of its content. https://hl7.org/fhir/R4/codesystem.html	
ConceptMap	*A statement of relationships from one set of concepts to one or more other concepts - either concepts in code systems, or data element/data element concepts, or classes in class models.* https://www.hl7.org/fhir/R4/conceptmap.html	N/A - see Mapping
ValueSet	*A ValueSet resource instance specifies a set of codes drawn from one or more code systems, intended for use in a particular context. Value sets link between CodeSystem definitions and their use in coded elements.* https://www.hl7.org/fhir/R4/valueset.html	N/A - see Concept Set
Domain	N/A - broadly aligned to the FHIR notion of binding a field/element to a ValueSet.	Defined by the tables and fields in the OMOP CDM that can contain Concepts describing all the various aspects of the healthcare experience of a patient.
Density	N/A	Measure(s) of a concept's prevalence in some data set.
Mapping	See ConceptMap and $translate	Also known as **a map**, is an association between a particular concept in one code system or dataset and code in another, rarely the same, code system that has the same (or similar) meaning. See MapsTo.

Term	FHIR definition	OHDSI definition
		Also, the process of transforming one concept into a Standard one.
Maps To	N/A - FHIR includes a broader range of mapping relationships and separates 'equivalence' from 'target is broader', also supporting 'target is narrower', 'target is disjoint', and other.	The **Maps To** relationship is from a Source Concept and to a Standard Concept. The semantics are full equivalence or an "uphill" mapping (mapping to a more general semantic category).
Maps To Value	N/A - Similar semantics can be represented with the 'product' component of a FHIR ConceptMap	An additional code to be used in the context of maps relating to measurements and observations.

2 Opportunities

This section describes some changes that could be made to the implementation of widely-used OHDSI tools which would improve the ability for OMOP users to take advantage of FHIR terminology services, and FHIR-based terminology platforms provided by QCTS and AHRA.

2.1 Enhancements to Atlas

Atlas is an open-source browser-based web application, with a dependency on another open-source backend component called WebAPI. The following enhancements could be made to Atlas and WebAPI to improve interoperability with FHIR and FHIR terminology services.

2.1.1 Import FHIR ValueSet from FHIR terminology server

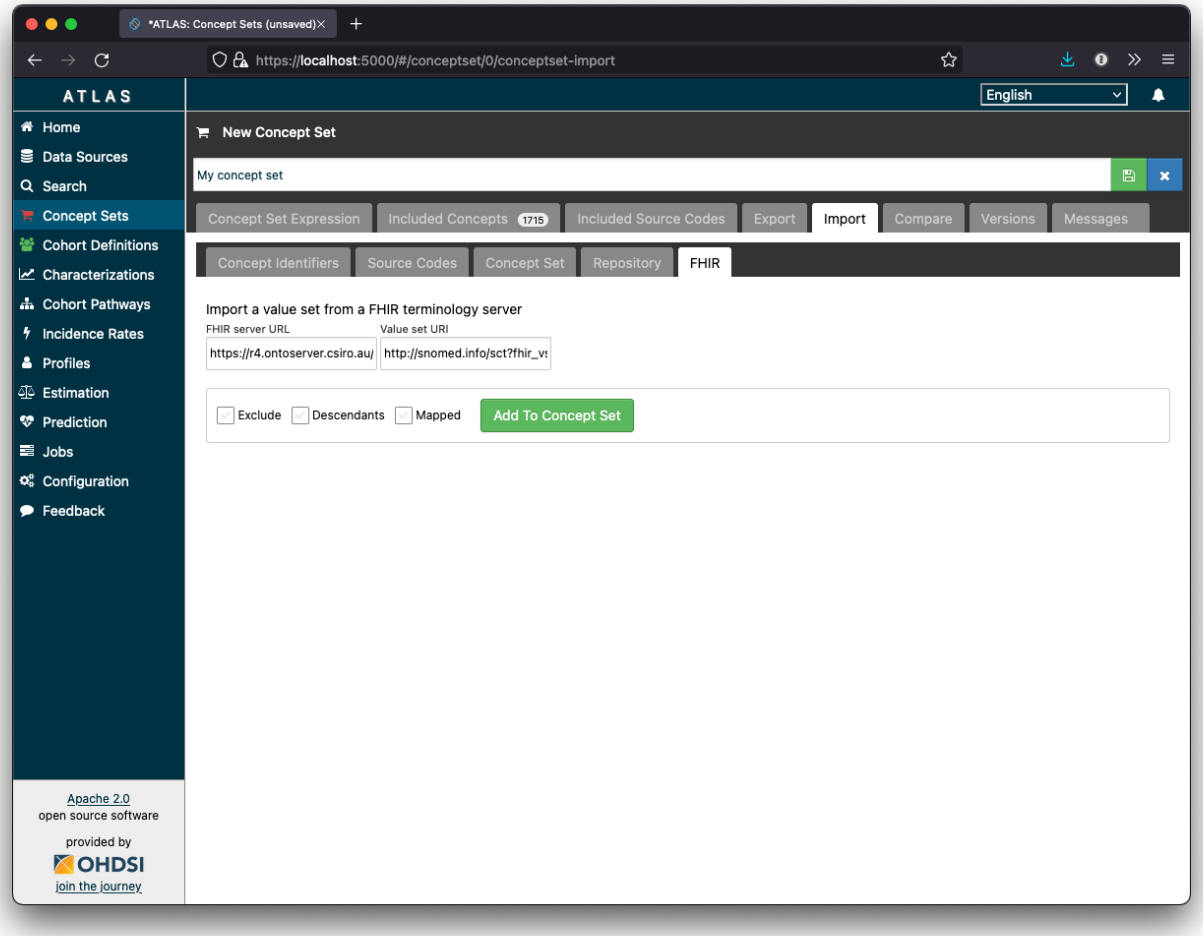


Figure 3 - Importing concepts from a FHIR terminology server in Atlas

FHIR includes some very powerful mechanisms for defining sets of terminology concepts, through the ValueSet resource and its “compose” element. In addition to this, additional expressive power has been defined within FHIR for many well-known code systems, such as SNOMED CT and LOINC.

One opportunity exists to allow users to import existing FHIR ValueSet expansions into Atlas concept sets. This would include:

- Explicitly defined ValueSet definitions published on platforms such as NCTS and QCTS;
- Locally maintained ValueSets, and;
- Implicit ValueSets (such as SNOMED CT reference sets).

An additional tab would be added to the import section of the concept set interface, allowing the user to specify:

5. A FHIR terminology server URL, and;
6. A URI identifying a ValueSet known to the terminology server.

One of the challenges with importing codes from FHIR relates to the identification of code systems, as OMOP uses a set of its own code system identifiers that are currently inconsistent with the identifiers that FHIR uses. Work is currently underway within HL7 to build and publish a table of OMOP Vocabularies and their corresponding FHIR URIs to address this problem. A working document is available here: https://docs.google.com/spreadsheets/d/1ykzbW-RMRJV4doIpZANibu45WveLFoFA/edit#gid=2038668198

Another key challenge is how to handle codes and code systems that a ValueSet includes, that are not yet present in the OMOP vocabulary tables, and thus do not have allocated surrogate keys.

A prototype implementation of FHIR ValueSet importing has been created, and is currently the subject of a pull request[1] with the Atlas GitHub repository.

2.1.2 Generate FHIR ValueSet from concept set

A concept set within Atlas is essentially a set of rules for inclusion and exclusion of concepts. These rules could be expressed using the FHIR ValueSet compose element.

The advantage of this would be that concept sets authored in Atlas could be represented using a format that is suitable for interchange, use with FHIR-enabled tools, and publishing via FHIR-enabled terminology platforms. Furthermore, the FHIR ValueSets would be expressed in terms of each code system's native code rather than non-portable surrogate keys.

2.1.3 Generate Pathling queries from cohort definition

Pathling is an open source FHIR server designed to facilitate analytic query over FHIR datasets. It uses FHIRPath expressions to allow users to describe queries and transformations of FHIR data, and it also supports integration with a FHIR terminology server for integration of advanced terminology operations within queries.

A cohort definition created within Atlas includes inclusion and exclusion criteria that is expressed using elements within the CDM, and also concept sets for leveraging terminology.

[1] https://github.com/OHDSI/Atlas/pull/2610

With the help of an OMOP to FHIR information model mapping, a cohort definition could be translated into a set of Pathling filter expressions that could be used to retrieve matching patients from a FHIR dataset. This could be useful for harmonising heterogeneous data sources, or for replicating and comparing analyses across different data models.

2.1.4 Generate CQL from cohort definition

Clinical Quality Language (CQL) is a standard for shareable definitions of cohorts, measures and indicators, and is data model independent in its design. It could be used to represent a cohort definition created within Atlas, enabling interoperability with tools designed to work with CQL.

Existing work to implement execution of CQL over OMOP could be leveraged to provide an integration point with FHIR terminology services, for complex terminology query that exceeds what is possible using only the vocabulary information within the CDM.

Additionally, an OMOP to FHIR information model mapping could be used to create CQL representations of cohort definitions that are expressed in terms of the FHIR data model, in a similar way to the previous opportunity around Pathling. This would enable cohort definitions authored in Atlas to be executed across FHIR datasets.

2.1.5 Improve concept search

Concept search within Atlas is implemented using OHDSI WebAPI, which executes queries against the vocabulary tables within the CDM database.

This feature could be altered to optionally use a FHIR terminology server to fulfil this capability, leveraging the improved quality of results and higher performance offered by FHIR terminology server implementations such as Ontoserver.

2.2 Enhancements to Usagi

Usagi is an open-source desktop application that has the ability to import source codes and export mappings suitable for use with OMOP CDM and its associated tooling ecosystem. The following changes could be made to Usagi to improve its interoperability with FHIR and FHIR terminology services.

2.2.1 Import FHIR CodeSystem from FHIR terminology server

Usagi could be enhanced to add the ability to import the expansion of a FHIR ValueSet from a FHIR terminology server. Upon receiving a FHIR terminology endpoint and a ValueSet URI from the user, Usagi could invoke the expand operation on the FHIR endpoint and import the resulting codes using the same mechanism that is currently used to import source codes from file.

2.2.2 Import FHIR ConceptMap from FHIR terminology server

Usagi could also be enhanced to enable it to import a FHIR ConceptMap resource as a starting point for a source code to OMOP mapping exercise. Target concepts from the FHIR ConceptMap

would be matched with OMOP standard concepts, and source codes would be imported from the source concepts in the ConceptMap.

FHIR ConceptMap equivalence values would need to be mapped to OMOP relationships. The resulting Usagi map would be left in a “Unchecked” state, facilitating subsequent human review of the result before export and load into the CDM database.

2.3 Enhancements of the Vocabulary Common Data Model

As articulated above, OMOP’s use of non-deterministically assigned surrogate keys is a significant stumbling block for stability across code system versions and interoperability between OMOP tooling deployments. This latter point is somewhat mitigated by the public availability of https://athena.ohdsi.org as a common shared source of the surrogate key assignment.

An alternative approach would be to replace the sequential assignment of surrogate keys by either: a) string-valued key that is the composite of the code system and the code, or b) a unique integer hash of the code system and code (or, in the case of code systems that do not exhibit concept permanence[2], a hash of the code system, code, and version).

While option (a) is attractive because it is transparent and the original information can be recovered from the key, option (b) is particularly attractive because it does not change the data type of the surrogate key, and thus does not change the rest of the Common Data Model schemas; it would appear to be a low-impact and backwards-compatible change.

[2] A code system that never re-uses a code for a different concept is said to exhibit *concept permanence*. SNOMED CT is one such code system. ICD10, on the other hand, does re-use codes and thus does not exhibit concept permanence.

3 Conclusion

A number of opportunities for strategic enhancements to tool implementations have been identified. These show promise for providing significant benefit to OMOP users with a need to integrate with FHIR data and FHIR terminology services. This work would also have wider benefits for improving the alignment of these complementary standards, and improving the quality of terminology within health data analytics more generally.

As part of this work, the team has forged links with key members of both the HL7 FHIR and OHDSI Vocabulary communities. It has also been a particularly well-timed exercise since OHSDI has now split out the OMOP Vocab Working Group as a separate entity to the OMOP Common Data Model Working Group with a specific “FHIR Terminology Alignment” agenda.

It is recommended that further evaluation of these proposals be conducted in parallel to the ongoing work that is going on to harmonise and align OMOP and FHIR at the standards level with strong engagement with this Working Group.

As Australia's national science agency and innovation catalyst, CSIRO is solving the greatest challenges through innovative science and technology.

CSIRO. Unlocking a better future for everyone.

Contact us

1300 363 400
+61 3 9545 2176
csiroenquiries@csiro.au
csiro.au

For further information

Health & Biosecurity
Michael Lawley
+61 7 3253 3609
michael.lawley@csiro.au
aehrc.com