The OMOP Annotator: A Database Agnostic Tool for Reviewing and Augmenting the Patient Record

Amy Yates, MS[1], Erik Benton, BA[1], Izabelle Humes, PT, DPT, MS[1], Matthew Lawhead, BS[1], Heath Harrelson, MS[1], Imogen Bentley, JD[1], Rumel Mahmood, PhD[1], William Hersh, MD[2], Steven Bedrick, PhD[2]

[1]Oregon Clinical and Translational Research Institute, Oregon Health & Science University, Portland, Oregon, United States; [2] Division of Informatics, Clinical Epidemiology and Translational Data Science, Oregon Health & Science University, Portland, Oregon, United States

Background

The Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) has had great success in standardizing healthcare data across institutions. It has a large, supportive community and 974 million unique records as of 2024.[1] However, many research projects need to collect related data that are not part of the patient's record and not appropriate for storage in OMOP. In the field of clinical phenotyping, for example, informaticians may want to test algorithms for identifying a cohort of patients. They need to record details of the cohort and perform manual chart review to rate an algorithm's effectiveness and attach a relevance judgment to the record.[2,3]

Typically chart review is performed using the EHR, but this has several disadvantages. First, the patient record is not static, and the abstractor must pay close attention to judge whether inclusion was appropriate at the time the algorithm was executed.[4] Secondly, the algorithm may not use all parts of the record (e.g., images and documents), so the abstractor can make different judgments based on additional information provided in the EHR. Finally, the abstractor must record their judgment and detail their supporting evidence in a different system.[5]

Using a CDM addresses some of the problems with EHR-based chart review. If a static database is used for both clinical phenotyping and evaluation, the abstractor can view exactly the same information used by the algorithm. However, to streamline the workflow, the abstractor still needs a way to examine each record and attach a judgment. The OMOP Annotator software fills this tooling gap by providing a user-friendly interface for reviewing and searching the patient record and augmenting it with annotations on specific data elements and an overall patient-level categorization.

Methods

The OMOP Annotator is a web application that is used to perform relevance assessment of clinical phenotyping algorithms executed outside the system. After the phenotyper has applied the algorithm and extracted the cohort from their desired dataset, the users would upload pools describing an experimental algorithm, topics defining the phenotype of interest (e.g., males aged 18-50 with an ER Visit for concussion), and cohorts of patient ids (e.g., medical record numbers [MRNs]) that are candidates for a topic match using the algorithm. Abstractors navigate through the cohort to review each patient's record and make a clinical determination of relevance to the phenotype criteria.

The OMOP Annotator is implemented using the Java Spring Boot framework with a combination of Mustache, Vue, and JavaScript for the User Interface (UI). It connects to any OMOP 5.3 database for retrieving patient information, employing a Hibernate Object/Relational Mapping (ORM) framework[6] that makes it agnostic to the underlying Relational Database Management System (RDBMS). It uses a secondary data source for storing user information, configuration, and other augmented data.

Results

The OMOP Annotator allows navigation through pools and topics to the judgment interface where users see a compact view of the patient's record. Data is organized by visits and searchable using Lucene[7] syntax. As the abstractor gathers evidence, they can use pins and notes to mark specific visits that support or contradict inclusion. These annotations provide evidence for the user to make a clinical determination of relevance and assign an overall judgment.

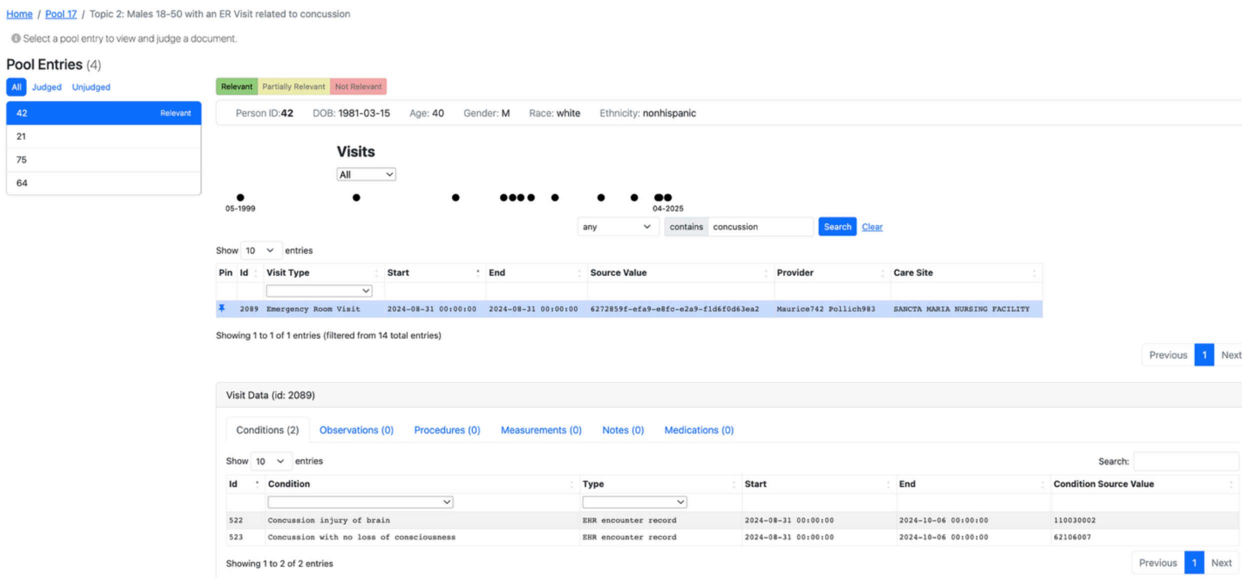
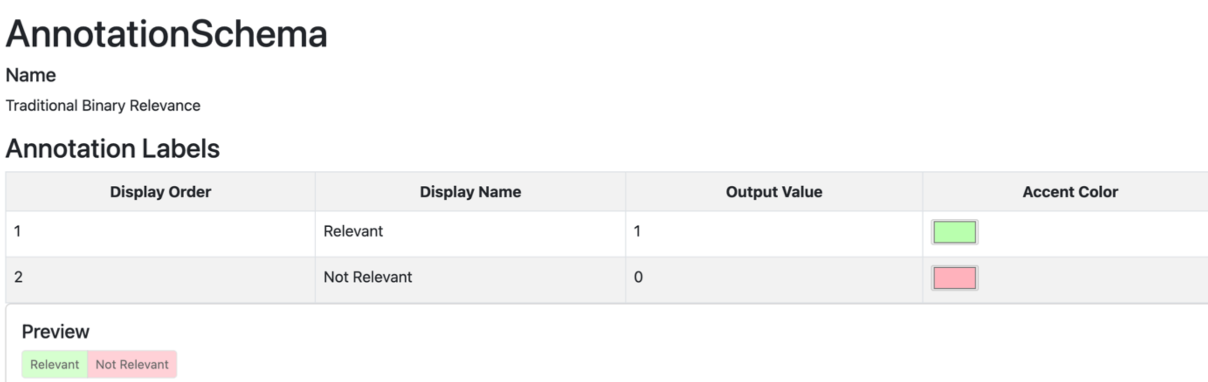


Figure 1. The patient browser filtered by the search term 'concussion'. Data source is the Synthea dataset.[8]

The list of judgment annotations available is configurable for the cohort of patients being processed, allowing for other use cases beyond relevance assessment. Users with administrative privileges can set up different annotation schemas that can be applied to any pool.



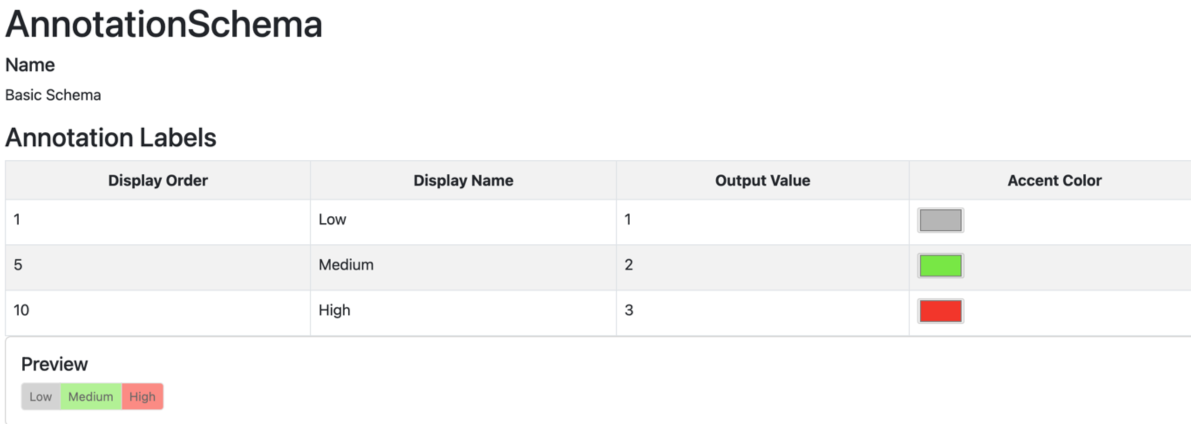


Figure 2. Configurable Annotation Schemas available for a judgment pool.

Initial testing was conducted against an Oracle OMOP instance at Oregon Health & Science University (OHSU) containing over 200K patient records and a Postgres instance of the MIMIC-III clinical database.[9] This testing revealed challenges with supporting multiple database technologies. While our software was designed to be database agnostic, we found differences in the schema definitions of OMOP that were not always compatible with assumptions in the ORM. For example, the field `quantity` in the table DRUG_EXPOSURE is declared as a `float` in Oracle and as a `NUMERIC` in Postgres. These are not compatible types in Java Database Connectivity (JDBC) Mappings, where the ORM expects a float to be declared as a Double in Java and a NUMERIC to be declared as a BigDecimal.[10] We overcame these schema differences by customizing the Hibernate type mappings. This approach should be generally applicable to other software needing this flexibility.

We also discovered institutional differences in how OMOP may be interpreted at each site, with some data being duplicated or empty at one site and not another. We implemented functionality so that administrators can configure which columns appear in the patient browser and whether that data is available for filtering.

OmopDisplayConfiguration List

Show 10 entries   Search: measurement

Entity name	Field name	Column display	Visible	Filter	Actions
MEASUREMENT	id	Id			Edit
MEASUREMENT	measurement	Measurement	✓	✓	(Not Editable)
MEASUREMENT	measurementDatetime	Date/Time	✓		Edit
MEASUREMENT	measurementType	Type	✓		Edit
MEASUREMENT	valueSourceValue	Source Value			Edit
MEASUREMENT	valueAsConcept	Value As Concept			Edit
MEASUREMENT	valueAsNumber	Value As Number	✓		Edit
MEASUREMENT	unit	Unit	✓		Edit
MEASUREMENT	measurementOperator	Operator			Edit
MEASUREMENT	rangeLow	Range Low			Edit

Showing 1 to 10 of 14 entries (filtered from 74 total entries)   Previous 1 2 Next

Figure 3. Configuration to control which columns are visible and filterable.

An unexpected benefit of building this application was uncovering data quality issues within our OMOP instance. For example, we found that a significant portion of visits at OHSU had no relational data. This led us to improve the interface so those visits could be excluded. We also found that many Observations of tobacco use were not related to visits and therefore not visible in the application. We provided feedback to our OMOP implementation team that they can use to make improvements.

The OMOP Annotator has been deployed at OHSU, Mayo Clinic, and the University of Texas Health Science Center at Houston and has been tested with OMOP instances backed by Oracle, Postgres, and SQL Server.

Conclusion

We have developed an RDBMS-agnostic platform for browsing and augmenting OMOP data. While the aim was to support relevance judgment of clinical phenotyping algorithms, we believe the software is applicable to many use cases within the OHDSI community. We have also used it to identify variance across institutions and to uncover potential data quality issues. Other uses might include categorization of clinical trial candidates identified through cohort discovery tools or augmentation of the patient's record with data relevant to a research question. Next steps include open-sourcing the application to GitHub to enable access and encourage adoption at other healthcare sites utilizing the OMOP common data model who may be interested in executing any of these use cases. We also host a public repository that contains a Java ARchive (JAR) file and instructions for deploying the application.[11]

Funding

This work was supported by Grant R01LM011934 from the National Library of Medicine and by the National Institutes of Health by Grant UL1TR002369 through the Clinical and Translational Science Award Program, run by the National Center for Advancing Translational Science.

References

1. Sachson C. Our Journey: Where the OHDSI community has been and where we are going [Internet]. 2024 [2025 June 18]. Available from: https://www.ohdsi.org/wp-content/uploads/2024/10/OurJourney2024.pdf
2. Chamberlin, S.R., Bedrick, S.D., Cohen, A.M., Wang, Y., Wen, A., Liu, S., Liu, H., Hersh, W.R., 2020. Evaluation of patient-level retrieval from electronic health record data for a cohort discovery task. JAMIA Open 3, 395–404. https://doi.org/10.1093/jamiaopen/ooaa026
3. Wu, S., Liu, S., Wang, Y., Timmons, T., Uppili, H., Bedrick, S., Hersh, W., Liu, H., 2017. Intrainstitutional EHR collections for patient-level information retrieval. Journal of the Association for Information Science and Technology 68, 2636–2648. https://doi.org/10.1002/asi.23884
4. Tamine L, Goeuriot L. Semantic information retrieval on medical texts: Research challenges, survey, and open issues. ACM Computing Surveys. 2021;54(7). 1-38. https://doi.org/10.1145/346247
5. Wu S, Timmons T, Yates A, Wang M, Bedrick S, Hersh W, Liu H. On developing resources for patient-level information retrieval [Internet]. 2016 [2025 June 18]. Available from: http://www.lrec-conf.org/proceedings/lrec2016/pdf/919_Paper.pdf
6. Hibernate. Hibernate ORM [Internet]. [cited 2025 June 18]. Available from: https://hibernate.org/orm/
7. Apache. Apache Lucene- Query Parser Syntax [Internet]. 2013 June 21 [2025 June 18]. Available from: https://lucene.apache.org/core/2_9_4/queryparsersyntax.html
8. Synthetic Health. Synthea Patient Generator [Internet]. [cited 2025 June 30]. Available from: https://github.com/synthetichealth/synthea
9. Johnson, A., Pollard, T., & Mark, R. MIMIC-III Clinical Database (version 1.4) [Internet]. PhysioNet. 2016 [2025 June 18]. RRID:SCR_007345. https://doi.org/10.13026/C2XW26
10. Andersen, L. Oracle JDBC 4.3 Specification [Internet]. 2017 July [2025 July 1]. Available from: https://download.oracle.com/otn-pub/jcp/jdbc-4_3-mrel3-eval-spec/jdbc4.3-fr-spec.pdf
11. Oregon Clinical and Translational Research Institute. OMOP Annotator Releases [Internet]. 2022 [cited 2025 July 1]. Available from: https://github.com/OCTRI/omop-annotator-release