

Phenotypes for All: The Phenotype Library Application

Ajit Londhe¹, Carmen Ng¹, Julia Ortmann¹, Andrej Snegiriov², Konstantin Iaroshovets², Lucy McDermott²

¹Boehringer Ingelheim, ²EPAM

Background

The OHDSI Phenotype Library has been an impactful effort to “foster the submission and retrieval of high-quality cohort definitions, cataloging of metadata, attribution and promotion of discovery and reuse in scientific research.”[1] The library itself is available as an R package, but there has not been a graphical user interface to make the content accessible to a larger audience. Similarly, within sites, though Atlas can be utilized to organize cohort definitions and concept sets for RWE-focused analysts, there are often challenges with making those reference standard definitions accessible to clinicians, epidemiologists, and other RWE-focused stakeholders.

The Phenotype Library Application (PLA) is a web-based system designed to govern submission and lifecycle of cohort definitions and concept sets as reference standards. The tool can be leveraged for the OHDSI community and within a site. The PLA enables submitters to upload phenotype definitions and define pathways to analytic implementation (including through OHDSI tools and abstractly, non-OHDSI tools). With this transparency, a larger audience can leverage the reference standard definitions in protocols, statistical analysis plans, and ultimately, as inputs to analytic data pipelines.

Methods

The application architecture and governance model were designed to support standardized, secure, and scalable phenotype submission and curation workflows. Engineering standards centered on OMOP CDM as the canonical data model for cohort definitions and concept sets, while also supporting non-OMOP formats through a data-grid input path.

Authentication was implemented using OpenID/OAuth 2.0 with pluggable identity providers, including organizational single sign-on, public providers such as Google, and local accounts for testing and open-source deployments. Access control combined role-based permissions (Viewer, Contributor, Librarian, Administrator) with attribute-based collaborator lists scoped to individual phenotypes or concept sets.

Workflow and data governance were supported through a defined lifecycle from Draft to Submitted, Approved or Rejected, Published, and Archived, with transitions governed by role and collaborator status. Submission required both mandatory and optional metadata, including therapeutic area or domain, tags, and descriptions, using controlled vocabularies where appropriate and free-text tagging for flexibility.

Attachment management for PDF and Word files was handled within the cohort and concept-set store rather than through ad-hoc file sharing. Validation and quality checks included automated OMOP JSON validation for structural integrity, Circe-be model conformance, required attributes, data types, and domain-specific rules; similarity detection across names, descriptions, and definitions to identify duplicate or near-duplicate submissions; and a Check Result Preview that consolidated warnings and errors into a single review pane for resolution prior to librarian review.

Results

The initial application has a working authenticated platform, end-to-end cohort workflow, and a Concept Sets foundation; Validation Feedback, and ATLAS Integration. Additional features will be implemented as part of Phase 2 scope. Deployment for the OHDSI community is planned for Fall 2026.

Sophisticated Transition Flow and Role-Based Restrictions

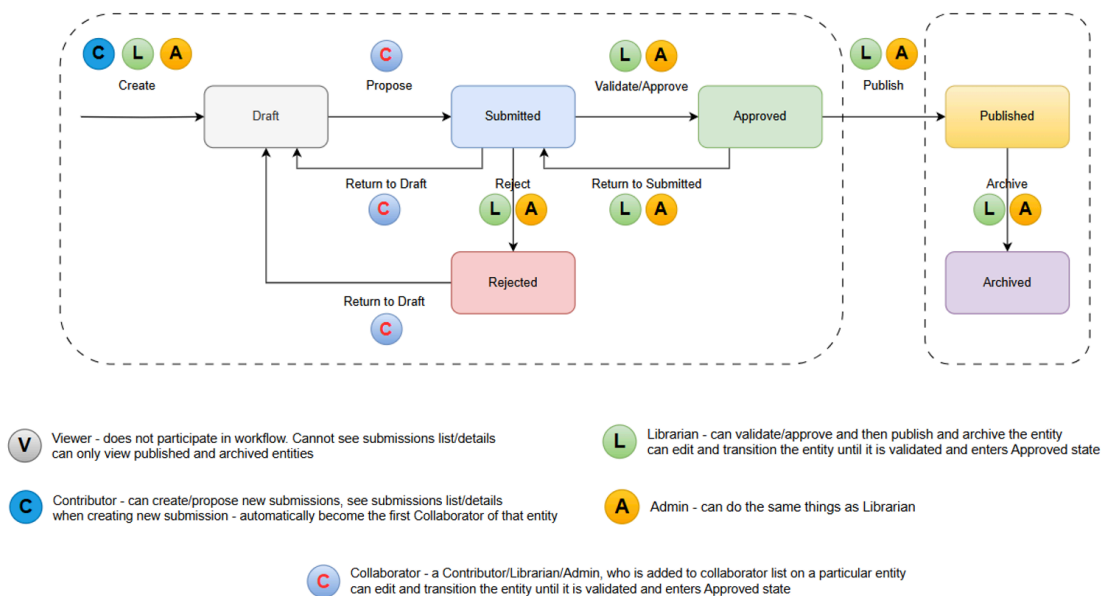


Figure 1. Transition Flow and Role-Base Restrictions

Heart disease

Heart disease - all events per person.
era collapse gap size: 1 day

PUBLISHED

Attributes

Therapeutic area: Cardiology

Cohort Definition (OMOP format)

Cohort Definition (OMOP Format) Viewer

Cohort Entry Event

Initial Event

Events having any of the following criteria:



a condition occurrence of Heart disease

with continuous observation of at least 0 days before and 0 days after event index date
Limit initial events to: all events per person.

Details

Created: Feb 10, 2026, 1:38 PM
Updated: Feb 10, 2026, 1:39 PM
Submitter: ms Viconia DeVir

Workflow History

- Published**
by Irenicus Archer
Feb 10, 2026, 1:39 PM
- Curated & Approved**
by Irenicus Archer
Feb 10, 2026, 1:39 PM
- Submitted for Review**

Figure 2. Cohort definition entry details

Cohort Definition Viewer

Initial Event
Events having any of the following criteria:
• a drug exposure of **Plaglitazone**
with continuous observation of at least 9 days before and 9 days after event index date
Limit initial events to: **earliest event** per person

Restrict Initial Event

Inclusion criteria

- Age between 21 and 70 years
- NASH confirmed by liver biopsy - Condition
having all of the following criteria:
• with at least 1 using all occurrences of:
a condition occurrence of **NASH - Condition**
where event starts between **all days before** and **6 days after** index start date
The index date refers to the event from the Cohort Entry criteria
- NASH confirmed by liver biopsy - Observation
having all of the following criteria:
• with at least 1 using all occurrences of:
an observation of **NASH - Observation**
where event starts between **all days before** and **6 days after** index start date
The index date refers to the event from the Cohort Entry criteria
- Impaired Glucose Tolerance or Type 2 Diabetes
- Hemoglobin ≥ 13 g/dL for males or ≥ 12 g/dL for females

Figure 3. Cohort Definition Viewer

Phenotypes
Browse published phenotypes and manage submissions

Library Submissions

Search by name, description, or domain...

Type: [] Domain: [] [E] Publication Date: [] [] Show Archived Phenotypes [X] Clear filters

NAME & DESCRIPTION	TYPE	DOMAIN	PUBLISHED
Lumbar Radiculopathy - multi-groups Condition: Circumferential, Lumbar Radiculopathy Drug Exposure: Plavix, Tylenol, Tylenol Control Spectrum Age: 18-65 years	GROUP	Neurology	by Mr. Anding Deng/Nov
T2DM - Non-insulin treated diagnosis (OMOP) enrollment inclusion criteria: Non-insulin treated T2D diagnosis + Age between 40 and 70 + HbA1c between 6.0% exclusion criteria: Diabetes diagnosis less than 3 years before the study initiation...	GROUP	*	by Mr. Anding Deng/Nov
Joint stiffness joint stiffness = condition occurrence Event will persist until fixed duration relative to initial event Event date to other from end date = 0 days	GROUP	Neurology	by John Norrish

Figure 4. Library View

Conclusion

The Phenotype Library Application will be an important conduit for the OHDSI community to share the rigorous phenotype designs with a larger audience. It will also support sites with fostering the FAIR principles (Findability, Accessibility, Reproducibility, and Interoperability) for reference cohort and concept set definitions.

References

- <https://ohdsi.github.io/PhenotypeLibrary/>