Athena-Client: A Community Python SDK for Programmatic Access to the OHDSI Athena Vocabulary Repository

Alvaro A. Alvarez[1,2], Farnoosh Sheikhi[1,2], Priya Desai[1,2]
[1] Stanford Health Care , [2]Stanford School of Medicine

Background

The Observational Health Data Sciences and Informatics (OHDSI) initiative distributes standardized clinical vocabularies through its **Athena** website. Although the site provides an interactive search interface, **no official public API exists**. Analysts who need scripted concept discovery therefore rely on manual downloads or ad-hoc scraping approaches that slow down iteration and jeopardise reproducibility. athena-client closes this gap by wrapping the JSON web-service endpoints consumed by the Athena UI in a resilient, developer-friendly SDK. The library transparently handles authentication, pagination, rate-limit back-offs, and deeply nested JSON, enabling rapid, automated vocabulary exploration and validation inside analytic pipelines [1].

Building on feedback from our earlier research-software demo, *Bridging the Language Gap: Generative Models for Efficient Medical Concept Discovery*—an OHDSI 2024 Collaborator-Showcase honouree (demo video [3]; source code [4]) we now extend our contribution to support more complex cohort-building workflows.

Methods

The athena-client library is built with a layered architecture to support robust and flexible interactions with OHDSI's Athena vocabulary service. Its core layers include secure transport with retry logic, a client interface for searching and retrieving concept details, and Pydantic-based data models for easy export to common formats. Higher-level modules like the CLI, *ConceptExplorer*, and *ConceptSetGenerator* enable rapid concept exploration, scripting, and validation against local OMOP databases.

athena-client layered architecture is listed below:

Layers	Functionality
Transport	Bearer-token + HMAC auth, exponential-back-off retries, and polite throttling
Core Client	search, details, relationships, graph, and summary
Data Models	Pydantic 2 objects such as export to dict, JSON, or pandas.DataFrame

Layers	Functionality
CLI	Mirrors core API for shell scripting
Exploration Modules	ConceptExplorer (breadth-first traversal) and ConceptSetGenerator (database-backed validation)

This library also adds endpoint discovery, schema validation, and adaptive retries since the underlying endpoints are not formally documented.

At startup, the client automatically grabs Athena's current API URL, and every JSON response is checked against a model; if the structure changes, you receive an immediate, readable error. Lastly, adaptive retries back off on HTTP 429 and re-authenticates when session tokens expire (Figure1).
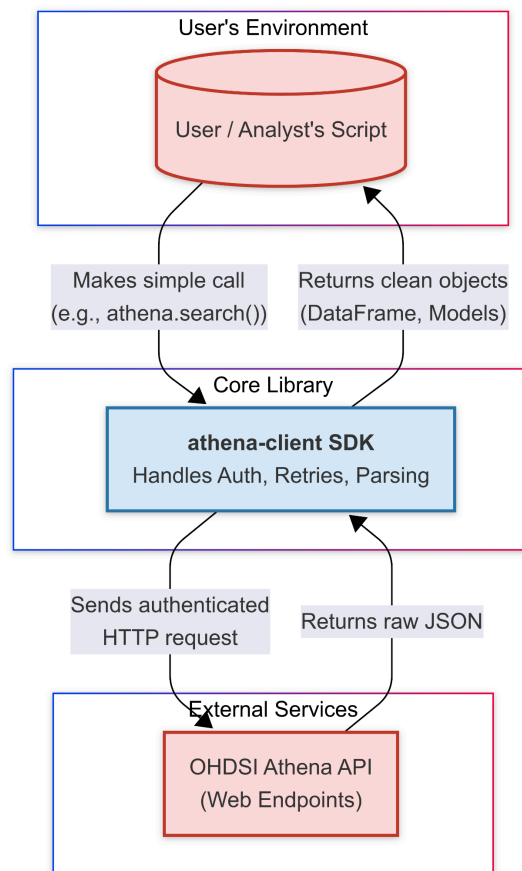

Figure 1: Depicts how the SDK mediates between the user and Athena endpoints

At Stanford, [athena-client](https://github.com/aandresalvarez/athena_client) has been successfully implemented and it is available to install (https://github.com/aandresalvarez/athena_client) in a Python framework (*pip install athena-client*). Additionally, this package has been properly tested and the usage examples are listed here.

Results

The [athena-client](#) library successfully streamlines access to OHDSI's standardized vocabularies by providing a fully programmatic interface to the Athena website. This package improves the scripted workflow that previously required manual navigation through the Athena UI by returning the results in both structured Python objects and tabular formats (e.g., pandas DataFrames). The *ConceptExplorer* and *ConceptSetGenerator* modules support complex concept set construction, validation, and hierarchical relationships, reducing the cognitive burden of concept discovery.

Python API Illustration

```Python
from athena_client import Athena
athena = Athena()

# Find aspirin (limit 3 hits)
hits = athena.search("aspirin", size=3)
for c in hits.all():
    print(f"{c.id=} {c.name=} {c.vocabulary=}")

# Inspect the top concept in depth
d = athena.details(1112807)      # 1112807 = Aspirin (RxNorm)
print(d.conceptClassId, d.standardConcept)
```

Console Output

```None
ID: 36835757, Name: 'ASPIRIN', Vocabulary: 'OMOP Invest Drug'
ID: 1112807,  Name: 'aspirin', Vocabulary: 'RxNorm'
ID: 46286425, Name: 'Aspirin declined', Vocabulary: 'SNOMED'
Class: Ingredient, Standard: Standard
```

CLI Equivalent

```
(.venv) → athena_client git:(main) x athena search "aspirin" --limit 3
[Responsible Usage] Please avoid excessive or automated requests. Abuse may result in your IP being blocked by the Athena API provider. Use filters and limits
where possible.
i Note: Query 'aspirin' matches many concepts (estimated 1,000+), but only 20 will be downloaded.
                                                Athena Results
 id        name              domain       vocabulary        className         standardConcept         code               invalidReason   score
 36835757  ASPIRIN           Drug         OMOP Invest Drug  Ingredient        ConceptType.NON_STANDARD R16C05Y76E         Valid           None
 1112807   aspirin           Drug         RxNorm            Ingredient        ConceptType.STANDARD     1191               Valid           None
 46286425  Aspirin declined  Observation  SNOMED            Context-dependent ConceptType.STANDARD     985721000000109    Valid           None
(.venv) → athena_client git:(main) x ▮
```

Conclusion

athena-client Python SDK fills a critical infrastructure gap in the OHDSI ecosystem by enabling fast, reproducible, and programmatic access to standardized vocabularies.
It streamlines multi-step, error-prone HTTP workflows into concise, readable Python or CLI commands, lowering the barrier for researchers to adopt OHDSI vocabularies programmatically. By offering a clean and reliable interface to Athena's backend services, the athena-client library simplifies key tasks like concept lookup, relationship navigation, and validation against local databases. These features support faster and more consistent cohort definition workflows and reduce the manual overhead typically involved in working with OMOP vocabularies.

Disclaimer. Because athena-client depends on endpoints that are not formally documented, maintainers monitor upstream changes and publish patches promptly. Users should pin versions and subscribe to repository notifications for uninterrupted service.

Key Takeaway

athena-client transforms Athena's web-only search capability into a scriptable, production-ready API, empowering the OHDSI community to embed vocabulary best practices directly into automated research pipelines.

References

1. Hripcsak G, Duke JD, Shah NH, *et al.* Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. *Stud Health Technol Inform.* 2015;216:574-578.
2. Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. *J Am Med Inform Assoc.* 2012;19(1):54-60.

3. Alvarez AA. *Bridging the language gap: Generative models for efficient medical concept discovery* [video]. YouTube; 2024. Available from: https://www.youtube.com/watch?v=vv2y48bLOJo
4. Alvarez AA. *concept-discovery* [software]. GitHub; 2024. Available from: https://github.com/aandresalvarez/concept-discovery