

Usagi-on-the-Web: Collaborative OMOP Concept Mapping with AI-quality Ranking but No LLM API Bills

Natthawut Adulyanukosol

Siriraj Informatics and Data Innovation Center (SiData+),
Faculty of Medicine Siriraj Hospital, Mahidol University, Thailand

Background

Standardized vocabularies are essential for harmonizing heterogeneous health data in the OMOP Common Data Model and enabling OHDSI network research across global sites [1]. In practice, mapping local source terms to standard OMOP concepts remains a high-friction task, particularly for institutions working with local languages, local drug catalogs, legacy coding systems, or limited informatics support. The original Usagi application has been widely used in the OHDSI community to support source-to-standard concept mapping through ranked candidate suggestions and manual review [2]. However, many mapping projects still require substantial human effort when fuzzy matching and lexical similarity are insufficient to distinguish clinically similar concepts.

Recent studies have explored large language models (LLMs), retrieval-augmented generation, and sentence-transformer approaches for OMOP concept mapping [3-6]. These methods show promise, but cloud LLM costs, privacy concerns, latency, and infrastructure requirements may limit adoption, especially for institutions without routine GPU access or permission to send local terminology data to external APIs. We previously developed THIRAWAT (**T**erminology **H**armonization using **L**ate-**I**nteraction **R**eranker **W**ith **A**lignment-tuned **T**ransformers), a ColBERT-style late-interaction drug mapping approach that combines SapBERT with novel BiMaxSim scoring and deterministic drug-aware reordering [7-9]. THIRAWAT improves candidate ranking to roughly 90% top-1 accuracy in drug concept mapping without requiring LLM calls. This reduces the human review effort to the remaining 10%.

In this software demonstration, we present Usagi-on-the-Web, a browser-based collaborative concept mapping workbench that integrates THIRAWAT ranking into a human-friendly review workflow for concept mapping projects.

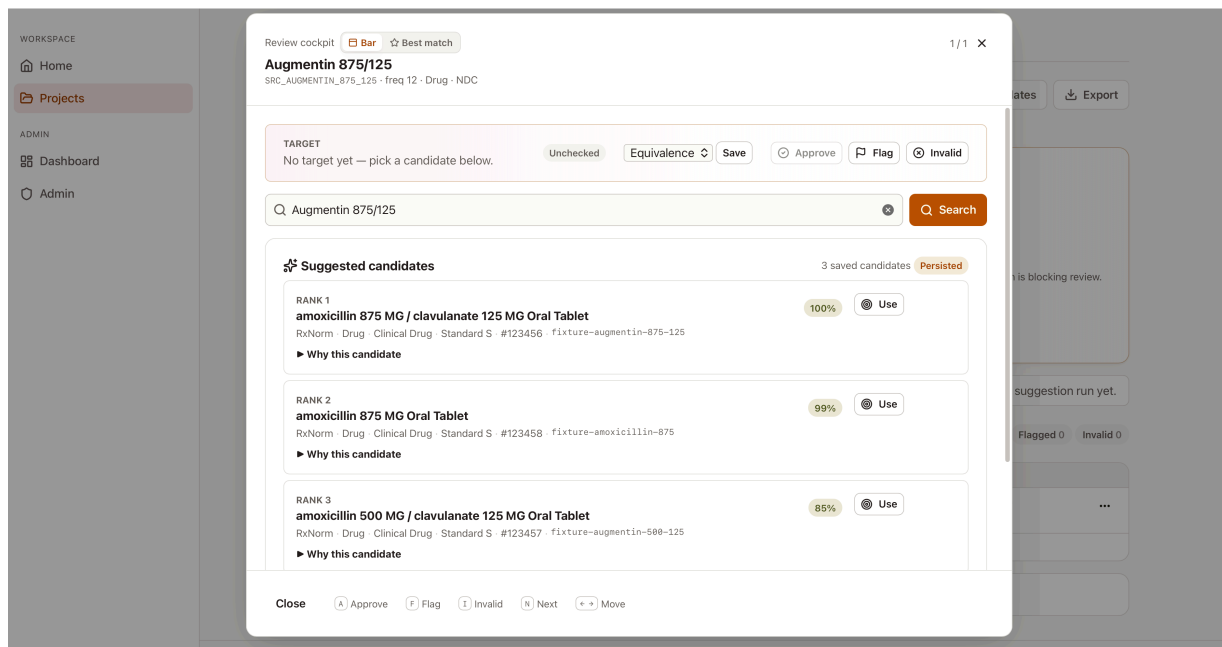


Figure 1. Review cockpit on Usagi-on-the-Web.

The review cockpit presents each source term in a focused modal interface that combines the current mapping decision, ranked candidate standard concepts, candidate scores, reviewer actions, history, and comments. Reviewers can search for alternative concepts, inspect suggested candidates, select a target concept, assign equivalence, approve or flag the mapping, and move through the review queue. This interface illustrates the system’s human-in-the-loop design: AI-assisted ranking accelerates candidate discovery, while mapping decisions, comments, and audit history remain explicitly reviewer-controlled.

Methods

Usagi-on-the-Web separates the collaborative review workflow from the ranking engine. The web application is implemented in Rails 8 and manages users, projects, memberships, imports, source terms, mapping decisions, candidate persistence, comments, audit events, engine job mirrors, and exports. Reviewers work entirely through the browser to create projects, import source terms from CSV or spreadsheet files, request candidate suggestions, inspect ranked candidates, select target

concepts, assign mapping equivalence, approve or flag mappings, discuss uncertain rows, and export reviewed results, see **Figures 1 and 2**.

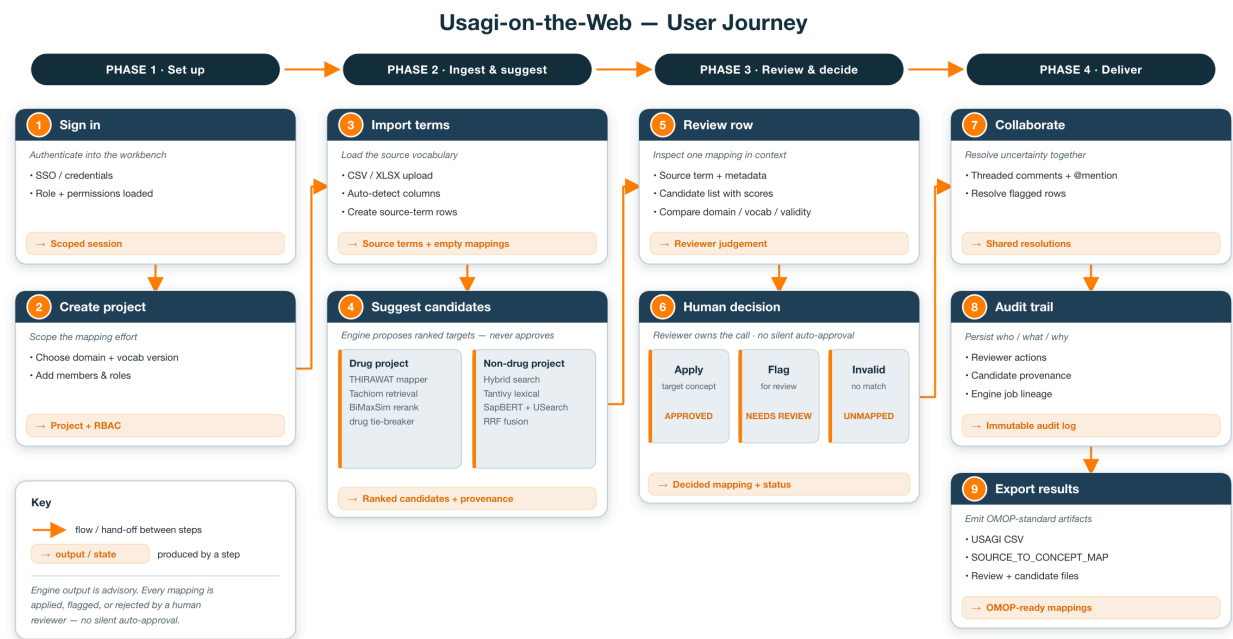


Figure 2. Human-centered mapping review journey.

The reviewer workflow begins with project creation and source term import, followed by AI-assisted candidate generation through either THIRAWAT-assisted Drug mapping or hybrid lexical and semantic search for other domains. Reviewers inspect ranked candidates, compare concept metadata and scores, select target standard concepts, assign equivalence, approve or flag mappings, add comments, and export reviewed results. The engine proposes candidates, but mapping authority remains with human reviewers; candidate provenance, job status, comments, and audit events are retained to support transparent and collaborative OMOP concept mapping.

Ranking and vocabulary functionality are exposed through Rust-based API services rather than browser-side file access. The catalog service stores valid standard OMOP concepts from Athena as the runtime vocabulary source of truth. The search service supports manual concept lookup using Tantivy lexical retrieval and SapBERT dense

retrieval stored in USearch approximate-nearest-neighbor indexes, combined through hybrid reciprocal rank fusion [8,10,11]. For Drug-domain source terms, the mapper service uses the THIRAWAT-SapBERT model to encode query tokens, retrieves candidates using Tachiom MaxSim-style multivector retrieval, applies exact BiMaxSim reranking, and then performs deterministic drug-specific tie-breaking based on clinically relevant attributes such as strength, ingredient combinations, dose form, route, release characteristics, and brand information [7,9,12]. The entire architecture is summarized in **Figure 3**.

The platform supports multiple deployment models. For local evaluation and demonstrations, all components can be launched using Docker Compose with prebuilt or fixture artifacts. For institutional deployments, the Rails application and Rust services can be deployed as containerized services within on-premises environments or private cloud infrastructure. Vocabulary indexes and embedding artifacts can be generated separately and reused across deployments, reducing operational overhead during routine mapping activities.

Human review remains explicit throughout the workflow. The engine proposes candidates, but it does not automatically approve mappings. Candidate scores, ranking methods, artifact identifiers, and model provenance are persisted so reviewers can understand why a suggestion was presented. Routine review and query-time mapping workflows operate without LLM API calls, while computationally intensive artifact generation can be performed independently when needed.

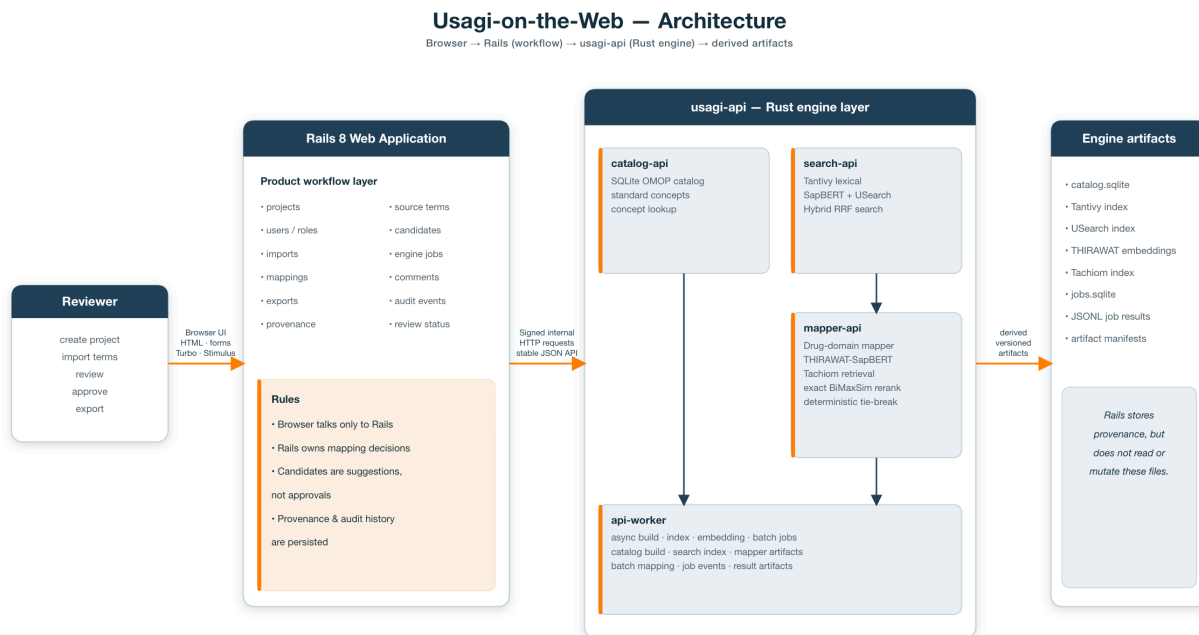


Figure 3. Usagi-on-the-Web system architecture.

Usagi-on-the-Web separates the collaborative review application from the ranking and vocabulary engine. The Rails 8 web application manages users, projects, imports, source terms, mapping decisions, candidate persistence, comments, audit events, job mirrors, and exports. Internal Rust API services provide the engine layer, including the OMOP standard concept catalog, Tantivy lexical search, SapBERT and USearch dense retrieval, hybrid reciprocal rank fusion, and the THIRAWAT Drug-domain mapper using Tachiom retrieval, exact BiMaxSim reranking, and deterministic drug-specific tie-breaking. The browser interacts only with Rails, while engine artifacts and indexes remain internal, preserving a clear boundary between human review workflow and reusable ranking infrastructure.

Results

In previously published drug mapping experiments, THIRAWAT achieved top-ranked candidate accuracy of 94% in the Branded Drug setting, 86% in the Clinical Drug setting, and 87% in the Thai Medicines Terminology setting [7]. Correct standard drug candidates appeared within the top five results for 94% of evaluated mappings, representing a substantial improvement over fuzzy search with TF-IDF reranking baselines commonly used for candidate generation. The same study found that LLM-based post-ranking provided only marginal additional gains while increasing computational cost and operational complexity [7].

The Usagi-on-the-Web demonstration will show this ranking approach embedded within a complete mapping workflow. Audience will see project creation, source term import, manual hybrid search, THIRAWAT-assisted Drug candidate generation, candidate inspection, reviewer selection, approval and flagging, comments, audit history, job status tracking, and export. The demonstration emphasizes review efficiency, transparency, and deployability.

Conclusions

Usagi-on-the-Web modernizes the Usagi-style OMOP mapping workflow for collaborative browser-based review while integrating late-interaction AI ranking for high-quality concept suggestions. By separating the review application from the ranking engine, preserving provenance, and avoiding routine dependence on cloud LLM APIs, the system aims to make accurate concept mapping more accessible to OHDSI sites with diverse infrastructure capabilities. Future work will focus on extending specialized ranking beyond drug concepts, improving the reviewer experience, and evaluating collaborative mapping quality across multilingual and multi-site projects.

References

- [1] Reich C, Ostropolets A, Ryan P, Rijnbeek P, Schuemie M, Davydov A, et al. OHDSI Standardized Vocabularies: a large-scale centralized reference ontology for international data harmonization. *J Am Med Inform Assoc.* 2024;31(3):583-590. doi:10.1093/jamia/ocad247.
- [2] OHDSI. Usagi [Internet]. GitHub; [cited 2026 Jun 5]. Available from: <https://github.com/OHDSI/Usagi>
- [3] Kimura E, Kawakami Y, Inoue S, Okajima A. Mapping drug terms via integration of a retrieval-augmented generation algorithm with a large language model. *Healthc Inform Res.* 2024;30(4):355-363. doi:10.4258/hir.2024.30.4.355.
- [4] Adams MC, Perkins ML, Hudson C, Madhira V, Akbilgic O, Ma D, et al. Breaking digital health barriers through a large language model-based tool for automated Observational Medical Outcomes Partnership mapping: development and validation study. *J Med Internet Res.* 2025;27. doi:10.2196/69004.
- [5] Zhou X, Dhingra LS, Aminorroaya A, Adejumo P, Khera R. A novel sentence transformer-based natural language processing approach for schema mapping of electronic health records to the OMOP common data model. *AMIA Annu Symp Proc.* 2025;2024:1332-1339.
- [6] Ostropolets A, Schuemie M. Ariadne: a Python toolkit for mapping source terminologies to standard concepts in the OHDSI Vocabulary [Internet]. GitHub; [cited 2026 Jun 5]. Available from: <https://github.com/OHDSI/Ariadne>
- [7] Adulyanukosol N, Chaisutyakorn K, Sombutjaroon S, Kanjanapong S, Suriyaphol P. Efficient drug terminology mapping with bidirectional late-interaction reranking and deterministic reordering. *Healthc Inform Res.* 2026;32(2):156-165. doi:10.4258/hir.2026.32.2.156.
- [8] Liu F, Shareghi E, Meng Z, Basaldella M, Collier N. Self-alignment pretraining for biomedical entity representations. In: *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*; 2021 Jun 6-11; Virtual Event. p. 4228-4238. doi:10.18653/v1/2021.naacl-main.334.
- [9] Khattab O, Zaharia M. ColBERT: efficient and effective passage search via contextualized late interaction over BERT. In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*; 2020 Jul 25-30; Virtual Event, China. p. 39-48. doi:10.1145/3397271.3401075.

[10] Quickwit Inc. tantivy [Internet]. Version 0.22.0. GitHub; 2024 Oct 17 [cited 2026 Jun 5]. Available from: <https://github.com/quickwit-oss/tantivy>. doi:10.5281/zenodo.13942948.

[11] Vardanian A. USearch by Unum Cloud [Internet]. Version 2.25.3. GitHub; 2026 [cited 2026 Jun 5]. Available from: <https://github.com/unum-cloud/USearch>. doi:10.5281/zenodo.7949416.

[12] Martinico S, Nardini FM, Rulli C, Venturini R. Efficient multivector retrieval with token-aware clustering and hierarchical indexing [Internet]. arXiv; 2026 [cited 2026 Jun 5]. Available from: <https://arxiv.org/abs/2604.28142>.