

***MindMeetsMachines Vocabulary***  
***Edition at OHDSI EU Symposium 2026***  
**参戦レポート**

- OHDSIのGenAIマッピング最前線 -  
2026/06/29

株式会社Yuimedi 井上 真吾

# 自己紹介

- 井上 真吾 / Shingo Inoue
- 株式会社Yuimedi 取締役 CTO
  - ex- CureApp
- CS - コンピューターサイエンス
  - JavaScript、データ処理、最近はLLMを用いたシステム開発など
- 医療情報の**技術者**としてOMOPに取り組んでいる
  - 医学分野は素人です



 Leko

## COI: 株式会社 Yuimedi の役員

ただし今日もYuimediの話ではなくOMOPそのものの話がメインです

# OHDSI Japan四度目の登壇です

## MEDIS標準病名マスターにおける ICD-10経由のボキャブラリマッピング 手法と実装

構造的課題の分析と実務的対応

株式会社Yuimedi 井上真吾、矢澤まり

2026/04/21 - OHDSI Japan イブニングカンファレンス2026年4月

## 公募ワークショップ9 連 合分析/連合学習の推進と 課題 振り返り

2024/11/26 株式会社Yuimedi 井上 真吾  
OHDSI Japan イブニングカンファレンス 2024/11



## コンセプトの海の泳ぎ方

株式会社Yuimedi 井上 真吾  
OHDSI Japan 2025年11月公開版

## 2026 EU Symposium

- 2026/04/20開催 @ Rotterdam (オランダ)
- [公式サイト](#) にポスターや現地の写真など色々あがってます
- あるワークショップにリモート参加してきました



EUROPEAN OHDSI  
SYMPOSIUM  
Rotterdam

# OHDSI EUROPE SYMPOSIUM 2026

*Continuous Collaboration for Living  
Evidence Generation*

Monday April 20<sup>th</sup>, 2026  
SS Rotterdam (NETHERLANDS)

Organized by:



# MindMeetsMachines Vocabulary Edition

- Anna Ostropolets氏がリードでForumで呼びかけ
  - <https://forums.ohdsi.org/t/mindmeetsmachines-vocabulary-edition-ohdsi-eu-symposium/25056>
- Human-arm vs Machine-armの精度勝負
  - “勝負”と言っているが、互いの強み・弱みを学び良いところ取りしましょうという趣旨
  - Human-arm: その場でチームを組み 1hかけてマッピング
  - Machine-arm: 事前に推論パイプラインを構築し WS当日に結果を提出
- ヨーロッパ圏の多言語で構成された診療行為コードがソース
  - マッピング先はSNOMED, LOINC, HCPCSなど標準コンセプト
  - 評価基準が設けられており、それに従って作成されたデータセットを元に成績を評価
  - 推論結果にはExact Match(1:1対応)、Broad Match(粒度を粗くしてマッチさせた)を付与する
- 木村先生からお誘いいただき参戦
  - アルゴリズム原案: 木村先生
  - アルゴリズム実装&精度改善: 井上

# Agenda

- 2026 EU Symposiumにて開催されたMindMeetsMachines Vocabulary Editionワークショップの参戦レポート
- ボキャブラリマッピングの半自動化の概論
  - UsagiやAthena(NLP)でのアプローチ
  - LLMでのアプローチ
- OHDSI GenAI WGの活動について
  - Hecate、Ariadneについて
- まとめ

# 当日の様子



- 20:00-22:00 JST 😊
- およそ100人がワークショップに参加
- Machine-arm: よくあるモデルコンペのようなスタイル
  - gold standard(正解付きデータ)が事前配布、開発と精度検証は各自が実施
  - 3人の専門家が独立にマッピングし合意形成、意見が割れたものは議論の結果を反映
  - WS前日にスコア評価用データが配布され、開発したモデルでの実行結果を提出
  - 実装したアプローチについて1-2スライドで簡単に発表
- Human-arm: 人間チームによるボキャブラリマッピング(専門性問わず)
  - 5チームで分担しMachine-armと同データを議論しながら1hマッピングし提出
  - Athena / Usagi は利用可。GenAIの使用は禁止
- 運営側が両arm/全チームの提出物を取りまとめ

# 振り返り: Jun 9のCommunity Call

### MindMeetsMachine Vocabulary edition set up

- Gold standard included codes collected from Norway, Netherlands, Austria, Greece (partially mapped)
- Mapping/mapping assessment was done by 2 independent vocabulary experts, disagreement was resolved by a 3<sup>rd</sup> expert
- Split into train set (n = 61) and test set (n = 291)
  - Train set sent to participants in advance, test set distributed before the meeting
  - Exact match and broad match (uphill)

### MindMeetsMachine Vocabulary edition set up

Human arm mapping conducted at the OHDSI EU Symposium

50+ people split into 5 teams with at least one vocabulary expert per group

232/291 codes mapped during 1-h session

### MindMeetsMachine Vocabulary Results

Reference standard was adjusted based on blinded review of combined results from the teams and human arm

Example:  
CT wrist-hand right originally mapped to “CT of hand and wrist”  
Added: mapping to “CT of right wrist” and “CT of right hand” as an alternative

291 terms, 25% have 2 or more alternative mappings, 198 exactMatch and 93 broadMatch


### MindMeetsMachine Vocabulary Results

Team	N codes	N not submitted	Agreement		
			Overall	Exact match	Broad match
AI_team_1	291	87	0.56	0.79	0.35
AI_team_2	291	4	0.60	0.60	0.38
AI_team_3	291	0	0.32	0.38	0.19
AI_team_4	291	0	0.55	0.66	0.32
AI_team_5	291	5	0.57	0.71	0.28
AI_team_6	291	0	0.58	0.68	0.35
AI_team_7	291	0	0.45	0.55	0.24
AI_team_8	291	0	0.70	0.83	0.43
AI_team_9	291	0	0.42	0.49	0.29
AI_team_10	291	0	0.70	0.86	0.51
AI_team_11	291	99	0.45	0.54	0.29
Humans	232	NA	0.58	0.64	0.44

- 01:00-02:00 JST 🥳
- 成績は3位でした 🎉
  - 人間チームに勝ちました
- 人間チーム
  - 50人以上が参加し1hで約80%ほどマッピング完了
- AIチーム
  - 正解・不正解の傾向分析を実施し得意不得意の分類を整理
- Usagi (AI\_team\_3)
  - どのチームより低いスコアとなった
- 総評
  - Exact MatchはLLMの方が総じて高めの傾向があり、Broad Matchは人間の方が高い傾向にあった

<https://www.ohdsi.org/community-calls-2026/>

## その他ワークショップで話されたこと

- このコンテストは単発イベントではなく、OHDSIの大きな取り組みの一部
  - 継続的な受け皿として GenAI & Foundational Models WG
  - ボキャブラリ構築をAIで加速するプロジェクトAriadne(今回はそのProcedure編)
- gold standard をめぐる議論
  -  から収集した352件の診療行為コード
    - 2人の専門家が独立にマッピング。意見が割れたら人目の意見を踏まえ決定
  - コミュニティ共通の gold standard を整備したい
    - 「100%正解」は存在しない/作った後も見直す前提
    - そもそもマッピングの“ルール”自体がまだ明文化されていない(整備中)
- AIと人間の役割分担
  - human-in-the-loop は必須、という共通認識
  - 現実解は「機械がスクリーニング → 専門家が最終決定」
- 残る論点(コミュニティの関心事)
  - confidence(確信度)の較正、open-source vs フロンティアモデル
  - 医療特化の埋め込み vs 汎用、評価ベンチマーク(MedHELM 等)
  - 「これは世界がまだ解けていない難問」「協力すれば前進できる」

# OHDSIのGenAI:Before = Usagi

- ボキャブラリマッピング補助ツール
- レビューやコメント機能がついてて便利
- 推薦精度は高くない
  - 古典的な文字列類似度に基づく
  - TF-IDFやLuceneを利用



Import codes from ICP2SNOMED.csv

Code	English term	Count	UMLS lookup	Dutch term
A09.00	General disease	5774012		Andere gegeneraliseerde niet gespecificeerde
K66.00	Hypertension uncomplicated	3887206		Essentiële hypertensie zonder orgaan
R44.00	Preventive Immunisations/Medications	3702922		Immunisatie/preventieve medicatie
T90.02	Diabetes mellitus type 2	2275799		Diabetes mellitus type 2
R05.00	Cough	1268829	4158493	Hoesten
R74.00	Upper respiratory infection acute	1061504		Acute infectie bovenste luchtwegen
A29.00	General symptom/complaint other	1035167		Andere algemene symptomen/klachten
L03.00	Low back symptom/complaint	998249		Lage rug pijn zonder uitstraling (ex. L8/S1)
L74.00	Cystitis/urinary infection other	970719		Cystitis/urinewegs infecties
A60.00	Results Tests/Procedures	903897		Uitslag onderzoek/verrichting
R07.00	Allergic rhinitis	892467	257007	Hoofkortst/allergische rhinitis
A97.00	No disease	868585		Geen ziekte
R96.00	Asthma	859684		Astma
P76.00	Depressive disorder	859660	440383	Depressie
P05.00	Sleep disturbance	771549		Slaapstoornis/andere slaapproblemen
R95.00	Chronic obstructive pulmonary disorder	764354	4188974	Emfysem/COPD
S88.00	Dermatitis contact/allergic	736689		Contact eczeem/vander eczeem
R78.00	Acute bronchitis/bronchiolitis	728202	4110022	Acute bronchitis/bronchiolitis
D12.00	Constipation	723752	75850	Oostpajle
A04.00	Weakness/fatigue general	721562		Moeheid/zwaakte
H81.00	Excessive ear wax	715031	374375	Overmatig cerumen
S03.00	Warts	671915	140641	Wraffen
L15.00	Noise symptom/complaint	623705		Kluis symptomen/klachten

Column mapping

Source code column: Code

Source name column: English term

Source frequency column: Count

Auto concept ID column: UMLS lookup

Additional info column: Dutch term

Filters

Filter by user selected concepts

Filter standard concepts

Include source terms

Filter by concept class: 2-dig nonbill code



Filter by vocabulary: ABMS

Filter by domain: Condition

Cancel Import

# 文字列類似度ベースのマッチング

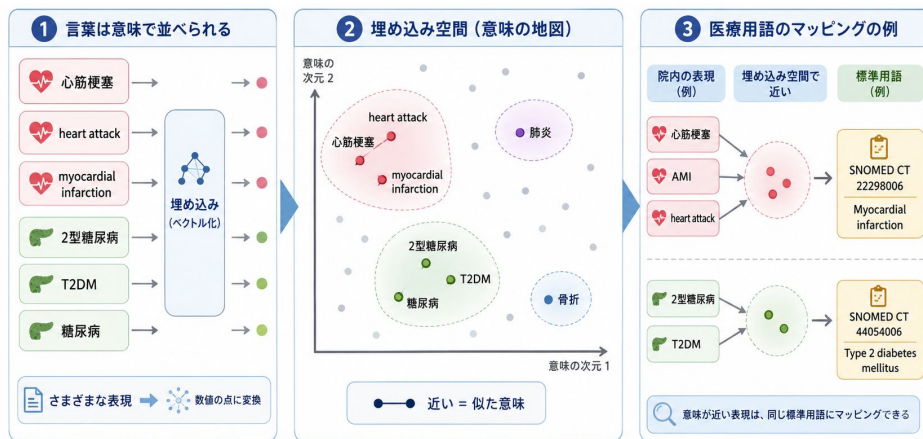
「文字の類似」に依存し類義語・多言語・否定・lateralityなどに弱い  
ただし結果の解釈性は高い(人間がみてなぜマッチしたか理解可能)

-  向いてる
  - Serum potassium → **Potassium** [Moles/volume] in **Serum/Plasma** (LOINC)
-  苦手
  - **同義語・言い換え**: Heart attack → Myocardial infarction
  - **略語**: CABG → Coronary artery bypass graft
  - **lateralityや部位の取り違い**:
    - X-ray of **wrist** → X-ray of **hand/forearm** 等に誤マッチ
    - **Left** total knee replacement → right・unspecified に誤マッチ
  - **否定・修飾**: Absence of spleen が Spleen/脾臓系にマッチ
  - **多言語**: ギリシャ語 ΣΤΕΦΑΝΙΟΓΡΑΦΙΑ、日本語 胃全摘術
    - そもそもUsagiは非英語コードを英語に翻訳しない(Google翻訳等での前処理を推奨)
- Athenaはまた別(全文検索エンジンSolr)

# 埋め込みベクトルベースのマッチング

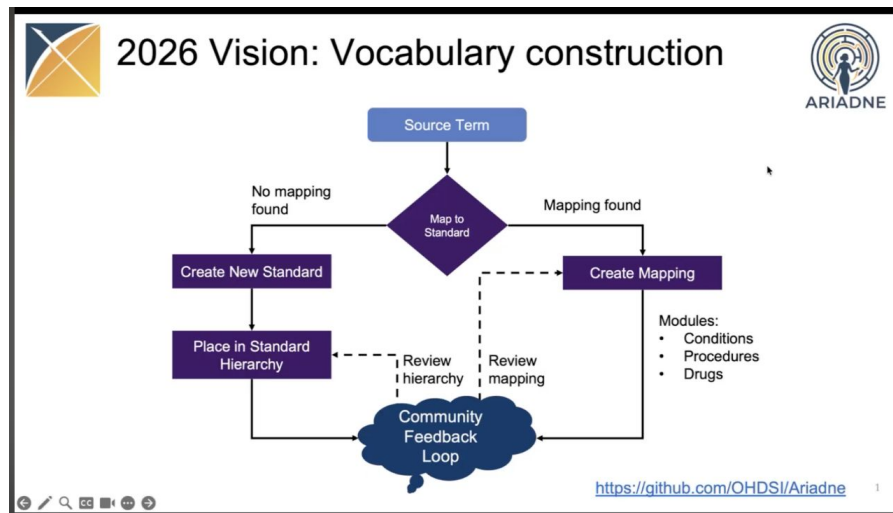
「ベクトルの類似度」に依存し類義語・多言語・否定・lateralityに強い

- LLMではRetrieval Augmented Generation(RAG)によく用いられている
- 与えた文がどういうベクトルになるか(良いマッチになるか)はモデル次第
  - 略語や同義語、多言語を学習させた LLMなら近いベクトルになる、etc
- 結果の解釈性は高くない。人間が結果を見ても直感に反することも



# OHDSIのGenAI 2026年のボキャブラリ構築ビジョン

- ソース語彙に対し、
  - 事前検証されたLLMパイプラインでマッピングを探索
  - ある→リレーションを定義
  - ない→標準コンセプトを策定し適切な階層に配置
- これをコミュニティのフィードバックループを受け改善していく



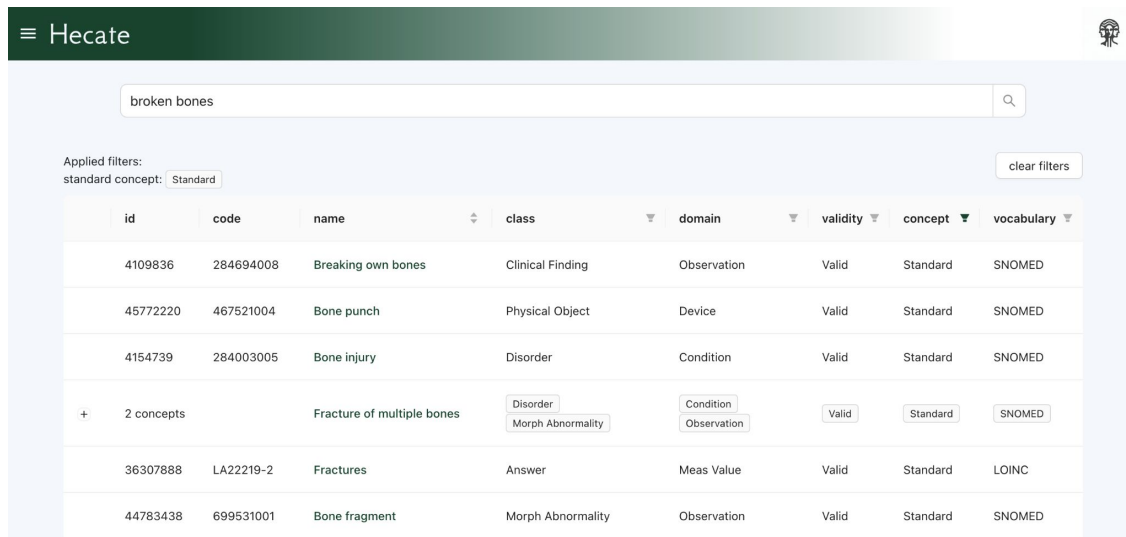
Ariadne: Automated vocabulary mappings using AI  
<https://www.youtube.com/watch?v=dj1m5Y22rAc>

# Hecate

- 埋め込みベクトルベースの標準ボキャブラリ検索エンジン
- GPT-o3でコンセプトの埋め込みベクトルを生成
- 検索UI / API / MCPを公開
- 認証不要、無料で使える

<https://hecate.pantheon-hds.com>

UsagiよりAthenaより”文字”ではなく”意味”を拾いやすい



The screenshot shows the Hecate search interface. At the top, there is a search bar containing the text "broken bones". Below the search bar, there are "Applied filters:" and "standard concept: Standard" with a "clear filters" button. The main content is a table of search results.

id	code	name	class	domain	validity	concept	vocabulary
4109836	284694008	Breaking own bones	Clinical Finding	Observation	Valid	Standard	SNOMED
45772220	467521004	Bone punch	Physical Object	Device	Valid	Standard	SNOMED
4154739	284003005	Bone injury	Disorder	Condition	Valid	Standard	SNOMED
+ 2 concepts		Fracture of multiple bones	Disorder Morph Abnormality	Condition Observation	Valid	Standard	SNOMED
36307888	LA22219-2	Fractures	Answer	Meas Value	Valid	Standard	LOINC
44783438	699531001	Bone fragment	Morph Abnormality	Observation	Valid	Standard	SNOMED

# ClaudeでHecate MCPを使ってみる

MCP = LLMが外部サービスと連携するための標準技術

Claude / ChatGPT / Geminiなどなど主要LLMプロバイダーは大体MCPをサポート

日→英でも検索できる。ただしノイズは多い

ガイド:<https://github.com/OHD/Sl/Hecate/blob/main/mcp/README.md>

Hecateで2型糖尿病・多発糖尿病性合併症ありに最も類似する標準コンセプトを探して推薦して

Hecateで類似する標準コンセプトを検索する戦略を立案した。 >

Hecateで検索します。これはMEDIS傷病名（2型糖尿病で複数の糖尿病性合併症を伴うもの）に相当する内容なので、標準コンセプト（SNOMED,Condition）に絞って探します。

多発合併症を捉えつつ、2型特異性を検索中 >

上位候補は出しましたが、原語が「2型糖尿病」と型を明示しているため、型を含むより特異的な標準コンセプトがないか追加で確認します。

標準コンセプトマッピングを検証し、ICD10CMコードを確認した。 >

「2型」と「多発合併症」の両方を完全に満たすSNOMED標準コンセプトは検索上位に現れていません。原語のMEDIS傷病名はICD-10のE11.7（2型糖尿病・多発合併症を伴うもの）に相当するため、ICD10CM側からの「Maps to」標準コンセプトを事実確認します。

医学用語の標準化マッピングを検討し、最適な概念を追求した。 >

ICD-10-CM（米国版）にはWHO ICD-10のE11.7に対応する単一コードが存在せず、ICD側からの直接の型一致は得られません。SNOMED側で「2型」かつ「多発合併症」を満たす標準コンセプトが実在するか、最有力候補（44793113）の階層を展開して子概念を確認します。

複数の概念を照合し、最適な標準コンセプトを特定中。 >

「多発合併症」を表すSNOMED標準コンセプト（44793113）は親が「Diabetes mellitus」のみで型を特定していません。最後に、「2型」と「多発合併症」を1概念で同時に満たす標準コンセプトが本当に存在しないかを、語彙制限を外して確認します。

標準コンセプト候補を検証し、トレードオフを提示した。 >

Hecateで複数の言い回しで検索した結果、事実として次のことが確認できました。

「2型」と「多発糖尿病性合併症」を1つで同時に満たすSNOMED標準コンセプトは存在しません。どの問い合わせでも、両方の意味を備えた単一の標準Disorderは返らず、トレードオフのある2つの方向に分かれます。

第一候補（意味的類似度が最高）

- **Diabetes mellitus with multiple complications / concept\_id 44793113** (SNOMED, code 385041000000108, Standard "S", Disorder)
- Hecateの類似度スコアが全クエリで最上位（最大0.94）。「多発合併症」という核を正確に表現します。
- ただし関係性を確認すると、この概念の上位は **Is a -> Diabetes mellitus (201820)** のみで、型（2型）を特定していません。「2型」の情報は欠落します。

型を優先する場合の候補

- **Type 2 diabetes mellitus / concept\_id 201826** (SNOMED, code 44054006, Standard "S")
- 「2型」は正確ですが「多発合併症」の情報が欠落します。OMOPで最も使用実績の多い概念です（record\_count約3.5億）。

「2型+合併症」を含む唯一の標準概念群はNebraska Lexiconの"Poorly controlled type II diabetes with complications"(3194082)などですが、いずれも「poorly controlled（コントロール不良）」という原語にない条件が付き、record\_countが0（実運用で未使用）のため非推奨です。

補足（マッピング経路の事実）：原語のMEDIS傷病名はWHO ICD-10のE11.7（2型糖尿病・多発合併症を伴うもの）に相当しますが、OMOPに記載されているのは米国版ICD10CMで、これにE11.7単一コードは存在せず（E11.6/E11.8等に分割）、ICD側から型一致の標準概念へ直接マップする経路もありませんでした。

推薦

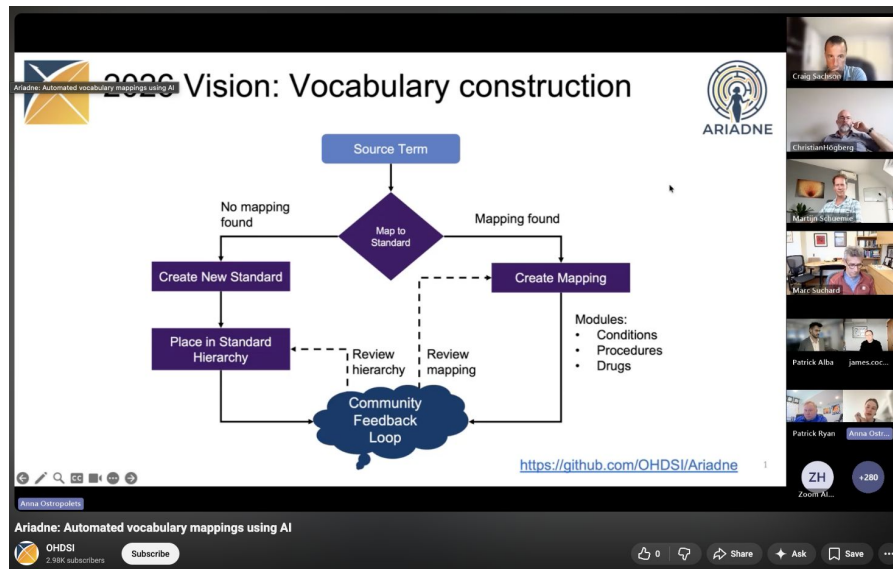
「最も類似する標準コンセプト」という観点（＝原語の主病態「多発合併症」の保持を優先）であれば、44793113「Diabetes mellitus with multiple complications」を第一候補として推薦します。Hecateの類似度でも最上位で、標準SNOMED Disorderとして使用実績もあります（record\_count,1,350）。

一方、Yuimediの傷病名→ICD-10→SNOMED CTマッピングのように「2型」という型情報の欠落が研究上問題になる用途では、201826（Type 2 diabetes mellitus）を主概念とし、合併症は別レコードでポストコーディネーション（個別の合併症概念を併記）する設計の方が情報損失が小さくなります。

どちらの方針（合併症優先か、型優先か）でいきますか。方針が決まれば、各合併症を個別概念に展開する場合の候補も同じくHecateで洗い出します。

# Ariadne: OHDSIのボキャブラリ構築プロジェクト

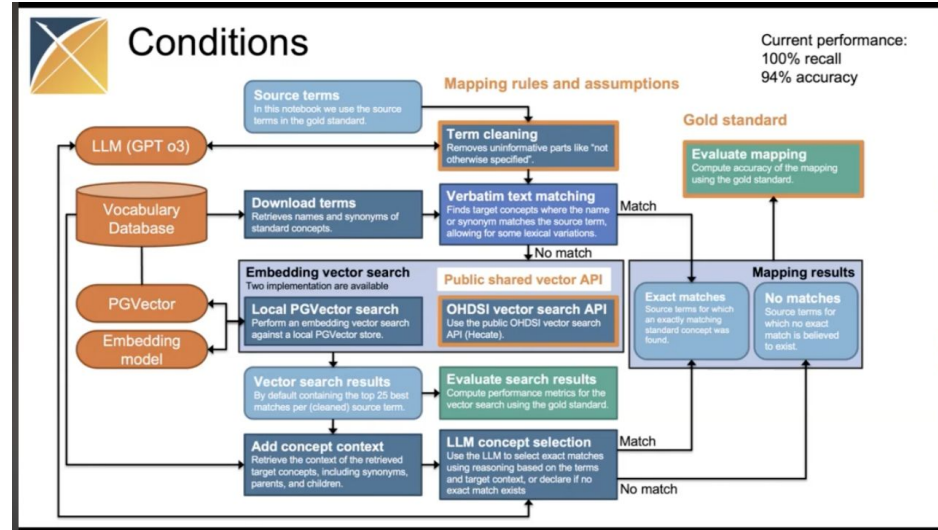
- ソース語彙
- →(Hecate等で)候補抽出
- → LLMがベストマッチ選択
- → 無ければ標準コンセプトを策定し適切な階層に配置(=RAG+人手フィードバック)
- 現状開発しているモジュール
  - Conditions
  - Procedures
  - Drugs



Ariadne: Automated vocabulary mappings using AI  
<https://www.youtube.com/watch?v=dj1m5Y22rAc>

# Ariadneのアプローチ: Conditions

- **Input:** gold standard内のソース語
- **Term cleaning:** 「not otherwise specified」のような非本質的な語を除去
- **Verbatim text matching:** 標準概念の名称・シノニムに一致すればマッチ、不一致ならRAG開始
- **Embedding vector search:**
  - Local PGVector(ローカルの pgvectorストアで検索)
  - OHDSI vector search API = Hecate(公開の共有ベクトル API)
- **Retrieval:** 1用語あたり上位25候補を取得 → gold standardで検索性能を評価
- **Add concept context:** 候補概念の文脈(シノニム・上位/下位概念)を付与
- **LLM concept selection(GPT o3):** 文脈をもとにLLMが exact matchを選択、無ければ「該当なし」と判定
- 出力は Exact matches / No matches に振り分け → gold standardで精度評価
- 現状: recall 100%、accuracy 94%

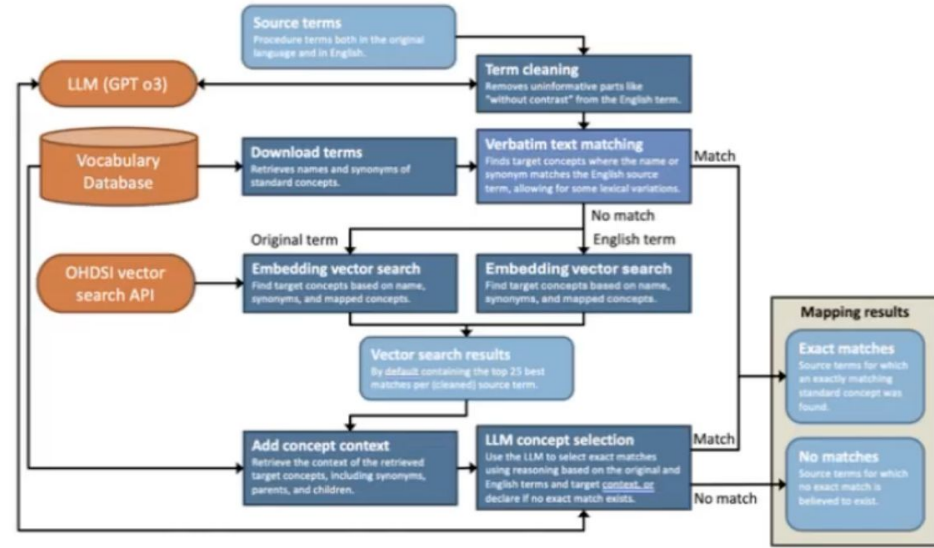


同動画、1:44~

<https://www.youtube.com/watch?v=dj1m5Y22rAc>

# Ariadneのアプローチ: Procedures

- **Input:** 原語と英語の両方のProcedureソース語(←Conditionsとちが  
いProcedureは多言語前提)
- **Term cleaning:**「without contrast」のような非本質的な語を英語から  
除去
- **Verbatim text matching:** 英語ソース語が標準概念の名称・シノニ  
ムに一致すればマッチ、不一致ならRAG開始
- **Embedding vector search を2系統並行:**
  - Original term(原語のまま)で検索
  - English term(英語訳)で検索
  - いずれも OHDSI vector search API(Hecate) を使用、名称・シノニ  
ム・mapped concept から候補抽出
- **Retrieval:** 既定で上位25候補
- **Add concept context:** 候補概念の文脈(シノニム・上位/下位概念)  
を付与
- **LLM concept selection(GPT o3):** 原語+英語+候補の文脈から  
推論してexact matchを選択、無ければ「該当なし」
- 出力は Exact matches / No matches に振り分け
- ※Conditionsとの違い=原語/英語の二経路検索で多言語に対応。た  
だしProcedureはマッピング規約が未整備で、性能の難所

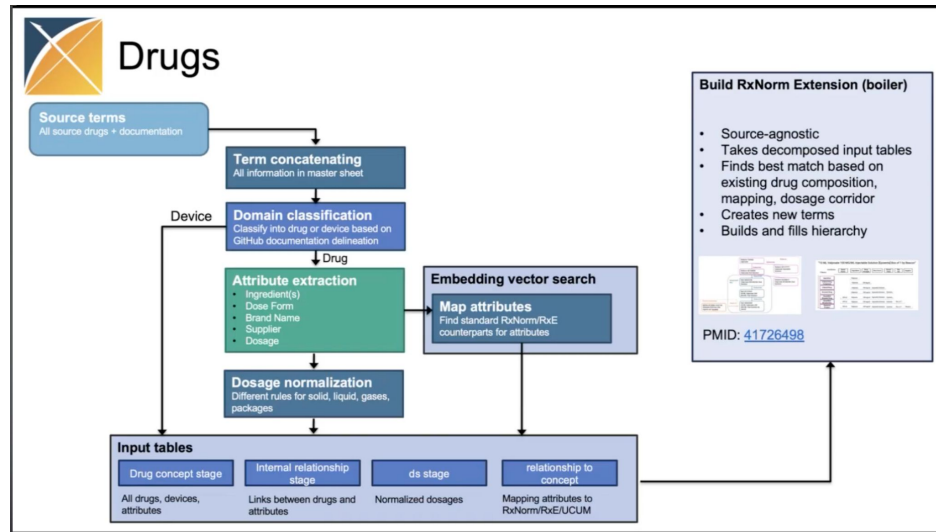


同動画、4:33~

<https://www.youtube.com/watch?v=dj1m5Y22rAc>

# Ariadneのアプローチ: Drugs

- **Input:** 全ソース医薬品語+ドキュメント→ Term concatenating(全情報をマスターシートに集約)
- **Domain classification:** GitHubの定義に基づき Device / Drug に分類
- **Attribute extraction(属性抽出):** Ingredient/Dose Form/Brand Name/Supplier/Dosage
- **Dosage normalization:** 固形・液体・気体・パッケージで異なるルールに従い用量を正規化
  - **Map attributes:** 各属性をEmbedding vector search で RxNorm / RxNorm Extension の標準属性に対応づけ(=文字列丸ごとでなく属性単位でマップ)
- **生成物=source-agnosticな入力テーブル群:**
  - Drug concept stage(全drug/device/属性)
  - Internal relationship stage(drugと属性の関連)
  - ds stage(正規化済み用量)
  - relationship to concept(属性→RxNorm/RxE/UCUM)
- **Build RxNorm Extension(boiler) に投入(決定的なSQL):**
  - source非依存/分解済み入力テーブルを受け取り、既存のdrug構成・属性マッピング・用量レンジから最適マッチを探索
  - 既存に無ければ新規概念を作成し、階層も構築・充填



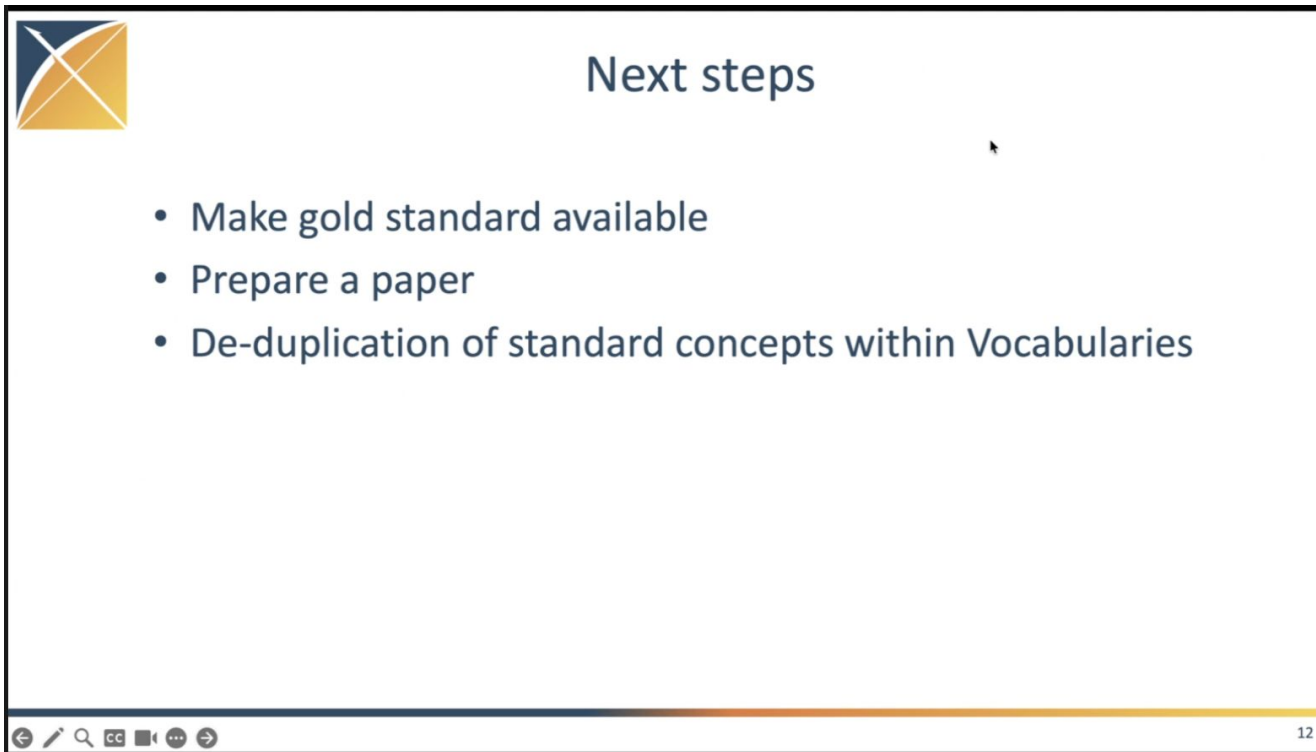
同動画、5:45~

<https://www.youtube.com/watch?v=dj1m5Y22rAc>

SQLの詳細: [Developing RxNorm Extension: A Step Toward Global Drug Data Harmonization in Observational Drug Research](https://pubmed.ncbi.nlm.nih.gov/41726498/)  
<https://pubmed.ncbi.nlm.nih.gov/41726498/>

## WSまとめ:今後の動き

- 

A slide titled "Next steps" with a logo in the top left corner. The logo is a square divided into four quadrants by a white 'X', with the top-left and bottom-right quadrants in blue and the top-right and bottom-left quadrants in orange. The slide contains a bulleted list of three items. At the bottom left, there is a navigation bar with icons for back, search, and other controls. At the bottom right, the number "12" is displayed.

Next steps

- Make gold standard available
- Prepare a paper
- De-duplication of standard concepts within Vocabularies

12

# まとめ

- OSVの再整備が動いている、人力の頑張りだけでなく仕組みを変える動き
- LLMにボキャブラリマッピングさせる方法論が徐々に確立されてきている
  - まだ領域ごとの正攻法は見え切っていない。依然として分野ごとの研究は必要
  - 機械的にBest-1を出す精度は70%程度がコンテスト参加者の限界
- **ただし人間によるレビュー・修正は現時点では必須**
  - Ariadneではコミュニティによるフィードバックループで駆動しようとしている
  - 井上はBest-1の選出精度だけでなく、いかにHuman-in-the-loopで効率的に高品質な表を生成できるかが現実解として求められている認識
- Gold standardの共用化は重要な一手、ただし作成が難しい
  - 共通の指標で画一的にパイプラインの性能を測れる
  - ”完璧”の定義は非常に難しく、多くの人手と時間が必要
- Ariadne、HecateによるOSVの品質向上の取り組みに期待
- Community CallやGenAIの活動にリアルタイム参加したいけど、タイムゾーンが合わない...