Hybrid approaches combining automation using LLM systems and expert validation enable scalable and governance-ready integration into vocabulary management and distribution system

Title: Jackalope Plus Adventures: Benchmarking Automated Mapping and Charting a Path to OHDSI vocabulary distribution

Objectives

- Benchmark automated vs expert mappings
- Evaluate agreement with a gold-standard dataset
- Propose a governance-ready framework for mapping browser and distribution system integration

Methods

Dataset: 749 source terms, multiple vocabularies (ICD, MedDRA, etc.)
Mapping approaches:

- Jackalope Plus (automated)
- 3 independent expert mappers

Evaluation:

- MATCH
- PARTIAL
- NON-OVERLAPPING

Only exact and broad matches allowed; hierarchical mappings excluded.

Results

Mapper	Time	MATCH	PARTIAL	NON-OVERLAP
Jackalope	16 min	13.08%	63.15%	23.77%
SME 1	16 h	67.82%	20.83%	11.35%
SME 2	7 h	62.62%	11.61%	25.77%
SME 3	12 h	54.34%	30.31%	15.35%

Mapping Interpretation

Category	Interpretation	Key Insight
MATCH	Exact agreement with gold standard concept set	High confidence mapping
PARTIAL	At least one overlapping concept with gold standard	Clinically meaningful but alternative representations exist
NON-OVERLAPPING	No shared concepts with gold standard	Does not imply incorrect mapping

Drivers of Differences

Alternative granularity (parent vs descendant concepts)

Sibling concepts within same hierarchy

Composite term decomposition strategies

Modeling strategy differences

Key findings

- **Automation = extreme speed (minutes vs hours)**
- **Lower exact match vs experts**
- **High PARTIAL overlap → clinically meaningful signals**
- **High variability between experts → mapping is not deterministic**

Limitations

- Evaluation is based on a single curated benchmark dataset, which may not fully capture real-world variability across all domains and vocabularies
- Strict set-based comparison may underestimate semantically valid alternative mappings
- Automated mapping was evaluated in a zero-tuning setting, without task-specific optimization
- Expert mappings exhibit inter-observer variability, reflecting the non-deterministic nature of concept mapping
- Hierarchical relationships were excluded from evaluation, limiting assessment of broader semantic alignment

Background

- High-quality, reproducible medical concept mapping is essential for OMOP-based research. Automated tools are increasingly used to accelerate ETL workflows, but their performance relative to expert mapping remains unclear.
- This study evaluates Jackalope Plus, an automated mapping system, against expert-generated mappings using a community-curated benchmark dataset aligned with OHDSI standards.

Mapping Workspace Hybrid Model

- The proposed workflow integrates automated mapping with independent expert validation in a structured, governance-aligned pipeline.
- Automated tools generate candidate concepts, which are subsequently reviewed, compared, and refined by experts, followed by documentation and controlled integration into vocabulary management system.
- This hybrid approach enables scalable mapping while ensuring semantic accuracy and reproducibility.

Operational Modes of the Mapping Workflow

The Mapping Workspace supports three operational modes corresponding to different levels of automation and governance:

- Manual mapping - fully expert-driven
- AI-assisted mapping - automation with optional expert involvement
- AI with structured QA - automation combined with formalized multi-expert validation

Conclusion

- Automated mapping is not sufficient alone
- Highly effective as first-pass accelerator
- Hybrid approach enables:
  - Scalability
  - Reproducibility
  - Governance-aligned contributions
- Proposed Mapping Workspace supports traceable and standardized vocabulary integration into vocabulary management system

Denys Kaduk
Polina Talapova
Maksym Trofymenko
Tetiana Nesmiian
Daryna Ivakhnenko
Anna Ostropolets
Mariia Pahur
Bohdan Khilchevskyi
Max Ved
Inna Ageeva

sciforce

OHDSI
Observational Health Data Sciences and Informatics