

An Ontology-Informed Strategy for Medication Interoperability in Real-World Data: From Analytic Feasibility to Conceptual Foundations

Haowen Hsu, PharmD, MPH¹, Gabriel Gazetta, MS², Chi-Hua Lu, PharmD, MS¹, David M. Jacobs, PharmD, PhD¹

¹ Department of Pharmacy Practice, School of Pharmacy and Pharmaceutical Sciences, University at Buffalo, Buffalo, New York, United States

² Department of Industrial and Systems Engineering, School of Engineering and Applied Sciences, University at Buffalo, Buffalo, New York, United States

Introduction

- Electronic health records (EHRs) frequently face systemic terminology heterogeneity, which could undermine analytic fidelity and risk propagating classification errors.¹
- Standardized vocabularies (e.g., RxNorm and Anatomical Therapeutic Chemical [ATC] classification) alone do not resolve interoperability challenges and ensure analytic readiness.²
- Structured harmonization is required to translate heterogeneous identifiers into unified, reproducible representations.

Objective

To develop and evaluate an ontology-guided framework that harmonizes heterogeneous medication identifiers into standardized ingredient- and pharmacologic-class representations.

Methods

Study design & data

- Retrospective EHR analysis (2020–2024)
- Data source: Buffalo General Medical Center (Kaleida Health)
- Population: Adults aged ≥65 years with ≥1 hospitalization.
- Data were extracted from **214,080 medication records** encoded with diverse identifiers across **29,123 subjects**.

Framework architecture

- Raw data: Medication records with diverse identifiers
- Two-layer semantic mapping framework**
 - Layer 1 (ingredient):** RxNorm ingredient concept (RxCUI [IN])
 - Layer 2 (therapeutic):** ATC classification

Mapping strategy

Strategies	Identifiers	Resources/Process
Code-based	NDC, RxCUI, Multum Drug ID	RxNorm full monthly release file & TriNetX reference file
String-based	Multum Drug Synonym ID	Exact + case-insensitive substring matching

Manual validation

- For string-based and RxCUI-to-ATC matching.
- Entries were first resolved to unique records by medication name, dose, and route of administration.
- Performed by a single pharmacy expert.
- Tools: RxMix API, *Clinical Pharmacology*, WHO ATC Index.
- Reasons for RxCUI [IN] or ATC correction were summarized.

Results

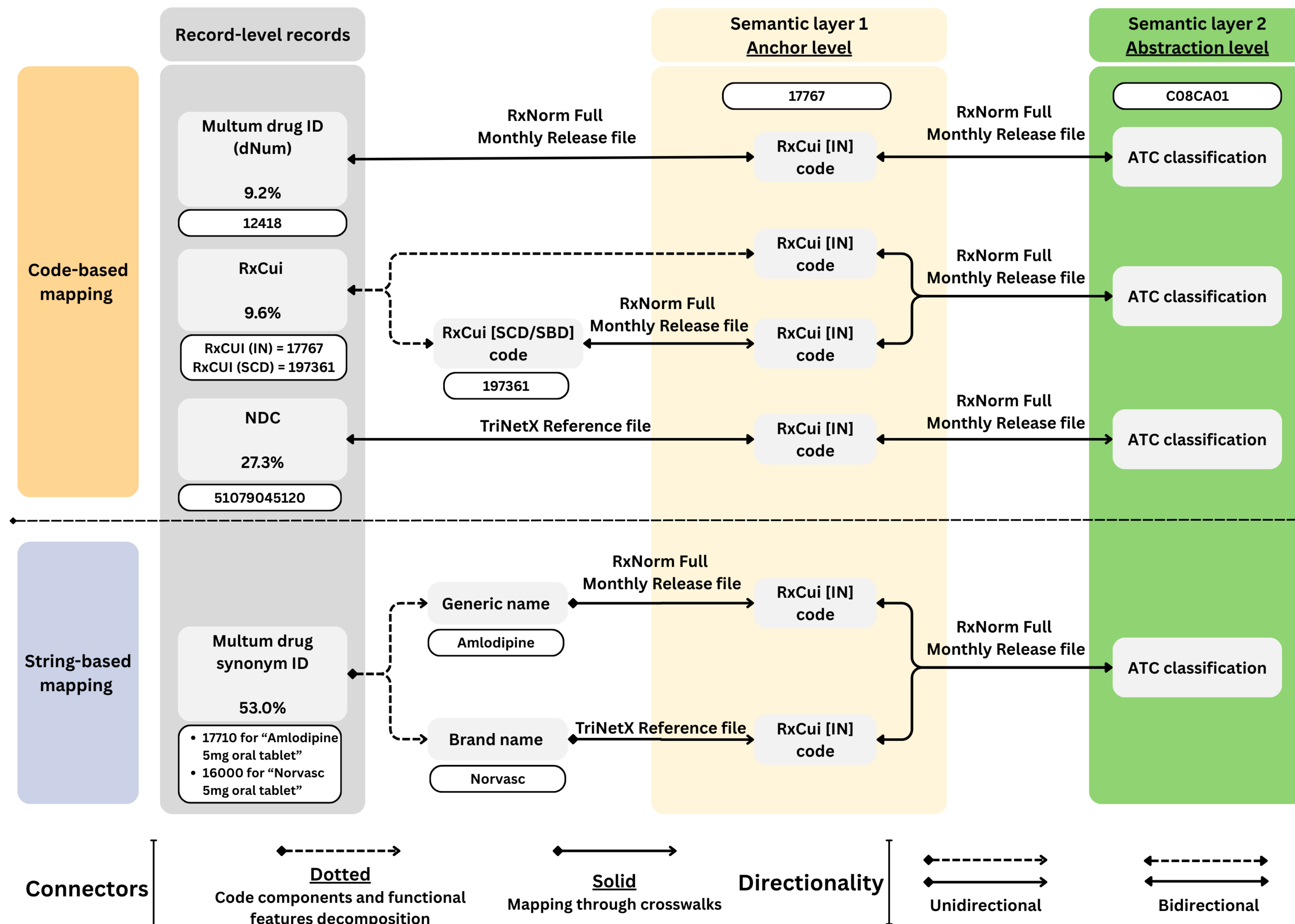
Table 1: The distribution of coding identifiers and their mapping and correction rate

Coding identifiers	Distribution		Code-based mapping rate		String-based mapping rate		Validation [Correction rate]	
	Raw	Unique	RxCUI [IN]	ATC	RxCUI [IN]	ATC	RxCUI [IN]	ATC
RxCUI	9.6%	9.1%	100%	100%	-	-	0%	31.2%
NDC	27.3%	25.4%	100%	100%	-	-	0%	32.6%
Multum drug ID (dNum)	9.2%	8.9%	100%	100%	-	-	0%	34.5%
Multum drug synonym ID	53.0%	56.6%	-	-	78.5%	75.5%	33.0%	57.4%
Missing	0.9%	-	-	-	-	-	-	-
Overall			100%	100%	78.5%	75.5%	18.7%	46.7%

Table 2: Reasons for RxCUI [IN] or ATC correction

Reasons	Layers / Mapping strategy	Definition	Examples
Crosswalk omissions	Layer 1 / String-based	Couldn't align brand names with TriNetX Reference file.	Failed to match Alivio to ibuprofen ("Alivio" did not exist in TriNetX reference file)
Lexical misclassification	Layer 1 / String-based	Inexact string-based matching caused erroneous ingredient attribution.	Recorded omeprazole for esomeprazole
Taxonomic ambiguity	Layer 2 / Both strategies	A single RxCUI [IN] may map to multiple ATC codes.	Moxifloxacin eye solution was assigned to J01MA14 (systemic) and S01AE07 (ocular anti-infectives)

Figure 1: The conceptual mapping framework for medication records harmonization: Amlodipine (Norvasc) case illustration



Discussion

- Real-world medication data is highly heterogeneous with limited interoperability when a guided framework is lacking.
- Mapping between identifiers is bidirectional. Can anchor to targeted identifiers for different analytic and clinical purposes.
- This framework integrates deterministic and string-based mapping, enhancing data completeness while preserving traceability and transparency.
- This framework requires human-based corrections and has not been fully automated.

Limitations:

- Only extracted a subset of medication records from a single medical center.
- Having a single researcher for manual validation may introduce subjective biases.

Conclusion

This ontology-informed framework:

- Establishes the architectural foundation necessary to transform raw EHR medication records into standardized representations at the ingredient and pharmacological levels.
- Provides a critical semantic foundation for transparent, transportable, and reproducible real-world medication analysis.

References

- Haendel et al., Classification, Ontology, and Precision Medicine. *N Engl J Med.* Oct 11 2018;379(15):1452-1462.
- Lester et al., *J Am Med Inform Assoc.* Aug 16 2022;29(9):1471-1479.

