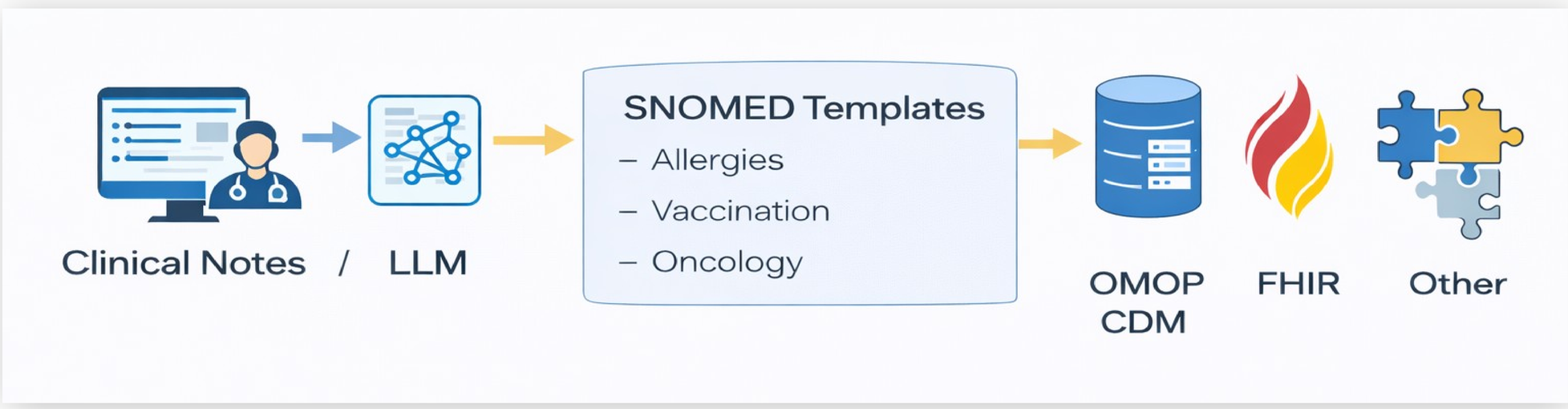
SNOMED CT Template–Based Interoperability for Multi-Standard Outputs

A semantic-first framework bridging unstructured clinical data and interoperable health data standards

Background: Traditional medical data integration relies on one-to-one mappings between source data and target models such as OMOP CDM or FHIR, often leading to loss of clinical context and limited reusability. We propose a novel interoperability framework using SNOMED CT templates as a central, model-agnostic semantic layer to preserve clinical meaning and enable multi-standard outputs.

Method

Clinical information is extracted from unstructured medical texts using a Large Language Model (Mistral Small 3), deployed on-premise to ensure data privacy. The model produces structured outputs (JSON schema) capturing entities, attributes, and relationships.



Extracted data are transformed into SNOMED CT expressions using domain-specific templates (allergies, vaccination, oncology). An automated SNOMED-isation process assigns concepts with confidence scores, supported by expert-annotated training data and validation thresholds.

Validated SNOMED expressions are then converted into multiple standardized formats, including OMOP CDM tables and FHIR resources. This pipeline is fully modular, allowing independent improvement of extraction, semantic modeling, and transformation components.

In addition, the use of structured templates ensures consistency across heterogeneous data sources while maintaining fine-grained clinical detail.

Results

The framework enables the generation of standardized, context-preserving clinical representations from free-text data. Compared to traditional mapping approaches, it significantly reduces information loss and improves data completeness.

The approach successfully captures complex clinical constructs across domains, including allergy reactions and causative agents, vaccination status and events, and oncological topography and morphology. The SNOMED-based layer allows seamless and reusable transformation into both OMOP CDM and FHIR outputs.

Preliminary evaluations show improved semantic fidelity and reduced need for manual post-processing compared to conventional pipelines.

The unified representation also facilitates extension to new domains without requiring redesign of downstream mappings.

Conclusion: This work introduces a scalable interoperability paradigm that decouples clinical meaning from technical representations. By leveraging SNOMED CT templates and LLM-based extraction, the framework preserves semantic richness while enabling flexible, multi-standard data integration. This approach supports both clinical research and secondary use of healthcare data and can be extended to additional domains and standards.

OHDSI

T. Helleputte, M. **Borshchivska**, T. **Klein**, G. **Vanhalst** , R. **Verschuren**, M. **Bastin**, O. **Latignies**, F. **Daue**, A. **Vandenberghe**