How metadata empowers MedDRA hierarchies and mappings

Mikita Salavei[1], Oleg Zhuk[1], Vlad Korsik[1], Alexander Davydov[1]
[1]Odysseus Data Services Inc., Cambridge, MA

Background: MedDRA, short for Medical Dictionary for Regulatory Activities, is a comprehensive medical terminology specifically developed for classifying adverse events and other crucial medical information, predominantly utilised within clinical trials and pharmacovigilance practice. Its integration with SNOMED CT, a cornerstone medical terminology in OHDSI Standardized vocabularies, offers substantial potential for augmenting healthcare research endeavours.

Lately, the compatibility between MedDRA and SNOMED in OMOP was hindered by a limited number of direct 'Maps to' links, constructed to align concepts based on their semantic similarity, often with varying levels of granularity. However, with the February 2024 OHDSI Vocabulary release[1], mapping metadata (inspired by the SSSOM model[2]) has been utilised to enhance the precision of mappings and to establish hierarchical links between MedDRA and SNOMED.

Methods: Our approach encompasses the utilisation of both internal resources within OHDSI, such as previously built MedDRA-SNOMED mappings and internal AI-augmented mapping approach, as well as tapping into external sources including the Unified Medical Language System thesaurus, MedDRA-SNOMED bidirectional mappings from the Maintenance and Support Services Organization, and ICD10-MedDRA mappings. To streamline the process, all identical MedDRA concepts with differing variants of potential mappings to SNOMED were methodically grouped and subjected to manual review by medical terminologists to discern the most optimal mapping choice with the help of a Common Data Environment (CDE)[3]. For each finalized variant of mapping, a unique relationship_id_predicate was defined as metadata, capturing three levels of semantic similarity: 'Maps to equivalent', 'Maps to uphill', and 'Maps to downhill' (Table 1).

The result of the mapping process involved the creation of hierarchical links between MedDRA and SNOMED based on the metadata, ensuring integration of two medical terminologies within the OMOP framework into a single hierarchy.

Table 1. Description of metadata relationship_id_predicates and typical examples of MedDRA-SNOMED mappings

Relationship_id full	Description	Concept name	Vocabulary	Relationship_id	Relationtionship_id_predicate	Concept code	Concept name	Vocabulary
Maps to equivalent	Standard full equivalent 'Maps to" with no data loss. The two terms are intended to refer the same thing	Adrenocortical insufficiency acute	MedDRA	Maps to	equivalent	766986002	Acute adrenal insufficiency	SNOMED
		Congenital pulmonary hypertension	MedDRA	Maps to	equivalent	1010627004	Pulmonary hypertension due to developmental abnormality	SNOMED
Maps to uphill	The source concept is a narrower term than the target concept. Data loss happens. Typical scenario when no exact match can be found	Anastomotic ulcer haemorrhage	MedDRA	Maps to	uphill	74474003	Gastrointestinal hemorrhage	SNOMED
		Anastomotic ulcer haemorrhage	MedDRA	Maps to	uphill	447408004	Ulcer of anastomosis	SNOMED
		Groin infection	MedDRA	Maps to	uphill	40733004	Disorder due to infection	SNOMED
		Groin infection	MedDRA	Maps to	uphill	118936007	Disorder of inguinal region	SNOMED
Maps to downhill	Rare scenario when the source concept is broader than the target concept. It should not happen generally if not stated otherwise	Suture rupture	MedDRA	Maps to	downhill	217008000	Suture failure during surgical operation	SNOMED
		Epstein-Barr virus test	MedDRA	Maps to	downhill	408219003	Epstein-Barr virus serology	SNOMED

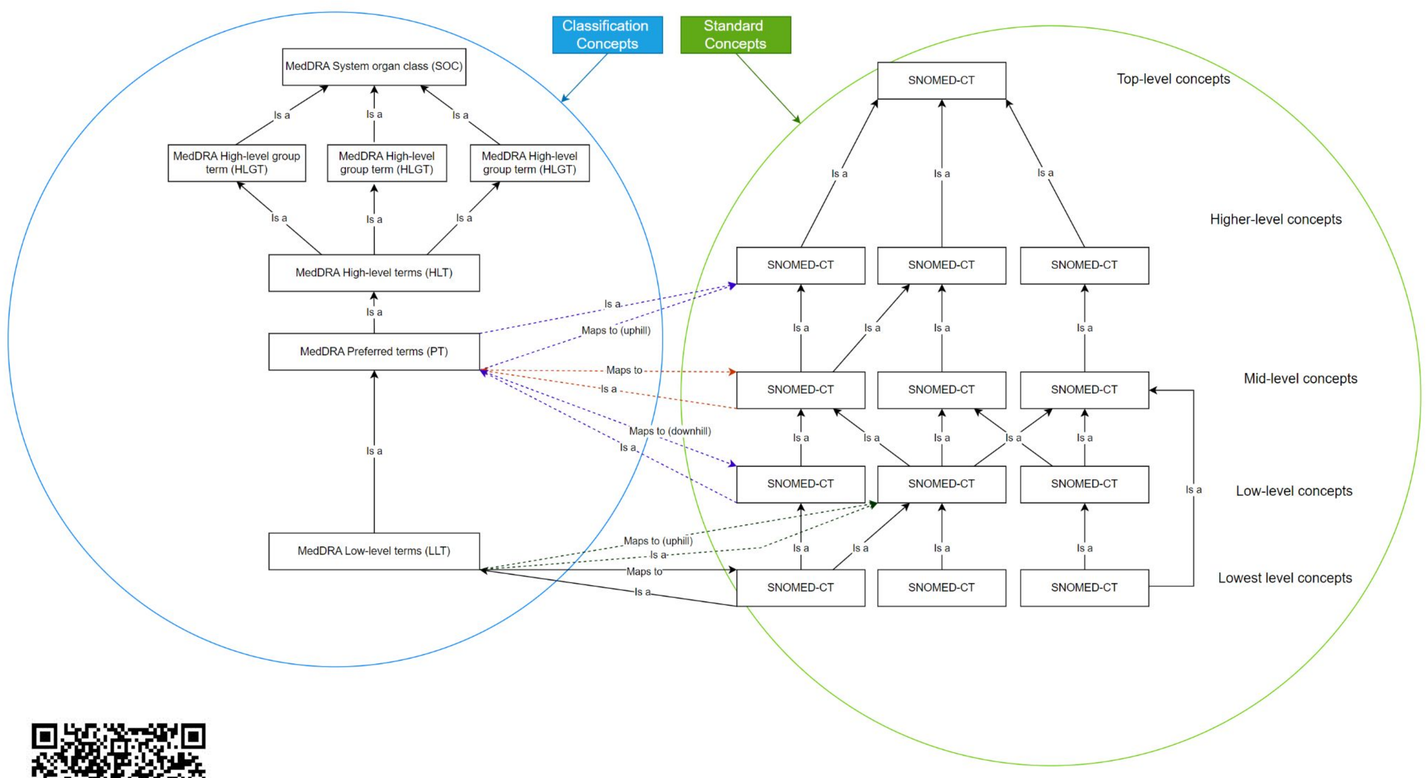
Results: Our primary focus centred on the Preferred Term (PT) level of MedDRA concepts due to their enhanced suitability for analytics needs. In addition, we add and improve Lowest Level Term (LLT) mapping. This work led to the addition of 10,189 PT and 2,874 LLT mappings of concepts to the OHDSI standardized vocabularies (as per February 2024 release).

Furthermore, hierarchical links, denoted by 'Is a' and 'Subsumes' relationships, were systematically constructed based on the following principles.

In cases of complete semantic correspondence between MedDRA and SNOMED concepts, the MedDRA concept was strategically positioned above its corresponding SNOMED counterpart. However, in cases of incomplete correspondence, diverse scenarios emerged wherein the MedDRA concept could be hierarchically higher (downhill mapping to SNOMED) or lower (uphill mapping to SNOMED) compared to the corresponding SNOMED concept.

Finally, a hierarchy between MedDRA and SNOMED / OMOP Extension has been constructed with 25,623 'Is a'/'Subsumes' links (Figure 1)

Figure 1. Hierarchical links between MedDRA and SNOMED vocabularies constructed based on metadata



Scan QR code to view eposter

Conclusion: The strategic emphasis placed on harnessing metadata to enhance MedDRA mappings and establish new hierarchies within the OMOP framework signifies a notable step forward in OMOP terminology improvement. MedDRA, retaining its status as a classification vocabulary, serves dual purposes within this framework. Firstly, it facilitates the mapping of source data from MedDRA codes to OMOP Standard concepts, leveraging 'Maps to' and 'Maps to value' links. Secondly, MedDRA codes are utilized as classification concepts for constructing concept sets based on hierarchical links. This is the first example of mappings metadata implementation in OHDSI vocabulary development process with a practical effect on the content in the official release.

References:

1. https://github.com/OHDSI/Vocabulary-v5.0/releases/tag/v20240229_1709217174.000000
2. O. Zhuk, A. Ostropolets, N. Matentzoglu, M. Haendel, D. Gabriel, A. Davydov, C. Reich Community Contribution to the OHDSI Vocabularies, User-Level QC and a New Entity Mapping System SSSOM, Conference: OHDSI European Symposium 2023.
3. I. Zherko, M. Nerovnya, M. Kallfelz, A. Davydov Common data environment for source vocabularies mapping, Conference: OHDSI European Symposium 2022.