Extending OHDSI with Preclinical Terminologies

Erik M. van Mulligen, Marcel de Wilde, Solène Grosdidier, Jan A. Kors, and Johan van der Lei

Dept of Medical Informatics, Erasmus University Medical Center, Rotterdam, The Netherlands

Background

The Enhancing TRANslational SAFEty Assessment through Integrative Knowledge Management (eTRANSAFE) project develops an integrative data infrastructure and innovative computational methods and tools that aim to drastically improve the feasibility and reliability of translational safety assessment during the drug development process. This infrastructure will be underpinned by development of open standards and robust policies widely accepted by stakeholders, including regulatory agencies and international organizations. Prerequisite of an improved translational safety assessment is a thorough understanding of the preclinical-clinical concordance. One of the challenges for such a concordance analysis is to combine preclinical (animal) data with clinical (human) data where both have their own typical terminologies. We describe here our approach to map the preclinical terminologies (Histopathology, Clinical Pathology, and selected SEND code lists) to the clinical terminologies (MedDRA).

Methods

The preclinical terminologies have been integrated in the OMOP's Common Data Model (CDM) and the clinical terminologies MedDRA and SNOMED-CT are already contained in the CDM. SNOMED-CT is a knowledge source that can decompose pre-coordinated concepts to e.g. *adverse event* and *organ* concepts. We use SNOMED-CT as an intermediary to bridge between the preclinical and clinical terminologies (see Figure 1). Suggested mappings have been obtained with OHDSI (Usagi), EBI (OLS, OxO, and ZOOMA), NLM (UMLS), BioPortal, and the CDM. Experts have manually verified these mapping suggestions for liver related adverse events prioritized by the frequency of Histopathology and Clinical Pathology concepts from eToxSys and MedDRA concepts from ClinicalTrials.gov.

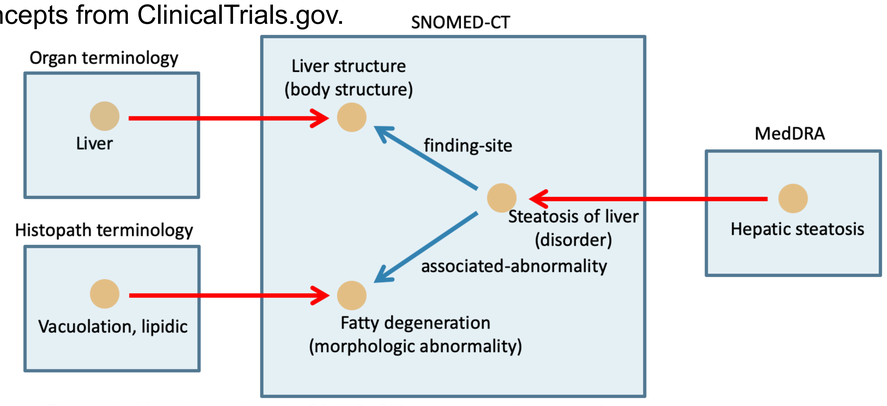

Figure 1. Mapping using SNOMED-CT as intermediary terminology

The preclinical vocabularies and their mapping have been integrated in the CDM with "*Maps to*" and "*Mapped from*" concept relations. On top of these we have developed a mapping service that provides two functions: (1) expansion of a concept with all its children and (2) returning mapped concepts for either preclinical or clinical concepts. The latter service contains the logic to exploit SNOMED-CT to perform the mapping between pre- and post-coordinated concepts.

Results

The results for the initial mapping are shown in Table 1 for those concepts present in the various clinical and preclinical databases. The last row shows the mapping to SNOMED-CT as an intermediary terminology.

Terminology	Size	Number of mapped terms (%)		
		Histopathology (n = 651)	Clinical pathology (n = 337)	MedDRA (n = 12,153)
Histopathology	1,281		NA	332 (2,8%)
Clinical pathology	337	NA		51 (0,4%)
MedDRA	80,909	184 (28,3%)	92 (27,3%)	
SNOMED-CT	454,722	334 (51,3%)	202 (59,9%)	7,358 (60,5%)

Table 1. Mapping results obtained from existing resources

We developed a mapping tool (see Figure 2) in which the on frequency prioritized adverse event concepts are shown with mapping suggestions. Experts can use this mapping tool to approve the initial mapping or define new mappings. The results for a first selection of liver related adverse event concepts are shown in Table 2. The mapped concepts are used in more than 99% of all records in preclinical and clinical databases.

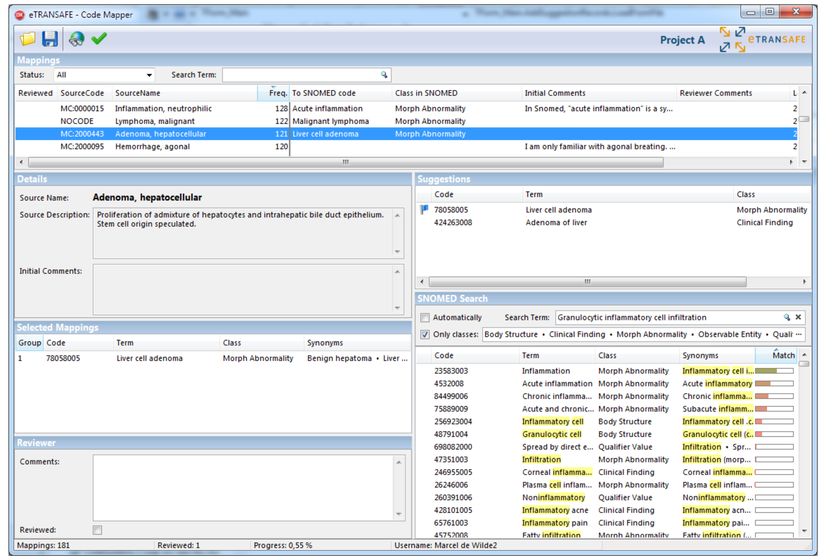

Figure 2. Mapping tool to refine mappings.

Terminology	Size	Concepts Mapped (%)	Data Mapped (%)
Histopathology	188	159 (87%)	>99%
Clinical pathology	337	279 (85%)	>99%
MedDRA	348	309 (89%)	>99%

Table 2. Mapping results after manual revision

Conclusion

The terminology database as contained in the OMOP CDM can be easily used to also store preclinical terminologies with their mappings. For the mapping we chose to use SNOMED-CT as an intermediary to bridge the differences between pre- and post-coordinated terminologies. Revision with the mapping tool brought the coverage from 50-60% for the suggested mappings to 85-90% for the liver domain. These preclinical terminologies and their mappings have been made available in a web service that provides query expansion and query translation.

Contact:e.vanmulligen@erasmusmc.nl

We acknowledge funding from eTRANSAFE which has received support from IMI2 Joint Undertaking under Grant Agreement No. 777365.