

## Background

The healthcare systems of the European Union constantly generate a myriad of disparate electronic medical data, which due to its incompatibility and interoperability is not easy to analyse using standardized and systematic approaches. This is the result of different languages, healthcare constellations, EHR systems and policies [1]. The Electronic Health Data in a European Network (EDHEN) consortium is aimed at addressing the problems of data fragmentation: It will provide a harmonised model to address the structural heterogeneity and the use of different coding standards [2].

Since the standardisation of health data to the OMOP Common Data Model and the adoption of analytical tools developed by OHDSI is central to EDHEN, efficient collaboration might only be possible in a close interaction with comprehensive planning, development and maintenance of European medical vocabularies.

The aims of this investigation are:

- assessment of the breadth of the European medical vocabularies;
- estimation of its current coverage in the OMOP Standardized Vocabularies;
- development of a plan on new vocabularies implementation;
- estimation of resources needed.

## Material and Methods

To collect the necessary information, the following web-resources were accessed:

- official websites of standard committees in 16 European countries;
- the World Health Organisation's international classifications page and implementation database [3, 4];
- other open sources [5, 6].

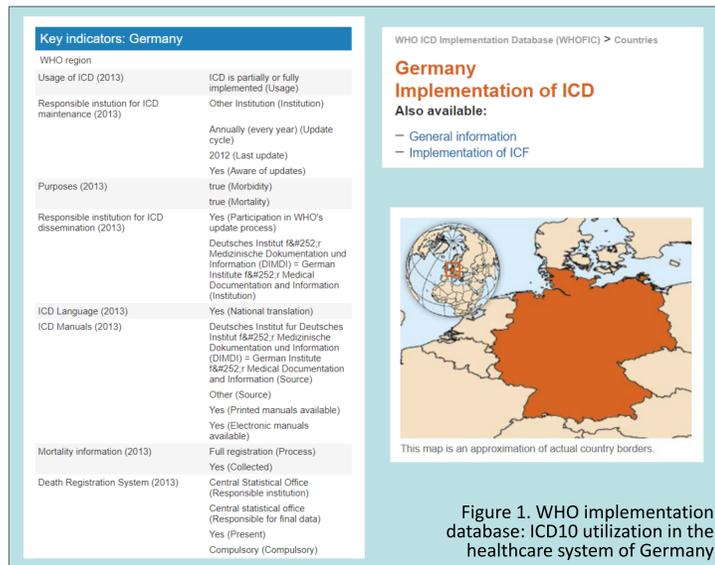


Figure 1. WHO implementation database: ICD10 utilization in the healthcare system of Germany

The latest available versions of open-access vocabularies were downloaded and compared to existing OMOP CDM concepts. To estimate the vocabularies coverage a series of SQL queries were used.

## Results

The general picture of European vocabularies coverage by OMOP CDM is characterised by the following patterns:

### 1. Lack of available structured information.

- Web resources are not perfect, often have just reduced English version.
- Not all the used vocabularies are precisely described or even mentioned.
- Even if mentioned, implementation level / application areas / current version often remains unclear.
- Source files are often not available or just text pdf version is provided.

### 2. Mostly sufficient coverage of widely used vocabularies.

- Some of them are represented as “linguistic variants”, i.e. translated to national languages without any semantic modifications (ICD9-CM, ICD10-CM, ICD-O-3, ATC, LOINC and MeSH). Once an international version of a vocabulary is implemented, all the national versions may be processed automatically storing the concept name interpretations as concept synonyms.
- Another variant of “national modification or extensions” use the original vocabularies as a basis and the application of a number of adjustments and alterations (ICD-10 and SNOMED). E.g., ICD10-GM, the German modification of ICD10-WHO is currently being revised annually and used for coding of diagnoses in the outpatient and inpatient care. Among 16,126 concept codes, 13,473 (83.5%) are equivalent to the original ICD10-WHO codes, 1,135 codes (7.0%) match to the US ICD-10-CM vocabulary (albeit with semantic variance), while 1,518 (9.4%) do not exist in any ICD vocabulary. The processing of this type of vocabularies generally needs some additional efforts, including reviewing, mapping, etc. But anyway this is much more easily than development of obscure one. It is noteworthy that an unaltered translation of ICD-10-WHO is simultaneously used in Germany for death reasons registration.

National ICD10 modifications were tested on codes equivalence considered that ICD10-WHO is taken as a basis (if ICD10-WHO code match, then it's completely equal). The US ICD10-CM can be helpful as well since codes are mostly the same, so it may be used in the future mapping (Figure 2).

## Results

Table 1. The list of European medical vocabularies and its current coverage in the OMOP CDM

Country	Domain					
	Condition	Procedure	Drug	Measurement	Device	Other/ mixed
Austria	ICD10-WHO-BMG ICD-O-3	APC APCC	ATC	LOINC-DE		LKF-PCS
Belgium	ICD9-CM ICD10-CM-BE ICD-O-3	ICD9-CM ICD10-PCS-BE	GGR ATC	LOINC-BE		APR-DRG SNOMED-CT-BE*
Czech Republic	ICD10-CZ (MKN) ICD11* ICF ICD-O-3-CV ORPHANET/OMIM/SSIEM TNM8		ATC SUKL		ZT GMDN	MeSH-CZ DRG-IR SNOMED-CT-CZ* MedDRA-CZE* VZP
Denmark	ICD10-DK (SKS) ICD-O-3	NCSP-DK	ATC	NPU-DK		SKS DRG-DK SNOMED-CT-DK SNOMED-CT-ET*
Estonia	ICD10-WHO-ET ICD-O-3	NCSP-ET	ATC	LOINC-ET	EUDAMED MSA GMDN	
France	ICD10-FR (CIM)	CCAM / ACPC CSARR	BDPM ATC	LOINC-FR	LPP	MeSH-FR
Germany	ICD10-GM ICD10-WHO-DE ICD11* ICD-O-3 ICF Alpha ID-SE	OPS	AMIS	LOINC-DE UCUM	EDMA UMDNS GMDN*	G-DRG MeSH-DE
Ireland	ICD10-WHO-AM	ACHI	ATC	LOINC		ACS SNOMED-CT* CMS-DRG
Italy	ICD10-WHO-IT ICF & ICF-CY ICPC2* ICD11*	ICD9-CM ICM* CIP1*		LOINC-IT	CND GMDN	
Netherlands	ICD10-NL ICF ICPC2-NL ICECI ICD-O-3	CMSV	MIB ATC	LOINC-NL	ISO 9999-MD GMDN	MeSH-DUT SNOMED-CT-NL
Poland	ICD10-PL (MKC) ICF ICNP	ICD9-CM	ATC	LOINC*	GMDN	SNOMED-CT-PL*
Portugal	ICD9-CM ICD10-CM-PT ICF & ICF-CY ICD-O-3 ESOD	ICD9-CM ICD10-PCS	CNPEM ATC	LOINC-PT SNOMED-CT*	CDM GMDN	
Spain	ICD-10-BAQ ICD-10-CAT ICF & ICF-CY ICD10-SE NSL ATC			LOINC-ES	GMDN	SNOMED-CT-ES
Sweden	ICD10-GM ICF ICD-O-3	CHOP	ATC SwissMedic	LOINC-DE		KVA/SVF SNOMED-SV KSI NordDRG-SV MeSH-SV SwissDRG
Switzerland	ICD10-GM ICF ICD-O-3	CHOP	ATC SwissMedic	LOINC-DE		
United Kingdom	ICD10-WHO ICD-O-3	OPCS-4 NICIP	dm+d Gemsript SNOMED-DrugExt	LOINC	dm+d	SNOMED-CT-UK Read

OMOPed: OMOPed but review / slightly changes are needed; partially OMOPed; not OMOPed, but approach was developed on similar vocabulary; not OMOPed and new approach is needed; \* – not widely used / planned to be implemented.

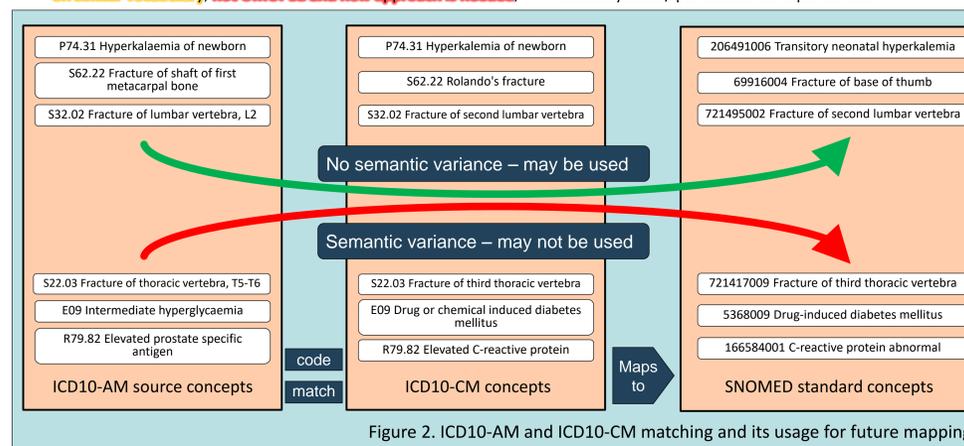
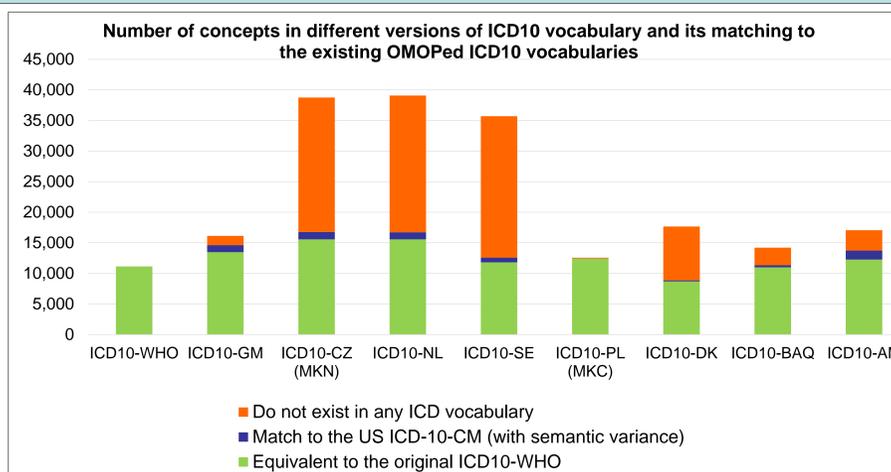


Figure 2. ICD10-AM and ICD10-CM matching and its usage for future mapping



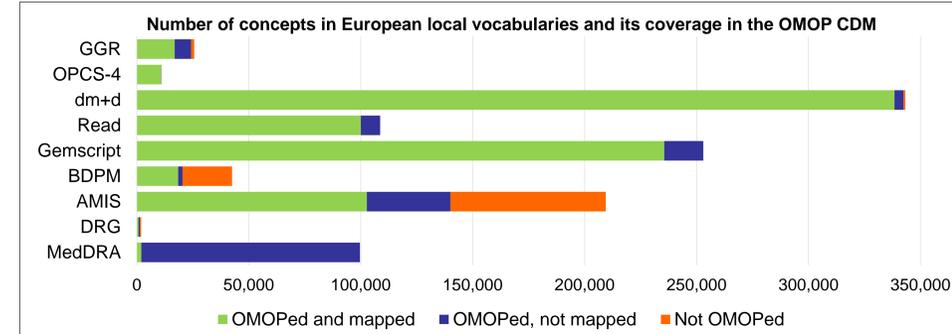
■ Do not exist in any ICD vocabulary  
■ Match to the US ICD-10-CM (with semantic variance)  
■ Equivalent to the original ICD10-WHO

## Results

### 3. Coverage of local vocabularies.

Due to utilization in existing European data, some vocabularies are covered by OMOP:

- Fully integrated with automatic refreshes and mapping - GGR (Belgium), OPCS-4 (multi-country), dm+d, Read and Gemsript (UK);
- Partial integration – BDPM (France), AMIS (Germany), DRG (multi-country) and MedDRA (multi-country).

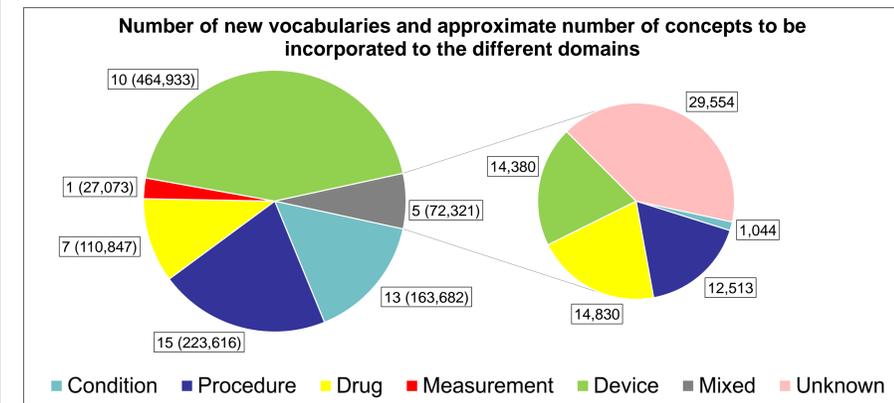


### 4. Heterogeneity in coverage among domains.

- **Measurement:** most covered, mainly because of LOINC wide usage in Europe.
- **Drug:** on a high level we can easily deal with any country drugs encoded by ATC, but for a Drug Product level we still need to OMOP a lot of national drug vocabularies.
- **Condition:** would be covered enough after the several international vocabularies (ICF, ICF-CY, TNM8) and ICD10 national extensions are processed.
- **Procedure:** the most heterogeneous group and still need to be mapped. The new mapping algorithm based on Procedure attributes might be applied to the vocabularies listed above.
- **Device:** still in a grey zone.

### 5. New challenges.

- An implementation of ICD-11 is a great challenge for the OMOP CDM, but is likely to greatly help the standardisation process in Europe.
- Not all the mentioned vocabularies (especially those with limited national implementation) need to be urgently incorporated into the OMOP CDM.
- Vocabulary-specific features that need to be taken into account. For example, the ICD-10 Dual Classification (Dagger and Asterisk) system allows recording of certain diagnostic subtleties like the combination of two codes: a primary and mandatory code for the underlying disease, marked with dagger sign (+), and an optional additional code for the manifestation in a particular organ or site, marked with an asterisk (\*). That is why vocabulary mapping of asterisk codes may be a challenge requiring a synchronous usage of both codes from combination.



## Conclusions

1. OHDSI dissemination to the European countries is a continuous process that will intensify after the EDHEN project has launched.
2. Patient data transformation to the OMOP CDM starts with medical vocabularies being mapped to standard concepts, which requires serious efforts and resources.
3. Although a lot of work has already been done, the medical vocabularies used in Europe for healthcare coding purposes are not covered completely by the OMOP CDM.
4. Heterogeneity of national systems and existence of local editions of international vocabularies (ICD, SNOMED, DRG) will require additional attention.
5. The current incomplete coverage of vocabularies by OMOP CDM, the lack of precise information or vocabulary source files result in a great bulk of work need to be done.
6. Having a large variety of methods, tools and practical experience, the OHDSI Vocabulary Team is prepared to take on the onset of European vocabulary incorporation and will need to prepare a comprehensive sequential development plan.

### References:

1. <http://www.edhen.eu/>
2. <https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/imi2-2017-12-04>
3. <https://www.who.int/classifications/en/>
4. <http://apps.who.int/gho/data/node.whofic>
5. <https://ec.europa.eu/cefdigital/wiki/pages/viewpage.action?pageId=52603401>
6. <https://pharm-infoclinks.webnode.cz/drug-databases/national-drug-databases-eu/>