Results

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

<table>
<thead>
<tr>
<th>Country</th>
<th>Condition</th>
<th>Procedure</th>
<th>Drug</th>
<th>Measurement</th>
<th>Device</th>
<th>Other/mixed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Belgium</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Czech Republic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Denmark</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estonia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>France</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ireland</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Netherlands</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poland</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Portugal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spain</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweden</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Switzerland</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Austria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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.

Conclusions

1. OHDSI dissemination to the European countries is a continuous process that will intensify after the EHDEN project has launched.
2. Patient data transformation to the OMOP CDM starts with medical vocabularies being mapped to standard terminology that requires various 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 necessity of European vocabulary incorporation and will need to prepare a comprehensive sequential development plan.

References:
http://www.google.de/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=0ahUKEwiosMm948tjAhVJrPQIHX74A3cQ_AjIBCA

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

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

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

1. Lack of available structural 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 (ICD10-CM, ICD10-AM, ICD-10, AT, LTCN, SNOMED). An international version of a vocabulary is implemented in all national versions but may be processed automatically storing the concept name interpretations as concept synonyms.
   - Another variant of "national modifications or extensions" are the original vocabularies as a basis and the application of a number of adjustments and additions (ICD-10 and SNOMED). E.g., ICD10-AM, the German modification of ICD10-WHO is currently being revised annually and used for coding of diagnoses in the outpatient and inpatient care (with about 1,670 concepts). 1,335 (81.6%) are equivalent to the original ICD10-WHO codes, 1,135 (7.0%) match to the US ICD-10-CM vocabulary (albeit with semantic variations), while 1,518 (9.4%) do not exist in any ICD system. The processing of this type of vocabularies generally needs some additional efforts, including reviewing, merging, etc. Anyways this is much more than development of obsolete one. It is noteworthy that an enriched translation of ICD-10-WHO is systematically 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, as it may be used in the future mapping (Figure 2).

Figure 3. Number of new vocabularies and approximate number of concepts to be incorporated to the different domains

Results

3. Coverage of local vocabularies

Due to the absence of centrally available European data, some vocabularies are covered by OMOP:

- Fully integrated with automatic refreshes and mapping - GRR (Belgium), OPCS-4 (multi-country), dm-d, Read and Gemvax (UK).
- Partial integration – BDPM (France), AMIS (Germany), DRG (multi-county) and MeDRA (multi-country).

4. Heterogeneity in coverage among domains.

- Measurement: most covered, mainly because of LOINC wide usage in Europe.
- Drug: on a high level 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 (IC, IF-CC, TINAB) 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.

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 syndromes 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 of a particular organ or site, marked with an asterisk (*). That is why vocabulary mapping of austrian codes may be a challenge requiring a synchronous usage of both codes from combination.

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 analyze using standardized and systematic approaches. This is the result of different languages, healthcare constellations, EHR systems and policies [2]. The Electronic Health Data in a European Health (EHDEN) consortium is aimed at addressing the problems of data fragmentation. It will provide a harmonized model to address the structural heterogeneity and the use of different coding standards [2].

Since the standardization of health data to the OMOP Common Data Model and the adoption of analytical tools developed by OHDSI is central to EHDEN, 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 24 EU countries
• The World Health Organisation’s international classifications page and implementation database [3, 4]
• Other open sources [5, 6]

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.