EMIF Deliverable 11.5: Second version of the harmonised information model and associated terminology mapping

Executive summary

Executive Summary

The activities reported in Deliverable 11.5 build on the trajectories of semantic harmonisation work documented the year before.

For **population data sets**, the decision was taken two years ago that the preferred common data model should be the OHDSI model version 5, which adopts selected terminology systems for different categories of health data. Over the past year mappings have been developed from several EMIF data sources to the corresponding OHDSI terminology systems. Deliverable 11.4 presented this common data model and D11.5 presents a brief update on the progress made: 4 data sources have now been completely mapped and another 5 are in progress. The challenges that have needed to be overcome have also been collated and shared with the OHDSI community at the most recent OHDSI Symposium (Washington, 2016).

A parallel effort of developing terminology mappings for **cohort data sets** was also reported last year, with a focus on concepts and clinical events relevant to EMIF Use Cases 10, 11 and 12. This has now been extended to Use Cases 9 and 13, and mapped from the terminologies used in 6 data sources. Deliverable 11.5 lists the mapped clinical events.

Work is reported on EMIF Knowledge Objects and on the **tooling to support semantic harmonisation** of the variables of interest for Alzheimer's Disease (AD). This especially focuses on transforming the data for locally-held variables at each AD data source, to the global variables of interest to researchers. Apart from expanding the library of actual global knowledge objects and their cross mappings to local knowledge objects, D11.5 introduces the concept of a "Switch-box" which holds the harmonisation rules and mappings to the global knowledge object store (represented using Stardog). This aims to this would reduce the mapping effort down to one ETL for the cohorts as a consequence. The pipeline from individual data sources through this harmonisation process to the EMIF Participant Selection Tool (PST) and Variable Selection Tool (VST) has been established this year, in addition to the use of knowledge objects to drive ETL into tranSMART.

Contacts

EMIF-Platform: Johan van der Lei – Nigel Hughes
j.vanderlei@erasmusmc.nl - nhughes@its.jnj.com