EMIF Deliverable 11.6: Final Report on Harmonisation and Semantics

Executive summary

Executive Summary

During EMIF, the partners involved in Work Package 11 have made important inroads into the challenges of harmonising the structure and semantics across different large and complex data sets. The approaches taken, and the results of working with both population and cohort data sets – and the lessons learned along the way – have been reported in previous WP11 deliverables. This deliverable, D11.6, is an update on recent work, and needs to be read in conjuction with D11.5, and ideally the earlier WP11 deliverables, since that background work is not repeated here.

Section 1 reports on the end of project status of mapping EMIF data sets to the OMOP single common data model. As reported previously, OMOP was chosen after careful consideration to be the most suitable target model for mapping the structure and semantics of diverse data sets. By adopting this approach, it has also been possible to leverage and contribute to the OHDSI tools development and community. Multiple EMIF data sets have now been mapped to OMOP, and to its in-built terminology bindings for conditions, drug exposure, measurements and observations (primarily to SNOMED CT and RxNORM). Not all of the data sets have a complete mapping yet.

For Alzheimer's Disease cohorts previous WP11 deliverables have reported on the design and definition of Knowledge Objects, which are a semantic rather than structural representation of research data concepts for populating the TranSMART research analysis tool. The semantics of selected EMIF AD cohorts has been mapped to Knowledge Objects. Section 2 summarises the latest developments of this thread of work, the Switchbox, that contains the end to end semantic harmonisation process. This Section presents the data model of the Switchbox. This will enable the Knowledge Objects and their cohort mappings to be more easily reused by other research groups working in Alzheimer's Disease research, after EMIF.

Several EMIF Use Cases have been defined by other WPs, to be run using EMIF cohort data. Previous WP11 deliverables have reported the semantic harmonisation work done to meet these requirements, expecially to establish the occurrence of clinical events that have to be inferred from multiple data points. Section 3 and Annex 1 provide an update on that work. The CodeMapper tool has been developed by the IMI ADVANCE project. With permission, this has been used to develop and hold these mappings. It will continue to be used beyond EMIF, by the partners who developed it, for future use cases.

For the population data sources it is expected that the new IMI project, the European Health and Evidence Data Network (EHDEN) , due to start in late 2018, will take up and build on the work of this EMIF WP, with some of the same partners.

Contacts

EMIF-PLAT: Dipak Kalra - dipak.kalra@eurorec.org