EMIF Deliverable 11.3: Extended Specification of the Framework of Reference

Executive summary

Executive Summary

This deliverable reports on advances made in harmonising the semantics of heterogeneous data sources when they are combined or analysed together to address a research question.

One challenge that has been identified as critical is the identification of target subpopulations within a cohort study or population health database. Most commonly, subpopulations are initially defined in terms of inclusion or exclusion conditions, and there are many situations in which such conditions are not explicitly and definitively coded within patient level data, or at least are not expressed at a sufficient level of detail for the research need. These conditions may therefore have to be inferred from a set of more precise inclusion and exclusion criteria, or through a triangulation of indicators such as a specific treatment that implies that the condition is present. This deliverable reports on work done to define a human and organisational workflow, and accompanying informatics tools, that can capture explicitly the definition of a clinical condition, using the worked example of 1)Type II diabetes mellitus, 2) Dementia and 3) non-alcoholic fatty liver disease (NAFLD) and non-alcoholic steatohepatitis (NASH), and the terminological tools that can facilitate mapping diverse representations of the condition or of its indicators. The tools have been implemented, drawing substantially on the cross terminology mappings within UMLS. This work has been undertaken jointly with WP12. Further work is now underway examining the applicability of the same workflow and tools to specialised diagnoses of dementia.

Another important area for this work package is the proposal of common data models. The business modelling work package has now identified a set of service levels which correspond to different levels of platform query functionality, with consequently different levels of detail needed in the underlying data. WP11 has re-examined the value of OMOP, which was initially summarised in D11.1 along with other research data models, to consider its role in underpinning EMIF dashboarding and profiling tools. A review of OMOP is given here, but the more detailed work of examining its suitability for these query tools will be undertaken in the year ahead, in particular how well it can support the structures and semantics required for the profiling tools.

Deliverable 11.2 introduced the concept of EMIF Knowledge Objects. This deliverable builds on that initial concept, through the pilot development of a partial ontology and Knowledge Object representation for dementia concepts. A clearer conceptual framework has been developed for (i) Knowledge Objects that are faithful to each underlying data source, to provide a consistent but not necessarily harmonised view of their data models, (ii) harmonised Knowledge Objects that are used to map heterogeneous faithful Knowledge Objects to standardised concepts and representations that are closer to the concepts that researchers wish to use within query specifications, and (iii) derived concepts that implement rules (inclusion criteria, data element combination or transformation rules), such as those referred to earlier in

This work has received support from the EU/EFPIA Innovative Medicines Initiative Joint Undertaking (EMIF grant n° 115372).

this report for type two diabetes. Examples of these three types have been implemented for dementia, at this stage as a proof of concept, using Protégé and the ONTOP Platform. This deliverable reports the progress made, but does not yet arrive at a definitive recommended structure for Knowledge Objects. This work is continuing, and will be reported more comprehensively next year.

Contacts

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

This work has received support from the EU/EFPIA Innovative Medicines Initiative Joint Undertaking (EMIF grant n° 115372).