Deliverable D4.3 Central tranSMART instance deployed as data warehouse for translational data including omics data from cohort studies – publishable summary

The aim of this deliverable is to make an omics analysis platform available that enables secure storage of omics data and that enables WP5 (data analytics) to perform translational research, meaning analysis of the omics data in combination with real-world data from observational population-based registries and clinical data from clinical trials and cohort studies. The second aim was to train WP5 (data analytics)on how to use the platform for cohort construction and data analysis.

Deliverable D4.4 Central OHDSI software deployed as data exploration and analysis platform for real world data and central OMOP v5 dictionaries set up – publishable summary

To optimally benefit from real-world population health data from multiple sources, a common data model (CDM) is required. Observational databases differ in both purpose and design. Electronic Medical Records (EMR) are aimed at supporting clinical practice at the point of care, while administrative claims data are built for the insurance reimbursement processes. Each has been collected for a different purpose, resulting in different logical organizations and physical formats, and the terminologies used to describe the medicinal products and clinical conditions vary from source to source. The CDM can accommodate both administrative claims and EHR, allowing users to generate evidence from a wide variety of sources. It also supports collaborative research across data sources and across countries, in addition to being manageable for data owners and useful for data users.

In the PIONEER project we decided to use the Observational Medical Outcomes Partnership (OMOP) CDM which allows us to collaborate with the Observational Health Data Sciences and Informatics (OHDSI) program. OHDSI is a multi-stakeholder, interdisciplinary collaborative to bring out the value of health data through large-scale analytics. OHDSI has established a fast-growing international network of researchers and observational health databases and is developing open source data analytics tools.

One of the strengths of the OMOP-OHDSI platform is that it supports federated data analysis. In this case an analysis script can be run on multiple local OMOP databases, generating stronger scientific evidence than would be possible on a single OMOP database.

In the PIONEER project federated installations of OHDSI-OMOP will be realised for data that may not leave the data providers premises while central installation of OHDSI-OMOP is realised for data that may leave the sources server or repository.

Deliverable D4.5 Mapping strategy for omics data and mapping tools deployed – publishable summary

The aim of this deliverable is to enable sharing and reuse of omics data by proposing the use of standards. Several organizations have expressed interest in sharing omics data with the PIONEER consortium. Omics data is highly dimensional, because it is the output of high-throughput biochemical

PIONEER is funded through the IMI2 Joint Undertaking and is listed under grant agreement No. 777492. This joint undertaking receives support from the European Union’s Horizon 2020 research and innovation programme and EFPIA.

assays, with almost all or a huge selection of molecular instances measured in the target space1. The three largest omics domains are genomics, transcriptomics and proteomics, in which respectively the genome, transcriptome and proteome being measured. In the current data platform setup data contributors can upload omics data to the central server via the Jupyter platform. There it can be integrated with the clinical data available in the OMOP CDM database. With JupyterLab, the combined data can then be analyzed accordingly. This works to a certain extent, but to fully enable data sharing and data reuse by different stakeholders (think FAIR -Findable, Accessible, Interoperable, Reusable2) in the long run, a common data format and appropriate metadata could be of great benefit. Using standard data formats also ensures that the current platform is not overloaded with huge data files in exotic file formats.