

## Deliverable D4.1 Detailed architecture for analytics platform that includes centralised and federated solution – publishable summary

PIONEER WP4 will develop a pan-European data-sharing platform and will adopt a two-pronged approach to address the project needs: a) a platform that can access population-based registry data, and b) a platform that can handle rich clinical and omics for translational analysis by WP5. To achieve this, the project will build upon and use approaches developed in a number of other IMI projects, such as IMI EMIF and IMI eTRIKS.

To cater for data integration and analysis for longitudinal PCa registries, we will use the OMOP and OHDSI technology, while for cohort studies that besides from deep phenotypic and clinical data also include omics data, we will use the transSMART technology. The same combination of transSMART and OHDSI has been used successfully in IMI EMIF.

All data will be harmonised (transSMART) or standardised (OHDSI-OMOP) before being loaded in the platform of choice. The open-source/available tools of the IMI eTRIKS and IMI EMIF projects will be used for harmonisation of data that are loaded in transSMART.

We have envisioned distinct possibilities depending on the nature of the data (centralised vs. decentralised/federated). In particular, central installation of transSMART and OMOP-ATLAS will be chosen for data that may leave the source's server or repository, while federated installations of OMOPATLAS will be chosen for data that may not leave the data provider's premises.

WP4 will also use the EMIF catalogue to facilitate centralised storage, management and sharing of metadata of available PCa data sets. It will provide a list of all the data sources registered by PIONEER, through a portal and search tools, to enable potential data users to discover data sources that are most relevant to their research needs, according to a variety of data source and dataset descriptors, and to support the access request process.

## Deliverable D4.2 Central open access catalogue deployed and customised according to consortium requirements – publishable summary

A centralised catalogue is provided that allows researchers to find PCa databases which fulfil their particular research study requirements in a quick and easy way. The catalogue provides relevant database descriptions using established metadata standards and supports requests for access to the databases.

We used the existing EMIF Catalogue which ensures the catalogues long-term sustainability and potential compatibility with other resources. In the EMIF Catalogue a separate PIONEER space (or community) is created for which permissions and access rights are managed by WP3.

The catalogue can be accessed via <https://emif-catalogue.eu/>. After creating an account via the log-in page a user will be able to access the PIONEER community pending approval by the work package 3 Community Managers.