

Background

Current clinical research is mostly hindered, on one hand, by limited and fragmented access to heterogeneous health data repositories and, on the other hand, by privacy issues due to legal, ethical and regulatory requirements. For the success of clinical translational research, it is imperative to develop solutions for distributed data homogenization and querying, without losing data and patients' privacy.

The **European Medical Information Framework (EMIF)**¹ is one of the most recent European projects, designed to meet the following goals:

- Integrate health data from disperse and heterogeneous EHR systems and cohorts;
- Facilitate the exploration and reuse of the integrated data, while preserving data privacy.

The **EMIF Catalogue**² allows researchers to browse information of biomedical databases at three conceptual levels: metadata (general characterization of databases), aggregated data and ultimately individual patient data in a remote controlled environment. It integrates **ATLAS**³ as a plugin that allows researchers to conduct scientific analyses on biomedical databases that have been converted to the OMOP Common Data Model (**CDM**). This interrelation led to the development of a steady methodology that allows clinical researchers not only to discover otherwise disperse and heterogeneous biomedical databases, as well as to query and provide them a secondary use, while keeping health data private in each healthcare institution.

Methods

Our methodology makes use of the EMIF Catalogue as a main entry and management solution, and assumes the use of the OMOP CDM. ATLAS, the open-source analytic tool developed by OHDSI, provides statistical and aggregated information on clinical digital data converted to CDM and it is an integrative part of this pipeline.

Three main actors are involved in this process:

- **Researcher (R)**, the person who wants to query one or several patient-level databases;
- **Data Custodian (DC)**, the person responsible for managing a database;
- **Study Manager (SM)**, the person who leads and manages the research study; additionally, the Study Manager moderates the tasks between the Researcher and the Data Custodian.

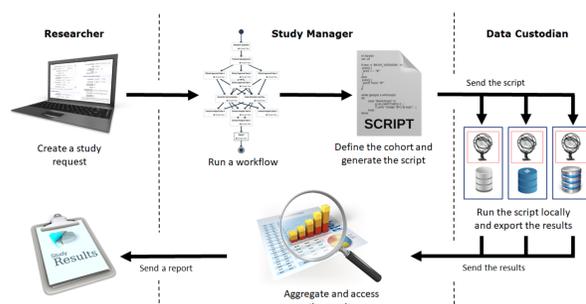
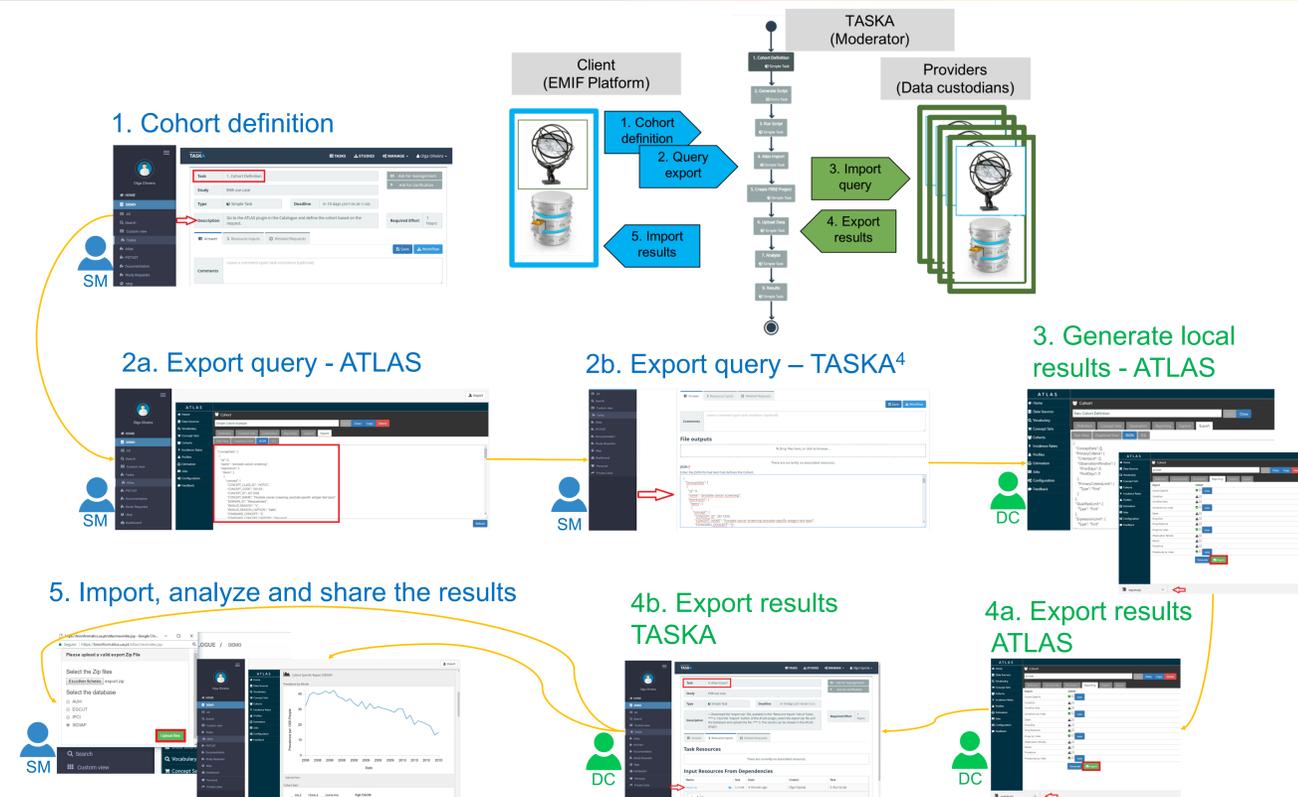


Figure 1: Workflow of the querying process.

Results



Conclusions

We developed a methodology to perform semi-automatic queries over distributed EHR databases. Our approach does not rely on any centralized data warehouse, but rather through a distributed model it assures that data custodians maintain control of their database and only share the data they consider to fulfill their legal, ethical and regulatory requirements.

The EMIF Catalogue together with open-source OMOP tools enables researchers to discover, query and reuse otherwise disperse and heterogeneous biomedical databases.

References:

- A. Trifan, J. van der Lei, C. Diaz, J. L. Oliveira. A methodology for fine-grained access control in exposing biomedical data. *Proceedings of Medical Informatics Europe (MIE18)*, Gothenburg, Sweden, April 2018 (in press).
- O. Fajarda, L. Silva, P. Rijnbeek, M. Van Speybroeck, J. L. Oliveira. A methodology to perform semi-automatic distributed EHR database queries. *Proceedings of the 11th International Conference on Health Informatics (HealthInf 2018)*, Funchal, Madeira, Portugal, January 2018.
- J. Almeida, R. Ribeiro, J. L. Oliveira. A modular workflow management framework. *Proceedings of the 11th International Conference on Health Informatics (HealthInf 2018)*, Funchal, Madeira, Portugal, January 2018.

¹ <http://www.emif.eu>
² <https://emif-catalogue.eu>
³ <http://www.ohdsi.org/web/atlas/#/home>
⁴ <https://emif-catalogue.eu/taska/>