Project ATHENA – consortium confidential

Privacy Preserving Analytics through Federated Data Networks

Michel Van Speybroeck | Janssen Pharmaceutica N.V.

November 16, 2023
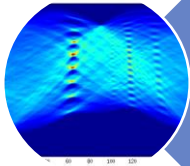
The ingredients of success

Analysis Methods

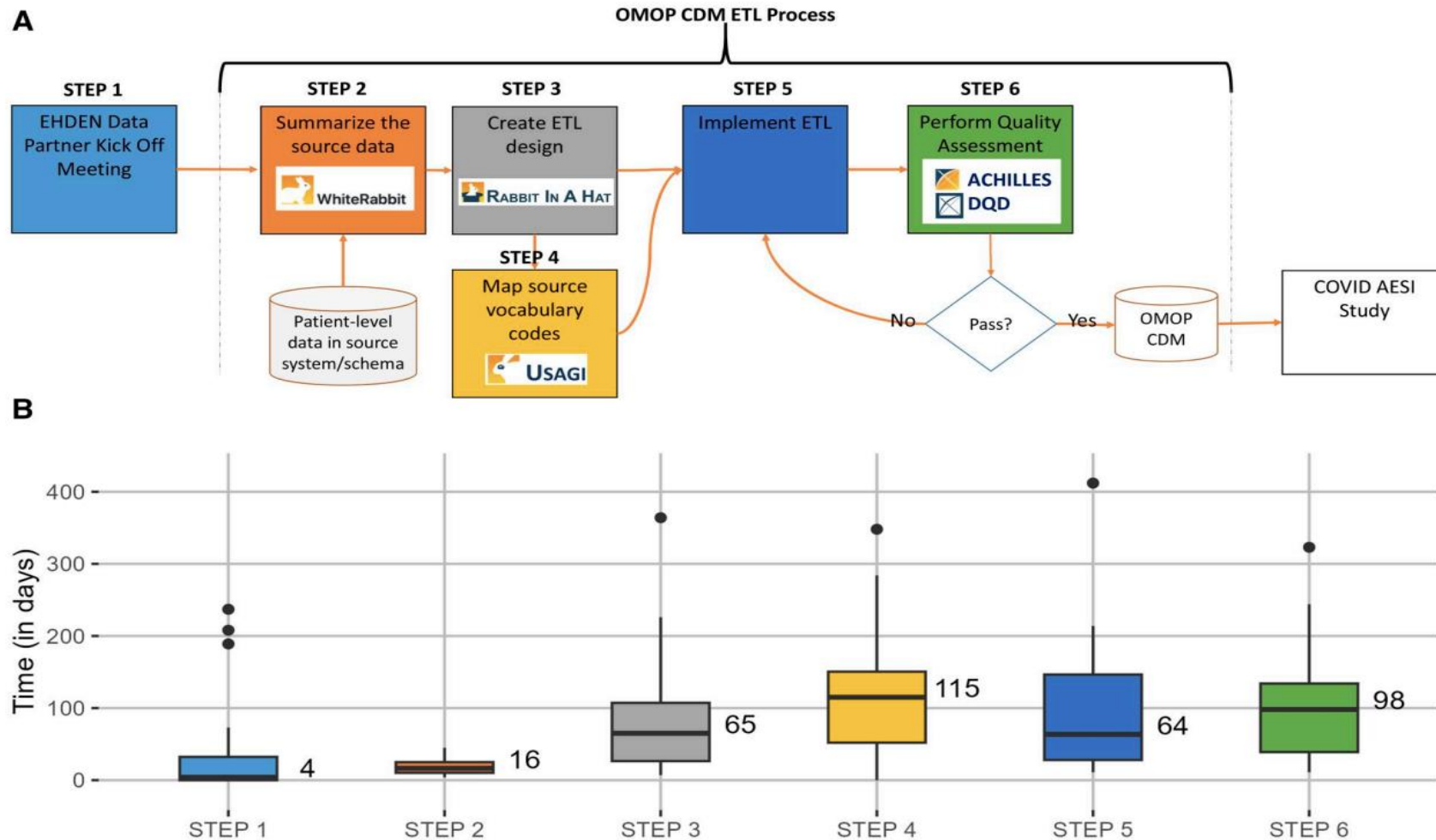Understand the Data and Map them correctly

Good and Complete Data

Trust and Willingness

The standard process for data harmonisation

Wrong assumptions

Technology as a substitute of trust

It's possible to get to good data with the right technology

Data in a database is of better quality than in an Excel spreadsheet

It's possible to make this work provided we have enough meetings about it.

Despite following the rules of the game....

Throughput times for mapping between 6 months and 2 years

Limited overlap in OMOP variables between centers in UC

Many iterations for complex variables

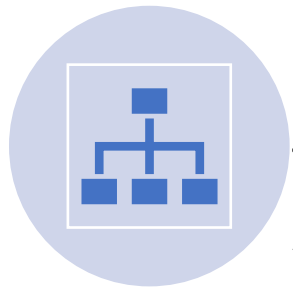Difficulty to incorporate other data modalities e.g. genomics

Mapping Status
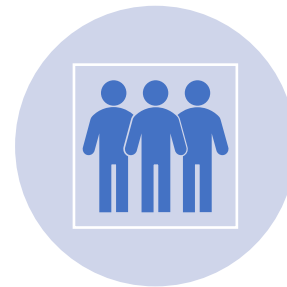
What works well…

Due diligence: complete data with the necessary attributes (incl. dates !)

Participation of people with knowledge of the data is critical

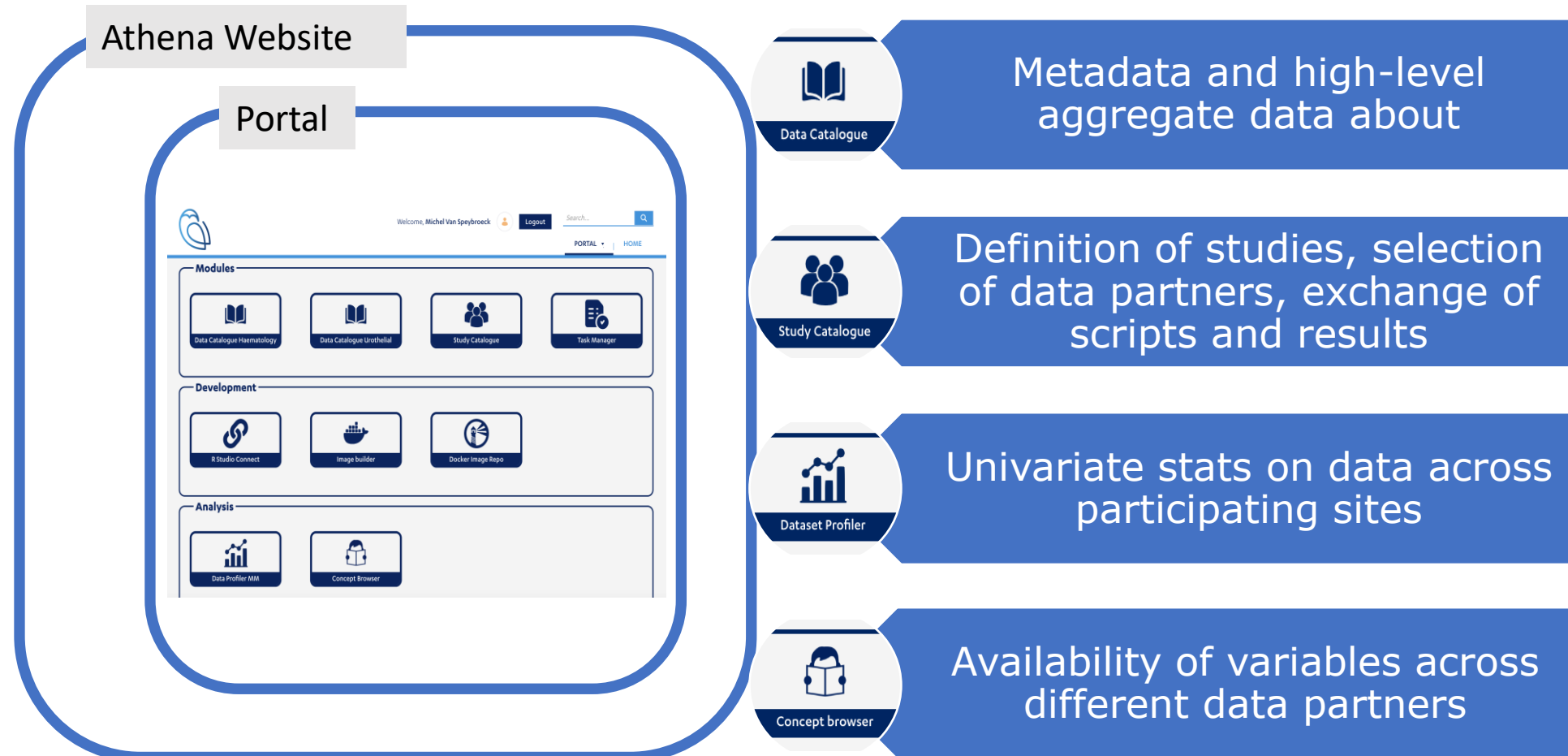COORDINATED harmonisation – with proper tooling / checks / processes
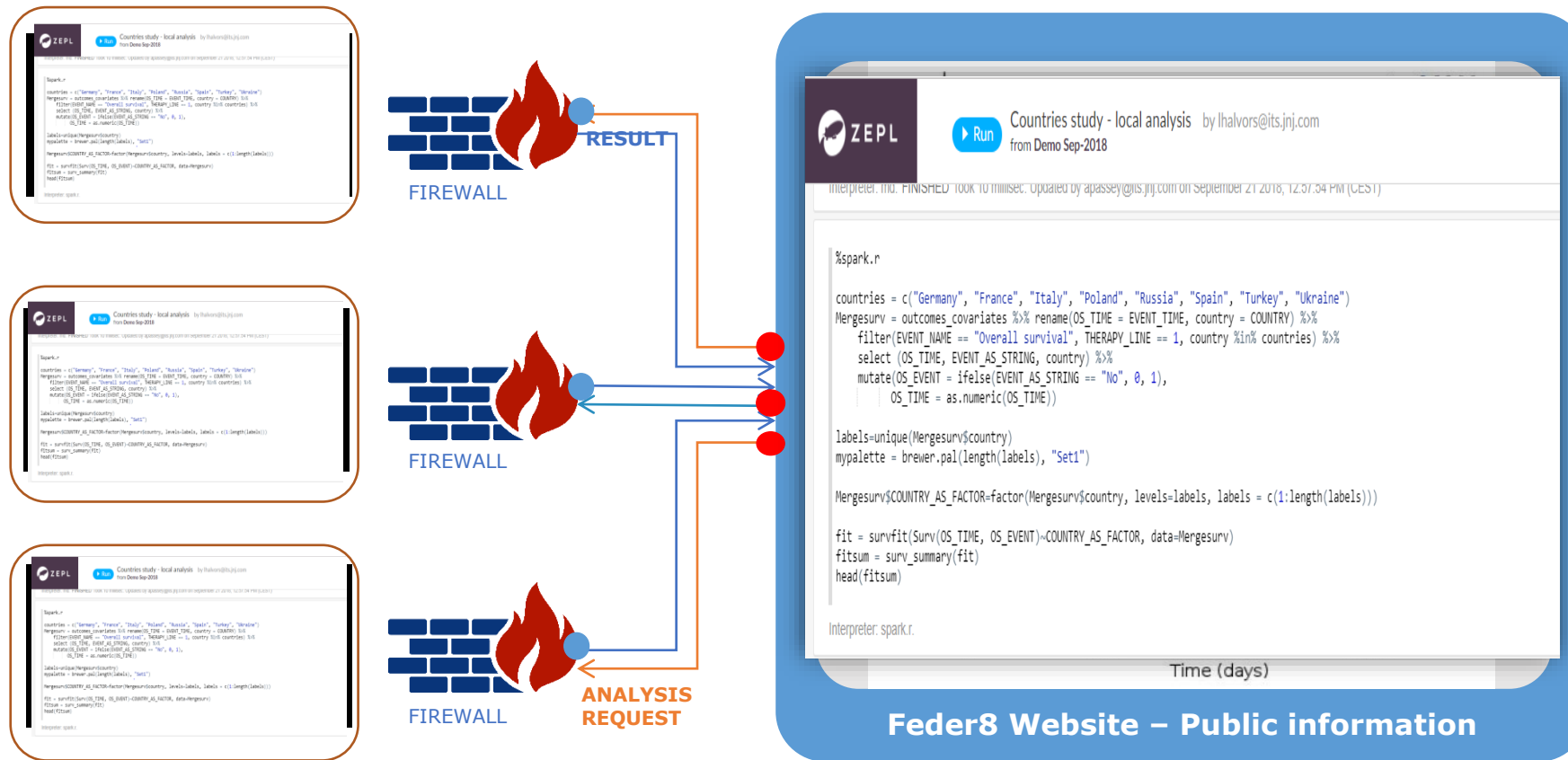
A cooperative IT organisation

Feder8

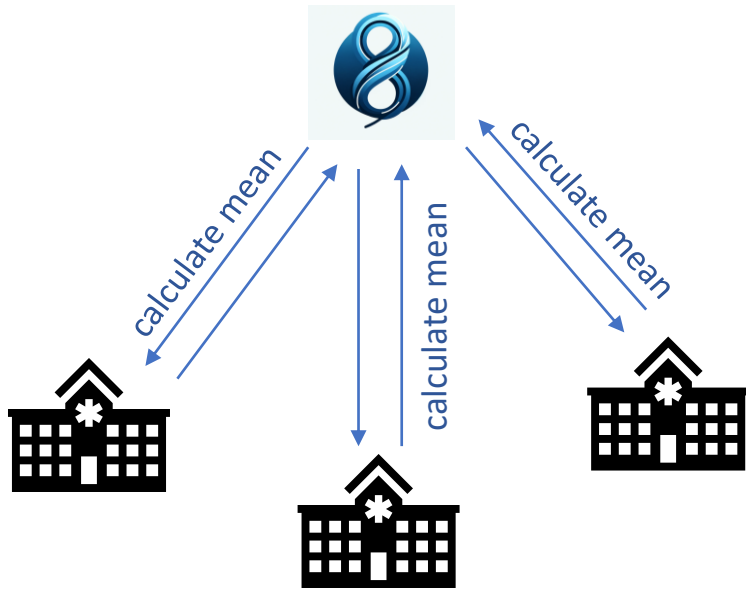A data management and analysis platform for federated data networks
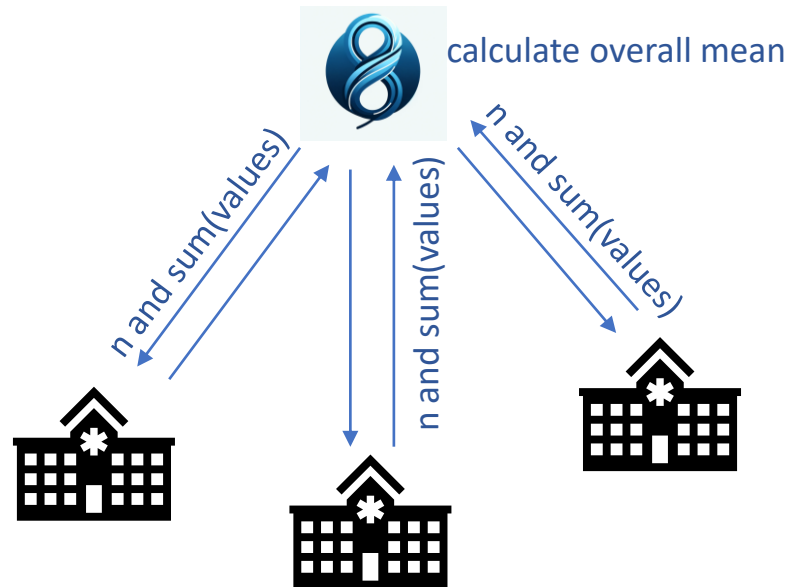
Central Infrastructure

Athena Website

Portal

Data Catalogue — Metadata and high-level aggregate data about

Study Catalogue — Definition of studies, selection of data partners, exchange of scripts and results

Dataset Profiler — Univariate stats on data across participating sites

Concept browser — Availability of variables across different data partners

The basic mechanism

Analysis Modalities in Feder8
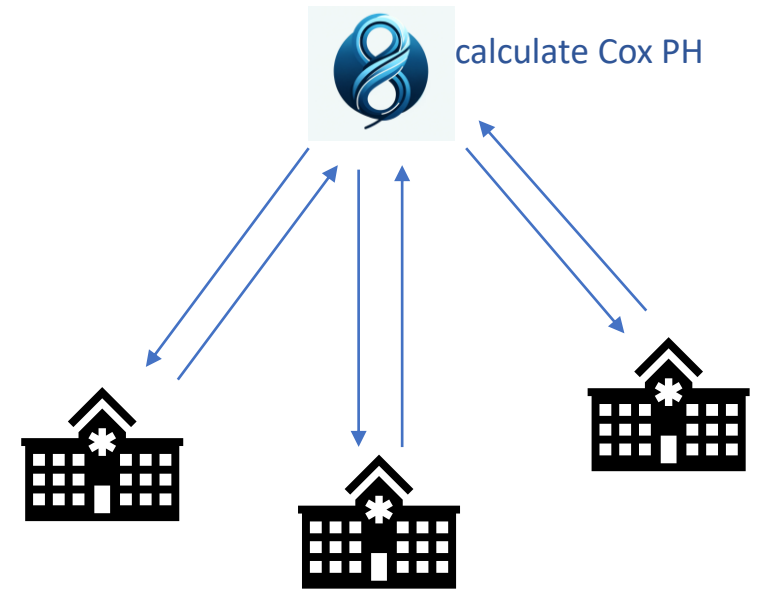
Self Service Analytics
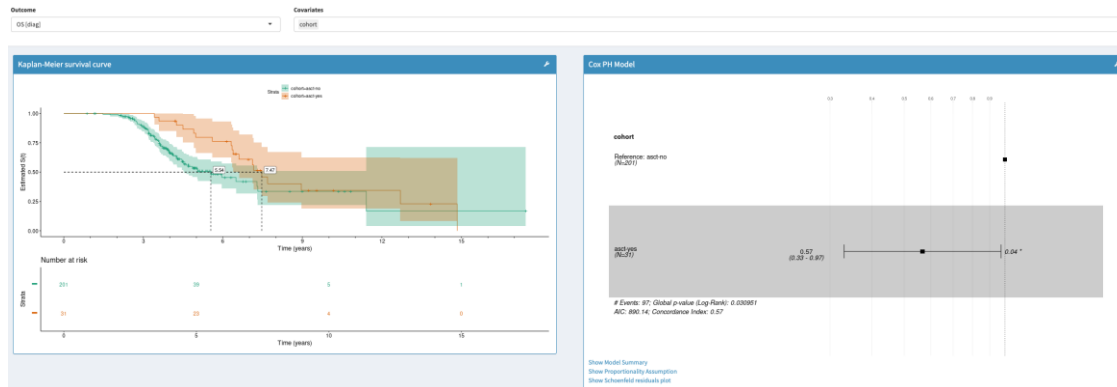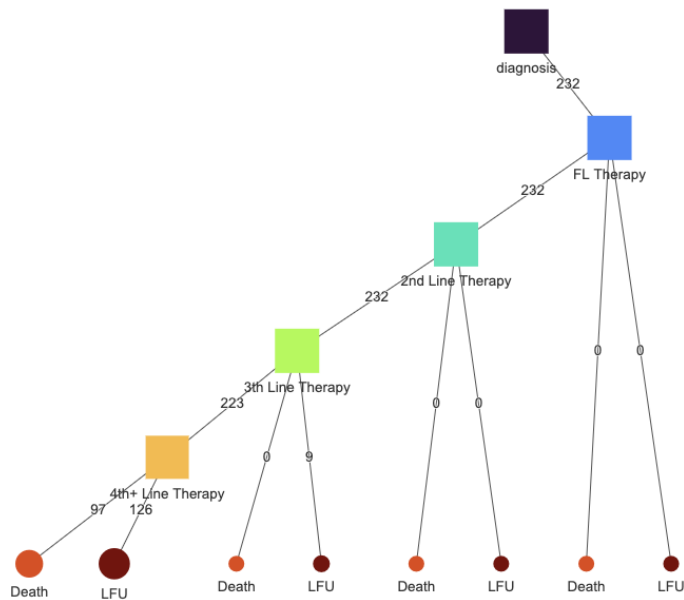
DiTrAn is a **disease-agnostic** framework for analysis and visualization of **clinical event models** (*disease trajectory*) based on OMOP data.

On basis of a custom clinical event model
OMOP data are re-mapped to this event model
Descriptive stats, treatment analysis, time-to-event analysis
Flexible cohort definitions
Export of results (pdf, html, csv)

Conclusion

- We have the ability to run distributed / federated analysis – statistical analysis and ML models

- Remaining work to be done on the validation of the OMOP'ed data

- Significant improvement opportunity exists to improve the process for getting ready to participate in federated data network programs.

- The challenge is not in the sophistication of (ML) models – it's in getting the basics right.

- We learned a lot about the data and the intrinsic heterogeneity in what gets (not) collected