Data harmonization and data use in a federated rare cancer network

EURACAN

Dr. Maaike van Swieten

Clinical Data Scientist (IKNL)

Data & Tech WP lead Blueberry

IDEA4RC

BlueBerry
a blueprint for the EURACAN registry

KWF

Funded by the EU

Outline

- EURACAN introduction

- Data harmonization strategy

- Steps to create a federated network

- Challenges

- Lessons learned

- Sneak Peek

Background

EURACAN (European Reference Network for Rare Adult Solid Cancers) is a **collaborative initiative** aimed at improving the diagnosis, treatment, and research of **rare adult solid cancers** across Europe.

- Low incidence (less than 6 per 100 000 people)

- 25% of all cancer diagnoses

- Diverse clinical manifestations

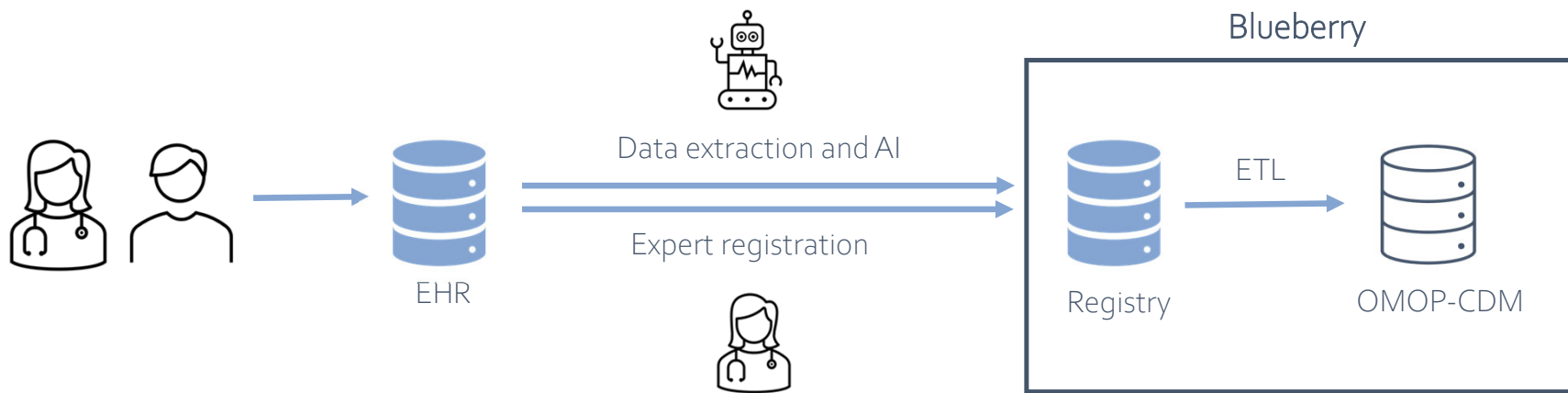》》 Lack of data interoperability

Two EURACAN projects

Blueberry

- Blueprint for a sustainable, effective, scalable EURACAN registry
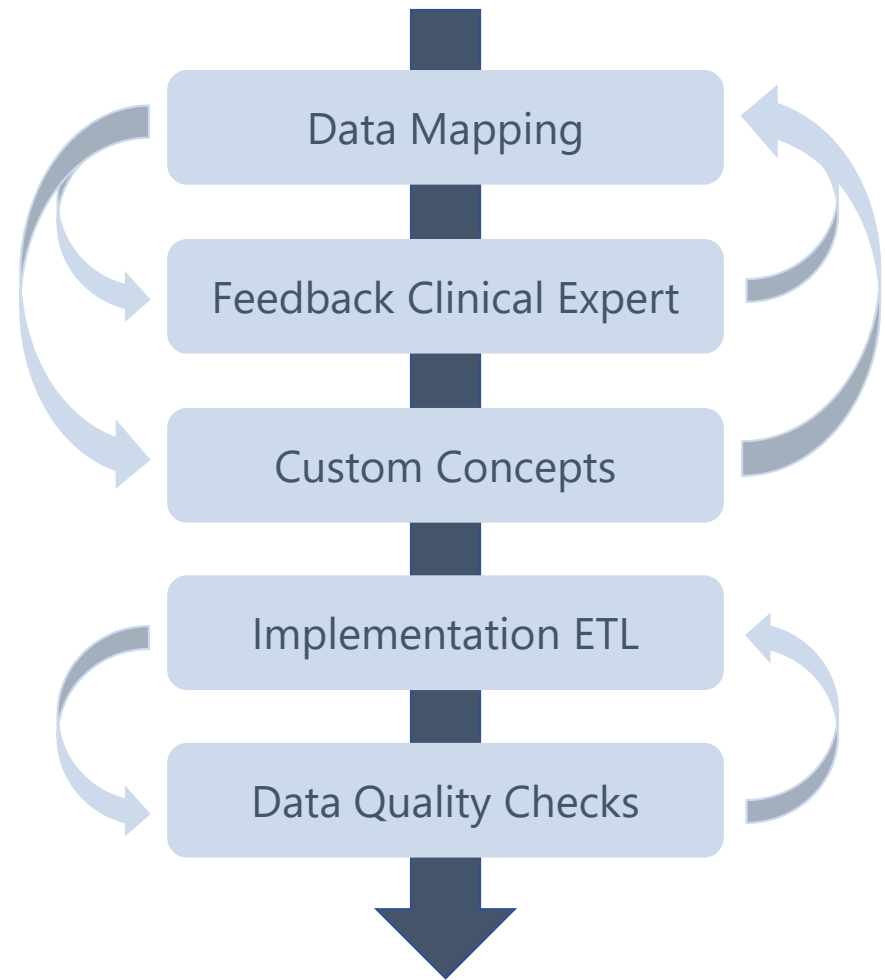
- 7 data partners

- OMOP-CDM

IDEA4RC

- An intelligent data ecosystem for rare cancers

- 25 partners from 12 European countries

- FHIR and OMOP-CDM
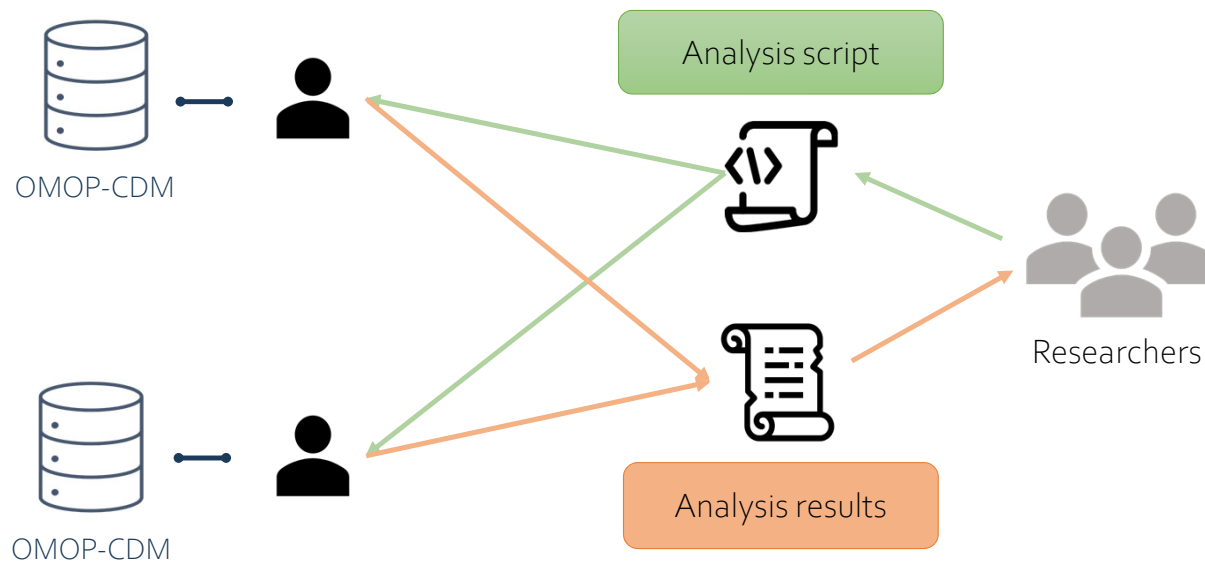
From health data to registry data

OMOP-CDM: Observation Medical Outcomes Partnership – Common Data Model

Data Harmonization Strategy

Data Mapping

Feedback Clinical Expert

Custom Concepts

Implementation ETL

Data Quality Checks

From registry data to federated learning network – step 1

OMOP-CDM

Analysis script
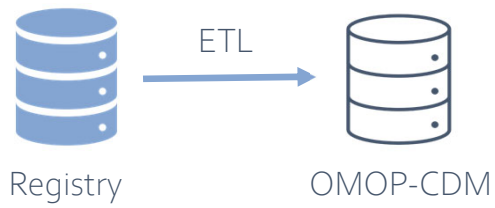
Researchers

Analysis results

OMOP-CDM

Each data partner runs the same script locally and sends back the results

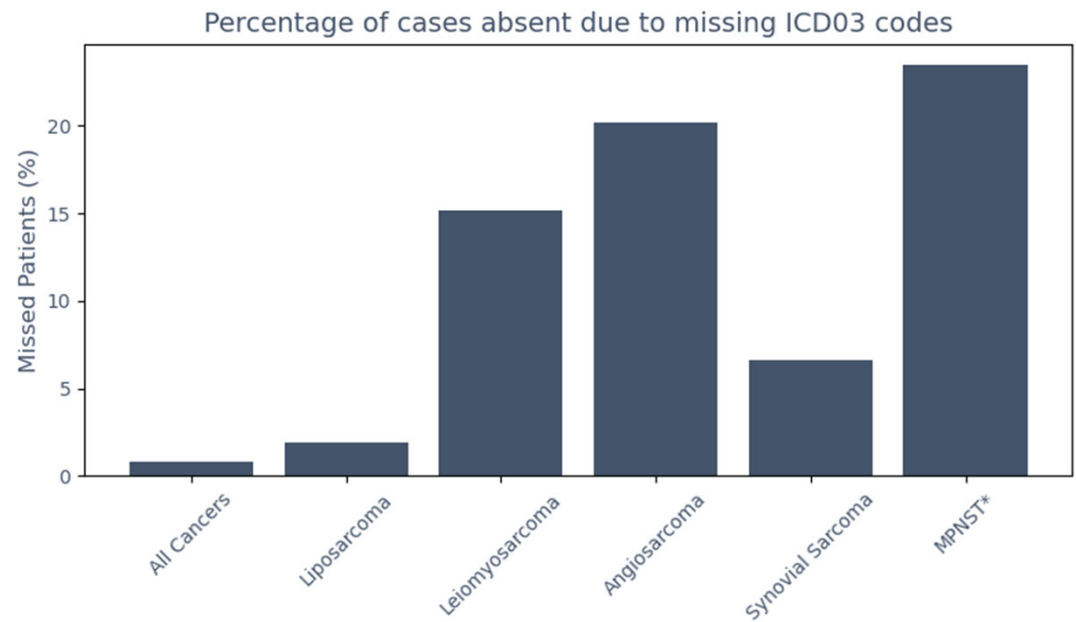# From registry data to federated learning network – step 2
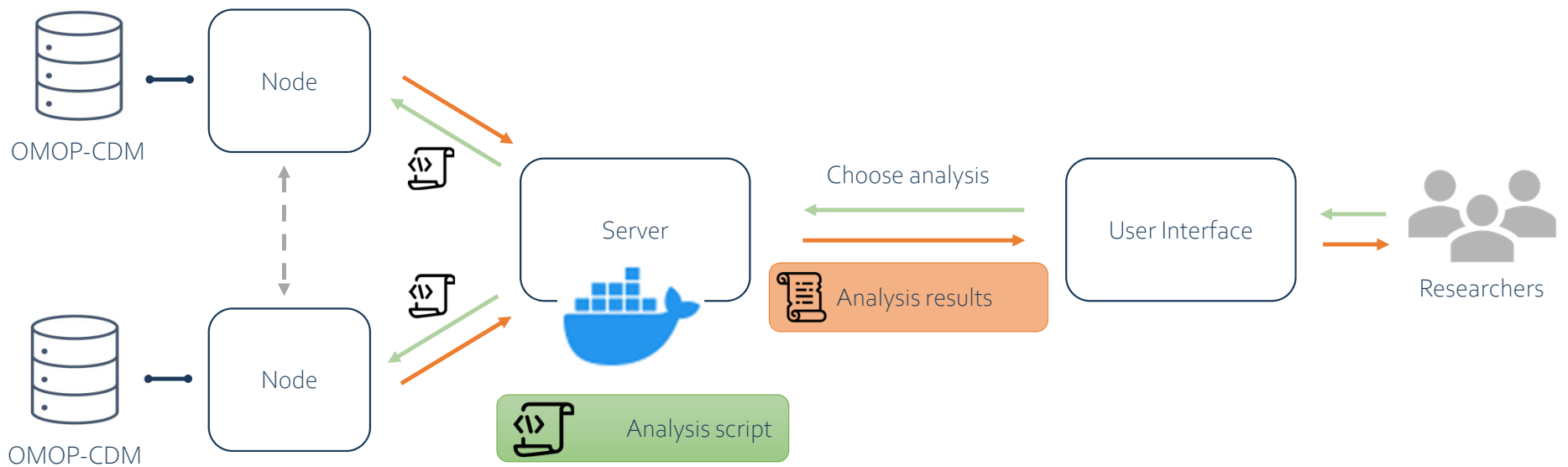
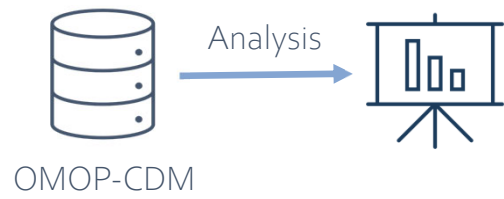'Grouped' analysis of different data sources as if data is stored locally

VANTAGE

https://distributedlearning.ai

Challenges

Registry → ETL → OMOP-CDM

- Missing diagnosis (ICDO3) codes

Percentage of cases absent due to missing ICD03 codes
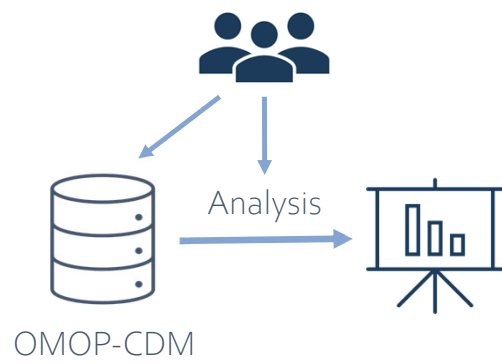
Challenges

OMOP-CDM

- Knowledge required about OMOP data format
- Cross-links in data

Challenges

OMOP-CDM

Analysis

- Balancing the wishes of clinicians with what is possible in OMOP-CDM

- Making the registry user-friendly for non-technical users

- Integrating the OHDSI tools into the federated learning network (vantage6)

https://python-ohdsi.readthedocs.io/en/latest/

Lessons learned

- **Data harmonization** should be done at **local and network level**

- Performing **data harmonization** and **fit-for-purpose** data quality checks in **parallel**

- Iterative **updating** of the **core dataset**

- Clinicians and future user **require training for the use** of the OMOP CDM

Sneak peek

Number of patients per histology group

Data harmonization and use in federated rare cancer network
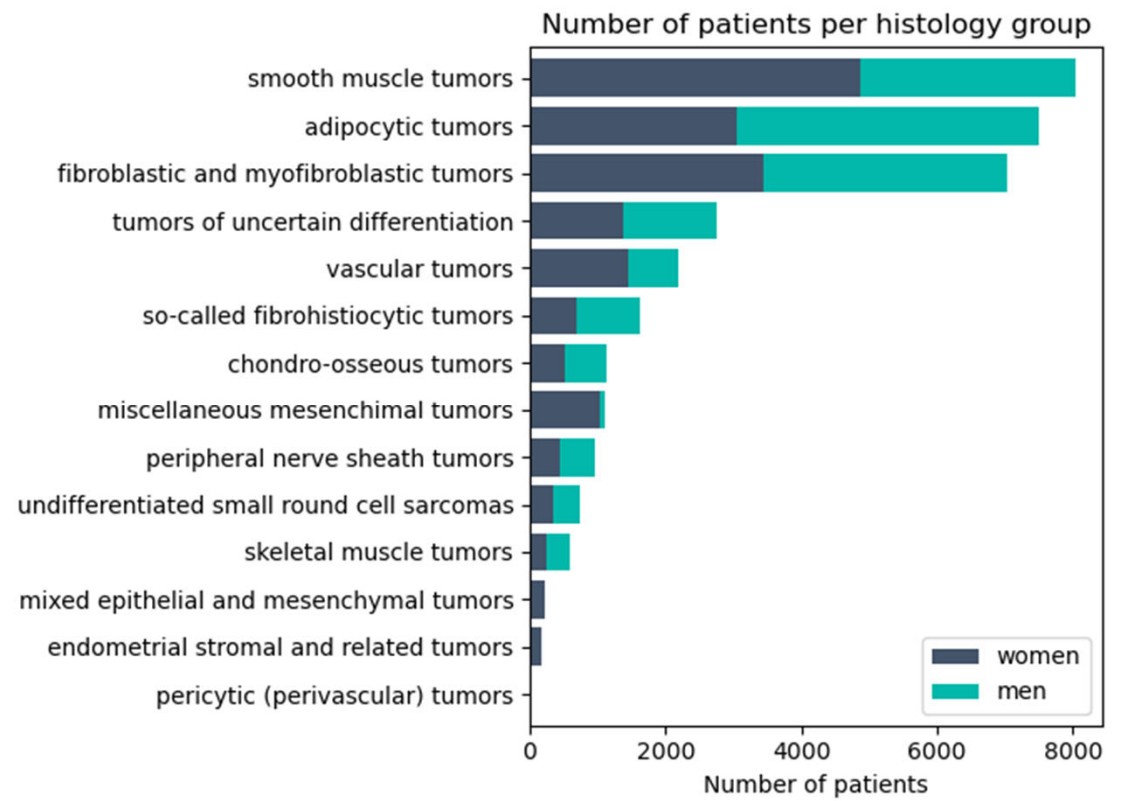
Acknowledgements

✉ m.vanswieten@iknl.nl