The IDEAL4OMOP Colorectal Cancer Project: Building Federated RWE in Oncology

Session 2 - Value of RWE in Cancer Research



Rita Rb-Silva
MD PhD, Clinical Hematologist

GEREGO, IPO-Porto Research Center
Porto Comprehensive Cancer Center (Porto.CCC)
University of Aveiro, Aveiro, Portugal

2 November, 2026

Presentation overview

01 Real-World Evidence in Oncology

02 Federated Real-World Evidence

03 The IDEAL4OMOP CRC Study

- Study design and objectives
- Implementation status
- Lessons learned
- Future perspectives

04 Key Takeaway

The presenter has no conflicts of interest to disclose.

01 | Real-World Evidence in Oncology

Clinical trials tell us what can work. RWE helps us understand what happens in practice.

Randomized Clinical Trials (RCT)



Real-world Evidence (RWE)



01 | Challenges in RWE in Oncology

Oncology generates rich clinical data, but much of its research potential remains untapped.

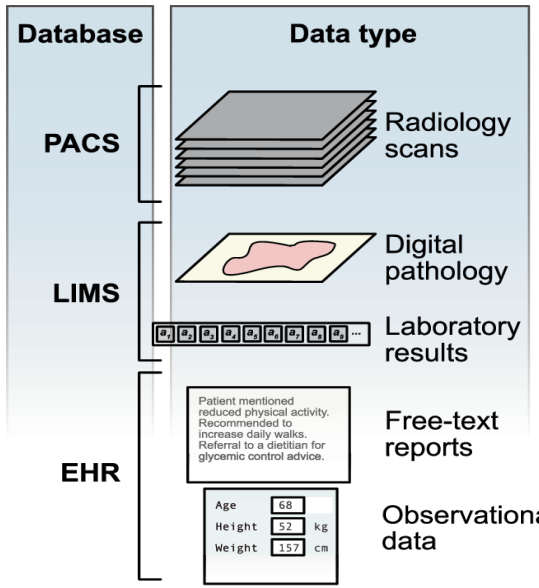
Oncology Data

Extremely rich clinical data

Generated throughout the cancer care pathway

- Diagnoses
- Pathology
- Imaging
- Biomarkers
- Staging
- Treatments and procedures
- Outcomes and follow-up
- Clinical notes

Primary use: Clinical care



Paverd et al., npj Precis. Onc. (2024)

Barriers to research use

Available data are not automatically usable

- **Fragmented:** Distributed across EHRs, registries, pathology and imaging systems
- **Unstructured:** Important variables documented in free-text clinical notes
- **Inconsistent:** Different formats, terminologies and documentation practices
- **Incomplete longitudinally:** Missing events, changing records and fragmented follow-up

01 | Challenges in RWE in Oncology

More data do not automatically mean better evidence.

Data Fragmentation
Incomplete reconstruction of patient journeys

Data Quality
Missingness, coding errors and temporal inconsistencies

Interoperability
Different meanings and representations of clinical concepts across institutions and countries

Bias and Confounding
Non-random treatment allocation and selection effects

Governance
Permissions, accountability and access to research infrastructure

Clinical Validity
Accurate phenotypes, exposures and outcomes

01 | Challenges in RWE in Oncology

More data do not automatically mean better evidence.

Data Fragmentation
Incomplete reconstruction of patient journeys

Data Quality
Missingness, coding errors and temporal inconsistencies

Interoperability
Different meanings and representations of clinical concepts across institutions and countries

Bias and Confounding
Non-random treatment allocation and selection effects

Governance
Permissions, accountability and access to research infrastructure

Clinical Validity
Accurate phenotypes, exposures and outcomes

01 | Challenges in RWE in Oncology

More data do not automatically mean better evidence.

Data Fragmentation
Incomplete reconstruction of patient journeys

Data Quality
Missingness, coding errors and temporal inconsistencies

Interoperability
Different meanings and representations of clinical concepts across institutions and countries

Bias and Confounding
Non-random treatment allocation and selection effects

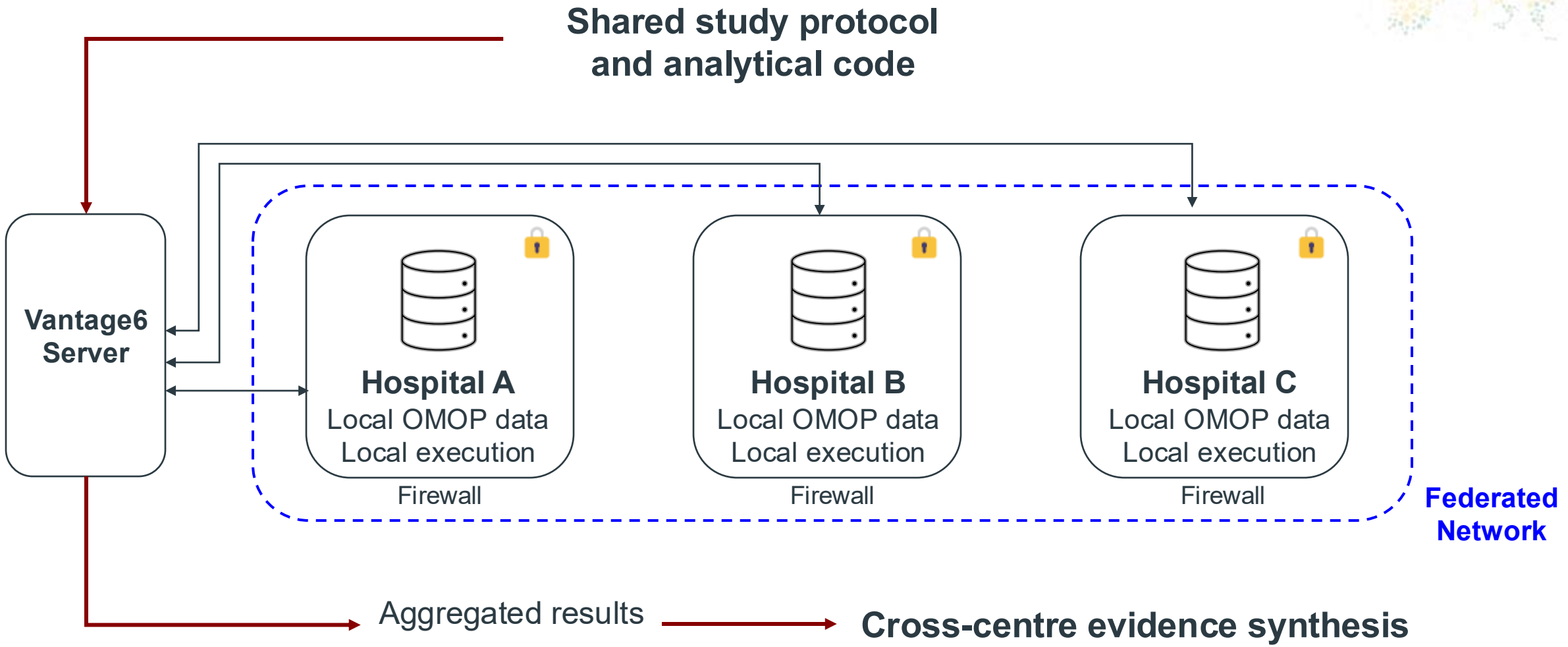
Governance
Permissions, accountability and access to research infrastructure

Clinical Validity
Accurate phenotypes, exposures and outcomes

Privacy-enhancing technologies (PETs) and approaches:
e.g. **Federated analysis**, Homomorphic encryption, Secure multi-party computation, Differential privacy, Data masking, Synthetic data, etc.

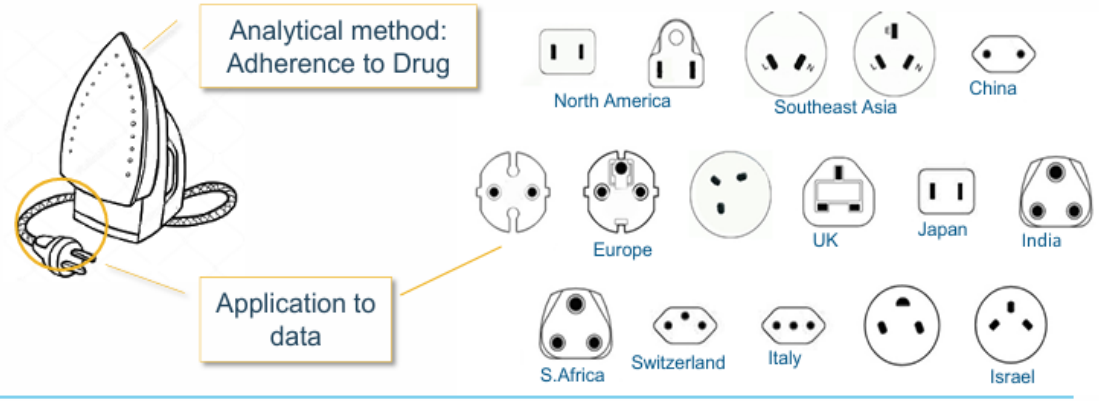
02 | Federated Real-World Evidence

Federated research connects institutions without sharing patient-level data.

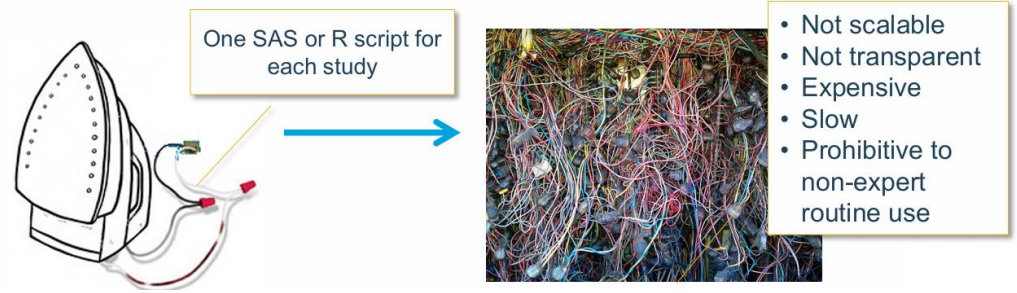


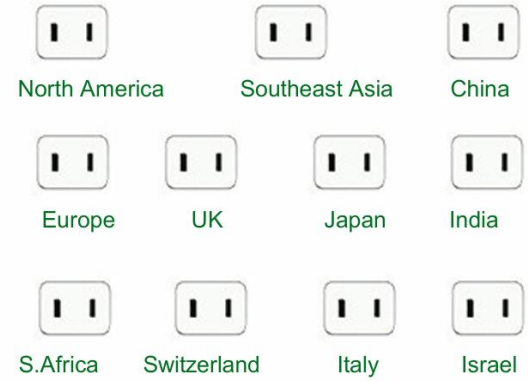
OMOP CDM gives different clinical data sources a common structure and vocabulary.

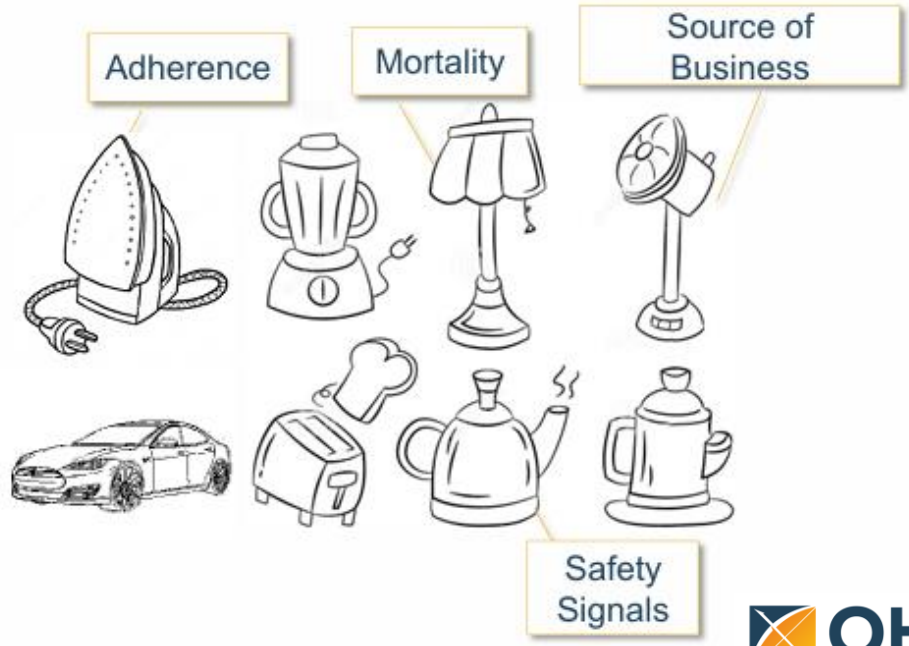
"What's the adherence to my drug in the data assets I own?"



Current solution:







02 | Federated Real-World Evidence

Keeping data local addresses one barrier, but does not remove the need for security, trust and governance.



From a European Research Network to a Funded CRC Study

DigiCore, Digital Institute for Cancer Outcomes Research

A European network advancing **digital oncology research** through collaboration, interoperable clinical data and common data standards.

IDEAL4OMOP 2024-2025

The IQVIA-DIGICORE Early-career Leadership Programme for combined technical RWE and OMOP training, research leadership development, and the creation of multinational proof-of-concept studies.

Portugal	Germany	United Kingdom	Netherlands	Lithuania
IPOPORTO	GOETHE UNIVERSITÄT FRANKFURT AM MAIN	The ROYAL MARSDEN NHS Foundation Trust; ICR The Institute of Cancer Research	Maastricht UMC+	NATIONAL CANCER INSTITUTE NCI

03 | The IDEAL4OMOP CRC

The IDEAL4OMOP CRC study brings five European cancer centres together around one clinical question.

How do early-onset and late-onset colorectal cancer differ in clinical characteristics, treatment patterns and survival?

- **Observational retrospective multicentre study**
- **Diagnosis period:** 2014-2023

18-49 years
Early-onset CRC

≥ 50 years
Late-onset CRC

- **Topography:** C18.0-C18.9, C19.9, C20.9
- **ICD-O-3:** 8140/3, 8210/3, 8211/3, 8255/3, 8261/3, 8263/3, 8265/3, 8480/3, 8490/3, 8510/3, 8560/3

- **Estimated sample size:**

Centre	EO-CRC	LO-CRC
Royal Marsden	≈ 1 900	≈ 8 000
Frankfurt	≈ 1 200	≈ 6 000
IPO-Porto	≈ 400	≈ 4 000
MUMC+	≈ 700	≈ 1 400
NCI, Vilnius	≈ 200	≈ 2 000
Sample size	**≈ 4 400**	**≈ 21 400**

A shared protocol turns one clinical question into comparable research objectives.

Study Objectives

Primary

I **Define comparable CRC cohorts**

Develop and validate EOCRC and LOCRC phenotypes

II **Characterise clinical practice and assess patient outcomes**

Demographics, comorbidities, treatment patterns and overall survival

Secondary

III **Explore Sepsis-Related Outcomes**

Occurrence, timing and potential clinical impact

Exploratory

IV **Explore Microsatellite Instability (MSI) Status**

Prevalence and potential impact on treatment outcomes

RIC-to-OMOP transformation was developed in collaboration with MTG Research Lab.

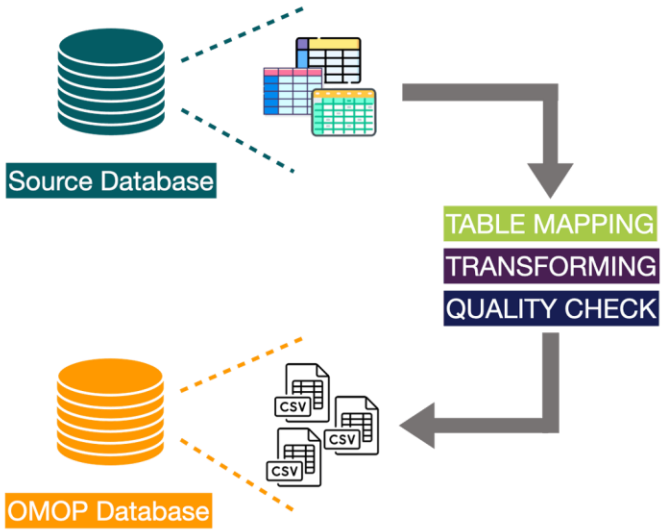
Source Data
- *"Repositório Integrado de Conhecimento"* (RIC)
- Other EHR sources

→

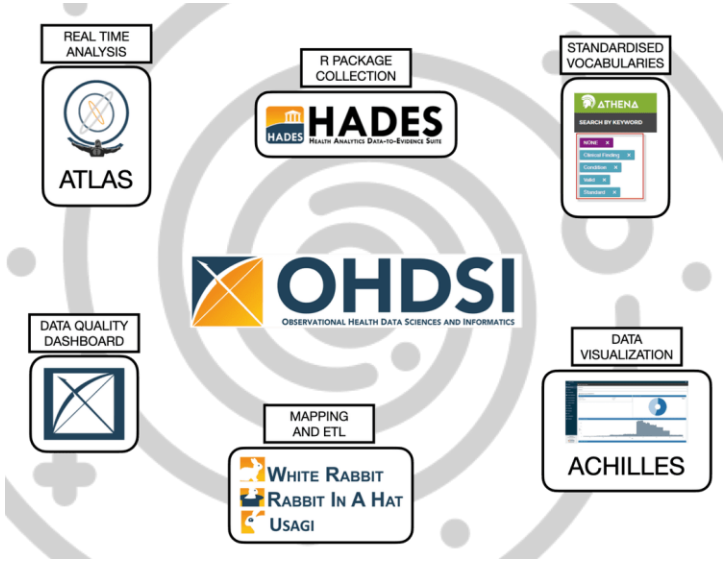
ETL & Mapping
- Extract, transform, load
- Mapping tools: White Rabbit and Rabbit-in-a-Hat

→

OMOP CDM
- Standardised representation
- Version 5.4
- Clinical validation



SciLifeLab
Precision Medicine Portal



OHDSI

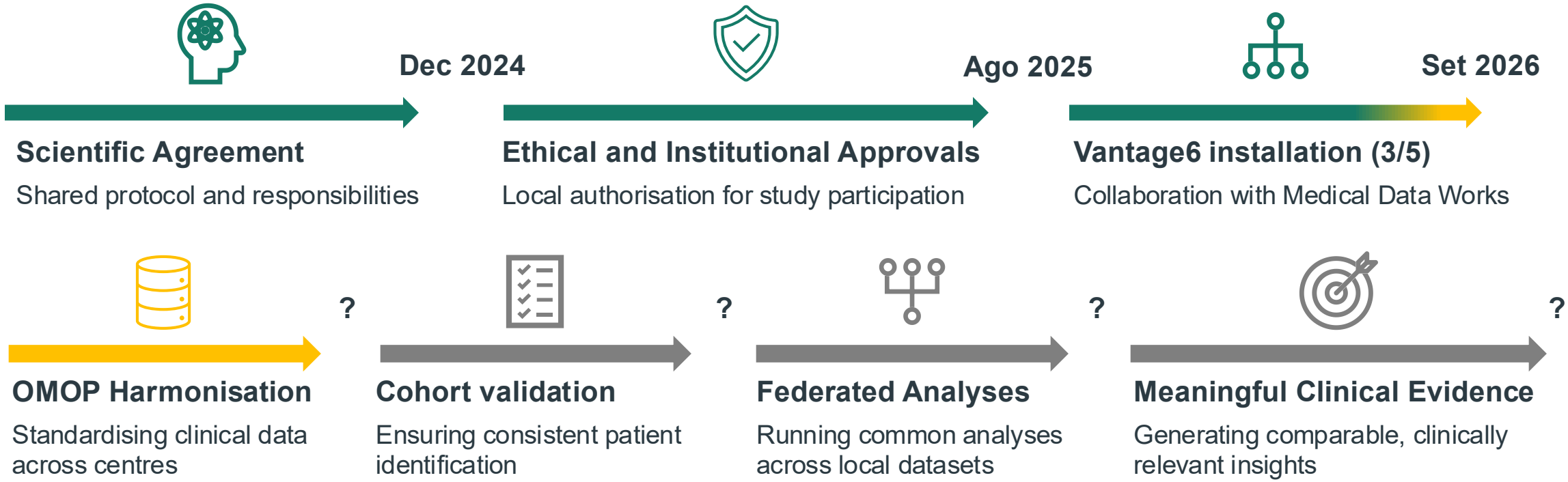
MTG RESEARCH & DEVELOPMENT LAB

DigiCore

03 | The IDEAL4OMOP CRC

From collaboration to clinical evidence, every milestone requires sustained effort.

Implementation Status



- **Complete**
- **In Progress**
- **Future**

Medical Data Works

DigiCore

03 | The IDEAL4OMOP CRC

IDEAL4OMOP CRC is one of several emerging federated research initiatives.

CancerWatch

DARWIN EU

oncovalue

CI4EU

ONCATS

promptly

REDE EVIDÊNCIA

IQVIA

AICIB AGÊNCIA DE INVESTIGAÇÃO CLÍNICA E INOVAÇÃO BIOMÉDICA

PATIENT FINDER

EUnetCCC

Funded by the European Union

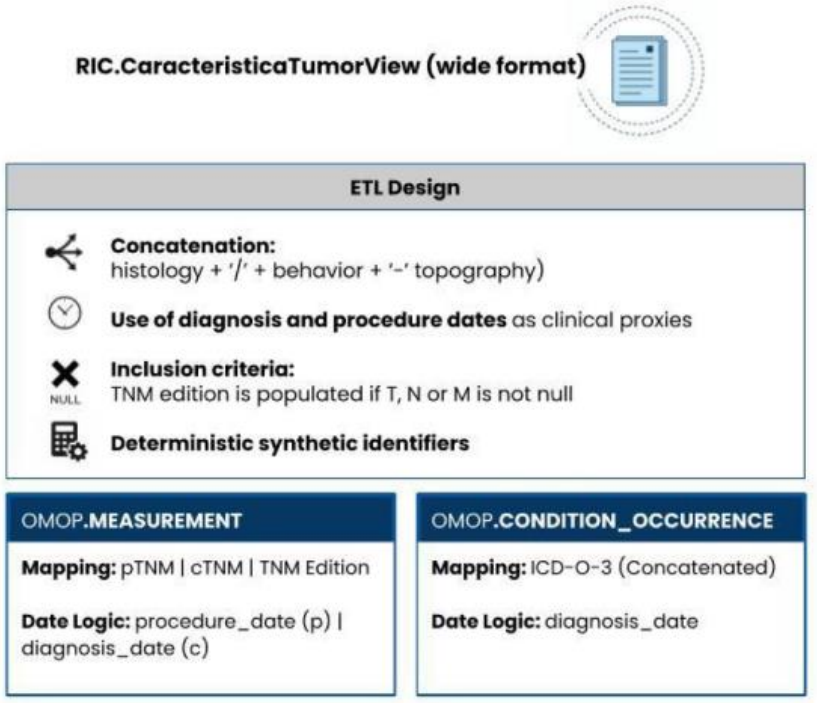
MTG RESEARCH & DEVELOPMENT LAB

VBHC/RWE

Harmonising oncology data requires more than mapping codes.

Lessons learned (I)

Clinical concept	Harmonisation challenge
Tumour diagnosis	Mapping ICD-O-3 morphology, topography and behaviour
Cancer stage	Representing TNM information consistently
Tumour characteristics	Transforming wide registry records into longitudinal observations
Event timing	Addressing missing dates and temporal uncertainty
Treatments	Extending source coverage to systemic therapies and procedures



Harmonised data must be validated before they can support reliable clinical comparisons.

Lessons learned (II)

1 - Data Quality Assessment

Completeness, plausibility, consistency and temporal integrity

2 - Clinical Phenotype Definition

Common CRC definitions, inclusion criteria and index dates

3 - Cohort Validation

Cross-centre diagnostics and clinical review

4 - Reproducible Federated Analysis

Shared analytical methods and reviewed outputs

Federated research timelines are often driven by institutional readiness, not technical complexity alone.

Original Plan

Activity	Status	Original target completion date
1. Protocol Review	**Complete**	15 May 2025
2. Development of the SAP	**Complete**	31 May 2025
3. Ethics Approval at All Sites	**Complete**	30 June 2025
4. Establishment of OMOP Databases and Vantage6 Infrastructure	In progress	31 July 2025
5. Data Analytics	In progress	30 Sept 2025
Database onboarding code	In progress	31 May 2025
Phenotyping code	**Complete**	
Primary objective: treatment patterns	Not started	31 July 2025
Secondary objective: characteristics	Not started	31 May 2025
Exploratory objective: survival	Not started	30 Sept 2025
6. Execution of Federated Data Analyses	Not started	31 Oct 2025
7. Dissemination of Results	Not started	31 Dec 2025

Scientific and regulatory readiness do not automatically translate into operational readiness.

Lessons learned (III)

- ☑ Research question
- ☑ Study team
- ☑ Protocol and analytical plan
- ☑ Funding and institutional agreements
- ☑ Ethical and regulatory approvals
- ☑ Available clinical data

≠

Operational readiness

- Dedicated IT capacity and ownership
- Tested federated infrastructure
- Harmonised and validated OMOP data
- Agreed execution and output-review procedures
- Synchronised readiness across centres
- Ongoing technical support
- Clear and continuous communication

Federation connects institutions. Governance must also connect projects within institutions.

Summary

	Challenge	Lesson Learned	Opportunity
1	**Parallel federated projects**	Independent implementations risk duplicating infrastructure and effort	**Develop a shared, institutionally governed OMOP infrastructure**
2	**Limited cross-project visibility**	Research teams need early coordination to identify common requirements	**Establish cross-project communication and joint planning**
3	**Competing IT priorities**	Federated research requires dedicated technical capacity and clear ownership	**Allocate IT resources and define institutional responsibilities**
4	**Project-specific harmonisation**	Separate mappings reduce reuse and comparability	**Reusable ETL pipelines, shared mappings and validation standards**

The next step is not just completing one study, but building a reusable research model.

Future Perspectives

Near Term

Complete harmonisation and validate CRC cohorts

Next milestone

Execute reproducible federated analyses and assess clinical outcomes

Long-Term Vision

Reusable OMOP infrastructure, federated workflows, governance and technical expertise.

From multiple projects to a sustainable institutional capability for federated research

1. Harmonisation is not merely a technical exercise.

OMOP makes common analyses possible, but clinical expertise is essential to make them meaningful.

2. Scientific readiness is not operational readiness.

Approvals, funding and a shared protocol are necessary, but they do not guarantee that all centres can execute analyses together.

3. Sustainable federation requires institutional coordination.

Shared infrastructure, dedicated technical capacity and reusable harmonisation processes can reduce duplication and support future studies.

References

- Wu P, Newby D, Alves-Rodrigues C, Okegunna Y, Schumann L, Dulskas A, Rb-Silva R. ***IDEAL4OMOP CRC Study Protocol***. 2025
- Canelas-Pais M, Silva R, Gomes S, Taveira-Gomes T, Rb-Silva R, Bento MJ, Garcia T. Poster ***Standardizing the "Repositório Integrado de Conhecimento" at IPO Porto: Design and Implementation of an OMOP CDM ETL for Oncology Data***. OHDSI Europe Symposium 2026
- Canelas-Pais M, Dienstmann R, Wang SV, Taveira-Gomes T. ***Enhancing learning health care systems with real-world data harmonization frameworks.*** ESMO Real World Data and Digital Oncology, **2026** Oct 12:14:100758. doi: 10.1016/j.esmorw.2026.100758. eCollection 2026 Dec.
- Verkerk K, Voest EE. ***Generating and using real-world data: A worthwhile uphill battle,*** Cell. **2024** Mar 28;187(7):1636-1650. doi: 10.1016/j.cell.2024.02.012. PMID: 38552611.
- Smits D, Van Beusekom B, Martin F, Veen L, Geleijnse G, Moncada-Torres A. ***An Improved Infrastructure for Privacy-Preserving Analysis of Patient Data.*** Stud Health Technol Inform. **2022** Jun 29;295:144-147. doi: 10.3233/SHTI220682. PMID: 35773828.

Thank you all for your attention!

E-mail: rita.silva@ipoporto.min-saude.pt

DigiCore

IPOPORTO

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

IQVIA™

PORTO. CCC PORTO COMPREHENSIVE CANCER CENTRE RAQUEL SERUCA

The ROYAL MARSDEN NHS Foundation Trust

Maastricht UMC+

RISE Health Research Network From the Lab to the Community

ICR The Institute of Cancer Research

NATIONAL CANCER INSTITUTE NCI 1937

MTG RESEARCH & DEVELOPMENT LAB

MDW Medical Data Works

UNIVERSITY OF OXFORD