MEDOC-ETL training

Introduction and outline

28th March 2024

DigiCore

Introducing today's speakers

The DIGICORE team

James Anderson

Leadership Development Advisor, DIGICORE

Will Sopwith

RWE Advisor, DIGICORE

The IQVIA OMOP team

James Brash

OMOP Adviser, IQVIA
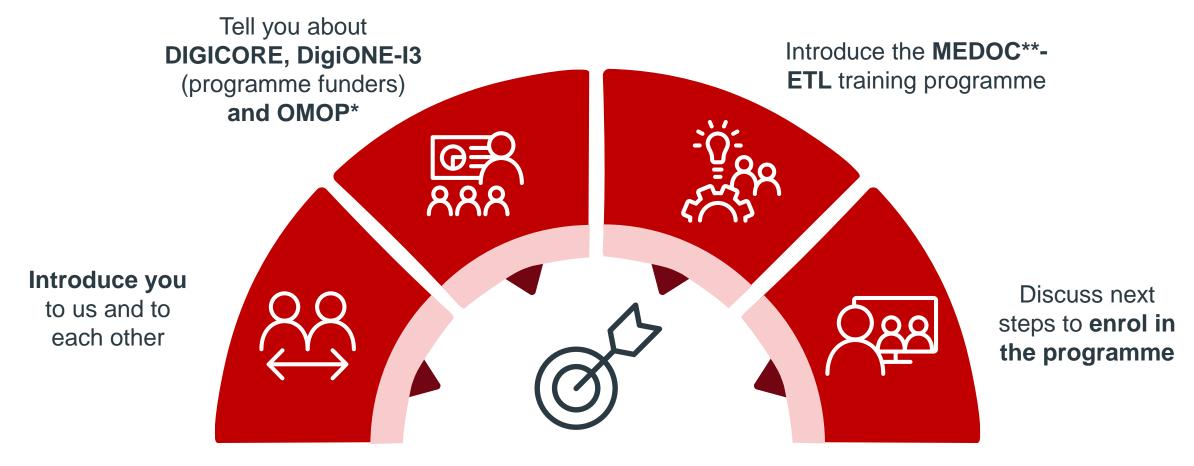
Anthony Reckard

OMOP Adviser, IQVIA

Manan Patel

OMOP Adviser, IQVIA

Please take a moment to introduce yourselves in the chat: your name, role, where you work, cancer specialism

DigiCore

The Digital Institute for Cancer Outcomes Research

MEDOC-ETL Training

Objectives for today

Tell you about
DIGICORE, DigiONE-I3
(programme funders)
and OMOP*

Introduce the **MEDOC****-
ETL training programme

Introduce you
to us and to
each other

Discuss next
steps to **enrol in
the programme**

*Observational Medical Outcomes Partnership
** Minimum Essential Description of Cancer

DIGICORE is an international consortium that aims to transform and digitise cancer outcomes research in Europe

Members

Academic cancer centres

unicancer

ALLIANCE AGAINST CANCER

Individual cancer centers

Industry

IQVIA

illumina®

DIGital Institute for Cancer Outcomes Research (DIGICORE)

Pan-EU research collaboration to study cancer outcomes, capitalizing on increase in **precision oncology**

Independent European Economic Interest Group (like OECI) with 40 cancer centres today in 17 countries

Benefits and rationale

- For **Cancer Centres**, interoperability of cancer data across sites for improved translational research

- For **Patients,** broader trial access and in future better outcomes

- For **Industrial Partners:** Drive commercial multi- centre, international RWE projects in precision oncology and drive precision trial recruitment

- Grow clinical evidence base for molecular diagnostic tests in improving outcomes and accelerate reimbursement for all vendors

To learn more, visit www.digicore-cancer.eu

DIGICORE now includes 40 cancer centres in 18 countries – Everyone welcome to join!

TAYS Cancer Centre

Vision Zero Cancer

Oslo University Hospital

Edinburgh
Leeds NHS Trust

Trinity St James

Maastricht UMC+

Saint-Luc

UCC Frankfurt

IPO Porto

Fundación Jiménez Díaz,
START Madrid,
Clínica Universidad de Navarra

Unicancer network

Institut Curie,
Centre Léon Bérard,
Centre Jean Perrin,
ICO

Tartu University Hospital

National Cancer Institute

Marie Skłodowska Curie

Masaryk Memorial

Inst of Oncology Ljubljana

Sestre Milosrdnice

Alliance against cancer network

IEO,
Gemelli,
INT Milan,
IRST,
San Matteo,
IFOM,
INT Pascale,
Regina Elena,
San Raffaele,
Humanitas Mirasole,
IRCCS Veneto

DigiCore

4

DigiONE I3 Consortium: €12.4m to upgrade 15 new sites to cancer OMOP

Country	Hospital
Belgium	Grand Hôpital de Charleroi Cliniques Universitaires Saint-Luc
Czechia	Masaryk Memorial Cancer Institute
Estonia	Tartu University Hospital
Germany	Charité Medical University Berlin Hospital
	Greifswald University Hospital
	Frankfurt University Hospital
	University Hospital Carl Gustav Carus Dresden
Ireland	Trinity St. James's Cancer Institute
Italy	Istituto Nazionale Tumori Regina Elena
	San Raffaele Hospital
Lithuania	Vilnius University Hospital Santaros Klinikos
Nether-lands	Maastricht University Medical Centre
	Maastro Clinic
	University Medical Center Groningen
Norway	Oslo university hospital
Poland	(+1 to be confirmed)
	Maria Skłodowska-Curie Institute of Oncology
UK	Leeds Teaching Hospital NHS Trust

Coordinator (EEIG)
DigiCore

Q-Helix Partner
EuroScan
International Network

Oncology Research Partner
IQVIA
IQVIA Cancer Research BV

Private Sector IT Partners

M DW

MAASTRO INNOVATIONS

edence Health

OncoDNA
THE CANCER THERANOSTIC COMPANY

CGM CompuGroup Medical

CliniNote

IQVIA
IQVIA Solutions BV
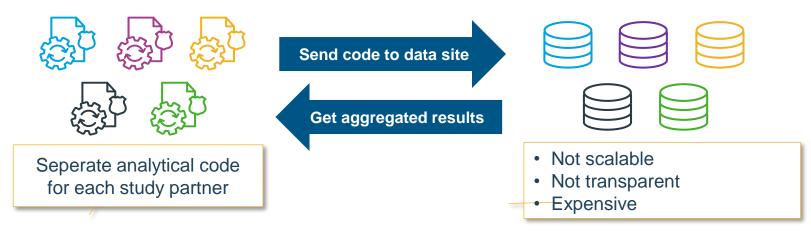
KAIROS
an IQVIA business

Create a federated, privacy-managed, high quality, digital research network built on open standards that links routine clinical data with routine molecular data at leading cancer centres across Europe

- Launched Nov 2023
- Partners invest 30% of project budget
- Pre-financing (70% of I3 funds) disbursed to partners on achieving project milestones; remainder after final report
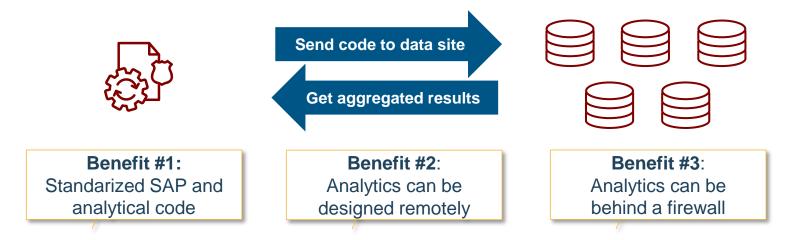
DigiCore

5

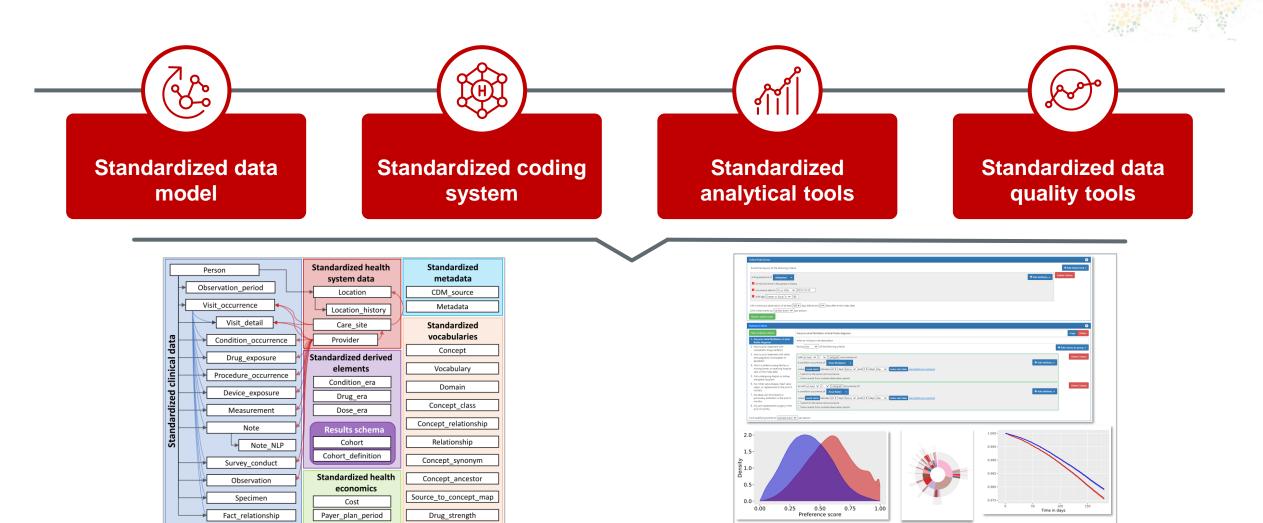# Data standardization enables reliable, reproducible and faster network studies

Classic network study:

Send code to data site

Get aggregated results

Seperate analytical code
for each study partner

- Not scalable
- Not transparent
- Expensive

Data standardized network study:

Send code to data site

Get aggregated results

Benefit #1:
Standarized SAP and
analytical code

Benefit #2:
Analytics can be
designed remotely

Benefit #3:
Analytics can be
behind a firewall

DigiCore

6

OMOP means standardization

| Standardized data model | Standardized coding system | Standardized analytical tools | Standardized data quality tools |

OMOP team publications

17 Infectious Disease publications

9 Cardiovascular publications

5 Oncology publications

19 Methods publications

4 Mental Health publications

3 Musculoskeletal publications

>40 Publications Since 2018

Frontiers

Dove Press

PDS

Pediatrics

JCO Clinical Cancer Informatics

Clinical Pharmacology Therapies

British Medical Journal

Science Reports

Pharmacoepidemiology and Drug Safety

Translational Psychiatry

JAMA

International Journal of Obesity

The Lancet

The Lancet Rheumatology

Frontiers in Pharmacology

PLoS One

Journal Of Biomedical Informatics

Annals of Internal Medicine

Science Direct

Rheumatology

Blood Advances

Hypertension

Academic Partners

EHDEN — European Health Data & Evidence Network

PIONEER — Health Data Research Hub

HARMONY

OHDSI — Observational Health Data Sciences and Informatics

COLUMBIA UNIVERSITY

UNIVERSITY OF OXFORD

Erasmus MC — University Medical Center Rotterdam

Publications by category

- Disease & Treatment Patterns
- Patient-level Clinical Evidence
- Healthcare Costs
- Policy Levers

Research areas

- Drug safety
- Drug efficacy
- Descriptive statistics
- Cohort characterization
- Risk analysis
- Comparative studies
- Method development and validation
- Combination therapy
- Vaccines

DigiCore

8

We have developed a vocabulary for DIGICORE cancer research: the "Minimum Essential Description of Cancer" ("MEDOC")

Demographics	
1.1	Date of birth (year)
1.2	Sex
1.3	Weight/*timestamp*
1.4	Height
1.5	Healthcare ID
1.6	Legal basis for data processing, e.g. consent or non-opposition

Clinical Phenotype	
2.1	Primary diagnosis and comorbidities (e.g. ICD10)
2.2	Charlson comorbidities index (derived from 2.1)
2.3	Date of primary diagnosis
2.4	Method of primary diagnosis
2.5	Performance status (e.g. ECOG, Karnofsky)/ *timestamp*
2.6	Disease stage (e.g. TNM, size, node and metastasis)
2.7	Histological cell type (e.g. ICD-O-3)
2.8	Menopausal status (breast cancer only)

Biomarkers	
3.1	Biomarker name, type
3.2	Biomarker measure and unit
3.3	Biological sample identifier, *timestamp*
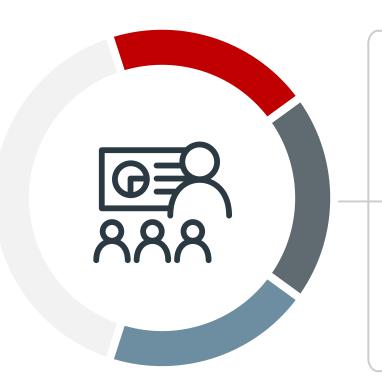
Treatment	
4.1	Line of therapy
4.2	Anti-cancer treatment (drug name)
4.3	Molecule generic name
4.4	Start date of drug treatment
4.5	Treatment dose
4.6	End date of the drug treatment
4.7	Radiotherapy type (e.g. procedure code of treatment)
4.8	Radiotherapy Start date
4.9	Radiotherapy dose
4.10	Radiotherapy end date
4.11	Surgery type (e.g. procedure code)
4.12	Surgery date
4.13	Participation in clinical trial
4.14	Date of trial consent

Outcomes	
5.1	Date of death (any location, in-hospital or from office for national statistics) Metastasis presence
5.2	Time to next treatment (derived from treatment dates)
5.3	Metastasis presence/absence
5.4	Metastasis location
5.5	Date of clinical visits (with cancer related differentiated)
5.6	Vital status (derived)
5.7	Extend of debulking surgery (gynaecological cancers only)

DigiCore

DigiONE I3 will not succeed without the right skills and experience

The people with the skills needed to conduct international digital research in OMOP **do not currently exist** in sufficient numbers

We need not only the skills, but **evidence of application of these skills, in proof-of-concept studies**, to persuade grant-makers and commercial partners

A cadre of OMOP-proficient leaders will **underpin future success for all partners**, academic and commercial

MEDOC-ETL is a critical part of this skills building

Training programmes to support i3

1

MEDOC/ETL training

To upskill <u>hospital IT staff and data scientists</u> to extract and configure data to MEDOC OMOP standard for cancer studies
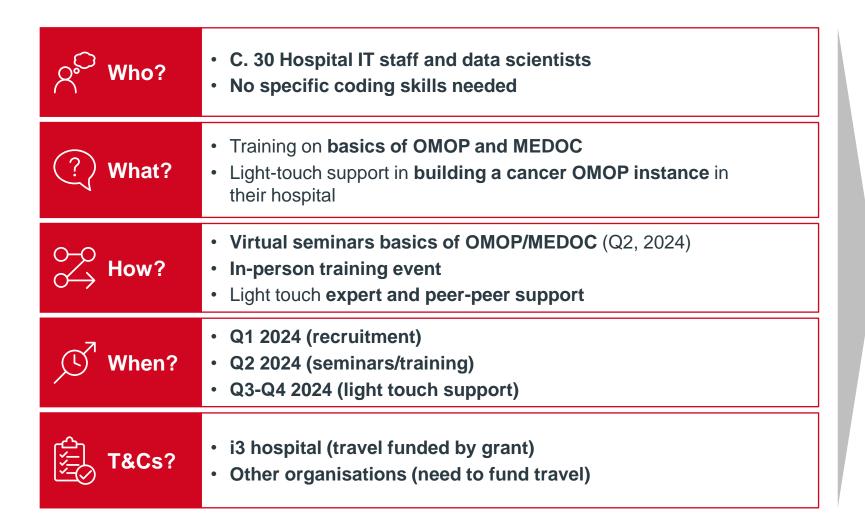
2

IDEAL4OMOP

To give <u>early career researchers (clinical and non-clinical)</u> the skills they need to design and run cancer OMOP studies

Funding from i3 award budget

DigiCore

The Digital Institute for Cancer Outcomes Research

1) MEDOC/ETL training scope and approach

Who?	• **C. 30 Hospital IT staff and data scientists** • **No specific coding skills needed**
What?	• Training on **basics of OMOP and MEDOC** • Light-touch support in **building a cancer OMOP instance** in their hospital
How?	• **Virtual seminars basics of OMOP/MEDOC** (Q2, 2024) • **In-person training event** • Light touch **expert and peer-peer support**
When?	• **Q1 2024 (recruitment)** • **Q2 2024 (seminars/training)** • **Q3-Q4 2024 (light touch support)**
T&Cs?	• **i3 hospital (travel funded by grant)** • **Other organisations (need to fund travel)**

Aims

By July 2024:

• **A cohort of 30+ OMOP trained hospital IT staff,** leading in-hospital implementation

• **A library of open source tools/documentation** for use by other hospitals (built off existing implementation guide)

By end 2024 and beyond:

• **Expanded** number of DIGICORE members with basic OMOP instance (suitable for e.g., counts)

DigiCore

12

Programme objectives

Understand OMOP Data Model

- Explore the structure and components of the OMOP Common Data Model (CDM)
- Learn about tables, relationships, and data representation

Explore OMOP Vocabulary

- Familiarize yourself with OMOP's standardized vocabulary, including concepts, codes, and terminology
- Understand how vocabulary mapping facilitates data harmonization

Convert Source Data to Basic OMOP CDM

- Learn the process of transforming source data (e.g., electronic health records) into the OMOP CDM format
- Understand the importance of standardization for research and analysis

Extend to OMOP Oncology Extension Model

- Explore the OMOP Oncology Extension, which enhances the CDM for oncology-specific data
- Understand how to incorporate cancer-related information into the model

Give you the skills, experience and tools to **build, maintain and improve an OMOP instance** in your own site

Supported with **practical exercises on synthetic data**

The programme centres around a 3-day working session in Frankfurt, Germany

Preparation
Remote working from home site

In-person working session
Frankfurt, 17-19 July*

Follow-up
Remote working from home site

- **Webinar 1** (April/May – 4 hours):
  - Basic concepts of OMOP and ETL
  - Homework assignment 1
- Participants conduct **homework**, with synthetic data
- **Webinar 2** (May/June – 4 hours)
  - Review homework assignment 1
  - Introduction to MEDOC data elements conversion
  - Homework assignment 2
- More **homework** for participants

Agenda to be refined, based on priorities identified from webinars/homework. Topics could include:

- Review of **homework assignment**
- Introduction to **OMOP oncology extension**
- Introduction/exploration of **DigiONE I3 studies**, and likely challenges posed by these
- **Deep dives in critical areas** such as quality assurance
- Further **practical exercises** on synthetic data

- Light touch support/advice from expert team (e.g., regular **drop-in clinics**)

Full OMOP instance up and running by Q4 2024

* Places limited to 1-2 people per site/participating institution

DigiCore

14

Webinar 1 overview

4 hours, April/May

Prerequisites

1. Technical Skills
- What is a database table?
- What is a column (field)?
- What is a row (record) in a database table?
- Relationships between tables
- SQL or any programing language is not required

2. Domain Knowledge
- Simple clinical terminology, such as procedure, provider, diagnosis
- Medical coding standards (e.g., ICD-9, ICD-10, RxNorm, LOINC).

Topics/activities

Topic 1: Basic OMOP Data Model and Vocabulary
- Why OMOP?
- OMOP data model introduction
- OMOP vocabulary introduction

Topic 2: ETL Process to Convert EHR Source Data into OMOP CDM
- Key concepts from team building to conducting ETL
- Common pitfalls and quality assurance

Topic 3: Convert EHR Source Data to CDM Demo
- Worked examples of going from source data to CDM

Topic 4: Homework Assignment
- Four practical assignments to test your knowledge; uses synthetic data

Webinars are open to all

DigiCore

The Digital Institute for Cancer Outcomes Research

Webinar 2 overview

4 hours, June

Prerequisites

1. Technical Skills

- Knowledge learned from Webinar 1
- Database basic: table, column, row and relationship
- SQL or any programing language is not required

2. Domain Knowledge

- Oncology related terminology, staging, metastasis, biomarker
- Medical coding standards (e.g., ICD-9, ICD-10, RxNorm, LOINC)

Topics/activities

Topic 1: Review of homework assignment in Webinar 1
- Answers, challenges and lessons learned

Topic 2: Convert MEDOC Source Data into OMOP CDM Demo
- Overview of MEDOC
- Working on examples of key MEDOC data elements (e.g., diagnosis, Charlson Comorbidity score, surgery, radiotherapy

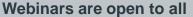**Topic 3: Cancer Modifier and Other MEDOC Data Elements Conversion Demo**
- Understand how cancer modifiers are modeled
- Convert cancer modifiers into CDM
- Conversion

Topic 4: Homework Assignment
- 4 practical homework assignments to conduct on synthetic data
- What to expect in Frankfurt

Webinars are open to all

Frankfurt F2F overview

3 days, 17-19 July

Prerequisites

1. Technical Skills

- Knowledge learned from Webinars 1 and 2
- Database basic: table, column, row and relationship
- SQL or any programing language is not required, **but may be helpful for the F2F meeting**

2. Domain Knowledge

- Oncology related terminology, staging, metastasis, biomarker
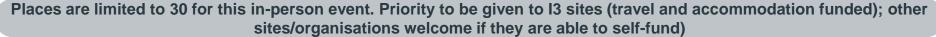- Medical coding standards (e.g., ICD-9, ICD-10, RxNorm, LOINC)

Topics/activities

Agenda tbc, but could include

- Review of **homework assignment**
- Introduction to **OMOP oncology extension**
- Introduction/exploration of **DigiONE I3 studies**, and likely challenges posed by these
- **Deep dives in critical areas** such as quality assurance
- Further **practical exercises** on synthetic data

Places are limited to 30 for this in-person event. Priority to be given to I3 sites (travel and accommodation funded); other sites/organisations welcome if they are able to self-fund)

DigiCore

Setting you up for future success OMOP-based real world research

Help your site/institution be at the cutting edge of the digital revolution

What you get

New skills

2 immersive webinar from ETL experts, with **practical homework** tasks

Face-to-face training

Follow-up support

Network of like-minded professionals

Practical advice on implementing an **OMOP instance in your site**

What you need to commit

Time

2-3 total days to July (webinars and homework)

3 days for in-person training in July

Enthusiasm for developing an OMOP instance in your site. No specific technical/programming skills required

The commitment of your employer to develop an **OMOP research instance in your site**

The Digital Institute for Cancer Outcomes Research

Next steps to register, and what to expect

- Slides and a recording of this introductory session will be made available to you in the coming days – **please tell others who may be interested**

- If you would be interested to apply for the program, please use the link to the right to access a short application form – **you will need agreement from your employer to give the necessary time**

- Registration deadline is **12th April 2024**

- Program will begin for successful applicants on **17th April 2024**

https://forms.office.com/r/Q4NLm7ReLS

DigiCore

19

Questions