ESMO GOOD SCIENCE BETTER MEDICINE BEST PRACTICE

ESMO REAL WORLD DATA AND DIGITAL ONCOLOGY

ORIGINAL ARTICLE

Establishing standards: harmonising coding principles for a minimal cancer dataset in the OMOP Common Data Model

A. Ajmal[1], O. Bouissou[2], J. Brash[1], S. Cheeseman[3], P. G. Banduge[4], A. L. Gomes[5], L. Revie[1], E. Ross[2], S. Theophanous[3], J. Thonnard[6], A. Van Maanen[6], A. Vengadeswaran[7], A. Wolf[8] & X. M. Fernandez[1,9]*

[1]IQVIA Ltd, London, UK; [2]Oslo University Hospital, Oslo, Norway; [3]Leeds Teaching Hospitals NHS Trust, Leeds, UK; [4]Department of Radiation Oncology (Maastro), Maastricht University Medical Center+, Maastricht; [5]Department of Radiation Oncology (Maastro), GROW Research Institution for Oncology and Reproduction, Maastricht, The Netherlands; [6]Cliniques Universitaires Saint-Luc, Brussels, Belgium; [7]Goethe University Frankfurt, Frankfurt; [8]University Cancer Centre, University Hospital Frankfurt, Frankfurt, Germany; [9]DIGICORE, Brussels, Belgium

Available online xxx

Background: Analysing clinical information across a network poses challenges due to heterogeneity of data collection, storage and availability. The Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) provides a standardised framework for clinical data, allowing network-level comparisons and the combination of data to enhance analytical power and increase research robustness. To capture specific oncology information across the Digital Oncology Network for Europe (DigiONE) using OMOP, we developed the Minimal Essential Description of Cancer (MEDOC) framework.

Materials and methods: MEDOC was developed through several iterations and was then utilised in DigiONE pilot studies. This was a community-driven process, made possible by discussions to distinguish differences in hospital data and by conducting deep-dive sessions to solve specific issues in aligning source data with the MEDOC structure.

Results: The initial version of MEDOC has been utilised in two DigiONE observational cancer studies to date with a further two studies in progress, and training resources including the implementation guide have been developed. Lessons learned in the development of our MEDOC to OMOP alignment include challenges in establishing diagnosis date, confirming metastasis location and tumour classification code due to granularity of available data, among other challenges specific to individual centres.

Conclusion: The utility of MEDOC has been evidenced in research applications and will be continually developed in line with both learnings from centres and developments in the field of oncology. The implementation of MEDOC in line with the OMOP CDM is timely, given European initiatives to harmonise health care data systems.

Key words: cancer research, Observational Health Data Sciences and Informatics (OHDSI), Common Data Model (CDM), Observational Medical Outcomes Partnership (OMOP)

INTRODUCTION

Cancer refers to an incredibly complex set of diseases with a long pathway from diagnosis to treatment, follow-up and beyond. This is reflected in the digital registrations for monitoring of cancer, in which challenges include data heterogeneity and a combination of semi-structured and unstructured data depending on the clinical setting of recording. In many centres, oncology data sources—composed of clinical, pathology, imaging, genomic and treatment data—are stored across multiple source systems and can be subject to missingness and inaccuracies.[1] However, even where data are structured, interoperability challenges are presented, such as variability in clinical coding between different health care settings [e.g. Systematized Nomenclature of Medicine (SNOMED) versus International Classification of Diseases for Oncology third edition (ICD-O-3)], and differences in database storage systems.

The Observational Medical Outcomes Partnership (OMOP) consortium was initiated in 2008 to improve the monitoring of drug safety using observational health care data. For this purpose, OMOP proposed a medical model, the OMOP Common Data Model (CDM), as a way of enabling consistent data analysis and facilitating large-scale observational research, since it enables the harmonisation of diverse datasets into a common format. The OMOP CDM is a comprehensive database definition designed to address some of these challenges in assessing clinical data

*Correspondence to: Dr Xose M. Fernandez, IQVIA Cancer Research, IQVIA Ltd, The Point, 37 North Wharf Road, London, UK. Tel: +44 (0) 7423 816 109
E-mail: xose.fernandez@iqvia.com (X. M. Fernandez).

by providing a standardised framework to which data can be mapped.[2] OMOP is equipped with >70 mapped terminologies and a set of tools provided by the Observational Health Data Sciences and Informatics (OHDSI) community.

The OHDSI community has since taken stewardship of the OMOP CDM, continuing its development to the recent OMOP oncology module.[3] The OMOP oncology module is being developed to support a more comprehensive characterisation of the complete disease pathway with the goal of providing a standardised cancer-specific data framework for observational studies. However, work on the module is ongoing to address all specific use cases required,[3] meaning that at present, OMOP for cancer research remains underdeveloped.

Despite the use of well-established models like the OMOP CDM to standardise clinical data, challenges in maintaining consistency across a network with local implementations persist, as reported in previous studies.[4-6] For example, previous work has reported that it is difficult to achieve the highest quality OMOP mapping without an understanding of the underlying medical practices at a centre, highlighting the importance of working with site-specific clinicians.[5]

The Digital Institute for Cancer Outcomes Research (DIGICORE) is a pan-European research community established in 2020 to accelerate the implementation of precision oncology. Over the past 2 years, DIGICORE has developed a Digital Oncology Network for Europe initiative (DigiONE), with the aim of creating and developing a privacy-preserving network of cancer centres with a core cancer dataset. The OMOP CDM has been selected to be used by the DigiONE centres for two reasons: (i) data from local systems employing differing clinical coding can be harmonised by using OMOP CDM inbuilt vocabularies, and (ii) OMOP's inherent extensibility can accommodate additional data elements that hold research significance.

The Minimal Essential Description of Cancer (MEDOC) was developed using the OMOP CDM to support the aims of DigiONE.[7] MEDOC is a collection of data concepts which was designed to cover key elements of demographics, clinical phenotype, biomarkers, treatments and patient outcomes, to provide a snapshot of a tumour. MEDOC does not aim at completeness (which often requires manual retyping), but delivery of insights based on universally available concepts across centres.

Randomised clinical trials are the cornerstone of evidence-based medicine, offering high-quality data on the effectiveness and safety of medical interventions.[8] However, their application in real-world clinical practice can go unrealised, resulting in uncertainties regarding their true impact on patient care. In light of this, providing an accessible framework such as MEDOC by which clinical data can be assessed and analysed in a consistent manner is paramount to producing high-quality and impactful real-world evidence (RWE).

To implement MEDOC, each DigiONE centre first collected the required data and then converted this to a local OMOP instance or extracted MEDOC concepts from an existing OMOP instance if already in place. When considering MEDOC on OMOP translation, additional challenges arise as there is an intricate relationship rather than a simple one-to-one mapping. Many discussions were conducted within the network to ensure OMOP implementations were consistent across participating centres/hospitals in line with OHDSI conventions and scientifically grounded.

This work focuses on our journey to develop the MEDOC on OMOP specification, which is practically outlined in our MEDOC implementation guide—a comprehensive and practical document to support centres in implementing MEDOC (see Supplementary materials, available at https://doi.org/10.1016/j.esmorw.2025.100179). This guide was crucial in ensuring consistency in data mapping across centres and thereby providing a firm foundation for observational cancer research within the network.

MATERIALS AND METHODS

To aid in the deployment of MEDOC, and to minimise interoperability issues across the network, we developed an implementation guide through a collaborative effort involving several iterations within the DigiONE network. Key contributors to its development included Cliniques Universitaries Saint-Luc (Belgium), Leeds Teaching Hospitals NHS Trust (UK), Frankfurt University Hospital (Germany), Maastricht University Medical Center+ (The Netherlands) and Oslo University Hospital (Norway), in partnership with IQVIA. This was a community-driven process aimed at ensuring the guide was comprehensive and applicable across different hospital archetypes in the network.

Once the components of MEDOC were agreed upon,[7] deep-dive sessions were conducted for each MEDOC concept to understand data availability at each hospital and to reach a consensus on the data transformation rules to OMOP. These sessions focused on both technical and scientific aspects, which involved collaboratively resolving topics such as agreeing the level of detail on mapping metastasis location across the network to ensure compatibility with all centres' data, and an algorithm for managing date of diagnosis. Data were considered both from a top-down perspective, looking at national standards, and a bottom-up perspective, examining electronic health records and ancillary databases to meet both adherence to national quality and local availability of concepts. Discussions were also held with the OHDSI cancer OMOP community to seek guidance and resolution on specific issues, such as the approach for distinguishing between somatic and germline mutations, or how new clinical trial drugs and placebos could be captured in OMOP. The output of these discussions formed the basis of the early implementation guide.

The implementation guide was created based on OMOP tables, with the first version only including the tables required for the first DigiONE pilot study on cancer diagnosis volumes during the coronavirus disease 2019 (COVID-

19) pandemic.[9] The development was iterative, requiring consistent feedback from the sites to ensure that the guidance was appropriate considering their data availability. The guide was refined further as part of the DigiONE pilot studies[9,10] and is currently in its second version, incorporating learnings on where greater specificity is required.

RESULTS

During the implementation of OMOP databases across the DigiONE centres, several challenges arose. We present the lessons learned during this process.

MEDOC concepts to OMOP

The OMOP CDM is a complex relational database model and MEDOC includes data concepts that require multiple underlying data items. This means there are several one-to-many relationships between single MEDOC concepts and multiple OMOP tables and/or columns, of which some source data have no obvious OMOP mapping. The simplest example is the MEDOC concept 'date of birth', which consists of up to three columns in the OMOP person table (year, month and day of birth).

In addition, a single OMOP table may contain several MEDOC concepts—for example, date of birth, sex and health care ID are three unique concepts contained within the OMOP person table. Moreover, some OMOP tables contain multiple records for the same patient and/or diagnosis; for example, multiple records per patient are present in the OMOP procedure table for every MEDOC procedure event (e.g. surgery, radiotherapy).

Importing data into the OMOP databases involved data transformation and extraction before mapping all source attributes and values from various sources of data within the hospitals into the OMOP CDM, as outlined in the MEDOC implementation guide (see Supplementary material, available at https://doi.org/10.1016/j.esmorw.2025.100179). To enable interoperability, the OMOP CDM relies on a curated set of standardised vocabularies, terminologies and accepted context that must be followed when importing data into the OMOP databases. During the data importing process, source codes are mapped to these standard concepts using the OMOP vocabulary system, which is maintained and regularly updated by the OHDSI community.

This extract, transform, load (ETL) process required both design and implementation phases, utilising OHDSI's 'Rabbit' toolkit including WhiteRabbit,[11] RabbitInAHat[12] and Usagi[13] applications and in some cases bespoke applications or tools that were already in place in centres under the guidance of clinical experts.

For the latter, data import was validated by counting rows and elements in the source dataset files and the OMOP database to ensure completeness, and OHDSI's Achilles[14] and Data Quality Dashboard software[15] were used to check correctness according to OMOP conventions. However, even following quality checking, the importance of harmonising on the OMOP vocabulary version should not be overlooked. OMOP mapping can be correct, and quality assured in individual centres, but when analysing across a network, differences in vocabulary versions can create discrepancies in data structure which then need to be accounted for in the analytical design.

When creating an OMOP instance, several columns are mandatory and must be included in the CDM. However, some of these required columns do not map to any MEDOC concept; for example, as per OMOP convention, 'race_concept_id', which denotes ethnicity, is a required column

Table 1. Minimal Essential Description on Cancer (MEDOC) concept list

MEDOC concept	OMOP table
1.1 Date of birth[a] 1.2 Sex 1.5 Health care ID	PERSON
2.1 Primary cancer diagnosis 2.3 Primary diagnosis date 2.2 Comorbidities (required for Charlson comorbidity index) 2.7 Histological cell type	CONDITION_OCCURRENCE
5.1 Date of death	DEATH
5.5 Date of visits	VISIT_OCCURRENCE
4.13 Participation in clinical trial 4.14 Date of trial consent 2.8 Menopausal status[a]	OBSERVATION
1.3 Weight[a] 1.4 Height[a] 2.5 Performance status[a] 2.6 Disease stage[a] 3.1 Biomarker name 3.2 Biomarker measure[a] 3.3 Biomarker sample ID 5.3 Metastasis presence/absence[a] 5.4 Metastasis location 5.7 Extent of debulking[a] 4.9 Radiotherapy dose	MEASUREMENT
2.4 Method of primary diagnosis 4.7 Radiotherapy type 4.8 Radiotherapy start date 4.10 Radiotherapy end date 4.11 Surgery type 4.12 Surgery date	PROCEDURE_OCCURRENCE
4.2 Anticancer treatment name 4.3 Molecule generic name 4.4 Start date for drug treatment 4.5 Treatment dose[a] 4.6 End date for drug treatment	DRUG_EXPOSURE, EPISODE & DRUG_STRENGTH
None	LOCATION CARE SITE PROVIDER OBSERVATION_PERIOD (not in MEDOC but crucial to define patient journey)
1.6 Legal basis for data processing 4.1 Line of therapy (derived) 5.2 Time to next treatment (derived) 5.6 Vital status (derived)	N/A one-off confirmation or derived centrally

N/A, not applicable.

MEDOC concept conversion to Observational Medical Outcomes Partnership (OMOP) standard tables.[b,c]

[a]Concepts which require multiple OMOP columns.

[b]Table details general mapping but it should be noted that some codes for specific concepts may deviate from this. For example, morphology can go to condition occurrence and observation tables. In these situations, sites should follow the rules set by the OMOP vocabulary conventions.

[c]OMOP vocabulary version and differences in OMOP mapping between centres can also influence the relation of MEDOC to OMOP concepts. These differences can be accounted for at the analysis design phase.

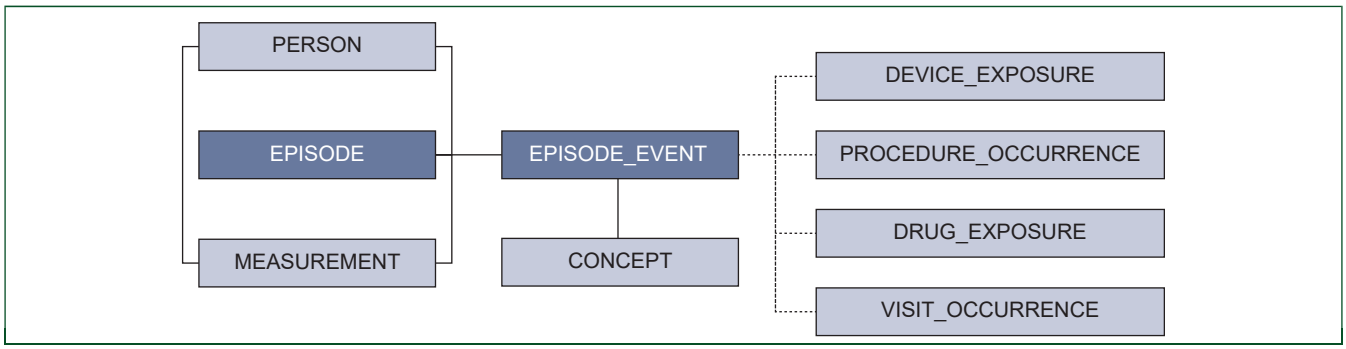

Figure 1. Episode and episode event in the OMOP CDM. The episode and episode event tables consolidate individual clinical events into groupings that represent disease phases, outcomes and treatments. Episode event connects qualifying clinical events to the correct episode.
CDM, common data model; OMOP, Observational Medical Outcomes Partnership.

within the person table. This OMOP table column is not included in the MEDOC framework, predominantly as this information is not collected universally by centres and/or typically has a high level of missingness. However, for centres where this information is not included in the source data, 'race_concept_id' can be '0' to allow the OMOP instance to be established according to convention.

Given these conventions of OMOP and MEDOC, and the ETL process involved for clinical centres to map to the CDM, multiple training resources were developed. These include the MEDOC implementation guide but extended to practical training courses and one-to-one technical support for centres delivered by OMOP experts. The resultant MEDOC to OMOP conversion is summarised in Table 1.

Although there are no MEDOC concepts stored in the OBSERVATION_PERIOD table, it is an important part of the OMOP CDM and should be populated. This table tracks the time period(s) containing a record of all the clinical events for a patient within a period, and is primarily used to (i) determine whether a patient had sufficient prior observation for a study, and (ii) determine the follow-up for any time-to-event analyses (see Supplementary materials, available at https://doi.org/10.1016/j.esmorw.2025.100179).

From OMOP CDM v5.4, the addition of the episode and episode_event table consolidates individual clinical events, which allows the organisation of complex patient journeys into grouping that represent disease phases, outcomes and treatments (Figure 1). This is particularly crucial for determining clinically relevant events for cancer patients—specific guidance for the population of the event tables within the MEDOC framework is provided in the implementation guide and aligns with guidance from the OHDSI community.[3] The integration of HemOnc ontology into the OMOP CDM to support the characterisation of cancer regimens has also been developed by the OHDSI community. While this ontology remains predominantly relevant to United States regimens, we continue to engage with the wider community on aligning MEDOC to OHDSI extensions where relevant.

Date of diagnosis and primary diagnosis

Initial conversion work detected that participating centres possess varying levels of detail on the local diagnostic pathway, therefore requiring a consensus approach to data mapping to attain consistency. This was achieved through discussions with technical and clinical experts, from which a universal approach to each MEDOC concept was agreed on.

The need for consensus is particularly evident when considering date of diagnosis. Cancer has a complex multimodal diagnostic work-up, with clinical assessment, imaging and pathology typically conducted over a period of several weeks. In some cancers, biochemical diagnosis is also possible, resulting in the 'triple date of diagnosis' issue in cancer—most patients have at least three possible different diagnosis dates (based on clinical assessments, imaging or pathology). Given this, the European Network of Cancer Registries (ENCR) has developed algorithmic rules[16] for the 'gold-standard' definition of date of diagnosis.

Highly digitally mature sites in our network were found to capture nuanced information including the date that the patient was informed of the diagnosis in structured records; however, this was rarely available across other sites. To address this, we first mapped the different levels of data granularity held by each of the centres. Comparing across the centres revealed that pathology and imaging data on date of diagnosis were universally available and thus formed the backbone of our approach. Where ENCR date of diagnosis was available, it was considered the gold standard and was used as the date of diagnosis; however where ENCR date was unavailable, the date of the pathology procedure was used, and where this was also unavailable the date of imaging procedure was used.

Rules were then agreed more widely for data assignment to concepts in each site's OMOP instance (Table 2) to ensure uniformity across the network—for example, use of 'condition_start_date' and 'condition_type_concept_id' to differentiate between dates originating from pathology and imaging with those from ENCR.

In addition, the preferred vocabulary type for cancer diagnosis codes in MEDOC is ICD-O-3, as it provides more granular detail than SNOMED (the OMOP standard for non-cancer diagnosis codes), although both coding systems are valid. However, not all centres' source data were coded using this vocabulary. ICD-O-3 requires the combination of topography and histology information which was not always

Table 2. Date of diagnosis in MEDOC

OMOP table	OMOP column	Source data	Type of source data for date of diagnosis used in order of priority		
			1. ENCR date of diagnosis	2. Pathology procedure and dates	3. Imaging procedure and dates
CONDITION_OCCURRENCE	condition_concept_id	Primary cancer diagnosis	Primary diagnosis		
	condition_start_date	Date of procedure that resulted in primary cancer diagnosis	ENCR date	Date of pathology procedure	Date of imagine procedure or report
	condition_type_concept_id	Provenance of record, e.g. claims from a billing system	Source system of the ENCR data	Source system of the pathology procedure	Source system of the imaging procedure
	condition_status_concept_id	Whether the diagnosis is a primary diagnosis or a recurrence	Primary cancer diagnosis (32902 primary diagnosis) or recurrence (32908 secondary diagnosis)		
PROCEDURE_OCCURRENCE	procedure_concept_id	Procedure used for the primary cancer diagnosis	N/A	Type of procedure	
	procedure_date	Date of procedure resulting in primary cancer diagnosis		Date of pathology procedure or report	Date of imagine procedure or report
	procedure_type_concept_id	Provenance of record, e.g. claims from a billing system		Source system of the pathology procedure	Source system of the imaging procedure

Priority mapping of date of diagnosis-related concepts in MEDOC to OMOP based on availability of data at local centres.
ENCR, European Network of Cancer Registries; MEDOC, Minimal Essential Description on Cancer; N/A, not applicable; OMOP, Observational Medical Outcomes Partnership.

readily available. Therefore, to achieve uniformity, we had to build a consistent mapping approach to standardise all diagnostic codes for the CDM. This workflow is illustrated in Figure 2. We do not discuss here the molecular characteristics of the tumour (as indicated in the European Society for Medical Oncology Clinical Practice Guidelines on the management of various indications) as these genomic biomarkers required specific development, following extensive consultation with the OHDSI Oncology Working Group.

Metastasis location mapping

After reviewing data sources across the network, metastasis location was found to be captured heterogeneously across the network, with some centres able to extract directly from structured data, while others required extraction solutions.

Furthermore, the granularity of data available was highlighted when initial analyses returned limited patients with metastasis located in the brain for some centres. Further review and discussion demonstrated that data heterogeneity across centres meant that in some cases metastases location was further sub-classified. For example, data for metastases of the brain could also be subdivided further into parenchymal disease of the less common leptomeningeal disease in some centres, resulting in seemingly limited patient numbers of 'brain metastases'. Following this, consensus was reached that MEDOC should be specified to capture the most granular data available in this context for specificity of future analyses, but this example demonstrated that studies are limited by the lowest common denominator of information captured across the network.

These lessons learned demonstrate the multi-faceted nature of aligning data in MEDOC. Discussions required consideration of both data format, prioritisation of data type and data availability across each concept in addition to careful system planning to ensure sustainability, bespoke training[17] and educational resources beyond the rich tapestry provided by OHDSI.

DISCUSSION

Cancer diagnoses are defined though attributes such as histology and grade, disease stage and primary location, as well as a growing number of biomarkers, which determine prognosis and treatment options. Treatments often involve complex personalised chemotherapy regimens, targeted therapies, immunotherapies, surgery or radiotherapy, making clinical data difficult to standardise within and between centres.

Using OMOP alone may not be sufficient in its current format to ensure consistent standardisation across a research network due to variations in local data practices. In this article, we have outlined the additional steps taken within the DigiONE network to ensure consistent setup of OMOP installations, which is a fundamental prerequisite for conducting high-quality RWE studies, and the implementation of MEDOC to provide a minimal essential framework for standardising oncology data.

The success of MEDOC as a foundational framework for observational cancer research can be evidenced in two studies within the DigiONE network to date.[9,10] This highlights the efficacy of the implementation guide and demonstrates proof-of-concept for MEDOC to be employed across a wider network of sites with varying data formats and systems. Moreover, within 14 months of DigiONE initiating the first study investigating time to diagnosis and treatment of cancers throughout the COVID-19 period,[9] initial findings were presented. This demonstrates MEDOC's translation to high-quality RWE across a network within a rapid time frame, providing a platform from which key cancer research questions can be readily addressed.

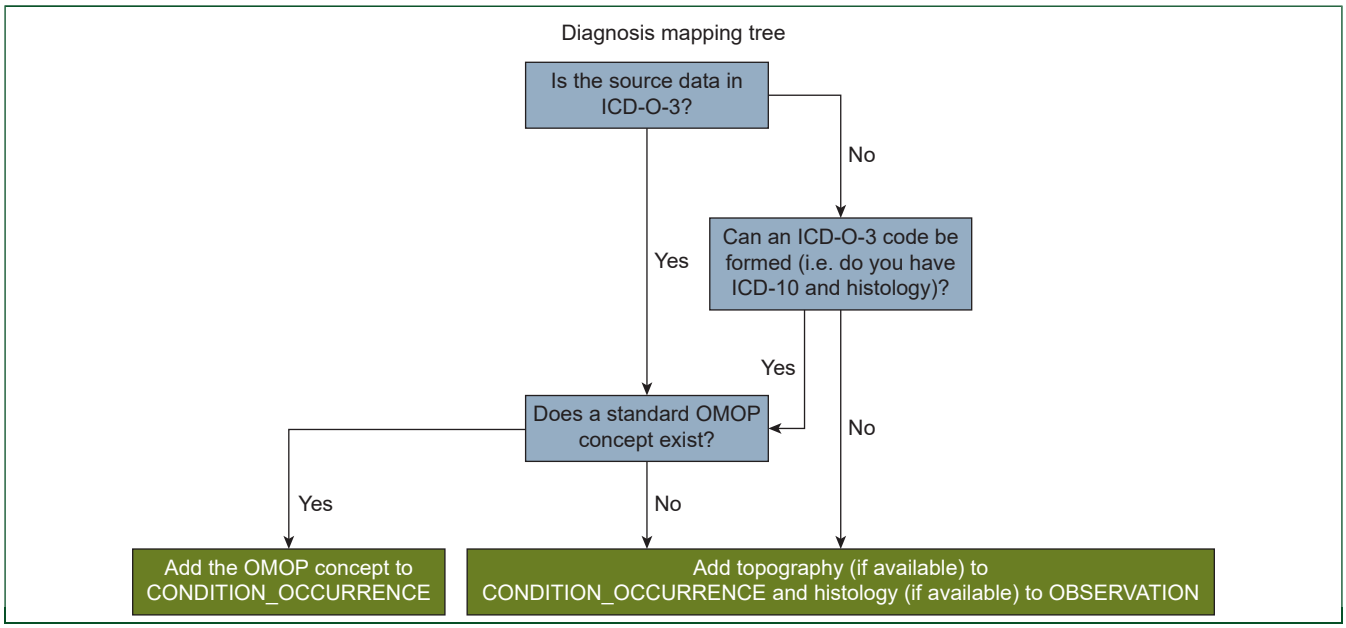

Figure 2. MEDOC Diagnosis code mapping tree. Diagnosis concept mapping workflow for standardising diagnostic codes to International Classification of Diseases for Oncology third edition (ICD-O-3) format for Observational Medical Outcomes Partnership (OMOP) frameworks. Cases where standard OMOP concept for ICD-O-3 does not exist include (i) the topography and morphology information required to form the ICD-O-3 form, a code which is not standard but maps to Systematized Nomenclature of Medicine (SNOMED); (ii) topology and histology does not map to standard codes and requires manual mapping.

It should be noted that there is sufficient technical effort and time commitment required in the process of creating an OMOP instance in centres which do not currently hold them. While this, and ultimately implementing MEDOC, provides returns in the ease of which research can be conducted centre wide, it requires initial investment of resources from centres.

Moreover, consideration must be given to the information governance of the OMOP installations. For example, OMOP vocabularies are subject to updates which must be carefully managed across the network to ensure consistency of vocabulary versions without causing disruption to ongoing studies.

One of the key next steps is to align as many centres as possible to the common MEDOC framework. This allows both standardised concepts to be accessed readily for cancer studies and allows centres to track key performance indicators independently and across clinical and research networks, to ultimately improve cancer outcomes on a wider scale. This is pertinent within European centres, given the European Health Data Space (EHDS) plan to unify health care data.[18] EHDS is a European initiative with the aim to facilitate the exchange of data for health care delivery across the EU and establish a consistent system for use of health date in research activities. This initiative highlights the increased importance for centres to align clinical data within a standardised framework across Europe, with frameworks such as MEDOC being pivotal to align at the earliest possible stage before data harmonisation on a wider scale.

The longer-term vision of MEDOC is to extend the framework to encompass additional concepts that are required for informative research, such as improving radiotherapy (as radioligands are adopted) or incorporating key concepts in haematological malignancies. Some of these concepts are currently under development in the OMOP oncology module, and once they are established MEDOC can be developed and mapped accordingly following the lessons learned from the original process. However, these developments will be balanced with the primary role of MEDOC as a minimal concept list—careful consideration will be required in future to determine the necessity of including additional concepts.

Conclusion

The first version of MEDOC (v1.05) is established in DigiONE sites at the time of writing, and we plan to develop MEDOC as the network grows to new indications and as changes in the field of oncology progress. In this way MEDOC will evolve in line with the clinical landscape to remain consistent with user requirements, while maintaining the primary aim of providing a minimal dataset aligned to the OMOP CDM. For many centres, MEDOC is the first step in the journey to implementing full OMOP coverage, extending the possibilities of high-quality RWE still further.

ACKNOWLEDGEMENTS

The OMOP analysts and scientists at IQVIA, and technical and clinical colleagues at individual centres provided valuable input and expertise which informed the work described here. The content of the paper and processes and experiences described are those of the authors and should not be attributed to those acknowledged.

FUNDING

This work was supported by IQVIA and Illumina (no grant number). Ongoing work expanding the network is supported by the European Commission under project 101115081.

DISCLOSURE

AA, JB, LR and XMF are IQVIA employees. All other authors have declared no conflicts of interest.

REFERENCES

1. Berenguer-Albiñana CB, Pallocca M, Fenton H, et al. Developing PRISM: a pragmatic institutional survey and bench marking tool to measure digital research maturity of cancer centers. *Appl Clin Inform*. 2024;15(4):743-750.
2. Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. *J Am Med Inform Assoc*. 2012;19(1):54-60.
3. Belenkaya R, Gurley MJ, Golozar A, et al. Extending the OMOP common data model and standardized vocabularies to support observational cancer research. *JCO Clin Cancer Inform*. 2021;5:12-20.
4. Papez V, Moinat M, Voss EA, et al. Transforming and evaluating the UK Biobank to the OMOP Common Data Model for COVID-19 research and beyond. *J Am Med Inform Assoc*. 2023;30(1):103-111.
5. Oja M, Tamm S, Mooses K, et al. Transforming Estonian health data to the Observational Medical Outcomes Partnership (OMOP) common data model: lessons learned. *JAMIA Open*. 2023;6(4):ooad100.
6. Biedermann P, Ong R, Davydov A, et al. Standardizing registry data to the OMOP Common Data Model: experience from three pulmonary hypertension databases. *BMC Med Res Methodol*. 2021;21:1-6.
7. Mahon P, Chatzitheofilou I, Dekker A, et al. A federated learning system for precision oncology in Europe: DigiONE. *Nat Med*. 2024;30(2):334-337.
8. Granholm A, Alhazzani W, Derde LP, et al. Randomised clinical trials in critical care: past, present and future. *Intensive Care Med*. 2022;48:164-178.
9. Theophanous S, Fenton H, Lobo Gomes A, et al. Harmonization of routine care data from hospitals in the Digital Oncology Network for Europe. Paper presented at the OHDSI Conference 2024. October, 2024. Available at https://www.ohdsi.org/wp-content/uploads/2024/10/86-Theophanous-Harmonization-of-routine-care-data-from-hospitals-in-the-Digital-Oncology-Network-for-Europe-Jie-Yeap.pdf. Accessed August 30, 2025.
10. Öjlert AK, Jain P, Hendriks LE, et al. Federated analysis of overall survival (OS) by location of metastases (mets) in patients (pts) with metastatic NSCLC (mNSCLC) from the Digital Oncology Network for Europe (DigiONE). *Ann Oncol*. 2024;35(2):859.
11. Observational Health Data Science and Informatics. WhiteRabbit. Available at https://ohdsi.github.io/WhiteRabbit/WhiteRabbit.html. Accessed February 5, 2025.
12. Observational Health Data Science and Informatics. RabbitInAHat. Available at https://ohdsi.github.io/WhiteRabbit/WhiteRabbit.html. Accessed February 5, 2025.
13. Observational Health Data Science and Informatics. Usagi. Available at https://ohdsi.github.io/Usagi/. Accessed February 14, 2025.
14. Huser V, DeFalco FJ, Schuemie M, et al. Multisite evaluation of a data quality tool for patient-level clinical data sets. *EGEMs*. 2016;4(1):1239.
15. Kahn MG, Callahan TJ, Barnard J, et al. A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data. *EGEMs*. 2016;4(1):1244.
16. European Network of Cancer Registries. ENCR Recommendation Basis of Diagnosis. October 24, 2022. Available at https://encr.eu/sites/default/files/Recommendations/ENCR%20Recommendation%20BoD_Oct2022_EN_241022.pdf. Accessed August 30, 2025.
17. Anderson J. DIGICORE, an international research consortium, is building infrastructure and skills to transform digital oncology research across Europe. *ESMO Real World Data and Digital Oncology*. 2024;3:100023.
18. European Commission. Regulation of the European Parliament and of the Council on the European Health Data Space. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52022PC0197. Accessed February 7, 2025.