DigiCore

DigiONE MEDOC Implementation Guide V1.06

Last Updated: 23 May 2025

CONTENT

A. Introduction to OMOP and the Implementation Guide

Goal: The purpose of this Implementation Guide is to provide comprehensive support to DigiONE partners on the transformation of source data related to the Minimal Essential Description Of Cancer (MEDOC) into the Observational Medical Outcomes Partnership (OMOP) common data model. This transformation leads to the creation of each partner's OMOP Research Data Repositories (RDR), which will be used for DigiONE studies.

It is crucial that each DigiONE OMOP RDR is developed using the conventions outlined in this document to ensure harmonisation across the network, which is a prerequisite for high-quality real world evidence (RWE) study execution.

Background on digital maturity

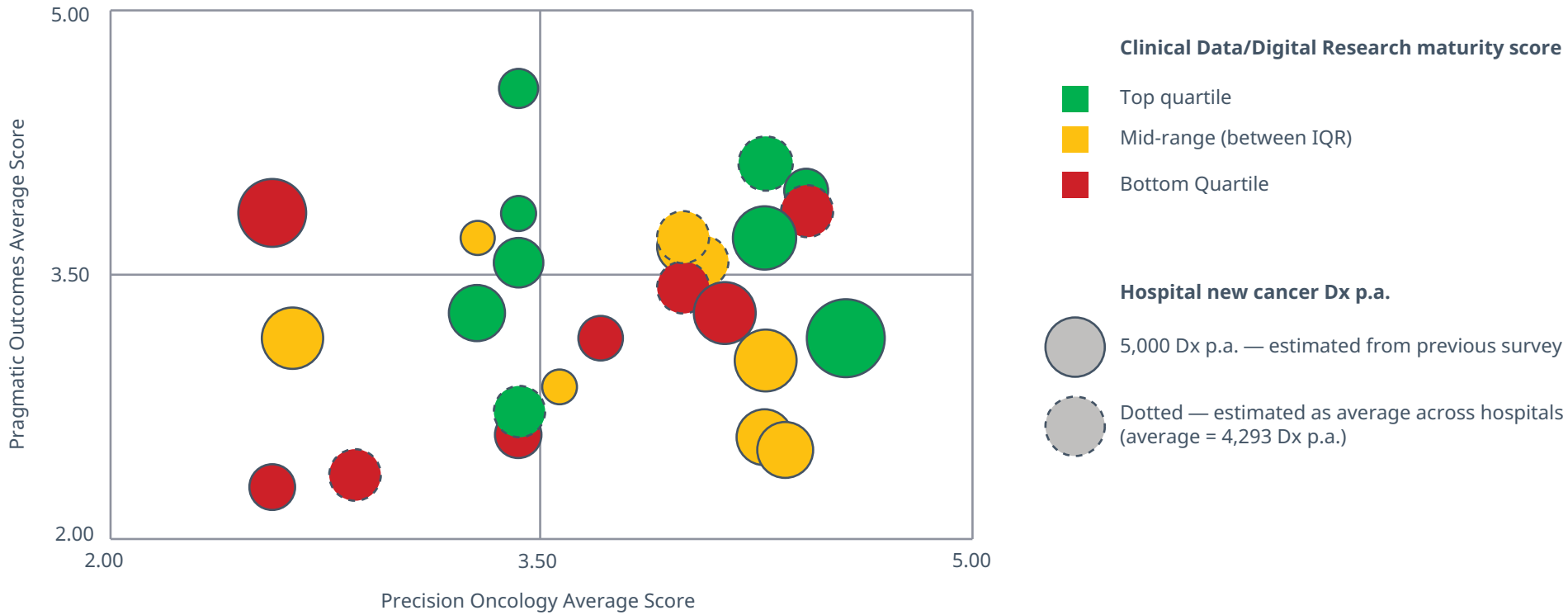
There is increasing international consensus that digital transformation is the way to tackle the challenges health systems face today. The transformation journey to achieve digital excellence passes by implementing and prioritising technologies and services to drive improved clinical outcomes.

Established assessment of digital excellence frameworks have mapped a number of stages of digital maturity, from absence of digital tools to fully digitalised hospitals. Yet, most frameworks focus exclusively on technology and not the human and organisational angle (which is the main reason behind failure in large IT projects in the healthcare setting). Within DIGICORE we have developed a way of measuring the level of digital maturity in our network, grouping well established digital centres with others taking their first steps in this journey.[1]

Focusing on data availability, we have explored four dimensions across our hospitals: precision oncology, clinical digital data, pragmatic outcomes and information governance and delivery. Figure 1 provides a snapshot picture of digital maturity across various DIGICORE members. As an example, hospitals in the top right quadrant would make good partners for clinical biomarker validation research, having both good outcome and molecular data availability. Members of the bottom right quadrant have data suitable for biomarker driven trial recruitment, but not a broader range of real-world research use cases. This notion of digital excellence serves an important purpose, providing a vision that can help motivate stakeholders and coordinate activities towards the pursuit of better outcomes.

[1] Berenguer Albiñana et al. 2024

Figure 1. Cancer digital maturity snapshot. [Berenguer Albiñana et al. 2024][2]



DigiONE research studies and the scope of your ETL

The DigiONE Scientific Committee prioritise cancer indications and research questions for the network, and encourage inputs from collaborators across DIGICORE. As per the DigiONE engagement rules hospital within the network have autonomy over which studies they participate in, taking into consideration the delivery effort, level of structured or unstructured data, funding and resources available.

By analysing the structured vs. unstructured dimension, a critical aspect to consider when assessing the digital maturity of a hospital we have identified different hospital archetypes, providing valuable information about a hospital's readiness to leverage data for decision-making, research, and patient care. We will find highly digital centres with very structured and therefore easily retrieved data, alongside other centres with unstructured data often trapped in PDF on the other side of the spectrum. Hospitals should consider their baseline digital maturity when designing their Extract-Transform-Load (ETL) process.

[1]Berenguer Albiñana et al. 2024

Archetype 1: extensive structured data availability in core Electronic Health Record (EHR), in all cancer indications. The level of effort for these hospitals is low and built into the ETL. We strongly recommend maximizing data integration by efficiently automating movement of structured data from source into OMOP. This approach is particularly valuable considering the diverse source systems targeted for ETL processes to generate Minimal Essential Description Of Cancer (MEDOC) for DigiONE studies – *more detail in* ***Context on OMOP and MEDOC*** Consequently, your OMOP Research Data Repository (RDR) could expand significantly, encompassing not only cancer but also all departments and diseases.

Archetype 2: limited structured data availability in core EHR, **with NLP capability**. The level of effort will be higher and may require tumour-specific ETL. If your hospital belongs to this archetype, you may have difficulty accessing the basic data set that you need for your analysis. You can use NLP and specialized sub-systems to extract and organize the relevant data from your EHR. We advise you to focus all your efforts on this task, as it is essential for your success.

Archetype 3: limited structured data availability in core EHR, **without NLP capability** OR **hospitals with limited resources**, we suggest initially just implementing OMOP for the delivery of the key concepts (discuss with the program team).

Context on OMOP and MEDOC

OMOP Common Data Model

The mission of the Observational Health Data Science and Informatics (OHDSI) community is to enhance patient healthcare by offering a collection of open-source software tools and methodologies dedicated to data standardization and analysis to support observation health research. This suite of technologies encompasses essential elements such as the OMOP common data model and its Standardized Vocabularies (e.g., SNOMED, RxNorm, ICD-O-3 etc.), which effectively harmonize the way observational data is recorded. Additionally, the OHDSI community provides guidelines and tools for the development of ETL processes to facilitate data transfer into the OMOP Common Data Model (CDM), as well as a robust framework tailored for data analytics.

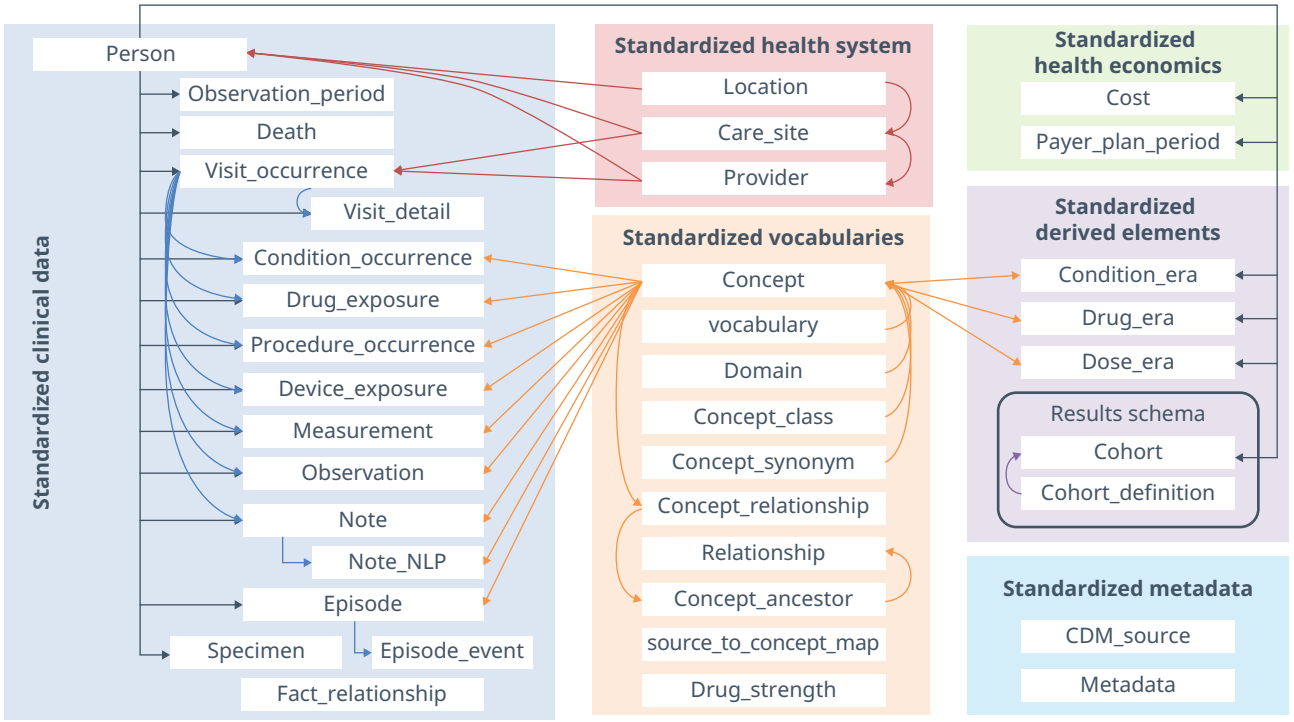
OMOP was selected as the CDM for DigiONE. Partners will need to build an OMOP RDR using the standard schema of the OMOP CDM (**Figure 2**) to create a local research database for federated research.

Key OMOP resources to use alongside the Implementation Guide:

- OMOP CDM v5.4 specification: https://ohdsi.github.io/CommonDataModel/cdm54.html
  - » This open-source specification provides ETL conventions for each OMOP Table and any constraints that should be followed. The Implementation Guide is aligned with and references the OHDSI guidance and conventions.
- The Book Of OHDSI: https://ohdsi.github.io/TheBookOfOhdsi
  - » This book serves as a central resource for all aspects of the OHDSI community and includes specific guidance on topics including the ETL process (Chapter 6)
- Join the OHDSI teams channel under your national node using these instructions

[3]from OHDSI

Figure 2. OMOP CDM (version 5.4) Standard Structure[3]



Minimal Essential Description Of Cancer (MEDOC)

In 2022, a collaborative effort involving 16 prominent European cancer centres came together for a consensus-driven initiative. The core objective was to delineate a minimal target dataset that would be conducive for wide-ranging outcomes research across Europe. This dedicated endeavour led to the creation of the Minimal Essential Description Of Cancer (MEDOC). MEDOC includes data concepts that i) describe cancer sufficiently, ii) are considered clinically important and iii) have reasonable electronic health record availability across Europe (even if not in structured data formats).

The first version, MEDOC V1.0, comprises 38 essential concepts. These concepts have been categorized into five domains: demographics, clinical phenotype, biomarkers, treatment, and outcomes. Each category plays an integral role in providing a comprehensive framework for accurate and insightful representations of cancer-related data (**Figure 3**). These concepts are clinically important for accurately describing cancer and enabling outcome research. The full MEDOC Data Dictionary V1.0 can be found in the Appendix (**Table 5**). MEDOC may be expanded in the future, with modules added to deal with aspects not covered in the current minimal dataset.

Figure 3. Consensus process results for MEDOC v1.0, as the "Derived in protocol" concepts[4]

1. Demographics	2. Clinical Phenotype	3. Biomarkers	4. Treatment	5. Outcomes
1.1 Date of birth (month)	**2.1 Primary cancer diagnosis** and comorbidities, typically in ICD	**3.1 Biomarker name**	**4.1 Line of therapy** (*derived algorithmically within each cancer type*)	**5.1 Date of death**, at any location
1.2 Sex	**2.2 Charlson comorbidity index** (derived from 17 comorbidities in 2.1)	**3.2 Biomarker measure**	**4.2 Anti-cancer treatment name,** including systemic treatment and supportive therapy	**5.2 Time to next treatment** (*derived form treatment start dates*)
1.3 Weight (with timestamp)	**2.3 Date of primary diagnosis**	**3.3 Biological sample ID**	**4.3 Molecule generic name**	**5.3 Metastasis presence/absence**
1.4 Height	**2.4 Method of primary cancer diagnosis**		**4.4 Start date for drug treatment**	**5.4 Metastasis location**
1.5 Healthcare ID (or other unique identifier)	**2.5 Performance status** (for example, codes by ECOG or Karnofsky standards)		**4.5 Treatmet dose**	**5.5 Date of clinical visits** (with cancer related visits separated from other visits)
1.6 Legal basis for data processing	**2.6 disease Stage** in a recognized standard such as TNIM		**4.6 End date for drug treatment**	**5.6 Vital status** (derived from visits or death linkage)
	2.7 Histological cell type, typically in ICD-O-3 standards		**4.7 Radiotherapy type**	**5.7 Extend of debulking surgery** (for example, for patitents with gynecological cancer)
	2.8 Menopausal status (for example, for patients with breast cancer)		**4.8 Radiotherapy Start date**	
			4.9 Radiotherapy dose	
			4.10 Radiotherapy end date	
			4.11 Surgery type	
			4.12 Surgery date	
			4.13 Participation in clinical trial	
			4.14 Date of trial consent	

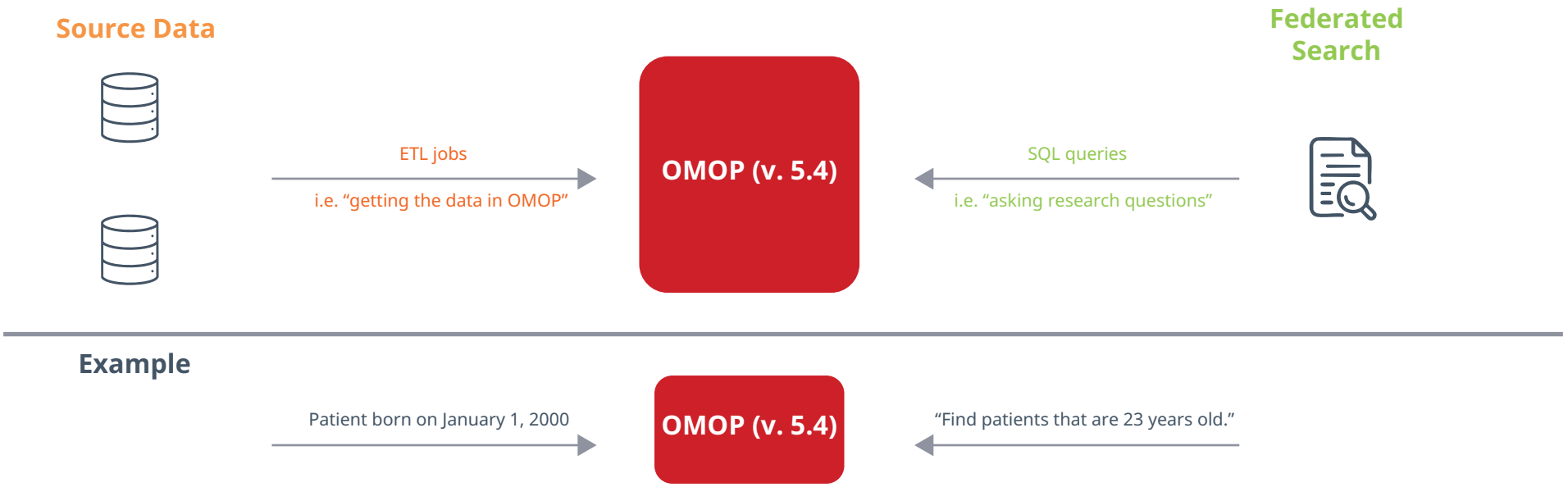
Note: Implementation of 1.5 and 5.1 is influenced by national regulations; 2.8 and 5.7 are essential only in some cancers for which risk-normalized audit or research cannot take place without their capture.

[4]Adapted from Mahon, P., et al. A federated learning system for precision oncology in Europe: DigiONE. Nat Med 30, 334–337 (2024).

It is important to distinguish the data concepts in the OMOP Research Data Repository (RDR) from the derived data concepts which will be defined in the protocols for particular research questions. In order to ingest MEDOC data from the source systems into OMOP RDR, bespoke ETL jobs must be developed. This process will entail creating software programs or scripts that facilitate the movement and transformation of data from source systems (in this case, the MEDOC data) into the target system, the OMOP RDR. This will involve first extracting data from databases, files, APIs, or other data sources where the MEDOC data resides. After extracting the data, it often needs to be transformed to fit the structure and requirements of the target system (OMOP RDR). Finally, the transformed data is loaded into the OMOP RDR or the target system.

Following the successful ingestion of data into the OMOP database, protocol-derived data extraction can be facilitated through e.g., SQL queries. It is important to note, however, that these are two different processes (**Figure 4**):

Figure 4. ETL vs. SQL queries for source data vs. protocol-derived variables



This guide concentrates exclusively on the extract, transform, and load processes required for RDR construction, excluding the derivation of protocol terms. **Table 1** illustrates several instances showcasing the distinction between these two. **These derivations will take place based on code provided centrally at the study level, sites do not need to derive these.**

Table 1. Examples of MEDOC concepts, their "pre-derivation" data partners in the RDR, and protocol-derived variables

MEDOC Data concept	MEDOC RDR for OMOP (MEDOC-in-OMOP)	Variables derived in protocol
Date of birth	Date of birth	Age at study index date in years
Charlson Co-morbidity Index (CCI)	Individual disease diagnosis beyond cancer, with a focus on the 17 co-morbidities of CCI	CCI calculated
Time to next treatment	Start and end dates of treatment, typically from visits for chemo etc.	Apply rules to calculate gap as TtNT
Date of visits	The specific date on which a patient attended a cancer-related in-person visit or telehealth consultation (e.g., phone consultation)	Apply rules to calculate date of last visit and vital status at a particular timepoint
Date/method of diagnosis	Multiple dates and methods may exist, depending on source data	Apply rules from protocol to select the date and method of diagnosis for primary cancer diagnosis

The relationship between MEDOC and OMOP:

MEDOC is small by design: collectively we can work on creating high completeness and quality data. MEDOC should be transformed into OMOP CDM for two different reasons: i) OMOP's inherent modular design provides a dedicated space for accommodating additional data elements that hold research significance; ii) The utilization of OMOP's vocabularies not only ensures enhanced interoperability but also facilitates the establishment of data harmonization across a diverse array of medical research datasets.

Because the OMOP CDM is a complex relational database model and MEDOC includes data concepts that require multiple underlying data items, the relationship between MEDOC and OMOP CDM is not as straightforward as a simple one-to-one mapping and may involve various combinations. Multiple factors contribute to this complex relationship:

1. A single MEDOC concept may need to be ingested into numerous OMOP tables and columns (one-to-many relationship)
   - E.g., MEDOC concept "Date of birth" consists of up to 3 columns in the PERSON table (PERSON.year_of_birth, PERSON.month_of_birth, PERSON.day_of_birth). Note that Date of birth is used to derive age in protocols and does not leave the RDR
2. A single OMOP table may contain several MEDOC concepts
   - » Within the same table e.g., the PERSON table 3 unique MEDOC concepts are contained (Date of birth, Sex, Healthcare ID)
   - » Repeated entries in tables for the same patient/diagnosis e.g., multiple records within the PROCEDURE table for every MEDOC concept procedure event (surgery, radiotherapy)
3. Deploying an OMOP instance involves a number of "required" columns with constraints, which may not map to MEDOC concepts, but essential to ensure communication with other OMOP instances and therefore 'hardcoded' in the CDM
   - » E.g., PERSON.race_concept_id is a required column that does not map to a single MEDOC concept. For countries where ethnicity is not allowed to be collected, or in centres where is collected but missing in some cases, race_concept_id column can be '0'.

Designing the ETL

Where to start when designing your ETL job? Because of the foreign key relationship, we recommend filling OMOP tables in the following sequence:

1. Location
2. Care_Site
3. Provider
4. Person
5. Visit_Occurrence
6. Event tables e.g., Condition_Occurrence, Procedure_Occurrence etc.)
7. Observation_Period

How to use the Implementation Guide (Section B)

The MEDOC data concepts have been categorised according to the OMOP table(s) where they will be stored. **It is important to note that there are rare deviations from the table below, in which case always map according to OMOP convention** (e.g., histologies are typically placed in the CONDITION table, however some codes may go in the OBSERVATION table). The Implementation Guide is organised according to OMOP tables (Section B). Each OMOP table section is written with the following structure:

1. Introduction to OMOP Table

- A brief description of the OMOP table(s) is provided, along with its significance, relevance, and interconnected OMOP tables.

2. MEDOC RDR Concepts & Definitions:

- This section lists the definitions for each MEDOC RDR concept that pertains to the OMOP table required for the minimum specification.
- Additionally, within this section, further resources will be provided for hospitals equipped with more extensive datasets. These resources will be linked to the OHDSI specification.

3. Implementing the OMOP Table

» This part provides detailed advice on how to deal with complex MEDOC concepts and how to convert them into the OMOP format.

» The OMOP table is presented with the following columns:

Column	Type	OMOP Required	DigiONE Required	Related to	ETL guidance for DigiONE nodes
[CDM Field]	integer	Yes	Yes		
[CDM Field]	integer	Yes	Yes		

The fields surrounded by a **box** (**grey**) are related to one or more MEDOC concepts.

- **Column:** contains all the fields in the OMOP table
- **Type:** the data format
- **OMOP Required:** Yes (**green**) means that there is a “not null” constraint
- **DigiONE Required:** Yes (**green**) means that this data point is required for DigiONE research
- **Related to:** indicates any linked OMOP table/field
- **DigONE User Guidance:** Guidance for DigiONE partners on source data to ingest and OMOP accepted concepts, aligned with OHDSI guidance

The MEDOC concept *Legal basis for data processing (1.6)* is not stored in the OMOP RDR. It is each hospital’s responsibility to confirm with their data privacy officers that there is adequate legal basis to use the data for research purposes and that no patients have specifically opted out of this. For regulatory purposes each centre would be expected to have traceability of the record of legal basis for data processing outside of OMOP.

B. Implementation Guide

QUICK GUIDE

MEDOC concepts	OMOP table
1.1 Date of birth	PERSON
1.2 Sex	
1.5 Healthcare ID	
2.1 Primary cancer diagnosis	CONDITION_OCCURRENCE
2.3 Primary diagnosis date	
2.2 Comorbidities (for CCI)	
2.7 Histological cell type	
5.1 Date of death	DEATH
5.5 Date of visits	VISIT_OCCURRENCE
4.13 Participation in clinical trial	OBSERVATION
4.14 Date of trial consent	
2.8 Menopausal status	
1.3 Weight	MEASUREMENT
1.4 Height	
2.5 Performance status	
2.6 Disease stage	
3.1 Biomarker name	
3.2 Biomarker measure	
3.3 Biomarker sample ID	
5.3 Metastasis presence/absence	
5.4 Metastasis location	
5.7 Extent of debulking	
4.9 Radiotherapy dose	

MEDOC concepts	OMOP table
2.4 Method of primary diagnosis	PROCEDURE_OCCURRENCE
4.7 Radiotherapy type	
4.8 Radiotherapy start date	
4.10 Radiotherapy end date	
4.11 Surgery type	
4.12 Surgery date	
4.2 Anti-cancer treatment name	DRUG_EXPOSURE, EPISODE & DRUG_ STRENGTH
4.3 Molecule generic name	
4.4 Start date for drug treatment	
4.5 Treatment dose	
4.6 End date for drug treatment	
None	LOCATION
	CARE SITE
	PROVIDER
	OBSERVATION_PERIOD (not in MEDOC but crucial to define patient journey)
1.6 Legal basis for data processing	N/A one-off confirmation or derived centrally
4.1 Line of therapy (derived)	
5.2 Time to next treatment (derived)	
5.6 Vital status (derived)	

PERSON

MEDOC Concept(s): Date of birth (1.1), Sex (1.2), Healthcare ID (1.5),

1. Introduction to OMOP Table: PERSON

- The PERSON table provides a unique identification number (person_id), and some demographic data for each individual in the database. Each individual in the dataset should have a single record in the PERSON table (where data is being used from multiple sources, efforts must be made to link individuals across sources and create a single record per person).

2. MEDOC Concept Definitions

- *Date of Birth (1.1):* The date on which a person was born or is officially deemed to have been born. If the exact date is not accessible, then use month & year of birth or if not available then use the year of birth.
  - » Note that date of birth is used to derive age in protocols and does not leave the RDR, nor will it be shared outside of the originating hospital
- *Sex (1.2):* In most countries this is the biological sex at birth. If not available (in some countries), then use a person's gender as self-declared (or inferred by observation for those unable to declare their sex).
- *Healthcare ID (1.5):* The "healthcare ID" refers to a local patient identifier.

3. Implementing the OMOP Table: PERSON

When importing your source data into the PERSON table, please ensure that you, at a minimum, fill out the "DigiONE Required" fields. However, considering the pivotal role of the PERSON table, we strongly recommend that you comprehensively complete all the remaining fields:

Column	Type	OMOP Required	DigiONE Required	Related to	ETL guidance for DigiONE nodes
person_id	integer	Yes	Yes		We recommend using a randomly generated integer ID of 7 digits for person_id. The hospital should maintain the link between the source system and the randomly generated ID. While it is technically feasible to utilize the ID from the source system as the person_ id, provided it is an integer, we strongly discourage this approach due to potential privacy concerns. Instead, we recommend placing the patient ID from the source system into the 'PERSON.person_source_value' column.
gender_concept_id	integer	Yes	Yes *Sex (1.2)*	CONCEPT.concept_id	Accepted gender concepts: • Female: 8532 • Male: 8507
year_of_birth	integer	Yes	Yes *Date of Birth (1.1)*		
Month_of_birth	integer	No	Yes *Date of Birth (1.1)*		
Day_of_birth	integer	No	No		If allowed by local privacy, we advise to include day of birth. If not, we advise setting day to 15th of month. Please note studies will generally obscure age to age in years at study index date.
Birth_datetime	datetime	No	No		
race_concept_id	integer	Yes	No	CONCEPT.concept_id	Required by OMOP but if race **cannot be** provided due to national regulations or not captured, enter '0'. If can be provided, please see Accepted race concepts.
ethnicity_concept_id	integer	Yes	No	CONCEPT.concept_id	Required by OMOP but if ethnicity **cannot be** provided due to national regulations or not captured, enter '0'. If can be provided, please see Accepted ethnicity concepts.
location_id	integer	No	No	LOCATION.location_id	In some cases, it might help to add the physical location of the person, but for DigiONE purposes the location of the person will not be relevant. See LOCATION table for further guidance

Column	Type	OMOP Required	DigiONE Required	Related to	ETL guidance for DigiONE nodes
provider_id	integer	No	Yes	PROVIDER.provider_id	This field can be used to link a person to the last known main care provider or general practitioner in the PROVIDER table. See PROIVDER table for further guidance
care_site_id	integer	No	Yes	CARE_SITE.care_site_id	The care site refers to where the provider typically provides the main care. In DigiONE care_site is defined as a particular medical department in the medical center. See CARE_SITE table for further guidance
person_source_value	varchar	No	Yes *Healthcare ID (1.5)*		This field is used to link back to the person in the source data. It will not be exposed to the network, and will remain in node. Can be used for error checking of ETL logic. Use Person ID value in as it appears in the source.
gender_source_value	varchar	No	No		*Optional* – We recommend filling for quality control. This field can be used to distinguish between with biological sex data and patients with self-declared gender data if desired
gender_source_concept_id	integer	No	No	CONCEPT.concept_id	*Optional* – We recommend filling for quality control.
race_source_value	varchar	No	No		If using race or ethnicity, we recommend filling in these linkages to allow quality control.
race_source_concept_id	integer	No	No	CONCEPT.concept_id	
ethnicity_source_value	varchar	No	No		
ethnicity_source_concept_id	integer	No	No	CONCEPT.concept_id	

OBSERVATION_PERIOD

1. Introduction to OMOP Table: OBSERVATION_PERIOD

The OBSERVATION_PERIOD table tracks the time period(s) for which you believe you are observing a patient, and contains a record of the clinical events that occurred in this period. If there are no records in this table for a certain period, it indicates that no clinical events were recorded during that time, and therefore, no analyses should be conducted for that period.

Each person must have at least one observation period. The start and end of an observation period is usually based on the first and last important health events. If there is only one event, the time period is one day long. A person can have multiple observation periods (even for the same diagnosis) as long as these periods do not overlap. If there are two or more time periods that are next to each other or overlap, they should be combined into one.

Although there are no MEDOC Concepts stored in the OBSERVATION_PERIOD table, it is important to remember that the Observational Medical Outcomes Partnership (OMOP) aims to enhance the use of healthcare databases for studying the effects of medical products, making observation periods a crucial element. This table is primarily used to (i) determine if a patient had sufficient prior observation for a study, (ii) determine the follow up for any time-to-event analyses.

2. Implementing the OMOP Table: OBSERVATION_PERIOD

While not directly linked to MEDOC concepts, these periods resemble the timeframes of MEDOC items. **While in OMOP there can be multiple observation periods for a patient, in DigiONE there should only be one observation period per patient unless in rare circumstances which would be agreed at the study-level** (e.g., if a cancer patient receives treatment at a hospital and then there is no further data for years until the patient relapses this could be considered as two separate observation periods). The observation period should start at the first event date available for the individual (baring date of birth; can include non-oncology data) and end on the last event data available. The full detailed convention to determine observation can be found here.

OBSERVATION_PERIOD should only be populated once all other tables have been filled.

Column	Type	OMOP Required	DigiONE Required	Related to	ETL guidance for DigiONE nodes
observation_period_id	integer	Yes	No		Use a unique observation_period_id to record each discrete observation period. A person may have multiple discrete observational periods.
person_id	integer	Yes	No	PERSON.person_id	
observation_period_ start_date	date	Yes	No		Where the concept of observation periods does not exist in source data, the start date can be defined as the earliest event date available for the person (excluding date of birth).
observation_period_end_ date	date	Yes	No		If the concept of observation periods does not exist in source data, the end date can be considered as the latest event date available for the person.
period_type_concept_id	integer	Yes	No	CONCEPT.concept_id	Type of observation period e.g., whether the period was determined from an insurance enrolment file, EHR healthcare encounters, or other sources. See Accepted Concepts.

CONDITION_OCCURRENCE

MEDOC Concept(s) : Primary cancer diagnosis (2.1), Date of primary diagnosis (2.3), Comorbidities (for the Charlson Comorbidity Index (*derived*) (2.2)

1. Introduction to OMOP Table: CONDITION_OCCURRENCE

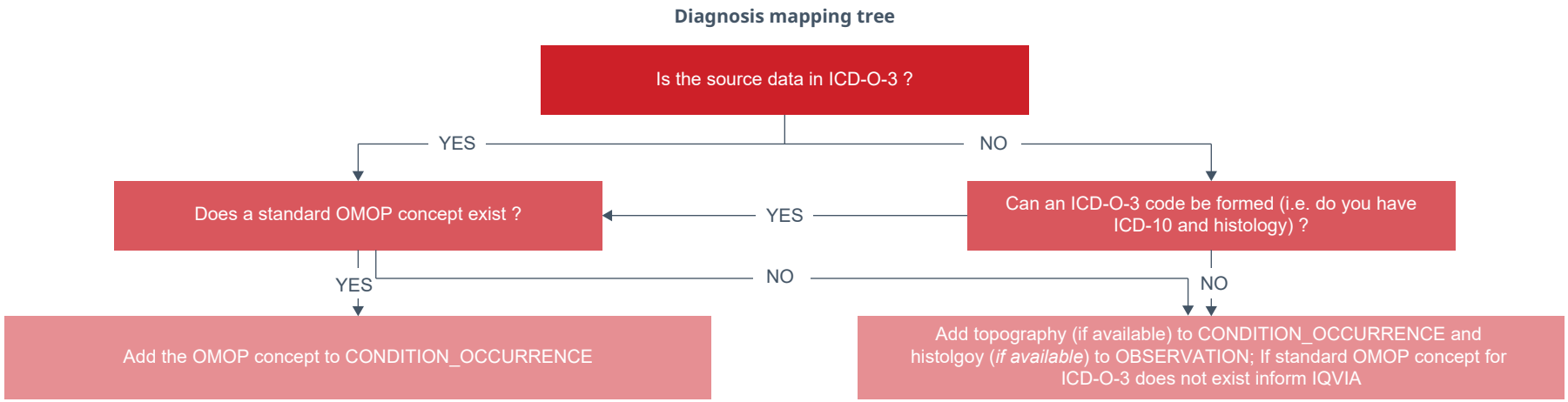
The CONDITION_OCCURRENCE table provides a record of the clinical events that suggest a person has a disease or medical condition including diagnoses, signs, symptoms and comorbidites. Most condition records are mapped from diagnostic codes (e.g. ICD10, SNOMED).

This table should not be used to record family history, historical diagnoses ('history of') and preliminary diagnoses in the process of establishing a diagnosis. Instead, these should be recorded in the OBSERVATION table.

2. MEDOC Concept Definitions

- *Primary cancer diagnosis (2.1)*: The primary diagnosis is the main condition treated or investigated during the relevant episode of healthcare. OMOP's core vocabulary for cancer diagnosis is ICD-O-3, with location (topography) and cell type (histology). The format and availability of this data can vary and different approaches should be followed depending on the scenarios outlined in the decision tree below:

Figure 5. Diagnosis mapping tree based on data availability



- *Date of primary diagnosis (2.3)*: The date of primary diagnosis of the main condition treated or investigated during the relevant episode of healthcare.
  - » Cancer diagnosis is more complex than many other diseases and often involves multi-modal investigations (imaging, biopsy, pathology etc). As a result, the diagnostic process requires careful representation in data, and we anticipate 2-3 modular implementations. A very basic implementation may just focus on a single date of diagnosis. A more complex one in a hospital with highly structured data may try and represent the broader diagnostic pathway. Implementation advice is provided in both situations so that across a network we can create like-for-like mappings
  - » Note: the original date of diagnosis must be provided, even if this occurred outside of your hospital site
- *Charlson Comorbidity Index (2.2)*: A weighted score of 17 conditions that predicts mortality risk and outcomes for patients. It was first developed in 1987 by Mary Charlson and colleagues (doi:10.1016/0021-9681(87)90171-8). It is the most widely used and validated measure of comorbidity level by researchers and helps clinicians to make informed decisions about procedures. It predicts the risk of mortality for a patient by assigning a score of 1, 2, 3, or 6, depending on the risk of dying associated with each comorbidity. Comorbidities will be limited to the 17 conditions included in CCI, for the minimum OMOP RDR.

3. Implementing OMOP Table: CONDITION_OCCURRENCE

On date of primary diagnosis (2.3): There are multiple options for date and method of diagnosis across different centres. It has been agreed by consensus in DigiONE to include the date of the pathology tissue biopsy (pathology date) and the imaging-based diagnosis (imaging date) in the OMOP RDR, where available. Some hospitals may also have access to the ENCR (European Network of Cancer Registries) date of diagnosis, which is defined by an algorithm by the relevant national cancer registries according to published priorities. Furthermore, some hospitals may have additional methods and dates for diagnosis capturing the full diagnostic pathway.

Given the availability of imaging and pathology diagnostic data, and in some cases, ENCR data, the CONDITION_OCCURRENCE table should be completed for each type of source data that is available. **Table 2** also demonstrates how the separately-filled PROCEDURE_OCCURRENCE table maps the procedure dates against the dates of diagnosis for imaging and pathology-based diagnosis i.e., method of diagnosis. Diagnosis method should be loaded into the Observation table with the following concepts where relevant:

1. 4338116 Pathology diagnosis
2. 4309119 Clinical diagnosis
3. 4101336 X-ray diagnosis
4. 4187810 Laboratory diagnosis
5. 4236136 Post-op diagnosis

Table 2. How to handle each Date of Diagnosis, including *mapping* to PROCEDURE_OCCURRENCE table

			Type of source data to be used for date of diagnosis in order of priority (based on data availability)		
OMOP Table	**OMOP Columns**	**Data description**	**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 the procedure that resulted in the primary cancer diagnosis	ENCR date	Date of pathology procedure	Date of imaging procedure or imaging report
	condition_type_ concept_id	The provenance of record, e.g., claims from a billing system	Source system of the ENCR data i.e., 32879 Registry	Source system of the pathology procedure e.g., 32835 EHR Pathology report	Source system of the imaging procedure e.g., 32841 EHR radiology Report
	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	The procedure used for the primary cancer diagnosis	N/A	Type of procedure	
	procedure_date	Date of the procedure that resulted in the primary cancer diagnosis		Date of pathology procedure or report	Date of imaging procedure or report
	procedure_type_ concept_id	The provenance of record, e.g., claims from a billing system		Source system of the pathology procedure	Source system of the imaging procedure

On Charlson Comorbidity Index (2.2) and comorbidities in general: At a minimum, we require the 17 comorbidities for the Charlson Comorbidity Index (CCI) calculation in the OMOP RDR (Refer to DigiONE data concept list file). If you capture comorbiditiy data in a structured format, we advise you to 'pump' all condition data into the CONDITION_OCCURRENCE table for future utility. If comorbidity data is captured as free text or in an unstructured format, implement your NLP solution to obtain the 17 comorbidities for CCI. In DigiONE, at the protocol-level a central script will be shared which will calculate CCI (there is no need to directly add CCI score to your OMOP RDR).

Cancer recurrence or progression: Recurrence for solid tumors, if captured in the source data, should be documented in the CONDITION_OCCURRENCE table e.g., 4201477 **Local recurrence of malignant tumor of breast**. If the appropriate concepts for the specific cancer are not available in source data, recurrence can be mapped to the generic concept_id **4097297 Recurrent tumor** and linked to the primary tumour diagnosis record using the Fact_Relationship table.

Similarly, if there is an EHR or registry source data for progression, this can be loaded into Condition_Occurrence table using concept_id **4168352 Tumor progression** and linked to the primary tumour diagnosis record using the Fact_Relationship table. Without an explicit record for tumour progression, but evidence of progression from tumour size or markers, these should be loaded into the MEASUREMENT table.

CONDTION_OCCURRENCE Table: High level overview

Column	Type	OMOP Required	DigiONE Required	Related to	ETL guidance for DigiONE nodes
condition_occurrence_ id	integer	Yes	Yes		The unique identifier for a condition record pertaining to an individual. Each occurrence of a condition within the source data necessitates assignment of this distinct key. A person might have multiple records of the identical condition during a single visit, while maintaining these duplications is acceptable, assigning distinct CONDITION_OCCURRENCE_ IDs to each is recommended.
person_id	integer	Yes	Yes	PERSON	
condition_concept_id	integer	Yes	Yes *Primary cancer diagnosis (2.1)* *Charlson comorbidity index (2.2)*	CONCEPT	The standard concept mapped from the source value which represents a condition. See Accepted Concepts. For instance, for the comorbidity “myocardial infarction” use the concept id “4329847” and for “congestive heart failure” use the concept id “319835”. If in doubt, use the concept ID with the greatest level of detail possible (e.g., 5562006 “Endometriosis of intestine” vs 433527 “Endometriosis (clinical)”)
condition_start_date	date	Yes	Yes *Date of primary diagnosis (2.3)*		See guidance above for date of primary cancer diagnosis, if multiple types of date of diagnosis are available
condition_start_ datetime	datetime	No	No		Not commonly used and not supported by OHDSI tools
condition_end_date	date	No	No		
condition_end_datetime	datetime	No	No		
condition_type_ concept_id	integer	Yes	Yes	CONCEPT	This provenance of the Condition record e.g., whether the condition was from an EHR system, registry, or other sources. See Accepted Concepts.
condition_status_ concept_id	integer	No	Yes	CONCEPT	This concept represents the type of diagnosis, including the timepoint during the visit that the diagnosis was given (admitting diagnosis, final diagnosis), the means of determining the diagnosis e.g., due to laboratory findings, if the diagnosis was exclusionary, or if it was a preliminary diagnosis, among others. See Accepted Concepts. See guidance above for date of diagnosis and linkage to method of diagnosis
stop_reason	varchar(20)	No	No		
provider_id	integer	No	No	PROVIDER	
visit_occurrence_id	integer	No	No	VISIT_ OCCURRENCE	

Column	Type	OMOP Required	DigiONE Required	Related to	ETL guidance for DigiONE nodes
visit_detail_id	integer	No	No	VISIT_DETAIL	
condition_source_value	varchar(50)	No	No		*Optional* – We recommend filling in with the source data value to allow QC
condition_source_ concept_id	integer	No	No	CONCEPT	*Optional* – Where the CONDITION_SOURCE_VALUE is coded in the source data with an OMOP supported vocabulary, we recommend listing the concept ID representing the source value here to allow QC
condition_status_ source_value	varchar(50)	No	No		*Optional* – We recommend filling in with the source data to allow QC

DEATH

MEDOC Concept(s): Date of death (5.1)

1. Introduction to the OMOP Table: DEATH

- The DEATH table contains information on the clinical event for how and when the individual died.
- Each person can have up to one death record per source system providing this data is available in the source data (e.g. condition in an administrative claim, explicit record in EHR). As such, looking across multiple systems, a patient may have multiple death records with the same data of death and different cause of death.

2. MEDOC Concept Definitions

- *Date of death (5.1)*: The specific date on which a patient died or is officially deemed to have died. It is preferable to have linkage to a death registry to capture death in any location and not limited to deaths that occur at the participating centre.

Note that the MEDOC concept "Vital status" will be derived from date of death in study-specific protocols.

3. Implementing the OMOP Table: DEATH

Column	Type	OMOP Required	DigiONE Required	Related to	ETL guidance for DigiONE nodes
person_id	integer	Yes	Yes	PERSON	
death_date	date	Yes	Yes		It is preferable to have the date of death linked to a death registry to capture death in any location and not limited to deaths that occur at the participating centre. Where the exact day or month is not known or cannot be reported, use December as the Month and the last day of the month as the day. Format: YYYY-MM-DD
death_datetime	datetime	No	No		
death_type_ concept_id	integer	No	Yes	CONCEPT	This is the provenance of the death record. See Accepted Concepts. • If death record is from national registry, use: 32879 Registry • If death record is from inpatient department in the hospital, use one of the EHR codes e.g., 32829 EHR inpatient note
cause_concept_id	integer	No	No	CONCEPT	*Optional* – We recommend filling in the standard concept representing the cause of death to allow QC
cause_source_value	varchar(50)	No	No		*Optional* – We recommend filling in the source code representing the cause of death to allow QC
Cause_source_ concept_id	integer	No	No	CONCEPT	*Optional* – If the cause of death is coded using an OMOP supported vocabulary, we recommend filling the concept id representing the source value here to allow QC

VISIT_OCCURRENCE

MEDOC Concept(s): Date of visits (5.5)

1. Introduction to the OMOP Table: VISIT_OCCURRENCE

The VISIT OCCURRENCE table captures an individual's encounters with the healthcare system. Visits are classified according to (i) whether the interaction is remote or in-person (ii) the type of medical staff seen and (iii) duration of the visit. Visit types in OMOP can be accessed here: Visit types in OMOP

Note that in MEDOC, only **cancer-related** in-person visits or telehealth consultations are required.

Note that Multidisciplinary Team Meetings (Team of doctors, nurses and other healthcare professionals from various fields meet in order to determine a patients' treatment plan) are captured in the OBSERVATION domain, with concept_id 44791544 (Multidisciplinary meeting).

Visit duration is determined by the difference between the VISIT_END_DATE and VISIT_START_DATE. Most visits have a duration of less than one day, except for inpatient visits and non-hospital institution visits. Additional details about visits, such as transfers between units during inpatient visits, can be found in the VISIT_DETAIL table.

2. MEDOC Concept Definitions

- *Date of Visits (5.5):* The specific date on which a patient attended a cancer-related in-person visit or telehealth consultation (e.g. phone consultation). This date will be used to derive the date of last visit/follow-up, or the last known date when the patient was seen or confirmed to be alive in the data. This information is critical for tracking patient follow-up, evaluating treatment outcomes, and assessing survival rates in cancer research and medical practice.

Note that "Date of last visit" will be derived by date of visits in study-specific protocols.

3. Implementing the OMOP Table: VISIT_OCCURRENCE

Column	Type	OMOP Required	DigiONE Required	Related to	ETL guidance for DigiONE nodes
visit_occurrence_id	integer	Yes	Yes		Unique visit_occurrence_id to record unique interactions between a person and the health care system. The ID links across to other OMOP event tables to link associated events with the visit. The OMOP database will automatically ensure a unique number is given to each visit_ocurrence_id
person_id	integer	Yes	Yes	PERSON	
visit_concept_id	integer	Yes	Yes	CONCEPT	The type of visit that took place e.g., "Inpatient Visit", "Outpatient Visit", "Ambulatory Visit". See Accepted Concepts. Include all type of visits (including in person or via telehealth). Cancer related visits should be flagged accordingly by filling the CARE_SITE_ID to identify visits that occurred within cancer facilities. PROVIDER_ID can also be used for tracking if provider provides cancer care.
visit_start_date	date	Yes	Yes		For inpatient visits, the start date is typically the admission date. For outpatient visits the start date and end date will be the same. In all other cases this should be the date of the patient-provider interaction. Format: YYYY-MM-DD
visit_start_datetime	datetime	No	No		
visit_end_date	date	Yes	Yes		If visit end dates are not available in the source data they should be derived in the following manner: • Inpatient Visit » Derive end date from the sudden decline in clinical events or absence of inpatient treatment. » If the person is an inpatient at the time of the data extract then set the visit_end_ date to the date of the data extract. • Non-hospital institution Visits: assume visit is for the duration of the month that it occurs in • Outpatient Visit, Emergency Room, All other visits: visit_end_date = visit_start_date Format: YYYY-MM-DD
visit_end_datetime	datetime	No	No		

Column	Type	OMOP Required	DigiONE Required	Related to	ETL guidance for DigiONE nodes
visit_type_concept_id	integer	Yes	Yes	CONCEPT	The provenance of the visit record, or where the record comes from. See Accepted Concepts. If possible, provide the concept with the highest level of detail e.g. 'EHR nursing report' (32832) vs 'EHR' (32817)
provider_id	integer	No	No	PROVIDER	If there are multiple providers associated with a visit, select the most appropriate one to include here. Additional providers can be stored in the VISIT_DETAIL table
care_site_id	integer	No	Yes	CARE_SITE	Each Visit should only have one Care Site associated with it. A care site is defined as a particular medical department in the medical center and will be required to distinguish cancer visits from other types of visits
visit_source_value	varchar(50)	No	No		
visit_source_concept_id	integer	No	No	CONCEPT	
admitted_from_concept_id	integer	No	No	CONCEPT	
admitted_from_source_value	varchar(50)	No	No		
discharged_to_concept_id	integer	No	No	CONCEPT	
discharged_to_source_value	varchar(50)	No	No		
preceding_visit_occurrence_id	integer	No	No	VISIT_OCCURRENCE	Optional: can link a visit with the visit immediately preceding it e.g., containing the primary cancer diagnosis.

PROCEDURE_OCCURRENCE

MEDOC Concept(s): Method of primary diagnosis (2.4), Radiotherapy type (4.7), Radiotherapy start date (4.8), Radiotherapy end date (4.10), Surgery type (4.11) Surgery date (4.12)

1. Introduction to OMOP Table: PROCEDURE_OCCURRENCE

The PROCEDURE_OCCURRENCE table records the activities or processes for a patient that are ordered by, or carried out by, a healthcare provider with a diagnostic or therapeutic purpose. Note that lab tests should be considered measurements rather than procedures.

When the ETL process is managing duplicate records, it should consider factors such as the same procedure, PROCEDURE_DATETIME, Visit Occurrence or Visit Detail, provider, and modifier for procedures. Source data mapped to Standard Concepts of the Procedure Domain should be recorded.

2. MEDOC Concept Definitions

- *Method of primary diagnosis (2.4):* The initial technique or approach used to identify and confirm the presence of cancer in an individual. Two options have been agreed for the method of primary diagnosis: pathology and imaging based diagnosis.
- *Radiotherapy type (4.7):* The specific approach or technique utilized in the administration of radiotherapy as a treatment for cancer. This term can be represented using a coded format, such as a procedure code, to accurately describe the method employed during the delivery of the radiotherapy treatment. It serves as an important classification system to categorize and document the various radiotherapy procedures used in cancer treatment, aiding in data analysis, research, and treatment planning.
- *Radiotherapy start date (4.8):* The specific date on which a patient begins their radiotherapy treatment for cancer. It marks the initiation of the radiotherapy procedure and is a crucial piece of information for tracking the treatment timeline, evaluating treatment outcomes, and maintaining comprehensive records of the patient's cancer therapy journey. This date is recorded to facilitate research, treatment planning, and data analysis related to radiotherapy interventions in cancer patients.
- *Radiotherapy end date (4.10):* The specific date on which the administration of radiotherapy treatment for the patient concludes, corresponding to the date of delivery of the last fraction of radiotherapy. It marks the completion of the prescribed course of radiotherapy treatment and is essential for accurately documenting the treatment duration, assessing treatment outcomes, and maintaining comprehensive records of the patient's radiotherapy journey for research, analysis, and treatment planning purposes.
- *Surgery type (4.11):* The specific classification or code used to describe the type of surgical procedure(s) performed on a patient. This classification should be represented in a coded format, with one entry per procedure. The "Surgery type" field is essential for accurately documenting the surgical interventions conducted on patients, facilitating data analysis, research, & treatment planning within the research group or healthcare institution. It enables a standardized and systematic approach to categorizing & studying various surgical procedures in the context of cancer research or medical treatment.

- *Surgery date (4.12):* The specific classification or code used to describe the type of surgical procedure(s) performed on a patient. This classification should be represented in a coded format, with one entry per procedure. The "Surgery type" field is essential for accurately documenting the surgical interventions conducted on patients, facilitating data analysis, research, and treatment planning within the research group or healthcare institution. It enables a standardized and systematic approach to categorizing and studying various surgical procedures in the context of cancer research or medical treatment.

3. Implementing OMOP Table: PROCEDURE_OCCURRENCE

On linkage to Date of Diagnosis procedures: For the dates of diagnosis in the CONDITION_OCCURRENCE that have been determined by imaging or pathology, these procedures must also be mapped and captured in the PROCEDURE_OCCURRENCE table (Table 3) and linked via the FACT_RELATIONSHIP table.

Table 3. Mapping Procedures to Date of Diagnosis

			Type of source data to be used for date of diagnosis in order of priority (based on data availability)		
OMOP Table	**OMOP Columns**	**Source data**	**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 the procedure that resulted in the primary cancer diagnosis	ENCR date	Date of pathology procedure	Date of imaging procedure or imaging report
	condition_type_concept_id	The provenance of record, e.g., claims from a billing system	Source system of the ENCR data i.e., 32879 Registry	Source system of the pathology procedure e.g., 32835 EHR Pathology report	Source system of the imaging procedure e.g., 32841 EHR radiology Report
	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	The procedure used for the primary cancer diagnosis	N/A	Type of procedure	
	procedure_date	Date of the procedure that resulted in the primary cancer diagnosis		Date of pathology procedure or report	Date of imaging procedure or report
	procedure_type_concept_id	The provenance of record, e.g., claims from a billing system		Source system of the pathology procedure	Source system of the imaging procedure

PROCEDURE_OCCURRENCE Table: High level overview

Column	Type	OMOP Required	DigiONE Required	Related to	ETL guidance for DigiONE nodes
procedure_occurrence_id	integer	Yes	Yes		The unique identifier given to each instance of a procedure occurrence in the source data. A person might have multiple records of the same procedure within the same visit.
person_id	integer	Yes	Yes	PERSON.person_id	
procedure_concept_id	integer	Yes	Yes	CONCEPT.concept_id	The standard concept mapped from the source value which represents a procedure. See Accepted Concepts.
procedure_date	date	Yes	Yes		The start date of the procedure.
procedure_datetime	datetime	No	No		
procedure_end_date	date	No	Yes		
procedure_end_datetime	datetime	No	No		
procedure_type_concept_id	integer	Yes	Yes	CONCEPT.concept_id	The provenance of the Procedure record e.g., whether the procedure was from an EHR system, insurance claim, registry, or other sources. See Accepted Concepts and the vocabulary wiki.
modifier_concept_id	integer	No	No	CONCEPT.concept_id	
quantity	integer	No	No		If left blank, a single procedure is assumed.
provider_id	integer	No	Yes	PROVIDER.provider_id	The provider associated with the Procedure
visit_occurrence_id	integer	No	No	VISIT_OCCURRENCE.visit_ occurrence_id	*Optional* – The visit during which the Procedure occurred.
visit_detail_id	integer	No	No	VISIT_DETAIL.visit_detail_id	*Optional* – The VISIT_DETAIL record during which the Procedure occurred
procedure_source_value	varchar	No	No		*Optional* – We recommend filling in with the source data value to allow QC
procedure_source_concept_id	integer	No	No	CONCEPT.concept_id	*Optional* – If the PROCEDURE_SOURCE_VALUE is coded in the source data using an OMOP supported vocabulary, we recommend filling the concept id representing the source value here to allow QC
modifier_source_value	varchar	No	No		

OBSERVATION

MEDOC Concept(s): Menopausal Status (2.8), Participation in clinical trial (4.13), Date of trial consent (4.14)

1. Introduction to OMOP Table: OBSERVATION

The OBSERVATION table is used to record clinical facts obtained through examinations, questioning, or procedures that cannot be represented by other domains.

Observations encompass a wide range of data, including social and lifestyle information, medical and family history, treatment needs, and healthcare utilization patterns. Unlike Measurements, Observations do not require standardized tests and can be stored as attribute value pairs, representing the Observation Concept and the clinical fact, which can be a Concept, numerical value, verbatim string, or datetime. It is recommended to record suggestive positive assertions as "Yes" in Observations, even though the null value is considered equivalent.

Note that Multidisciplinary Team Meetings (Team of doctors, nurses and other healthcare professionals from various fields meet in order to determine a patients' treatment plan) are captured in the OBSERVATION domain, with concept_id 44791544 (Multidisciplinary meeting).

2. MEDOC Concept Definitions:

- *Menopausal Status (2.8):* This is a cancer specific data concept only required for women with breast cancer. Menopausal status could be inferred from a structured field or notes in the EMR (patients report that their periods have ceased), blood test results (such as FSH and oestrogen), combination of treatments received and surgical history (removal of the ovaries). The PROVIDER and CARE_SITE tables should be used to populate this concept for breast cancer patients only.
- *Participation in clinical trial (4.13):* The involvement of a patient in an interventional drug trial in oncology. This concept encompasses patients who have volunteered to be part of a clinical trial, following specific protocols, and allows researchers to gather valuable data to assess the investigational drug's potential benefits and risks.
- *Date of trial consent (4.14):* The specific date on which a patient provides informed consent to participate in a clinical trial. It marks the point in time when the patient formally agrees to volunteer for the research study after receiving comprehensive information about the trial's purpose, procedures, potential risks, and benefits.

3. Implementing the OMOP Table

On menopausal status (2.8): If your centre captures menopausal status on a structured format, input it directly into the OBSERVATION table with observation_concept_id 4295261 (postmenopausal), 45757505 (perimenopausal) or 4331463 (premenopausal). If menopausal data is not available, it will be derived centrally based on the variables included in the table below. Use the provider_id to extract data from breast cancer specialists only for this MEDOC Concept.

Table 4 Data requirements for documenting menopausal status

Data required	OMOP table	Implementation guidance
Other reported status	Observation	Observation_concept_id – add entries for any of the following observations present: • Postmenopausal – 4295261 • Perimenopausal – 45757505 • Premenopausal – 4331463
Level of follicle-stimulating hormone (FSH)	Measurement	Measurement_concept_id – 4149280 (Follicle stimulating hormone measurement) Value_as_number – add the numerical value Unit_concept_id – add the unit used (*e.g., for mIU/mL use 9550 (milli-international unit per milliliter), for IU/L use 8923 (international unit per liter*))
Level of oestradiol	Measurement	Measurement_concept_id – 4292703 (Estradiol measurement) Value_as_number – add the numerical value Unit_concept_id – add the unit used (e.g., *for pg/mL use 8845 (picogram per milliliter), for picomole per liter use 8729 (picomole per liter*))
Date of birth	Person	Year_of_birth – add year of birth as integer e.g., 2000 Month_of_birth – add month of birth as integer e.g., 01
Aromatase inhibitors and LHRH agonist treatment	Drug_exposure	Drug_concept_id – add entries for any of the following drugs if present in the source data: • Anastrozole – 1348265 • Letrozole – 1315946 • Exemestane – 1398399 • Fulvestrant – 1304044 • Goserelin – 1366310 • Leuprolide – 1351541
Removal of ovaries	Procedure_ occurrence	Procedure_concept_id – add entries for any of the following procedures: • Bilateral oophorectomy – 4297990 • Right oophorectomy – 4120985 • Left oophorectomy – 4114635

On Participation in clinical trial (4.13) and Date of trial consent (4.14): When it comes to sponsored studies, often pharmaceutical clients do not want to include in a routine care study a drug as a comparator that was delivered within an oncology clinical trial. The patient does not need to be removed from the study entirely, however the date of consent provided to participate in the clinical trial can be used to exclude the subsequent treatments prescribed within 6-12 months (for example). We expect for some treatments we will not know the molecule received if it was delivered within a blinded clinical trial before data release. We noted in the protocols that 'Clinical Trial' could be entered in this situation instead. Where we find SACT entries that are unusual in OMOP and we expect the treatment was not part of routine care, we can check proximity to the date of clinical trial consent. Thus, for the purposes of DigiONE, participation in an oncology clinical trial should be recorded. To record participation in an oncology clinical trial, input into the OBSERVATION table with observation_concept_id 4323504 (Clinical Trial Participant). Input the Date of trial consent as the observation_date.

OBSERVATION Table: High level overview

Column	Type	OMOP Required	DigiONE Required	Related to	ETL guidance for DigiONE nodes
observation_id	integer	Yes	Yes		The unique identifier given to each instance of an observation record for a Person. A person might have more than one observation during the same visit.
person_id	integer	Yes	Yes	PERSON.person_id	
observation_concept_id	integer	Yes	Yes	CONCEPT.concept_id	No specified domains required, however records with concepts in the condition, procedure, drug, measurement, or device domains must go to the corresponding table. Detailed guidance on which concept_ids to use for each MEDOC concept provided in the table above.
observation_date	date	Yes	Yes		The date of the observation.
observation_datetime	datetime	No	No		
observation_type_concept_id	integer	Yes	Yes	CONCEPT.concept_id	The provenance of the Observation record e.g., whether the measurement was from an EHR system, insurance claim, registry, or other sources. See Accepted Concepts.
value_as_number	float	No	No		
value_as_string	varchar	No	No		
value_as_concept_id	integer	No	No	CONCEPT.concept_id	
qualifier_concept_id	integer	No	No	CONCEPT.concept_id	
unit_concept_id	integer	No	No	CONCEPT.concept_id	
provider_id	integer	No	No	PROVIDER.provider_id	
visit_occurrence_id	integer	No	No	VISIT_OCCURRENCE. visit_occurrence_id	
visit_detail_id	integer	No	No	VISIT_DETAIL.visit_ detail_id	
observation_source_value	varchar	No	No		*Optional* – We recommend filling in with the source data value to allow QC

Column	Type	OMOP Required	DigiONE Required	Related to	ETL guidance for DigiONE nodes
observation_source_concept_id	integer	No	No	CONCEPT.concept_id	*Optional* – Where OBSERVATION_SOURCE_VALUE is coded in the source data in an OMOP supported vocabulary, we recommend filling the concept id representing the source value here to allow QC
unit_source_value	varchar	No	No		*Optional* – We recommend filling in with the source data value to allow QC.
qualifier_source_value	varchar	No	No		*Optional* – We recommend filling in with the source data value to allow QC
value_source_value	varchar	No	No		*Optional* – We recommend filling in with the source data value to allow QC
observation_event_id	integer	No	No		If the observation record is related to another record in the database use this field is the primary key of the linked record.
obs_event_field_concept_id	integer	No	No	CONCEPT.concept_id	If the Observation record is related to another record in the database, this field is the CONCEPT_ID that identifies which table the primary key of the linked record came from.

MEASUREMENT

MEDOC Concept(s): Weight (1.3), Height (1.4), Performance Status (2.5), Disease Stage (2.6), Biomarker name (3.1), Biomarker measure (3.2), Biomarker sample ID (3.3), Radiotherapy dose (4.9), Metastasis presence/absence (5.3), Metastasis location (5.4), Extent of debulking (5.7)

1. Introduction to OMOP Table: MEASUREMENT

The MEASUREMENT table stores structured results from examinations and tests including laboratory tests, vital signs, and pathology findings. Measurements are represented as attribute value pairs (e.g., attribute as the measurement concept and value as the result). It is important to note that measurements require a standardised test or activity to generate a quantitative or qualitative result, and if no result is available it indicates that the test was conducted but the result was not captured.

When working with measurements, focus on lab tests and be aware of operator_concept_ids indicating comparisons (<, >, =, etc.). Only include records in the MEASUREMENT table where the source value maps to a concept in the measurement domain. Although a result is not always mandatory, recording the fact that a measurement was performed can still provide valuable information for certain use cases. Additionally, specific measurement concepts may include the result in the test itself, represented through concept mappings and relationships in the CONCEPT_RELATIONSHIP table. Where required, units for measurements should be captured in unit_concept_id.

2. MEDOC Concept Definitions

- *Weight (1.3):* The measurement of the body weight.
- *Height (1.4):* The height of a person refers to the linear measurement of an individual's stature. It is often recorded along with weight to calculate body mass index (BMI), which helps assess a person's weight status relative to their height.
- *Performance status (2.5):* An evaluation of patient situation. It can be Karnofsky performance status or ECOG performance status (also called WHO or Zubrod performance status, description below):
  - » 0 = Fully active, able to carry on all pre-disease performance without restriction
  - » 1 = Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature, e.g., light house work, office work
  - » 2 = Ambulatory and capable of all selfcare but unable to carry out any work activities. Up and about more than 50% of waking hours
  - » 3 = Capable of only limited selfcare, confined to bed or chair more than 50% of waking hours
  - » 4 = Completely disabled. Cannot carry on any selfcare. Totally confined to bed or chair
  - » 5 = Dead

- *Disease stage (2.6):* The classification or categorization of cancer based on its extent of spread in the body at a specific point in time. The stage of cancer is typically described using a numerical system (such as stages I to IV) or a combination of letters and numbers (such as TNM staging system, Gleason score, FIGO score).
- *Biomarker name (3.1):* The quantification or assessment of a specific biomarker in a biological sample at a particular point in time.
- *Biomarker measure (3.2)*: The measurement of a biomarker involves collecting a biological sample, such as blood, urine, tissue, or saliva, and analysing it using various laboratory techniques. The result of the biomarker measurement indicates the level, concentration, or presence of the biomarker in the sample, which can be indicative of a particular condition, disease progression, or treatment effect.
- *Biomarker sample ID (3.3):* The sample of the biomarker measurement.
- *Radiotherapy dose (4.9):* The amount of radiation delivered to a specific area or target within the body as part of a radiotherapy treatment. It can be measured in Gray (Gy) and represents either the prescribed total radiation dose for the entire course of treatment or the actual dose delivered to the patient. This information is crucial for evaluating treatment effectiveness, monitoring patient responses, and conducting research in the field of cancer therapy [Note on EQD2 as the dose delivered in 2Gy fractions that is biologically equivalent to a total dose]
- *Metastasis presence/absence (5.3):* The indication of whether metastatic disease is observed in a patient's cancer diagnosis or not. Metastasis involves the spread of cancer cells from the primary tumour to other parts of the body, forming distant growths in different organs. Patients can present with metastatic disease at the time of diagnosis or at a later stage. This concept is essential for accurately characterizing the stage and severity of cancer, guiding treatment decisions, and monitoring disease progression in cancer research and clinical practice.
- *Metastasis location (5.4):* The specific site or organ in the body where cancer has spread from its original or primary site. It denotes the secondary locations where cancer cells have formed distant growths, resulting from the migration of malignant cells from the primary tumour. Recording the "Metastasis location" is crucial for understanding the extent of cancer spread, determining the stage of the disease, and tailoring appropriate treatment strategies in cancer research and clinical management*
- *Extent of debulking (5.7):* This is a cancer specific data concept only required for women with a gynaecological cancer. "Extent of debulking" refers to the outcome of debulking surgery such as complete resection, optimal, and sub-optimal. The measure is largest diameter residual disease in centimetres.

*Note we are aware that OMOP cannot currently handle oligometastases – we are working with the OHDSI community to ensure this is added and it will be reflected in a subsequent release of the implementation guide.

3. Implementing OMOP Table: MEASUREMENT

On Disease stage: TNM and FIGO (in ovarian cancer) should be loaded into the measurement table as indicator of the disease stage. Cancer staging will be derived from TNM and FIGO scores.

On Metastasis presence/absence and location: The TNM system is a widely accepted method for staging cancer, assessing factors like tumour size, the spread of cancer to lymph nodes, and the extent of metastasis. In this system, M0 denotes the absence of metastasis, whereas M1 signifies that cancer has spread to other parts of the body. This structured framework of TNM provides a clear way to determine the presence or absence of metastasis. In the context of data capture, when M1 is recorded, it should be interpreted as indicating the presence of metastasis in the MEASUREMENT table. Conversely, when metastasis is absent, this field will remain blank.

In some cases, hospitals may enter TNM data into local or national cancer registries or incorporate metastasis information into tumour registration databases that can be integrated into the OMOP system. Additionally, metastasis data may also exist in unstructured, free-text formats within pathology reports, multidisciplinary team (MDT) letters, radiology reports, or clinical notes. To address this variability, it is essential to implement an NLP solution to effectively capture the presence of metastasis.

In certain instances, the presence of metastasis is recorded alongside the location of its spread. When this location data is available, it is captured in the Cancer Modifiers in OMOP, which provides a high level of detail for the measurement_concept_id. For example, it might specify “Metastasis to brain.”

Emergent metastases i.e., metastases that are not present at diagnosis should also be identified in the source data. This is relevant in some cancers e.g., approximately 20% of NSCLC are not metastatic at first presentation. TNM stage may miss patients who develop metastases later in their disease course; the methods of detection and source of emergent metastases may vary across the network but should be normalised by capture in the MEASUREMENT table. Some centres may capture these in a coded system, while others will have to curate from unstructured data, proxies from treatment information or multidisciplinary meetings.

On Extent of debulking: The concept of debulking is a significant consideration within the MEDOC 1.0 framework, primarily focusing on gynaecological cancer. It pertains to the surgical outcome of debulking procedures, with outcomes falling into distinct categories. These categories encompass complete resection (indicating the absence of any residual disease), optimal resection, and suboptimal resection. The criteria for optimal versus suboptimal outcomes are determined based on the largest diameter of residual disease, measured in centimetres.

In addition to debulking outcomes, resection margins are another crucial aspect of this framework. They are categorized as R0 (denoting clear margins with no disease), R1 (indicating the presence of microscopic residual disease), and R2 (indicating the presence of macroscopic disease). While the extent of debulking primarily relates to gynaecological cancer, the classification of resection margins holds broader applicability across various cancer types, including gynaecological cancer, and is more typically a structured concept. Furthermore, the R2 category can be further subdivided based on measurements by diameter in centimetres. For instance, it can be divided into categories such as < 1cm, 1-2cm, and >2cm. This subdivision is particularly useful to meet the specific surgical requirements for gynaecological debulking procedures. It’s worth noting that information regarding resection margins and the extent of debulking can be recorded in different formats. These details may be found

in free-text sections of surgery and pathology reports or within structured Electronic Health Records (EHRs). We therefore recommend using resection margins as a structured concept from source data, while extent of debulking measurements can be captured in value_as_number in the same entry. Use the provider_id to extract data from gynaecological specialists only for this MEDOC concept.

Progression: if there is evidence of tumour progression by measurable markers e.g., tumour size or PSA levels increasing, these measurements can be loaded into the MEASUREMENT table and use the measurement_event_id and meas_event_field_concept_id columns to link to the primary cancer diagnosis in the CONDITION_ OCCURENCE table.

The tables below provide guidance on how to fill the key fields for weight, height, performance status, disease stage, metastasis presence/absence, metastasis location and extent of debulking in the MEASUREMENT table.

For *weight, height, performance status, disease stage*:

OMOP Column	*Weight*	*Height*	*Performance status*	*Disease stage*
measurement_concept_id	Please distinguish between measured and self-reported weight: • 4099154 (Body weight) when weight is a measurement. • 37205098 (self-reported body weight)	Use 607590 (Body height)	Use the Athena code corresponding to your performance status: • 36305384 (ECOG Performance Status score) • 36303287 (Karnofsky Performance Status score)	Please use TNM or FIGO classification. TNM is divided in separate measurement tables for T, N & M. Use the measurement_ concept_id corresponding to the source value. See a pre-filtered list of CONCEPT_ID here; where method is available (e.g., pathological, clinical) ensure a code specifying this is used e.g. 1633693 (AJCC/UICC pathological T1b Category)
value_as_number	E.g, 75.5	E.g. 182	NULL	NULL
value_as_concept_id	NULL	NULL	E.g. 42530851	NULL
unit_concept_id	Use the concept id required for your units. Please use kilograms or any of its fractional units. E.g. 9529 (kilogram)	Use the concept id required for your units. Please use centimetres or any of its fractional units. E.g. 8582 (centimetre)	NULL	NULL
measurement_source_value	'Weight'	'Height'	E.g. 'ECOG 3'	E.g. 'pT1b'. Use abbreviated notation.

For metastasis presence/absence, metastasis location, extent of debulking:

OMOP Column	*Metastasis presence/absence*	*Metastasis location*	*Extent of debulking*
measurement_concept_id	E.g., 36768130 (Generalized metastases)	E.g.,36768862 (Metastasis to brain)	1634643 R0: No residual tumour 1633801 R1: Microscopic residual tumour 1634484 R2: Macroscopic residual tumour
value_as_concept_ID	E.g., 373066001 (Yes), 4188540 (No)	NULL	Enter the largest diameter of residual disease
unit_concept_id	NULL	NULL	cm
measurement_source_value	'Metastasis presence' or 'Yes' TNM = M1	NULL	E.g., 'Residual disease'

On Biomarkers: Biomarkers are genes, proteins, or other molecules that are found in tumour cells or body fluids providing information about cancer (from presence to type or stage) which can also be used to assess treatment effectiveness and guide treatment decisions. Since biomarker information is captured by measurements e.g., PSA levels, platelet count in blood creatine or even gene amplification, we will use the OMOP MEASUREMENT table. In this iteration of the OMOP RDR, map genetic variations that are tested and typically captured in LOINC. Any key mutations not included in LOINC can be requested to be included in the next iteration of OMOP. In the case of genomic panels, we can include information about the genomic test used in the PROCEDURE_OCCURRENCE table (procedure_type_concept_id). OMOP is evolving and further iterations should enable us to accommodate additional metadata information.

The recommended OMOP mapping for key biomarkers of interest, categorised by Genomic-led testing and Biochemistry-led testing are listed in the DigiONE Data Concepts List file. Note that some biochemistry markers will be essential for calculating other fields in the OMOP database e.g., serum creatinine or creatinine clearance as a marker for kidney function to inform regimen dosing.

Biomarker examples (not all OMOP columns are shown):

OMOP Column	ALK Test *(e.g., 25 U/L, interpreation = normal)*	ALK Test *(e.g., Negative by FISH)*	HER2 *(e.g., IHC, 2+/equivocal)*	HER2 *(e.g., FISH, Positive)*
measurement_concept_id	3006923 ('Alanine aminotransferase [Enzymatic activity/volume] in Serum or Plasma')	21492143 ('ALK gene rearrangements [Presence] in Blood or Tissue by FISH')	4161390 ('Human epidermal growth factor receptor 2 gene detection by immunohistochemistry')	1617486 ('ERBB2 gene duplication in Tumor by FISH')
value_as_concept_id	4069590 ('Normal')	9189 ('Negative')	45881916 ('2+') or 4172976 ('Equivocal')	9191 ('Positive')
value_as_number	25	NULL	NULL	NULL
operator_concept_id	4172703 ('=')	NULL	NULL	NULL
unit_concept_id	8645 (Unit per liter')	NULL	NULL	NULL

In some clinical settings, borderline results may require further validation, for instance when assessing *ERBB2* amplifications, ICH might yield a 2+ value which would be considered as equivocal and would require further validation by FISH (in this case the cut-off value is a ratio >2.0). If we can capture raw values reported by pathologists (value_source_value) we can come back to equivocal results as new scenarios become available for HER2-low cases. HER2 low is defined as immunohistochemical (IHC) score of 1+ or IHC 2+ with non-amplified/negative fluorescence in-situ hybridisation (FISH). These two testing events can be linked in the MEASUREMENT table by measurement_event_id and meas_event_field_concept_id.

MEASUREMENT Table: High level Overview

Column	Type	OMOP Required	DigiONE Required	Related to	ETL guidance for DigiONE nodes
measurement_id	integer	Yes	Yes		The unique identifier given to each Measurement record for a Person. It is possible for a person to have multiple records of the same measurement as part of the same visit.
person_id	integer	Yes	Yes	PERSON.person_id	
measurement_concept_id	integer	Yes	Yes	CONCEPT.concept_id	The CONCEPT_ID that the MEASUREMENT_SOURCE_CONCEPT_ID maps to.
measurement_date	date	Yes	Yes		The date of the measurement
measurement_datetime	datetime	No	No		
measurement_time	varchar	No	No		
measurement_type_concept_id	integer	Yes	Yes	CONCEPT.concept_id	The provenance of the Measurement record e.g., whether the measurement was from an EHR system, insurance claim, registry, or other sources. See Accepted Concepts and the vocabulary wiki.
operator_concept_id	integer	No	No	CONCEPT.concept_id	Concept 4172703 for '=' is identical to omission of a OPERATOR_CONCEPT_ ID value. See Accepted Concepts.
value_as_number	float	No	Yes		The numerical value of the Measurement result, if available.
value_as_concept_id	integer	No	No	CONCEPT.concept_id	Categorial result for measurements are captured and mapped to standard concepts in the "Meas Value" domain. This field may be used to capture information on HER2Low status
unit_concept_id	integer	No	No	CONCEPT.concept_id	There are no standard units recommended for individual measurements, please choose the most plausible unit for each measurement
range_low	float	No	No		

Column	Type	OMOP Required	DigiONE Required	Related to	ETL guidance for DigiONE nodes
range_high	float	No	No		
provider_id	integer	No	No	PROVIDER.provider_id	The provider associated with measurement record, e.g. the provider who ordered the test or the provider who recorded the result.
visit_occurrence_id	integer	No	No	VISIT_OCCURRENCE. visit_occurrence_id	*Optional* – The visit during which the Measurement occurred.
visit_detail_id	integer	No	No	VISIT_DETAIL.visit_ detail_id	*Optional* – The VISIT_DETAIL record during which the Measurement occurred
measurement_source_value	varchar	No	No		*Optional* – We recommend filling in with the source data value to allow QC
measurement_source_concept_id	integer	No	No	CONCEPT.concept_id	*Optional* – Where MEASUREMENT_SOURCE_VALUE is coded in the source data using an OMOP supported vocabulary, we recommend filling the concept id representing the source value here to allow QC. If source data is not an OMOP support code, this will be 0
unit_source_value	varchar	No	No		*Optional* – We recommend filling in with the source data value to allow QC
unit_source_concept_id	integer	No	No	CONCEPT.concept_id	*Optional* – Where UNIT_SOURCE_VALUE is coded in the source data using an OMOP supported vocabulary, we recommend filling the concept id representing the source value here to allow QC
value_source_value	varchar	No	No		*Optional* – We recommend filling in with the source data value to allow QC
measurement_event_id	integer	No	No		If the Measurement record is related to another record in the database, this field is the primary key of the linked record.
meas_event_field_concept_id	integer	No	No	CONCEPT.concept_id	Where the Measurement record is related to another record in the database, this field is the CONCEPT_ID that identifies which table the primary key of the linked record originated from.

DRUG_EXPOSURE, EPISODE & DRUG_STRENGTH

MEDOC Concept(s): Anti-cancer treatment name (4.2), Molecule generic name (4.3), Start date for drug treatment (4.4), Treatment dose (4.5), End date for drug treatment (4.6)

1. MEDOC Concept Definitions

- *Anti-cancer treatment name (4.2):* A prescribed systematic form of treatment such as chemotherapy or a combination of chemotherapy and immunotherapy, and supportive therapy such as bisphosphonates. It encompasses a systemic anti-cancer treatment (SACT) regimen, which consists of one or more cycles of anti-cancer drugs and different schedules Additionally, it may include information regarding the treatment intent, such as adjuvant or neoadjuvant, serving as a new field to specify the therapeutic purpose of the treatment.*
- *Molecule generic name (4.3):* The molecule generic name refers to the name of a specific anti-cancer drug or supportive therapy used within a systemic anti-cancer treatment (SACT) regimen. It can be represented either as free text or using a coded format. In the case of combination therapies within the SACT regimen, multiple entries may need to be created, each corresponding to the specific molecule or drug utilized in the treatment.
- *Start date for drug treatment (4.4):* The start date for drug treatment refers to the specific date on which a patient receives the first dose of the systemic anti-cancer treatment (SACT) regimen. It marks the initiation of the anti-cancer therapy, and it can be recorded as either the date the SACT regimen is applied or prepared for administration, or the actual date when the patient begins receiving the anti-cancer treatment.
- *Treatment dose (4.5):* The specific amount of an anti-cancer drug or therapeutic substance administered to a patient as part of their systemic anti-cancer treatment (SACT) regimen. This dose is carefully prescribed and calculated based on the patient's medical condition, drug characteristics, and treatment goals to achieve optimal therapeutic effects while minimizing potential side effects. It is typically measured in milligrams (mg) per square meter of body surface area and is given in a single application during the course of anti-cancer treatment.
- *End date for drug treatment (4.6):* The specific date on which a patient completes their prescribed course of systemic anti-cancer treatment (SACT) regimen, marking the conclusion or discontinuation of the administration of anti-cancer drugs or therapeutic substances. This date is essential for tracking the treatment duration, assessing treatment outcomes, including response to therapy and potential adverse effects, and ensuring comprehensive and accurate records of the patient's treatment journey.

*Note: treatment intent information is recorded in the observation table

Data for individual treatments will be captured in OMOP RDR per the guidance in this section. Treatment regimens can subsequently be calculated with SQL statements. Data on regimens does not need to be captured directly into OMOP.

2. OMOP Tables: DRUG_EXPOSURE and EPISODE

For MEDOC items related to **treatment name, start and end dates,** entries need to be made in two distinct OMOP tables: DRUG_EXPOSURE and EPISODE.

2.1 Introduction to OMOP Table: DRUG_EXPOSURE

The DRUG_EXPOSURE table stores records of drug exposure, including medicines, vaccines, and biologic therapies, but excludes radiological devices. The purpose of the table is to indicate the exposure to a specific drug active ingredient, and it includes information about prescriptions, dispensed drugs, and drugs administered by a provider.

DRUG_EXPOSURE Table: High level overview

Column	Type	OMOP & DigiONE Required	Related to	ETL guidance for DigiONE nodes
drug_exposure_id	integer	Yes		The unique identifier given to each instance of a drug dispensing or administration.
person_id	integer	Yes	PERSON.person_id	
drug_concept_id	integer	Yes	CONCEPT.concept_ id	The CONCEPT_ID that the drug maps to. Note: If only the drug class is known, the DRUG_CONCEPT_ID field should contain 0. See Accepted Concepts.
drug_exposure_start_date	date	Yes		Use this date to determine the start date of the drug record
drug_exposure_start_ datetime	datetime	No		
drug_exposure_end_date	date	Yes		In the event that treatment end dates are not available, the planned treatment/number of cycles could be used as a proxy (e.g., if a drug in X regimen is usually given for 3 months, assume the end date is 3 months from the start date). If the planned treatment end date is in the future at the time of your data export, use the date of export as the end date
drug_exposure_end_ datetime	datetime	No		
verbatim_end_date	date	No		
drug_type_concept_id	integer	Yes	CONCEPT.concept_ id	The provenance of the record e.g. from a prescription. See Accepted Concepts and vocabulary wiki. There may be variability in data availability of planned vs. delivered drugs. Use a hierarchical approach i.e., use delivered record if available e.g., 32825 EHR dispensing record; and planned if not e.g., 32837 EHR planned dispensing record

Column	Type	OMOP & DigiONE Required	Related to	ETL guidance for DigiONE nodes
stop_reason	varchar	No		
refills	integer	No		
quantity	float	No		
days_supply	integer	No		
sig	varchar	No		
route_concept_id	integer	No	CONCEPT.concept_ id	
lot_number	varchar	No		
provider_id	integer	No	PROVIDER.provider_ id	
visit_occurrence_id	integer	No	VISIT_OCCURRENCE. visit_occurrence_id	
visit_detail_id	integer	No	VISIT_DETAIL.visit_ detail_id	
drug_source_value	varchar	No		*Optional* – We recommend filling in with the source data value to allow QC
drug_source_concept_id	integer	No	CONCEPT.concept_ id	*Optional* – Where the DRUG_SOURCE_VALUE is coded in the source data in an OMOP supported vocabulary, we recommend filling the concept id representing the source value here to allow QC
route_source_value	varchar	No		
dose_unit_source_value	varchar	No		

2.2 Introduction to OMOP Table: EPISODE

The EPISODE table consolidates individual clinical events (e.g. VISIT_OCCURRENCE, DRUG_EXPOSURE) into groupings that represent disease phases, outcomes, and treatments.

The EPISODE table is present in the OMOP CDM but it is undergoing updates to improve useability. **In DigiONE we recommend not filling this table in at the time of your OMOP ETL, and instead populate this at the study level based on the study-specific guidance provided.**

EPISODE Table: High level overview

Column	Type	OMOP & DigiONE Required	Related to	ETL guidance for DigiONE nodes
episode_id	integer	Yes		Unique identifier for each Episode
person_id	integer	Yes	PERSON.person_id	
episode_concept_id	integer	Yes	CONCEPT.concept_id	The CONECPT_ID that represents the kind abstraction related to the disease phase, outcome or treatment. See Accepted Concepts
episode_start_date	date	Yes		Use this date to determine the start date of the Episode
episode_start_datetime	datetime	No		
episode_end_date	date	Yes		
episode_end_datetime	datetime	No		
episode_parent_id	integer	No		
episode_number	integer	No		
episode_object_concept_id	integer	Yes	CONCEPT.concept_id	
episode_type_concept_id	integer	Yes	CONCEPT.concept_id	The provenance of the Episode record e.g., whether the episode was from an EHR system, insurance claim, registry, or other sources. See Accepted Concepts and vocabulary wiki.
episode_source_value	varchar	No		*Optional* – We recommend filling in with the source code data value to allow QC
episode_source_concept_id	integer	No	CONCEPT.concept_id	

Example of treatment represented in the DRUG_EXPOSURE and EPSIODE table:

Source data example:	OMOP data example
01/01/2010	**DRUG_EXPOSURE:** drug_exposure_start_date = ‚2010-06-01‘ drug_exposure_start_datetime = ‚2010-06-01 00:00:0000‘ or NULL
	EPISODE: episode_start_date = ‚2010-06-01‘ episode_start_datetime = ‚2010-06-01 00:00:0000‘ or NULL

Source data example:	OMOP data example
Crizotinib	**DRUG_EXPOSURE:** drug_concept_id = 40242675 [Crizotinib] drug_source_concept_id = 0 drug_source_value = ‘Crizotinib’
	EPISODE: episode_object_concept_id = 35806424 [Crizotinib monotherapy]

Illustrative example of linking drug_exposure events in an episode event:

EPISODE

Field	Content
episode_id	9900851
person_id	John Smith
episode_concept_id	Treatment Cycle
episode_start_datetime	August 1, 1996
episode_end_datetime	August 28, 1996
episode_parent_id	9900850
episode_number	1
episode_object_concept_id	Chemotherapy Treatment
episode_type_concept_id	EMR
episode_source_value	PACLITAXEL + CARBOPLATIN
episode_source_concept_id	

EPISODE_EVENT

Field	Content		
episode_id	9900851	9900851	9900851
event_id	9900145	9900146	9900147
event_domain_id	drug_exposure	drug_exposure	drug_exposure

DRUG_EXPOSURE

Field	Content
drug_exposure_id	9900145
person_id	John Smith
drug_concept_id	Cyclophosphamide
drug_exposure_start_datetime	August 1, 1996
drug_exposure_end_datetime	August 1, 1996
drug_exposure_type_concept_id	EMR
drug_exposure_source_value	Cyclophophamide 1000 MG Injection

DRUG_EXPOSURE

Field	Content
drug_exposure_id	9900146
person_id	John Smith
drug_concept_id	Doxorubicin hydrochloride
drug_exposure_start_datetime	August 4, 1996
drug_exposure_end_datetime	August 4, 1996
drug_exposure_type_concept_id	EMR
drug_exposure_source_value	Doxorubicin Hydrochloride 50 MG Injection

DRUG_EXPOSURE

Field	Content
drug_exposure_id	9900147
person_id	John Smith
drug_concept_id	Dexamethasone acetate
drug_exposure_start_datetime	August 7, 1996
drug_exposure_end_datetime	August 7, 1996
drug_exposure_type_concept_id	EMR
drug_exposure_source_value	Dexamethasone acetate 8 MG/ML Injectable

Source: IQVIA OMOP team

2.3 Introduction to OMOP Table: DRUG_STRENGTH

The DRUG_STRENGTH table captures information on the amount of a specific drug ingredient received by a person.

Ingredient strength information is provided either as absolute amount (usually for solid formulations) or as concentration (usually for liquid formulations):

- If the absolute amount is provided (for example, ‚Acetaminophen 5 MG Tablet') the amount_value and amount_unit_concept_id are used to define this content (in this case 5 and 'MG').
- If the concentration is provided (for example ‚Acetaminophen 48 MG/ML Oral Solution') the numerator_value in combination with the numerator_unit_concept_id and denominator_unit_concept_id are used to define this content (in this case 48, 'MG' and 'ML').

DRUG_STRENGTH Table: High level overview

Column	Data Type	OMOP Required	DigiONE Required	Related to	ETL Guidance for DigiONE nodes
drug_concept_id	integer	Yes	Yes	CONCEPT.concept_id	Concept ID for the Branded Drug
ingredient_concept_id	integer	Yes	Yes	CONCEPT.concept_id	Concept ID for the active ingredient contained within the drug product. Note that combination drugs will have one record per active ingredient in this table. This table is of particular importance when dealing with biosimilar drugs
amount_value	float	No	Yes		The amount of active ingredient contained within the drug product.
amount_unit_concept_id	integer	No	Yes	CONCEPT.concept_id	Concept ID for the unit of measure for the amount_value.
numerator_value	float	No	Yes		Concentration of active ingredient in the drug product.
numerator_unit_concept_id	integer	No	Yes	CONCEPT.concept_id	Concept ID for the Unit of measure for numerator_value
denominator_value	float	No	No		The amount of total liquid.
denominator_unit_concept_id	integer	No	Yes	CONCEPT.concept_id	Concept ID for the unit of denominator_value
box_size	integer	No	No		
valid_start_date	date	Yes	Yes		Date when the specific concept ID was first recorded. By default this is 1-Jan-1970.
valid_end_date	date	Yes	Yes		Date when then Concept became invalid.
invalid_reason	varchar	No	No		

LOCATION

1. Introduction to OMOP Table: LOCATION

The LOCATION table records the location or address of persons and care sites i.e. an address book for persons and care sites

2. MEDOC Concept Definitions

While not directly linked to MEDOC concepts, it has the potential to enrich care site information (e.g., for de-duplication of patients treated at two DigiONE hospitals). Patient addresses are not required for DigiONE, and we strongly discourage the extraction, transformation, and loading (ETL) of address data associated with individuals. View this table as a "nice-to-have", if you want to add information on care sites, departments, etc. Please be aware that the current implementation of the LOCATION table is US centric. Until a major release to correct this, certain fields can be used to represent different international values.

3. Implementing the OMOP Table: LOCATION

Column	Type	OMOP Required	DigiONE Required	Related to	ETL guidance for DigiONE nodes
location_id	integer	Yes	No		Unique identifier for each location.
address_1	varchar	No	No		
address_2	varchar	No	No		
city	varchar	No	No		
state	varchar	No	No		
zip	varchar	No	No		
country	varchar	No	No		
location_source_value	varchar	No	No		
country_concept_id	integer	No	No	CONCEPT.concept_id	The concept ID representing the country. Here are the Accepted Concepts.
country_source_value	varchar	No	No		
latitude	Float	No	No		
longitude	Float	No	No		

CARE_SITE

1. Introduction to OMOP Table: CARE_SITE

The CARE_SITE table compiles a comprehensive record of distinct institutional units where healthcare services are administered, encompassing spaces like offices, wards, hospitals, and clinics. The intricacies of the relationships between Care Sites can be established within the FACT_RELATIONSHIP table. The CARE_SITE table will give us granularity to handle indication-specific concepts (e.g. extent of debulking, menopausal status).

2. MEDOC Concept Definitions

Care site data doesn't form a direct connection with any MEDOC data concept. It does provide, however, valuable information on departments. This can help to achieve two objectives: effective benchmarking and easy creation of outcome measures. Effective benchmarking will help compare care providers across different types of cancer care settings, such as specialized cancer centres and general hospitals. To do this, it is essential to avoid confounding factors by differentiating the visit frequencies between these settings. Easy creation of outcome measures means generating useful metrics such as visit counts and durations.

3. Implementing the OMOP Table: CARE_SITE

Column	Type	OMOP Required	DigiONE Required	Related to	ETL guidance for DigiONE nodes
care_site_id	integer	Yes	Yes		Unique identifier for each care site.
care_site_name	varchar	No	Yes		Add the centre ID assigned to your centre by DigiONE e.g. DE01
place_of_service_concept_id	integer	No	Yes	CONCEPT.concept_id	This concept represents the setting of care. Provide as much granularity as possible. See Accepted Concepts.
location_id	integer	No	No		
care_site_source_value	varchar	No	No		
place_of_service_source_ value	varchar	No	No		

PROVIDER

1. Introduction to OMOP Table: PROVIDER

The PROVIDER table comprises an exclusive list of uniquely identified individual healthcare providers (e.g. phyiscians, nurses, pharmacists).

2. MEDOC Concept Definitions

Provider data establishes no direct link with any MEDOC data concept. Nevertheless, it holds potential for two key goals: enabling effective benchmarking and facilitating the creation of outcome measures. Effective benchmarking is pivotal for comparing care providers within diverse cancer care settings, spanning specialized cancer centers to general hospitals. This necessitates meticulous consideration of confounding factors, chiefly differing visit frequencies across these settings. Furthermore, provider data streamlines the generation of useful outcome metrics, like visit counts and durations, simplifying the process of creating insightful performance indicators. It is important to note that data on individual providers will not be necessary for DigiONE. Only data on institutional providers and data on specialty are relevant.

3. Implementing the OMOP Table: PROVIDER

Column	Type	OMOP Required	DigiONE Required	Related to	ETL guidance for DigiONE nodes
provider_id	integer	Yes	Yes		Unique ID for each provider.
provider_name	varchar	No	No		
npi	varchar	No	No		
dea	varchar	No	No		
specialty_concept_id	integer	No	No	CONCEPT.concept_id	See Accepted Concepts.
care_site_id	integer	No	No	CARE_SITE.care_site_id	The CARE_SITE that the provider primarily practices in.
year_of_birth	integer	No	No		
gender_concept_id	integer	No	No	CONCEPT.concept_id	
provider_source_value	varchar	No	No		*Optional* – We recommend filling in with the source data value to allow QC.
specialty_source_value	varchar	No	No		*Optional* – We recommend filling in with the source data value to allow QC.
specialty_source_concept_id	integer	No	No	CONCEPT.concept_id	
gender_source_value	varchar	No	No		
gender_source_concept_id	integer	No	No	CONCEPT.concept_id	

Additional Information

MEDOC Data Dictionary

Table 5. MEDOC Data Dictionary V1.0

	MEDOC ID	MEDOC Data Concept	Definition summary
Demographics	1.1	Date of birth (month)	The date on which a person was born or is officially deemed to have been born. If the exact date is not accessible, then use month & year of birth or if not available then use the year of birth.
	1.2	Sex	In most countries this is the biological sex at birth. If not available (in some countries), then use a person's gender as self-declared (or inferred by observation for those unable to declare their sex).
	1.3	Weight (with timestamp)	The measurement of the body weight.
	1.4	Height	The height of a person refers to the linear measurement of an individual's stature (in metres). It is often recorded along with weight to calculate body mass index (BMI) or body surface areas (BSA).
	1.5	Healthcare ID (or other unique identifier)	A local patient identifier.
	1.6	Legal basis for data processing	The lawful justification or legal grounds under which personal data of an individual can be processed by an organization or entity.
Clinical phenotype	2.1	Primary cancer diagnosis and comorbidities, typically in International Classification of Disease standards such as ICD10, ICD9 or ICD-O-3	The primary diagnosis is the main condition treated or investigated during the relevant episode of healthcare.
	2.2	Charlson comorbidity index x (*derived from 17 comorbidities in 2.1*)	A weighted score of 17 conditions that predicts mortality risk and outcomes for patients. It was first developed in 1987 by Mary Charlson and colleagues (doi:10.1016/0021-9681(87)90171-8). It is the most widely used and validated measure of comorbidity level by researchers and helps clinicians to make informed decisions about procedures. It predicts the mortality for a patient by assigning a score of 1, 2, 3, or 6 to each condition depending on the risk of dying associated with each comorbidity. Comorbidities will be limited to the 17 conditions included in CCI.
	2.3	Date of primary cancer diagnosis	The date of diagnosis of the main condition treated or investigated during the relevant episode of healthcare. When available, ENCR (European Cancer Registry Network) should be used as the date of primary diagnosis. If ENCR is not available, it was agreed at network level to use date of the pathology tissue (pathology date) biopsy or imaging based diagnosis (imaging date).

	MEDOC ID	MEDOC Data Concept	Definition summary
	2.4	Method of primary cancer diagnosis	The initial technique or approach used to identify and confirm the presence of cancer in an individual. Two options have been agreed for the method of primary diagnosis: pathology and imaging based diagnosis.
	2.5	Performance status (for example, coded by ECOG or Karnofsky standards)	An evaluation of patient situation. It can be Karnofsky performance status or ECOG performance status (also called WHO or Zubrod performance status).
	2.6	Disease stage in a recognized standard such as TNM	The classification or categorization of cancer based on its extent of spread in the body at a specific point in time. The stage of cancer is typically described using a numerical system (such as stages I to IV) or a combination of letters and numbers (such as TNM staging system).
	2.7	Histological cell type, typically in ICD-O-3 standards	The specific type of cells that make up a tumour or cancerous growth, based on the International Classification of Diseases for Oncology, 3rd Edition (ICD-O-3). For example, common histological cell types in cancer include adenocarcinoma, squamous cell carcinoma, and melanoma, each associated with different types of tissues and exhibiting unique features under microscopic examination. ICD-O-3 mappings may not be exhaustive in OMOP, but missing ones can be reported if deemed critical.
	2.8	Menopausal status (for example, for patients with breast cancer)	This is a cancer specific data concept only required for women with breast cancer. Menopausal status will be identified from a structured field or notes in the EMR (patients report that their periods have ceased), blood test results (such as FISH and oestrogen) and surgical history (removal of the ovaries).
Bio-markers	3.1	Biomarker name	The quantification or assessment of a specific biomarker in a biological sample at a particular point in time.
	3.2	Biomarker measure	The measurement of a biomarker involves collecting a biological sample, such as blood, urine, tissue, or saliva, and analysing it using various laboratory techniques. The result of the biomarker measurement indicates the level, concentration, or presence of the biomarker in the sample, which can be indicative of a particular condition, disease progression, or treatment effect.
	3.3	Biological sample ID	The sample of the biomarker measurement.

	MEDOC ID	MEDOC Data Concept	Definition summary
Treatments	4.1	Line of therapy (*derived algorithmically within each cancer type*)	Line of Therapy is algorithmically derived using anti-cancer treatment name, surgery type and radiotherapy type including the start dates of these treatments. The definitions will be in the protocols. Implementation of a consistent Line of Therapy definition across the network will allow multi-country analysis of treatment patterns and outcomes
	4.2	Anti-cancer treatment name, including systemic treatment and supportive therapy	A prescribed systematic form of treatment such as chemotherapy or a combination of chemotherapy and immunotherapy, and supportive therapy such as bisphosphonates. It encompasses a systemic anti-cancer treatment (SACT) regimen, which consists of one or more cycles of anti-cancer drugs. Additionally, it may include information regarding the treatment intent, such as adjuvant or neoadjuvant, serving as a new field to specify the therapeutic purpose of the treatment.
	4.3	Molecule generic name	The molecule generic name refers to the name of a specific anti-cancer drug or supportive therapy used within a systemic anti-cancer treatment (SACT) regimen. It can be represented either as free text or using a coded format. In the case of combination therapies within the SACT regimen, multiple entries may need to be created, each corresponding to the specific molecule or drug utilized in the treatment.
	4.4	Start date for drug treatment	The specific date on which a patient receives the first dose of the systemic anti-cancer treatment (SACT) regimen. It marks the initiation of the anti-cancer therapy, and it can be recorded as either the date the SACT regimen is applied or prepared for administration, or the actual date when the patient begins receiving the anti-cancer treatment.
	4.5	Treatment dose	The specific amount of an anti-cancer drug or therapeutic substance administered to a patient as part of their systemic anti-cancer treatment (SACT) regimen. This dose is carefully prescribed and calculated based on the patient's medical condition, drug characteristics, and treatment goals to achieve optimal therapeutic effects while minimizing potential side effects. It is typically measured in milligrams (mg) per square meter of body surface area and is given in a single application during the course of anti-cancer treatment.
	4.6	End date for drug treatment	The specific date on which a patient completes their prescribed course of systemic anti-cancer treatment (SACT) regimen, marking the conclusion or discontinuation of the administration of anti-cancer drugs or therapeutic substances. This date is essential for tracking the treatment duration, assessing treatment outcomes, including response to therapy and potential adverse effects, and ensuring comprehensive and accurate records of the patient's treatment journey.
	4.7	Radiotherapy type	The specific approach or technique utilized in the administration of radiotherapy as a treatment for cancer. This term can be represented using a coded format, such as a procedure code, to accurately describe the method employed during the delivery of the radiotherapy treatment. It serves as an important classification system to categorize and document the various radiotherapy procedures used in cancer treatment, aiding in data analysis, research, and treatment planning.
	4.8	Radiotherapy start date	The specific date on which a patient begins their radiotherapy treatment for cancer. It marks the initiation of the radiotherapy procedure and is a crucial piece of information for tracking the treatment timeline, evaluating treatment outcomes, and maintaining comprehensive records of the patient's cancer therapy journey. This date is recorded to facilitate research, treatment planning, and data analysis related to radiotherapy interventions in cancer patients.

	MEDOC ID	MEDOC Data Concept	Definition summary
	4.9	Radiotherapy dose	The amount of radiation delivered to a specific area or target within the body as part of a radiotherapy treatment. It can be measured in Gray (Gy) and represents either the prescribed total radiation dose for the entire course of treatment or the actual dose delivered to the patient. This information is crucial for evaluating treatment effectiveness, monitoring patient responses, and conducting research in the field of cancer therapy. [Note on EQD2 as the dose delivered in 2Gy fractions that is biologically equivalent to a total dose]
	4.10	Radiotherapy end date	The specific date on which the administration of radiotherapy treatment for the patient concludes, corresponding to the date of delivery of the last fraction of radiotherapy. It marks the completion of the prescribed course of radiotherapy treatment and is essential for accurately documenting the treatment duration, assessing treatment outcomes, and maintaining comprehensive records of the patient's radiotherapy journey for research, analysis, and treatment planning purposes.
	4.11	Surgery type	The specific classification or code used to describe the type of surgical procedure(s) performed on a patient. This classification should be represented in a coded format, with one entry per procedure. The „Surgery type" field is essential for accurately documenting the surgical interventions conducted on patients, facilitating data analysis, research, and treatment planning within the research group or healthcare institution. It enables a standardized and systematic approach to categorizing and studying various surgical procedures in the context of cancer research or medical treatment.
	4.12	Surgery date	The specific date on which a surgical procedure was performed on a patient as part of their cancer treatment. It represents the day when the surgery took place and is crucial for tracking treatment timelines, evaluating treatment outcomes, and maintaining comprehensive records of surgical interventions for research and analysis within the research group.
	4.13	Participation in clinical trial	The involvement of a patient in an interventional drug trial in oncology. This concept encompasses patients who have volunteered to be part of a clinical trial, following specific protocols, and allows researchers to gather valuable data to assess the investigational drug's potential benefits and risks.
	4.14	Date of trial consent	The specific date on which a patient provides informed consent to participate in a clinical trial. It marks the point in time when the patient formally agrees to volunteer for the research study after receiving comprehensive information about the trial's purpose, procedures, potential risks, and benefits.

	MEDOC ID	MEDOC Data Concept	Definition summary
Outcomes	**5.1**	Date of death, at any location	The specific date on which a patient died or is officially deemed to have died. It is preferable to have linkage to a death registry to capture death in any location and not limited to deaths that occur at the participating centre
	5.2	Time to next treatment (*derived from treatment start dates*)	The derived duration, measured in days, between the completion of a previous treatment and the initiation of the subsequent treatment in a patient's cancer therapy journey. This concept is calculated based on existing data and may depend on the specific study protocol being conducted. It is essential for monitoring the interval between different treatment phases, understanding treatment patterns, and assessing the impact of delays or modifications in the treatment schedule on patient outcomes in cancer research.
	5.3	Metastasis presence/absence	The indication of whether metastatic disease is observed in a patient's cancer diagnosis or not. Metastasis involves the spread of cancer cells from the primary tumor to other parts of the body, forming distant growths in different organs. Patients can present with metastatic disease at the time of diagnosis or at a later stage. This concept is essential for accurately characterizing the stage and severity of cancer, guiding treatment decisions, and monitoring disease progression in cancer research and clinical practice.
	5.4	Metastasis location	The specific site or organ in the body where cancer has spread from its original or primary site. It denotes the secondary locations where cancer cells have formed distant growths, resulting from the migration of malignant cells from the primary tumor. Recording the „Metastasis location" is crucial for understanding the extent of cancer spread, determining the stage of the disease, and tailoring appropriate treatment strategies in cancer research and clinical management.
	5.5	Date of clinical visits (with cancer related visits separated from other visits	The specific date on which a patient attended a cancer-related in-person visit or telehealth consultation (e.g. phone consultation). This date will be used to derive the date of last visit/follow-up, or the last known date when the patient was seen or confirmed to be alive in the data. This information is critical for tracking patient follow-up, evaluating treatment outcomes, and assessing survival rates in cancer research and medical practice.
	5.6	Vital status (*derived from visits or death linkage*)	The current known state of a patient's life at the time of data extraction. It indicates whether the patient is currently alive or deceased. This information is derived from the available data and is essential for tracking patient outcomes, conducting survival analyses, and understanding the current status of patients in cancer research or medical studies.
	5.7	Extent of debulking surgery (for example, for patients with gynecological cancer)	This is a cancer specific data concept only required for women with a gynaecological cancer. „Extent of debulking" refers to the outcome of debulking surgery such as complete resection, optimal, and sub-optimal. The measure is largest diameter residual disease in centimetres.

Charlson Comorbidity Index — 17 Conditions

#	Condition	Suggested ICD-10 Codes*
1	Myocardial Infarction	I21.x, I22.x, I25.2
2	Congestive Heart Failure	I09.9,I11.0, I13.0, I13.2, I25.5, I42.0, I42.5-I42.9, I43.x, I50.x, P29.0
3	Peripheral Vascular Disease	I70.x, I71.x, I73.1, I73.8, I73.9, I77.1, I79.0, I79.2, K55.1, K55.8, K55.9, Z95.8, Z95.9
4	Cerebrovascular Disease (CVD)	G45.x, G46.x, H34.0, I60.x-I69.x
5	Dementia	F00.x-F03.x, F05.1, G30.x, G31.1
6	Chronic Obstructive Pulmonary Disease (COPD)	I27.8, I27.9, J40.x-J47.x, J60.x-J67.x, J68.4, J70.1, J70.3
7	Rheumatologic Disease	M05.x, M06.x, M31.5, M32.x-M34.x, M35.1, M35.3, M36.0
8	Peptic Ulcer Disease	K25.x-K28.x
9	Diabetes	E10.0, E10.l, E10.6, E10.8, E10.9, E11.0, E11.1, E11.6, E11.8, E11.9, E12.0, E12.1, E12.6, E12.8, E12.9, E13.0, E13.1, E13.6, E13.8, E13.9, E14.0, E14.1, E14.6, E14.8, E14.9
10a	Diabetes with Complications	E10.2-E10.5, E10.7, E11.2-E11.5, E11.7, E12.2-E12.5, E12.7, E13.2-E13.5, E13.7, E14.2-E14.5, E14.7
10b	Paralysis (Hemiplegia or Paraplegia)	G04.1, G11.4, G80.1, G80.2, G81.x, G82.x, G83.0-G83.4, G83.9
12	Mild Liver Disease	B18.x, K70.0-K70.3, K70.9, K71.3-K71.5, K71.7, K73.x, K74.x, K76.0, K76.2-K76.4, K76.8, K76.9, Z94.4
13a	Moderate/Severe Liver Disease	I85.0, I85.9, I86.4, I98.2, K70.4, K71.1, K72.1, K72.9, K76.5, K76.6, K76.7
13b	Renal Disease	I12.0, I13.1, N03.2-N03.7, N05.2-N05.7, N18.x, N19.x, N25.0, Z49.0-Z49.2, Z94.0, Z99.2
14	Solid tumour	C00.x-C26.x, C30.x-C34.x, C37.x-C41.x, C43.x-C58.x, C60.x-C80.x
15	Leukaemia	C91.x-C95.x
16	Lymphoma	C81.x- C86.x, C88.x
17	AIDS	B20.x-B22.x, B24.x

*Quan et al., Coding Algorithms for Defining Comorbidities in ICD-9-CM and ICD-10 Administrative Data, Med Care 2005;43: 1130–1139

Copyright 2022 IQVIA Ltd.

Licensed under the Apache License, Version 2.0 (the “License”); you may not use this software except in compliance with the License. You may obtain a copy of the License at http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS” BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

All files in this Supplementary Materials are implicitly licensed under the Apache License, Version 2.0, unless otherwise explicitly stated.

Apache License
Version 2.0, January 2004
http://www.apache.org/licenses/