

# Enhancing healthcare process analysis through object-centric process mining: Transforming OMOP common data models into object-centric event logs

Gyunam Park<sup>a</sup>, Yaejin Lee<sup>b</sup>, Minsu Cho<sup>b,\*</sup>

<sup>a</sup>Process and Data Science Group (PADS), RWTH Aachen University, Ahornstraße 55, Aachen, 52074, North Rhine-Westphalia, Germany

<sup>b</sup>School of Information Convergence, Kwangwoon University, Kwangwoon-ro 20, Seoul, 01897, South Korea

---

## Abstract

**Objectives.** This study aims to enhance the analysis of healthcare processes by introducing Object-Centric Process Mining (OCPM). By offering a holistic perspective that accounts for the interactions among various objects, OCPM transcends the constraints of conventional patient-centric process mining approaches, ensuring a more detailed and inclusive understanding of healthcare dynamics.

**Methods.** We develop a novel method to transform the Observational Medical Outcomes Partnership Common Data Models (OMOP CDM) into Object-Centric Event Logs (OCELs). First, an OMOP CDM4PM is created from the standard OMOP CDM, focusing on data relevant to generating OCEL and addressing healthcare data's heterogeneity and standardization challenges. Second, this subset is transformed into OCEL based on specified healthcare criteria, including identifying various object types, clinical activities, and their relationships. The methodology is tested on the MIMIC-IV database to evaluate its effectiveness and utility.

**Results.** The application of our proposed method to the MIMIC-IV dataset effectively produces OCELs, allowing for the implementation of OCPM in the healthcare industry. We rigorously evaluate the comprehensiveness and level of abstraction to validate our approach's effectiveness. Additionally, we create diverse object-centric process models intricately designed to navigate the complexities inherent in healthcare processes.

**Conclusion.** Our approach introduces a novel perspective by integrating multiple viewpoints simultaneously. To the best of our knowledge, this is the inaugural application of OCPM within the healthcare sector, marking a significant advancement in the field.

**Keywords:** Health Informatics, Healthcare Process, Object-Centric Process Mining, OMOP Common Data Model, Object-Centric Event Log

---

## 1. Introduction

Process mining is a field that extracts process-related knowledge, including process discovery, conformance checking, predictive process monitoring, and action-oriented improvements, based on event logs collected in information

---

\*Corresponding author

Email addresses: gnpark@pads.rwth-aachen.de (Gyunam Park), 191yaejin@gmail.com (Yaejin Lee), mcho@kw.ac.kr (Minsu Cho)

systems [1, 2]. In particular, it has contributed to addressing healthcare process issues due to the increasing availability of medical event data [3, 4, 5, 6, 7, 8, 9]. For example, traditional process mining has effectively utilized in diverse processes [10], including outpatient care [11], surgical process [12], medical treatment [13], emergency room processes [14], or specific disease-oriented processes [15].

Recently, the application of object-centric process mining in healthcare has gained attention due to its benefits [16]. However, its practical implementation in the healthcare domain remains limited due to the requirement for a new data format, i.e., Object-Centric Event Logs (OCELs). To establish OCELs for healthcare, we posit that two pivotal challenges must be addressed: (i) a comprehensive specification of OCEL for healthcare is imperative for effectively utilizing OCPM, in other words, it is required to determine a diverse array of object types, abstraction levels for both object types and clinical activities, and the intricate relationships between object types and clinical activities, and (ii) the heterogeneity in data structure and vocabularies of healthcare information systems limits the standardization and reproducibility of results when applying object-centric process mining techniques.

In this study, we present a novel approach for converting an OMOP Common Data Model (OMOP CDM) [17] into an Object-Centric Event Log (OCEL), facilitating the application of object-centric process mining (OCPM) in healthcare. The method consists of two phases. In the first phase, we create an OMOP CDM4PM from a standard OMOP CDM. The OMOP CDM4PM is a subset of an OMOP CDM that is relevant to creating an OCEL. This addresses the challenge (ii). The starting point for our method is the OMOP CDM, a standard clinical data model designed explicitly for multi-center research, enabling collaborative studies through consistent reference terminologies across multiple healthcare institutions [17]. Leveraging the OMOP CDM as our starting point allows healthcare organizations with data collected in hospital information systems using this model to readily implement our method and systematically apply OCPM.

In the second phase, we suggest a method of transforming OMOP CDM4PM to OCEL based on the specification of OCEL. More specifically, we delineate all the critical elements integral to formulating OCEL in Healthcare. This entails a detailed exploration for assigning clinical activities and the systematic generation of object types, with a comprehensive consideration of the abstraction levels inherent in these processes. This addresses the challenge (i) mentioned earlier.

The proposed method contributes significantly to both the research and practical standpoints. From a research perspective, this work effectively addresses the identified challenges from the distinguished characteristics and challenges of process mining within the healthcare domain, notably the dominant patient-centric focus and the necessity to consider data at multiple abstraction levels [10]. Additionally, this research introduces a pioneering avenue by demonstrating the applicability of the OMOP CDM in resolving process-related issues involving diverse objects. Furthermore, in alignment with the challenges offered in [3], this research addresses the existing deficiency in applying standardized terminologies within the process mining field, laying the groundwork for using standardized data. Also, this work is the first to apply object-centric process mining in healthcare to the best of our knowledge. From a practical viewpoint, the proposed method broadens the scope of process mining techniques within the healthcare sphere,

enhancing their applicability and utility.

The rest of the paper is organized as follows. In Section 2, we motivate the importance of object-centric process mining in healthcare. Next, Section 3 introduces the OMOP CDM and OCEL, i.e., the input and output of our proposed method. Then, Figure 4 presents the method that transforms OMOP CDM to OCEL. Afterward, Section 5 evaluates the effectiveness and utility of the proposed method by applying it to real-life OMOP CDM datasets. Finally, Section 6 provides discussions and concludes the paper.

| Statement of significance |                                                                                                                                                                                                                                                    |
|---------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Problem                   | The application of object-centric process mining in healthcare is crucial for the complete transparency of healthcare processes.                                                                                                                   |
| What is Already Known     | Traditional process mining techniques are well applied to healthcare processes, but there is a lack of research on building object-centric event logs and applying object-centric process mining.                                                  |
| What This Paper Adds      | We propose an approach for converting an OMOP Common Data Model into an object-centric event log: (i) creating OMOP CDM4PM that focuses on data relevant to generating OCEL and (ii) transforming into OCEL based on specific healthcare criteria. |

## 2. Background

Traditional process mining methods in healthcare have conventionally adopted a patient-centric approach [4, 3, 18]. As such, a single entity, i.e., typically the patient or their visit, represents the entire process instance, and the related clinical events form the event logs.

A core application of traditional process mining is the discovery of the actual flow of patient processes. This is often achieved through the construction of clinical process models, such as the directly-follows graph (DFG) [19] depicted in Figure 1(a), which illustrate the sequence of care a patient receives. Existing research has explored a range of techniques for extracting process models from clinical event logs. In [20], the authors applied process mining to discover patient pathways in appendicitis treatment, aiding clinicians in visualizing common care sequences. To tackle the inherent complexity and variability of healthcare processes, Najjar et al. [21] employed clustering methods to discover heart failure treatment processes in elderly patients from clinical event logs, which assist in pinpointing treatment patterns linked with lower mortality or requiring more intensive monitoring. Additionally, another study [22] implemented topic modeling to condense the multitude of specific activities into a generalized overview of breast cancer post-surgical care, offering clinicians a high-level view of the treatment process.

In addition to process discovery, process mining has been employed to compare actual patient care processes with established clinical guidelines. For instance, a technique proposed in [23] identifies deviations in patient treatment for unstable angina, which could suggest necessary updates to clinical pathways or indicate a need for better compliance measures. Similarly, in [24], the authors investigated how the treatment of colon cancer patients aligns with clinical pathways by examining where care diverged from expected practices.

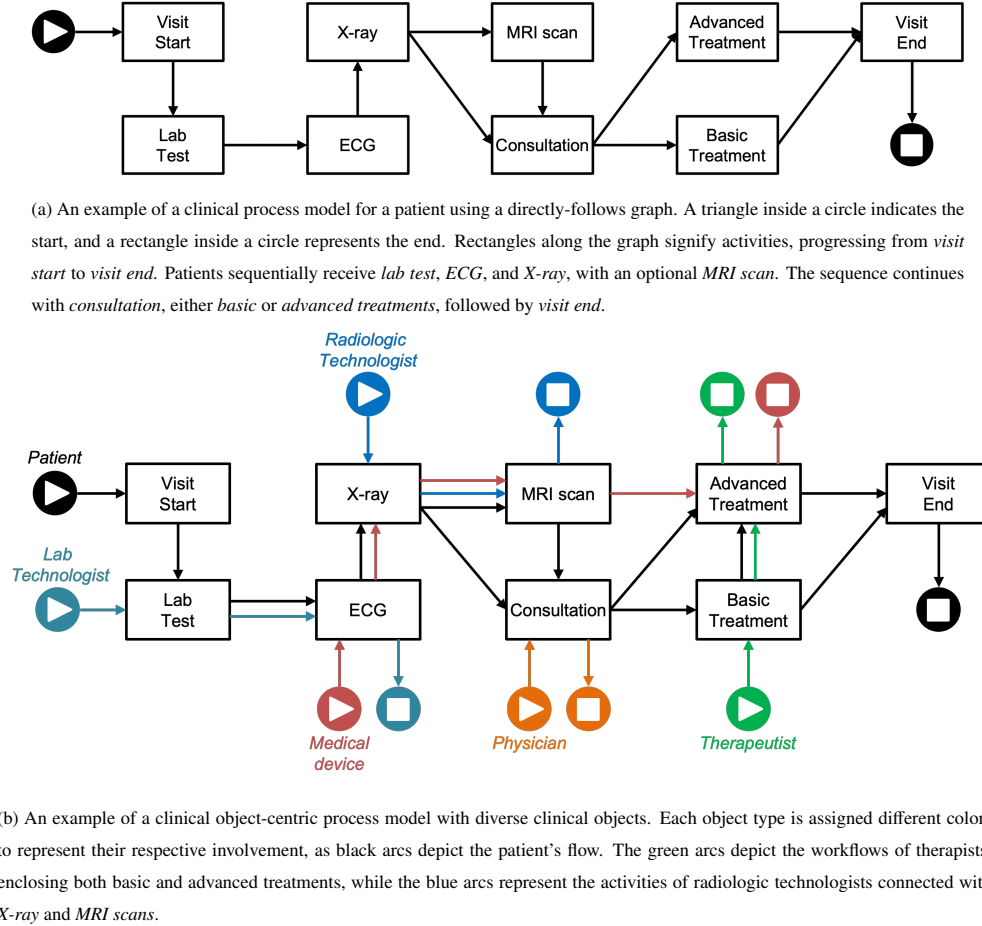


Figure 1: Examples of clinical process models with different viewpoints

In the healthcare system, however, processes of providing clinical services, e.g., outpatient or inpatient processes, include multiple groups of entities, i.e., object types. For example, it can include patient-related, provider-related, and device-related object types. Here, the patient-related object types correspond to a group of objects that receive clinical services within a process, while provider-related and device-related object types are connected to the objects that provide clinical services.

Each of the object types has its own process, i.e., a process for a patient, a process for a doctor, etc. More importantly, these clinical object types may be simultaneously involved in a specific clinical event. For example, in the case of an X-ray procedure, three entities are synchronously engaged: the patient who undergoes the X-ray

procedure, the radiologic technologist who administers it, and the X-ray equipment that carries out the operation. Figure 1(b) is the result of expanding the process model depicted in Figure 1(a), incorporating perspectives from other object types, e.g., lab technologist, radiologic technologist, medical device, physician, and therapist, in addition to the patient's viewpoint. This type of process model is referred to as an object-centric process model, which is generated by applying object-centric process mining (OCPM) from object-centric event logs (OCEL) [25].

In healthcare systems, such an analysis offers significant advantages. It provides complete transparency into healthcare processes, enabling a comprehensive view of clinical activities and their associations with clinical object types. This approach also facilitates the identification of relationships between clinical object types based on their synchronous involvement in activities, allowing for a deeper understanding of object-to-object dependencies within the healthcare process. Moreover, it offers a fresh perspective by simultaneously considering multiple viewpoints, enabling conformance checking, performance measurement, and root cause analysis.

### 3. Preliminaries

#### 3.1. Observational Medical Outcomes Partnership Common Data Model (OMOP CDM)

OMOP CDM<sup>1</sup> consists of 6 table groups: clinical data, health systems data, derived elements, health economics, metadata, and vocabularies. Each table group has two or more tables, and there are various relationships between tables within a table group as well as between different table groups. It can be formalized directly and straightforwardly. Note that not all table groups or tables are included, but only the essential components introduced earlier.

**Definition 1** (OMOP Common Data Model (CDM)). *Let  $Pe$  be the set of person records, i.e., person table,  $VO$  the visit occurrence table,  $PO$  the procedure occurrence table,  $DrE$  the drug exposure table,  $DE$  the device exposure table,  $CO$  the condition occurrence table,  $Me$  the measurement table,  $Ob$  the observation table,  $Lo$  the location table,  $CS$  the care site table,  $Pr$  the provider table,  $Co$  the concept table,  $Vo$  the vocabulary table,  $Do$  the domain table, and  $CC$  the concept class table. An OMOP common data model is a tuple  $CDM = (CDC, HSC, SVC, Rel)$  where*

- $CDC = \{Pe, VO, PO, DrE, DE, CO, Me, Ob\}$  is a set of clinical tables;
- $HSC = \{Lo, CS, Pr\}$  is a set of health system tables;
- $SVC = \{Co, Vo, Do, CC\}$  is a set of vocabularies tables;
- $Rel \subseteq \{(x, y) \in X \times Y \mid X \subseteq \{CDC, HSC, SVC\} \wedge Y \subseteq \{CDC, HSC, SVC\}\}$  is a set of relationships between data from the tables.

---

<sup>1</sup><https://www.ohdsi.org>

Table 1: A fragment of an object-centric event log

| Event ID | Activity           | Timestamp           | Patient | Doctor     | Therapist |
|----------|--------------------|---------------------|---------|------------|-----------|
| e92      | Consultation       | 2023-01-14 09:00:00 | {Adams} | {Schuster} |           |
| e93      | Advanced treatment | 2023-01-14-13:00:00 | {Adams} |            | {Huang}   |
| ...      | ...                | ...                 | ...     | ...        | ...       |

### 3.2. Object-Centric Process Mining (OCPM) and Object-Centric Event Logs (OCEL)

In the field of process mining, we can distinguish between traditional and object-centric techniques. Traditional process mining operates under the assumption that each event in event logs corresponds to a single object of a unique type, often referred to as a case. For instance, in the process of a patient's hospital visit, a patient's check-in at the hospital could be an example of such a case. In this event log, an event is defined by three attributes: a case, an event type (i.e., activity), and a timestamp. For example, a patient (the case) registers at reception (the activity) at 10:00 AM on 01.Jan.2023 (the timestamp). Conversely, object-centric process mining techniques expand on the traditional single-case assumption to consider that an event may be associated with multiple objects of various types. For example, a patient's consultation with a doctor can involve not just the patient but also the doctor and perhaps a diagnosis or a prescribed medication.

An object-centric event log is the starting point of process analysis based on object-centric process mining. The Object-Centric Event Log (OCEL) standard [26] provides a comprehensive framework for these logs, addressing events, objects, their attributes, and their interrelationships. Before we define these logs, we begin by introducing several universes.

**Definition 2** (Universes).  $\mathbb{U}_{ei}$  is the universe of event identifiers;  $\mathbb{U}_{oi}$  is the universe of object identifiers;  $\mathbb{U}_{act}$  is the universe of event types (i.e., activities);  $\mathbb{U}_{time}$  is the universe of timestamps;  $\mathbb{U}_{ot}$  is the universe of object types;  $\mathbb{U}_{attr}$  is the universe of attributes;  $\mathbb{U}_{val}$  is the universe of values;  $\mathbb{U}_{map} = \mathbb{U}_{attr} \rightarrow \mathbb{U}_{val}$  is the universe of attribute-value mappings. For any  $f \in \mathbb{U}_{map}$  and  $x \notin \text{dom}(f)$ ,  $f(x) = \perp$ .

Based on the universes, we give a formal definition of an object-centric event log as follows.

**Definition 3** (Object-Centric Event Log). An object-centric event log is a tuple  $L = (E, O, R, \mu)$ , where  $E \subseteq \mathbb{U}_{ei}$  is a set of events,  $O \subseteq \mathbb{U}_{oi}$  is a set of objects,  $\mu \in (E \rightarrow \mathbb{U}_{map}) \cup (O \rightarrow \mathbb{U}_{map})$  is a mapping, and  $R \subseteq E \times O$  is a relation, s.t. for any  $e \in E$ ,  $\mu(e)(act) \in \mathbb{U}_{act}$  and  $\mu(e)(time) \in \mathbb{U}_{time}$ , and for any  $o \in O$ ,  $\mu(o)(type) \in \mathbb{U}_{ot}$ .  $\mathbb{U}_L$  is the set of all possible object-centric event logs.

For convenience, we will express  $\mu(o)(x)$  as  $\mu_x(o)$  and  $\mu(e)(x)$  as  $\mu_x(e)$ . Table 1 depicts a subset of an event log  $EL_1 = (O_1, E_1, R_1, \mu_1)$  with  $O_1 = \{Adams, Schuster, Huang, \dots\}$ ,  $E_1 = \{e_{92}, e_{93}, \dots\}$ ,  $R_1 = \{(e_{92}, Adams), (e_{92}, Schuster), \dots\}$ ,  $\mu_{act}(e_{92}) = \text{Consultation}$ ,  $\mu_{time}(e_{92}) = 2023-01-14 09:00:00$ ,  $\mu_{type}(Adams) = \text{Patient}$  etc.

## 4. Method

As depicted in Figure 2, our transformation method consists of two phases. In the first phase, we create an OMOP CDM for Process Mining (OMOP CDM4PM). This entails organizing the data across three categories: event tables, object tables, and a concept table. The associations between events and objects are outlined in event-object relationships, while concept relationships link events and objects to their respective concepts. The events, objects, and relations, essential for constructing an OCEL are directly derived from the corresponding event tables, object tables, and event-object relationships, whereas the event and object type mappings are derived from the concept of each event and object.

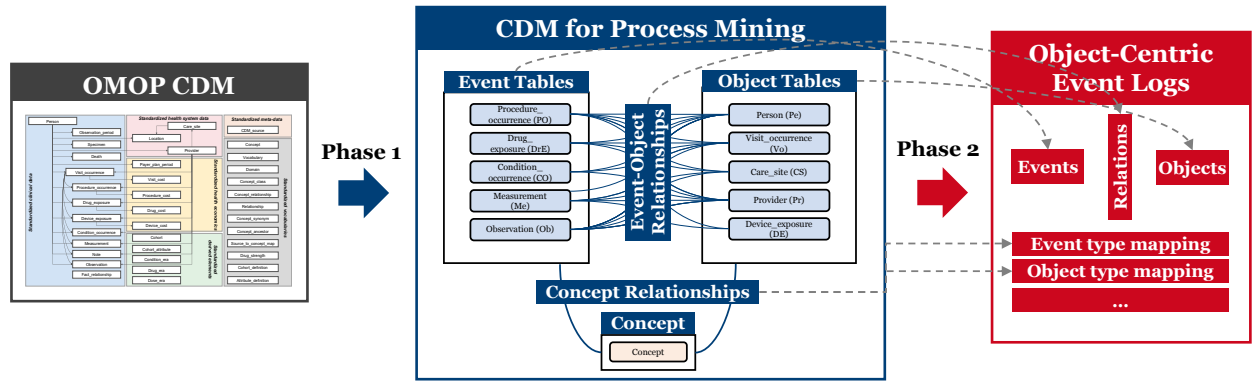


Figure 2: Overview of the two-phase transformation method from OMOP CDM to OCEL. Phase 1 organizes data into event, object, and concept tables with interrelationships, while phase 2 generates the necessary components and mappings for an OCEL.

### 4.1. Phase 1: Creating OMOP CDM4PM

As outlined in Subsection 3.2, an object-centric event log primarily encompasses three record types: events, objects, and the relationships that exist between these events and objects. Although the OMOP CDM inherently houses these three elements, it does not organize data on events, objects, and their interrelationships in a format that is readily applicable to object-centric process mining methods.

#### 4.1.1. Events in OMOP CDM

Our initial step involves pinpointing the tables in OMOP CDM that contain event concepts. We outline the following tables as primary sources of event records in the OMOP CDM schema:

- *Procedure\_occurrence*: This table records the instances of procedures (i.e., events) performed on a patient during their course of care.
- *Drug\_exposure*: This table records the instances of a person ingesting or being exposed to a drug (i.e., events).
- *Condition\_occurrence*: This records the instances of a person experiencing a condition (i.e., events), i.e., a disease, ailment, or medical issue that the person has been diagnosed with or is considered to have.

- *Measurement*: This table records measurements (i.e., events), e.g., lab results, vital signs like blood pressure or heart rate, scores from assessment scales, etc.
- *Observation*: This table records observations (i.e., events) about a person obtained in the course of healthcare delivery, e.g., lifestyle choices (like smoking status or alcohol use), patient-reported outcomes, survey data, etc.

Note that additional tables can be incorporated based on the analysis objectives and users' preferences. For instance, the *visit\_occurrence* table, which contains information on the start date and visit date, allows for the generation of supplementary events, such as *visit\_start* and *visit\_end*.

For example, each row of *procedure\_occurrence* table in the OMOP CDM represents a unique procedure occurrence (see Table A.4). *Procedure\_Occurrence\_ID* is a unique identifier for each procedure occurrence, while *Person\_ID* is a unique identifier for each patient in the database. The *Concept\_ID* corresponds to the identifier for the concept in the OMOP CDM vocabulary that represents the procedure. The *DateTime* column records the date and time when the procedure took place. The *Provider\_ID* represents a unique identifier for the provider (e.g., doctor) who performed the procedure, and *Visit\_Occurrence\_ID* is a unique identifier for the visit occurrence during which the procedure was performed. *Quantity* indicates the number of times the procedure was performed during the visit.

#### 4.1.2. Objects in OMOP CDM

Subsequently, we proceed to determine the tables in OMOP CDM that house object concepts. The following tables are identified as containing object records within the OMOP CDM schema:

- *Person*: This table records demographic information about patients in the database, such as their gender, year of birth, race, and ethnicity.
- *Visit\_occurrence*: This table records the instances of a person's encounters with the healthcare system, which may take place in various settings, such as inpatient (hospital), outpatient (clinic), emergency room, etc.
- *Care\_site*: This table holds information about places where healthcare services are delivered.
- *Provider*: This table stores information about healthcare professionals who provide care to patients, including doctors, nurses, therapists, or other types of providers.
- *Device\_exposure*: This table stores information about clinical devices that directly enable the clinical care of patients.

For instance, each row of the *provider* table from the OMOP CDM corresponds to a unique provider (see Table A.5). The *Provider\_ID* is the unique identifier for each provider. The *Concept\_ID* column represents the unique identifier for the concept in the OMOP CDM vocabulary, which represents the specialty of the provider. The *Care\_Site\_ID* represents the unique identifier for the location where the provider delivers care. The *NPI* (National Provider Identifier) is a unique identification number for healthcare providers in the United States. The *DEA*

(Drug Enforcement Administration) column holds the DEA registration number, which is a unique identifier assigned to a health professional who prescribes controlled substances.

#### 4.1.3. Event to Object Relationship in OMOP CDM

The OMOP CDM inherently defines relationships between tables based on primary and foreign key associations. Leveraging these pre-established connections, we draw upon the identified event and object tables to construct explicit event-to-object relationships. For example, the first row in the event table depicted in Table A.4 is linked to the first row in the object table depicted in Table A.5, as they are connected through `Provider_ID`.

#### 4.1.4. Concepts of Events and Objects

The event and object tables in the OMOP CDM contain unique identifiers, such as `Procedure_Occurrence_ID` and `Provider_ID`. These identifiers refer to specific instances of events or objects in the healthcare process. However, to understand the nature of these specific instances, we must reference the *concept* table in the OMOP CDM, which provides standardized definitions for these identifiers. In essence, the concept table offers detailed descriptions of these entities, such as what the *Procedure Occurrence 1278* refers to or the specialization of *Provider 514*.

The concept table consists of the following columns (see Table A.6): The `Concept_ID` column is a unique identifier for each concept; The `Concept_Class_ID` designates the category to which a concept belongs, while the `Domain_ID` provides information about the data domain that a particular concept belongs to; The `Concept_Name` is the human-readable term or phrase that represents the concept; In addition to this, the `Valid_Start_Date` and `Valid_End_Date` denote the period of validity for each concept;

#### 4.1.5. OMOP CDM for Process Mining (OMOP CDM4PM)

Drawing from the specifications identified in the preceding subsections, we present our newly defined model, OMOP CDM4PM. This refined model constitutes a targeted subset of the larger OMOP CDM schema, explicitly designed to enable process mining applications in healthcare data analysis.

OMOP CDM4PM comprises five principal components: Firstly, *event tables* are derived from specific tables within the OMOP CDM that encapsulate event records. Secondly, *object tables* originate from the OMOP CDM tables that hold object records. Thirdly, an *event-object relationship table* is derived from existing relationships within the OMOP CDM, predominantly based on primary and foreign key connections. Fourthly, *concept table* is the concept table from the OMOP CDM. Lastly, *concept relationship table* represents the link from events and objects to concepts.

**Definition 4** (OMOP CDM4PM). Let  $CDM = (CDC, HSC, SVC, Rel)$  be an OMOP CDM, where  $Co \in SVC$  is a concept table. An OMOP CDM4PM is a tuple  $CDM_{PM} = (ETab, OTab, R_{EO}, Co, R_{Co})$  where

- $ETab = \{PO, DrE, CO, Me, Ob\} \subseteq CDC$  is a set of event tables;
- $OTab = \{Pe, VO, CS, Pr, DE\} \subseteq CDC \cup HSC \cup SVC$  is a set of object tables;

- $R_{EO} \subseteq \{(x, y) \in X \times Y \mid X \in ETab \wedge Y \in OTab\}$  is a set of relationships between event data and object data;
- $R_{Co} \subseteq \{(x, y) \in X \times Co \mid X \in ETab \cup OTab\}$  is a set of concept relationships.

For any table  $T \in ETab \cup OTab$  and its entry  $x \in T$ ,  $\pi_y(x)$  returns  $x$ 's  $y$  attribute. For example, for  $pr_1 \in Pe$ ,  $\pi_{name}(pr_1)$  returns the name of patient  $pr_1$ .

#### 4.2. Phase 2: Transforming OMOP CDM4PM to Object-Centric Event Log (OCEL)

In this section, we detail the conversion process from OMOP CDM4PM into OCEL. The primary aim is to generate all necessary components of an OCEL: namely, the events ( $E$ ), objects ( $O$ ), relationships ( $R$ ), and mapping functions ( $\mu$ ), using an OMOP CDM4PM.

Given that each event in an OCEL needs to possess an event type (i.e., activity) and timestamp, and each object should have an object type (cf. Definition 3), our attention is primarily centered on creating the essential mapping functions. These include event type mapping ( $\mu_{act}$ ), timestamp mapping ( $\mu_{time}$ ), and object type mapping ( $\mu_{type}$ ).

Assuming the availability of the following functions: 1) a function to assign an identifier to each table entry ( $id$ ), 2) a function to assign an event type to each event table entry ( $act$ ), 3) a function to allocate a timestamp to each event table entry ( $time$ ), and 4) a function to assign an object type to each object table entry ( $otype$ ), we can then formalize the transformation function. The detailed implementation of each of these functions will be discussed in the subsequent parts of this section.

**Definition 5** (Transformation). *Let  $CDM_{PM} = (ETab, OTab, R_{EO}, Co, R_{Co})$  be an OMOP CDM4PM. For any  $T \in ETab \cup OTab$  and  $x \in T$ ,  $id(x)$  denotes the identifier of  $x$ . For any  $T \in ETab$ ,  $act(x)$  and  $time(x)$  denote the event type and timestamp associated with  $x$ . For any  $T \in OTab$ ,  $otype(x)$  denotes the associated object type with  $x$ .  $trans(CDM_{PM}) \in \mathbb{U}_L$  is a transformation of an OMOP CDM to an OCEL, i.e.,  $trans(CDM_{PM}) = (E, O, R, \mu)$  where*

- $E = \{id(x) \mid T \in ETab \wedge x \in T\}$ ,
- $O = \{id(x) \mid T \in OTab \wedge x \in T\}$ ,
- $R = \{(id(x), id(y)) \mid (x, y) \in R_{EO}\}$ ,
- $\mu_{act} = \{(id(x), act(x)) \mid T \in ETab \wedge x \in T\}$ ,
- $\mu_{time} = \{(id(x), time(x)) \mid T \in ETab \wedge x \in T\}$ , and
- $\mu_{type} = \{(id(x), otype(x)) \mid T \in OTab \wedge x \in T\}$

Implementing the identifier assignment functions ( $id$ ) is relatively straightforward. As illustrated in Table A.4 and Table A.5, every event and object table within the OMOP CDM4PM has an ID field. We leverage this field to generate identifiers for both events and objects, which means the function  $id$  maps table entries to identifiers. For instance, in

reference to the first row of the procedure occurrence table shown in Table A.4, labeled as  $po_1$ , the output of the identifier function is  $id(po_1) = 1277$ .

Similarly, the timestamp assignment functions ( $time$ ) can be implemented by employing the `DateTime` field present in each event table. For example, for the first row of the procedure occurrence table denoted as  $po_1$ , the timestamp function assigns  $time(po_1) = 2023-01-14 00:00:00$ .

Despite the straightforward implementation of identifier and timestamp assignment functions, the task of implementing the event type assignment functions and object type assignment functions remains a more complex problem. The methods for implementing these functions are detailed in the following subsections.

#### 4.2.1. Event Type Assignment

Event type refers to event type; generally, each event has precisely one event type. Different from other domains, however, in the healthcare system, clinical activities hold a systematic hierarchy on the basis of ontology mapping. It implies that a single clinical event can be mapped to different clinical activities based on the level of granularity.

A representative example of establishing the hierarchy of clinical activities is *SNOMED* clinical terms [27]. Figure B.6 presents an example of partial *SNOMED* clinical terms hierarchy. *SNOMED* clinical terms define the health terms and their relationships with the IS-A concept for representing the roles of parents and children.

In this work, we determine clinical activities with three abstraction levels suited for the healthcare environment so that one event can have up to three levels of activities at the same time. *High-level* clinical activities represent the largest aggregate with the homogeneous property. Therefore, they are the most commonly expressed and can be connected to the direct behaviors of clinical service providers, such as observation, drug, and procedure. *Low-level* clinical activities stand for specific clinical terms, concepts, and orders, whereas middle-level clinical activities are somewhere between high and low levels. For example, checking for alcohol abuse is engaged in low-level event type. In such a step, surveys, i.e., the means to check for alcohol abuse, are *middle-level*, while the high-level of this event type becomes observations.

Based on the concept, we define three distinct functions for generating event types: *high-level*, *middle-level*, and *low-level* event type assignment functions. These functions leverage the concept table ( $Co$ ) in the OMOP CDM4PM for implementation. The concept table provides various degrees of explanatory detail for each event listed in the event table.

The high-level event type assignment function utilizes the most comprehensive explanation available in the `Domain_ID` field. The middle-level event type assignment function employs an intermediate level of detail from the `Concept_Class_ID` field. Finally, the low-level event type assignment function extracts the most granular explanation provided in the `Concept_Name` field.

**Definition 6** (Event Type Assignment Functions). *Let  $CDM_{PM} = (ETab, OTab, R_{EO}, Co, R_{Co})$  be an OMOP CDM4PM. We denote  $ecmap$  to be a event concept mapping that assigns concepts to events, i.e.,  $ecmap = \{(x, c) \in T \times Co \mid T \in ETab \wedge (x, c) \in R_{Co}\}$ . We define three event type assignment functions:*

- $act_{high} = \{(x, \pi_{Domain\_ID}(ecmap(x))) \mid T \in ETab \wedge x \in T\}$  is a high-level event type assignment function,
- $act_{middle} = \{(x, \pi_{Concept\_Class\_ID}(ecmap(x))) \mid T \in ETab \wedge x \in T\}$  is a middle-level event type assignment function,
- $act_{low} = \{(x, \pi_{Concept\_Name}(ecmap(x))) \mid T \in ETab \wedge x \in T\}$  is a low-level event type assignment function,

Considering the second row of the procedure occurrence table as shown in Table A.4, labeled as  $po_1$ , its corresponding concept, namely  $ecmap(po_1)$ , aligns with the second row of the concept table as illustrated in Table A.6. The high-level event type assigned to  $po_1$ , formally  $act_{high}(po_1)$ , is *Procedure*, as the DOMAIN\_ID field of  $po_1$ 's associated concept, that is, 1989694, is defined as *Procedure*. In contrast, the low-level event type for  $po_1$ , i.e.,  $act_{low}(po_1)$ , is *Cone beam CT Teeth* because the respective Concept\_Name of  $po_1$ 's associated concept, i.e., 1989694, is *Cone beam CT Teeth*.

#### 4.2.2. Object Type Assignment

Similar to the multi-leveled approach for assigning event types to events, we employ a two-level approach for assigning object types to objects. High-level object types include categories such as patients, visit occurrences, providers, and care sites. These types correspond directly to the object tables found in an OMOP CDM4PM. Low-level object types, on the other hand, are specified for providers and care sites. The low-level object types for providers reflect various provider specialties, which include designations such as *Nurses*, *Physicians*, and *Pharmacists*. For care sites, low-level object types are established based on the diverse facilities and services provided, encompassing classifications like *Magnetic Resonance Imaging*, *Insulin pumps*, and *Beds*.

The high-level object type assignment function draws on the Domain\_ID field, while the low-level object type assignment function utilizes the Concept\_Name field.

**Definition 7** (Object Type Assignment Functions). *Let  $CDM_{PM} = (ETab, OTab, R_{EO}, Co, R_{Co})$  be an OMOP CDM4PM where  $OTab = \{Pe, VO, CS, Pr, DE\}$ . We denote  $ocmap$  to be a object concept mapping that assigns concepts to objects, i.e.,  $ocmap = \{(x, c) \in T \times Co \mid T \in OTab \wedge (x, c) \in R_{Co}\}$ . We define two object-type assignment functions:*

- $otype_{high} = \{(x, \pi_{Domain\_ID}(ocmap(x))) \mid T \in ETab \wedge x \in T\}$  is a high-level object type assignment function and
- $otype_{low} = \{(x, \pi_{Concept\_Name}(ocmap(x))) \mid T \in ETab \wedge x \in T\}$  is a low-level object type assignment function.

Taking the first row of the provider table depicted in Table A.5, denoted  $pr_1$ , as an example, its corresponding concept, denoted  $ocmap(pr_1)$ , matches the first row of the concept table depicted in Table A.6. The high-level object type associated with  $pr_1$ , i.e.,  $otype_{high}(pr_1)$ , is *Provider*, since the Domain\_ID of  $pr_1$ 's related concept, i.e., 32577, is listed as *Provider*. The low-level object type for  $pr_1$ , denoted  $otype_{low}(pr_1)$ , is *Physician*, given that the Concept\_Name field of  $pr_1$ 's related concept, i.e., 32577, is *Physician*.

## 5. Evaluation

In the evaluation, we use the publicly available MIMIC-IV dataset [28]. MIMIC-IV is a clinical database compiled from the electronic health records of the Beth Israel Deaconess Medical Center, spanning the years 2008 to 2019 and containing the records of 257,000 patient visits. We convert the MIMIC-IV dataset to a OMOP CDM using the systematic transformation approach [29].

The full MIMIC-IV dataset, while comprehensive, contains a wide array of healthcare interactions that vary not only by type (such as emergency, outpatient, and inpatient care) but also across different time periods. Analyzing such a heterogeneous dataset in its entirety could obscure insights, derived from applying process mining techniques, due to its complexity and the non-homogeneity of the processes. Therefore, to maintain analytical clarity and focus, we selected three distinct sub-datasets that represent varied yet specific healthcare settings, each constituting a different case study:

- MIMIC-IV-S that comprises selected records from 100 patients across all visit types;
- MIMIC-IV-IP limited to records from inpatient visits of 1,088 patients with the visit concept codes 262 (Inpatient Visit from Emergency Room) and 9201 (Inpatient Visit);
- MIMIC-IV-ED consisting of records from emergency department visits of 913 patients with the visit concept code 8870 (Emergency Department).

The source code and manual for the transformation is publicly available at <https://github.com/gyunamister/omop2ocel>.

### 5.1. Transforming MIMIC-IV datasets to OCELS

First, we create an OMOP CDM4PMs from each of the three OMOP CDM datasets, as described in Subsection 4.1. The OMOP-CDM4PM-S, derived from MIMIC-IV-S, encompasses 364,532 events and 541 objects, whereas the OMOP-CDM4PM-IP and OMOP-CDM4PM-ED contain 1,650,306 events with 2,414 objects and 1,580,208 events with 2,122 objects, respectively.

Next, we transform the OMOP CDM4PMs into OCELS, as described in Subsection 4.2. During the transformation process, we assign event types across three tiers: high-, middle-, and low-level activities as defined in Subsubsection 4.2.1. Specifically, events (for example, those within the measurement table) are classified according to their domains (such as ‘measurement’) at the high-level, their concept classes (like ‘laboratory class’) at the middle-level, and their specific concept names (such as ‘Lactate in blood’) at the low-level.

Finally, we generate object types as presented in Subsubsection 4.2.2. The resulting OCEL contains eight object types, which we categorize into three groups for a presentation purpose: patient-related, provider-related, and device-related. The patient-related object type includes *patients* and *visits*, while the provider-related encompasses *physicians*, *advanced health professionals*, and *pharmacists*. The device-related object types are grouped as *dermal*

*drug delivery* which include items such as Gelclair gel, Gelclair oral gel, and Hydrocerin cream, *injection* which encompasses items such as Fat emulsion, Insulin pump, and Gadobutrol injection, and *diagnostic and supplements* which involves items such as Flour-I-Strip AT, Aqua-e liquid, and therapeutic bed.

## 5.2. Evaluation on Data Quality

To assess the effectiveness of our transformation method, we employed the widely adopted data quality framework suggested by Kahn et al. [30], which categorizes data quality into completeness, conformance, and plausibility. Completeness measures whether all necessary data elements are included in the data set. Conformance evaluates the data's adherence to specified data standards and expected data formats, ranges, and behaviors. Plausibility assesses whether the data values are believable and realistic within the context.

The completeness of the transformation is analyzed by ensuring that the resultant data model retains all essential information from the source data model without any data loss during the transformation. This involves verifying that all expected events and objects are present and appropriately accounted for in the transformed dataset.

For conformance, we assess if the transformed data adheres to expected standards by comparing basic statistical measures such as the number of patients, the number of visits, and the duration of hospital stays, before and after the transformation. This comparison helps to ascertain if the transformation process maintains the integrity and accuracy of the data as it conforms to the predefined expectations.

Table 2: Results demonstrating the completeness and conformance of the transformation

|                | Completeness |           |           |       | Conformance |       |          |       |            |       |
|----------------|--------------|-----------|-----------|-------|-------------|-------|----------|-------|------------|-------|
|                | # events     |           | # objects |       | # patients  |       | # visits |       | LOS (days) |       |
|                | before       | after     | before    | after | before      | after | before   | after | before     | after |
| OMOP-CDM4PM-S  | 364,532      | 364,532   | 541       | 541   | 100         | 100   | 378      | 378   | 5.1        | 5.1   |
| OMOP-CDM4PM-IP | 1,650,306    | 1,650,306 | 2,414     | 2,414 | 1,088       | 1,088 | 1,141    | 1,141 | 15.3       | 15.3  |
| OMOP-CDM4PM-ED | 1,580,208    | 1,580,208 | 2,122     | 2,122 | 913         | 913   | 1,155    | 1,155 | 10.6       | 10.6  |

Table 2 shows the results of our experiments that demonstrate the transformation's effectiveness in maintaining data quality across different datasets:

- OMOP-CDM4PM-S: The number of events and objects remained constant at 364,532 and 541 respectively, before and after transformation. Similarly, the number of patients and visits remained at 100 and 378, with the length of stay (LOS) consistently at 5.1 days.
- OMOP-CDM4PM-IP: Here too, we observed no change in the number of events (1,650,306), objects (2,414), patients (1,088), visits (1,141), or LOS (15.3 days).
- OMOP-CDM4PM-ED: Consistency was maintained with 1,580,208 events, 2,122 objects, 913 patients, and 1,155 visits both before and after transformation, and the LOS remained at 10.6 days.

### 5.3. Evaluation on Abstractness

Subsequently, we assess the abstractness of the transformation. The relevance of event abstraction in process mining lies in its ability to distill complex event logs into more manageable, higher-level views that are easier to understand and analyze [31]. By abstracting events, we can focus on the most significant aspects of the process flows, enabling clearer insights into the performance, bottlenecks, and variations in healthcare processes.

Table 3: Results demonstrating the effectiveness of abstractions in different levels

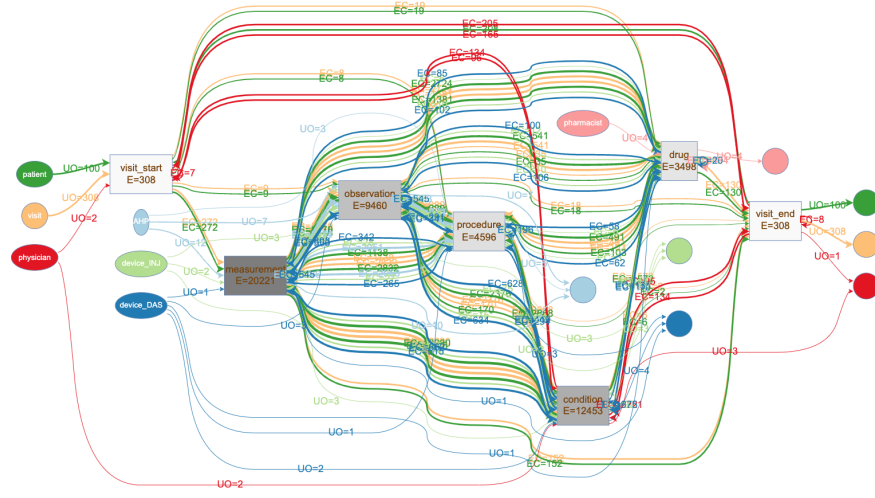
|                | Low-level     |               | Middle-level  |               | High-level    |               |
|----------------|---------------|---------------|---------------|---------------|---------------|---------------|
|                | # events      | Abstraction % | # events      | Abstraction % | # events      | Abstraction % |
| OMOP-CDM4PM-S  | 364,532       | -             | 68,804        | 81.13%        | 52,104        | 85.71%        |
| OMOP-CDM4PM-IP | 1,650,306     | -             | 133,143       | 91.93%        | 74,250        | 95.50%        |
| OMOP-CDM4PM-ED | 1,580,208     | -             | 123,155       | 92.21%        | 70,959        | 95.51%        |
|                | # event types | Abstraction % | # event types | Abstraction % | # event types | Abstraction % |
| OMOP-CDM4PM-S  | 2,393         | -             | 25            | 98.96%        | 7             | 99.71%        |
| OMOP-CDM4PM-IP | 4,156         | -             | 29            | 99.30%        | 7             | 99.83%        |
| OMOP-CDM4PM-ED | 4,227         | -             | 29            | 99.31%        | 7             | 99.83%        |

The results presented in Table 3 demonstrate a significant reduction in both the number of events and event types as we abstract from low-level to high-level. Specifically, the OMOP-CDM4PM-S dataset shows an 81.13% reduction in the number of events from low to middle-level abstraction, increasing to 85.71% at high-level abstraction. The number of event types also sees a significant decrease, from 2,393 types at the low level to only 25 at the middle level—a 98.96% reduction—and further down to 7 at the high level, indicating a 99.71% abstraction. Similarly, OMOP-CDM4PM-IP exhibits a 91.93% reduction in the number of events at the middle level and 95.50% at the high level, with the number of event types reduced by 99.30% and 99.83%, respectively. Lastly, for OMOP-CDM4PM-ED, there is a 92.21% reduction in the number of events at the middle level, which increases to 95.51% at the high level, while the number of event types is condensed by 99.31% at the middle level and 99.83% at the high level.

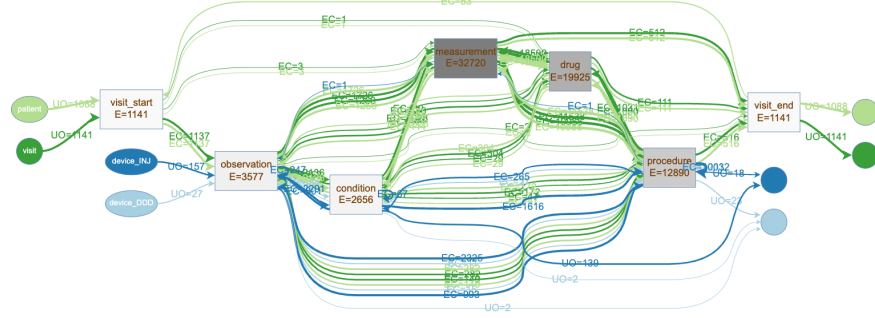
These high percentages of abstraction suggest that our transformation effectively simplifies the event logs, thereby enhancing the clarity and manageability of the data for process mining purposes.

### 5.4. Evaluation on Transformation Utility

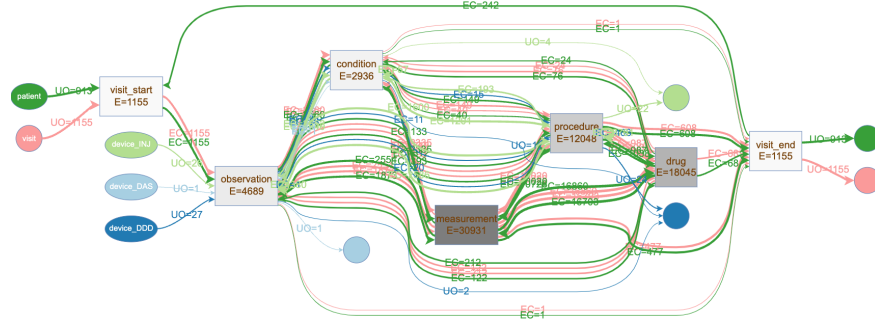
Next, we evaluate the utility of the transformation by analyzing the transformed OCELS (i.e., OCEL-S, OCEL-ED, and OCEL-IP) with object-centric process mining techniques. Figure 3 depicts the object-centric process models with the high-level activities from three datasets. The derived process models include the whole object types and high-level clinical activities without filtering. Here, the discovered process models take the form of directly-follow graphs, where the structure is determined by precedence and following relationships between events.



(a) The object-centric process model from OCEL-S



(b) The object-centric process model from OCEL-IP



(c) The object-centric process model from OCEL-ED

Figure 3: The discovered object-centric process models from three datasets using a DFG format. Circles symbolize objects, squares denote activities, and connecting lines illustrate the flow of objects related to activities. Different colors distinguish clinical objects. Numbers inside the squares and on lines connecting tasks indicate the quantities of events, while lines between tasks and objects show the quantities of objects.

However, it is evident that all these models exhibit a notable level of complexity. This complexity arises from the presence of high-level activities and the engagement of the whole object types. For instance, all three models include devices as object types, further divided into DDD (Dermal Drug Delivery), DAS (Diagnostic and Supplements) and INJ (Injection) categories. These device object types contribute significantly to the increased complexity observed in

the models.

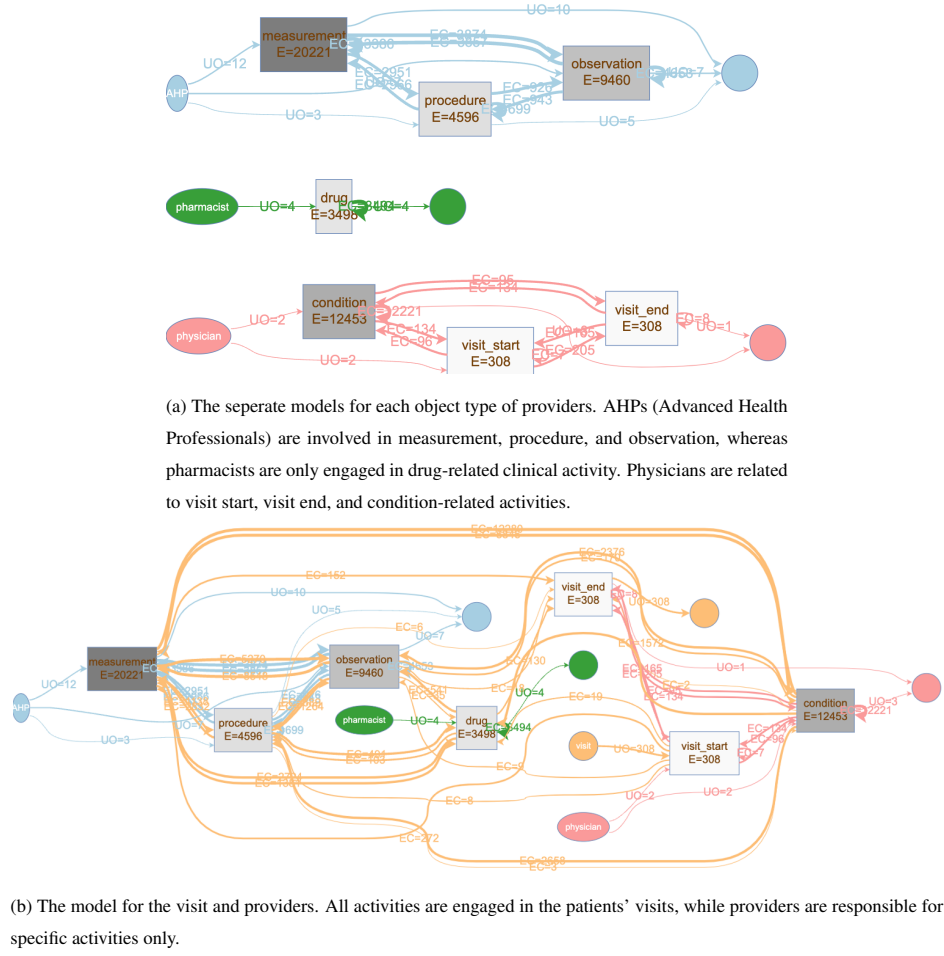
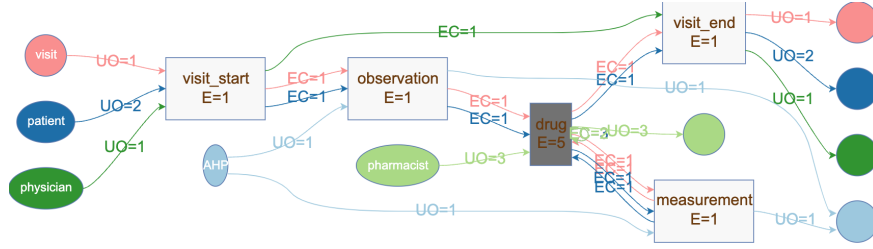


Figure 4: The object-centric process models derived from OCEL-S using object type filtering. (a) focuses exclusively on the provider-related object types, whereas (b) encompasses the visit and providers.

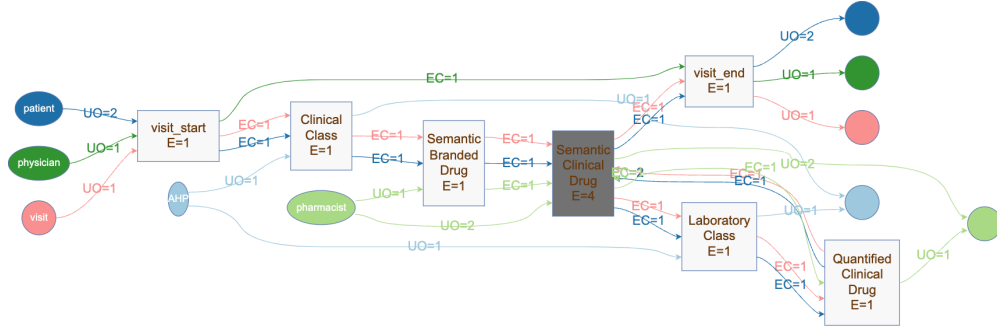
To reduce the complexity, we filter object types in various ways and analyze the corresponding process models. First, Figure 4(a) depicts the individual flow for each provider-related object type. Specifically, the model comprises AHPs, pharmacists, and physicians, each playing a unique role within the clinical process. Consequently, the single model is divided into three distinct child models to reflect these roles. Pharmacists and physicians are each focused on a single activity—drug and condition, respectively (aside from visit start and end), whereas AHPs engage in multiple activities. This distinction allows for further categorization of AHPs into specialists and generalists based on their range of activities. Second, Figure 4(b) includes these three object types and the visit. In this model, we can determine which providers are involved in the sequential activities of patient visits. This representation clearly illustrates how individual object types operate within the broader context of a visit.

Next, we compare process models from different abstraction levels. Figure 5 shows the process models for a single

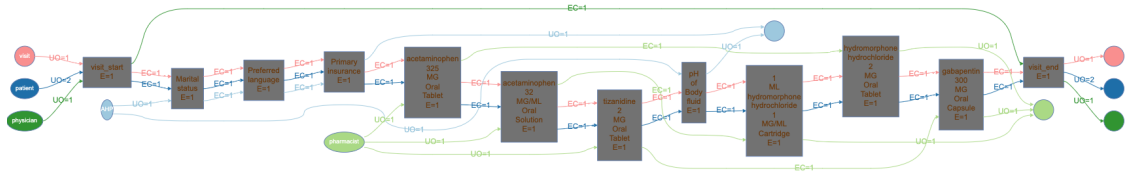
process instance from OCEL-S. The model becomes more complex as we transition from high-level to low-level activities due to the increasing diversity of activities at the lower levels. Specifically, we observed that the drug activity at the high level becomes semantic or quantified clinical drugs at the middle level and eventually narrows down to the specific orders, e.g., acetaminophen and tizanidine, at the low level. Furthermore, loops are discovered at the high and middle levels, such as the self-loop involving a drug and the loop between drug and measurement activities in Figure 5(a), contrasting with the low level where activities are arranged sequentially. Therefore, observable structural differences in the models emerge alongside variations in complexity, which stem from the differing number of activities across abstraction levels.



(a) The high-level OCPM for a single case. This single case features a sequence of activity records that includes visit start, followed by observation, drug, measurement, and concludes with visit end.



(b) The middle-level OCPM for a single case. Compared to (b), each activity in the model is mapped to a middle-level activity. Notably, drug-related activities are categorized into three groups: semantic branded drugs, semantic clinical drugs, and quantified clinical drugs.



(c) The low-level OCPM for a single case. Compared to (a) and (b), all activities in the model are mapped with fine-grained order activities, allowing for detailed identification of clinical examinations, drugs, and other specific clinical actions.

Figure 5: The OCPMs with the different levels of activities for a single case from OCEL-S

## 6. Discussion

This work introduces a method to convert an OMOP Common Data Model (OMOP CDM) into an Object-Centric Event Log (OCEL), a crucial pivot toward harnessing OCPM’s potential. The method consists of two phases: 1) creating an OMOP CDM4PM and 2) transforming the OMOP CDM4PM into an OCEL. We demonstrated the effectiveness of the method through its successful application on the MIMIC-IV, converting it to an OCEL.

The evaluation presented in the prior section underscores the successful transformation of the well-known MIMIC-IV as an OMOP CDM to an OCEL. Moreover, the object-centric process models discovered from the transformed OCELS provide reliable representations of the underlying healthcare processes. These process models, illustrated as directly-follow graphs, provide transparency over the healthcare process, showing the interplay between patient-related, provider-related, and device-related object types.

Our evaluation has highlighted the intrinsic complexity pervasive in healthcare processes. The models derived from all three datasets reveal this intricacy, attributable to factors such as the engagement of numerous object types and the varying abstraction of activities. A closer examination of specific object types, like those related to providers, reveals valuable insights into the distinct workflows of various healthcare practitioners. Additionally, we have studied the variations in observed behaviors by adjusting the abstraction levels.

Our work has important implications from both academic and practical viewpoints. First, our approach is rooted in the standard healthcare data model, particularly leveraging the OMOP CDM. Given the current trend of healthcare organizations increasingly adopting the OMOP CDM model, our methodology stands to be highly relevant and applicable. This compatibility with an industry-standard augments the potential of our model for widespread deployment and utilization. Second, object-centric process mining has emerged as a focal point of interest both in academic circles and industry practices. The successful demonstration and evaluation of our proposed approach contribute significantly to this emerging field, offering a potential blueprint for future endeavors in similar application domains. Finally, one of the standout aspects of process mining is the abundance of tooling support available. These tools are incredibly user-friendly, provided the data is on hand. Given our approach’s adaptability to standard healthcare data formats, it fits perfectly into this ecosystem, ensuring ease of use and a smoother adoption curve for practitioners.

A notable limitation in our evaluation is the absence of discussions with domain experts. Such interactions typically enrich interpretations, as experts often provide context, nuance, and practical implications. Future endeavors should actively seek and incorporate domain-specific insights to enhance the reliability and applicability of results. Moreover, our approach does not provide guidance in determining the optimal level of event type abstractions. There exists an intrinsic trade-off between complexity and detail. While high-level abstractions provide streamlined views and reduced complexity, they might inadvertently sacrifice crucial details that low-level data offers. A meaningful direction for future work would be to devise strategies for pinpointing appropriate abstraction levels, particularly designed for the healthcare domain. Furthermore, ensuring clarity in data preprocessing is crucial, particularly regarding data missingness and inaccuracy. Although preprocessing details are beyond the scope of this study, we addressed

tasks such as handling duplicate events based on attributes like activities, timestamps, and objects; managing codes not mapped to established vocabularies; and managing events outside the designated visit range, i.e., before the visit starts or after it ends. Given the significant implications of data cleansing in real medical datasets, establishing a formalized process is essential. Also, while our approach has shown promise in MIMIC-IV, its scalability and adaptability to diverse healthcare scenarios remain to be rigorously tested. Different healthcare institutions might have varying workflows, protocols, and patient demographics, which could introduce new layers of complexity not addressed in our current study. To resolve these issues, we plan to include collaboration with the OHDSI community or other medical institutions in future work.

### **CRedit authorship contribution statement**

**Gyunam Park:** Conceptualization, Methodology, Software, Writing - original draft. **Yaejin Lee:** Data curation, Software, Investigation. **Minsu Cho:** Conceptualization, Methodology, Supervision, Writing – original draft, review & editing.

### **Declaration of competing interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### **Acknowledgements**

This work was supported by the National Research Foundation of Korea(NRF) grant funded by Ministry of Science and ICT (No. NRF-2021R1G1A1094019) and the Research Grant of Kwangwoon University in 2023.

### **References**

- [1] W. M. P. van der Aalst, Process mining, *Communications of the ACM* 55 (8) (2012) 76–83.
- [2] W. M. P. van der Aalst, Process mining: a 360 degree overview, in: *Process Mining Handbook*, Springer, 2022, pp. 3–34.
- [3] N. Martin, J. D. Weerdt, C. Fernández-Llatas, A. Gal, R. Gatta, G. Ibáñez, O. A. Johnson, F. Mannhardt, L. Marco-Ruiz, S. Mertens, J. Munoz-Gama, F. Seoane, J. Vanthienen, M. T. Wynn, D. B. Boilève, J. Bergs, M. Joosten-Melis, S. Schretlen, B. B. V. Acker, Recommendations for enhancing the usability and understandability of process mining in healthcare, *Artificial Intelligence in Medicine* 109 (2020) 101962.
- [4] E. Rojas, J. Munoz-Gama, M. Sepúlveda, D. Capurro, Process mining in healthcare: A literature review, *Journal of biomedical informatics* 61 (2016) 224–236.
- [5] E. De Roock, N. Martin, Process mining in healthcare—an updated perspective on the state of the art, *Journal of biomedical informatics* 127 (2022) 103995.
- [6] R. S. Mans, M. Schonenberg, M. Song, W. M. P. van der Aalst, P. J. Bakker, Application of process mining in healthcare—a case study in a dutch hospital, in: *Biomedical Engineering Systems and Technologies: International Joint Conference, BIOSTEC 2008 Funchal, Madeira, Portugal, January 28–31, 2008 Revised Selected Papers 1*, Springer, 2009, pp. 425–438.

- [7] K. Park, M. Cho, M. Song, S. Yoo, H. Baek, S. Kim, K. Kim, Exploring the potential of omop common data model for process mining in healthcare, *Plos one* 18 (1) (2023) e0279641.
- [8] H. Baek, M. Cho, S. Kim, H. Hwang, M. Song, S. Yoo, Analysis of length of hospital stay using electronic health records: A statistical and data mining approach, *PloS one* 13 (4) (2018) e0195901.
- [9] K. Chen, F. Abtahi, J.-J. Carrero, C. Fernandez-Llatas, F. Seoane, Process mining and data mining applications in the domain of chronic diseases: A systematic review, *Artificial Intelligence in Medicine* (2023) 102645.
- [10] J. Munoz-Gama, N. Martin, C. Fernandez-Llatas, O. A. Johnson, M. Sepúlveda, E. Helm, V. Galvez-Yanjari, E. Rojas, A. Martinez-Millana, D. Aloini, et al., Process mining for healthcare: Characteristics and challenges, *Journal of Biomedical Informatics* 127 (2022) 103994.
- [11] M. Cho, M. Song, S. Yoo, A systematic methodology for outpatient process analysis based on process mining, in: *Asia Pacific Business Process Management: Second Asia Pacific Conference, AP-BPM 2014, Brisbane, QLD, Australia, July 3-4, 2014. Proceedings 2*, Springer, 2014, pp. 31–42.
- [12] F. Gonzalez-Lopez, N. Martin, R. de la Fuente, V. Galvez-Yanjari, J. Guzmán, E. Kattan, M. Sepúlveda, J. Munoz-Gama, Prodem: A process-oriented delphi method for systematic asynchronous and consensual surgical process modelling, *Artificial Intelligence in Medicine* 135 (2023) 102426.
- [13] W. Chomyat, W. Premchaiswadi, Process mining on medical treatment history using conformance checking, in: *2016 14th International Conference on ICT and Knowledge Engineering (ICT&KE)*, IEEE, 2016, pp. 77–83.
- [14] E. Rojas, M. Sepúlveda, J. Munoz-Gama, D. Capurro, V. Traver, C. Fernandez-Llatas, Question-driven methodology for analyzing emergency room processes using process mining, *Applied Sciences* 7 (3) (2017) 302.
- [15] K. Baker, E. Dunwoodie, R. G. Jones, A. Newsham, O. Johnson, C. P. Price, J. Wolstenholme, J. Leal, P. McGinley, C. Twelves, et al., Process mining routinely collected electronic health records to define real-life clinical pathways during chemotherapy, *International journal of medical informatics* 103 (2017) 32–41.
- [16] W. M. P. van der Aalst, Object-centric process mining: Dealing with divergence and convergence in event data, in: P. C. Ölveczky, G. Salaün (Eds.), *Software Engineering and Formal Methods - 17th International Conference, SEFM 2019, Oslo, Norway, September 18-20, 2019, Proceedings, Vol. 11724 of Lecture Notes in Computer Science*, Springer, 2019, pp. 3–25.
- [17] G. Hripsak, J. D. Duke, N. H. Shah, C. G. Reich, V. Huser, M. J. Schuemie, M. A. Suchard, R. W. Park, I. C. K. Wong, P. R. Rijnbeek, et al., Observational health data sciences and informatics (ohdsi): opportunities for observational researchers, *Studies in health technology and informatics* 216 (2015) 574.
- [18] E. Batista, A. Solanas, Process mining in healthcare: A systematic review, in: N. G. Bourbakis, G. A. Tsihrintzis, M. Virvou (Eds.), *9th International Conference on Information, Intelligence, Systems and Applications, IISA 2018, Zakynthos, Greece, July 23-25, 2018*, IEEE Computer Society, 2018, pp. 1–6.
- [19] S. J. Leemans, E. Poppe, M. T. Wynn, Directly follows-based process mining: Exploration & a case study, in: *2019 International Conference on Process Mining (ICPM)*, IEEE, 2019, pp. 25–32.
- [20] A. W. Kempa-Liehr, C. Y. Lin, R. Britten, D. Armstrong, J. Wallace, D. Mordaunt, M. O’Sullivan, Healthcare pathway discovery and probabilistic machine learning, *Int. J. Medical Informatics* 137 (2020) 104087.
- [21] A. Najjar, D. Reinharz, C. Girouard, C. Gagné, A two-step approach for mining patient treatment pathways in administrative healthcare databases, *Artif. Intell. Medicine* 87 (2018) 34–48.
- [22] L. Chiudinelli, A. Dagliati, V. Tibollo, S. Albasini, N. Geifman, N. Peek, J. H. Holmes, F. Corsi, R. Bellazzi, L. Sacchi, Mining post-surgical care processes in breast cancer patients, *Artif. Intell. Medicine* 105 (2020) 101855.
- [23] Z. Huang, W. Dong, L. Ji, L. Yin, H. Duan, On local anomaly detection and analysis for clinical pathways, *Artif. Intell. Medicine* 65 (3) (2015) 167–177.
- [24] G. Geleijnse, H. Aklecha, M. Vroling, R. Verhoeven, F. N. van Erning, P. A. Vissers, J. C. A. M. Buijs, X. A. Verbeek, Using process mining to evaluate colon cancer guideline adherence with cancer registry data: a case study, in: *AMIA 2018, American Medical Informatics Association Annual Symposium, San Francisco, CA, November 3-7, 2018*, AMIA, 2018.

- [25] W. M. P. van der Aalst, Object-centric process mining: Unraveling the fabric of real processes, *Mathematics* 11 (12) (2023) 2691.
- [26] A. F. Ghahfarokhi, G. Park, A. Berti, W. M. P. van der Aalst, Ocel: A standard for object-centric event logs, in: *European Conference on Advances in Databases and Information Systems*, Springer, 2021, pp. 169–175.
- [27] M. Q. Stearns, C. Price, K. A. Spackman, A. Y. Wang, Snomed clinical terms: overview of the development process and project status., in: *Proceedings of the AMIA Symposium*, American Medical Informatics Association, 2001, p. 662.
- [28] A. E. Johnson, L. Bulgarelli, L. Shen, A. Gayles, A. Shammout, S. Horng, T. J. Pollard, S. Hao, B. Moody, B. Gow, et al., Mimic-iv, a freely accessible electronic health record dataset, *Scientific data* 10 (1) (2023) 1.
- [29] N. Paris, A. Lamer, A. Parrot, Transformation and evaluation of the mimic database in the omop common data model: development and usability study, *JMIR Medical Informatics* 9 (12) (2021) e30970.
- [30] M. G. Kahn, T. J. Callahan, J. Barnard, A. E. Bauck, J. Brown, B. N. Davidson, H. Estiri, C. Goerg, E. Holve, S. G. Johnson, et al., A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data, *Egms* 4 (1) (2016).
- [31] S. J. van Zelst, F. Mannhardt, M. de Leoni, A. Koschmider, Event abstraction in process mining: literature review and taxonomy, *Granular Computing* 6 (2021) 719–736.

## Appendix A. Example Tables in OMOP CDM

Table A.4 presents a fragment of *procedure\_occurrence* table in OMOP CDM.

Table A.4: A fragment of *procedure\_occurrence* table

| Procedure_Occurrence_ID | Person_ID | Concept_ID | DateTime            | Provider_ID | Visit_Occurrence_ID | Quantity | .... |
|-------------------------|-----------|------------|---------------------|-------------|---------------------|----------|------|
| 1277                    | 1         | 4127451    | 2023-01-14 00:00:00 | 514         | 740                 | 0        | ...  |
| 1278                    | 2         | 1989694    | 2023-01-14 03:00:00 | 514         | 750                 | 1        | ...  |
| 1279                    | 3         | 45929955   | 2023-01-14 06:00:00 | 515         | 770                 | 0        | ...  |
| ...                     | ...       | ...        | ...                 | ...         | ...                 | ...      | ...  |

Table A.5 depicts a fraction of *provider* table from OMOP CDM.

Table A.5: A fraction of a *provider* table

| Provider_ID | Concept_ID | Care_Site_ID | NPI        | DEA    | ... |
|-------------|------------|--------------|------------|--------|-----|
| 514         | 32577      | 101          | 0495058675 | 059684 | ... |
| 515         | 32415      | 101          | 9683910487 | 604956 | ... |
| 516         | 33005      | 102          | 5038174052 | 603948 | ... |
| ...         | ...        | ...          | ...        | ...    | ... |

Table A.6 presents a fraction of *concept* table from OMOP CDM.

## Appendix B. Event Type Assignment

Figure B.6 shows SNOMED clinical terms hierarchy that can be used for deriving event type assignments.

Table A.6: A fraction of *concept* table

| Concept_ID | Domain_ID | Concept_Class_ID     | Concept_Name       | Valid_Start_Date | Valid_End_Date | ... |
|------------|-----------|----------------------|--------------------|------------------|----------------|-----|
| 32577      | Provider  | Physician Specialty  | Physician          | 2010-01-01       | 2030-01-01     | ... |
| 1989694    | Procedure | Clinical Observation | Cone beam CT Teeth | 2010-01-01       | 2030-01-01     | ... |
| ...        | ...       | ...                  | ...                | ...              | ...            | ... |

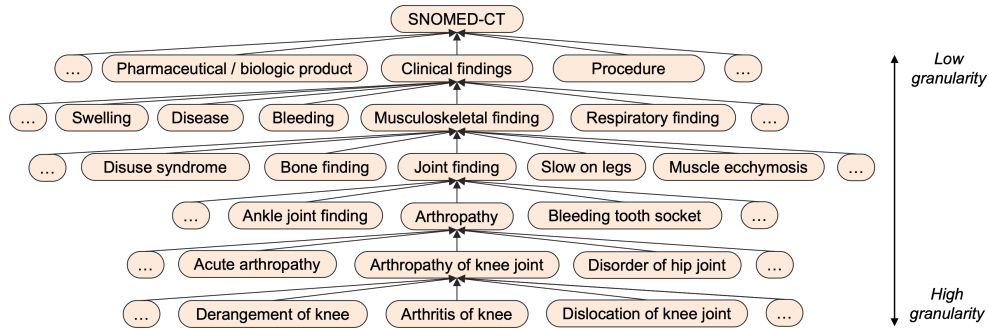


Figure B.6: A partial example of SNOMED clinical terms hierarchy. It refers to clinical activities with higher granularity towards the top and lower granularity towards the bottom.

## Appendix C. Evaluation

Table C.7 shows the results for the completeness of our transformation method for the three transformed OCEs used in the evaluation.

Table C.7: The detailed results for the completeness of our transformation method for three transformed OCELS

|         |                                                       | OCEL-S   |          | OCEL-IP   |          | OCEL-ED   |          |
|---------|-------------------------------------------------------|----------|----------|-----------|----------|-----------|----------|
|         |                                                       | Quantity | Rate (%) | Quantity  | Rate (%) | Quantity  | Rate (%) |
| Events  | High-level Activities                                 | 52,104   | 100.00   | 74,250    | 100.00   | 70,959    | 100.00   |
|         | Visit Start                                           | 378      | 0.73     | 1,141     | 1.54     | 1,155     | 1.63     |
|         | Condition                                             | 12,735   | 24.44    | 2,662     | 3.59     | 2,936     | 4.14     |
|         | Drug                                                  | 3,914    | 7.51     | 19,925    | 26.84    | 18,045    | 25.43    |
|         | Measurement                                           | 20,328   | 39.01    | 32,802    | 44.18    | 30,931    | 43.59    |
|         | Observation                                           | 9,751    | 18.71    | 3,586     | 4.83     | 4,689     | 6.61     |
|         | Procedure                                             | 4,620    | 8.87     | 12,993    | 17.50    | 12,048    | 16.98    |
|         | Visit End                                             | 378      | 0.73     | 1,141     | 1.54     | 1,155     | 1.63     |
|         | Middle-level Activities                               | 68,804   | 100.00   | 133,143   | 100.00   | 123,155   | 100.00   |
|         | Clinical Finding                                      | 12,896   | 18.74    | 3,217     | 2.42     | 3,512     | 2.85     |
|         | Clinical Class                                        | 21,029   | 30.56    | 44,802    | 33.65    | 40,233    | 32.67    |
|         | Laboratory Class                                      | 6,935    | 10.08    | 30,730    | 23.08    | 27,705    | 22.50    |
|         | Observable Entity                                     | 5,589    | 8.12     | 164       | 0.12     | 404       | 0.33     |
|         | Surveys                                               | 3,430    | 4.99     | 1,287     | 0.97     | 1,659     | 1.35     |
|         | Quantified Clinical Drug                              | 2,636    | 3.83     | 7,941     | 5.96     | 6,810     | 5.53     |
|         | Semantic Clinical Drug                                | 2,348    | 3.41     | 11,760    | 8.83     | 10,672    | 8.67     |
|         | Others                                                | 13,941   | 20.26    | 33,242    | 24.97    | 32,160    | 26.11    |
|         | Low-level Activities                                  | 364,532  | 100.00   | 1,650,306 | 100.00   | 1,580,208 | 100.00   |
|         | Respiratory rate                                      | 19,234   | 5.28     | 114,879   | 6.96     | 112,753   | 7.14     |
|         | Heart rate                                            | 18,502   | 5.08     | 113,919   | 6.90     | 111,465   | 7.05     |
|         | Oxygen saturation in Arterial blood by Pulse oximetry | 17,607   | 4.83     | 111,317   | 6.75     | 109,424   | 6.92     |
|         | Mean blood pressure                                   | 9,639    | 2.64     | 46,579    | 2.82     | 36,396    | 2.30     |
|         | Systolic blood pressure                               | 9,560    | 2.62     | 46,493    | 2.82     | 36,284    | 2.30     |
|         | Diastolic blood pressure                              | 9,559    | 2.62     | 46,486    | 2.82     | 36,266    | 2.30     |
|         | Body temperature                                      | 4,031    | 1.11     | 38,697    | 2.34     | 34,402    | 2.18     |
|         | Others                                                | 276,400  | 75.82    | 1,131,936 | 68.59    | 1,103,218 | 69.81    |
| Objects | Patient-related Object Types                          | 478      | 100.00   | 2,229     | 100.00   | 2,068     | 100.00   |
|         | Patients                                              | 100      | 20.92    | 1,088     | 48.81    | 913       | 44.15    |
|         | Visits                                                | 378      | 79.08    | 1,141     | 51.19    | 1,155     | 55.85    |
|         | Provider-related Object Types                         | 30       | 100.00   | -         | -        | -         | -        |
|         | Physicians                                            | 4        | 13.33    | -         | -        | -         | -        |
|         | Advanced Health Professionals                         | 22       | 73.33    | -         | -        | -         | -        |
|         | Pharmacists                                           | 4        | 13.33    | -         | -        | -         | -        |
|         | Device-related Object Types                           | 33       | 100.00   | 185       | 100.00   | 54        | 100.00   |
|         | Dermal Drug Delivery                                  | -        | -        | 28        | 15.14    | 27        | 50.00    |
|         | Injection                                             | 11       | 33.33    | 157       | 84.86    | 26        | 48.15    |
|         | Diagnostic and Supplements                            | 22       | 66.67    | -         | -        | 1         | 1.85     |