

A Semantic Encoding of Object Centric Event Data

Saba Latif¹, Fajar J. Ekaputra², Maxim Vidgof²,
Sabrina Kirrane², and Claudio Di Ciccio³

¹ Sapienza University of Rome, Rome, Italy, saba.latif@uniroma1.it

² Wirtschaftsuniversität Wien, Vienna, Austria, firstname.lastname@wu.ac.at

³ University of Utrecht, Utrecht, The Netherlands, c.diciccio@uu.nl

Abstract. The Object-Centric Event Data (OCED) is a novel meta-model aimed at providing a common ground for process data records centered around events and objects. One of its objectives is to foster both interoperability and process information exchange. In this context, the integration of data from different providers, the combination of multiple processes, and the enhancement of knowledge inference are novel challenges. Semantic Web technologies can enable the creation of a machine-readable OCED description enriched through ontology-based relationships and entity categorization. In this paper, we introduce an approach built upon Semantic Web technologies for the realization of semantic-enhanced OCED, with the aim to strengthen process data reasoning, interconnect information sources, and boost expressiveness.

Keywords: Object centric event log · OCED · Semantic web · Ontology · Knowledge graphs · Process mining

1 Introduction

Process mining is the discipline aimed at extracting, analyzing, and enhancing knowledge of business processes from event data stored by information systems in the form of event logs [2]. Over the last few years, the focus of process mining has experienced a gradual drift from the historically established activity-centric view, interpreting process execution logs as sequences of actions [34]. The spotlight is moving towards the information artifacts, namely *objects*, that activity executions alter or read. This viewpoint shift is testified by the surge of object-centric process mining [3] and object-centric event log formats [24].

The IEEE Task Force on Process Mining leads the standardization process for a new event data paradigm, with the aim to overcome the previous established format of XES (eXtensible Event Stream, [34]): The Object Centric Event Data (OCED) meta-model [22]. An accurate semantic ontology of the novel meta-model is still under discussion. Ontologies function as explicit conceptual models representing domain knowledge, making it accessible to information systems. They are essential to the vision of the semantics, offering the semantic vocabulary needed to annotate websites so that machines can meaningfully interpret [25].

In this paper, we provide a framework that defines the semantics of OCED to endow the new standard with a consistent, extensible, and interoperable representation. We

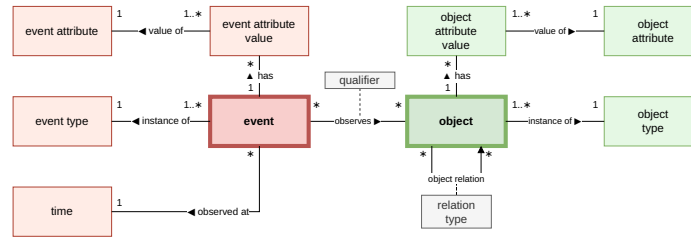


Fig. 1: The OCED meta-model core [22]

achieve this by leveraging the body of knowledge of the semantic web [11]. On top of that, we introduce an approach inspired by the literature in information integration [15] that resorts to semantic web description languages to establish links between the model at the meta-level (described by the OCED Ontology, OCEDO), customizable and domain-specific representations at an intentional level (via Domain-specific extensions of OCED, OCEDD), and the data at an extensional level (as knowledge graphs of OCED Resources, OCEDR). We provide a demonstration of how raw process data can be automatically endowed with machine-readable semantics in an object-oriented fashion through a proof-of-concept prototype that automatically extracts OCEDR knowledge graphs from activity-centric XES logs. Our semantic framework encoding and tools are openly available at github.com/wu-semsys/ocedo.

Our long-term goal is manifold. Among others, the foreseeable opportunities opened up by our approach range from an improvement of OCED data reasoning, to the knowledge inference from semantically enriched data, from an increase in expressiveness, to the interconnection of information sources. Regarding the last point, we remark that under our semantic umbrella, other (not necessarily process-specific) existing datasets can be linked with the resulting potential of unleashing unprecedented contextual enrichment for (semantic reasoning in) process mining.

The remainder of the paper is as follows. [Section 2](#) outlines the notions and technical bases of our investigation. [Section 3](#) proposes our three-layered conceptual framework for the semantic encoding of OCED. [Section 4](#) discusses the implications of our endeavor. [Section 5](#) overviews related work. [Section 6](#) draws conclusions for this work.

2 Background

Our investigation aims to forge semantic layers for object-centric event data using the OCED standard as the anvil, and the conceptual tools and frameworks provided by semantic web technologies as the hammer. In this section, we provide a description of our conceptual forgery's items.

2.1 The OCED Meta Model

OCED aims to overcome the limitations of the well-established IEEE standard XES [1], which emerged in its ultra-decennial adoption at the time of writing. These limitations

include lack of generalizability, complexity of the data structure, and memory expensive-ness of the storage format [34]. Unlike XES, OCED shifts the focus from sequences of activities recorded in information systems, trace after trace, to the lifecycle and conceptual interconnection of business objects that these systems handle.

A clear example of how semantically-aware object-centric formats can be useful to the representation of process execution records comes a public real-world event log tracing the process for incident and problem management at Volvo [33]. Although the event log is natively stored in XES, it comes endowed with several pieces of information pertaining to the treated business objects, and rich documentation for the data representation and contents.⁴ Furthermore, it demonstrates how raw, activity-centric event logs can be turned into semantically rich, object-centric information stores. We shall use it as a running example in this paper, revisited under the lens of object-centric representations.

Figure 1 graphically depicts the key elements of the OCED meta-model and their interrelation. Events and objects (marked with a thick bounding box in the figure) represent the first-class citizens of OCED. An *event* represents a point-in-time occurrence of an action. The *objects* are entities of which the event may report the creation, change, deletion, or mere reading (i.e., that the object *observes* in a way that the *qualifier* clarifies). For example, every status change of an incident (e.g., the one identified by the ticket number 1-364285768) is an event (e.g., signaling that it closes the incident by acquiring a *Completed* status and *Closed* sub-status [32]). The reported incident, and hence the event, pertains to a product (e.g., PROD582). The status change is operated by a responsible (e.g., Siebel) of a support team (e.g., V5 3rd) within an IT function division (e.g., the one of A2.5). of a service center (e.g., the one of Org line A2). The incident, the product, the responsible, and the team are all objects in OCED. Notice that there are relations between two objects in this example (Siebel works in a team, which is part of an IT function division, which is in turn within a service center). In OCED, the concept of *object relation*, labelled by a *relation type* (*works in*, *is part of* and *is within*, in our example) bears this notion.

Both events and objects can come endowed with attributes. In OCED, attributes are akin to name-value pairs; therefore, an object (like the incident) is associated to an *object attribute value* for an *object attribute* (e.g., *High* for the *Impact* of the problem), and an event (like the status change) is associated to an *event attribute value* for an *event attribute* (e.g., *Completed* for the new incident's *status* and *Closed* for its *sub_status*). A *time-stamp* is a special attribute since, unlike the other attributes, it has fixed semantics: In particular, it indicates when the event was registered (e.g., 2012-05-11T01:26:15+02:00).

2.2 Semantic Web Technologies

Semantic web technologies provide a comprehensive set of standards for the representation, linking, and processing of semantically explicit information. For our intents and purposes, of particular relevance in the semantic web technology stack is the Resource Description Framework (RDF), which we use as a uniform model to represent object-centric event logs and the information they bear. RDF is a data representation model published by the World Wide Web Consortium (W3C) as a set of recommendations and working

⁴<https://ais.win.tue.nl/bpi/2013/challenge.html>. Accessed: 26/09/2025.

group notes.⁵ It provides a standard model for expressing information about *resources*, which in our paper represent events, objects, attributes, relations, etc. An RDF dataset consists of a set of statements about these resources, expressed in the form of triples (s, p, o) where s is an RDF-*subject*, p is a *predicate*, and o is an RDF-*object*; s and o represent the two resources being related whereas p represents the nature of their relationship. For instance, with RDF we can declare that ev_1 (RDF-*subject*) is *observed at* (predicate) 2012-05-11T01:26:15+02:00 (RDF-*object*). A key advantage of RDF as a data model is its extensible nature, i.e., additional statements about s , p , and o can be added to link concepts and predicates from various additional, potentially domain-specific vocabularies. To this end, the encoding of a resource is associated to a Uniform Resource Identifier (URI), and grouped into namespaces that are used as prefixes to clarify the vocabulary they belong to. The notion of RDF-object itself, for instance, is encoded as `rdf:object`, whereby the namespace `rdf` is associated to the URI `http://www.w3.org/1999/02/22-rdf-syntax-ns#`. New vocabularies can be defined and connected.

Notice that RDF can cross the boundaries of information integration levels. The triple $(ev_1, \textit{observed at}, 2011-02-03T08:28:58+01:00)$ is exerted on RDF-*subject* and RDF-*object* at the extensional level, whilst *observed at* is a meta-model element. We can also claim that ev_1 is an event, having the RDF-*object* at the meta-level. The RDF Schema (RDF/S)⁶ is built upon the RDF vocabulary and predicates over resources, classes categorizing resources, and the relations among classes. For instance, with RDF/S we can express that *observes* relates events (*domain*) to objects (*range*). It is thus possible to automatically deduce that if $(ev_1, \textit{observes}, obj_1)$ is a declared triple, then ev_1 is an event and obj_1 an object. Ontology specification languages such as the Web Ontology Language (OWL)⁷ can be used to more closely describe the semantic characteristics of terms in use. For example, with RDF and OWL we can clarify that *event* is a `owl:Class`, and that *observed at* is an `owl:DatatypeProperty`. RDF can be serialized via different formats, including RDF/XML, JSON-LD, and the text-based Terse RDF Triple Language (Turtle).⁸ Without loss of generality, we adopt Turtle in the remainder of the paper due to its compactness.

A clear benefit of the aforementioned technologies is interoperability: information expressed in RDF using shared vocabularies and ontologies can be exchanged between applications without loss of meaning. Furthermore, it makes it possible to apply a wide range of general purpose RDF parsing, mapping, transformation, and query processing tools. Once transformed into RDF, information assumes the form of *Knowledge Graphs* that can be published online, interlinked, and shared between applications and organizations, which is particularly interesting in the context of collaborative processes. Next, we will see how we use semantic web technologies as a building block for a semantic encoding of OCED that allows for an explicit stratification of domain-specific intensional layers over the meta-model.

⁵<http://www.w3.org/TR/rdf11-primer/>. Accessed: 30/09/2025.

⁶RDF/S: www.w3.org/TR/rdf-schema. Accessed: 30/09/2025.

⁷OWL: www.w3.org/TR/owl2-primer. Accessed: 30/09/2025.

⁸RDF/XML: [w3.org/TR/rdf-syntax-grammar](http://www.w3.org/TR/rdf-syntax-grammar); JSON-LD: [w3.org/TR/json-ld](http://www.w3.org/TR/json-ld); Turtle: [w3.org/TR/turtle](http://www.w3.org/TR/turtle). Accessed: 30/09/2025.

3 A Semantic Stratification of OCED

The OCED meta-model allows for flexibility in data representation, as it is meant to be *domain-agnostic*. Notice, e.g., that no restrictions are exerted on the possible relations that an object of a given type may or may not have with other objects, the values that attributes can take, or what (types of) objects are observed by which events. For example, OCED does not natively provide users with means to enforce that each incident pertains to a product, enumerate what the possible status changes can be, or indicate that every status change event affects an incident. It is also worth noticing that this meta-model removes a constraint that activity-centric standards like XES exerted, i.e., that every event be associated to a case, namely a process instance [31,10]. In the OCED meta-model, the case is not considered as a relevant, let alone mandatory, concept –however, it can be represented as an object.

While this structural looseness is desirable to encompass the amplest plethora of process data in a meta-model, domain experts should be allowed to inject knowledge on top of the meta-model, to provide additional details on the structure of the recorded data, promote traceability of process information provenance, and help the linkage of different data sources [16,14,30]. With this paper, we thus advocate the adoption of an overarching ontological approach to cover three conceptual layers, following a well-known approach in the field of information integration [15]: From the *meta-level* (represented by the *OCED* Ontology, OCEDO) to the *intensional level* (the *OCED* Domain-specific extension, OCEDD) to the *extensional level* (the *OCED* Resources' knowledge graph, OCEDR). To define and interconnect the three levels, we resort to semantic web technologies, as we describe in the remainder of this section.

3.1 The Meta-Level: OCED Ontology (OCEDO)

Figure 2 depicts our reference implementation of OCED's meta-model and intensional level in the form of an RDF semantic network. The OCED original meta-model entities and relations from Fig. 1 occur below in the figure, as a reference. With dashed blue lines, we link those to the corresponding semantic encodings in our new OCED Ontology (OCEDO, in the middle). OCEDO is split in two parts. *OCEDO Core* (prefixed with *oced*, namespace <https://w3id.org/ocedo/core#>) encompasses the first-class citizens of OCED, namely *oced:Event* and *oced:Object*.⁹ *OCEDO Auxiliary* (*aux* prefix, namespace <https://w3id.org/ocedo/aux#>) contains the representations of the other entities and relations in the original meta-model, namely the event type, event attribute, object type, object attribute, and the reified relations between objects, and from events to objects (*observe*).

Here we provide a brief overview of the OCED meta-model's encoding with semantic web technologies. A full specification is available in our open code repository.¹⁰ Listing 1 shows an excerpt of the OCEDO schema serialized in Turtle. In the listing, we colored the background of keywords compatibly with the color scheme of Fig. 2 for understandability

⁹Notice that the prefix linked to URIs allows for disambiguation: The notion of the *rdf:object*, defined by W3C, does not match that of an *oced:Object*.

¹⁰<https://semsys.ai.wu.ac.at/ocedo>. Accessed: 30/09/2025.

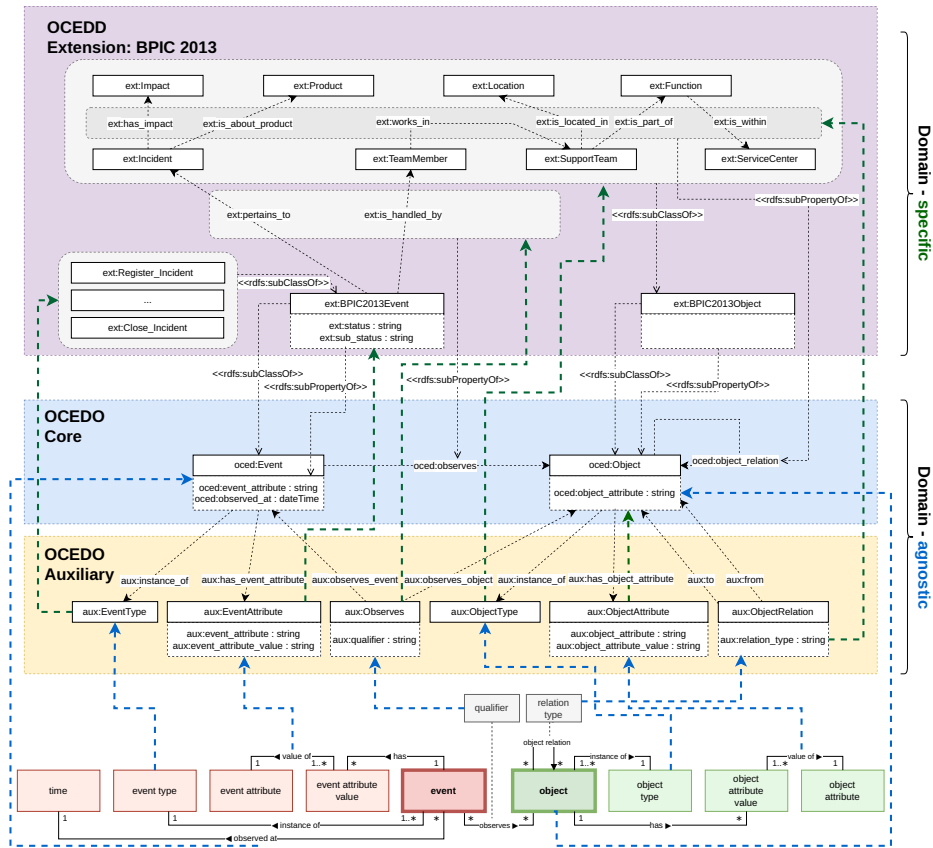


Fig. 2: OCED meta-model with RDF representation and additional domain-specific layer

purposes. After the import of existing ontologies via prefixes and namespaces (like the aforementioned RDF, RDF/S, OWL, see Sect. 2.2) we declare that what is preceded by `oced:` and `aux:` identifies concepts for our core and auxiliary ontology parts of the meta-model, respectively. Thereafter, starting from line 8, we state that `oced:Event` is a class to be labeled “Event”, to be regarded as a representation of the event class (see the comment directive; both `label` and `comment` are terms of the `rdfs` vocabulary). Similarly, we introduce the concept of `oced:Object` (line 11).

Among the concepts of the auxiliary section of OCEDO, we report here that of `aux:ObjectAttribute`, and of the reified relation types `aux:ObjectRelation` and `aux:Observe` (see lines 23, 25 and 21, respectively). Notice that the mappings from objects to their attributes, from objects to the relation to other objects, and from events to the observation of objects are all defined in the form of an `owl:ObjectProperty`, specifying the `rdfs:domain` and `rdfs:range` of such mappings (see Sect. 2.2), as can be noticed on lines 28-31, 32-41, and 42-51. Since the timestamp of an event is a scalar value of a type provided by the well-established XML-schema (`xsd`) vocabulary, we

```

1 @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
2 @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
3 @prefix owl: <http://www.w3.org/2002/07/owl#> .
4 @prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
5 @prefix oced: <https://w3id.org/ocedo/core#> .
6 @prefix aux: <https://w3id.org/ocedo/auxiliary#> .
7 # [...] OCED core
8 oced:Event a owl:Class ;
9   rdfs:label "Event" ;
10  rdfs:comment "Representation of the concept of Event" .
11 oced:Object a owl:Class ;
12   rdfs:label "Object" ;
13   rdfs:comment "Representation of the concept of Object" .
14 # [...] OCED core - data properties
15 oced:observed_at a owl:DatatypeProperty ;
16   rdfs:label "observed_at" ;
17   rdfs:comment "timestamp of an event" ;
18   rdfs:domain oced:Event ;
19   rdfs:range xsd:dateTime .
20 # [...] OCED auxiliary - classes
21 aux:Observe a owl:Class ;
22   rdfs:label "Observe" .
23 aux:ObjectAttribute a owl:Class ;
24   rdfs:label "Object Attribute" .
25 aux:ObjectRelation a owl:Class ;
26   rdfs:label "Object Relation" .
27 # [...] OCED auxiliary - object properties
28 aux:has_object_attribute a owl:ObjectProperty ;
29   rdfs:label "has_object_attribute" ;
30   rdfs:domain oced:Object ;
31   rdfs:range aux:ObjectAttribute .
32 aux:from a owl:ObjectProperty ;
33   rdfs:label "from" ;
34   rdfs:domain aux:ObjectRelation ;
35   rdfs:range oced:Object ;
36   rdfs:comment "source Object" .
37 aux:to a owl:ObjectProperty ;
38   rdfs:label "to" ;
39   rdfs:domain aux:ObjectRelation ;
40   rdfs:range oced:Object ;
41   rdfs:comment "target Object" .
42 aux:observe_object a owl:ObjectProperty ;
43   rdfs:label "observe_object" ;
44   rdfs:domain aux:Observe ;
45   rdfs:range oced:Object ;
46   rdfs:comment "Relation between Observe and related Object" .
47 aux:observe_event a owl:ObjectProperty ;
48   rdfs:label "observe_event" ;
49   rdfs:domain aux:Observe ;
50   rdfs:range oced:Event ;
51   rdfs:comment "Relation between Observe and related Event" . # [...]

```

Listing 1: A Turtle representation of the OCEDO ontology (excerpt)

indicate that `observed_at` is an `owl:DatatypeProperty` and specify that the RDF/S domain and range of it are the `oced:Event` class and the `xsd:dateTime` datatype, respectively (see the block starting on lines 15-19).

3.2 The Intensional Level: Domain-specific Extension (OCEDD)

OCEDO is domain-agnostic. Regardless of the organization registering the process data, and the structure thereof, it does not vary. The members of the OCED Working Group encourage work for enriching process data semantics with domain-specific knowledge in their white paper [22]. To cater for this, we leverage the extensibility of RDF to postulate the addition of another semantic layer at the intensional level: The domain-specific extensions which we collectively name as OCEDD (henceforth prefixed with `ext`). The abstraction step is akin to the passage from the notion of entity and relationship types as modeling concepts to the conceptual schema of a relational database, with actual entities and relationships representing the business domain of data. Notice, however, that our modeling objective is descriptive and not normative like an ER diagram.

While OCEDO is meant to encode the semantics of the OCED meta model based on the directives of the XES/OCED working group [22], OCEDD is intended as a customizable addendum to create dialects of OCED that are domain- or case-specific. OCEDD extensions allow domain experts, analysts, but also data mining tools, to enrich the information brought by recorded process data with additional knowledge tailored for the situation in use in a controlled fashion, thus propelling *extensibility*. However, OCEDD classes are intended to extend the existing, domain-agnostic OCEDO counterparts, so as to foster retro-compatibility and *interoperability*.

Let us consider again the example of BPIC 2013. Some concepts therein solely pertain to the domain-specific nomenclature of the terms (like incidents, products, responsible person, support teams) and to the constraints binding the concepts the information system

```

1 @prefix ext: <https://w3id.org/ocedo/domain#> .
2 # [...] Domain-specific event and object types
3 ext:BPIC2013Event a owl:Class ;
4   rdfs:subClassOf oced:Object .
5 ext:BPIC2013Object a owl:Class ;
6   rdfs:subClassOf oced:Object .
7 ext:Incident a owl:Class ;
8   rdfs:subClassOf ext:BPIC2013Object .
9 ext:TeamMember a owl:Class ;
10  rdfs:subClassOf ext:BPIC2013Object .
11 ext:SupportTeam a owl:Class ;
12  rdfs:subClassOf ext:BPIC2013Object .
13 ext:Completed_Resolved a owl:Class ;
14  rdfs:subClassOf ext:BPIC2013Event .
15 # [...] Domain-specific event-object and object-
    object relations
16 ext:pertains_to a owl:ObjectProperty ;
17   rdfs:label "ext:pertains_to" ;
18   rdfs:domain oced:Event ;
19   rdfs:range ext:Incident .
20 ext:is_handled_by a owl:ObjectProperty;
21   rdfs:label "ext:is_handled_by" ;
22   rdfs:domain oced:Event ;
23   rdfs:range ext:TeamMember .
24 ext:works_in a owl:ObjectProperty ;
25   rdfs:label "ext:works_in" ;
26   rdfs:domain ext:TeamMember ;
27   rdfs:range ext:SupportTeam .
28 # [...] Domain-specific event-attribute relations
29 ext:status a owl:DatatypeProperty ;
30   rdfs:subPropertyOf oced:event_attribute .
31 ext:substatus a owl:DatatypeProperty ;
32   rdfs:subPropertyOf oced:event_attribute .

```

Listing 2: A Turtle representation of an OCEDD extension for BPIC 2013 (excerpt)

reports on (every incident pertains a product, and every status update thereof is handled by a responsible person who works in a support team). Therefore, they do not apply to OCED as a whole, but to the specific extension thereof to use in this case. In what follows, we provide an example of a possible domain-specific encoding of the BPIC 2013 data linked with OCEDO. Our reconstruction is tentative and based on the dataset documentation⁴ and the analysis of the challenge submissions like [32] for exemplification purposes. It is by no means intended to be comprehensive or devoid of interpretation mistakes, but serves the purpose of clarifying the rationale of OCEDD.

Listing 2 shows an excerpt of an OCEDD extension for the BPIC2013 log. An overview of it is graphically depicted in the topmost layer of Fig. 2, with dashed green arcs highlighting some implicit connections between OCEDD elements and the related OCEDO concepts below whenever entailed, but not explicitly enforced, by RDF primitives. The full encoding is publicly available in our codebase.¹⁰ Here we focus on a few salient characteristics. We begin defining a sub-class of `oced:Event` and `oced:Object` to characterize the concepts of pertinence for the specific log, here `ext:BPIC2013Event` (line 3) and `ext:BPIC2013Object` (line 5), respectively. Event types categorize the possible `ext:BPIC2013Events` via `rdfs:subClassOf` (see line 13 for the status transition of the incident to being completed and resolved). Objects of interest in this domain include `ext:Incident`, `ext:TeamMember`, and `ext:SupportTeam` (the names recall the ones mentioned in Sect. 2.1; see lines 7, 9 and 11 in Listing 2).

In our example scenario, all events are bound to an incident. We express this concept by means of the `owl:ObjectProperty` we adopted in OCEDO (see Sect. 3.1): thereby, the `pertains_to` relationship binds `oced:Events`, the domain, to an `ext:Incident` (range; see lines 16-19). Similarly, we indicate that an `oced:Event` is handled by an `ext:TeamMember` (lines 20-23), and that an `ext:TeamMember` works in an `ext:SupportTeam` (lines 24-27). Through the aforementioned `owl:DatatypeProperty`, we can associate events to given attributes like the `ext:status` and `ext:substatus` by declaring those as an `rdfs:subPropertyOf` of the `oced:event_attribute` (see lines 29-30 and 31-32, respectively).

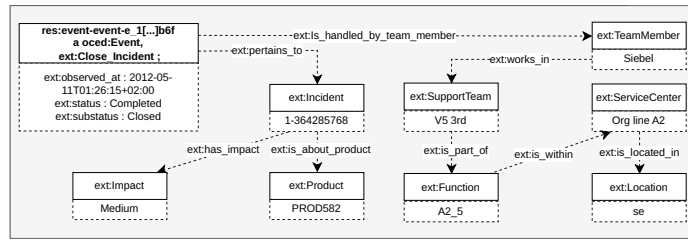


Fig. 3: A section of the OCEDD knowledge graph from BPIC2013

```

1 @prefix res: <https://w3id.org/ocedo/resource/> .
2 # [...] An event
3 res:event-e_1[...]b6f a oced:Event, ext:
  Close_Incident ;
4   oced:observed_at "2012-05-11T01:26:15+02:00"^^
  xsd:dateTime ;
5   ext:status "Completed" ;
6   ext:pertains_to res:object-o_5[...]374 ;
7   ext:is_handled_by res:object-o_e[...]011 ;
8   ext:substatus "Closed" .
9 # [...] An object of type Incident
10 res:object-o_5[...]374 a oced:Object, ext:
  Incident ;
11   ext:ticket_number "1-364285768" .
12   ext:is_about_product res:object-o_7[...]846 ;
13 # [...] An object of Type TeamMember
14 res:object-o_e[...]011 a oced:Object, ext:
  TeamMember ;
15   ext:team_member_name "Siebel" ;
16   ext:works_in res:object-o_8[...]430,
17 # [...] An object of Type SupportTeam
18 res:object-o_b[...]430 a oced:Object,
19   ext:SupportTeam ;
20   ext:team "V5 3rd" ;
21 # [...] An object of type Product
22 res:object-
  o_7cb5c41b1041659e8f4b2d1286c58459145ac846 a
  oced:Object, ext:Product ;
23   ext:product_number "PROD582" .

```

Listing 3: A Turtle representation of an OCEDR knowledge graph for the BPIC 2013 dataset (excerpt)

3.3 The Extensional Level: Resources’ (OCEDR) Knowledge Graphs

Equipped with the description of the domain-specific information about events and objects in the process data, we proceed with the third and last, extensional level. To this end, we introduce the OCEDR layers.

Figure 3 shows an excerpt of the BPIC2013 event log turned into an OCEDR knowledge graph, revolving around a specific event from our running example in Sect. 2.1. In Listing 3, we show a fragment of the corresponding Turtle encoding. In the latter, we refer to the entries instantiating the concepts in OCEDO and OCEDD generically as resources (prefixed with res). On line 3, e.g., we indicate that event-e_1[...]b6f is an event of type ext:Close_Incident. In the indented lines that follow (till 8), we specify its status and substatus attributes, its timestamp, and the relations to two more resources instantiating objects: among others, the ext:Incident it pertains to (namely the one with ticket number 1-364285768, see line 10), and the ext:TeamMember (whose name attribute is Siebel, see line 14). Some relations among objects are shown on lines 16 (to indicate that Siebel works in team V5 3rd, see line 19) and 12 (to indicate that the incident pertains to product number PROD582, see line 22).

Thus far, we have discussed with examples the core concepts underpinning our three-layered conceptual framework for the semantic description of OCED data. Next, we discuss ongoing work and future implications for its possible adoption and enhancement.

4 Potential Impact and Future Work

Our proposed framework is a first building block, on which diverse research endeavors can be based. The core contribution here is the systematization of semantic object-centric process data modeling with distinct strata: Meta-level, intensional level, and extensional level. To generate the examples in this paper, and have a preliminary proof-of-concept testing our framework, we created a tool for the automated extraction of OCEDR knowledge graphs out of flattened event logs in XES format. It is openly available at github.com/wu-semsys/ocedo. It takes as input an XES event log, an OCEDD document, and a tabular descriptor linking XES entry names to OCEDO and OCEDD concepts. The descriptor, e.g., dictates that the `org:resource` and `org:group` XES event attributes respectively contain the values of the `name` attribute of an `ext:TeamMember` in `ext:is_handled_by` relation with that `oced:Event`, and the `team` attribute of an `ext:SupportTeam` object in `ext:works_in` relation with the latter. The tool leverages RDF Mapping language (RML)¹¹ for the mapping from input to output resources. Further tests conducted with other real-world event logs in the financial and governmental domains can be found there. A description of the tool goes beyond the scope of this paper. However, it serves as a clarion call for further implementations that link not only XES but other formats including raw CSVs, relational databases, OCEL files to OCEDR graphs. The more the connectors, the higher the impact that a semantic framework can yield. Furthermore, they can allow for the linkage of heterogeneous data repositories from multiple sources, enhancing knowledge inference capabilities, and enriching expressiveness. Information integration is indeed a driving factor that motivates our work.

The idea underneath our proof of concept takes inspiration from noticeably sophisticated approaches proposed for the automated extraction of activity-centric event logs like ONPROM [17]. Our vision is that the transfer of knowledge from those notable endeavors to the object-centric paradigm intertwined with our semantic encoding can unleash unprecedented results. Among those, we foresee a novel holistic mining approach: While object-centric process mining stands out as the natural extension towards information extraction from OCEDR knowledge graphs, an intriguing direction points at the use of machine learning approaches to automatically infer event-to-object and object-to-object relations based on raw event data. It is also worth mentioning that RDF datastores cater for query engines that retrieve, aggregate and manipulate semantically rich information via SPARQL Protocol and RDF Query Language (SPARQL),¹² opening up the opportunity for an advanced intermediate layer between data and process mining algorithms.

While the creation of OCEDD documents can be a way to encode the knowledge of knowledge experts, visual languages and model-driven approaches should be developed so as to allow their definition without necessarily being fluent in RDF and derivatives. Furthermore, OCEDD documents could be complemented with normative specifications dictating the structure of process data (e.g., dictating that every team member *must be* associated to exactly one support team). To this end, existing standardized languages like Shapes Constraint Language (SHACL) could come handy (a first attempt in this

¹¹<https://rml.io/specs/rml/>. Accessed: 30/09/2025

¹²<https://www.w3.org/TR/sparql11-query/>. Accessed: 30/09/2025.

direction was shown in [19] for compliance checking). Reasoning on description and prescription documents looking for redundancies and contradictions, as well as to deduce connections among concepts, is a core task for semantic web and the foundational track of process management.

5 Related Work

Our investigation relates with the semantic enrichment of process data, and to object-centric approaches to process mining. Next, we overview some notable pieces of work investigating those two areas.

The efforts to standardize an exchange format for event logs began in 2003¹³ with the XML-based format MXML (Mining eXtensible Markup Language) [20]. The well-known open-source process mining toolkit ProM [5] used MXML as the language of choice for input event logs. In their paper, the authors postulated key concepts for process mining data treatment: The instantaneous nature of an event, without duration, the typing of events to classify the activity they report on, the association of data to a process, and the association of a process instance (case) to an event. Soon after, Alves de Medeiros et al. [29] advocated the need for a semantic stratum to lay on data to allow mining techniques to consolidate, link and reason on concepts rather than strings. Hence their proposal: Semantically Annotated Mining eXtensible Markup Language (SA-MXML), annotating MXML entries with ontology references. We share their goal of semantically enriching the notions included in an event log, although we aim to operate at a deeper level by representing the whole standard as a multi-layered extensible knowledge graph. Bertoli et al. [13] use an OWL knowledge base to represent business processes with semantic annotations. They advance prior research by encoding data artifacts, integrating execution traces, and enhancing semantic modeling. Furthermore, they improve on collaborative modeling and execution analysis using semantic reasoning techniques [26]. The cross-fertilization of those disciplines has already brought about fruitful results (cf. [19,9,8]), which we seek to further spur.

Object-centric process mining has been gaining traction in recent years [6]. Several approaches to object-centric process mining have been proposed, by enhancing techniques for imperative [28,4] and declarative languages [18], or for completely new paradigms like the Object-Centric Behavioral Constraints (OCBC) [7,27]. Interestingly, the notion of knowledge graph was presented in the process mining community for a more comprehensive view on multi-dimensional analyses, covering control flow, data flow, resources, and time, with the event knowledge graphs [21]. By adding a semantic layer to the event data network introduced thereby, we approach can be integrated with that.

Before and concurrently to OCED, other object-centric event log languages have been introduced, such as eXtensible Object-Centric (XOC) logs [28], and Object-Centric Event Logs (OCEL) [24]. Especially OCEL has observed a production of several utilities to handle and process it, including a visualization tool [23], case and variant comparison [6],

¹³Quoting the words of the initiator, Wil van der Aalst, on the occasion of the 10th anniversary of XES: <https://www.tf-pm.org/newsletter/newsletter-stream-4-12-2020/10-years-of-xes>. Accessed: 28/09/2025.

filtering and sampling [12]. Our framework could amply leverage this body of knowledge to promote semantic web reasoning tasks on OCED datastores.

6 Conclusion

Recent advances in object-centric process mining underscore the need for a formal semantic definition of Object-centric Event Data (OCED) to ensure consistency, extensibility, and interoperability. By leveraging semantic web technologies, we presented a machine-readable semantic framework that enhances ontology-based relationships and entity categorization for OCED. Our approach involves a three-level knowledge representation, having OCEDO at the meta-level, OCEDD domain-specific extensions at the intensional level, and OCEDR knowledge graphs at the extensional level. We demonstrated how our framework can be employed to automatically endow raw event logs with data semantics, with the help of a proof-of-concept tool we implemented to this end. Through iterative refinements, we seek to build upon the presented results and deliver new techniques to improve OCED data reasoning, link heterogeneous data sources, and enhance knowledge inference and expressiveness.

References

1. IEEE Standard for eXtensible Event Stream (XES) for achieving interoperability in event logs and event streams (2023)
2. van der Aalst, W.M.P.: *Process Mining - Data Science in Action*, Second Edition. Springer (2016)
3. van der Aalst, W.M.P.: Object-centric process mining: Dealing with divergence and convergence in event data. In: SEFM 2019, vol. 17, pp. 3–25. Springer (2019)
4. van der Aalst, W.M.P., Berti, A.: Discovering object-centric petri nets. *Fundam. Informaticae* **175**(1-4), 1–40 (2020)
5. van der Aalst, W.M.P., van Dongen, B.F., Günther, C.W., Rozinat, A., Verbeek, E., Weijters, T.: ProM: The process mining toolkit. In: BPM Demos (2009), <http://ceur-ws.org/Vol-489/paper3.pdf>
6. Adams, J.N., Schuster, D., Schmitz, S., Schuh, G., van der Aalst, W.M.P.: Defining cases and variants for object-centric event data. In: ICPM. pp. 128–135 (2022)
7. Artale, A., Kovtunova, A., Montali, M., van der Aalst, W.M.P.: Modeling and reasoning over declarative data-aware processes with object-centric behavioral constraints. In: The 17th International Conference on Business Process Management (BPM 2019). pp. 139–156 (2019)
8. Bachhofner, S., Kiesling, E., Revoredo, K., Waibel, P., Polleres, A.: Automated process knowledge graph construction from BPMN models. In: DEXA (1). pp. 32–47 (2022)
9. Baumann, N., Hinkelmann, K., Montecchiari, D.: Supporting reuse of business process models by semantic annotation. In: CAiSE. pp. 29–35 (2023)
10. Bayomie, D., Di Ciccio, C., Mendling, J.: Event-case correlation for process mining using probabilistic optimization. *Information Systems* **114**, 102167 (Mar 2023)
11. Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web. *Scientific American* **284**(5), 34–43 (2001)
12. Berti, A.: Filtering and sampling object-centric event logs. *CoRR* **abs/2205.01428** (2022)
13. Bertoli, P., Corcoglioniti, F., Di Francescomarino, C., Dragoni, M., Ghidini, C., Pistore, M.: Semantic modeling and analysis of complex data-aware processes and their executions. *Expert Systems with Applications* **198**, 116702 (2022)

14. Calegari, D., Delgado, A.: A model-driven engineering perspective for the object-centric event data (OCED) metamodel. In: BPM Workshops. pp. 508–520 (2023)
15. Calvanese, D., De Giacomo, G., Lenzerini, M., Nardi, D., Rosati, R.: Information integration: Conceptual modeling and reasoning support. In: CoopIS. pp. 280–291 (1998)
16. Calvanese, D., Kalayci, T.E., Montali, M., Tinella, S.: Ontology-based data access for extracting event logs from legacy data: The onprom tool and methodology. In: BIS. pp. 220–236 (2017)
17. Calvanese, D., Kalayci, T.E., Montali, M., et al.: Conceptual schema transformation in ontology-based data access. In: EKAW. pp. 50–67 (2018)
18. Christfort, A.K., Rivkin, A., Fahland, D., Hildebrandt, T.T., Slaats, T.: Discovery of object-centric declarative models. In: ICPM. pp. 121–128 (2024)
19. Di Ciccio, C., Ekaputra, F.J., Cecconi, A., Ekelhart, A., Kiesling, E.: Finding non-compliances with declarative process constraints through semantic technologies. In: CAiSE Forum 2019. pp. 60–74 (2019)
20. van Dongen, B.F., van der Aalst, W.M.P.: A meta model for process mining data. In: EMOI-INTEROP (2005), <https://ceur-ws.org/Vol-160/paper11.pdf>
21. Fahland, D.: Process mining over multiple behavioral dimensions with event knowledge graphs. In: van der Aalst, W.M.P., Carmona, J. (eds.) Process Mining Handbook, pp. 274–319. Springer (2022)
22. Fahland, D., Montali, M., Leberz, J., et al.: Towards a simple and extensible standard for Object-Centric Event Data (OCED) - Core model, design space, and lessons learned. CoRR **abs/2410.14495** (2024)
23. Ghahfarokhi, A.F., van der Aalst, W.M.P.: A python tool for object-centric process mining comparison. CoRR **abs/2202.05709** (2022)
24. Ghahfarokhi, A.F., Park, G., Berti, A., van der Aalst, W.M.P.: OCEL: A standard for object-centric event logs. In: ADBIS Short Papers. pp. 169–175 (2021)
25. Grimm, S., Abecker, A., Völker, J., Studer, R.: Ontologies and the semantic web. In: Handbook of Semantic Web Technologies, pp. 507–579. Springer (2011)
26. Kampik, T., Mansour, A., Boissier, O., Kirrane, S., Padget, J., Payne, T.R., Zimmermann, A.: Governance of autonomous agents on the web: Challenges and opportunities. ACM Transactions on Internet Technology **22**(4), 1–31 (2022)
27. Li, G., de Carvalho, R.M., van der Aalst, W.M.P.: Automatic discovery of object-centric behavioral constraint models. In: BIS. pp. 43–58 (2017)
28. Li, G., de Murillas, E.G.L., de Carvalho, R.M., van der Aalst, W.M.P.: Extracting object-centric event logs to support process mining on databases. In: CAiSE Forum. vol. 30, pp. 182–199 (2018)
29. Alves de Medeiros, A.K., van der Aalst, W.M.P., Pedrinaci, C.: Semantic process mining tools: Core building blocks. In: ECIS. pp. 1953–1964 (2008)
30. Piccirilli, E., Di Ciccio, C., Montali, M., Peñaloza, R., Pontieri, L., Ricca, F.: Explainable knowledge-aware process intelligence. KI - Künstliche Intelligenz (2025)
31. Pourmirza, S., Dijkman, R.M., Grefen, P.: Correlation miner: Mining business process models and event correlations without case identifiers. Int. J. Cooperative Inf. Syst. **26**(2), 1742002:1–1742002:32 (2017)
32. Van den Spiegel, P., Dieltjens, L., Blevi, L.: Applied process mining techniques for incident and problem management. In: BPI (2013), <https://ceur-ws.org/Vol-1052/paper12.pdf>
33. Steeman, W.: BPI Challenge 2013 (2013). <https://doi.org/10.4121/uuid:a7ce5c55-03a7-4583-b855-98b86e1a2b07>
34. Wynn, M.T., Leberz, J., van der Aalst, W.M.P., Accorsi, R., Di Ciccio, C., Jayarathna, L., Verbeek, H.M.W.: Rethinking the input for process mining: Insights from the XES survey and workshop. In: ICPM Workshops. pp. 3–16 (2021)

This document is a pre-print copy of the manuscript
([Latif et al. 2026](#))
published by Springer (available at link.springer.com).

The final version of the paper is identified by DOI: [10.1007/978-3-032-17618-9_37](https://doi.org/10.1007/978-3-032-17618-9_37)

References

Latif, Saba, Fajar J. Ekaputra, Maxim Vidgof, Sabrina Kirrane, and Claudio Di Ciccio (2026). “A Semantic Encoding of Object Centric Event Data”. In: *Mining a Scientist’s Process*. Ed. by Jan Mendling, Sander Leemans, Boudewijn F. van Dongen, and Hajo Reijers. Vol. 16480. Lecture Notes in Computer Science. Springer, pp. 569–582. ISBN: 978-3-032-17618-9. DOI: [10.1007/978-3-032-17618-9_37](https://doi.org/10.1007/978-3-032-17618-9_37).

BibTeX

```
@InCollection{
  author      = {Latif, Saba and Ekaputra, Fajar J. and Vidgof, Maxim and
                Kirrane, Sabrina and Di Ciccio, Claudio},
  booktitle   = {Mining a Scientist’s Process},
  title       = {A Semantic Encoding of Object Centric Event Data},
  year        = {2026},
  pages       = {569--582},
  crossref    = {Wil60},
  doi         = {10.1007/978-3-032-17618-9_37},
  eprinttype  = {arXiv}
}
@Proceedings{
  title       = {Wil60,
                Mining a Scientist’s Process: Essays Dedicated to Wil van
                der Aalst on the Occasion of His 60th Birthday},
  year        = {2026},
  editor      = {Mendling, Jan and Leemans, Sander and van Dongen,
                Boudewijn F. and Reijers, Hajo},
  isbn        = {978-3-032-17618-9},
  publisher   = {Springer},
  series      = {Lecture Notes in Computer Science},
  volume      = {16480}
}
```