Analysis of XES and OCEL Data Schemas: Towards Multidimensional Process Mining of Intertwined Construction Processes

Araham Jesus Martinez Lagunas[1] and Mazdak Nik-Bakht[2]

[1]Dept. of Building, Civil & Environmental Engineering (BCEE), Concordia University, Montreal, QC, Canada
[2]Compleccity Lab, Dept. of BCEE, Concordia University, Montreal, QC, Canada
araham.martinez@mail.concordia.ca, mazdak.nikbakht@concordia.ca

Abstract –

Despite current digital transformation attempts, productivity in the construction industry has remained stagnant for decades. Construction organizations generate vast amounts of data from their day-to-day operations. However, these valuable digital footprints frequently remain underutilized and isolated in different IS (Information Systems). To tackle this issue, process mining, a novel and robust technology, provides organizations with the means and methods to automatically monitor and improve the efficiency of business processes by exploiting event process data. Since construction projects are process-heavy, process mining adoption becomes essential to overcome productivity stagnation. To enable process mining capabilities, process data is required to be captured in the form of event logs. In this regard, the existing body of knowledge in the construction domain yet lacks a system-agnostic framework for event log generation that considers current standard data schemas such as XES (eXtensible Event Stream) and OCEL (Object Centric Event Logs) to ensure event logs' soundness and machine readability. Therefore, to address this limitation, this work aims to facilitate the production of event logs with proper syntax and semantics by: (i) developing an ETL (Extract, Transform, Load) framework to harness process data from IS;(ii) analyzing and describing the XES and OCEL's relational data structures; (iii) providing a comparative analysis of both event log data standards. The results include a functional demonstration for constructing these data schemas applied to the Change Order Management Process (COMP) implemented in a commercial office high-rise building project. Construction change orders have the potential to either contribute to construction projects' success (if managed properly) or yield to failure otherwise.

Keywords –

Construction Management; ETL process; Event Log Generation; Data Schemas; Construction Operations; Process Mining; Process Automation

1 Introduction

The digital intensity of a country's economic sectors is an important metric in quantitatively assessing its digital economic well-being. Among the leading sectors, wholesale, finance, and manufacturing have increased their digital intensity four-fold during the last three decades by adopting cutting-edge ICT (Information and Communication Technology) developments. In this vein, there exists a strong correlation of $r \approx +0.7$ between digitization indices and productivity growth levels reported by several countries [1, 2]. In other words, the higher the digital intensity of an economic sector is, the greater its productivity and economic growth levels are. The labor productivity of digitally intensive sectors grew 22.1% in the past two decades. These leading sectors also presented a greater resilience than non-digitally intensive sectors during the COVID-19 pandemic thanks to the adoption of latest ICT developments [3-5]. Process mining, a recent technological development with exponential market growth [6], has become a crucial asset for several industries in supporting their digital transformation as its capabilities include automated process model discovery as well as monitoring, managing and improving business operations by analyzing actual operational performance behavior from event data logs[7]. To properly implement this technology, these operational event logs must be generated with formal syntax (i.e., proper notation) and clear semantics (i.e., data structures with meaningful data relationships).

In this context, the event log generation process can take up to 80% of the process mining implementation efforts [8]. Therefore, it becomes essential to understand the event log's composition including its conceptual definitions. In this regard, an event is described by three primary components: (i) a case, which is referred to as a as a process instance or execution; (ii) an event activity, often referred to as an action, transaction or unit of work happening at a particular point in time; and (iii) timestamps associated with each executed activity [7]. Event logs comprise a collection of events that can be enriched with other attributes such as objects, resources, cost, etc. These enriched event logs allow to analyze specific business operational processes from different perspectives or dimensions [7].

Unlike digitally intensive sectors, construction is part of the lagging non-digitally intensive sectors facing stagnant productivity over the past three decades [3]. Most construction companies use Information Systems (IS) and software platforms to manage their construction projects and daily business operations that vary in size and complexity. These IS store vast amounts of data related to those projects and business processes. The event logs event logs can be retrieved and reconstructed from multiple data sources including SQL relational databases [9]; Enterprise Resource Planning (ERP) systems, [10]; and Project Management Information systems (PMIS). such as Procore[11], Rhumbix [12], and Oracle NetSuite [13]. Othe data sources entail Workflow Management Systems (WfMS), Application Programming Interfaces (APIs), sensors, Building Information Modeling (BIM), 3D point clouds, etc., which are enabled by the so-called 'Internet of Events' [7, 14]. However, the stored data in those IS frequently remains underexploited, in this context, the construction domain faces three major challenges that hinders successful process mining implementations: (i) need for data integration (i.e., siloed data in sparse IS); (ii) lack of suitable/quality event logs due to poor understanding of main data schemas for event log generation according to XES and OCEL; and (iii) lack of visibility and transparency over the ripple effect of intertwined processes (i.e., multidimensional process mining).The latter is particularly important given that several construction processes are by nature interrelated with one another across construction project life cycle phases [15].

Aiming at streamlining the event log generation process necessary for process mining deployment in the construction domain, this paper has the following main objectives: (i) to develop an ETL (Extract, Transform, Load) framework for data collection; (ii) to perform a comparative analysis of XES and OCEL event log data standards while analyzing and describing their relational data structures; and (iii) to apply both data schemas into a construction Change Order Management Process as a functional demonstration. This key process, if poorly managed, is one of the main root causes of cost overruns, delays, productivity decline, and legal claims. Previous studies report expected cost growth of 10% due to change orders for most construction projects, and in some projects, the overrun far exceeds [16, 17]. Thus, it turns out essential to investigate the capabilities that more automated methods such as process mining can enable to model, monitor, audit, and manage construction processes in a more efficient and productive manner.

2 Event Log Standards - Related Works

To provide formal syntax and semantics to event logs, it is necessary to consider the most current worldwide adopted XES [18] and OCEL [19] standards for event log generation and their Event Schema Definitions (XSD).

2.1 XES Standard

The XES standard was created in 2009 by the Architecture of Information Systems (AIS) research group from the Eindhoven University of Technology (TU/E) [20] and in 2010 was adopted by the Institute of Electrical and Electronics Engineers (IEEE) [21]. Nowadays, it is a well-recognized international standard for structuring, storing, and interchange event logs in a machine-readable representation primarily based on the Extensible Markup Language (XML) suitable for process mining implementation [22]. The XES standard is founded on the concept of a single case notion, meaning that events and their related attributes belong to one process instance. Under this standard event logs are generated by selecting the perspective of interest as the event activity related to a process sequence (i.e. process control flow, organizational, cost, time, etc.). Van der Aalst (2003) applied this case-handling approach in the building industry for first time to enable automated process modeling of unstructured construction process involving a lot of uncertainty by extracting construction operational data from WfMS of semi-prefab concrete floor elements and heating, ventilation, and air-conditioning HVAC installations.

2.2 OCEL Standard

The XES standard relies on a single case notion to describe process executions, which is useful and needed to describe processes from a single perspective/angle at a time as required by the process analyst. However, for real-life processes such as those present in construction projects where the executed business operations might be composed of several cases (i.e., interrelated processes), the XES structure does not suffice to represent multiple cases within the same process model. The OCEL 1.0 standard, developed by the AIS research group in 2020, is an object-centric even log structure that empowers business analysts with the capability of structuring and storing multidimensional business processes. OCEL provides a more realistic view of business process behavior and normally stands between the source data coming from IS and the XES event log extraction. Multidimensional OCEL logs can be flattened into XES logs as needed for further analysis on specific process perspectives/views, yet one should be aware that the flatting approach can result into discovering false process behavior due to convergence (i.e., duplicated events) and divergence (i.e., considering or omitting events that are not part of the selected perspective) problems [23].

3 Methodology

Successful process mining implementations are highly dependent on well-structured event logs. Thus, a suitable method to generate machine-readable high-quality event logs is of paramount importance.

3.1 System-agnostic ETL Framework

To this end, a system-agnostic framework for event log generation is proposed and depicted in Figure 1. Regardless of the data source system, the ETL process framework comprises (i) the extraction of the event data from source IS such as construction management systems, ERPs, and SQL databases. This data can be manually extracted in the form of comma-separated values (CSV) files. However, as this data is stored in siloed systems, the manual extraction is time-consuming and resource intensive. Thus, more automated extraction methods through representational state transfer application programming interface (RESTful API) calls or Structured Query Language (SQL) as Java Script Object Notation (JSON) or CSV file formats are advised; (ii) the transformation step is two-fold, first, the data resulted from API calls or SQL queries is stored, and second, the extracted raw operational data is automatically transformed into event logs according to either XES or OCEL standard structure through SQL queries or coding pipelines and (iii) the loading step consist in storing the constructed event logs into a data storage service (i.e., cloud-based SQL database) for process mining enablement.

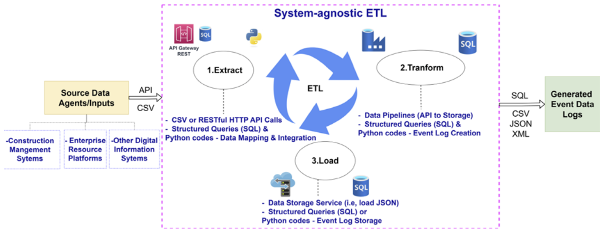

Figure 1. ETL framework for event log generation

3.2 XES Data Structure Analysis

Raw process-related data can be transformed (i.e., step 2 from ETL framework) into standard event logs that comply with XES standard. Figure 2 shows the Unified Modeling Language (UML) class diagram created based on the XES Standard [22]. The main components that should be kept in the event log schema definition include the log, extensions, global attributes, classifiers, traces, events, attributes, and data types[24].

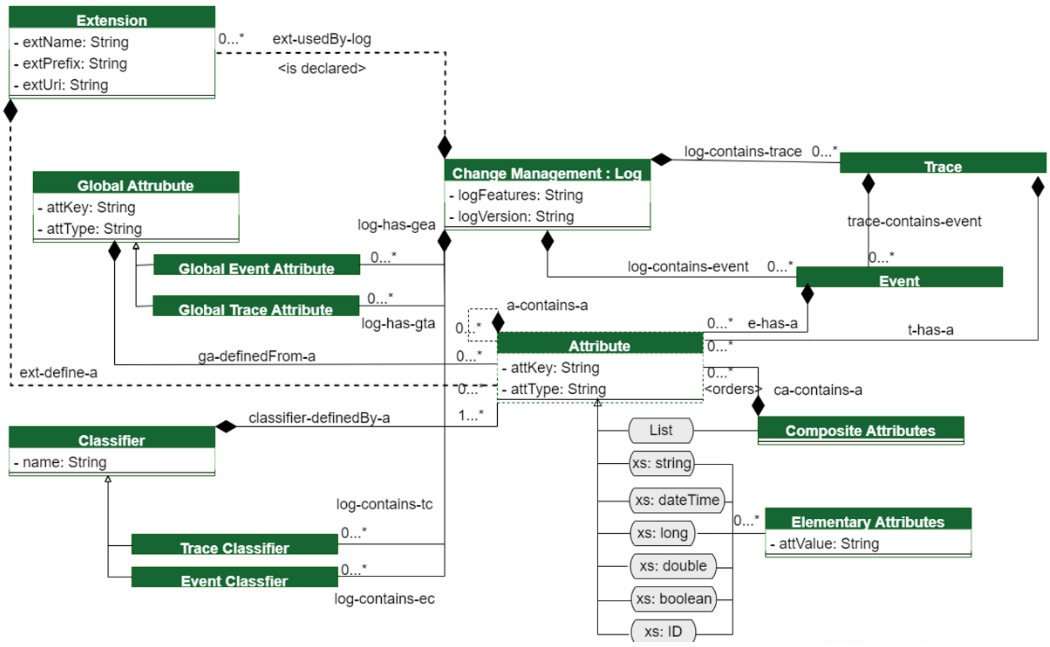

Figure 2. XES UML class diagram [22]

Event Log – An event log refers to a historical record of activities or transactions happening across processes' lifecycle. Events are labeled with execution timestamps and grouped by a unique case identifier (i.e., process instance or trace). An event log can contain one or several event traces, and it should declare/contain any required extensions to semantically describe the process.

*Extensions –*The extensions provide formal semantics and structure to the event log components by considering/assigning predefined attributes at various levels of the event log (i.e., log, trace, event). There are seven standard extensions in the 2016 version of the XES standard. New extensions can also be defined for domain-specific event log developments.

Global Attributes – This class refers to the declaration of global attributes used when certain predefined process-related information needs to be contained in the log. These global attributes are assigned to every trace and event within the log.

Classifiers – An identity can be assigned to each trace and event in the whole log using classifiers. These classifiers act as labels for traces and events that allow grouping them to compare against one another. An example of an event classifier shall contain two main attributes the event activity/instance name and the lifecycle transition of that activity (i.e., "Create Event – Complete").

Traces – They store the event activities related to a process instance or case. Each trace can contain several event objects. A trace can be seen as a list of activities during the execution of a process, the event activities are often chronologically ordered with the use of timestamps.

Extension	Attribute Key	Attribute Type	Attribute Level	Description	Source Entity	Attribute Value
Concept	name	string	log, trace, event	process name; process instance; & event activities	Change_Events Change_Orders	Change Management Process; Change Orders; Create Event
	instance	string	event	Identifier of the executed event activity instance	Change_Events	Create Event-Created
Lifecycle	model	string	log	The adopted transactional model for lifecycle transitions of the events in the log	XES Standard	Standard
	transition	string	event	Lifecycle transition for each event in the log	Applicable to each executed activity/event	Complete (i.e., from InProgress to Closed)
	state	string	event	Lifecycle state for each event in the log	Applicable to each executed activity/event	Completed
Organizational	resource	string	event	The resource name or identifier who performed the event activity	Workers - "WorkerId"	Worker Id
	role	string	event	The role of the resource	ProjectRoles - "RoleLabel"	PM Approver
	group	string	event	The organizational group where the worker belongs	Workers- "UnionCode"	WorkerUnion
Time	timestamp	datetime	event	The date and time at which the event occurred	Change_Events Change_Orders – "Date Created"	2017-01-09T09:18:00.000+01:00
Semantic	model reference	string	log, trace, event, meta	It refers to model classes of a certain ontology	No currently available	Administrative Processes
ID	id	id	log, trace, event, meta	Unique identifier for the element	Change_Events Change_Orders	ChangeOrdersId
Cost	total	double	trace, event	The total cost of a trace or an event	Change_Orders	$20,000
	currency	string	trace, event	The currency of all incurred costs	Projects	CAN
	drivers	list	trace, event	List of Cost Drivers	Change_Events Change_Orders - "Description"	ChangeEventReasons Descriptions
	amount	double	meta	The value amount for a cost driver	Change_Events Change_Orders - "Amount"	$500 for Design Change
	driver	string	meta	Identifier of cost driver	Change_Events Change_Orders - "Change Id"	ChangeEventsId
	type	string	log, trace, event	Type of cost	Projects – "BillingType"	Fixed

Figure 3. Sample Standard XES Extensions

Figure 3 provides a list of the standard extensions with a description of their attributes indicating the level at which they shall be declared. The "attribute value" column in Figure 3 also shows actual attributes' values that can be included within each of the XES extension for a COMP, while the "Source Entity" column specifies the SQL table names from which the key-value pairs for each attribute can be extracted.

3.3 OCEL Data Structure Analysis

Raw digital footprints from business process operations can also be transformed (i.e., step 2 from ETL framework) into multidimensional event logs as per OCEL standard. The initial version of the Unified Modeling Language (UML) class diagram based on OCEL 1.0 event log schema definition. This schema comprises the log, global components, objects, events, elements, and data types [23].

*OCEL Log–*Similar to XES Standard, but without the trace and extensions class definitions.

Global Components– They are different from the XES global attributes because these global attributes are not directly assigned to events, but the class element in the log. Even when in the case that there are no global attributes defined, these classes should be defined with the "value=INVALID" as it is required for the log composition. In other words, these classes act as higher-level containers/placeholders, but the real execution process information comes from Objects, Events, and Elements/Attributes.

Objects– This is the most important component of Object Centric Event Logs. Instead of traces for single case notions. OCEL is composed of a list of object types that can be seen/defined as multi-case event logs to discover a multidimensional process model. OCEL event logs can be flattened to XES logs via the objects.

Events– Similar to XES, but they are related to objects rather than to traces.

Elements/Attributes– Similar to XES, but related to objects and events (i.e., not to traces).

Aiming to improve and simplify OCEL 1.0, OCEL 2.0 standard has been recently released to facilitate the schema definition of multidimensional event logs in the form of a Common Data Model (OCDM). Figure 4 shows its UML class diagram. The main OCEL 2.0 components are the log, events, objects, event types, object types, object-to-object relationships, event-to-object relationships and their related attribute-value pairs [25].OCEL 2.0 data structure also sets aside the global classes previously considered in OCEL 1.0.

Once the event logs have been generated according to either XES or OCEL standards following the proposed system agnostic ETL framework, process mining capabilities can be leveraged and implemented for key processes in the construction domain.

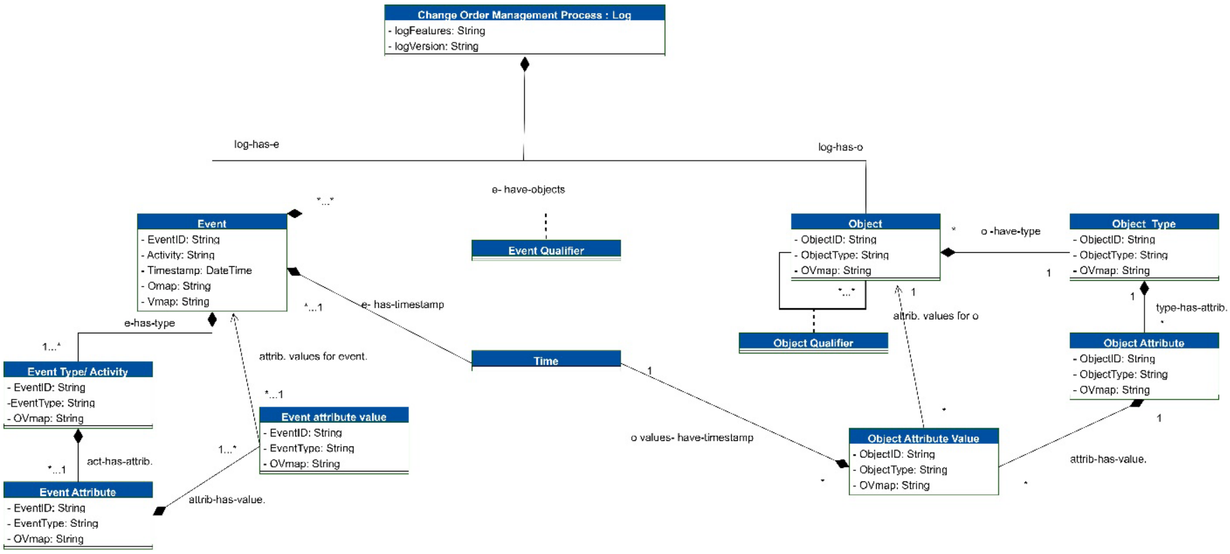

Figure 4. OCEL 2.0 UML class diagram [25]

4 Event Log Generation Process

The functional demonstration of the event log generation process mainly adopts the perspective from a General Contractor (GC) during the construction phase of a commercial office high-rise building project whose prime contracts (i.e., between owner and general contractor), budget, commitment contracts (i.e., between the general contractor and subcontractor), construction changes including Time and Materials (T&M) tracking, and their corresponding invoices are being managed on Procore construction management platform.

Data Extraction - The event log generation from digital footprints is not a trivial task, in fact it can take up to 80% of the process implementation efforts[15]. The first step consists of data extraction and collection from siloed data sources.

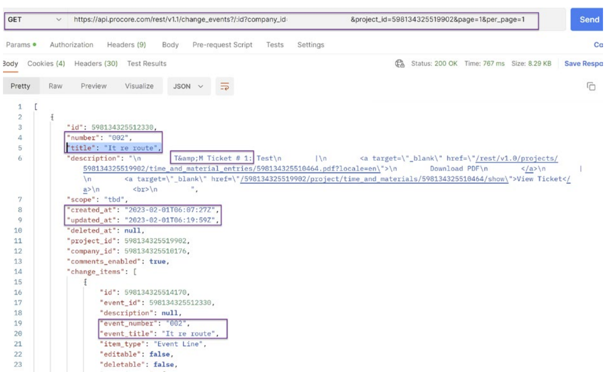

Figure 5. Data extraction Procore API - Postman [26]

In this case, the event activity timestamps, the event activity names, and their corresponding case identifiers resided on Procore, thus we extracted this information by developing a connection data application and calling the Procore API endpoints using the RESTful framework as shown in Figure 5.

4.1 XES Event Log Generation

Data Transformation to XES - The transformation step depicted in Figure 1 consisted in changing the extracted raw process event data into XES event log format through SQL queries as depicted in Figure 6.

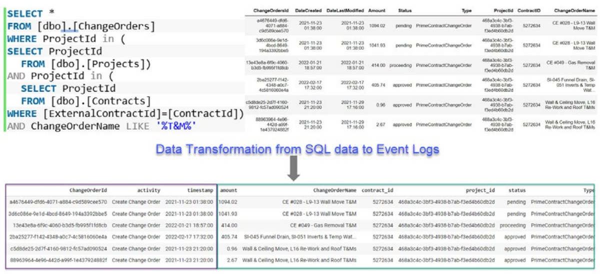

Figure 6. COMP Data Transformation to Event Log

A tabular excerpt of the generated event log for 2 executed traces/instances of the construction change order management process is shown in Figure 7.

CaseID	Activity Name	Timestamp	ActorRole
1	Create Project	2023-01-02 0:59	PM - GC
1	Create Contract	2023-01-02 1:15	PM - GC
1	Open Field Correspondence Extra Work	2023-01-02 1:05	Foreman - GC
2	Open Field Correspondence Extra Work	2023-03-20 17:34	Foreman - GC
1	Sign Correspondence Authorization to Proceed	2023-01-02 10:25	CM or Owner
1	Create Change Event from Correspondance	2023-01-02 1:07	PM - GC
1	Create T&M - In Progress	2023-01-02 1:08	Foreman - GC
1	Fill-in T&M - Ready for Review	2023-01-02 1:12	Foreman - GC
1	GC Approval/Sign T&M	2023-02-02 11:08	Foreman - GC
1	Request Customer/Client Signature T&M	2023-02-02 11:09	Foreman - GC
1	Client Approval/Sign T&M - Field Verified	2023-02-06 9:09	CM or Owner
2	Create T&M - In Progress	2023-03-21 0:53	Foreman - GC
2	Fill-in T&M - Ready for Review	2023-03-23 16:50	Foreman - GC
1	Add T&M to an existing Change Event	2023-02-07 17:12	PM - GC
1	Create Potential Change Order	2023-02-08 9:15	PM - GC
1	Create Change Order	2023-02-08 17:10	PM - GC
1	Generate Invoice	2023-02-09 10:20	Financial Manager
1	Pay Invoice	2023-02-09 9:30	Financial Manager
2	GC Approval/Sign T&M	2023-03-21 1:37	Foreman - GC
2	Client Approval/Sign T&M - Field Verified	2023-03-21 1:37	CM or Owner
2	Create Change Event from T&M	2023-03-21 1:44	PM - GC

Figure 7. COMP Tabular XES Event Log

4.2 OCEL Event Log Generation

Data Transformation to OCEL - This transformation step consisted in changing the extracted raw process event data into multidimensional OCEL event log format though API calls and SQL queries. An excerpt of the generated tabular OCEL event log is shown in Figure 8. It is worth noting that OCEL format allows for multiple case ids also known as process objects that can be interrelated with one another and correspond to specific process events.

event_id	event_activity	event_timestamp	ProjectID	ContractID	ActorID
1	Create Project	2023-01-02 0:59	['ProjectID_598134325515452']		['ActorID_PMGC9326505']
2	Create Contract	2023-01-02 1:15	['ProjectID_598134325515452']	['ContractID_598134325515452']	['ActorID_PMGC9326505']
3	Open Field Correspondence Extra W	2023-01-02 1:05	['ProjectID_598134325515452']		['ActorID_FGC9326506']
4	Open Field Correspondence Extra W	2023-03-20 17:34	['ProjectID_598134325515452']		['ActorID_FGC9326506']
5	Draft Field Correspondence	2023-03-21 0:53	['ProjectID_598134325515452']		['ActorID_CMO9326507']
6	Sign Correspondence Authorization	2023-01-02 10:25	['ProjectID_598134325515452']		['ActorID_CMO9326507']
7	Create Change Event from Correspo	2023-01-02 1:07	['ProjectID_598134325515452']		['ActorID_PMGC9326505']
8	Create T&M - In Progress	2023-01-02 1:08	['ProjectID_598134325515452']		['ActorID_FGC9326506']
9	Fill-in T&M - Ready for Review	2023-01-02 1:12	['ProjectID_598134325515452']		['ActorID_FGC9326506']
10	GC Approval/Sign T&M	2023-02-02 11:08	['ProjectID_598134325515452']		['ActorID_FGC9326506']
11	Request Customer/Client Signature	2023-02-02 11:09	['ProjectID_598134325515452']		['ActorID_FGC9326506']
12	Client Approval/Sign T&M - Field Ve	2023-02-06 9:09	['ProjectID_598134325515452']		['ActorID_CMO9326507']
13	Create T&M - In Progress	2023-03-21 0:53	['ProjectID_598134325515452']		['ActorID_FGC9326506']
14	Fill-in T&M - Ready for Review	2023-03-23 16:50	['ProjectID_598134325515452']		['ActorID_FGC9326506']
15	Add T&M to an existing Change Eve	2023-02-07 17:12	['ProjectID_598134325515452']		['ActorID_PMGC9326505']
16	Create Potential Change Order	2023-02-08 9:15	['ProjectID_598134325515452']	['ContractID_598134325515452']	['ActorID_PMGC9326505']
17	Create Change Order	2023-02-08 17:10	['ProjectID_598134325515452']	['ContractID_598134325515452']	['ActorID_PMGC9326505']
18	Generate Invoice	2023-02-09 10:20		['ContractID_598134325515452']	['ActorID_FM9326508']
19	Pay Invoice	2023-02-09 9:30		['ContractID_598134325515452']	['ActorID_FM9326508']
20	GC Approval/Sign T&M	2023-03-21 1:37	['ProjectID_598134325515452']		['ActorID_FGC9326506']
21	Client Approval/Sign T&M - Field Ve	2023-03-21 1:37	['ProjectID_598134325515452']		['ActorID_CMO9326507']
22	Create Change Event from T&M	2023-03-21 1:44	['ProjectID_598134325515452']		['ActorID_PMGC9326505']
23	Create Potential Change Order	2023-03-22 2:22	['ProjectID_598134325515452']	['ContractID_598134325524448']	['ActorID_PMGC9326505']
24	Create Change Order	2023-03-22 15:20	['ProjectID_598134325515452']	['ContractID_598134325524448']	['ActorID_PMGC9326505']
25	Generate Invoice	2023-03-27 9:15		['ContractID_598134325524448']	['ActorID_FM9326508']
26	Pay Invoice	2023-03-27 9:20		['ContractID_598134325524448']	['ActorID_FM9326508']

Figure 8. COMP Tabular OCEL Event Log

4.3 XES and OCEL Event Log Schemas

The excerpt in Figure 9a presents an example of the generated event log in XES file format showing the main schema components described in Section 3.1. CSV tabular event logs can be transformed into XES using ProM [27], XESame[28],OpenXES[29] and PM4Py [30]. While Figure 9b presents an example of the generated event log in OCEL file format showing the main schema components described in Section 3.2. OCEL files can be converted from CSV into JSON OCEL using the Object Centric Process Analysis (OCPA) python library[31].

5 Discussion

Most medium-sized and large construction companies often have their IT architecture up and running to store and manage their construction project data for instance in SQL servers. However, their current data-oriented IT architectures lack process awareness. To address this limitation, the proposed process-oriented framework allows for automated data extraction from construction business operations being managed in siloed IS. The extracted raw data can be stored in SQL relational databases, yet it need to be transformed in the form of event logs to enable process mining capabilities.

The results from analyzing the event log transformation process for XES and OCEL help to identify some of the main differences and similarities between both standards as well as their main benefits and limitations. For instance, even though currently the XES standard is globally predominant, a trend exists towards adopting object centric event logs for multidimensional process analysis by generating a common data model OCDM of intertwined construction processes not only to analyze the actual process performance from single process perspective but also to be able to detect how the performance of a process of interest (i.e., process efficiency of COMP) behavior is impacted by interrelated process instances/objects (i.e., T&M, Change Events, Contracts, Invoices, etc.). As per their similarities include the fact that both shall include case (s), events and timestamps. Figure 10 provides a comparison table that results from the analyses described in Section 3.


```xml
<?xml version="1.0" encoding="UTF-8" ?>
<log xes.version="1.0" xes.features="nested-attributes" openxes.version="1.0RC7">
    <extension name="Lifecycle" prefix="lifecycle" uri="http://www.xes-standard.org/lifecycle.xesext"/>
    <extension name="Time" prefix="time" uri="http://www.xes-standard.org/time.xesext"/>
    <extension name="Concept" prefix="concept" uri="http://www.xes-standard.org/concept.xesext"/>
    <classifier name="Event Name" keys="concept:name"/>
    <classifier name="(Event Name AND Lifecycle transition)" keys="concept:name lifecycle:transition"/>
    <string key="concept:name" value="identify_ce_s.csv"/>
    <trace>
        <string key="concept:name" value="1"/>
        <event>
            <string key="amount" value="4392.990005"/>
            <string key="Resource_Role" value="ConstructionManager"/>
            <string key="lifecycle:transition" value="complete"/>
            <date key="time:timestamp" value="2022-05-18T00:00:00.000-04:00"/>
            <string key="project_id" value="9b246618-d64a-4654-af39-311ae0259cb5"/>
            <string key="cost_code" value="1-1-1-1-0000"/>
            <string key="change_description" value="(CEN-948) ASI-581, MJI-148, Bldg G, Irrigation L5 Terrace &amp; Supplemental,
            <int key="contract_id" value="1501578"/>
            <date key="timestamp_s" value="2022-05-18T00:00:00.000-04:00"/>
            <string key="concept:name" value="Identify Change Event"/>
        </event>
        <event>
            <string key="amount" value="4392.990005"/>
            <string key="Resource_Role" value="Project Manager_GC"/>
            <string key="lifecycle:transition" value="complete"/>
            <date key="time:timestamp" value="2022-05-19T00:00:00.000-04:00"/>
            <string key="project_id" value="9b246618-d64a-4654-af39-311ae0259cb5"/>
            <string key="cost_code" value="1-1-1-1-0000"/>
            <string key="change_description" value="(CEN-948) ASI-581, MJI-148, Bldg G, Irrigation L5 Terrace &amp; Supplemental,
            <int key="contract_id" value="1501578"/>
            <date key="timestamp_s" value="2022-05-19T00:00:00.000-04:00"/>
            <string key="concept:name" value="Create Change Event"/>
        </event>
```


Figure 9a. XES Event Log Schema


```json
{
    "ocel:global-log": {
        "ocel:attribute-names": [
            "event_Payment",
            "event_CaseID",
            "event_ContractID",
            "event_Location",
            "event_id",
            "event_Invoice",
            "event_index",
            "event_ProjectID",
            "event_CompanyID"
        ],
        "ocel:object-types": [
            "PCO",
            "FCorrespondence_ID",
            "ChangeOrder",
            "ChangeEvent",
            "TMTicket",
            "ActorID"
        ],
        "ocel:version": "1.0",
        "ocel:ordering": "timestamp"
    }
    "ocel:events": {
        "0": {
            "ocel:activity": "Create Project",
            "ocel:timestamp": "2023-01-02T00:59:40",
            "ocel:omap": [
                "ActorID_PMGC9326505"
            ],
            "ocel:vmap": {
                "event_Payment": NaN,
                "event_CaseID": 1,
                "event_ContractID": NaN,
                "event_Location": NaN,
                "event_id": 1,
                "event_Invoice": NaN,
                "event_index": 1,
                "event_ProjectID": "['ProjectID_598134325515452']",
                "event_CompanyID": "['CompanyGC_598134325510176']",
                "start_timestamp": "2023-01-02 00:59:40"
            }
        },
        "1": {
            "ocel:activity": "Create Contract",
            "ocel:timestamp": "2023-01-02T01:15:37",
            "ocel:omap": [
                "ActorID_PMGC9326505"
            ],
            "ocel:vmap": {
                "event_Payment": NaN,
                "event_CaseID": 1,
                "event_ContractID": "['ContractID_598134325515452']",
                "event_Location": NaN,
                "event_id": 2,
                "event_Invoice": NaN,
                "event_index": 2,
                "event_ProjectID": "['ProjectID_598134325515452']",
                "event_CompanyID": "['CompanyGC_598134325510176']",
                "start_timestamp": "2023-01-02 01:15:37"
```


Figure 9b. OCEL 1.0 Event Log Schema

Figure 9. XES vs. OCEL Event Log Schema Definitions

Criteria	XES	OCEL 1.0
Perspective	To analyze a single perspective at a time (i.e., control flow)	Multi-perspective (i.e., event control flow & actors)
Process Mining Approach	Traditional Process Mining (Current state-of-the-art)	Multidimensional Process Mining (Industries shifting towards Multi-level Process Mining)
Quality Aspects	Less transparency over events and precedence relationships	Preserves Entities/Objects, Events, and relations
Suitable for	Workflow executions	E2E Process Executions
Schema Basis	Process Instance-based	Entity/object-centric based
Main Schema Components	Log, Traces, Events	Logs, Objects/Entities, Events
Case Notion	Single Case Notion (i.e., Change Events)	Multiple Case Notion (i.e., Change Events & Orders)
Conversion	XES to OCEL (Requires domain knowledge of the E2E Process)	OCEL to XES (One should deal with convergence and divergence problems)
Conformance	Enables Conformance Checking	Enables Conformance Checking
Transparency	Promotes Transparency	Enables Higher E2E Transparency
Knowledge Management	Knowledge of process instance (i.e., less complex networks)	Higher E2E Knowledge Capture (i.e., real complex networks)
Interoperability	Less focus on interoperability	Higher focus on Interoperability

Figure 10. Comparison XES vs OCEL standards.

6 Conclusions and Future Work

Current IT data architectures of several construction companies lack process awareness in the sense that they have poor visibility and transparency over the performance behavior of actual process executions. To address this limitation, this study proposed an ETL process-oriented framework for event log generation which extracts raw construction data footprints. from IS and transform it into process event logs. The reconstructed event logs enable process mining capabilities to model, monitor and manage the actual performance of their construction business operations through automated means and methods. Besides, this study contributed to the body of knowledge by shedding light on the XES and OCEL data standard schemas understanding and their application specific for the construction domain.

This research facilitates the event log generation process with proper syntax and semantics necessary for a successful process mining implementation that can support construction stakeholders during process audits and to streamline business operations that result in significant productivity gains. Despite the stated benefits, this study has the following limitations: (i) although the proposed framework aims to be system-agnostic, this study is mainly focused on Procore as the data source for event log generation, future studies should further investigate the framework's validity when considering other data sources/IS; (ii) the scope of this study is limited to the event log generation phase, thus future work should elaborate on the process analytics phase through process mining algorithms; and (iii) the functional application is limited to a single project, so future analysis should include a benchmarking comparison of process performance behavior between two or more similar projects in different location and managed by the same or different companies.

In this context, parallel studies are underway to demonstrate process mining automating and improving capabilities in construction [32].It is also important to note that it is necessary to use formal process modeling notations such as Business Process Modeling and Notation (BPMN) to represent the automatically discovered actual process executions, which can also enrich companies' Standard Operating Procedures (SOPs) .Last but not least, future work in the construction domain should be focused on extending current descriptive and diagnostic process mining capabilities into predictive and prescriptive analytics by leveraging novel simulation and AI technologies[15]

References

[1] J. Mollins and T. Taskin, "Digitalization: Productivity," Bank of Canada, 1914-0568, December 2023 2023. [Online]. Available: https://doi.org/10.34989/sdp-2023-17

[2] A. Charles Atkins *et al.*, "Rekindling US productivity for a new era," McKinsey Global Institute, December 2023 2023. [Online]. Available:https://www.mckinsey.com/mgi/our-research/rekindling-us-productivity-for-a-new-era

[3] McKinsey Global Institute, "DIGITAL AMERICA: A TALE OF THE HAVES AND HAVE-MORES," 2015. [Online]. Available: www.mckinsey.com/mgi.

[4] H. Liu, "Economic performance associated with digitalization in Canada over the past two decades Economic and Social Reports," Statistics Canada, December 2023 2021. [Online]. Available: https://doi.org/10.25318/36280001202100200001-eng

[5] F. Calvino, C. Criscuolo, L. Marcolin, and M. Squicciarini, "A TAXONOMY OF DIGITAL INTENSIVE SECTORS," OECD, December 2023 2017. [Online]. Available: https://one.oecd.org/document/DSTI/CIIE/WPIA(2017)2/en/pdf

[6] Grand View Research. "Process Mining Software Market Size Report, 2021-2028." https://www.grandviewresearch.com/industry-analysis/process-mining-software-market-report (accessed December, 2023).

[7] W. M. P. Van der Aalst, *Process Mining Data Science in Action*. Berlin, Heidelberg: Springer (in en), 2016.

[8] W. M. P. van der Aalst, "Process Mining: A 360 Degree Overview," in *Lecture Notes in Business Information Processing*, vol. 448: Springer

Science and Business Media Deutschland GmbH, 2022, pp. 3-34.
[9] "Cloud Computing Services | Microsoft Azure." https://azure.microsoft.com/en-us (accessed February20, 2023).
[10] SAP. "What is SAP? | Definition and Meaning." @SAP.https://www.sap.com/about/company/what-is-sap.html (accessed September 26, 2022).
[11] Procore Technologies. "What is Procore? | Construction Management Software." https://www.procore.com/en-ca/what-is-procore (accessed September 26, 2022).
[12] Rhumbix. "RMBX." https://www.rhumbix.com/ (accessed September 26, 2022).
[13] ORACLE. "ORACLE NETSUITE " Oracle. https://www.netsuite.com/portal/products/erp.shtml (accessed September 26, 2022).
[14] S. Kouhestani and M. Nik-Bakht, "IFC-based process mining for design authoring," (in en), *Automation in Construction,* vol. 112, p. 103069, 2020/04//2020, [Online]. Available: https://doi.org/10.1016/j.autcon.2019.103069
[15] W. M. P. Van der Aalst, J. Carmona, J. Mylopoulos, S. Ram, M. Rosemann, and C. Szyperski, *Process Mining Handbook*. Springer, 2022.
[16] A. M. Eldeep, M. A. M. Farag, and L. M. Abd El-hafez, "Using BIM as a lean management tool in construction processes – A case study: Using BIM as a lean management tool," *Ain Shams Engineering Journal,* vol. 13, no. 2, 2022/3// 2022, doi: 10.1016/j.asej.2021.07.009.
[17] W. Ibbs, "Update on Quantitative Analysis of Change and Loss of Productivity," *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction,* vol. 13, no. 1, 2021/2// 2021, doi: 10.1061/(asce)la.1943-4170.0000447.
[18] *Standard for eXtensible Event Stream (XES) for Achieving Interoperability in Event Logs and Event Streams*, 9781504424219, C. I. S. IEEE, 2016. [Online]. Available: https://ieeexplore.ieee.org/servlet/opac?punumber=7740856
[19] *OCEL: A Standard for Object-Centric Event Logs*, 9783030850814, A. F. Ghahfarokhi, G. Park, A. Berti, and W. M. P. van der Aalst, 2021. [Online]. Available: http://dx.doi.org/10.1007/978-3-030-85082-1_16
[20] Eindhoven University of Technology. "Analytics for Information Systems - AIS." https://www.win.tue.nl/ais/doku.php (accessed September 29, 2022).
[21] IEEE. "IEEE - The world's largest technical professional organization dedicated to advancing technology for the benefit of humanity." https://www.ieee.org/ (accessed September 28, 2022).
[22] IEEE, *IEEE Standard for eXtensible Event Stream for Achieving Interoperability in Event Logs and Event Streams*. 2023.
[23] A. F. Ghahfarokhi, G. Park, A. Berti, and W. M. P. van der Aalst, *OCEL: A Standard for Object-Centric Event Logs* (Communications in Computer and Information Science). Springer International Publishing, 2021, pp. 169-175.
[24] J. C. A. M. Buijs, "Mapping Data Sources to XES in a Generic Way," Master, Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, 2010. [Online]. Available:http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Mapping+Data+Sources+to+XES+in+a+Generic+Way#0
[25] A. Berti *et al.*, "OCEL (Object-Centric Event Log) 2.0 Specification," 2023. [Online]. Available:https://www.ocel-standard.org/2.0/ocel20
[26] Postman. https://www.postman.com/. (accessed March 14, 2024)
[27] ProM. "ProM." http://www.promtools.org/doku.php?id=prom69 (accessed September 21, 2023).
[28] J. C. A. M. Buijs, "Mapping Data Sources to XES in a Generic Way," in *Chelsea*, ed, 2010, pp. 123-123.
[29] C. Günther and E. Verbeek, "OpenXES Developer Guide," Eindhoven University of Technology, The Netherlands, 2012. [Online]. Available:https://www.xes-standard.org/_media/openxes/openxesdeveloperguide-1.8.pdf
[30] FIT. "PM4py - Process Mining for Python." https://pm4py.fit.fraunhofer.de/ (accessed September 14, 2023.
[31] A. Berti, M. Montali, and W. M. P. van der Aalst, "Advancements and Challenges in Object-Centric Process Mining: A Systematic Literature Review," 2023/11// 2023. [Online]. Available: http://arxiv.org/abs/2311.08795.
[32] A. J. Matinez Lagunas and M. Nik-Bakht, "Enabling Process Mining in the Construction Industry: An Event Log Schema for Change Management Process," in *CSCE/CRC International Construction Specialty Conference. (Manuscript accepted for publication)*, 2023, pp. 1-13.