This work was submitted to:

Chair of Process and Data Science (PADS - Informatik 9), RWTH Aachen University

Integrating Sustainability Data into Event Logs for Object-Centric Process Mining

Bachelor's Thesis

Author: **Marco Heinisch**

Student ID: **422307**

Supervisor: Nina Graves M.Sc.

Examiners: Prof. Dr. Wil van der Aalst
Prof. Dr. Ulrik Schroeder

Registration Date: 2023-11-23

Submission Date: 2024-03-25

Abstract

In response to environmental regulations and societal expectations, companies are adopting sustainable practices through innovative strategies, tools, and technologies. In the context of manufacturing, Process Mining (PM) is recognized as a technique to support and automate sustainability assessments. However, the existing literature either discusses the use of PM for sustainability assessments without providing concrete guidance or is limited to very case-specific applications. To address this gap, we present a novel approach by structuring sustainability data for integration into Object-Centric Event Logs (OCEL). We propose a standardized method for integrating sustainability data as attributes of objects and events in OCELs, and organizing them into classes to incorporate domain knowledge required for assessment. To encourage further research in process mining for sustainability, an artificial example event log is generated by simulating a manufacturing process using a colored Petri net model. We enriched the log with sustainability data received from Climatiq API. Based on the provided artificial example log, we evaluated how sustainability information is included and how it provides the required information for assessment. Our research demonstrates the potential of utilizing OCELs for integrating object-centric process mining and sustainability assessment.

Contents

Abstract **ii**

1 Introduction **1**

1.1 Motivation 2

1.2 Problem Statement 3

1.3 Research Questions 3

1.4 Research Goals 3

1.5 Contributions 4

1.6 Thesis Structure 4

2 Preliminaries **5**

2.1 Sustainability Assessment 5

2.1.1 Life Cycle Assessment 5

2.1.2 Carbon Accounting 6

2.2 System Boundary 7

2.2.1 Material and Energy Flows 7

2.3 Object-Centric Event Logs 8

3 Related Work **10**

4 Integrating Sustainability Data **12**

4.1 Identifying Relevant SD in Manufacturing 12

4.1.1 Search Method 12

4.1.2 Data Extraction 13

4.2 Classes of Sustainability Data 16

4.2.1 Impact Indicators 16

4.2.2 Impact Parameters 17
4.2.3 Impact Scores 17
4.2.4 Data Relation 18
4.3 Including Sustainability Data into OCEL: sOCEL 19
4.3.1 Object Classes in sOCEL 19
4.3.2 Sustainability Attributes in sOCEL 20
4.3.3 Formal Definition 22
4.3.4 Example Application 23
4.3.5 Example sOCEL 24

5 Implementation **27**
5.1 Designing a Manufacturing Process Model 27
5.1.1 A Simplified Hinge Production 28
5.1.2 Modeling and Generation using CPN-Tools 30
5.2 Tool Implementation 32
5.2.1 Usage 33
5.2.2 Architecture 35

6 Evaluation of sOCEL **37**
6.1 Generating sOCEL 37
6.2 Providing Insights 37
6.2.1 Material Input and Output 38
6.2.2 Workpiece Flow 39
6.2.3 Energy 41
6.2.4 Not Allocatable or Missing Impacts 41
6.3 How sOCEL Supports Carbon Accounting 41

7 Discussion **44**

8 Conclusion **46**
8.1 Future Work 46

Bibliography **52**

Acknowledgements **53**

Chapter 1

Introduction

Sustainable development is commonly defined as "development that meets the needs of the present generation without compromising the ability of future generations to meet their own needs" [1]. With the adoption of the United Nations 2030 Agenda, the global community has committed itself to striving for sustainable development [2]. Consequently, environmental legislation and societal expectations require businesses to reduce their carbon emissions [3]. To achieve this, they must adopt effective strategies, tools, and technologies [3] and as a result, there's a growing need to develop approaches that provide process-level insights into the environmental impact of business activities [4]. Such transparency makes it possible to identify reduction opportunities, assess environmental impacts and support the process of reducing a company's carbon footprint [4]. As 23% of the global CO_2 emissions are allocatable to traded goods and manufacturing causes significant environmental impacts, it is recognized that sustainability of the manufacturing industry is important in order to achieve sustainable development [5, 6].

Process Mining (PM) is recognized as a valuable tool for analyzing and improving complex processes based on event logs [7]. Recent literature has recognized PM's potential for supporting Life-Cycle Assessment (LCA), which is a standardized method for evaluating the environmental impacts of a product throughout its life cycle [8]. A product life cycle includes resource extraction, production or manufacturing, distribution, usage, and end-of-life treatment. Conducting an LCA allows assessment and identification of reduction potentials of environmental impacts caused by the system under study. As comprehensive assessments, like LCAs on manufacturing processes, represent a high manual effort, the potential of automatization by facilitating PM for Sustainability Assessment (SA) has been identified [9, 10].

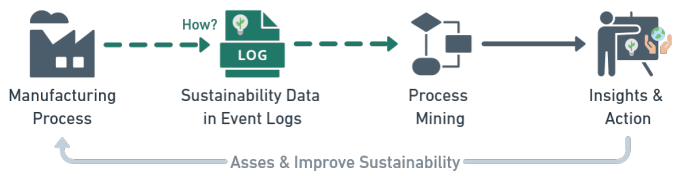

Figure 1.1: Asses and Improve Sustainability of Manufacturing Processes using PM

However, current literature shows a significant gap in terms of specific, low-level guidelines for incorporating sustainability data (SD) into event logs to support assessment like LCA [8, 10]. Moreover, existing studies have generally not considered state-of-the-art process mining techniques, such as object-centric event logs (OCEL) [8]. The potential of Object-Centric Process Mining (OCPM) to improve sustainability is recognized [11]. Therefore, it is essential to investigate how sustainability data can be effectively integrated into these log formats and to provide artificial examples to enable further research on this topic integrating sustainability data into event logs enables the possibility to apply adjusted algorithms to support continuous life-cycle assessment [12]. This could in turn enable reduction of environmental impacts of manufacturing processes.

In this work, we demonstrate how state-of-the art OCELs can be facilitated for sustainability assessment, as visualized in Figure 1.1 focusing on manufacturing domain. We provide a specification on how SD is integrated into event logs, named sOCEL, and generate and provide an example event log according to this specification. To achieve this, CPN-Tool is used to simulate a manufacturing process using a colored Petri net model. We enrich our CPN model with sustainability data provided by Climatiq API, using a for this purpose implemented Python tool.

1.1 Motivation

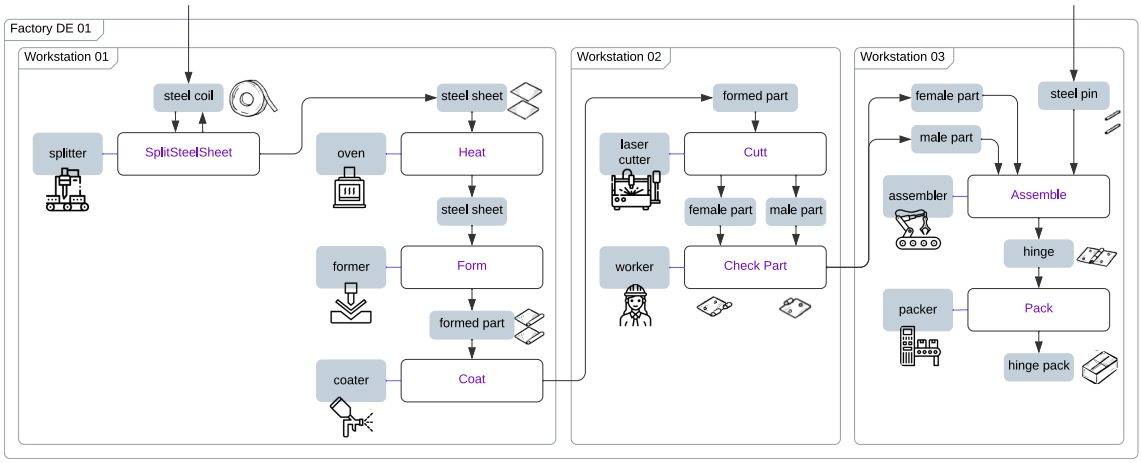

Figure 1.2: Example Manufacturing Process: A Hinge Production Line

Exploring the use of object-centric process mining for sustainability assessment in manufacturing, we first consider a running example we will refer to in subsequent chapters. The example is a simple production line for manufacturing hinges. This process is located in a facility in Germany using various machines grouped in three workstations as shown in Figure 1.2. The process uses steel pins and steel coils as input. The steel coil is split into steel sheets at the first workstation, which are then formed into male and female parts at the second workstation. At the third workstation, the hinge leaves are assembled with a steel pin to create the hinges. The final step is packaging the hinges into a hinge pack.

As the objective is to apply process mining for supporting sustainability assessment of this process, the first step is to determine which data from this production line is extractable

as well as which data is needed for a comprehensive assessment. In this example, we focus on energy consumption of machines and material consumptions, as hypothetically identified by reviewing assessment standards. To create an OCEL in the next step, we decide to represent workpieces, discrete materials, and workers as objects and process operations as events. Continuing with data extraction, questions arise such as whether to model machines as attributes or as objects with associated consumption data [8]. It is also unknown how to account for continuous resource consumption that is not event related, such as idle power consumption.

By reviewing additional literature or consulting experts, we might solve these and additional questions for this example process. However, the resulting event log and the included Sustainability Data (SD) is not standardized and therefore case or organization specific. To enable interoperability between specialized assessment algorithms and event logs of different origin, a standardized approach for including SD in OCEL is needed [12, 13]. Thus, for supporting sustainability assessment using object-centric process mining, it is essential to define which data should be considered and in detail how assessment related information is stored in OCELs.

1.2 Problem Statement

Ideally, a standard exists and is being used to incorporate SD into state-of-the-art event logs for PM. Currently, however, there is no specific low-level guideline specifically for object-centric process mining [10]. As a result, there are no specialized applications and algorithms that work with SD in a standardized way. The full potential of the multi-perspective view that object-centric process mining could provide has not been explored. In this work, we address this research gap and propose a specified way to integrate sustainability data used in assessments into OCELs.

1.3 Research Questions

This work focuses on the integration of sustainability data into event logs, and therefore aims to answer the following research questions:

1. How can sustainability data be included in object-centric event logs to support sustainability assessment methods like life cycle assessment?
   (a) Which information, classes of data are relevant for (automated) sustainability assessment?
   (b) How can as relevant identified information be integrated into OCEL2.0 to support sustainability assessment?

1.4 Research Goals

Aligned with our research questions, it is the objective of this thesis is to contribute to the interdisciplinary field of OCPM and sustainability assessment, with a focus on manufacturing processes. The specific research goals are:

1. Identify data relevant to sustainability assessments in manufacturing and that is applicable in process mining (RQ 1a).
2. Develop a specific guideline for integrating sustainability data into OCEL2.0 (RQ 1.b).
3. Additionally, generate an artificial OCEL2.0 dataset containing sustainability data.

1.5 Contributions

In this work, we provide an overview and structure of sustainability data that can be facilitated for SA in OCPM in the context of manufacturing. Based on these findings, we specify a framework for including SD into OCEL2.0 called sOCEL. We generate and provide an example event log according to this specification. In this process, we implemented a python tool for inclusion of sustainability data from APIs such as Climatiq.

1.6 Thesis Structure

The remainder of this thesis is structured as follows. In chapter 2, we present relevant concepts and terms of both fields, PM and sustainability assessment. In chapter 3, we discuss related work. In chapter 4, we review literature on assessments and structure for assessment required information and propose an approach of integrating SD into OCEL. In chapter 5 we demonstrate the applicability of this approach in a case study and generate an event log enriched with sustainability data, using an implemented tool for integrating SD from external sources. Further, we use the generated example event log to evaluate our approach in chapter 6. Last, we discuss the results in chapter 7 and give a conclusion and outlook in chapter 8.

Chapter 2

Preliminaries

In this chapter, we introduce terms and concepts from both domains, object-centric process mining and sustainability assessment, which we refer to in the remaining work.

2.1 Sustainability Assessment

Sustainable development is understood to be based on three main pillars: economic prosperity, social equity and environmental quality [1]. In the context of sustainability assessment, an impact therefore describes a positive or negative consequences of a process in the context of economic environmental or social dimension [14]. Further, in this work, individual causes of impacts, such as "steel waste" and "methane emissions", are referred to as *impact aspects*. These impact aspects are further categorized into *impact elements*, like "material output".

The field of sustainability assessment is broad, encompassing various methodologies, relying on different approaches and standards for measuring and assessing sustainability or impacts [15]. This work focuses on two prevalent approaches: Life Cycle Assessment (LCA), widely adopted in the manufacturing domain [16], and carbon accounting.

2.1.1 Life Cycle Assessment

Life Cycle Assessment (LCA) is a standardized method for evaluating the environmental impact of a product system or service throughout its entire life cycle, as specified in ISO

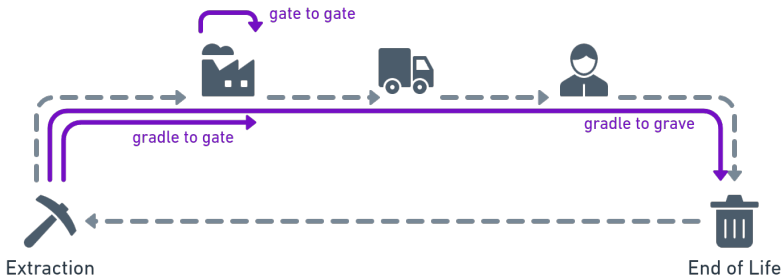

Figure 2.1: A Selection of Life Cycle Stages of a Product

14044:2006 [17]. The life cycle of a product in the context of LCA is presented in Figure 2.1. The LCA process involves several iterative steps conducted for assessment [17]:

1. In the **Goal and Scope Definition** phase, the system under study and its boundaries are defined. Typically, scopes are defined as gradle-to-gate, gate-to-gate and similar life cycle phases. Furthermore, a reference flow is defined, describing what an assessment result relates to, i.e., a single hinge that is produced.
2. In the **Life-Cycle-Inventory (LCI) Analysis** phase, data on inputs and outputs of a system under study is collected and structured into a dataset, called LCI.
3. In the **Life-Cycle-Impact-Assessment (LCIA)** phase, the collected LCI data is translated into a set of sores as defined by an applied LCIA standard such as ReCiPe16 [18] or ILCD [19]. Resulting scores are, for example, "climate change" and "land transformation".
4. In the **Interpretation** phase, findings from the life cycle inventory (LCI) and life cycle impact assessment (LCIA) are used to draw conclusions, make recommendations and decisions according to the original goals and scope.

During the LCI creation phase, not all information on inputs and output is available. For example, for a gradle-to-gate LCA (see Figure 2.1), it may be unknown to a manufacturing company what impacts the production of input goods or extraction of impact resources already caused. Therefore, common practice involves referring to previous LCA results on such goods as reference values. Collections of LCAs on different products, goods, and processes in different scopes are gathered in LCI Databases (LCI-DB) [19].

In the context of LCA, a unit process is defined as the smallest component of a manufacturing process that is taken into account in the life cycle inventory analysis with quantifiable input and output data [20].

2.1.2 Carbon Accounting

Carbon Accounting describes the effort for systematically quantifying and analyzing the overall Greenhouse Gas (GHG) emissions of a system or organization. Different publication standardize this method, i.e. ISO 14064 or the GHG Protocol which serves as a practical guide. Services such as Climatiq-API[1] provide reference GHG emission values for emission factors, like consumed resources or produced wastes. We use the service to demonstrate automatic integration of sustainability data into an artificial event log for PM.

2.1.2.1 Process Mining for Carbon Accounting

A qualitative study by Brehm et al. identified the following three main requirements and two potentials for the application of PM for carbon accounting. The main challenge and requirement is *Retrieving and Integrating Carbon-Related Data*. For analysis, event logs must contain information about carbon emissions, i.e., an event causes. This information is typically not available. However, APIs are identified as useful tools to receive carbon data related to recorded events. The necessary information required for the use of APIs, like mass and geometrics of a product, is difficult to obtain. Additionally, *Technical and Process Know-How* are required. Technical expertise is necessary to integrate relevant

[1] www.climatiq.io

data into event logs, while process know-how is needed to draw conclusions from insights that PM could provide and take action to reduce emissions. The authors suggest that PM has the potential to *Create Transparency About As-Is*. The application of PM allows for the analysis of carbon emissions in a given process and additionally provides context and relationship of emissions to the process. This analysis has the potential to *Identify Levers to Reduce CO2 on an Operational Process Level*. Unlike a top-down approach where a company aims to reduce its total energy consumption in general, PM relies on bottom-up data and provides operational process level insights that enable the identification of reduction potentials. Finally, the potential of PM for CA is identified as *Linking Operational Efficiency and Sustainability (Multidimensional Analysis)*. Process mining, can provide insights into both carbon emissions, economic KPIs or other dimensions, providing a multi-perspective view on processes [4].

2.2 System Boundary

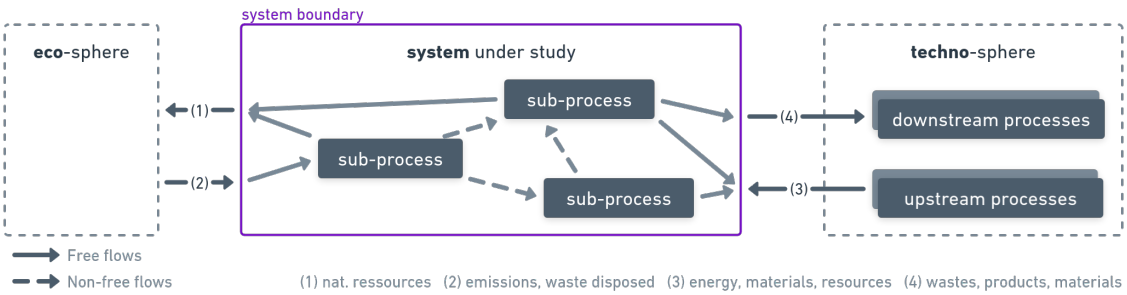

Figure 2.2: In- and Outflows between a System, Ecosphere and Technosphere [21]

For assessments like LCA, a clear scope and system definition is required [17]. A system, for example a production process under study, is set into context to the technosphere and ecosphere as visualized in Figure 2.2. The ecosphere describes the natural environment, like the atmosphere, where GHG are emitted to. A system, however, commonly interacts with other technical systems, i.e., electricity production if it requires electricity as input. The entirety of the systems with which a system under study can interact is referred to as the technosphere [19]. This includes the system under study itself. An environmental impact is caused through interaction between technosphere and ecosphere [19].

2.2.1 Material and Energy Flows

Standards defining LCI formats used for LCIA like ILCD distinguish multiple input or output flow classes, such as waste, energy, or products. In this work, we refer to a general concept of flows in assessment context, as proposed by Carlson et al. and visualized in Figure 2.2. The authors only differ between *inflow* and *outflow*. Further, flows that are not connected to other flows in a system are referred to as *free flows*. These flows occur between nature and the technosphere, such as natural resources, emissions, and products. Flows that occur within a system and between sub-processes, operations, or events are referred to as *non-free flows* and can be interpreted as a network between sub-processes.

In the context of sustainability assessment, a typical flow considered is material flow, which includes material input and output of a system studied. The therm refers to both

substances and goods, including (natural) resources, (raw) materials, (co-, by-, in-between) products and similar [22].

2.3 Object-Centric Event Logs

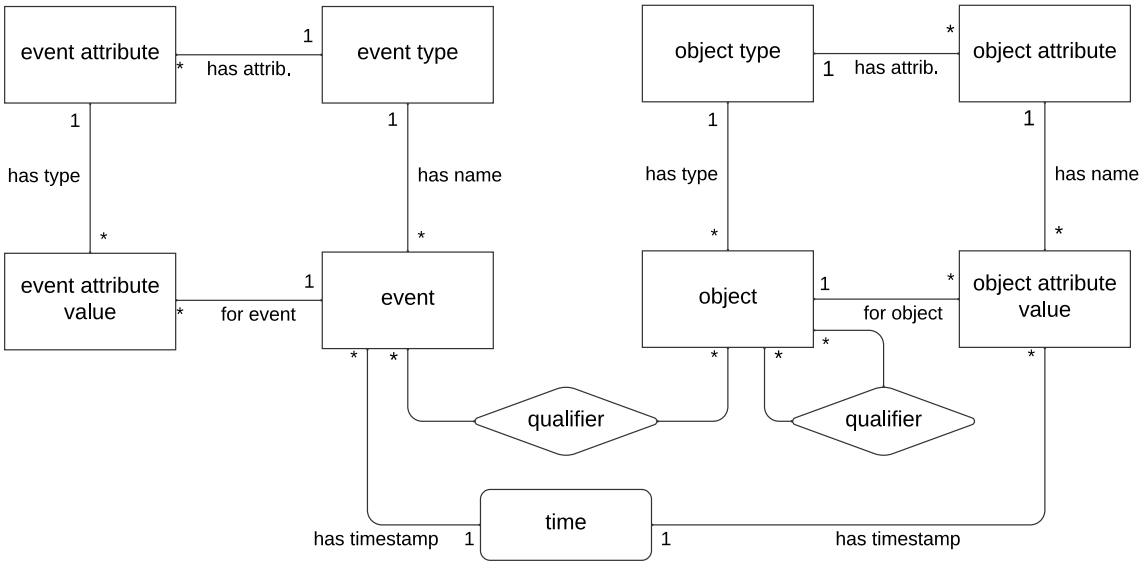

Figure 2.3: OCEL2.0 Metamodel [23]

Process mining [24] involves analyzing event data (logs) from real-world processes to gain data-based insights into processes. Event logs are traditionally based on a case notation. Each log entry consists of a timestamp, an activity (event type) and a case ID. However, this approach does not fully capture the multidimensional nature of processes, where multiple cases (objects) can be linked to a single event. In the context of manufacturing, for example, machines, workpieces, workers, and other elements interact within a single production operation.

To overcome these limitations, Object-Centric Event Logs (OCEL[2]) were introduced and further developed into OCEL version 2.0 [23]. OCEL2.0 allows events to relate to multiple cases, called objects. Objects have optionally qualified relationships with each other, the event to object relationship is qualified, and object attributes can be updated and changed. The relationships are clearly defined, and a metamodel is provided in Figure 2.3, illustrated in a simplified UML-like class notation. In the following, we detail the formalization of this approach.

To define object-centric event logs, we first need to give some preliminary definitions:

Definition 2.1 (Universes)**.** Let $\mathbb{U}_\Sigma$ be the universe of strings. We define the following pairwise disjoint universes:

- $\mathbb{U}_{ev} \subseteq \mathbb{U}_\Sigma$ *is the universe of events.*
- $\mathbb{U}_{etype} \subseteq \mathbb{U}_\Sigma$ *is the universe of event types (i.e., activities).*
- $\mathbb{U}_{obj} \subseteq \mathbb{U}_\Sigma$ *is the universe of objects.*

[2] www.ocel-standard.org

- $\mathbb{U}_{otype} \subseteq \mathbb{U}_{\Sigma}$ *is the universe of object types.*
- $\mathbb{U}_{attr} \subseteq \mathbb{U}_{\Sigma}$ *is the universe of attribute names.*
- $\mathbb{U}_{val}$ *is the universe of attribute values.*
- $\mathbb{U}_{time}$ *is the universe of timestamps, with* $0 \in \mathbb{U}_{time}$ *as the smallest element and* $\infty \in \mathbb{U}_{time}$ *as the largest element.*
- $\mathbb{U}_{qual} \subseteq \mathbb{U}_{\Sigma}$ *is the universe of qualifiers.*

Based on these preliminary definitions, we can now define an OCEL2.0:

Definition 2.2 (OCEL2.0)**.** An Object-Centric Event Log (OCEL2.0) [23] is a tuple $L = (E, O, EA, OA, evtype, time, objtype, eatype, oatype, eaval, oaval, E2O, O2O)$ with

- $E \subseteq \mathbb{U}_{ev}$ *is the set of events.*
- $O \subseteq \mathbb{U}_{obj}$ *is the set of objects.*
- $evtype : E \to \mathbb{U}_{etype}$ *assigns types to events.*
- $time : E \to \mathbb{U}_{time}$ *assigns timestamps to events.*
- $EA \subseteq \mathbb{U}_{attr}$ *is the set of event attributes.*
- $eatype : EA \to \mathbb{U}_{etype}$ *assigns event types to event attributes.*
- $eaval : (E \times EA) \nrightarrow \mathbb{U}_{val}$ *assigns values to event attributes (not all the attributes are mapped for each event).*
- $objtype : O \to \mathbb{U}_{otype}$ *assigns types to objects.*
- $OA \subseteq \mathbb{U}_{attr}$ *is the set of object attributes.*
- $oatype : OA \to \mathbb{U}_{otype}$ *assigns object types to object attributes.*
- $oaval : (O \times OA \times \mathbb{U}_{\text{time}}) \nrightarrow \mathbb{U}_{val}$ *assigns values to object attributes.*
- $E2O \subseteq E \times \mathbb{U}_{qual} \times O$ *are the qualified event-to-object relations.*
- $O2O \subseteq O \times \mathbb{U}_{qual} \times O$ *are the qualified object-to-object relations.*

such that

- $\text{dom}(eaval) \subseteq \{(e, ea) \in E \times EA \mid evtype(e) = eatype(ea)\}$ *to ensure that only existing event attributes can have values.*
- $\text{dom}(oaval) \subseteq \{(o, oa, t) \in O \times OA \times \mathbb{U}_{time} \mid objtype(o) = oatype(oa)\}$ *to ensure that only existing object attributes can have values.*

We use this definition in subsection 4.3.3 to define Sustainability enriched Object-Centric Event Logs (sOCEL).

Chapter 3

Related Work

This chapter discusses related work in the field that combines process mining with sustainable development. There is a lack of academic research on this topic, as noted by several authors [4, 8]. However, it is becoming increasingly important [8].

Graves et al. recently performed a comprehensive literature review on the topic of process mining and sustainability, including Life-Cycle Assessment and other assessment methods. The review summarizes existing literature and proposes the PM4S framework. This framework outlines usage of elements of object-centric process mining, such as objects and event log entries. In this framework, objects represent products, raw materials, waste, and resources. They include as attribute, information on what they are, where they are located, and waste-resource indicators. Events represent sourcing, production, logistics, and maintenance operations. Additionally, they include information on emissions and energy as event attribute. PM4S outlines how OCPM can be used for sustainability assessment. They state, that challenges identified in literature regarding the use of PM for SA can be addressed by using OCPM. This object-centric approach overcomes the limitations of traditional process mining, which relies on one-dimensional case notation in event logs [8]. State of the art OCPM can provide multiple perspectives, e.g., material flow and control flow. Further, Graves et al. suggest integrating sustainability related indicators into objects and events using APIs in an additional preprocessing step. They also mention additional attributes for objects entering (and leaving the system) and suggest the consideration of inter-attribute dependencies. The authors emphasize the necessity for more detailed research on the practical implementation of these concepts.

Brehm et al. analyze the application of process mining for carbon accounting. They conducted a qualitative study, interviewing experts to identify the potential and requirements of process mining for carbon accounting. They highlighted a gap in existing studies on the potential of process mining for sustainable development and carbon reduction. The authors identified the need for more automated data import and efficient mapping of energy-related data to process models.

The authors Kaniappan Chinnathai and Alkan propose a framework for digital lifecycle management and sustainable smart manufacturing. The framework utilizes process discovery to gain insights into as-is processes. It combines process mining, process analysis, and energy consumption analysis using a simulation model. However, the integration of sus-

tainability data into event logs and PM is not explicitly mentioned, and traditional event logs are utilized. The article highlights the potential of process mining and simulation models in creating digital representations of manufacturing processes.

Hagedorn et al. have developed an approach for process mining-based value stream analysis. They also suggest the use of process mining for life cycle assessment in production processes. Although they have developed a concept for combining process mining with value stream analysis to evaluate environmental indicators, they conclude that further refinement is necessary for practical implementation. The authors highlight the challenge of data availability and extraction.

Ortmeier et al. present a framework for integrating process mining into life cycle assessment and identified potential application of PM in the LCI phase where PM is used for data extraction and mining. In their case study, they included production data, such as electricity, as event-level metadata in a conventional event log. They demonstrated that discovered process models can be used to visualize impacts, such as energy consumption, in the identified process model. The authors noted that direct input of energy and machine consumption is required. However, the proposed framework is a high-level methodological approach and does not facilitate state-of-the-art PM.

Similarly, Constantin Houy et al. discusses visualizing the impact on process models and annotating activities with resource. In their approach, the authors annotate activities in process models with resource consumption or greenhouse gas emissions and suggest a visualization-based analysis. However, the article does not address the challenges related to system boundaries.

In a study on Green Business Process Management (BPM), Hoesch-Klohe and Ghose enrich process models with sustainability information and discuss the aggregation and allocation of sustainability data using carbon dioxide as a primary example indicator. The authors propose guidelines for SD-based process improvement and redesign. However, it presupposes an as-is process model, which potentially is discovered using process mining.

Research on process mining and sustainability assessment sees PM's potential in sustainability assessment [4, 8, 12]. Yet, studies offer either broad frameworks or specific case insights without a general but also specific method to integrate SA into PM for comprehensive assessments, revealing a research gap to be addressed.

Chapter 4

Integrating Sustainability Data

As shown, there is no existing mapping of SD of manufacturing processes to OCEL. In this chapter, we first identify and structure data which is relevant for sustainability assessment. Second, we propose a mapping of such data to OCEL and provide guidance on how to include it.

4.1 Identifying Relevant SD in Manufacturing

To identify data relevant for assessment, we conduct an exploratory literature review. We gather an overview of the data and information considered in assessments, based on case studies of manufacturing operations, as well as previous literature reviews published and extract information related to manufacturing (unit) processes.

4.1.1 Search Method

The main objective is to gather data for assessment purposes. We do not intend to present a complete overview of all assessments found in literature, as there are numerous published for manufacturing, as there are numerous publications on the subject. Our methodology aims to identify and collect the most relevant and exemplary data necessary for sustainability assessment, as well as to provide an overview of how data is accessed and collected in studies. We therefore use two search queries:

In a first search, we are interested in data used for assessments of manufacturing operations. Therefore, we considered a list of typical operations and specified the corresponding keywords for the search query. To ensure comprehensibility, we used DIN 8580's categorization of operations. For each of the six main operation categories, we selected a keyword for the search. For instance, "casting" represents *primary shaping*, similarly, "CNC"and "cutting" are used for *dividing*, and "heating" is employed for *modifying material properties*. The categories of *forming*, *joining*, and *coating*, as defined by DIN 8580, are used as keywords without modification. All keywords resulting from this process are listed in the figure Figure 4.1. For each keyword, a search was conducted using the first operation-specific query, with the keyword inserted in the placeholder for ⟨Operation⟩. Additionally, a second query is used to identify collections of sustainability data relevant for assessments in the manufacturing industry.

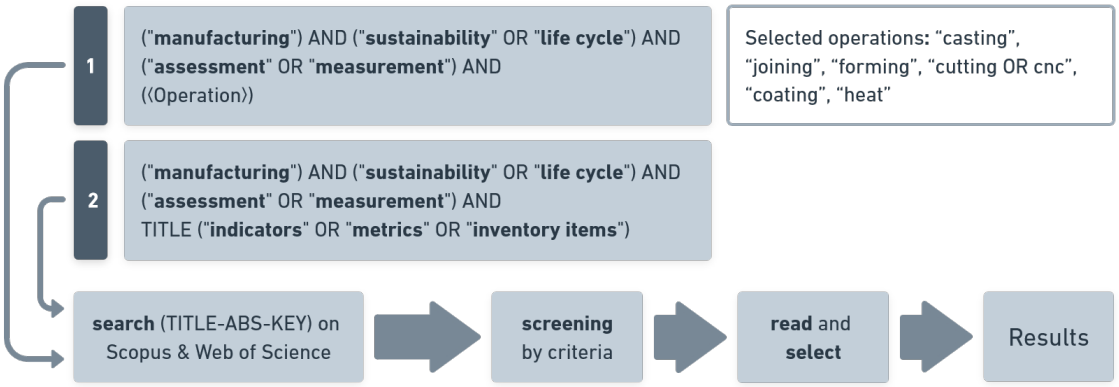

Figure 4.1: Overview of the Literature Screening and Selecting Process

The search strings stated in Figure 4.1 are applied to Web of Science and Scopus, and selection criteria for the results are applied in the following order: first, the title and abstract are considered. If a decision cannot be made, the full text of the papers is reviewed. Only papers with a title and abstract that indicate a focus on the type of manufacturing operation being searched for, as well as a sustainability rating, are selected. We exclude papers that are not written in English or German and those that are not available in full text. Only papers that clearly state the data, measures, or lists of indicators used in the assessment and discuss how these items were determined or measured are included.

4.1.2 Data Extraction

A review of the selected case studies, presented in Table 4.1 reveals a pattern in the data used in the assessments, as visualized in Figure 4.2. These assessments aim to quantify a relative value, such as Global Warming Potential (GWP), to assess the overall impact of an operation, for example, to compare the impact of two different welding techniques. This is done by identifying and quantifying impact aspects (also called impact indicators or life cycle inventory items) that represent a sustainability aspect that causes an impact, such as electrical energy consumed or greenhouse gases emitted. As some of these dimensions are difficult to measure, calculation parameters or reference values from the literature are used. For example, intermediate parameters such as welding length or processing time are used to determine or estimate indicators such as material consumption. Example parameters and relative value, referred to as impact scores, are listed in Table 4.2.

The literature describes various methods for collecting relevant data. One common approach involves using metering devices, such as sensors installed on machinery, to monitor specific metrics, such as a device's electricity consumption [31]. One method for obtaining data involves extracting it from Enterprise Resource Planning (ERP) systems, which may include information on both procured and consumed materials or workpieces [28]. Other data is obtained through indirect means, such as estimates or simulations [35] [37]. For example, the quantity of welding material required can be calculated based on the length of the weld [28]. Additionally, some studies use data from previous literature and LCI-databases [38]. LCI databases provide reference values for impact indicators and measures that are averaged for a specific context, such as a particular product, country, region, and year.

	Product	Pilar	S.Element	Data Category	Data Relation	Data Source
[28]	Metal welded workpiece	Env	Eergy Input, Material Input, Material Output (Emissions, Waste)	Indicators, Parameters	Output Workpiece batch	Mathematical model, (LCI) databases, WPS, CAD, Welding map, Material certificates
[29]	Die casted metal part	Env	Energy Input, Material Input, Material Output	Scores, Indicators	Output Workpiece Unit	Literature, Computational model
[30]	Laser cutted workpiece	Env, Eco	Eergy Input, Material Input, Material Output, Costs of Materials, Costs Waste, Investment Costs	Scores, Indicators, Parameters	Output Workpiece batch	Calculation, unstated
[31]	Incremental formed workpiece	Env, Eco	Eergy Input, Costs, Material Output	Scores, Indicators, Parameters	Output Workpiece batch	Calculation, Simulation, Sensor, Literature
[32]	Formed workpiece	Env	Energy Input, Material Input	Scores, Indicators, Parameters	Output Workpiece batch	Sensor, time study
[33]	CNC cutted workpiece	Env	Energy Input, Material Input, Material Output, Noise	Scores, Indicators, Parameters	Output Workpiece Unit	Bill of material, Mathematical model, Production data, Literature review
[34]	Plasma coated workpiece	Env	Energy Input, Material Input, Material Output	Parameters, Indicators	Relative to Coated Area	Experimental data, Mathematical model
[35]	Heat treated workpiece (induction)	Env, Eco, Soc	Energy Input, Costs, Material Input, Material Output, Worker	Scores, Indicators, Parameters	Output Workpiece Unit	Mathematical model, (LCI) databases
[36]	Heat treated workpiece (gas)	Env	Energy Input, Material Input	Scores, Indicators, Parameters	Output Workpiece Unit	Experimental data

Table 4.1: Selected Case-Studies

P/S	Parameter/Score	Data Relation	Data Source	Study
P	Workpiece material specification	Workpiece	Data sheet, Measured	[32, 35, 36]
	Machine mode or configuration	Operation	-	[31, 32, 34]
	Processing time	Operation	-	[31, 32, 34]
	Process location	Operation	-	[30, 35]
	Welding length	Workpiece, Operation, Machine	-	[28]
S	Material Footprint	Whole Process and Output Unit	Calculated	[36]
	Carbon Footprint			[32, 36]
	Cumulative Energy Demand			[33, 36]
	Land use			[30]
	Toxicity to humans			[30]

Table 4.2: Examples of Impact Scores and Parameters

Pillar	S.Element	Sustainability Aspect	Data Relation	Data Source	Query 1	Query 2
Economic	Costs	Disposal costs	(Material Output)	*	[30]	
		Energy costs	(Energy Input)	Literature	[30]	[39, 40]
		Investment costs	Machine, Facility, ...	*	[30]	[40, 41]
		Material costs	(Material Input)	*	[30]	[39–42]
		Labor costs	Process Time and Employee	Estimation	[30]	[39, 40, 42]
		Transportation cost, warehouse cost	Operation, Facility	**		[39, 40]
		Defective/returned products loss, warranty cost	Product	**		[39, 40]
	Benefits	Sales price, profit	Output unit, Organisation	**		[39–41]
Social	Employees	Injury rate and others i.e. absenteeism	Employee	Organisation data, (LCI)-Database	[35]	[42]
		Lost work days	Employee	Organisation data, (LCI)-Database	[35]	
		Exposures (noise, mist/dust, chemical concentration, temperature, hazardous, threats)	Facility and Employee	*	[33]	[39, 42]
		Health and safety impacts (Chronic illness rate)	Employee	Organisation data, (LCI)-Database	[35]	[39, 41]
Environmental	Other	Emitted heat, light (outside), noise (outside) and radioactive emissions	Facility	**		[40]
	Material Output	Cutting fluid loss	Machine	Production Data, Calculated	[33]	
		Emissions to air (GHG/Pollutant)	Operation	(LCI)-Database	[29, 33, 35]	[39–42]
		Liquid waste (hydraulic fluid, wastewater, lubricant waste)	Operation, Machine	*	[29, 30]	[39, 40, 42]
		Solid waste (steel, disposal/recycle)	Operation, Output unit	Production Data, Calculated	[29, 30, 33, 34]	[39, 40, 42]
	Material Input	Material used in product (molten metal)	Input Unit	*	[29]	[39, 40]
		Material, resources (water, die lubricant, hydraulic fluid, plunger lubricant, compressed air), processing gas (nitrogen), alcohol, powder coat, oil	Operation (and Machine)	*	[29, 30, 34–36]	[39–42]
	Energy Input	Burned Natural Gas	Machine	*	[29, 32, 36]	[39, 40]
		Energy consumption (by source, from renewable and nonrenewable sources)	Operation	Time Series, Bill of materials, Measured, Calculated	[29–31, 31–34, 36]	[39–42]
		Electric energy idle/standby/maintain consumption	Machine	Measured, Bill of materials, Calculated	[31, 31–33]	[39–42]

*not clearly stated **only obtained from second query

Table 4.3: Impact Indicators Listed or Used in Assessments and Relating to Manufacturing Processes

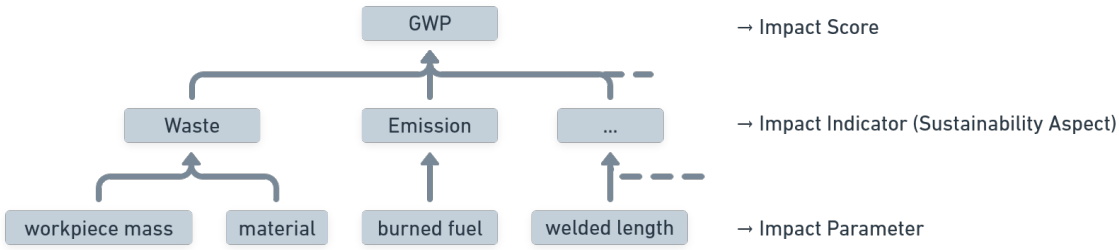

Figure 4.2: Identified Classes (or Hierarchies) of SD in Case Studies

In a next step, we focus on impact aspects. In addition to the selected case studies, we also consider four studies resulting from the second search query. These studies present lists of indicators themselves, without focusing on specific operations or assessment procedures. They do, however, provide a broad overview of relevant sustainability aspects. We collect and group all identified indicators and filter out those that do not relate to manufacturing (unit) processes. Table 4.3 provides a list of the remaining indicators. Indicators presented in this list can be cross dimensional. For example, measures such as "noise" can be considered as impact on a surrounding ecosystem or attributed as impact on workers at the same time [40].

It is noteworthy that there are various indicators, the most important of which are material input, output, and energy consumption as identified in Table 4.3. Therefore, we will focus on these aspects of sustainability in the remainder of this work.

4.2 Classes of Sustainability Data

In reviewed literature we identified three classes of sustainability data, as already shown in Figure 4.2, potentially includable as attributes. In the following sections, we motivate *impact indicators*, *impact parameters*, and *impact scores*. These classes of data contain all sustainability-related information needed for assessment.

4.2.1 Impact Indicators

The reviewed literature assesses the overall impact by collecting and aggregating different flows that indirectly cause impacts in upstream processes or on the spot, such as electric energy, steel waste, or emissions through gas combustion (see collection in Table 4.3). Therefore, we introduce the class of *impact indicators*.

Impact indicators, as presented in Figure 4.3, contain information on all different impacts caused by an event, object, or process. For example, if a LCA is conducted, the set of indicators represents the set of LCI-Items. To give a specific example in the context of LCA, we present the definition of a life cycle inventory item following ReCiPe2016 [18], one of many LCIA methods. The ReCiPe2016 standard requires, for example, two aspects to identify an emission inventory item for LCIA: The identification of the substance is done using a chemical identification string (i.e. "P" for Phosphorus) followed by the identification of the emission compartment (i.e. fresh water, agricultural soil). Based on these aspects, an identification string can be used as an impact indicator in OCEL (i.e. `emission_p_freshwater`). All flows at the process level must be translated to this item

standard, for example, by using reference values obtained from LCI-DBs.

To provide another example, considering carbon accounting, an alternative automated approach involves utilizing APIs such as Climatiq. This platform provides an API Specification that outlines unique IDs for various impact aspects across different scopes, i.e. `id_electricity-supply_grid-source_nuclear`. An Entry specified by ID has specific metadata describing the assessed context, such as location, year of assessment or covered scope (i.e. gradle-to-gate), that must be considered. If an assessment is based on this, impact indicators relate bijective to available API entries.

Figure 4.3: Example Indicators of a Process

4.2.2 Impact Parameters

Revisiting the example impact indicator of created waste, as shown in Figure 4.3. For event-based assessment, we measure the amount of waste per event to quantify the indicator attribute during log creation. However, this is often not feasible due to current data availability presented in the reviewed studies. Additionally, some indicators, such as on-the-spot emitted CO_2 from gas combustion, are challenging to measure, and other impacts, such as idle-time energy consumption, are not always attributable to single events [4]. To address data availability and allocation issues and further enable the inclusion of estimations and models, we introduce the class of *impact parameter*.

Impact parameters provide data to estimate impact indicators indirectly based on domain and process knowledge. These process-specific parameters include, i.e., material, mass, volume, geometric properties, and operation duration. In the example provided in Figure 4.4, a process parameter is the weight of the workpiece measured before and after the waste-creating operation (i.e., during a quality check event). Alternatively, waste could be calculated based on the geometrical specifications of input and output workpieces. Additionally, by considering varying or non-allocatable impacts, parameters can reference external time series data of machine consumptions. In contrast to discrete attribute values, time series data can cover continuous measures and varying consumption, such as electricity consumption, allowing for its consideration in process analysis and assessment.

Figure 4.4: Example Parameter of a Process

4.2.3 Impact Scores

While impact indicators are used to represent specific impact factors, understanding the relative magnitude of these impacts requires a higher level of abstraction. For example, using only impact indicators, it is unclear whether a process that consumes 1 kg of steel

and 1 kWh of electricity has a greater impact on global warming than a process that is otherwise the same except that it consumes 0.5 kg of steel and 5 kWh of electricity. Thus, the class of *impact scores* is introduced to describe the overall effect of a system or event. Multiple impact scores for different impact categories, such as climate change or toxicity, can be calculated and represented as impact scores. This process involves aggregating and weighting individual impact indicators according to a standardized methodology, such as ReCiPe2016. The scores can represent various levels of abstraction, such as midpoint impacts (e.g., climate change) or endpoint impacts (e.g., damage to ecosystems) [43], or both.

For example, Climatiq calculates one single metric, GWP, measured in CO_2 equivalent (CO_2e). The API defines an ID for impact indicators such as electrical energy consumption. This allows the calculation of CO_2e from a given amount of electrical energy consumed through requests that meet the requirements for Climatiq API calls. The resulting GWP CO_2e value represents an impact score, as shown in Figure 4.5.

Figure 4.5: Example Score of a Process

4.2.4 Data Relation

Considering the example of implicit waste presented in Figure 4.5, the storage and locations for parameters and indicators must be specified in OCEL to distinguish for assessment relevant from irrelevant information, to potentially identify sustainability data gaps and to enable the development and application of assessment tools or algorithms. In the reviewed literature, indicators, scores, and parameters relate to entities, such as a product unit, an operational event, or a machine. The related entities and the identified impact elements are mapped to OCEL in Table 4.4. We include impact elements solely as attributes of the objects or events to which they relate.

Event	Event. Attr.	Object	Object. Attr.
Operation	Time	In-/Output Unit	Material In-/Output
	Material In-/Output	Employee	Energy Input
	Energy Input	Machine	
		Facility, Location	

Table 4.4: Initial Event and Object Attributes Mapping

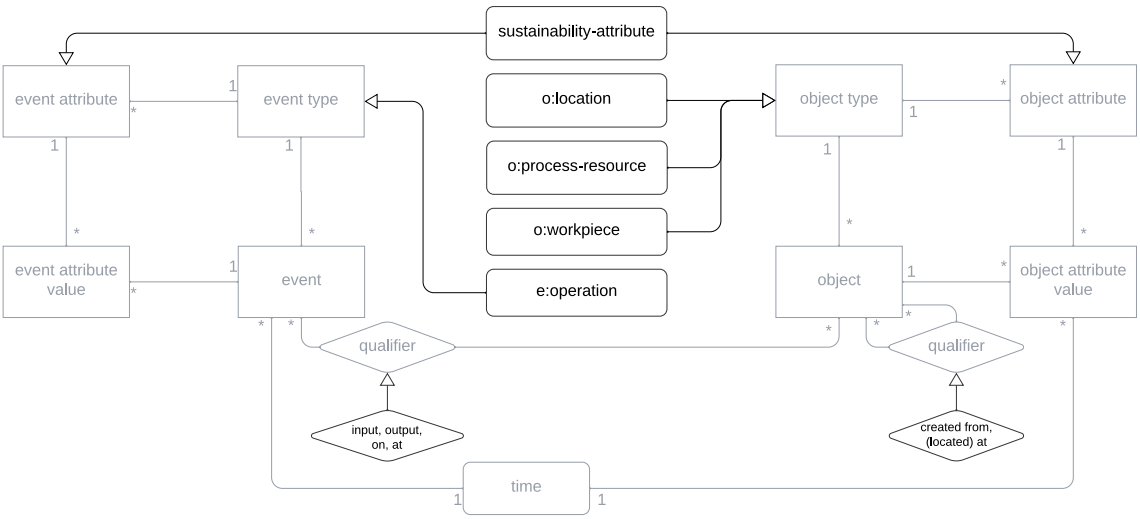

Figure 4.6: sOCEL Classes Integrated into the OCEL2.0 Standard

4.3 Including Sustainability Data into OCEL: sOCEL

Generalizing the initial mapping, extend OCEL2.0 to Sustainability enriched Object-Centric Event Log (sOCEL) by introducing new classes for object types, event types, and attributes as presented in Figure 4.6. We also specify qualifiers for object-to-object and event-to-object relationships, outlined in Figure 4.7 and propose storage locations for essential sustainability attributes and outline their processing workflow, as shown in Figure 4.8. This inclusion reflects domain knowledge, relevant in assessments of manufacturing processes.

Within sOCEL, we define an event-type class called *operation*, which represents manufacturing operations. All events, that represent an impact causing event, are assigned this class. This enables the separation of events that are relevant for sustainability assessments, for tracing of material flow and potentially linking sustainability data.

4.3.1 Object Classes in sOCEL

In sOCEL three classes of objects are defined: *location*, *process resources*, and *workpiece*. Each operation event can be associated with a location object (Where is the operation performed?). An operation event can be associated with a process resource object, such as a machine (What is an operation performed with?). Finally, operation events are optionally linked to workpieces, specifying the target of the operation (What was the operation performed on?).

Process Resource (PR): This class includes entities involved in the operation without being part of the primary material flow. Examples include workers, machines, and tools. Since consumption on unit process level is primarily associated with machines and other process resources and less with workpieces, impact data is primarily recorded related to objects of the process resource class. Linking operation events to process resource objects allows attribution of machine-caused impacts to events. sOCEL uses the “with” qualifier to explicitly define the relationship between an operation event and a process resource.

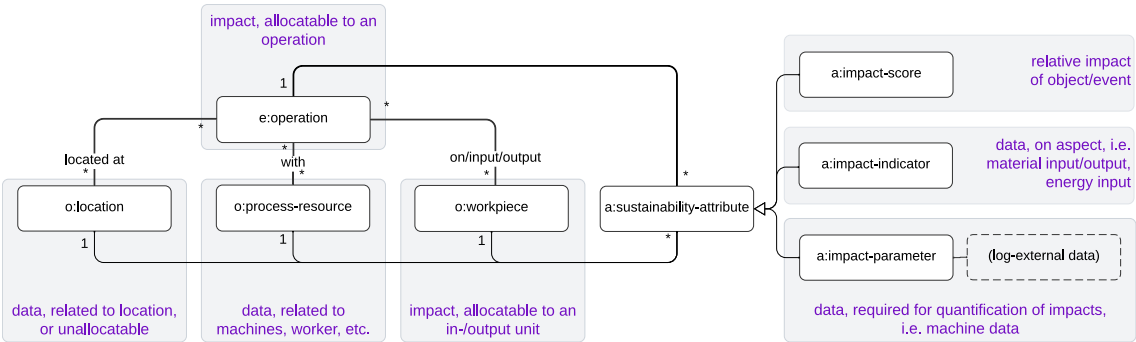

Figure 4.7: sOCEL Classes and Relationships

Location: Impacts that are not directly related to the operation or its materials, such as those from infrastructure (e.g., facilities), are to be included as an attribute of objects of the location class. While location information is considerable as an attribute of other events and objects, this approach enables to store location related sustainable impacts, that are not allocatable elsewhere. These impacts, often related to utilities such as lighting and heating, support various events and are difficult to attribute directly to specific operations. The location class supports the usage of sOCEL as a single source of truth. In sOCEL, the located "at" qualifier is used for operation-to-location relationships, with the potential to extend these qualifiers to include logistical events (i.e. "from", "to").

Workpiece: Objects of this class represent discrete physical entities that are the subject of operations that either modify, consume, or create them. Objects of this class can explicitly describe raw materials, discrete resources, as well as intermediate, co-products, by-products, and similar items. Workpiece-objects either enter a system, are newly created, consumed or leave in the context of a manufacturing process. Tracking workpiece objects is crucial for assessing material inflows and outflows to and from a system, and for distinguishing between the various states of materials. For example, it must be clear whether an object involved in an operation is only an input (i.e., it existed before and then was consumed), only an output (i.e., it was newly created and now exists), or both (i.e., it existed before and now still exists). Within OCEL2.0, this information is not always extractable, particularly for objects that are only associated with a single event. To cover this, sOCEL introduces qualifiers for operation-to-workpiece relationships, namely "input", "output", and "on", where "on" is used to indicate a dual role of "input"and "output". Furthermore, we recommend using the object-to-object qualifier "created from" to encode material origins when one workpiece object is transformed into another. This enables tracing of material origins, even if multiple workpiece objects relate to an event.

4.3.2 Sustainability Attributes in sOCEL

All impacts are represented and quantified in sOCEL as attributes. We classify these attributes as motivated in Section 4.2 into *impact parameters*, *impact indicators*, and *impact scores*. As OCEL2.0 does not support storing metadata for attributes explicitly and to easier recognize entries in a log and our examples, we employ a specific attribute string format representing these classes. To easier differentiate between categories, the prefix identifiers ("p_", "i_", and "s_") are used in the following examples. Units of attributes are defined by the postfix, "[⟨UNIT ⟩]" i.e. "[kg]".

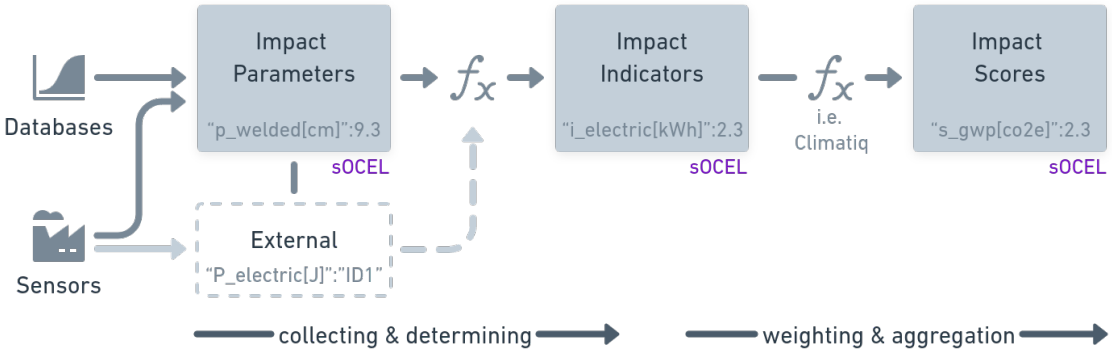

Figure 4.8: Processing of Sustainability Data Attributes in sOCEL Preprocessing

Impact Scores, such as `s_gwp[co2e]`, are attributes of operation events or workpiece objects and qualify and quantify the relative impact caused by an event or caused due to the existence of an object. A score is calculated based on impact indicators of this event or object instance, as defined by an applied assessment standard. An unquantified impact score indicates that there is missing information required for its calculation.

Impact Indicators, such as `i_waste_steel[kg]`, are attributes of operation events or workpiece objects. An impact indicator qualifies and quantifies an impact aspect, which the instance causes. An unquantified impact indicator indicates missing information.

Impact Parameters, like `i_welded-length[cm]`, can be attributes of operation events, location objects, workpiece or process-resource objects. Compared to the first two attribute classes, it is used in a in sOCEL unspecified manner, to bridge data gaps, allow workarounds and include required data, not integrable otherwise.

The placement of an impact indicator and score in a OCEL encodes an allocation of this (quantified) impact, as summarized in Figure 4.9. Is an indicator quantified larger than zero for an operation event, then this represents a caused impact of this event. Similar for workpiece objects: A quantified indicator represents an impact attributed to the (physical) object's existence, which can be both, due to entering or leaving the system under study. This is distinguishable by the indicator identifier. In example, `i_waste[kg]` implies the impact of an object leaving and `i_inital_material[kg]` implies the upstream impact of an object entering. To avoid double counting, quantified impacts are only allocated once to an event or object. An indicator or score with an attribute value of zero indicates that this instance doesn't cause an impact (aspect). If an impact is not allocatable directly, it can be measured or stored as an impact parameter of location objects, to be accessible for later analysis. In the context of manufacturing, we list impact parameters and indicators typically considered for evaluation in Figure 4.9.

The process for integrating this data into process mining is outlined in Figure 4.8. Indicators are initially specified by the assessment method. During log creation, parameters, and indicators are quantified, but potentially not entirely. In preprocessing, parameters are used to determine unquantified indicators, which are then used to calculate scores for a more comprehensive understanding. This conversion enables the inclusion of different parameters for heterogeneous data that requires preprocessing, for example measurements stored as time series.

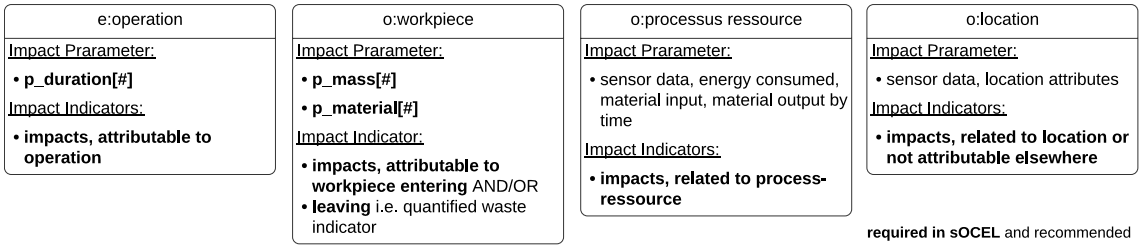

Figure 4.9: Recommended Impact Parameters and Indicators

4.3.2.1 Impact Parameter and Time Series Machine Data

According to research, resources such as CNC machines display continuous consumption patterns that vary across idle, active, and other operational states [37]. When manufacturing processes at the production process level are logged and impact data is stored as discrete attributes, continuous consumption and idle state consumption are not captured. To address the issue, we describe the possibility to use impact parameter in sOCEL process-resource objects, to link to a specific instance of log external time series data (machine data). Tacke Genannt Unterberg et al. formalized a time series machine data format that connects *objects* specifying where data is captured to *measurements*, as well as *event* entries defining what happened. These entries are linked to timestamps indicating when an entry is recorded. In the context of our example splitting operation, the log format stores *events* such as startup, fast processing started, slow processing started, empty, and power off, as well as respective power consumption *measurements* by timestamps.

By linking sOCEL process resources to machine data records, consumption patterns can be allocated to events in a preprocessing step as well as analyzed in detail, improving the granularity and accuracy of impact assessments. Consumption can be translated into discrete event attributes. In sOCEL, machine data records can be linked to process resource objects, by defining an impact parameter representing a pointer. In our examples, we use the prefix "P_" for parameter, representing a pointer.

4.3.3 Formal Definition

After describing the concept of attributes and classes we introduce, we provide the following formalization of sOCEL, based on definition 2.2 and definition 2.1. The definition supplements the previously explained requirements for use and application. It adds the requirement of a qualified relationship between operations and workpieces, which is necessary to track the entire material flow.

Definition 4.1 (sOCEL)**.** Let $AC = \{\text{parameter, indicator, score}\}$ be the set of attribute classes and $EC = \{\text{operation}\}$ the set of event-type classes, and $OC = \{\text{pr, workpiece, location}\}$ be the set of object-type classes. Further, let $Q = \{\text{created_from, on, input, output, at}\} \subseteq \mathbb{U}_{qual}$ be a set of sOCEL qualifiers. We define a sustainability data enriched Object-Centric Event Log (sOCEL) as a tuple $(L, \mathit{attclass}, \mathit{objclass}, \mathit{evclass}, Q)$ with

- $L = (E, O, EA, OA, \mathit{evtype}, \mathit{time}, \mathit{objtype}, \mathit{eatype}, \mathit{oatype}, \mathit{eaval}, \mathit{oaval}, E2O, O2O)$ as defined in definition 2.2 and
- $\mathit{attclass} : (EA \cup OA) \nrightarrow AC$ assigns an attribute class to an attribute (not all the attributes are mapped),

- $objclass : \mathbb{U}_{otype} \nrightarrow OC$ assigns object classes to object types (not all the object types are mapped),
- $evclass : \mathbb{U}_{etype} \nrightarrow EC$ assigns event classes to event types (not all the event types are mapped),

such that

- for any $e_{op} \in \{e \in E \mid evtype(e) \in dom(evclass)\}$ and any $o_{wp} \in \{o \in O \mid objclass(otype(o)) = \text{workpiece}\}$: $(e_{op}, q, o_{wp}) \in E2O \implies q \in \{\text{on}, \text{input}, \text{output}\}$ *ensuring all relationships between operations and workpieces to be qualified, specifying whether they are an input, output or being operated on.*

Although it is possible to formalize the operation event class using a construct like $EC \subseteq \mathbb{U}_{evtype}$, the definition presented here takes a different approach to encourage extensibility by highlighting the potential for adding other event classes. This could be useful for evaluating non-manufacturing processes, such as the use phase or logistics, that are not in the focus of this work.

Further, we describe two different variants of an sOCEL, we refer to in the following work.

- A **virgin sOCEL** describes a sOCEL as defined above, and all indicators and scores are identified and included as attributes, representing the relevant (physical) impacts. It is not required that indicators and scores are quantified.
- A **complete sOCEL** describes a virgin sOCEL where all scores and indicators are quantified.

4.3.4 Example Application

To be able to analyze the impact of processes a sOCEL is created from, we strive towards quantified impact scores serving as key (environmental) performance indicators. Ideally, an OCEL is enriched with scores, as shown in an abstract representation of a sOCEL with three different recorded events in Figure 4.10. Impacts are quantified with a comparable measure, CO_2e. We can identify hotspots in a process model, like the two workpiece objects crossing the system boundary. Scores, however, are not measured, but calculated based on impact indicators, that are attributes of the same entities in a sOCEL. Indicators of this example are visualized in Figure 4.11.

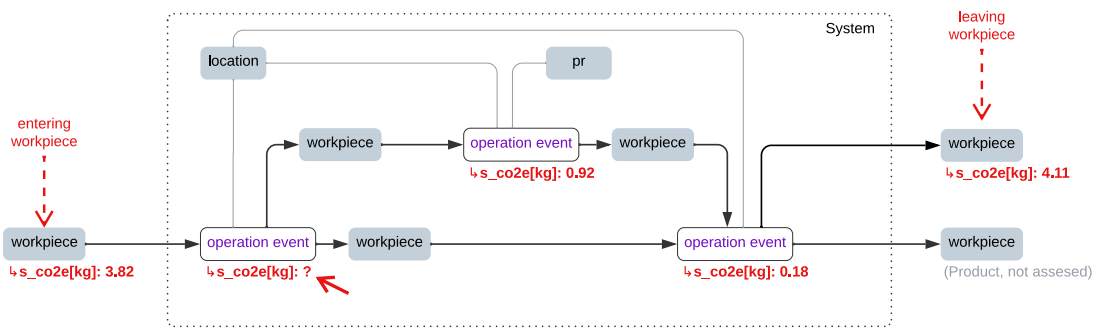

Figure 4.10: Three-Event-sOCEL Example - Scores

Considering the marked event, we see that steel waste is created but currently unquantified. Therefore, we consider the impact parameter `p_weigth[kg]` from input and output objects of this event, as shown in Figure 4.12. Based on the differences in mass, we calculate in a preprocessing step the missing information on waste. Based on the quantified missing indicator value, we calculate the missing score and as a result, we have a complete version of our primer virgin sOCEL.

In addition, the total energy consumption is logged as a time series and linked with a location object in this example. This data could be used to validate the relevance of insights gained from process analysis by comparing, i.e., system total energy usage with energy usage tracked in sOCEL events. Further, idle energy consumption can be considered that is not allocated to events.

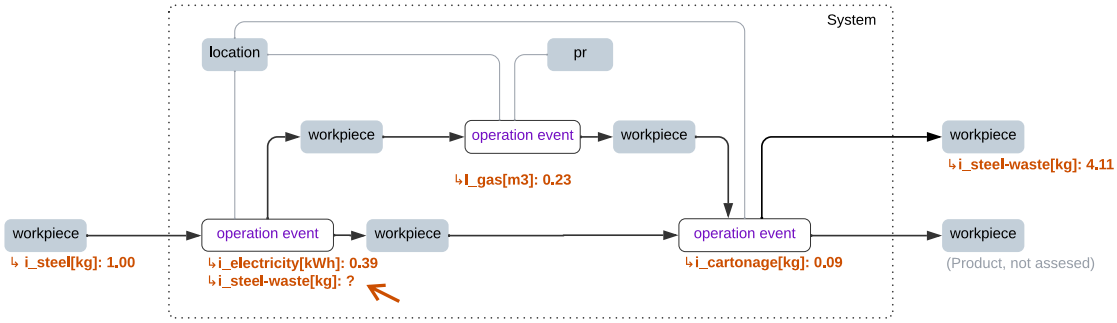

Figure 4.11: Three-Event-sOCEL Example - Indicators

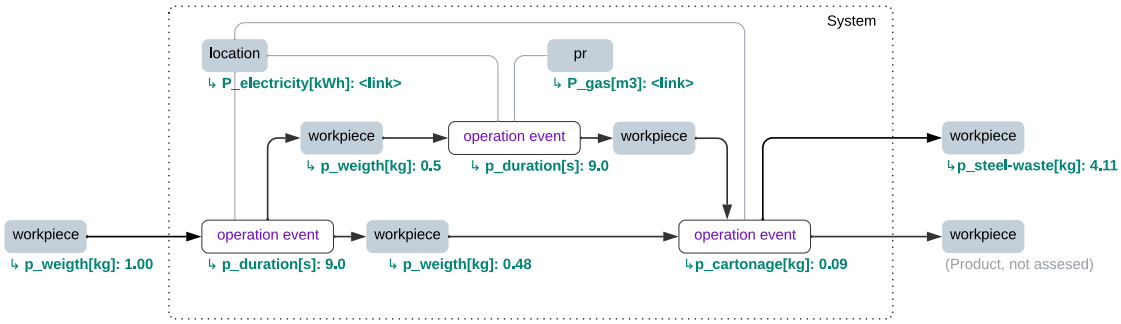

Figure 4.12: Three-Event-sOCEL Example - Parameters

4.3.5 Example sOCEL

To illustrate the proposed guideline and a resulting sOCEL on log level, we present a simplified manufacturing sub-process of two events originating from our running example. This segment is shown in Figure 4.13. Given this two event process, the first step is to define both the impact indicators and the corresponding impact scores based on the applied assessment method. In this example, Climatiq API is used for assessment. Therefore, an impact score is solely the element `gwp-co2e[kg]` as defined by the API. The identification of the indicator involves reviewing the collection of in general relevant indicators listed in Table 4.3, identifying those that are relevant to the process under study, and then selecting the most appropriate API entry. In this case, indicators are (shortened) activity-ids of these API-entries.

This example involves the continuous measurement of electricity, nitrogen gas used by a laser (N_2), and natural gas through sensors. Except for continuous consumption of electricity at facility level (i.e. including lighting, heating, etc.), it is assumed that consumptions are attributable to operation events. Steel consumption is recorded by logging workpiece objects, noting their initial weight and material as an impact parameter. The impact of combustion emissions is measured in burned gas by Climatiq and is therefore equated to gas consumption. Steel waste is not directly measured and requires additional impact parameters to be logged to determine its impact. To address this, we take parameters for steel weight into account, `p_weight-DC01[kg]` for steel coils and steel sheet workpiece objects, which allows us to compute the difference in steel that is part of the inputs and outputs. The resulting log is represented as an ER diagram in Figure 4.14.

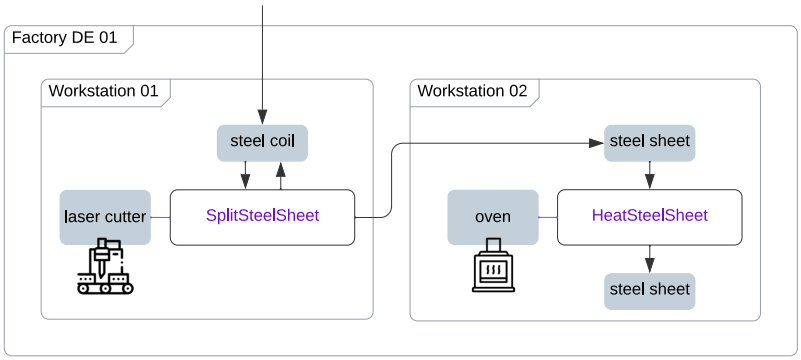

Figure 4.13: Two-Event-Process Example

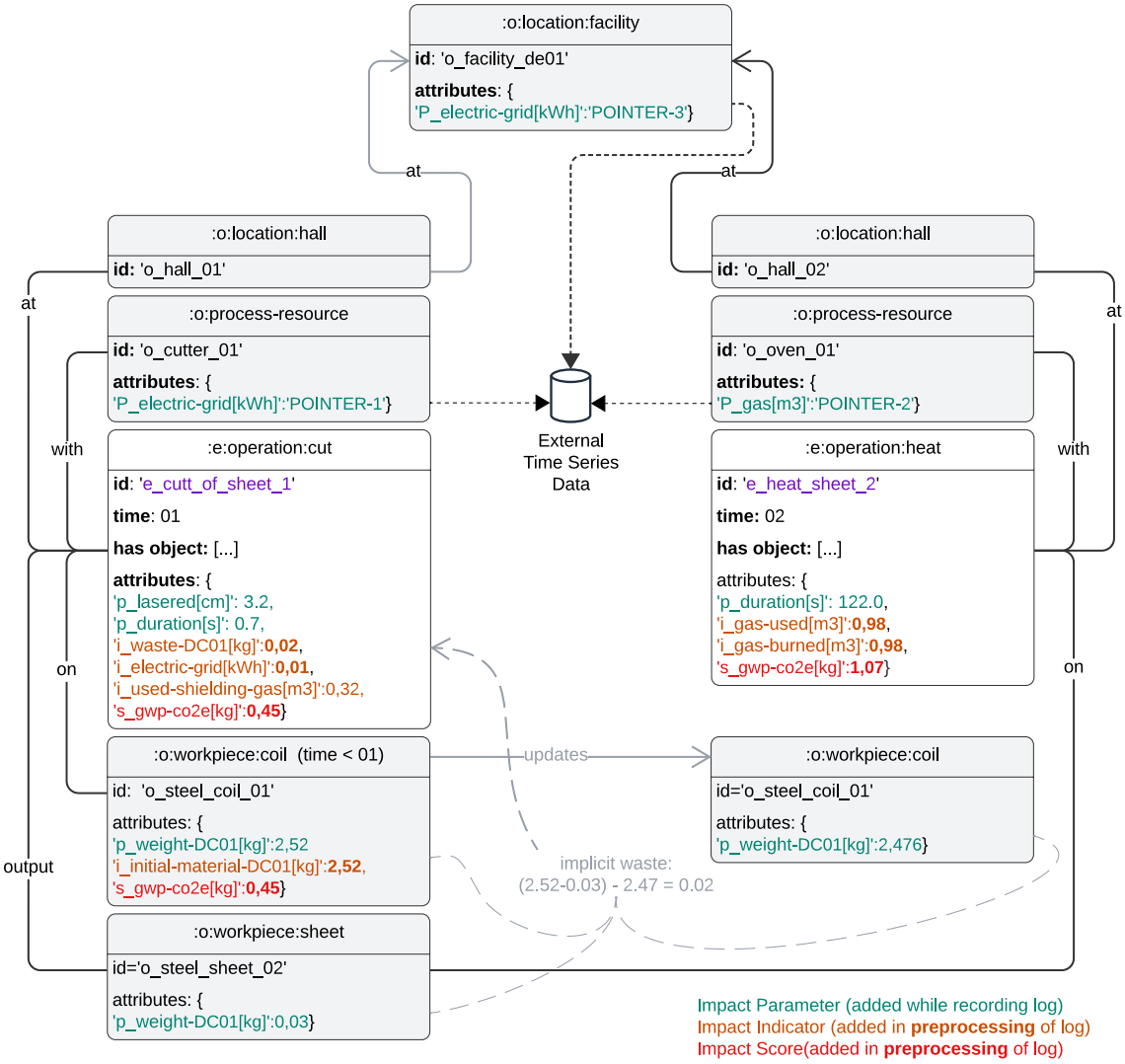

Figure 4.14: Two-Event-Process: sOCEL with two recorded event instances

Chapter 5

Implementation

After specifying a standard for the inclusion of sustainability data into OCEL, we demonstrate in a next step the generation and application of an sOCEL. In this chapter, we outline our methodology for creating an example event log. We detail the associated manufacturing process model, explain our design approach, and describe the method used for event log generation. Further, we demonstrate how APIs can be used to automatically enrich sOCEL with sustainability data and describe a Python tool implemented for this purpose.

5.1 Designing a Manufacturing Process Model

To generate a log, we designed a process model. Our result is a simplified hinge production line, modeled as a Colored Petri Net (CPN) in CPN-Tools[1], a tool for editing and creating colored Petri nets. This model is an idealized representation of an exemplary manufacturing process. It is designed to examine and present the use case of SD in OCPM and does not realistically represent a specific manufacturing process in detail.

The model's design process was structured into four steps. First, the workpiece flow was outlined, identifying the sub-products used in hinge production, as shown in Figure 5.1. Next, control flow elements were incorporated, adding batching and queuing behaviors as well as quality control procedures. Sustainability data relevant to this example were integrated into events and objects following the sOCEL specification. Finally, the simulation of values such as waiting times and weights now includes randomness and deviations.

[1] cpntools.org

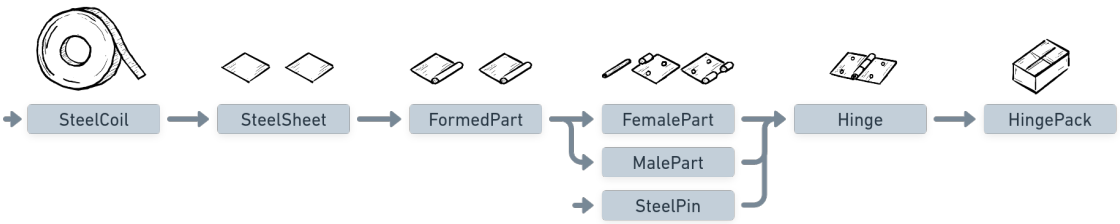

Figure 5.1: Workpiece Objects and Workpiece Flow

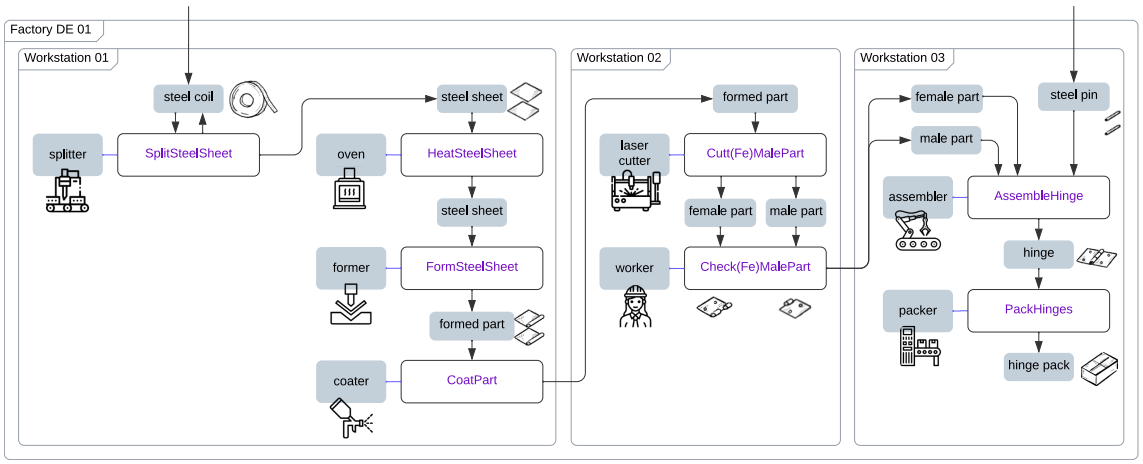

Figure 5.2: Example Manufacturing Process: A Hinge Production Line

In a manufacturing, DIN 8580 provides categorizes of operations used in the initial literature review. To add a modeling perspective to those operations, we consider a study by Erasmus et al. identifying typical operations and model fragments of manufacturing processes. To enhance our model's ability to represent common process structures, we incorporated elements corresponding to these model fragments. Erasmus et al. define four operation categories, production, inventory, quality and maintenance operations, where each has multiple subcategories. Our model includes operation representing all production operations (mass reducing, mass conserving shaping, assembly and non-shaping or coating) and inventory operations (packing, handling, storing, transporting). As it increases model complexity without justified relevance, we do not model process quality operations as well as maintenance operations.

5.1.1 A Simplified Hinge Production

In this section, we present the resulting manufacturing process as well as constructed and purposely included flaws and bottlenecks related to sustainability aspects that could potentially be discovered by process mining techniques. The manufacturing process is already introduced as a running example and additionally presented in Figure 5.2. Impact indicators related to this process, as well as unquantified indicators requiring preprocessing, are listed in Figure 5.3.

In this example scenario, hinges are produced in a German facility. The production process is divided into three workstations, all supervised by a single worker. The production line is started by this employee workstations for workstations every morning at 7:30. The employee pauses work for a lunch break at 12:00 pm and then continues working from 1:00 pm until 3:00 pm. The production process starts with two primary inputs: steel coils and steel pins. The steel pins, also known as steel rods, are each 3.0 cm long and are used to hold the leaves together. The steel coils hold a rolled steel strip that is approximately 0.3 centimeters thick and 3.0 centimeters wide. This steel strip is then processed into hinge leaves through the following steps.

In the first operation of our model, one steel coil at a time is processed at the first workstation. Each coil is continuously cut into steel sheets that are 3 cm in length. As

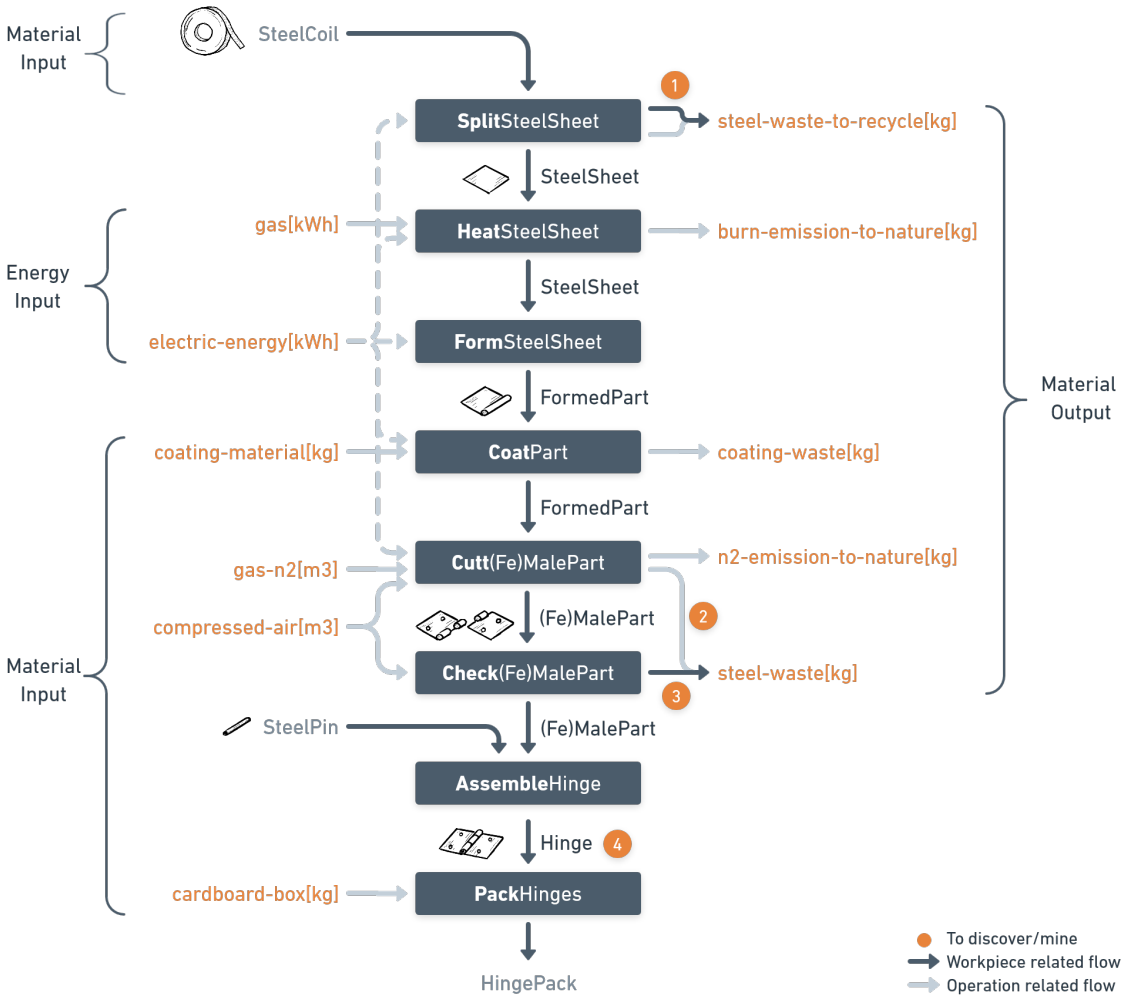

Figure 5.3: Operations and Simulated Impact Indicators

these strips are cut, we assign each resulting steel sheet an identification number. This process continues until there is only an unusable steel remainder of the steel coil, which we do not track further in this process. This remainder needs to be classified as waste (1) in preprocessing. The amount of waste created during the production of rolled steel strips varies depending on the initial length and could be optimized. The cut steel sheets are handled in a line and heat-treated in a gas oven. A different oven technology could reduce impacts. Then, each 3.0 cm steel sheet is rolled on one side to form a barrel for the insertion of hinge pins. This operation is performed using a hydraulic press powered by electricity. In our model, we refer to these produced parts as *FormedParts*. Finally, in a very energy intense step, plasma coating is conducted using a specialized machine. The *FormedParts* are automatically collected in a transport carriage. If the carriage reaches full capacity, an alert signals the operator to manually move the batch of *FormedParts* to the second workstation.

At the second workstation, *FormedParts* are shaped into alternating male and female hinge leaves (*MalePart* and *FemalePart*) using a laser cutting machine. This is achieved by cutting out specific sections from each part. Notably, the method generates more

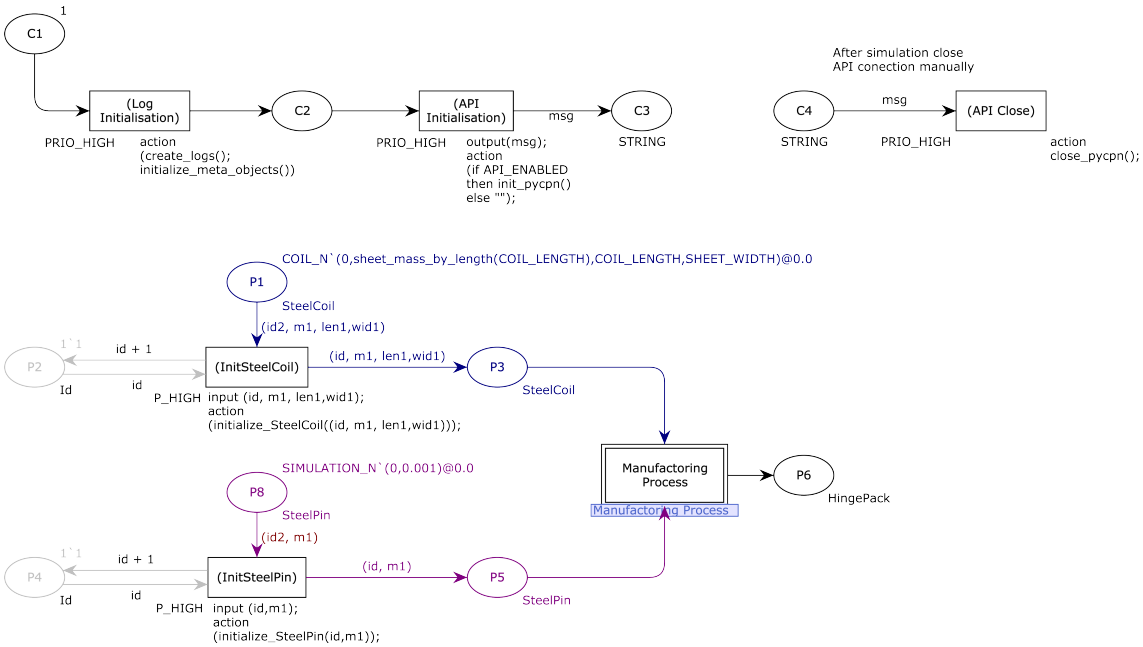

Figure 5.4: CPN Model - Meta Layer

waste compared to an alternative process where the male and female parts are directly cut out from the coil and then shaped further. The metal waste containing the coat is hypothetically not recyclable, which is reflected in a separate impact indicator. The waste created during this operation (2) is not recorded, however, the weight of the workpiece object before and after the material removal is recorded.

In the next step, the worker manually checks both *FemalePart* and *MalePart* for quality criteria, which is mass, representing various possible measurable attributes, including geometrics and material properties. If everything is within tolerance, the parts are placed on a moving band to the third workstation. If not, the part represents waste (3), which again is to be quantified in a preprocessing step. At the next station, an assembling machine uses one *MalePart*, one *FemalePart*, and one *SteelPin* to create a Hinge. The material origin of a hinge or its parts could potentially be discovered (4). A set of 10 hinges is automatically buffered and placed into a cardboard box by a packaging machine.

5.1.2 Modeling and Generation using CPN-Tools

Modeling a manufacturing process, the workpiece flow and logic or control flow must be included. To achieve this, substitution transitions are utilized, and the model is divided into three layers: the meta level, the manufacturing process layer, and the logic layer. The meta level shown in Figure 5.4 initializes a socket connection to our python tool we introduce separately in the following section. It also initializes input objects, *SteelCoil* and *SteelPin*. The manufacturing process layer provides an abstracted view of process steps shown in Figure 5.5 and the logic layer contains control flows describing worker and machine behavior, as shown for splitting operation in Figure 5.6. The resulting model consists of 28 transitions, 130 arcs, and 45 place instances in total. A simulation for four scenario days requires approximately 57.000 fired transitions.

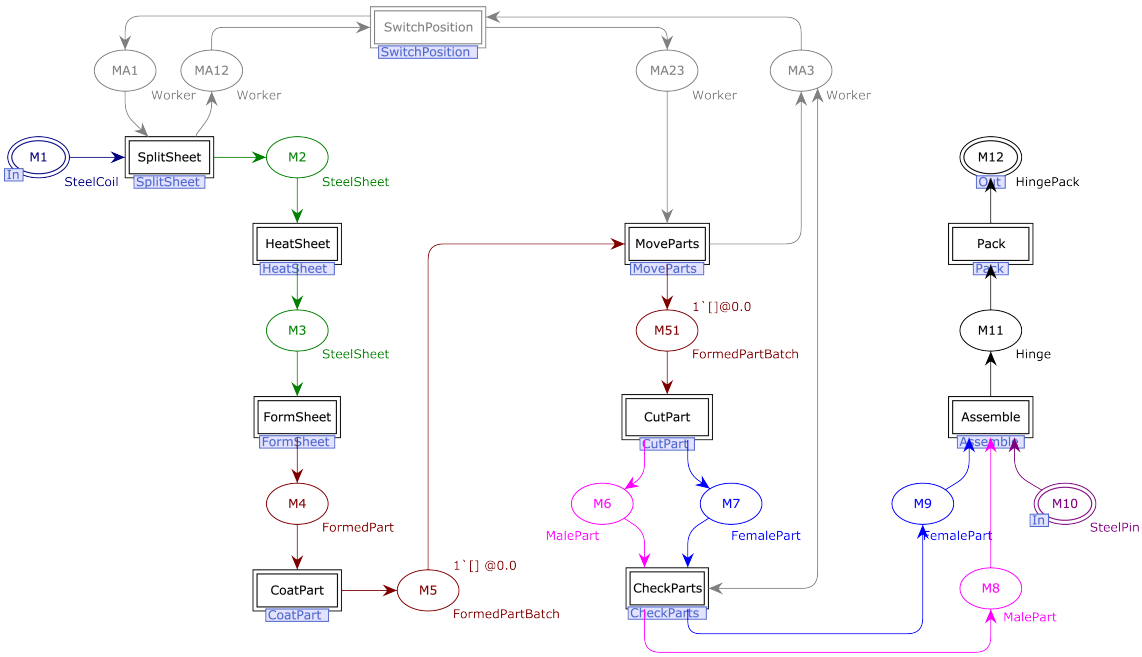

Figure 5.5: CPN Model - Manufacturing Process Layer

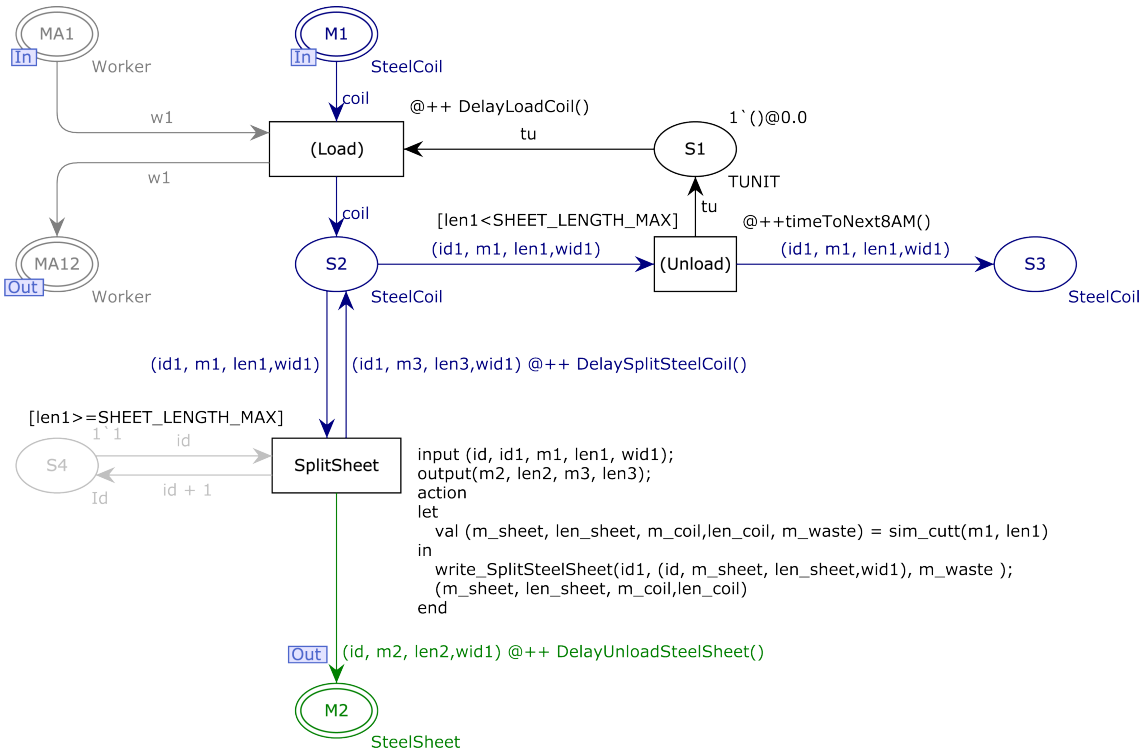

Figure 5.6: CPN Model - Logic Layer (SplitSheet)

Object Class	Object Types
workpiece	"SteelCoil", "SteelSheet", "FormedPart", "MalePart", "FemalePart", "SteelPin", "Hinge", "HingePack"
location	"Workstation", "Facility"
pr	"Machine", "Worker"

(a) *objclass*

Event Class	Event Types
operation	"SplitSteelSheet", "HeatSteelSheet", "FormSteelSheet", "CoatPart", "CuttMalePart", "CuttFemalePart", "CheckMalePart", "CheckFemalePart", "AssembleHinge", "PackHinges", "MoveParts"

(b) *evclass*

Table 5.1: Excerpt of sOCEL: Objects and Event Class Mapping

ocel_id	**ocel_time**	**p_duration[s]**	**s_co2e[kg]**	**i_electric[kWh]**	**i_steel-waste-rec[kg]**
e_split_1	2023-04-03 07:38:25	6.45	0.0068	0.0099	0.00042
e_split_2	2023-04-03 07:38:31	6.55	0.0067	0.0098	0.00042
e_split_3	2023-04-03 07:38:38	6.67	0.0070	0.0103	0.00042
e_split_4	2023-04-03 07:38:45	6.63	0.0069	0.0101	0.00042
e_split_5	2023-04-03 07:38:52	6.71	0.0070	0.0103	0.00042
e_split_6	2023-04-03 07:38:58	6.67	0.0071	0.0103	0.00042

Table 5.2: Excerpt of sOCEL: Events of Type *SplitSteelSheet*

In CPN Tools, Standard ML, a functional programming language, can be integrated into colored Petri nets modeled, to handle complex calculations and behavior. The OCEL2.0 log is generated by integrating SML code into the CPN model as shown in Figure 5.6, which provides the functionality to write log entries on transition execution to CSV-files. The SML code was initially based on a similar model created by Benedikt Knopp and Nina Graves. After simulating until a specified number of starting objects, represented as tokens in the colored Petri net, are consumed, the CSV files are converted to a XML file according to the OCEL2.0 standard [23]. The event log generated contains ∼3850 events of 11 different types, as well as ∼23700 objects from 12 different object types. To extend this OCEL2.0 to sOCEL, we define event and object classes required in Table 5.1. Attribute classes are implicit mapped by the corresponding prefix. Excerpts of the resulting log with a focus on the first event *SplitSteelSheet* are presented in Table 5.2, Table 5.3 and Table 5.4.

Due to the lack of REST functionality in SML, a connection was made to a Python backend server as described in the next chapter to enable computation via REST API requests. The backend server provides values that are subsequently used and logged in the CPN simulation.

5.2 Tool Implementation

To generate a sOCEL we need to insert and quantify, impact indicators, parameters, and scores. While parameters and indicators are calculated during simulation as designed and modeled in CPN-Tools and SML code, the impact scores are included automatically into the event log using values received from Climatiq API. However, as we use Standard ML and CPN-Tools for simulation and this setup does not support REST-API requests, a

ocel_id	ocel_time	p_mass[kg]	p_material-[EN10130]	p_init-len[cm]	p_width[cm]	i_material-steel[kg]	s_co2e[kg]
o_coil_1	07:30:00	31.91076	DC01	4522.5	3.0	31.9107	80.7343
o_coil_2	07:30:00	31.91076	DC01	4522.5	3.0	31.9107	80.7342
o_coil_3	07:30:00	31.91076	DC01	4522.5	3.0	31.9107	80.7342
o_coil_4	07:30:00	31.91076	DC01	4522.5	3.0	31.9107	80.7342

Table 5.3: Excerpt of sOCEL: Objects of Type SteeCoil

ocel_source_id	ocel_target_id	ocel_qualifier
o_workstation_1	o_facility_1	located at
o_machine_01	o_workstation_1	located at
o_steelsheet_1	o_coil_1	created from
o_steelsheet_2	o_coil_1	created from
o_steelsheet_3	o_coil_1	created from

(a) O2O Relationships

ocel_event_id	ocel_object_id	ocel_qualifier
e_split_1	o_coil_1	on
e_split_1	o_steelsheet_1	output
e_split_1	o_workstation_1	located at
e_split_2	o_coil_1	on
e_split_2	o_steelsheet_2	output

(b) E2O Relationships

Table 5.4: Excerpt of sOCEL: Objects and Events Relationships

Python tool is implemented that allows SML to trigger API calls during simulation.

5.2.1 Usage

The implemented tool is used in two phases, in a configuration phase and a simulation phase. The configuration phase corresponds to the modeling process of a colored Petri net. If the process modeler adds objects or events with impact indicator attributes to the model, they can additionally add a method evaluating these indicator values, returning a corresponding impact score. To be able to use this method, they need to set a `request_id` connecting a method call in the code of the CPN-Simulation to a specified Climatiq API request that is to be executed by the tool. Using the Climatiq Data Explorer[2] the modeler browses available impact entries and selects a most fitting one and links the specified API ID as well as additional parameter Climatiq requires to the `request_id` using a provided Jupyter notebook. We showcase these usage steps in Figure 5.8 and Figure 5.9.

In the simulation phase, the CPN is simulated using CPN-Tools. While impact indicator values are calculated during simulation, predefined calls are performed, containing a `request_id` and value as shown in Figure 5.7. In this phase, our tool serves as a python backend and receives these request via socket communication. Depending on the configured method linked to the `request_id` a local calculation or a configured API requests is performed. The returning impact scores are sent back via socket communication and are now accessible in the CPN simulation.

[2] www.climatiq.io/data

```
val cardboard_material_kg = 12.4;
val co2e = call("id_cardboard_material_kg", cardboard_material_kg);
```

Figure 5.7: SML Code Snippet of API-Call via Python Backend

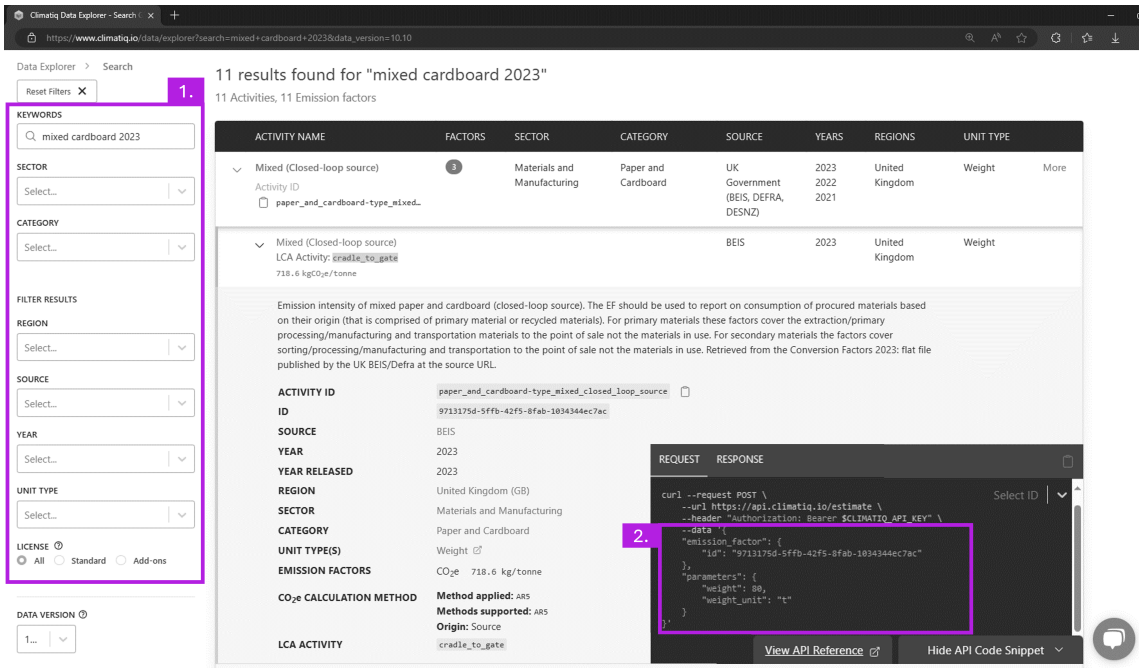

Figure 5.8: Screenshot of Usage 1/2: **1.** Search and select the best fitting Climatiq API-Entry. **2.** Copy API-request ID & parameter (optionally adjust these).

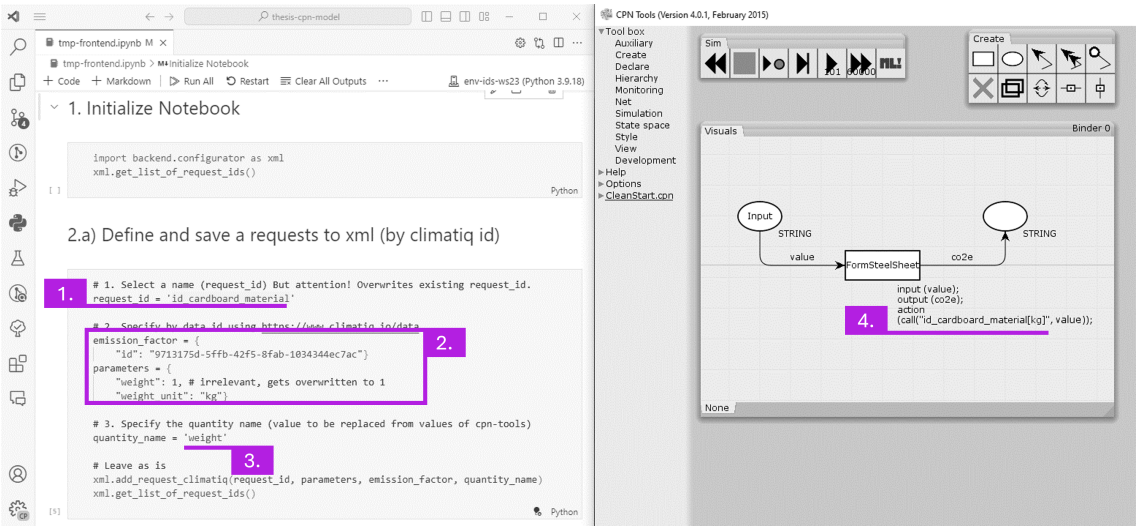

Figure 5.9: Screenshot of Usage 2/2: **1.** Set `request_id` **2.** Insert API-request ID & parameter (optionally adjust) **3.** Define which parameter is variable and run the code in this cell. **4.** Use `call` with `request_id` in CPN simulation.

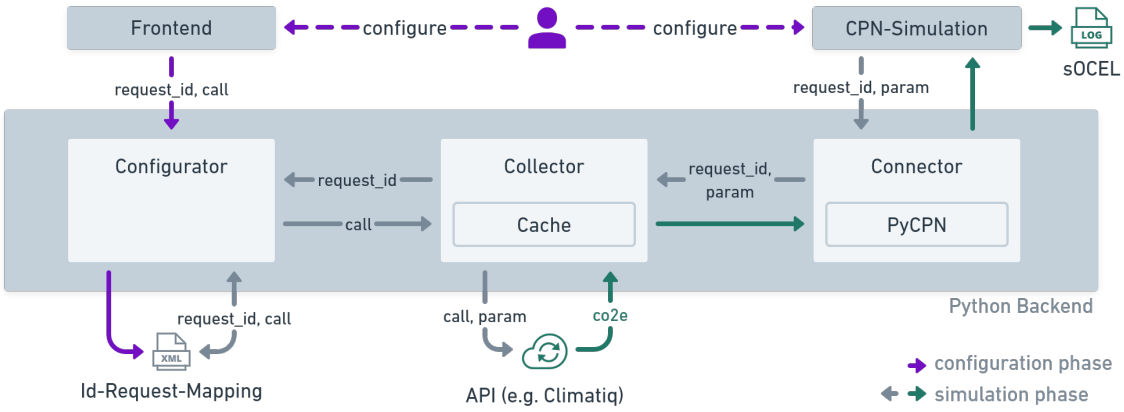

Figure 5.10: Tool Architecture & Integration with CPN Tools

5.2.2 Architecture

The tool architecture reflects this usage workflow and consists out of four main modules: A Jupyter notebook serving as a pragmatic frontend, a *Configurator*, *Collector* and a *Connector* module, as visualized in Figure 5.10.

Initially, the user interacts via a Jupyter notebook with the *Configurator* module. Since modeling involves working with SML code, a sophisticated user interface would not add any value in this stage of tool development. The *Configurator* module provides methods to write a configured link between Climatiq-API calls and SML calls to a configuration file. This XML configuration file maps impact indicators (CPN-model) to Climatiq-API requests.

In *Connector* the communication with a running CPN-Tools simulation is organized. We use PyCPN[3] to establish a socket communication, where the tool acts as a server and the running SML simulation as a client triggering requests and awaiting response. The *Connector* waits for calls and forwards any requests to the *Collector* for processing.

As Climatiq limits API-requests and our simulation triggers approximately 50.000 requests per run, the *Collector* module is introduced and caching is implemented. For a given `request_id` the *Collector* determines where the factor is accessible by looking up this information in the configuration XML managed by the *Configurator* module. If configured, an API-request is performed as specified to receive a factor corresponding to the request, potentially received by cache. The storage and usage of custom factors stored locally is supported.

This architecture and tool-design result in requirements that a user must consider. Indicator values linked with calls via `request_id` in the CPN-Model must be equivalent to the impact refers to in Climatiq API, ensuring realistic assessment. Moreover, each sustainability value counter must be used and modified in such a way that units are used consistently.

Our tool code is publicly available on GitHub[4]. The current implementation is geared towards Climatiq-API. The tool is, however, designed to be extendable for further APIs,

[3] github.com/vgehlot/PyCPN
[4] github.com/marcoheinisch/thesis-sOCEL

as well as complex local calculations not implemented in Standard ML. Future development could create a user interface that simplifies the process of selecting and linking `request_id` to Climatiq API calls. Alternatively, it could be adopted in a way that allows it to be used on previously created event logs to enhance them, rather than interacting with a simulation.

Chapter 6

Evaluation of sOCEL

In this chapter, we evaluate the sOCEL framework's ability to include sustainability data. First, we summarize that sOCEL are generatable. Then, we demonstrate what information sOCEL can contain and what is extractable. Third, using the example sOCEL scenario of a hinge production, we evaluate how sOCEL performs regarding requirements and potentials of PM for carbon accounting identified by Brehm et al..

6.1 Generating sOCEL

In the previous chapter, we demonstrate that sOCEL is generatable for a manufacturing process. Assuming recorded machine data and workpiece specifications or measurements at the workstation, sustainability data relevant for the example process is includable in sOCEL. We have demonstrated that services such as Climatiq can be utilized during log creation to automatically include sustainability data, such as impact scores.

In Table 6.1 we show and list all identified sustainability aspects from our literature review as well as how they can be included in sOCEL. While we focus on the three main sustainability elements, material input, output, and energy input, for the specification of sOCEL and the generation of the example log, the list shows that various aspects can be additionally included if it is relevant for assessment. Further, we see that all data relations identified in the literature review can be mapped in sOCEL. For example, economic indicators like costs attributed to specific operations or the costs associated with an input workpiece are includable as impact parameters and indicators, i.e. `i_cost-of-used-steel[EUR]`.

Last, we note that the generated example lacks on log-external machine data that is referenced by parameters in the event log. Future work is required for specifying and implementing the integration of both, sOCEL and machine data.

6.2 Providing Insights

As sOCEL extends OCEL2.0 without changing the core, available OCPM algorithms and tools[1] are already applicable for log-based analysis. In this section, we assess the

[1] pm4py.fit.fraunhofer.de, ocelot.pm or ocpm.info

Sustainability Aspect	In OCEL*
Disposal costs	pr parameter, op indicator, wp indicator
Energy costs	loc/pr parameter, op indicator
Investment costs	loc/pr parameter
Material costs	pr parameter, op indicator, wp indicator
Labor costs	pr parameter
Sales price, profit	wp parameter/indicator
Transportation cost, warehouse cost	loc/pr parameter
Defective/returned products loss, warranty cost	wp object
Injury rate and others e.g. absenteeism	loc/pr parameter
Lost work days	loc/pr parameter
Exposures (noise, mist/dust, chemical concentration, temperature, hazardous, threats)	loc/pr parameter
Health and safety impacts (chronic illness rate)	loc/pr parameter
Emitted heat, light (outside), noise (outside) and radioactive emissions	loc/pr parameter
Cutting fluid loss	pr parameter, op indicator
Emissions to air (GHG/Pollutant)	pr parameter, op indicator
Liquid Waste (hydraulic Fluid, wastewater, lubricant Waste)	pr parameter, op indicator
Solid Waste (steel, disposal/recycle)	pr parameter, op indicator, wp indicator
Material used in Product (molten metal)	wp indicator
Material, resources (water, die lubricant, hydraulic Fluid, plunger lubricant, compressed air), processing gas (nitrogen), alcohol, powder coat, oil	pr parameter, op indicator, wp indicator
Burned Natural gas	pr parameter, op indicator
Energy consumption (by source, from renewable and nonrenewable sources)	loc/pr parameter, op indicator
Electric energy idle/standby/maintain consumption	loc/pr parameter

* wp:workpiece object, pr: process resource object, op: operation event

Table 6.1: Mapping of Sustainability Aspects to sOCEL

extractability of information from the artificially generated log, such as material flow and energy consumption. We also identify missing capabilities of our framework.

6.2.1 Material Input and Output

In sOCEL, materials (including material resources) that enter or leave the system under study are integrated using impact indicators like `i_coating_material[kg]` and respective impact parameters to determine them. These indicates are attributes of operation events or workpiece objects. If material inflow or outflow is included as a quantified impact indicator attribute of an operation event, as shown in a log excerpt in Table 6.2, this information can be extracted easily.

On the other hand, if a workpiece object represents a material inflow or outflow, interpretation is ambiguous without qualifiers. Considering the manufacturing operation event *SplitSteelSheet* that is part of our example process, we discover a workpiece flow shown in Figure 6.1. Without standardized relationship qualifiers in this example, it is unclear how object *SteelCoil* and *SteelSheet* relate to the system under study. Specifying qualifiers sOCEL defines ("input" , "output" , "on"), enables differencing between objects crossing

ocel_id	ocel_time	s_co2e[kg]	i_electric[kWh]	i_coating[kg]	i_coat-waste[kg]
e_coa_1	23-04-03 07:56:00	0.1751	0.25	0.0004	0.0001
e_coat_5	23-04-03 07:56:25	0.1764	0.25	0.0005	0.0001
e_coat_4	23-04-03 07:56:36	0.1750	0.25	0.0004	0.0001

Table 6.2: Excerpt of sOCEL: Event CoatPart

the system boundary (*SteelCoil* is input) where initial impact needs to be considered and newly created objects, like the output *SteelSheet*, that originate from an already assessed object *SteelCoil* and is qualified as output (but not input) of this event. Also note that this distinguishing between objects crossing the system boundary or not is not mineable here.

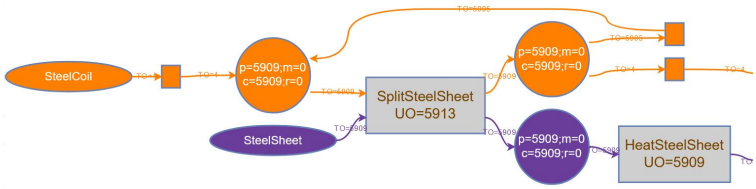

Figure 6.1: Excerpt of Discovered Object-Centric Petri Net

6.2.2 Workpiece Flow

For non-free flows, where workpiece objects occur in further operation events of a process, the information whether an object is input, output or both can be partly mined if not provided in the log by discovering a workpiece flow. In the example in Figure 6.1 *SteelSheet* is used in the subsequent event *HeatSteelSheet* so it must be output of *SplitSteelSheet*.

The workpiece-flow is of interest as it provides insight into production processes and the dependencies between operations, impacts and related workpieces. The required class definition in sOCEL predefines different process perspectives that can be adopted in process discovery. If we only consider workpiece objects for OCPM, a workpiece flow, as shown in Figure 6.2, is discoverable.

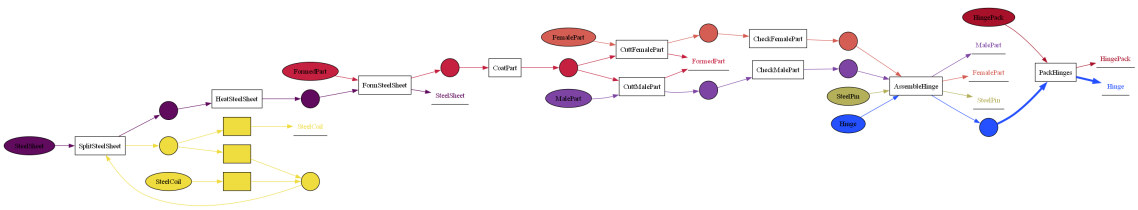

Figure 6.2: Discovered Object-Centric Petri Net - Only Workpiece Objects

6.2.2.1 Workpiece Inflow and Outflow (Wastes)

In sOCEL, each object entering the system boundary is associated with an initial impact indicator, such as `i_material-cold-rolled-steel-coil[kg]`, which reflects the environmental impact incurred during its upstream production processes.

Material outflows are includable and extractable using impact indicators, like `i_emitted-n2-gas[m3]` or `i_steel-waste-to-recycling[kg]` in workpiece objects. Waste is an important subtype of material outflow. It can be quantified either directly, based on the measured weight of the waste, such as steel chips, or indirectly. In the case of a mass-reducing operation like *SplittSteelSheet*, an indirect approach is to calculate the weight difference between the input workpiece(s) and the mass of the (multiple) output workpiece(s) of the same material. Depending on the process, a second measurement of weight may occur at a later operational event, which must be considered in the preprocessing step.

6.2.2.2 Material Origin

If multiple objects relate to a manufacturing operation in an event log, it is not specified how material origins of workpiece objects are encoded. As sOCEL specifies the "created from" object-to-object relationship qualifier, tracking of material origins is even in individual cases possible, as visualized in Figure 6.3. However, this approach does not encode a quantification, which therefore must be mined separately. In the context of manufacturing, this qualifier supports the discovery of Bill of Materials (BOM) of a given end product. The BOM is a list of the subassemblies, intermediates, materials, subcomponents, parts, and quantities required to manufacture a product, representable as network [46].

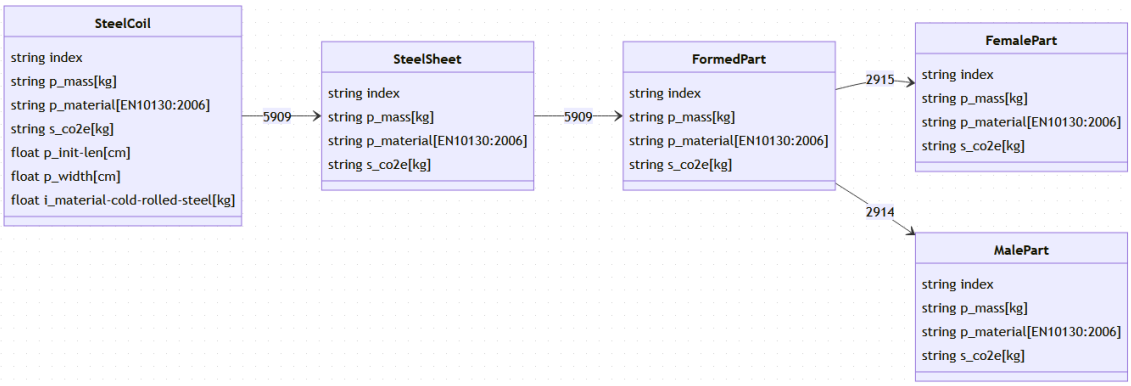

Figure 6.3: "created from"-Relationships Visualized Log Based using ocelot.pm

6.2.2.3 Transport and Logistic

Both explicit and implicit modeling of transportation activities is possible within the sOCEL framework and included in the generated example for intra facility movements. If a workpiece participates in two operation events one after another and these operations are linked to two distinct location objects, movement is implicated, such as between two workstations (second and third workstation in the generated example). This movement can be accounted for in the pre-processing phase, either by assigning impacts to implicit transport activities or by adding corresponding events explicitly to the log.

Explicitly, transport events include start and end locations or a distance as parameters as well as transport modes, all of which are required for estimating environmental impacts, for example, using Climatiq. However, it's important to note that our work is primarily focused on the gate-to-gate processes and does not extensively cover assessment of logistics operations.

6.2.3 Energy

Similar to material flow, is information on energy consumption included as an impact indicator. Therefore, it is aggregatable for an operation event and given time frame or usable for further analysis and OCPM. Energy consumption can also be included as parameter of location objects and process resources, i.e., as time series of consumption measurements. The generated log only theoretical includes such links. Further research and specification on this topic is needed.

6.2.4 Not Allocatable or Missing Impacts

In sOCEL, impact indicators and evaluations can be assigned to either input or output workpieces or operations, meaning that this impact is also assigned and allocated to this object or event. The location object class enables the linking of impacts that cannot be directly allocated to events and workpiece objects and could not otherwise be mapped in the log. This makes such effects accessible for later analysis, where they can be assigned to specific activities or interpreted.

Some impact indicators considered out of scope for automated assessment using PM. This is typically due to the long-time scope of impacts like land use or due to company internal data availability and typically unquantified aspects, like child labor. As these impacts are however relevant for assessment, sOCEL provides with location objects the integration of such data, even if it is inaccessible for PM as such.

The specification of event, object, and attribute classes in sOCEL as well as recommended relationships of these, include domain knowledge of manufacturing and sustainability assessment. Validating these requirements, allows the discovery of violated or conflicting aspects, like missing input or output event-to-object relationships and missing or unquantified impact indicators. While PM by its nature depends on data quality and completeness [4], the in sOCEL introduced specifications contribute to improving these aspects in preprocessing.

6.3 How sOCEL Supports Carbon Accounting

Brehm et al. performed an empirical study on how process mining can support carbon accounting, resulting in the identification of the following requirements and potentials. For each, we discuss, how sOCEL supports these as relevant identified aspects.

Retrieving and Integrating Carbon-Related Data: As PM requires and relies on data and is highly dependent on data quality and completeness, data availability is identified as a key requirement. In particular, information on carbon emissions is considered to be typically unavailable. APIs are mentioned as a potential data source to fill such data gaps. Event logs according to sOCEL provide impact indicators and parameters needed to perform API requests or calculate missing information. In our case-study, we show in practice how APIs such as Climatiq are used to supplement external sustainability information, in this case impact scores (`s_co2e[kg]`) that add a relative measure of sustainability to sOCEL. Further, it is also mentioned that basic data such as weight or size of a workpiece are often not available in production processes. This issue cannot be fully addressed by sOCEL as data availability is an issue due to the nature of PM. Tho concept of three classes of sustainability attributes, as well as impact parameters in sOCEL

provides potential to link or include data sources that require preprocessing.

Technical and Process Know-How: The authors note that PM for carbon accounting alone does not reduce emissions. For "transformation and contextualization of the data" domain experts are needed, first a technical expert to incorporate data, second a process expert to discover reduction potentials. SOCEL itself does not address this requirement directly. Further, the combination of primary and secondary data is identified as important for the quality of the insights PM provides. In sOCEL, given three classes of sustainability attributes, primary (measured) impacts and impacts calculated based on secondary sources (reference values) can be integrated as an impact attribute. Keeping the three layers separate also allows corrections to be made even after the event log has been created.

Create Transparency about As-Is: In sOCEL discovering control flow or workpiece flow is archived by filtering respective classes. Discovered log-based process models providing an as-is process perspective on production lines as demonstrated in Figure 6.2 or Figure 6.4. In future work, discovered models, can visualize impacts or can be annotated with impacts caused by events [12]. Tracing individual production steps of a product, its material origins, and, i.e., an aggregation of related impacts of a specific production process as demonstrated in Figure 6.5 is feasible with sOCEL.

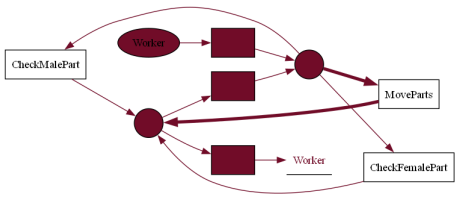

Figure 6.4: Discovered Object-Centric Petri Net - Only Worker Object

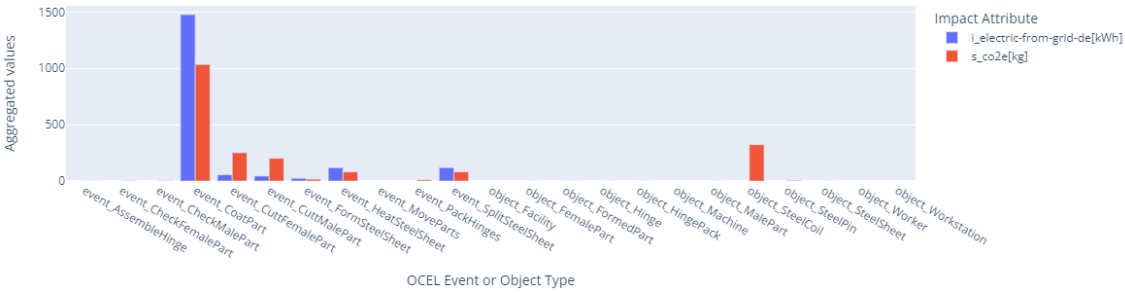

Figure 6.5: Accumulated Total of an Impact Score (CO_2e) and Electricity Impact Indicator (kWh): The high impact score and power consumption of event *CoatPart* are noteworthy.

Identify Levers to Reduce CO_2 on an Operational Process Level: In sOCEL, impact parameters and indicators can be included based on measures facilitated on production level in sOCEL. This provides button-up data insights, as indicators and scores that are logged on event or object level can be easily aggregated, which enables, i.e. impact hotspots identification, as shown for our example process in Figure 6.5. Combined with other PM techniques for analysis and visualization, sOCEL provides this operational process level insight. In addition, using attributes of location objects allows linkage and storage of top-down data (and time series data) like overall facility electricity consumption,

which allows comparison of electricity allocated in log and a real total. This gives potentially an additional measure for validity of insights generated by PM techniques.

Link Operational Efficiency and Sustainability: The object and attribute classes introduced in sOCEL allow the distinction between events, objects, and attributes that represent direct sustainability impacts and are evaluated or analyzed by specialized algorithms and those that are included to analyze other operational goals. SOCEL can serve as a centralized log format that supports analysis and inclusion of data for both traditional process improvement and (standardized) sustainability assessment. In Figure 6.6 we present a log-based perspective on our example manufacturing process, offering operational insights, i.e., on supposed idle times and pauses.

Figure 6.6: Ordered Distribution of Events by Timestamp for One Day: Patterns in recorded events are recognizable, i.e., long idle phases after event *MoveParts*.

In summary, it can be said that all issues can be addressed. However, sOCEL and OCPM themselves can better tackle not so much the need for technical expertise, but rather the potential for creating transparency about the current situation by sOCEL being able to include relevant sustainability data.

Chapter 7

Discussion

In this chapter, we discuss our methodology and results for integrating sustainability data into OCEL. Initially, we conducted an exploratory literature review, focusing on selected examples to gain an overview of assessments and LCAs conducted in the manufacturing domain. The results only pertain to metal workpieces, presumably because of the high-energy intensity of related operations and the potential for reducing their impact. Furthermore, the selection of examples, limits the validity of our results, but makes the amount of published work and LCAs in manufacturing manageable. Conducting a comprehensive literature review of the data required for specific manufacturing operations and qualitative interviews with LCA practitioners, may provide further robust insights.

In our work, we show that OCEL2.0 can store and provide data and sustainability metrics for various assessment strategies. Containing material and energy flows as well as additional process-related metrics, sOCEL can be utilized as central storage for LCI and sustainability monitoring data (including traditional business related KPIs). It thereby supports LCA and material flow analyses. Due to the potential of dynamic creation of event logs, sOCEL enables the integration of PM in continuous life cycle or sustainability assessment.

However, missing or wrong information and included information with high uncertainty influence the quality of insights of PM based on sOCEL. While we assume the existence of live sensor data or measures at production operation level, this is often not the case [4]. Therefore, the open specification of impact parameters allows heterogeneous ways to include information, i.e., referencing external time series data, providing workarounds. It is important to note that not all impacts can be assessed within PM or through automated means, and therefore may not be documented within sOCEL for mining purposes. To maintain the principle of a single source of truth, we recommend storing or at least linking these impacts as unspecified metadata to objects or location objects via attributes or impact parameters.

PM is a technology that works with log data. In sOCEL, parameters can link external data sources to provide a comprehensive data basis for assessment. The inclusion and quantification of impact indicators in sOCEL also requires external domain knowledge, both in manufacturing and sustainability assessment. Additionally, a preprocessing step is necessary, for example, to fill missing data gaps.

In this work, we focus on gate-to-gate process, an organization typically monitors. Assessments like LCA, however, must be done on the whole supply chain if they want to identify and reduce impacts. Otherwise, instead of reduction, a shift of impacts is possible that must be recognized early on [10]. Consequently, an interesting future research topic is to utilize OCEL2.0 and sOCEL for storing product and material flows on a higher abstraction layer (facilities and organizations, instead of single operations) mapping entire supply chains.

Last, we note, that the inclusion and extraction of SD into event logs represents a high manual effort, as there is a wide range of possible impact aspects to consider for a wide range of different manufacturing operations with different sustainability aspects that have to be considered. Development of supporting tools for utilizing impact parameters to calculate indicators as well as to add data after log creation is needed, increasing the applicability of sOCEL.

Chapter 8

Conclusion

In this thesis, we introduce sOCEL, a guideline and specification for the inclusion of sustainability data in object-centric event logs, focusing on the manufacturing process domain. The framework addresses the need for standardized integration of sustainability information into event logs for process mining.

We identify three classes of sustainability attributes, impact scores, impact indicator and impact parameter, required for comprehensive sustainability assessments using PM techniques. We present a detailed specification for integrating sustainability data as attributes into OCELs. This specification acts as a guideline for the integration of sustainability information and distinguishes our work from others that lack detailed instructions for integrating SD into OCELs.

To demonstrate the application and feasibility of the framework, a manufacturing process model is designed and simulated as a case study. We developed a Python tool interacting with CPN Tools to automatically integrate sustainability data using emission data from Climatiq API, thereby enhancing the authenticity of the impact scores inserted in the simulation model. To evaluate our approach, we generate a sample sOCEL log and discuss how relevant information is included and extractable. To encourage further research, the example sOCEL as XML file, the simulation model, and the Python tool are publicly available on GitHub[1].

8.1 Future Work

Future work results from our evaluation and discussion: In this work, we designed sOCEL with a strong focus on gate to gate manufacturing process. As sOCEL with the proposed concept of object, event, and attribute classes is designed extendable, it emphasizes adding classes depicting domain knowledge for assessment of further product life cycle phases, like logistics and distribution or the product use-phase. Additionally, it is necessary to study the application of federated process mining for SA. This enables the utilization of process mining across organizations, and enable the mapping of supply chains and assessment of their sustainability impacts.

[1] github.com/marcoheinisch/thesis-sOCEL

Based on the provided sOCEL specification and the generated example event log, further work is needed to develop specialized algorithms and tools working on sOCELs and that contributes to:

- Identification, mining and (manual) completion of missing or unquantified impact indicators (based on parameters) in sOCEL.
- Visualization and analysis of impact indicators and scores, already quantified and provided in sOCEL.
- Specification of referencing time series and machine data in sOCEL as well as allocating, i.e. consumption data, to sOCEL events.
- Challenge of representing and dealing with non-allocable impacts included in sOCEL.

Bibliography

[1] World Commission on Environment and Development. Report of the world commission on environment and development: Our common future, 1987.

[2] General Assembly United Nations. Transforming our world: The 2030 agenda for sustainable development. resolution a/res/70/1, 2015.

[3] Stefan Schaltegger and Maria Csutora. Carbon accounting for sustainability and management. status quo and challenges. *Journal of Cleaner Production*, 36:1–16, 2012. ISSN 09596526. doi: 10.1016/j.jclepro.2012.06.024.

[4] Lars Brehm, Jessica Slamka, and Andreas Nickmann. Process mining for carbon accounting: An analysis of requirements and potentials. In Jens Dibbern, Jens Förderer, Thomas Kude, Franz Rothlauf, and Kai Spohrer, editors, *Digitalization Across Organizational Levels*, Progress in IS, pages 209–244. Springer International Publishing, Cham, 2022. ISBN 978-3-031-06542-2. doi: 10.1007/978-3-031-06543-9_9.

[5] Steven J. Davis, Glen P. Peters, and Ken Caldeira. The supply chain of co2 emissions. *Proceedings of the National Academy of Sciences of the United States of America*, 108 (45):18554–18559, 2011. doi: 10.1073/pnas.1107409108.

[6] Lianhui Li, Ting Qu, Yang Liu, Ray Y. Zhong, Guanying Xu, Hongxia Sun, Yang Gao, Bingbing Lei, Chunlei Mao, Yanghua Pan, Fuwei Wang, and Cong Ma. Sustainability assessment of intelligent manufacturing supported by digital twin. *IEEE Access*, 8: 174988–175008, 2020. doi: 10.1109/ACCESS.2020.3026541.

[7] Wil van der Aalst. Process mining: Overview and opportunities. *ACM Transactions on Management Information Systems*, 3(2):1–17, 2012. ISSN 2158-656X. doi: 10. 1145/2229156.2229157.

[8] Nina Graves, István Koren, and Wil M.P. van der Aalst. Rethink your processes! a review of process mining for sustainability. In *2023 International Conference on ICT for Sustainability (ICT4S)*, pages 164–175. IEEE, 2023. ISBN 979-8-3503-1109-9. doi: 10.1109/ICT4S58814.2023.00025.

[9] Naren Sundaravaradan, Manish Marwah, Amip Shah, and Naren Ramakrishnan. Data mining approaches for life cycle assessment. In *Proceedings of the 2011 IEEE International Symposium on Sustainable Systems and Technology*, pages 1–6. IEEE, 2011. ISBN 978-1-61284-394-0. doi: 10.1109/ISSST.2011.5936863.

[10] Gregor Hagedorn, Peter Kalmus, Michael Mann, Sara Vicca, Joke van den Berge, Jean-Pascal van Ypersele, Dominique Bourg, Jan Rotmans, Roope Kaaronen, Stefan

Rahmstorf, Helga Kromp-Kolb, Gottfried Kirchengast, Reto Knutti, Sonia I. Seneviratne, Philippe Thalmann, Raven Cretney, Alison Green, Kevin Anderson, Martin Hedberg, Douglas Nilsson, Amita Kuttner, and Katharine Hayhoe. Nachhaltigkeitsfokussierte digitale wertstromanalyse: Konzept zum einsatz von process mining für die nachhaltigkeitsfokussierte wertstromanalyse. *Science (New York, N.Y.)*, 364(6436): 139–140, 2019. doi: 10.1126/science.aax3807.

[11] Wil M. P. van der Aalst. Twin transitions powered by event data: Using object-centric process mining to make processes digital and sustainable. In Luão, Robert Lorenz, Jan Martijn E. M. van der Werf, and Sebastiaan J. van Zelst, editors, *Joint Proceedings of the Workshop on Algorithms & Theories for the Analysis of Event Data and the International Workshop on Petri Nets for Twin Transition co-located with the 44th International Conference on Application and Theory of Petri Nets and Concurrency (Petri Nets 2023), Caparica, Portugal, June 25-30, 2023*, volume 3424 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2023.

[12] Christian Ortmeier, Nadja Henningsen, Adrian Langer, Alexander Reiswich, Alexander Karl, and Christoph Herrmann. Framework for the integration of process mining into life cycle assessment. *Procedia CIRP*, 98:163–168, 2021. ISSN 22128271. doi: 10.1016/j.procir.2021.01.024.

[13] Leah Tacke Genannt Unterberg, István Koren, and Wil M.P. van der Aalst. Maximizing reuse and interoperability in industry 4.0 with a minimal data exchange format for machine data. In *Modellierung 2024*, pages 103–118. Gesellschaft für Informatik e.V, Bonn, 2024. ISBN 978-3-88579-742-5. doi: 10.18420/modellierung2024_011.

[14] Ian C. Garretson, Mahesh Mani, Swee Leong, Kevin W. Lyons, and Karl R. Haapala. Terminology to support manufacturing process characterization and assessment for sustainable production. *Journal of Cleaner Production*, 139:986–1000, 2016. ISSN 09596526. doi: 10.1016/j.jclepro.2016.08.103.

[15] Manish Kumar and Monto Mani. Sustainability assessment in manufacturing for effectiveness: Challenges and opportunities. *Frontiers in Sustainability*, 3:837016, 2022. ISSN 2673-4524. doi: 10.3389/frsus.2022.837016.

[16] Andreas Fritsch, Johanna von Hammerstein, Clemens Schreiber, Stefanie Betz, and Andreas Oberweis. Pathways to greener pastures: Research opportunities to integrate life cycle assessment and sustainable business process management based on a systematic tertiary literature review. *Sustainability*, 14(18):11164, 2022. doi: 10.3390/su141811164.

[17] Iso 14040 international standard: Environmental management - life cycle assessment - principles and framework, 2006.

[18] Mark A. J. Huijbregts, Zoran J. N. Steinmann, Pieter M. F. Elshout, Gea Stam, Francesca Verones, Marisa Vieira, Michiel Zijp, Anne Hollander, and Rosalie van Zelm. Recipe2016: a harmonised life cycle impact assessment method at midpoint and endpoint level. *The International Journal of Life Cycle Assessment*, 22(2):138–147, 2017. ISSN 0948-3349. doi: 10.1007/s11367-016-1246-y.

[19] ILCD. Ilcd handbook - general guide on lca - provisons and action steps, 2010.

[20] Guide on life cycle inventory (lci) data generation for the environmental footprint, 2018.

[21] Raul Carlson, Anne-Marie Tillman, Bengt Steen, and Göran Löfgren. Lci data modelling and a database design. *International Journal of Life Cycle Assessment*, 3: 106–113, 1998. doi: 10.1007/BF02978500.

[22] Paul H. Brunner and Helmut Rechberger. *Practical handbook of material flow analysis*, volume 1 of *Advanced methods in resource and waste management*. CRC/Lewis, Boca Raton FL, 2004. ISBN 1566706041.

[23] Alessandro Berti, Istvan Koren, Jan Niklas Adams, Gyunam Park, Benedikt Knopp, Nina Graves, Majid Rafiei, Lukas Liß, Leah Tacke Genannt Unterberg, Yisong Zhang, Christopher Schwanen, Marco Pegoraro, and Wil M. P. van der Aalst. Ocel (object-centric event log) 2.0 specification, 2024.

[24] Wil M. P. van der Aalst. *Process Mining: Data Science in Action*. Springer Berlin Heidelberg and Imprint: Springer, Berlin Heidelberg, 2nd ed. 2016 edition, 2016. ISBN 9783662498514.

[25] Malarvizhi Kaniappan Chinnathai and Bugra Alkan. A digital life-cycle management framework for sustainable smart manufacturing in energy intensive industries. *Journal of Cleaner Production*, 419:138259, 2023. ISSN 09596526. doi: 10.1016/j.jclepro.2023.138259.

[26] Constantin Houy, Markus Reiter, Peter Fettke, and Peter Loos. Towards green bpm - sustainability and resource efficiency through business process management. In *Business Process Management Workshops*, 2010.

[27] Konstantin Hoesch-Klohe and Aditya Ghose. Environmentally aware business process improvement in the enterprise context. In San Murugesan and G. R. Gangadharan, editors, *Harnessing Green It*, pages 265–281. Wiley, 2012. ISBN 9781119970057. doi: 10.1002/9781118305393.ch13.

[28] Claudio Favi, Federico Campi, Michele Germani, and Marco Mandolini. A data framework for environmental assessment of metal arc welding processes and welded structures during the design phase. *The International Journal of Advanced Manufacturing Technology*, 105(1-4):967–993–967–993, 2019. ISSN 0268-3768. doi: 10.1007/s00170-019-04278-w.

[29] Megan F. Watkins, Mahesh Mani, Kevin W. Lyons, S. K. Gupta. Sustainability characterization for die casting process. *Proceedings Of The Asme International Design Engineering Technical Conferences And Computers And Information In Engineering Conference*, 2a, 2013.

[30] Stefano Guarino, Gennaro Salvatore Ponticelli, and Simone Venettacci. Environmental assessment of selective laser melting compared with laser cutting of 316l stainless steel: A case study for flat washers' production. *CIRP Journal of Manufacturing Science and Technology*, 31:525–538, 2020. ISSN 17555817. doi: 10.1016/j.cirpj.2020.08.004.

[31] Fuyuan Liu, Yanle Li, Shahid Ghafoor, Zinan Cheng, Fangyi Li, and Jianfeng Li. Sustainability assessment of incremental sheet forming: a review. *The International*

Journal of Advanced Manufacturing Technology, 119(3-4):1385–1405, 2022. ISSN 0268-3768. doi: 10.1007/s00170-021-08368-6.

[32] Giuseppe Ingarao, Paolo C. Priarone, Francesco Gagliardi, Rosa Di Lorenzo, and Luca Settineri. Subtractive versus mass conserving metal shaping technologies: an environmental impact comparison. *Journal of Cleaner Production*, 87:862–873, 2015. ISSN 09596526. doi: 10.1016/j.jclepro.2014.10.018.

[33] Junli Shi, Junyu Hu, Mingyang Ma, and Huaizhi Wang. An environmental impact analysis method of machine-tool cutting units based on lca. *Journal of Engineering, Design and Technology*, 19(5):1192–1206, 2021. ISSN 1726-0531. doi: 10.1108/JEDT-06-2020-0247.

[34] Liming Wang, Xueju Ran, Yanle Li, Fangyi Li, Jing Liu, Jiyu Du, Xingyi Zhang, and Xiaoxia Qi. Energy consumption model of plasma spraying based on unit process life cycle inventory. *Journal of Materials Research and Technology*, 9(6):15324–15334, 2020. ISSN 22387854. doi: 10.1016/j.jmrt.2020.11.007.

[35] Michael D. Eastwood and Karl R. Haapala. An induction hardening process model to assist sustainability assessment of a steel bevel gear. *The International Journal of Advanced Manufacturing Technology*, 80(5-8):1113–1125, 2015. ISSN 0268-3768. doi: 10.1007/s00170-015-7053-y.

[36] Wiebke Hagedorn, Alexander Gramlich, Kathrin Greiff, and Ulrich Krupp. Alloy and process design of forging steels for better environmental performance. *Sustainable Materials And Technologies*, 34:e00509, 2022. doi: 10.1016/j.susmat.2022.e00509.

[37] Daeyoung Kong, Seungchoun Choi, Yusuke Yasui, Sushrut Pavanaskar, David Dornfeld, and Paul Wright. Software-based tool path evaluation for environmental sustainability. *Journal of Manufacturing Systems*, 30(4):241–247, 2011. ISSN 02786125. doi: 10.1016/j.jmsy.2011.08.005.

[38] Huimin Liu, Qiqiang Li, Guanguan Li, and Ran Ding. Life cycle assessment of environmental impact of steelmaking process. *Complexity*, 2020:1–9, 2020. ISSN 1099-0526. doi: 10.1155/2020/8863941.

[39] Shamraiz Ahmad, Kuan Yew Wong, and Srithar Rajoo. Sustainability indicators for manufacturing sectors. *Journal of Manufacturing Technology Management*, 30(2): 312–334, 2019. ISSN 1741-038X. doi: 10.1108/JMTM-03-2018-0091.

[40] Mohannad Shuaib, Daniel Seevers, Xiangxue Zhang, Fazleena Badurdeen, Keith E. Rouch, I.S. Jawahir. Product sustainability index (prodsi): A metrics-based framework to evaluate the total life cycle sustainability of manufactured products. *Journal of Industrial Ecology*, 18(4):491–507, 2014. doi: 10.1111/jiec.12179.

[41] Che B. Joung, John Carrell, Prabir Sarkar, and Shaw C. Feng. Categorization of indicators for sustainable manufacturing. *Ecological Indicators*, 24:148–157, 2013. ISSN 1872-7034. doi: 10.1016/j.ecolind.2012.05.030.

[42] André Luís Helleno, Aroldo José Isaias de Moraes, and Alexandre Tadeu Simon. Integrating sustainability indicators and lean manufacturing to assess manufacturing processes: Application case studies in brazilian industry. *Journal of Cleaner Production*, 153:405–416, 2017. ISSN 09596526. doi: 10.1016/j.jclepro.2016.12.072.

[43] Lars-Gunnar Lindfors, Tomas Ekvall, Elin Eriksson, Kristian Jelse, Tomas Rydberg. The ilcd handbook in a nutshell: A brief analysis of the ilcd handbook and the draft guidance on product environmental footprint. *IVL Report B2020*, 2020.

[44] Jonnro Erasmus, Irene Vanderfeesten, Konstantinos Traganos, and Paul Grefen. Using business process models for the specification of manufacturing operations. *Computers in Industry*, 123:103297, 2020. ISSN 01663615. doi: 10.1016/j.compind.2020.103297.

[45] Benedikt Knopp and Nina Graves. Container logistics object-centric event log. oct 2023. doi: 10.5281/zenodo.8428084.

[46] H.M.H. Hegge and J.C. Wortmann. Generic bill-of-material: a new product model. *International Journal of Production Economics*, 23:117–128, 1991. doi: 10.1016/0925-5273(91)90055-X.

Additionally, DeepL Write and ChatGPT are used to enhance grammar and language.

Acknowledgments

I would like to thank my supervisor, Nina Graves, for her dedicated support and insights into the current research on process mining and sustainability. I would also like to thank my family for their support during the last months, as well as my friends and colleagues, particularly Lennart, Franziska, and Maxim for their valuable feedback on this thesis!