Can We Leverage Process Data from ERP Systems for Business Process Sustainability Analyses?

Dominik Schäfer[1,3], Finn Klessascheck[2,4], Timotheus Kampik[3,5], and Luise Pufahl[2,4]

[1] Technical University of Munich, School of CIT, Munich, Germany
dominik.maximilian.schaefer@tum.de
[2] Technical University of Munich, School of CIT, Heilbronn, Germany
firstname.lastname@tum.de
[3] SAP Signavio, Berlin, Germany
timotheus.kampik@sap.com
[4] Weizenbaum Institute, Berlin, Germany
[5] Umeå University, Umeå, Sweden

Abstract. Sustainability is an increasingly important issue, which organizations need to take into account when assessing and improving their business processes. Doing so can contribute to enhancing an organisation's overall sustainability. *Green Business Process Management* is a line of research concerned with supporting organisations to integrate a sustainability perspective into their processes. However, existing approaches that assess sustainability on activity and process levels using, for instance, *Life-Cycle Assessment* (LCA) are often time-consuming and complex. Therefore, this work explores whether *Key Ecological Indicators* (KEIs) used to assess the sustainability of a business process can be calculated using data already available within an organisation. Following a case study methodology, we analyse nine real-world datasets extracted from a business process analysis system of a large enterprise software vendor. Results indicate that current data availability is insufficient for exact assessments. To address this issue, we introduce a high-level conceptual model and provide recommendations for action based on the observations of the case study.

Key words: Sustainability, Green Business Process Management, Key Ecological Indicators, Process Data Analysis

1 Introduction

Five of the top ten global risks over a ten-year period can be assigned to environmental risks, highlighting the urgency of sustainability [1]. In addition, customers and society increase cultural pressures to drive sustainability [2]. Given the critical role of sustainability in today's world, organisations are increasingly incentivised to identify their environmental impact [3]. Organisations can be viewed as collections of business processes designed to generate business value, and hence can improve their environmental footprint by enhancing these processes [4].

One promising approach for achieving this is *Green Business Process Management* [3]: it extends traditional Business Process Management (BPM), which focuses on concepts, methods, and techniques to design, execute and analyse business processes [5], by integrating a sustainability dimension and establishing sustainability as a business objective [6]. Central to this are Key Ecological Indicators (KEIs), e.g. Energy Consumption, which represent an extension of traditional Key Performance Indicators (KPIs) [7] and allow a quantitative assessment of business process sustainability [6]. KEIs are essential as, if reported and used correctly, they influence strategic decisions inside an organisation and thus can help to improve environmental performance [3].

Green BPM approaches include the introduction of *new notations*, such as emission annotations [8,9] to express sustainability aspects in process models, and the use of *Activity-based Costing (ABC)* approaches [8,9] or the integration with *Life Cycle Assessment (LCA)* [10], to quantify the environmental footprint of business processes. Notably, for many approaches which quantify the environmental footprint of business processes, the availability of data remains a critical issue: the Greenhouse Gas (GHG) protocol highlights that methodological rigour does not compensate for poor data quality [11]—indeed, related work such as [8,9] explicitly assume that key data such as carbon dioxide (CO_2) emissions of activity instances or power consumption data of machines involved in activity executions are known beforehand. However, it is unclear whether sufficient data of appropriate quality is actually available in practice.

To address this challenge, LCA databases have been developed, providing a wide range of data [12]. Notably, conducting an LCA study is usually complex and costly due to extensive data collection, stakeholder involvement, necessary reviews and updates [13]. These complexities underscore a significant research gap, particularly in situations where conducting a full LCA is not feasible, which is addressed in this work by answering the following Research Questions (RQs):

RQ1: To what extent is the data needed to calculate KEIs of Business Processes already present in large Business Process Analysis Systems?
RQ2: How could commonly found KEIs be approximated using real-world data?

To address these RQs, this work conducts a case study [14] in cooperation with one of the world's largest enterprise software vendors. The study performs a meta-data analysis of real-world process data to evaluate the feasibility of calculating common KEIs from this data. The process data is extracted from a Business Process Analysis tool, which sources its data from an Enterprise Resource Planning (ERP) system.

The paper is structured as follows: first, Section 2 introduces the theoretical background and related work, after which we identify common KEIs used in Green BPM in Section 3. Subsequently, in Section 4, the methodology followed in the case study is presented and in Section 5, the results are described. Section 6 interprets these results and introduces a conceptual model for integrating KEIs in Green BPM while taking into account a lack of suitable data. Finally, we conclude the work and present future work in Section 7.

2 Background and Related Work

In the following, the research discipline *Green BPM* is introduced and common approaches to include a sustainability perspective into traditional BPM are presented.

2.1 Green BPM

Green Business Process Management is an emerging discipline that has increased in relevance over the years [4,7]. It extends traditional BPM by integrating a sustainability dimension [6] and putting this dimension into focus [3,6,13]. This work analyses business process data with respect to KEIs and evaluates whether commonly used KEIs are calculable with the given data. KEIs, also referred to as *Environmental Performance Indicators* [3,6], are indicators that organisations use to assess their environmental performance and to quantify their environmental impact [3].

2.2 Approaches for Incorporating Sustainability

Activity-based Costing. ABC aims at increasing the accuracy of product cost estimates by converting indirect costs in the traditional accounting system into direct costs [15]. This is achieved by allocating resource costs such as wages to cost objects based on their activity consumption [15]. It has found adoption in Green BPM, with multiple contributions leveraging ABC-based approaches to incorporate a sustainability perspective into business processes (e.g., [8–10]).

Life Cycle Assessment. LCA is a tool used to evaluate the environmental impact and resource utilisation across the entire life cycle of a product or service [12]. Fundamental to LCA is the *cradle-to-grave* perspective, meaning that it considers the entire life cycle of the process or product when assessing the environmental footprint [12]. In practice, LCA is widely used and specified by two international standards [12]. The cradle-to-grave perspective is also subject to some critique, as it is challenging to obtain all necessary data: LCA is extremely data-intensive and the reliability of its results highly depends on the data used [12]. To provide this data, several LCA databases have been developed, which provide environmental data on numerous products and essential services that are required in many LCAs [12].

Approaches for Green BPM. A tertiary literature review [13] clusters Green BPM approaches into the six capability areas of *modelling*, *deployment*, *optimisation*, *management*, *culture* and *structure*. For this work, especially the *modelling*, *deployment* and *management* areas are considered relevant. *Modelling* includes the three approaches [7] extending notations (e.g., [8,9]), adding notations (e.g., [16,17]), and adding patterns (e.g., [18]). *Deployment* and *management* both deal with KEIs, while the main difference is that the deployment area deals with the *application* of KEIs and the management area deals with the *definition* of them [13]. Relevant works in the *deployment* capability area include the measuring and controlling of emissions (e.g., [19]), whereas the *management* area includes the extension of the traditional business process lifecycle into the green business process lifecycle (e.g., [4]).

However, research so far does not explicitly address to which extent the data necessary for incorporating sustainability into business processes by calculating KEIs is available in average companies *without* relying on LCAs. Usually, the data needed to incorporate the sustainability dimensions is considered as given or is elicited manually (see, e.g., [8,9]). Even though using LCA databases to obtain the required data (see e.g., [10]) is feasible, a LCA study is still complex and costly, requiring extensive data collection, stakeholder involvement, and continuous reviews and updates [13]. To the best of our knowledge, there are no relevant studies focusing on whether KEIs, known from literature and useable for assessing the sustainability of a business process, can be calculated using data available within companies. Thus, this work addresses this research gap and focuses on KEIs. Focusing on KEIs has the advantages that they *i)* enable the assessment of whether a business process improves its sustainability over time [6]; *ii)* can serve as powerful tools that can influence the strategic decisions of organizations [3]; iii) align with the call made in [3] for research aimed at identifying the *right KEIs*.

3 Key Ecological Indicators in Green BPM

To analyse the availability of business process data needed for calculating KEIs, we first identify which KEIs are commonly used in Green BPM by drawing on existing Systematic Literature Reviews (SLRs).

KEI Identification. In a tertiary literature review, Fritsch et al. [13] observe that only three SLRs they analysed, namely [3, 6, 20], explicitly provide an overview of KEIs. However, as [3, 20] display the same sustainability aspects only with a slightly different number of corresponding primary papers [13], only [3] is included here for further analysis. In addition, examining the references of the aforementioned SLRs led to the inclusion of [21]. Consequently, we identify KEIs commonly used in Green BPM based on the SLRs [3, 6, 21].

Table 1 presents the top three KEIs from each of these SLRs and their reported occurrences. To identify similarities across the SLRs, we assigned codes to unique KEIs. While the authors of [3] and [6] explicitly identify the top three KEIs and provide the percentage of their occurrence in the reviewed papers, Gräuler and Teuteberg [21] only display an absolute number of references per identified sustainability metric. Thus, in the latter case, the number of references per KEI was counted and divided by the total number of relevant papers identified by [21] and the top three KEIs with the most references are displayed. In total, [3] analysed 49 sources, [6] 56 sources, and [21] 31 sources.

Roohy Gohar & Indulska [3]			**Hernández González et al. [6]**			**Graeuler & Teuteberg [21]**		
Code	**Top KEIs**	**[%] ↓**	**Code**	**Top KEIs**	**[%] ↓**	**Code**	**Top KEIs**	**[%] ↓**
EC	Energy Consumption	31%	EM	Emissions	71%	EC	Energy Efficiency/ Consumption in e.g. kWh/unit	29%
C2	CO_2 footprint	22%	EC	Energy consumption	61%	GH	Emissions of greenhouse gases, ozone-depleting substances or other emissions in e.g. CO_2-equivalents	26%
GH	other GHG emissions	20%	UM	Use of materials	29%	WG	Waste Generation in e.g. kg/unit	13%

Table 1. The top three KEIs identified in each SLR [3, 6, 21]

By examining how often each KEI occurred in Table 1, we aggregate the KEIs of the three Green BPM SLRs and arrive at a final set of common KEIs. *Energy Consumption* (*code: EC*) was found in all three SLRs and is a final KEI. GHG Emissions, Emissions and CO_2 (*codes: C2, GH & EM*) all refer to air quality and are summarised under the final KEI *Emission*. Material Use & Waste Generation (*codes: UM & WG*) both refer to physical goods and are related to each other as material use can also lead to waste, and thus are summarised under the final KEI *Material Use & Waste Generation*. The three *final KEIs* are only referred to as *KEIs* in the remainder of this paper.

KEI Calculation. For being able to assess whether the three KEIs identified above can be calculated using data from the case study, we first present approaches commonly used in the literature for calculating them.

Energy Consumption. To calculate the energy consumption of an activity, [9] examines energy-related parameters such as the wattage a machine utilises to perform a specific action expressed in kW, the total apparent power (KVA) or the ratio between both (power factor). Similarly, the GHG Protocol [11] suggests methods to gather energy consumption data, such as through meter readings or invoices. Once this data is obtained, energy data can further be used to calculate the resulting CO_2 emissions as demonstrated in [8]. They used data for energy consumption combined with a calculation tool provided by the GHG protocol [11] to calculate the resulting CO_2 emissions.

Emissions. When discussing emissions, this study understands *emissions* as GHG as used in the GHG-Protocol [11]. It is a broad definition and includes both CO_2 *emissions* and *other emissions* as previously found in the literature (see Table 1). In general, to calculate the GHG emissions, the GHG Protocol [11] provides over 14 different calculation tools. For example, to calculate direct and indirect CO_2 emissions from the stationary equipment or to calculate the CO_2 emissions from the calcination process in cement production. More abstractly, [9] calculates GHG emissions of an activity as *activity data * emission factor*. Emission factors in this context are ratios that indicate how much, e.g. GHG emissions, per base unit or activity consumption, e.g. hour of use, are emitted [11].

Material Use & Waste Generation. To calculate material use & waste generation, again, it is essential to have access to data tracking the actual use or waste. In [8] the authors calculate the emission of paper consumption by using the formula *weight of paper * emission factors for manufacture of paper*. Thus, in order to calculate the environmental footprint, it is again necessary to determine an emission factor.

4 Methodology

To address the research questions we outline in Section 1, we conduct a *case study*, following the six-step methodology proposed by Recker [14].

For *planning* the study, we aimed to identify the extent to which data needed for calculating KEIs of business process is already present in real-world business process analysis systems and investigate how far common KEIs can be calculated using this data. With this, we justify the use of a case study methodology since case studies are well suited if the goal is to intensively study a contemporary case in its natural setting to understand its complexity [14], which applies to this work.

For the *design step*, we chose an *exploratory* research objective, a *positivist* epistemology, and a *holistic single case* design [14]. As a data source, we selected the pseudonymous *System Alpha*, which is a business process analysis system used worldwide by customers of the enterprise software vendor where the case study is conducted. It provides fast insights into the performance of over 100 defined business processes.

The data required to display these business processes is exclusively collected from the underlying ERP system by using a data collector, which regularly gathers predefined process data from the different ERP tables.

As to the *preparation* stage, we first performed a preliminary inspection of the entire data inside System Alpha. Out of 104 available processes, nine were chosen with the support of internal experts. These nine processes focus on the domains of *transportation* and *logistics*, which significantly impact a company's environmental footprint [22], and are, according to internal experts of the case study's company, among the most commonly used transportation processes by customers.

The *collecting and analysing* stages were conducted iteratively, with each step informing the next. Initially, nine datasets in the form of CSV files were extracted from System Alpha for nine selected processes. To create the final meta-dataset, only the column names were retained from each process dataset, as these indicate which types of data are stored in each column, representing the data the customer has access to in their system. After data cleaning, all column headers were consolidated into a single meta-dataset, with each original column header transformed into an entry in the new meta-data analysis table. To enhance the interpretability of each original column, demo values were added by accessing an internal demo version of System Alpha containing example values for each column. Afterwards, the frequency of each column was counted and analysed.

After this first *collecting-analysing* iteration, a second iteration was done. First, a company-internal version of the Large Language Model (LLM) model of OpenAI, *GPT-4-Turbo*, was used to describe each column header in exactly one sentence and further enhance their interpretability. Subsequently, a categorisation of the data regarding the *Informative Value* was done. For this, a manual assessment was conducted identifying whether the information stored in an individual column contains useful information for calculating KEIs *by itself* (*e.g., the field 'delivery quantity (lfimg)' with the value '3.0'*) or whether the information must be *combined* with data from other systems—such as ERP systems—to become useful (*e.g., the field 'Company Code (BUKRS)' with the value 'F010'*). Additionally, a second categorisation, called *data-type* categorisation, was done by following a qualitative, inductive coding approach to identify the different data types existing in the meta-dataset, such as *Dates & Timestamps*. Using this categorisation, relevant units, such as *kWh* for *energy*, were searched in the analysis in Section 5.2. Finally, the meta-dataset was searched for relevant synonyms potentially containing information for calculating KEIs. These synonyms were generated using two online thesauruses,[1] and the meta-dataset was searched for these synonyms and their combinations. The detailed search protocol can be found online in a GitHub repository.[2]

After finishing the preceding steps, the meta-dataset was complete, with an excerpt provided in Table 2. It consists of six columns: *Column Name* stores the names of the original columns extracted from System Alpha, and is the primary key of this table. [#] counts the frequency with which each column occurs in the original nine datasets. *Demo Values* provides three, ideally non-null, demo values for each

[1] See `https://dictionary.com`, `https://www.collinsdictionary.com/dictionary/english-thesaurus` [Accessed: 20/08/2024]

[2] `https://github.com/dominik-maximilian-schaefer/icpm2024` [Accessed: 22/08/2024]

entry. *LLM-Description* contains one sentence describing each entry. *Informative Value* holds the assigned categorisation value for this entry. *Data Type* contains the assigned categorisation value for this entry.

Finally, the *sharing* stage, the sixth and final step proposed by Recker [14], is represented by the following sections displaying the case study's results and findings.

Column Name	# ↓	Demo Values	LLM Description	Informative Value	Data Type
Currency (WAERS)	3	'EUR', 'RUB', 'USD'	Represents the currency used in the financial transactions, like USD or EUR.	direct	Amounts, Values & Units

Table 2. Excerpt of the final meta-dataset

5 Results

This section describes the result of our case study by first describing the dataset we created and subsequently analysing it regarding the question of which common KEIs can be calculated based on it.

5.1 Final Meta-Dataset

The final meta-dataset with 422 entries is provided in the GitHub repository linked above. When examining the *frequency* (#) of each column described, it is notable that approximately 32% of all column names appear in more than one of the original nine datasets. The majority of columns (68%) appear only once, while about 22% occur twice, 5% are found in three datasets, and the remaining 5% occur in four or more datasets.

The results of the *informative value categorisation* show that overall, 44% of all entries are considered to have direct value, while 56% are considered to have indirect value. This means that 56% of all 422 entries are considered to be not directly interpretable, and additional data sources, such as an ERP system or an LCA database, are required for deriving meaningful insights. Conversely, 44% of all entries are considered to have direct value, which means that the information is inherently interpretable.

The *data type categorisation* resulted in the following values: *Unique Identifiers (UIDs) (30%)*, *Status & Indicators (28%)*, *Dates & Timestamps (25%)*, *Amounts, Values & Units (16%)* and *Ambiguous (1%)*. The significant proportion of information categorised under UIDs suggests that much of the data may not be directly useful for KEI calculation, as it cannot be further interpreted without additional context. This categorisation is mostly used to filter for relevant units in the next step.

5.2 Key Ecological Indicators

After describing the final meta-dataset, we now investigate it regarding data needed to calculate the KEIs of *Energy Consumption*, *Emissions* and *Material Use & Waste Generation*. The analysis is structured in two parts: First, based on the data type categorisation; second, based on the synonym analysis.

Analysis Based on Data Type. To identify energy-related parameters like *kWh* or other relevant units and quantities used for calculating the three KEIs (see Section 3), the meta-dataset was filtered according to the dimension *Amounts, Values & Units*. Of the 68 entries identified, 55 pertained to quantities or numerical data, while 13 related to units. However, none of these units were relevant to the KEIs, as they mainly referred to currencies or basic measurement units like *kilograms* or *liters*. The 55 entries on quantities were mainly pertinent to the KEI *Material Use & Waste Generation*, with fields like *withdrawal quantity (enmng)*, *goods receipt quantity (gr_amount)*, and *delivery quantity (lfimg)* providing insights into material usage. Fields like *confirmed scrap (iasmg)* and *unplanned goods issue (unplanned_gi)* were significant for understanding waste generation.

Analysis Based on Synonyms. The second approach for identifying relevant data for calculating the three KEIs involved using synonyms, as detailed in Section 4. In the GitHub repository linked above, the two tables depicting the results from the synonym search are provided. The first table depicts the synonyms (e.g. *power* and *electricity*) found for each single term (e.g. *Energy* and *Consumption*) of which the single KEIs (*e.g. Energy Consumption*) are composed of, while the second table depicts the synonyms that were found for each compound term of KEIs (e.g. *Energy Consumption*). Further, the final search protocol is provided there, outlining exactly which synonym sets were searched within the meta-dataset, with the search conducted using substring matching across the columns *Column Name*, *Demo Values*, and *LLM Description*. An excerpt of the search protocol is shown in Table 3. The first column is a continuous numbering, the second column indicates the corresponding KEI, the third column depicts which synonym set was used for the search (e.g. synonym set S1 contains all synonyms for energy; if two or more codes are combined with x, the cartesian product of the synonym sets was built, and subsequently, a search using all the resulting elements was done), the fourth column indicates an example word contained in the synonyms set, the last column contains the number of results found in the meta-dataset when searching for all synonyms under the respective code.

Nr.	KEI	Code	Example	[#]
1	Energy Consumption	S1	"Energy"	29
7	Emissions	S5	"Emission"	30
21		S9	"Material"	142
23	Material Use &	S11	"Waste"	4
25	Waste Generation	S9 x S10	"Material" x "Use"	2
27		S20	"Material Use"	1

Table 3. Excerpt of the search protocol

Analysing the results shows that for the KEI *Energy Consumption*, S1 found 29 unique column names. Examining all these matches shows that no column actually refers to energy consumption. Indeed, only matches for *potential* and *service* (as synonyms of Energy) were found. For the KEI *Emission*, only S5 has more than zero results. Examining these results, only matches for *issue, release, ejection and venting* were found, no match providing relevant information to calculate the KEI *emissions*.

For the KEI *Material Use & Waste Generation*, *S9, S11, S9 x S10* and *S20* are relevant. The matches support the findings of Section 5.2, where a substantial number of columns were found by analysing *quantities*. All fields found there are also found examining the matches here. Thus, the preliminary conclusion can be drawn that data needed to calculate the KEI of *Material Use & Waste Generation* is available in the meta-dataset.

6 Discussion

This section discusses our results in the light of existing research, provides recommendations for action based on the findings of this case study and highlights limitations.

6.1 KEIs in Literature vs. Practice

Regardless of which KEI is considered, the literature typically calculates them by first gathering relevant *activity data* for each KEI, and then applying an *emission factor* to calculate the environmental footprint (see Section 3). The following discussion is structured according to these two steps:

Activity Data. Gathering *activity data* is the basis for the calculation of the KEIs and *must* be incorporated into the business process data so that it can effectively be used [23]. Primary activity data, particularly in the context of environmental footprints of products, is usually captured through equipment control sensors that measure i.a. flow rates and temperatures; however, this data is often *not* collected within the EPR system due to being perceived as having little value [24]. Examining the *activity data* available for *Energy Consumption* and *Emissions*, we find that the analysed meta-dataset does not provide useful data. Neither the analysis using the data type categorisation nor the analysis based on synonyms found any relevant fields that contain potential data that indicate one of these KEIs. For *Material Use & Waste Generation*, a substantial amount of data is found in the meta-dataset.

Emission Factors. Examining the *emission factors* which are used in literature when quantifying the environmental footprint of KEIs, the lack of data becomes even more salient: Data representing concrete emission factors is not found for any of the three KEIs. This possibly explains the success of LCA databases, as they provide the necessary emission factors [12], enabling such calculations to be performed (e.g., [10]).

From this case study, we can conclude that the availability of data required to calculate KEIs in large business process analysis systems is limited. Out of the three KEIs, *activity data* was only found for *Material Use & Waste management*, while data useful for *emission factors* was not found.

6.2 Recommendations for Action

Based on our findings, we propose three recommendations. First, use *indirect approximation* to quickly evaluate business process sustainability with *available* data, avoiding complex, time-consuming methods such as LCA. Second, *refine data collectors* to use more relevant data that is *available* in the ERP systems but not used in business process analysis systems. Third, *leverage expert knowledge* to improve the understanding of complex and partially ambiguous ERP data. These recommendations also address RQ2 and are explained in the following.

Indirect Approximation of KEIs. Our analysis indicates that data for exact sustainability analysis is not necessarily available in ERP systems. To mitigate this issue, it may be feasible to devise heuristics-based approaches that focus on inferring actions for sustainability improvements, given the data that is available. Therefore, we outline a conceptual model in Figure 1 that might assist in improving the sustainability of business processes by using only available data.

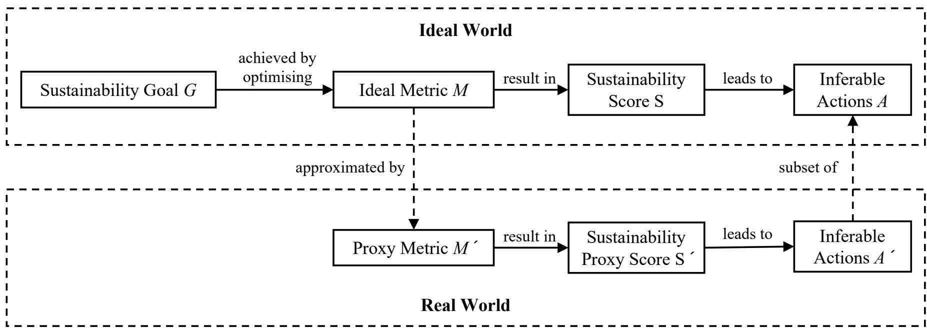

Fig. 1. Conceptual model for heuristics-based sustainability improvement in Green BPM.

The proposed solution involves setting a sustainability goal G, such as improving the environmental sustainability of a specific process. In an *ideal world*, this would be achieved by optimising *ideal metrics*, like the CO_2 emissions from transporting goods. These metrics form a *sustainability score* S which guides *inferable actions*. However, due to data limitations in the *real world*, *ideal metrics* often cannot be calculated. Instead, *proxy metrics* are used, such as the mode of transportation, to approximate the CO_2 footprint. By inferring the mode of transportation, such as inferring that a plane is used for transportation between Singapore and Munich based on a travel time of 20 hours and the distance, we can make informed inferences about the environmental impact of the process, even without detailed CO2 data or data specifying which transportation mode was used.

Refine Data Collectors. Enhancing the *data collectors*, which regularly collect data for systems such as System Alpha, is crucial: primary data central to sustainability assessments often does not penetrate the ERP system due to its historically low perceived value [24]. Additionally, about 50% of current entries offer only indirect informative value (see Section 5). By refining the data collector, we can increase the data's utility for sustainability assessments. For instance, extracting specific material details (e.g., type, weight) from the ERP system could enable rules that correlate larger or heavier items with a higher environmental footprint. Similarly, understanding delivery priorities could inform sustainability decisions, such as recognising that higher priority deliveries may have a greater environmental impact.

Leverage Expert Knowledge. Finally, evaluating the sustainability of business processes is complex and requires in-depth understanding of the data involved. For example, consultation with process experts revealed that depending on the process, 'scrap' can either be waste or be reintegrated into production, meaning it is no waste at all. This underscores the need for careful interpretation of process data by leveraging expert knowledge. When starting a new project, especially in specific sectors like transportation, it is essential to first understand the processes in detail, often through expert consultations.

6.3 Threats to Validity

This study is not without limitations: First, the data stored within each column is demo-data. However, as this case study analysed the column names and not the data within them, these demo-values are of secondary value and were only used to increase interpretability of each column. Second, the data analysis was done primarily by the first author. To mitigate this, the methodology is clearly documented, with additional materials provided in the GitHub repository linked above. Third, an LLM was used to enhance the data model with column descriptions. While LLMs can produce inaccurate information or hallucinate, this work mitigates that risk by using the LLM output as just one of five input sources, thoroughly reviewing it and cross-checking it against known attributes.

7 Conclusion

Our findings indicate that data for sound and complete sustainability analyses is not necessarily readily accessible in business process analysis systems, which extract their data from ERP systems. From a practice-oriented perspective, this limitation is crucial to acknowledge, as it highlights the importance of continued initiatives to increase sustainability data maturity and suggests that the data quality of existing sustainability reporting may be problematic. To still support organisations in assessing the sustainability of their business processes, we introduce a high-level conceptual model using only available data, which may serve as a starting point. Future work can: *i)* further expand research into data availability and quality for sustainability analysis, for example by analysing a variety of source systems; *ii)* devise and evaluate frameworks and metamodels for the extraction and analysis of sustainability data from ERP systems; *iii)* further refine and evaluate heuristics-based approaches for sustainability analyses.

References

1. World Economic Forum: The global risks report 2024
2. Seidel, S., Recker, J., vom Brocke, J.: Sensemaking and sustainable practicing: Functional affordances of information systems in green transformations. MIS Quarterly **37** (2013) 1275–1299

3. Roohy Gohar, S., Indulska, M.: Environmental sustainability through green business process management. Australasian Journal of Information Systems **24** (2020)
4. Maciel, J.: The Core Capabilities of Green Business Process Management – A Literature Review. (2017)
5. Weske, M.: Business Process Management: Concepts, Languages, Architectures. 3rd edn. Springer, Berlin, Heidelberg (2019)
6. Hernández González, A., Calero, C., Pérez Parra, D., Mancebo, J.: Approaching green bpm characterisation. Journal of Software: Evolution and Process **31**(2) (2019) e2145
7. Couckuyt, D., van Looy, A.: A systematic review of green business process management. Business Process Management Journal **26**(2) (2020) 421–446
8. Recker, J., Rosemann, M., Hjalmarsson, A., Lind, M.: Modeling and analyzing the carbon footprint of business processes. In vom Brocke, J., Seidel, S., Recker, J., eds.: Green Business Process Management. Springer, Berlin, Heidelberg (2012) 93–109
9. Wesumperuma, A., Ginige, A., Ginige, J., Hol, A.: Green activity based management (abm) for organisations. ACIS 2013 Proceedings (2013)
10. Klessascheck, F., Weber, I., Pufahl, L.: Sopa: A framework for sustainability-oriented process analysis and re-design in business process management
11. World Resources Institute and World Business Council for Sustainable Development: The greenhouse gas protocol: A corporate accounting and reporting standard. revised edition.
12. Finnveden, G., Hauschild, M.Z., Ekvall, T., Guinée, J., Heijungs, R., Hellweg, S., Koehler, A., Pennington, D., Suh, S.: Recent developments in life cycle assessment. Journal of environmental management **91**(1) (2009) 1–21
13. Fritsch, A., von Hammerstein, J., Schreiber, C., Betz, S., Oberweis, A.: 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) (2022)
14. Recker, J.: Scientific research in information systems: A beginner's guide. 2nd edn. Progress in IS. Springer, Cham (2021)
15. Tsai, W.H.: Activity-based costing model for joint products. Computers & Industrial Engineering **31**(3-4) (1996) 725–729
16. Houy, C., Reiter, M., Fettke, P., Loos, P., Hoesch-Klohe, K., Ghose, A.: Advancing business process technology for humanity: Opportunities and challenges of green bpm for sustainable business activities. In vom Brocke, J., Seidel, S., Recker, J., eds.: Green Business Process Management. Springer Berlin Heidelberg, Berlin, Heidelberg (2012) 75–92
17. K. Hoesch-Klohe, A. Ghose, Lam-Son Le: Towards green business process management. In: 2010 IEEE International Conference on Services Computing. (2010) 386–393
18. A. Nowak, F. Leymann: Green business process patterns – part ii (short paper). In: 2013 IEEE 6th International Conference on Service-Oriented Computing and Applications. (2013) 168–173
19. Ghose, A., Hoesch-Klohe, K., Hinsche, L., Le, L.S.: Green business process management: A research agenda. Australasian Journal of Information Systems **16**(2) (2010)
20. Roohy Gohar, S., Indulska, M.: Business Process Management: Saving the planet? (2015)
21. Gräuler, M., Teuteberg, F.: Experimental evaluation of a process benchmarking tool in a green business process management context. Wirtschaftsinformatik 2013 Proceedings (2013) 1083–1097
22. Van Woensel, T., Creten, R., Vandaele, N.: Managing the environmental externalities of traffic logistics: The issue of emissions. Prod. Oper. Manag. **10**(2) (2001) 207–223
23. Graves, N., Koren, I., van der Aalst, W.: Rethink your processes! a review of process mining for sustainability (postprint). (2023)
24. de Soete, W.: Towards a multidisciplinary approach on creating value: Sustainability through the supply chain and erp systems. Systems **4**(1) (2016)