Schmitt, Thomas; Bejarano Rodriguez, Ronal; Assuad, Carla

Challenges and opportunities of automated data pipelines for environmental sustainability applications in industrial manufacturing

Published in:
Procedia CIRP

DOI:
10.1016/j.procir.2024.01.089

Published: 01/01/2024

Document Version
Publisher's PDF, also known as Version of record

Published under the following license:
CC BY-NC-ND

Please cite the original version:
Challenges and opportunities of automated data pipelines for environmental sustainability applications in industrial manufacturing

Thomas Schmitta,b,*, Ronal Bejarano c,d, Carla Assuad e

aScania CV AB, Smart Factory Lab, Verkstadsvägen 17, 15138 Södertälje, Sweden
bUppsala University, Department of Civil and Industrial Engineering, Lägerhyddsvägen 1, 752 37 Uppsala, Sweden
Wärtsilä Oy, Manufacturing Accelerator Center, Frilundintie 7, 65170 Vaasa, Finland
cWärtsilä Oy, Manufacturing Accelerator Center, Frilundintie 7, 65170 Vaasa, Finland
dAalto University, Maurinitie 8, 02150 Espoo, Finland
Norwegian University of Science and Technology, Teknologivg 21, 2815 Gjøvik, Norway

Abstract
Energy efficiency analyses and life cycle assessments are two examples of data-driven applications focused on environmental sustainability in the industrial manufacturing context. Both use methodologies based on aggregating operational data, extracted from multiple data sources along a value chain. However, the possibility to source, utilize, and share data is often obstructed by heterogeneous or non-transparent data operations across organizations, as well as poor availability of digital interfaces for data collection. In practice, the underlying application processes lack fidelity and granularity, as a trade-off to simplify the modeling and data collection tasks. Therefore, new opportunities arise for automated workflows on data collection (data pipelines) as a relevant component of every digital transformation strategy. This study reflects on the challenges experienced by two industrial actors collecting operational data for an energy efficiency analysis and a Life-cycle Assessment (LCA), each with a different manufacturing context and architectural approach. Finally, it presents the opportunities for new technologies at affordable costs that potentially ease the development and operation of data pipelines to solve such challenges.

© 2024 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the 31st CIRP Conference on Life Cycle Engineering (LCE 2024)

Keywords: energy efficiency; life cycle analysis; digitalization; data pipelines; environmental sustainability

1. Introduction
Environmental regulation motivated by the sustainability crisis and economic competitive advantages are drivers towards more energy and resource conservation in industry. For example, environmental documentation such as Digital Product Passport, Environmental Product Declaration, ISO 14001, and Eco-Management and Audit Scheme (EMAS) are becoming more relevant and demanded by consumers of certain products [1]. The success of these initiatives, aiming to reduce environmental impact and enhance energy and resource efficiency throughout the product life cycle (Ibid., p. 98), hinges on efficient information flow horizontally across supply chains and vertically across organizational levels within the companies.

* Corresponding author. Tel.: +46-85535-1291 E-mail address: thomas.schmitt@scania.com

Shifting towards environmentally sustainable manufacturing requires Environmental Impact Analyses (EIA), often employing Life Cycle Assessment (LCA) to evaluate the environmental impact throughout a product or process life cycle [2, 3]. LCA includes metrics such as carbon footprint and energy consumption [4, p.62]. Energy efficiency (EE) measures, a key aspect of EIAs, focus on reducing energy consumption while maintaining or increasing output (Ibid.).

However, collecting a data inventory of the environmental impacts of a product’s life cycle is a time-consuming and rigid process [5]. To facilitate EIAs, various studies have proposed methodologies and frameworks aimed at automating data collection processes and reducing uncertainties, utilizing digital technologies such as the industrial internet of things (IIoT) [6, 7]. Despite these advancements, a significant scarcity of practical, cost-effective data architecture solutions is observed when it comes to addressing this matter within large-scale manufacturing companies. This highlights a gap between the aca-
demic literature and the actual utilization of Industry 4.0 technologies for conducting EIAs in practice.

This paper explores this gap by investigating and comparing the experience when collecting data for an emission calculation based on LCA and an energy efficiency analysis in two industrial actors. Besides, the paper presents opportunities for new technologies at affordable costs that potentially ease the development and operation of data pipelines to solve such challenges.

2. Literature Background

Digital transformation creates opportunities for organizations to utilize data from heterogeneous sources for multiple purposes, including EIAs such as LCA and EE analysis [4]. However, a structured approach to working with manufacturing and environmental data is crucial to guarantee reliable information for decision-makers. This necessitates the establishment of reliable operations for data collection, integration, storage, analysis, and visualization [8]. With the increasing use of sensors in manufacturing processes, standardized and automated collection of data on output, time, and resources spent becomes feasible, forming the foundation for more accurate data analysis for EIAs and efficiency improvements. This is made possible through process integration and automated data flows, referred to as data pipelines by the authors in this paper.

2.1. Life Cycle Assessment (LCA)

By definition, a LCA is a methodological framework for estimating and assessing environmental impacts attributable to the life cycle of a product. The contents presented in this subsection are based mainly on the work published by Rebitzer and Pennington et al. in [9, 10]. Usually, a LCA provides an estimation of emissions and waste per functional unit within a defined boundary of the life cycle, being the results from a derived unit, for instance m³ of engine block cast packed or an unit of a product type "X". The functional unit can be used as a base unit, to compare emissions and waste across variants in products or processes. As the life cycle consists of a chain of processes, a product system is defined which is made up of individual unit processes representing the activities of interest (e.g. resource extraction, manufacturing, transportation).

LCA consists of three phases. Firstly, the goal and scope definition describe the product system, delineating system boundaries and establishing the functional unit. The second phase involves the Life Cycle Inventory (LCI), where data from the product system is compiled about environmental exchanges and energy flows. Finally, the Life Cycle Impact Assessment (LCIA) calculates and interprets indicators of the potential impact on the natural environment, considering specific categories and additional contextual information.

There are two common goals of a LCA. First, to describe a product system and its environmental exchanges, also known as attributional LCA. Second, to assess the impact of changes in the system on its environmental exchanges, known as consequential LCA. The results of an attributional LCA are mostly used for comparing processes and products, while a consequential LCA informs the impact of introducing new processes or technologies.

2.2. Energy Efficiency Analysis

Energy efficiency optimization involves improving the ratio of energy used to the output of a system [11, p.30]. EE is one of the goals directly connected to environmental exchanges in LCA. An energy management systems is seen as a key enabler in both industry and academia to systematically improve EE [12]. The industry standard ISO 50001 specifies an energy management system based on Plan-Do-Check-Act cycle, which includes the identification of the main energy users, setting goals and actions for energy reductions, and monitoring their effect (Ibid.).

One challenge in energy-saving activities in manufacturing is the lack of knowledge regarding existing energy waste in the system. Literature also cites additional obstacles, such as a lack of awareness of energy-efficient measures on the shop floor and a low prioritization of energy in general [13, p.36]. Therefore, it is crucial to monitor energy and resource flows to establish a baseline or "normal situation." The level of monitoring detail needs to be decided based on the application goals; for example, the sampling rate relevant for billing purposes is 15 minutes (in Germany), or the correlation of energy consumption to different machine states for discrete-event simulation [11, p.138].

Data is generated by sensors attached to the unit of interest, typically through an energy management system or the existing data communication interface of the machine [12]. Subsequently, energy consumption can be aggregated across organizational levels, such as unit, line, and factory, with the application of different energy efficiency strategies [14]. Figure 1 illustrates environmental and data exchanges across organizational levels, along with their analysis and application.

2.3. Challenges on conducting LCA and EE

Implementing an LCA and EE with reliable results depends on ensuring the accuracy of the product system modeling and guaranteeing well-defined process units, aligned with the general goal. However, challenges around data collection techniques have been reported by academics and practitioners during the last two decades. One challenge is the elaboration of the LCI, including monitoring energy flows. This task uses a data set populated by compiling a selection of environmental exchanges (inputs and outputs). Often these tasks are identified as the most time-consuming steps for several reasons. Some of such reasons presented by [9, 11] are:

- Lack of multidisciplinary competence and cooperation needed when dealing with the complexity of mapping energy and materials flow
- Lack of adequate software tools and measurement points
- Trade-offs are necessary and can lead to less rigorous practices
Fig. 1. A descriptive diagram of Life Cycle Assessment (LCA) components and terminology, partially adapted from [11, p. 11]. Initially, a functional unit (product) and a product system (in this case, the manufacturing firm’s factory, which can be broken down into multiple sub-processes) are determined. Subsequently, the requirements of the EIA and the data sources are defined. These requirements influence decisions regarding abstractions, simplifications, and approximations. The data is then input into the applications.

- Analysis results come too late for timely action
- Difficulties on keeping consistent nomenclature on data
- Heterogeneous landscape of strategies for EIAs
- Absence of perceived need or priority

The complexity, duration, and low prioritization of data collection processes for EIAs have resulted in a gap between available and expected data inputs. Consequently, industry has resorted to adopting simplifications in methodologies. These simplifications often hold estimations that prove ineffective for meaningful environmental impact analysis. Moreover, the outlined challenges discourage not only SMEs and start-ups, but also larger organizations from applying EIAs beyond mere compliance reasons.

3. Research methodology

This work comprises a qualitative case study of two innovation work groups, each in a different multinational corporation part of a cross industry collaboration network in the Nordics [15]. The groups are responsible for investigating digital technologies and implementing proof of concepts. Recently, a branch within the groups started working to adopt digital technologies for EIAs. Each study case will be described briefly and connected to a specific EIA, EE analysis in 3.1 and LCA in 3.2. This study aims to analyze the current challenges on implementing EIAs, and discuss opportunities to ease the implementation of EIAs by automating the data collection process. The research questions of this paper are:

1. What operation and information technologies (OT - IT) are used today for environmental analysis in the study cases?
2. What are the common challenges in using data flow automation for EIAs (EE + LCA)?
3. What are the mechanisms that enable data flow automation for effective EIAs?

3.1. Study case A: Automotive manufacturing

Situated within the automotive sector, today only 6% of the company’s emissions occur within production, whereas the largest share is contributed by vehicles on the road. With the transition from internal combustion engines to battery-electric vehicles, and an improvement in the electricity mix, this number is expected to increase to 85% by 2030 [16, p. 17]. The case company established a laboratory with the purpose to test new digital technologies for production operations. Additionally, a central group within the company is responsible for translating high level organisational goal on energy consumption onto department goals, and assist in rolling out energy initiatives. One such initiative is developing an energy management system, to enable data sharing and scaling energy dashboards across different production units. For this, digital tools are tested to achieve a more detailed and automated energy efficiency assessment of production processes.

3.2. Study case B: Power unit manufacturing

In this case, a company involved in the manufacturing of power units for marine and energy applications leads an inno-
The initiative targets to accelerate the adoption of digital transformation and environmental sustainability practices in engineering and manufacturing of industrial equipment across value chains in the mechanical engineering sector. The initiative of interest for the scope of this paper is the creation of a digital tool to support de-carbonization and environmental sustainability by providing inputs for a partial attributional LCI over the cradle to gate scope (manufacturing) [17]. This case in particular addresses concerns about emissions connected to the sourcing of raw materials, since every product unit has a total mass of more than 20 metric tons.

3.3. Research design

The data collection process involved engaging in discussions with environmental sustainability experts from both companies, analyzing company documents, and conducting proof-of-concept studies within the respective innovation workgroups. A digital transformation framework was used to map the current and future practices and infrastructure for implementing the EIAs in both study cases. The authors considered the Smart CE Framework presented by Kristoffersen et al. in [18] as a comprehensive example built upon relevant references.

4. Results

4.1. Mapping current tools for EIAs

For the energy efficiency analysis in study case A, the energy group maintains an energy management system with different types of metering devices to collect data on electrical power, compressed air, gas, heating (and water). The goal is to break down the energy used per produced unit. Across its three production locations in Sweden, the company maintains around 600 meters, with the main site alone housing 350 electricity meters (280 within industrial operations). Most of the electricity meters are attached to the transformers, which typically supply electricity to multiple production lines including both production and auxiliary equipment like elevators and lights. The meters automatically forward an hourly consumption value to a SQL database, integrated with a billing and reporting tool. At the end of the month, staff query and compile the data into a digital spreadsheet, aggregating monthly consumption values for attribution to respective departments. The monthly values serve as indicators to the production units, highlighting consumption trends and allowing the identification of significant deviations. Staff faced challenges with the old energy management system, including installation and maintenance costs, scalability, and stability. To address these issues, the company invested in a new energy management infrastructure using LoRa network and modular software tools (see Figure 2). Ten employees from seven internal departments, external firms for component installation, and consultants for server setup were involved. The electricity meter’s data collection frequency increased from one hour to five minutes, to enhance stability and reduce the risk of missing data. This totals 36.8 million sensor readings from electricity meters annually.

In study case B, data collection relies on a digital spreadsheet form template, often designed with assistance from a consultant company guiding the entire data collection process. The primary data source for in-house produced parts is the Product Data Management (PDM) team, which maintains an information system (SQL database) and a reporting website. PDM data originates from engineering design documentation (R&D). External parts are sourced from suppliers, who complete and submit input forms via email exchanges, as other digital interfaces are not available. Given that each product can consist of numerous components, the strategy is to concentrate on components with the highest mass or economic value until the collected inputs represent approximately 70% of the total mass or value of the final product. The dataset includes 21 components. According to personnel, the entire data collection, processing, and analysis for a product take approximately 7 months to 1 year. In some instances, not all values were available or had legible records, leading to trade-offs and other practices described in the sources from Section 2, summarized in the following section.

4.2. Challenges in utilizing data pipelines

Both case studies encountered similar challenges regarding the use of reliable data for EIAs. A list of the most common challenges are listed below:

1: Lack of multidisciplinary competence and cooperation. To define standards and implement an effective data collection process, it requires collaboration between operators, production engineers, IT - OT experts, innovation teams, and environmental teams. Disjoint efforts could lead to the development of ineffective tools and sub-optimizing a process.

2: Existing metering devices are insufficient. Collected environmental data, including energy data, lacks the granularity and sample rate from control (SCADA) or execution (MES) process data, which provides real-time performance indicators for each machine. The data permits only high-level analysis of multiple consumers, making it challenging to track anything beyond significant deviations, thereby making continuous environmental improvement difficult.

3: Trade-offs are necessary and can lead to less rigorous practices. Given that it is not feasible to collect all environmentally relevant data, methodological choices must be made regarding the allocation of inputs per functional unit. This impacts the accuracy of data inputs for EIAs. For large companies with a complex product portfolio, deciding on how to simplify, what factors are important, and how to balance environmental aspects against each other is a challenge.

4: Lack of adequate and interconnected software tools. Despite the support provided by software tools, the current setup still demands manual extraction and migration between databases and spreadsheets, making data logging unreliable and non-transparent. In one of the study case, the lead time for the last project that involved internal data collection was three months, resulting in delays in feedback loops. This causes that
Fig. 2. The present status of the study cases mapped to the CE framework presented in [18]. The current EIA tools of both companies are highlighted in the red box, and the planned data pipelines in EIA tools are depicted within the dashed line.

detailed EIAs are sometimes perceived as "beyond the reach of potential users," particularly when fast decision-making is necessary.

5: Uncertainty in data and terminology. This arises from variability, poor specification, practical errors (e.g., typos), incomplete reports, and the use of rounded-off values. These factors lead to discrepancies between the product system and the real system. For instance, achieving consistency in nomenclature and measuring units across the entire data collection stack (factory, model, database, software) is challenging. In one study case, it was found that from 27 possible classifications of metal materials from the LCA database, none matched the 14 possible values in the material description from the bill of materials.

4.3. Opportunities for utilizing data pipelines

The opportunities are framed as new data pipelines extending from data collection through data integration to data analysis. Figure 2 presents an updated architecture (box outlined in red dashed lines), incorporating distributed computing. The potential opportunities arising from a new toolset utilizing data pipeline technologies are:

1: Smart low-cost sensors and controllers to enable connection to OT/IT gateways or edge servers and facilitate automatic data collection. The goal is to minimize or eliminate the reliance on spreadsheets and manual inputs. Companies should establish general specifications for metering and data collection infrastructure.

2: OT/IT integration and competence. In-house OT/IT expertise and open-source tools can address digital interface needs, facilitating the creation of software capable of automatically transmitting data to secure storage and using it further. This involves developing software adapters, connectors, and edge modules based on industrial communication standards such as OPC-UA and Sparkplug. Distributed computing using single-board computers like Raspberry Pi, along with attachments (HATs), embedded boards, or small form factor computers, optimizes infrastructure and facilitates the translation of sensor data between (legacy) protocols.

3: Visualization and analysis tools allow to merge and analyze large data sets. Dashboards accessible to the production personnel can support the inclusion of environmental aspects into decision-making. Tools such as Grafana and Power BI reduce time to access data and build dashboards, allowing for timely and autonomous decision making. It requires the production units to gain competence in data querying and visualization.

4: Environmental data requirements throughout the procurement and installation phases of products and equipment can enhance the effectiveness of EIAs. This demands standards for the procurement process and user-friendly digital interfaces capable of collecting, e.g., emissions data from component sourcing. These standards are defined by various teams, including environmental, automation, and production teams. Decisions regarding simplifications and assumptions are made.
based on several factors, including the desired level of detail, the acceptable level of uncertainty and the available resources. It is essential to consider the applications that process the data, to ensure that the collected information aligns with the overall goals and standards set for the EIA.

5. Discussion

Data pipeline technologies in EIA tools are vital for meeting regulations and enhancing resource and energy efficiency for competitive advantage. Accurate information is crucial for EIAs, and effective data pipelines play a key role. While numerous approaches exist [6, 7] standards and industrial practices are not fully developed from an IT perspective.

Digitalization has dismantled previous information silos, eliminating the necessity for simplifications and shortcuts in environmental analyses. EIAs, varying in scope, share a common facilitator—the use of data pipelines, serving multiple purposes (Figure 2). The common goal between both use cases is transitioning from purely attributional or descriptive EIAs to a highly dynamic consequential EIA integrated into day-to-day operations by building data pipelines with specifications defined with expertise from various domains.

The literature showcases diverse approaches to data pipeline infrastructure, with large companies often engaging in uncoordinated efforts to establish a unified data architecture. This is attributed, in part, to the proliferation of communication protocols like MQTT and LoRa in smart sensors. The continuous evolution of software and hardware further complicates the quest for a standardized setup, necessitating modular architectures to avoid vendor lock-in. The strategy adopted by the two cases, as illustrated in Figure 2, involves digital tools serving specific purposes in the data pipeline infrastructure. This modular approach allows for the replacement of tools. Tools bridging and analyzing process and environmental data enable companies to set relative consumption goals, avoiding reliance on guesswork for absolute saving goals. However, these applications may not yield direct profits or are challenging to quantify in financial terms. Without regulatory mandates, justifying their implementation to the business remains a significant challenge.

6. Conclusion

This study highlights the current use of digital technologies for EIAs in the industry. Comparing the technological approaches of two industrial players can assist in developing standardized data pipelines for EIAs. Both cases experience the enduring challenges described by academia and industry over the last two decades. An opportunity to enhance data collection is introduced by proposing data operations automation through cutting-edge technologies such as IIoT and distributed computing (Edge and Cloud resources), all encompassed under the “data pipelines” umbrella.

Envisioned as part of daily operations, data flows for EIAs are anticipated to facilitate continuous data exchange and processing, supporting decision-making. The expectation is that EIAs will evolve into live applications, consuming data generated by each unit produced, as opposed to static reports executed only once for a product type.

New discussion topics are expected to arise about whether there is merit to invest on improving the level of automation on data collection for EIAs, especially given the current short lived life cycles of certain products. Measuring the impact on EIAs given high quality data input, or comparing the impact of decisions made while ignoring factors that are too hard to integrate into the data operations infrastructure, are remaining open questions for the scientific community.

References