MAY 2024

BRIGHTWOLVES

Data Collection in Life Cycle Assessments

Whitepaper

WRITTEN BY: ARTHUR VAN DORPE

WWW.BRIGHTWOLVES.COM

CONTENTS

Foreword

Welcome to our LCA Whitepaper Series, each dedicated to a specific aspect of Life Cycle Assessment (LCA). As a critical tool in the field of sustainability, LCA provides a comprehensive view of the environmental impacts associated with all the stages of a product's life, from raw material extraction through materials processing, manufacture, distribution, use, repair and maintenance, and disposal or recycling. This series aims to delve into the intricate aspects of LCA, shedding light on the various methodologies, standards, and frameworks that guide its application.

In each whitepaper, we will focus on a specific topic, be it ISO standards that govern LCA, industry-specific standards such as Global Feed LCA Institute (GFLI), various allocation methods, ReCiPe method vs. Product Environmental Footprint (PEF), and many more. Our goal is to provide a clear, comprehensive, and accessible understanding of these complex topics, enabling you to apply this knowledge in your sustainability journey.

Whether you are a seasoned professional in the field of sustainability or a newcomer looking to understand the intricacies of LCA, these whitepapers will serve as a valuable resource. We invite you to join us in this exploration of LCA, as we strive to contribute to a more sustainable future.

LCA Data Collection: Primary versus Secondary Data

Introduction

Before diving into more practical topics in this series around life cycle assessments, we wanted to cover one more theoretical part: the importance of and the difference between primary and secondary data. As the foundation of any robust LCA lies in the meticulous collection and analysis of data, this paper delves into the intricacies of data collection within the framework of LCAs. More importantly, we will examine the underlying quality of an LCA, which ultimately shapes the reliability of the assessment's conclusions.

To navigate the complexities of data collection, this paper will dissect the dichotomy between primary data and secondary data—two pivotal types of information that feed into the LCA process. Primary data, characterized by its specificity and direct measurement, offers a granular view of the product system, often at the expense of greater resource investment. Conversely, secondary data, sourced from existing databases and studies, provides broader accessibility and efficiency, albeit with potential trade-offs in precision and context specificity.

Data Collection in Life Cycle Assessments

Data collection is the cornerstone of LCAs, providing the empirical foundation upon which the entire analysis is constructed. The International Organization for Standardization (ISO) outlines the principles and framework for LCAs under ISO 14040 and 14044, emphasizing the critical role of data collection in achieving a reliable and comprehensive assessment. The data collected informs the Life Cycle Inventory (LCI), which is the second phase of an LCA after the goal and scope definition.

The LCI is an exhaustive compilation of the energy and material inputs, as well as product and environmental outputs, associated with a product system throughout its life cycle. The LCI phase involves two primary components: the compilation of data and the calculation of life cycle inventory results. The data encompasses every relevant input and output, quantifying the flows from and to nature, such as raw materials, energy consumption, emissions to air, water, and soil, and waste generation.

Data collection in LCAs is a systematic process that begins with the definition of the goal and scope. This initial phase sets the boundaries of the study, determining what will be included in the LCA and to what extent. The system boundaries delineate the unit processes that will be part of the LCA, which can range from "cradle-to-grave" or be limited to a specific life cycle stage.

Once the boundaries are established, the data collection process can begin. This involves identifying and quantifying all relevant inputs and outputs within the system boundaries. Inputs include resources such as raw materials, water, and energy, while outputs encompass products, co-products, waste flows and emissions. Background processes, which support the system but are not directly involved in the product life cycle, such as transportation and energy production, are also considered.

Documenting the data collection methods, assumptions, and sources is a critical step in ensuring the LCA's transparency and reproducibility. This documentation includes keeping track of the units of measurement, conversion factors, and allocation methods used to handle multi-output processes. The ISO standards do not prescribe specific data collection methods, but they do require that the methods be consistent with the goal and scope of the study.
Verification of the data is equally important to ensure its accuracy, completeness, and consistency. This may involve cross-referencing data sources, conducting expert reviews, and identifying any gaps or uncertainties in the data. The verification process helps to build confidence in the LCA results and supports the credibility of the conclusions drawn from the study.

The importance of rigorous data collection in LCAs cannot be overstated. The quality of the LCA results is directly tied to the quality of the data collected. Inaccurate or incomplete data can lead to misleading conclusions, which can have significant implications for decision-making processes. Therefore, a well-defined data collection strategy, aligned with ISO standards, is essential for conducting LCAs that are robust, reliable, and capable of informing sustainable practices.

In summary, data collection in LCAs is a meticulous and structured process that is vital to the integrity of the assessment. By adhering to ISO standards and best practices, LCA practitioners can ensure that their studies accurately reflect the environmental impacts of the product systems under examination and provide valuable insights for sustainability.

Primary Data

Primary data plays an indispensable role in the foreground system of LCAs, offering a level of specificity and detail that is unmatched by other data types. This section will explore the definition of primary data, its collection methods, the inherent challenges and limitations, and the benefits it brings to the credibility and specificity of LCAs.

Primary data refers to information that is directly collected from specific processes or activities within the product life cycle. It is raw, original data gathered firsthand by the LCA practitioner or the organization conducting the assessment. In the context of LCAs, primary data is used to model the foreground system, which includes all processes that are directly controlled or influenced by the organization responsible for the product. This can range from the extraction of raw materials to manufacturing processes and product use.

The collection of primary data is typically conducted through on-site measurements and observations. This involves directly measuring inputs and outputs of a process, such as energy consumption, material usage, and emissions. On-site data collection ensures that the information is as accurate and relevant to the specific system as possible. Statistical sampling is another method used to collect primary data. This approach involves selecting a representative subset of data from a larger population, which can then be used to draw conclusions about the whole system. Statistical sampling is particularly useful when it is impractical or too resource-intensive to measure every single instance within a process.

Incorporating stakeholder engagement into the data collection process is essential for ensuring the comprehensiveness and accuracy of the data. Engaging with suppliers provides access to specific information about the origins and production methods of materials, as well as the logistics involved in the supply chain, all of which are critical for a thorough lifecycle analysis. Similarly, dialogue with customers can yield valuable insights into the actual use and eventual disposal of products, which can influence the assessment of the product's overall environmental impact. By involving stakeholders, companies can also foster transparency and build trust, which can lead to more collaborative and sustainable practices. Moreover, stakeholders can often contribute experiential knowledge and contextual information that might not be captured through quantitative methods alone, thereby enriching the data collection process and enhancing the validity of the findings.

Advanced techniques such as electronic data collection sheets and tools like Microsoft Excel or specialized LCA software are often employed to record and analyze the data. These tools facilitate the organization, calculation, and interpretation of large datasets, making it easier to manage the complexity of the information gathered. Despite its advantages, collecting primary data is not without challenges. One significant limitation is the resource intensity of the process. It can be time-consuming and costly to gather primary data, especially for complex systems or processes that are geographically dispersed. Access to data can also be a barrier. Organizations may face difficulties obtaining primary data due to proprietary concerns, confidentiality agreements, or logistical obstacles. Additionally, the quality of primary data is only as good as the measurement devices and methodologies used, which can vary in their precision and reliability.

The use of primary data in LCAs offers several benefits. It provides a high level of specificity, allowing for a more accurate representation of the product system. This specificity can lead to more reliable and authentic insights into the environmental impacts of the product life cycle. Primary data enhances the credibility of an LCA, as it reflects actual measurements rather than estimates or averages. This can be particularly important for organizations looking to make strong, evidence-based claims about their products or processes. Furthermore, primary data gives companies ownership over their LCAs, enabling them to tailor the assessment to their specific needs and objectives. It allows for a deeper understanding of the environmental hotspots within their operations and can inform targeted improvements.

Secondary Data

Secondary data constitutes an integral component of LCAs, particularly within the background system. This section will define secondary data, elucidate its role in LCAs, describe the sources from which it is derived, discuss the trade-offs associated with its use, and examine the methods for verifying its relevance and accuracy.

Secondary data is information that has been previously collected and is available for use in new research contexts. In the realm of LCAs, secondary data pertains to information that is not directly collected from the product's life cycle processes but is obtained from existing databases, scientific literature, and previous studies. It is typically used to model the background system, which includes processes that are not directly controlled by the organization, such as the production of raw materials, energy generation, and waste treatment services.

Secondary data is sourced from a variety of repositories and databases that compile life cycle inventory (LCI) data. These databases, such as EcoInvent, provide aggregated data on environmental impacts associated with materials, energy, and processes based on a wide range of studies. Scientific literature, industry reports, and government publications are also rich sources of secondary data, offering insights into average values and standard practices within industries and regions. We provide an overview of the importance of LCA databases and some examples in the figure below.

The use of secondary data in LCAs presents several trade-offs that must be carefully considered. One of the primary advantages of secondary data is its accessibility. It allows for the rapid creation of LCAs, as the data is readily available and does not require the time and resources associated with primary data collection. This can be particularly beneficial for small organizations or for preliminary assessments where time and cost constraints are significant. However, the specificity and accuracy of secondary data may not match that of primary data. Secondary data often represents average conditions or standard processes that may not precisely reflect the unique attributes of the specific product system under study. This can result in a less accurate portrayal of the actual environmental impacts, potentially leading to generic insights that are not entirely aligned with reality.

Ensuring the relevance of secondary data is crucial for the integrity of an LCA. Practitioners must assess whether the data is appropriate for the geographic, technological, and temporal scope of their study. For instance, data from a different country or a slightly different process may require adjustments or may not be suitable at all. Verification methods for secondary data include benchmarking against other similar studies or databases to ensure consistency. Cross-referencing with industry experts or academic research can also provide validation for the data used. When primary data is available, it can serve as a benchmark to validate the secondary data, offering a comparison to ensure that the secondary data is within a reasonable range of accuracy.

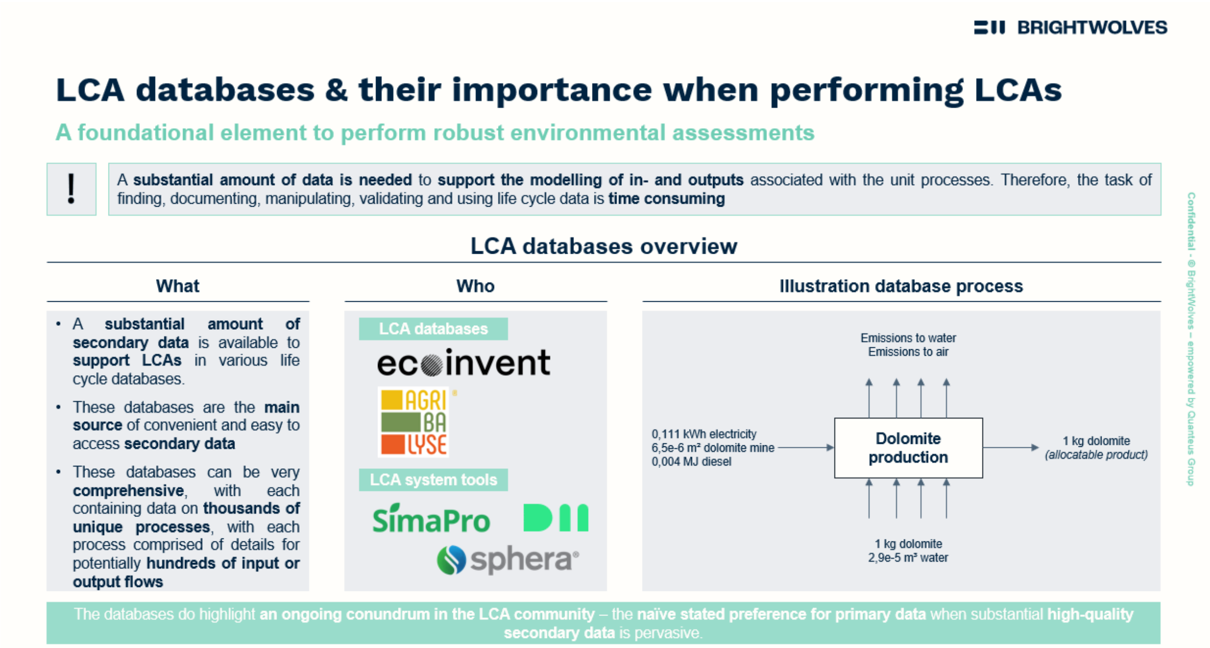

Figure 1: Secondary Data - LCA Databases

Comparison of Primary and Secondary Data

The selection of data types in LCAs is a critical decision that influences the accuracy, credibility, and applicability of the study's outcomes. Primary and secondary data each have distinct characteristics that can affect the LCA in different ways. This section will compare these data types in terms of accessibility, credibility, and representativeness, and discuss their implications on the speed and authenticity of LCAs. A detailed example will illustrate the practical differences between primary and secondary data in a real-world context. Finally, we provide an illustration that compares both data types on a high-level basis.

Primary data is often less accessible than secondary data due to the resources required for its collection. Gathering primary data can be time-consuming and may involve significant financial costs, specialized equipment, and access to proprietary processes. In contrast, secondary data is generally more accessible as it is already compiled and available in databases or literature, allowing for quicker and more cost-effective LCA development.

Credibility is a measure of the trustworthiness of the data used in an LCA. Primary data is typically considered more credible because it is collected specifically for the purpose of the study, ensuring that it closely reflects the actual processes and systems being assessed. Secondary data, while useful, may be perceived as less credible due to its generic nature and the potential for it to be outdated or not entirely applicable to the specific context of the LCA.

Representativeness refers to how well the data reflects the true nature of the system being studied. Primary data provides a high degree of representativeness, as it is process-specific and tailored to the product's life cycle under investigation. Secondary data, however, may lack this level of representativeness, as it is often derived from average conditions or standard processes that may not align perfectly with the unique attributes of the system in question.

The use of primary or secondary data can significantly impact the speed at which an LCA can be conducted. Secondary data allows for a faster assessment process since it is readily available and can be quickly integrated into the LCA model. However, this speed may come at the cost of authenticity, as the data may not accurately represent the specific system being studied. Primary data, while slower to collect, contributes to a more authentic LCA, providing a genuine reflection of the environmental impacts associated with the product's life cycle.

Consider the example of baking a vanilla cake, using either primary data or secondary data. In the context of the recipe, primary data offers you detailed ingredients and their quantities, relevant equipment and clear instructions based on multiple measurements, observations and documentation that is collected firsthand by the practitioner. Therefore, the primary data offers the advantage of accuracy, credibility, and a significant amount of representativeness for the individual baking experience that can be used for accurate replication and analysis. However, the collection of primary data is time-consuming and resource-intensive, as it requires direct measurement and observation, and may be less accessible due to the need for specific equipment, such as a digital kitchen scale or an energy meter. Conversely, secondary data provides a more accessible and convenient alternative, drawing on a wealth of existing information from cookbooks, culinary websites, and expert advice. It is easy to use and can be rapidly sourced, saving time and effort in the initial stages of recipe development or research. Secondary data offers a broad overview and can be beneficial for gaining quick insights or understanding standard practices in cake baking. However, it may lack the specificity and context of primary data, potentially leading to less precision and a greater risk of relying on outdated or non-specific information that may not perfectly align with the unique conditions of the current baking scenario.

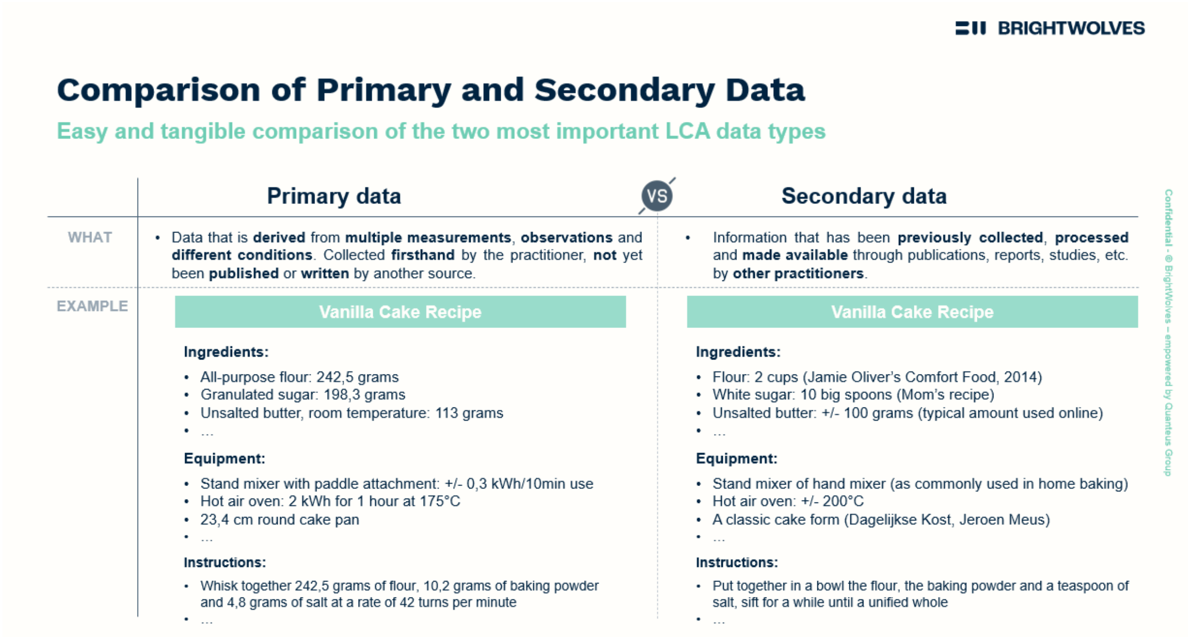

Figure 2: Easy and Tangible Comparison of Primary and Secondary Data

Data Quality in LCAs

Primary and secondary data play a pivotal role in the final quality of your data, and thus the final result of your assessment. Data quality is a multifaceted concept in LCAs that encompasses the characteristics of data which relate to their ability to satisfy stated requirements. According to ISO standards, particularly ISO 14044:2006, data quality is integral to the credibility and validity of an LCA. This section will discuss the importance of data quality, outline the requirements and indicators, explain Data Quality Goals (DQGs), describe the Data Quality Pedigree Matrix, and explore the impact of data quality on the final impact assessment.

Importance of Data Quality According to ISO Standards

The International Organization for Standardization (ISO) emphasizes the importance of data quality in LCAs, recognizing that the reliability of an LCA's conclusions is directly linked to the quality of the underlying data. ISO 14044:2006 provides guidance on LCA methodology, including the need for clear statements about data quality requirements such as age, geographical reach, completeness, and sources.

Data Quality Requirements and Indicators

Data quality requirements are predefined criteria that characterize the fundamental expectations of the data used in an LCA. These requirements are considered during the goal and scope definition phase and are essential for ensuring that the data collected will meet the needs of the study. ISO identifies ten key categories for addressing data quality:

1. Time-related coverage.
2. Geographical coverage
3. Technology coverage
4. Precision
5. Completeness
6. Representativeness
7. Consistency
8. Reproducibility
9. Sources of the data
10. Uncertainty of the information

Data quality indicators are summary metrics used to assess these requirements. They provide a means to evaluate the data against the established criteria, ensuring that it is fit for the intended use.

Data Quality Goals (DQGs)

DQGs are defined during the goal and scope phase of the LCA and are used to describe the ideal representativeness and process completeness for the project. They are specific to the study and may include temporal, geographical, and technological aspects. For example, a temporal DQG would specify the age of the data and the time period it should cover, while a geographical DQG would define the area from which data should be collected. Establishing DQGs is critical for aligning the data collection process with the study's objectives.

Data Quality Pedigree Matrix

The Data Quality Pedigree Matrix, or ISO scale, is a tool used to assess and communicate the quality of data in an LCA. It rates data quality on a scale from 1 (very good) to 5 (very bad), with specific criteria for each rating. For example, a score of 1 for precision indicates highly reliable data, while a score of 5 would suggest significant uncertainty. The matrix evaluates aspects such as flow reliability, representativeness, and the robustness of data collection methods.

Impact of Data Quality on the Final Impact Assessment

The quality of data used in an LCA has a profound impact on the final impact assessment. High-quality data leads to more accurate and reliable results, which in turn can influence decision-making and stakeholder confidence. Conversely, poor data quality can result in misleading conclusions, potentially harming the credibility of the study and the organization conducting it.

To validate data, practitioners may compare primary data with similar secondary data or validate it over time. Secondary data can be benchmarked against other reports or cross-referenced by experts. These validation steps are crucial for ensuring that the data meets the quality requirements and serves the goals of the LCA effectively.

Key takeaways

- Rigorous data collection is the foundation of LCAs, essential for ensuring the accuracy and integrity of the assessment's results. Adherence to ISO standards, meticulous documentation, and thorough verification are critical to producing reliable LCAs that can improve your credibility and decision-making.
- Primary data is crucial for accurately modeling the foreground system in LCAs, providing unmatched specificity, and enhancing the credibility of the assessment. Despite the challenges of resource intensity and data accessibility, the use of primary data allows for precise measurement of environmental impacts and supports evidence-based claims, enabling organizations to make informed decisions for targeted sustainability improvements.
- Secondary data is a vital element for modeling the background system in LCAs, offering the advantage of accessibility and efficiency. However, it may lack the specificity of primary data and requires careful verification to ensure its relevance and accuracy for the LCA's scope. Practitioners must judiciously assess and adjust secondary data to accurately reflect the environmental impacts of the product system being studied.
- The quality of both primary and secondary data is critical to the validity and credibility of Life Cycle Assessments (LCAs), as outlined by ISO 14044:2006. Adherence to data quality requirements and indicators, the establishment of Data Quality Goals (DQGs), and the use of the Data Quality Pedigree Matrix are essential practices for ensuring that the data used in LCAs meets the necessary standards, leading to accurate and reliable impact assessments that can confidently inform decision-making.

Are you inspired by the topic of LCAs or curious to know more about Digit Mint and its LCA tool? Stay tuned for our next paper of this LCA Whitepaper Series or do not hesitate to contact **Peter-Jan Roose** or **Vincent Govaers**!

References

- Data Collection:

Matthews, H. & Hendrickson, Chris & Matthews, Deanna. (2014). Life Cycle Assessment: Quantitative Approaches for Decisions that Matter.

BrightWolves' LCA Academy

https://www.linkedin.com/advice/1/what-best-methods-collecting-data-life-3vj0c

- Primary and Secondary Data:

Matthews, H. & Hendrickson, Chris & Matthews, Deanna. (2014). Life Cycle Assessment: Quantitative Approaches for Decisions that Matter.

https://ecochain.com/blog/primary-vs-secondary-data-in-lca/

BrightWolves' LCA Academy

https://www.iso.org/obp/ui/en/#iso:std:iso:14040:ed-2:v1:en

- Data Quality:

Edelen, A. AND W. Ingwersen. Guidance on Data Quality Assessment for Life Cycle Inventory Data. U.S. Environmental Protection Agency, Washington, DC, EPA/600/R-16/096, 2016.

https://www.apec.org/docs/default-source/Publications/2004/2/Life-Cycle-Assessment-Best-Practices-of-International-Organization-for-Standardization-ISO-14040-Ser/04_cti_scsc_lca_rev.pdf

BRIGHTWOLVES